



Master Thesis

The lepton+jets Selection and Determination of the Lepton Fake Rate with the Full RunIIb Data Set

Author(s):

Meister, Daniel

Publication Date:

2013

Permanent Link:

<https://doi.org/10.3929/ethz-a-009789965> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

The lepton+jets Selection and Determination of the Lepton Fake Rate with the Full RunIIb Data Set

Master Thesis
at the Swiss Federal Institute of Technology Zurich (ETH)
and the University of Illinois at Chicago (UIC)

Daniel Meister <dmeister@student.ethz.ch>

supervised by

Prof. Cecilia E. Gerber (UIC), Prof. Günther Dissertori (ETH), Dr. Andreas W. Jung (Fermilab)

January 13th, 2012

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

UIC

Abstract This thesis presents the combined single top and $t\bar{t}$ lepton+jets selection for the full RunIIb dataset of the DØ detector. The selection uses the newest software versions including all standard central object identifications and corrections and has various additions and improvements compared to the previous 7.3 fb^{-1} $t\bar{t}$ selection and the previous single top selection in order to accommodate even more different analyses.

The lepton fake rate ϵ_{QCD} and the real lepton efficiency ϵ_{sig} are estimated using the matrix method and different variations are considered in order to determine the systematic errors. The calculation has to be done for each run period and every set of analysis cuts separately.

In addition the values for the exclusive jet bins and for the new single top analysis cuts have been derived and the thesis shows numerous control plots to demonstrate the excellent agreement between data and Monte Carlo.

Contents

Abstract	3
1 Introduction	9
2 Theory	13
2.1 The Standard Model of Particles and Fields	13
2.1.1 Fermions	13
2.1.2 Gauge Bosons / Interactions	14
2.1.3 Higgs	15
2.1.4 Symmetries	16
2.2 The Top Quark	16
2.2.1 Properties	16
2.2.2 Production	16
2.2.3 Decay	17
3 Methods	21
3.1 Lepton Fake Rate & Real Lepton Efficiency	21
3.2 Multivariate Analysis Methods	22
3.2.1 Boosted Decision Tree	23
3.2.2 Neural Networks	23
4 Experimental Setup	25
4.1 The Tevatron	25
4.2 The DØ Detector	25
5 Analysis	31
5.1 Methodology	31
5.1.1 Reconstruction	31
5.1.2 Monte Carlo Samples	31
5.2 Analysis Chain	32
5.2.1 Initial Selection	32
5.2.2 Analysis Cuts	34
5.2.3 Control Plots	37
5.2.4 Fake Rate	37
5.3 Variables	37
6 Results	41
6.1 Control Plots	41
6.1.1 Muon Track Curvature Significance	41
6.2 W Scale Factors S_W	52
6.3 Lepton Fake Rate & Real Lepton Efficiency	52

6.4	Fake Rate and Efficiency Systematic Errors	53
6.4.1	Statistical Errors	53
6.4.2	Systematic Errors	53
7	Conclusion	63
8	Summary & Outlook	65
8.1	Fake Rate and Efficiency Dependencies	65
8.2	Using the full Dataset	67
8.3	General Code for Reweighting	68
8.4	Better Automatic Evaluation of Data / Monte Carlo Agreement	68
	Literature	69
	Acknowledgement	73
A	More Result Values	75
B	More Control Plots	81

List of Figures

1.1	Work flow of the selection and the analyses	10
2.1	Graphical overview of the SM particles	13
2.2	Single top production modes	17
2.3	Graphical overview of the $t\bar{t}$ decay channels	18
2.4	The $t\bar{t}$ decay channels	19
4.1	The Fermilab accelerator chain	26
4.2	Delivered and recorded integrated luminosity for RunIIb	26
4.3	Cross section view of the full DØ detector	28
4.4	Cross section view of the tracking system of the DØ detector	28
5.1	Analysis process	32
5.2	$d\phi(\ell, \cancel{E}_T)$ vs \cancel{E}_T distribution for data and Monte Carlo signal	35
5.3	The different triangle cuts	38
5.4	The trkcurvsig cut	39
6.1	H_T plots for RunIIb e+jets using the $t\bar{t}$ selection cuts	42
6.2	H_T plots for RunIIb μ +jets using the $t\bar{t}$ selection cuts	43
6.3	H_T plots for RunIIb e+jets using the single top cuts	44
6.4	H_T plots for RunIIb μ +jets using the single top cuts	45
6.5	$d\phi(\ell, \cancel{E}_T)$ plots for RunIIb4 μ +jets using the $t\bar{t}$ selection cuts	48
6.6	$d\phi(\ell, \cancel{E}_T)$ plots for RunIIb4 μ +jets using the $t\bar{t}$ selection cuts and the cut on track curvature significance	49
6.7	Lepton η plots for RunIIb4 μ +jets using the $t\bar{t}$ selection cuts	50
6.8	trkcurvsig plots for RunIIb4 μ +jets using the $t\bar{t}$ selection cuts	51
6.9	\cancel{E}_T plots for RunIIb3 e+jets channel	54
8.1	ϵ_{QCD} plots for the inclusive 2 jet bin in RunIIb2 derived from lepton p_T	65
8.2	ϵ_{sig} plots for the inclusive 2 jet bin in RunIIb2 derived from lepton p_T	66
8.3	ϵ_{sig} plots for the inclusive 2 jet bin in RunIIb2 derived from lepton η	66
8.4	ϵ_{sig} plots for the inclusive 2 jet bin in RunIIb2 derived from the instantaneous luminosity	67
B.1	Histograms for ϵ_{QCD} calculation for RunIIb1 e+jets using $t\bar{t}$ selection cuts	84
B.2	Histograms for ϵ_{QCD} calculation for RunIIb1 μ +jets using $t\bar{t}$ selection cuts	85
B.3	ϵ_{QCD} plots for RunIIb1 e+jets using $t\bar{t}$ selection cuts	86
B.4	ϵ_{QCD} plots for RunIIb1 μ +jets using $t\bar{t}$ selection cuts	87
B.5	Histograms for ϵ_{QCD} calculation for RunIIb2 e+jets using $t\bar{t}$ selection cuts	88
B.6	Histograms for ϵ_{QCD} calculation for RunIIb2 μ +jets using $t\bar{t}$ selection cuts	89
B.7	ϵ_{QCD} plots for RunIIb2 e+jets using $t\bar{t}$ selection cuts	90
B.8	ϵ_{QCD} plots for RunIIb2 μ +jets using $t\bar{t}$ selection cuts	91

B.9	Histograms for ϵ_{QCD} calculation for RunIIb3 e+jets using $t\bar{t}$ selection cuts	92
B.10	Histograms for ϵ_{QCD} calculation for RunIIb3 μ +jets using $t\bar{t}$ selection cuts	93
B.11	ϵ_{QCD} plots for RunIIb3 e+jets using $t\bar{t}$ selection cuts	94
B.12	ϵ_{QCD} plots for RunIIb3 μ +jets using $t\bar{t}$ selection cuts	95
B.13	Histograms for ϵ_{QCD} calculation for RunIIb4 e+jets using $t\bar{t}$ selection cuts	96
B.14	Histograms for ϵ_{QCD} calculation for RunIIb4 μ +jets using $t\bar{t}$ selection cuts	97
B.15	ϵ_{QCD} plots for RunIIb4 e+jets using $t\bar{t}$ selection cuts	98
B.16	ϵ_{QCD} plots for RunIIb4 μ +jets using $t\bar{t}$ selection cuts	99
B.17	H_T plots for RunIIb1 e+jets using the $t\bar{t}$ selection cuts	100
B.18	H_T plots for RunIIb1 μ +jets using the $t\bar{t}$ selection cuts	100
B.19	H_T plots for RunIIb2 e+jets using the $t\bar{t}$ selection cuts	101
B.20	H_T plots for RunIIb2 μ +jets using the $t\bar{t}$ selection cuts	101
B.21	H_T plots for RunIIb3 e+jets using the $t\bar{t}$ selection cuts	102
B.22	H_T plots for RunIIb3 μ +jets using the $t\bar{t}$ selection cuts	102
B.23	H_T plots for RunIIb4 e+jets using the $t\bar{t}$ selection cuts	103
B.24	H_T plots for RunIIb4 μ +jets using the $t\bar{t}$ selection cuts	103
B.25	H_T plots for RunIIb1 e+jets using the single top cuts	104
B.26	H_T plots for RunIIb1 μ +jets using the single top cuts	104
B.27	H_T plots for RunIIb2 e+jets using the single top cuts	105
B.28	H_T plots for RunIIb2 μ +jets using the single top cuts	105
B.29	H_T plots for RunIIb3 e+jets using the single top cuts	106
B.30	H_T plots for RunIIb3 μ +jets using the single top cuts	106
B.31	H_T plots for RunIIb4 e+jets using the single top cuts	107
B.32	H_T plots for RunIIb4 μ +jets using the single top cuts	107

List of Tables

2.1	The known leptons and their properties	14
2.2	The known quarks and their properties	15
2.3	Gauge bosons and their properties	15
5.1	Integrated luminosities for the different run periods	31
5.2	Relevant Monte Carlo samples used for this selection	33
6.1	Total yields for the $t\bar{t}$ selection cuts in RunIIb for the e+jets channel .	46
6.2	Total yields for the $t\bar{t}$ selection cuts in RunIIb for the μ +jets channel	47
6.3	S_W scale factors for the $t\bar{t}$ selection cuts	52
6.4	Jet bin dependent ϵ_{QCD} values for the $t\bar{t}$ selection cuts	55
6.5	Jet bin dependent ϵ_{sig} values for the $t\bar{t}$ selection cuts	56
6.6	Lepton fake rate and real lepton efficiency for the $t\bar{t}$ selection cuts .	57
6.7	Results of systematic error analysis for ϵ_{QCD} using $t\bar{t}$ selection cuts – part 1	59
6.8	Results of systematic error analysis for ϵ_{QCD} using $t\bar{t}$ selection cuts – part 2	60
6.9	ϵ_{QCD} derived from different variables using $t\bar{t}$ selection cuts – part 1 .	61
6.10	ϵ_{QCD} derived from different variables using $t\bar{t}$ selection cuts – part 2 .	62
A.1	S_W scale factors for the single top cuts	75
A.2	Total yields for the single top cuts in RunIIb for the e+jets channel .	76
A.3	Total yields for the single top cuts in RunIIb for the μ +jets channel .	77
A.4	Jet bin dependent ϵ_{QCD} values for the single top cuts	78
A.5	Jet bin dependent ϵ_{sig} values for the single top cuts	79

1. Introduction

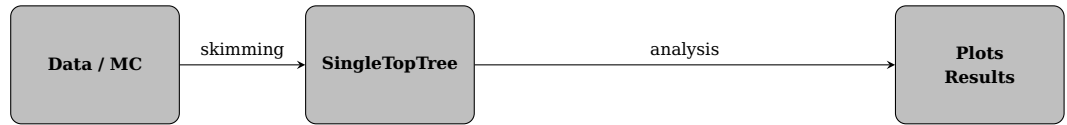
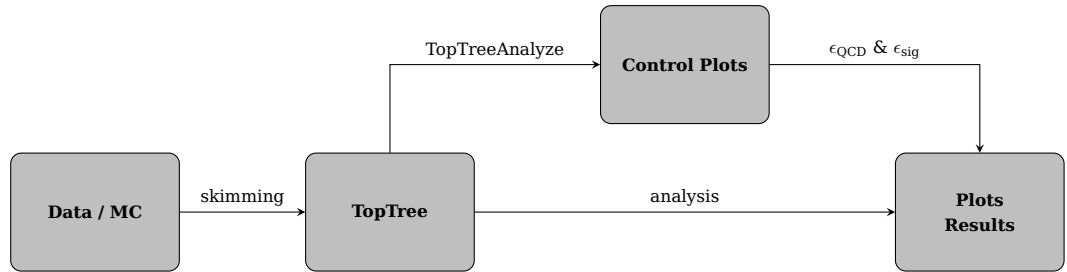
The goal of this Master thesis project was to combine the single top and $t\bar{t}$ lepton+jets selection and update it to include the full RunII data set of the DØ detector. Various technical details in the storage format of this final selection and in the software used to process it needed to be improved and updated to the latest versions.

In a second step the shape and the scale of the multijet background are calculated for each run period individually using the matrix method including various studies to estimate the statistical and systematic errors of the fake rates and the efficiencies; in addition it had to be verified that the matrix method also works for the updated multivariate electron identification that does not use different cut-based rules anymore but only one cut on the output of a process involving a boosted decision tree and an artificial neural network.

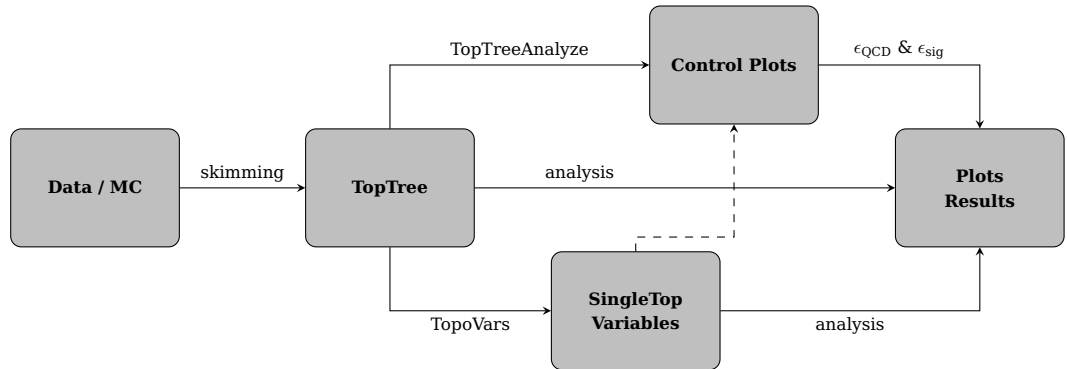
There are many advantages in using a common selection for single top and for the $t\bar{t}$ lepton+jets channel. First of all the selection criteria are relatively similar and using a common selection saves manpower as some steps in the work flow can be combined. Furthermore it saves disk space as the whole tree only needs to be produced and stored once. Saving resources in preparing the selection is particularly important because these resources can be used for analyses or other projects.

Having a more consistent overall setup and one software tool-chain less also helps when introducing new people that want to work with the TopTree format or when passing on work to other people. All that can be achieved without setting any additional restrictions on the analysis selections that are later derived from the centrally produced trees. Setting all the cuts loose enough allows to accommodate both of the previous selections and the additional variables usually needed for single top analyses can be easily added as a second layer to the trees without the need to rerun all the samples or change the original files. For a more detailed description of the new work-flow see figure 1.1. When preparing the previous analyses both groups derived their selection independently and in many cases there were two tools basically performing the same task; for example the skimming was done separately and the lepton fake rate had to be derived for each selection. For the new selection common steps were identified and the work-flow was changed in order to avoid doing the same thing twice.

Understanding the background of an analysis channel is arguably one of the most important steps of any analysis. For small signals expected like especially in single top or in the search for new physics it is crucial to have as much information about the background as possible. In general there are two kinds of backgrounds. The easier one to control is the so-called *physics* background consisting of other hard-scattering events that have a similar signature in the detector (i.e. for the lepton+jets channel that means events containing a real lepton). For lepton+jets this can be for example W+jets (Z+jets) events where the W (Z) decays leptonically,



(a) Old work flow using separate selections



(b) New, optimized work flow using the combined selection

Figure 1.1: Work flow of the selection and the analyses (old and new version)

diboson events, and for single top analyses even the $t\bar{t}$ is considered to be background. These backgrounds can be estimated precisely using Monte Carlo simulation with various corrections for all the different possible processes. The produced Monte Carlo events then run through a simulation of the detector resulting in reconstructed events that have the exact same format as data from measurement but with additional information like particle IDs, etc.

The more difficult background is *instrumental* background originating from events with no real lepton in the final state but in which one of the jets is misidentified in the detector as a lepton and thus passes the selection criteria (in addition to the misidentified electron there usually is a wrongly reconstructed missing energy so the event passes the cut on that variable as well). This background is derived from data using the matrix method. Basically this means that two different identification levels for a real lepton (one sample called *loose* with loose criteria and a *tight* subsample with more stringent requirements) are used and then the shape of the multijet background is constructed by subtracting those two samples. The scaling is calculated using special samples with an inverted cut on \cancel{E}_T (more details in the section 3).

In the end the real data events are compared to the sum of events from all the Monte Carlo samples (i.e. physics background and simulated signal) and the correctly scaled multijet (i.e. instrumental) background by plotting them in histograms for different variables. If the detector is modelled correctly the two distributions should match each other within statistical errors.

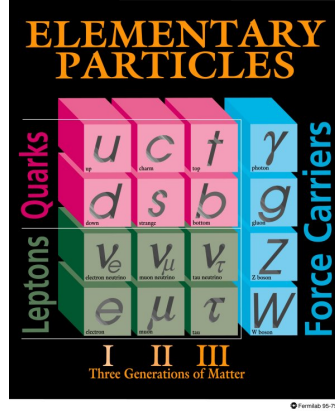


Figure 2.1: Graphical overview of the SM particles [4].

2. Theory

This section describes some of the important theoretical foundations for this thesis including an introduction to the standard model of particle physics with an emphasis on the top quark and important properties of the top quark.

2.1. The Standard Model of Particles and Fields

The standard model of particles and fields (SM) is the theory describing the fundamental particles and their interactions (excluding gravity which can not be included in this framework). The SM correctly predicted multiple unknown particles that were subsequently observed (the bottom quark was discovered in 1977 [1], the top quark in 1995 [2], and the τ -neutrino in 2000 [3]). Due to numerous experiments and analyses the SM is one of the best and most accurately tested scientific theories. This section gives a short introduction into the particles, described by the SM. An overview of all these particles is shown in figure 2.1.

2.1.1. Fermions

The fundamental spin one-half particles described by the SM are called fermions and obey Fermi-Dirac statistics. They can be organized into three generations where each generation consists of two *quarks* and two *leptons*. The known quarks and leptons are listed in tables 2.2 and 2.1¹, respectively.

The classification into quarks and leptons is according to the possible interactions of the particles. All the fermions carry a weak isospin and therefore interact through the weak interaction. All of the quarks and one lepton from each generation carry an

¹All the numbers are given in natural units with $\hbar = c = 1$.

Generation	Lepton	Mass [MeV]	Charge [e]
1	(e^-) electron	$0.510998910 \pm 0.000000013$	-1
	(ν_e) electron neutrino	$< 2 \cdot 10^{-6}$	0
2	(μ^-) muon	105.658367 ± 0.000004	-1
	(ν_μ) muon neutrino	< 0.19	0
3	(τ^-) tau	1776.82 ± 0.16	-1
	(ν_τ) tau neutrino	< 18.2	0

Table 2.1: The known leptons and their properties [6]

electric charge and interact with other charged fermions through the electromagnetic interaction. The three leptons that have no electric charge and can therefore only interact through the weak interaction are called *neutrinos*. Quarks carry an additional color charge and thus also interact through the strong interaction.

The Dirac equation that describes these massive spin one-half particles correctly predicts the existence of an antiparticle for every charged massive fermion (i.e. for all the quarks and all the charged leptons) with the exact same mass but opposite charge and opposite magnetic moment relative to the direction of the spin and allows for the existence of an anti-neutrino. [5]

Of the three charge carrying leptons only one of them (the electron e^-) and its antiparticle (the positron e^+) are stable while the other two (and their antiparticles) decay via the weak interaction. The finite lifetimes and the masses seem to be the only difference (in addition to the different flavours) between the charged leptons in each generation; all of them appear to be elementary particles.

Of the six quarks only the up quark and the down quark (and their antiparticles) are stable while the other decay via the weak interaction into first generation quarks. The first generation is special insofar as all (stable) matter consists of first generation elementary particles. Quarks have never been observed as isolated particles but always inside compound structures that extend over approximately 1 fm. The simplest quark systems are *mesons* consisting of one quark and one antiquark, and *baryons* consisting of three quarks (or *antibaryons* consisting of three anti-quarks) [7]. It is however possible to measure properties of single quarks using deep inelastic scattering (DIS) [8] and the asymptotic freedom (i.e. that the strong interactions becomes arbitrarily small at short distances) of the corresponding gauge theory.

2.1.2. Gauge Bosons / Interactions

The interactions between the fermions are mediated by spin one particles called gauge bosons and obey Bose-Einstein statistics. The properties of the known gauge bosons are listed in table 2.3.

The *photon* is the electromagnetic force carrier and mediates the electromagnetic

Generation	Quark	Mass [GeV]	Charge [e]
1	(u) up	$0.0017 - 0.0031$	$2/3$
	(d) down	$0.0041 - 0.0057$	$-1/3$
2	(c) charm	$1.29^{+0.05}_{-0.11}$	$2/3$
	(s) strange	$0.1^{+0.03}_{-0.02}$	$-1/3$
3	(t) top	$172.9 \pm 0.6 \pm 0.9$	$2/3$
	(b) bottom	$4.19^{+0.18}_{-0.06}$	$-1/3$

Table 2.2: The known quarks and their properties [6]

Boson	Mass [GeV]	Charge [e]	Interaction
(γ) photon	$< 1 \cdot 10^{-27}$	$< 1 \cdot 10^{-35}$	electromagnetic
(g) gluon	0 (theoretical)	-	strong
(W^+)	80.399 ± 0.023	1	weak
(W^-)	80.399 ± 0.023	-1	weak
(Z)	91.1876 ± 0.0021	0	weak

Table 2.3: Gauge bosons and their properties [6]

interaction; it is described by quantum electrodynamics (QED) [9] which is one of the best tested field theories.

The eight *gluons* mediate the strong interaction between the quarks. Because they carry an effective color charge themselves they also self-interact. The gluons are described by quantum chromodynamics (QCD) [10].

The W and Z bosons mediate the weak interaction. These two bosons are massive, which limits the reach of the weak interaction. The W bosons also carry an electric charge and thus couple to the electromagnetic interaction, furthermore they only couple to left-handed fermions and right-handed anti-fermions while the electrically neutral Z bosons couples to all charged leptons but only to left-handed neutrinos and right-handed anti-neutrinos. The three weak force carriers can also interact with each other.

2.1.3. Higgs

The *Higgs* boson is the only SM particle that has not been observed yet. The theorized Higgs boson is a scalar elementary particle that is a quantum of the theoretical Higgs field, it is the only scalar particle predicted by the SM. The associated Higgs mechanism explains how the vector bosons get a mass (that violates the local gauge symmetry) through spontaneous electroweak symmetry breaking [5].

2.1.4. Symmetries

Theoretically the SM can be described by its internal local gauge group symmetry

$$SU(3)_C \times SU(2)_L \times U(1)_Y$$

that gives rise to the conservation of color charge, weak isospin, weak hypercharge, and electric charge by Noether's theorem [11].

In addition the full SM Lagrangian is invariant under baryon phase, electron phase, muon phase, and tau phase (accidental global symmetries) which in turn lead to the conservation of baryon number and lepton number (for each generation individually).

2.2. The Top Quark

The top quark is the heaviest elementary particle known today. It is interesting for several reasons. Due to its short lifetime (considerably shorter than the timescale of the strong interaction) it is the only quark that does not hadronize and thus allows for the measurements of free quark properties like for example spin that are passed on to its decay products. Due to its high mass it has the strongest Yukawa-coupling to the Higgs boson of all fermions, which makes it interesting to search for processes involving interactions between these two particles.

2.2.1. Properties

The top quark forms a weak isospin doublet together with the bottom quark, where the top quark has $I_3 = \frac{1}{2}$. It has an electromagnetic charge of $\frac{2}{3}e$ and a mean lifetime of approximately $5 \cdot 10^{-25}$ s. The current world average for the top mass is

$$172.9 \pm 0.6 \text{ (stat.)} \pm 0.9 \text{ (syst.) GeV [6].}$$

The error is already dominated by systematics which means that even with the much higher statistics that will be available within the next few years it will need considerable effort to reduce the overall uncertainty.

2.2.2. Production

The top quark can be produced via strong or electroweak interaction. At the Tevatron with $\sqrt{s} = 1.96$ TeV the $t\bar{t}$ pair production via strong interaction is the dominant process. The production of a single top quark (see figure 2.2) can occur via the s-channel process $q\bar{q} \rightarrow W^+ \rightarrow t\bar{b}$ or via the t-channel process $gq \rightarrow tq\bar{b}$ (where $t\bar{b}$ always refers to $t\bar{b}$ and $\bar{t}b$ and $tq\bar{b}$ to $tq\bar{b}$ and $\bar{t}qb$).

The theoretically predicted $t\bar{t}$ production cross section (NNLO approximation) at $\sqrt{s} = 1.96$ TeV for a top mass of 172.5 GeV is

$$\sigma_{t\bar{t}} = 7.46^{+0.48}_{-0.67} \text{ pb [12] or } \sigma_{t\bar{t}} = 6.41^{+0.51}_{-0.42} \text{ pb [13] .}$$

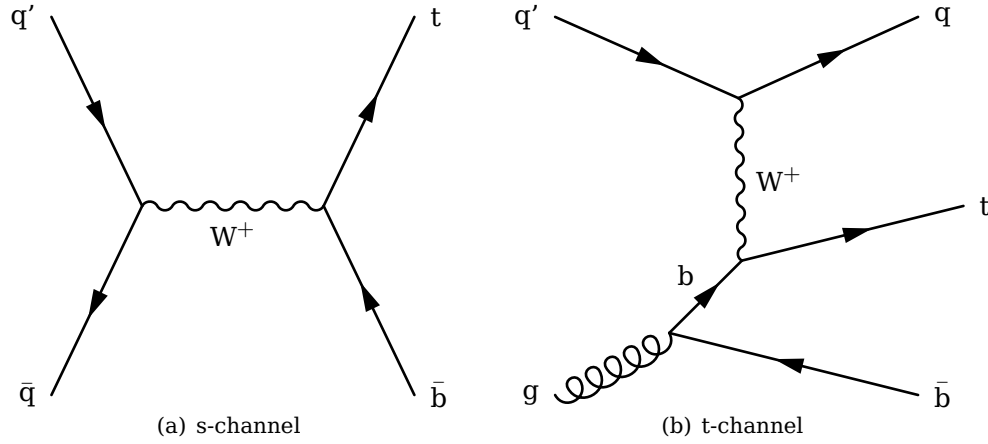


Figure 2.2: The two single top production modes that occur at the Tevatron collider. (a) shows *s-channel* single production and (b) shows *t-channel* production

The theoretically predicted single top production cross sections are (assuming a top mass of 170 GeV and using a NNLL-NLO matching)

$$\sigma(p\bar{p} \rightarrow t\bar{b} + X) = 1.12 \pm 0.04 \text{ pb} \quad \text{and} \quad \sigma(p\bar{p} \rightarrow tqb + X) = 2.34 \pm 0.12 \text{ pb} \quad [14].$$

Comparing these cross sections motivates the use of more advanced multivariate statistical tools like neural networks or decision trees for the single top search. Also, while the $t\bar{t}$ signal is concentrated in the four jet bin, most of the single top signal is found in the two and three jet bins that suffer from much higher background contamination compared to the four jet bin.

2.2.3. Decay

The SM predicts a decay of the top quark via weak interaction into a quark and a W boson; measurements show that

$$\frac{\Gamma(t \rightarrow Wb)}{\Gamma(t \rightarrow Wq)} = 0.99^{+0.09}_{-0.08} \quad \text{where } q \in \{b, s, d\} \quad [6]. \quad (1)$$

Thus the decay channels of the top quark are classified according to the decay of the two W bosons from the decaying $t\bar{t}$ pair.

A W^+ boson (from a t decay) can decay hadronically into $u\bar{d}$ or $c\bar{s}$ (while the W^- can obviously decay into the charge conjugates) or leptonically into one of the three anti-leptons and the corresponding neutrino (or in the case of W^- a lepton and the corresponding anti-neutrino).

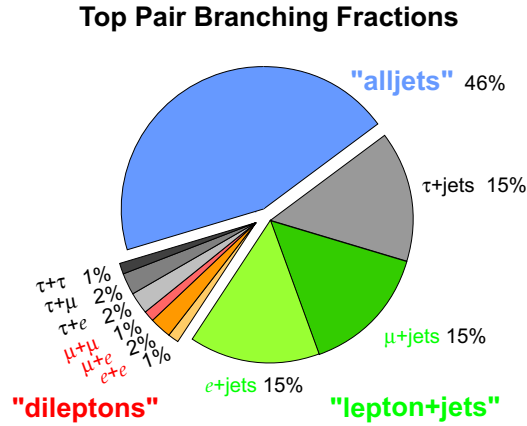


Figure 2.3: Graphical overview of the $t\bar{t}$ decay channels and the corresponding branching ratios [15].

Assuming lepton-universality the branching ratio for a W boson decaying into a lepton is given by

$$\frac{\Gamma(W^+ \rightarrow \ell^+ \nu)}{\Gamma_{\text{full width}}} = (10.80 \pm 0.09)\% \quad \text{where } \ell^+ \in \{e^+, \mu^+, \tau^+\} \quad [6]. \quad (2)$$

Summing up these branching ratios results in the approximate 2 : 1-ratio for hadronic to leptonic decays one would expect when factoring in the three lepton families as well as the QCD color factors. A graphical overview of the branching ratios is plotted in figure 2.3.

This leads to three possible “channels” for the decay of a $t\bar{t}$ pair (see figure 2.4). The *alljets* channel where both of the W decay hadronically (about 46% of all events), the *lepton+jets* channel where one of the W decays hadronically and one leptonically (about 44% of all events), and the *dilepton* channel where both W’s decay into a lepton and a neutrino (about 11% of all events).

Considering the detector inefficiencies in identifying the τ and its various decay modes (both leptonically and hadronically) the branching ratio calculation for the observed quantities is actually more complicated.

The alljets channel is the least accessible because of the overwhelming background from other multijet events. The dilepton channel is the most pure decay channel (i.e. it has a very clear signature and low background contamination), but is suffering from low statistics due to the small branching ratio. The lepton+jets channel has almost as high statistics as the alljets channel while the contamination from various background processes is much lower due to its clear signature from the lepton.

For single top analyses there is only one decay of a t or \bar{t} and thus only the lepton+jets channel is reasonable because the background of the alljets channel is not manageable for the even lower statistics we have here.

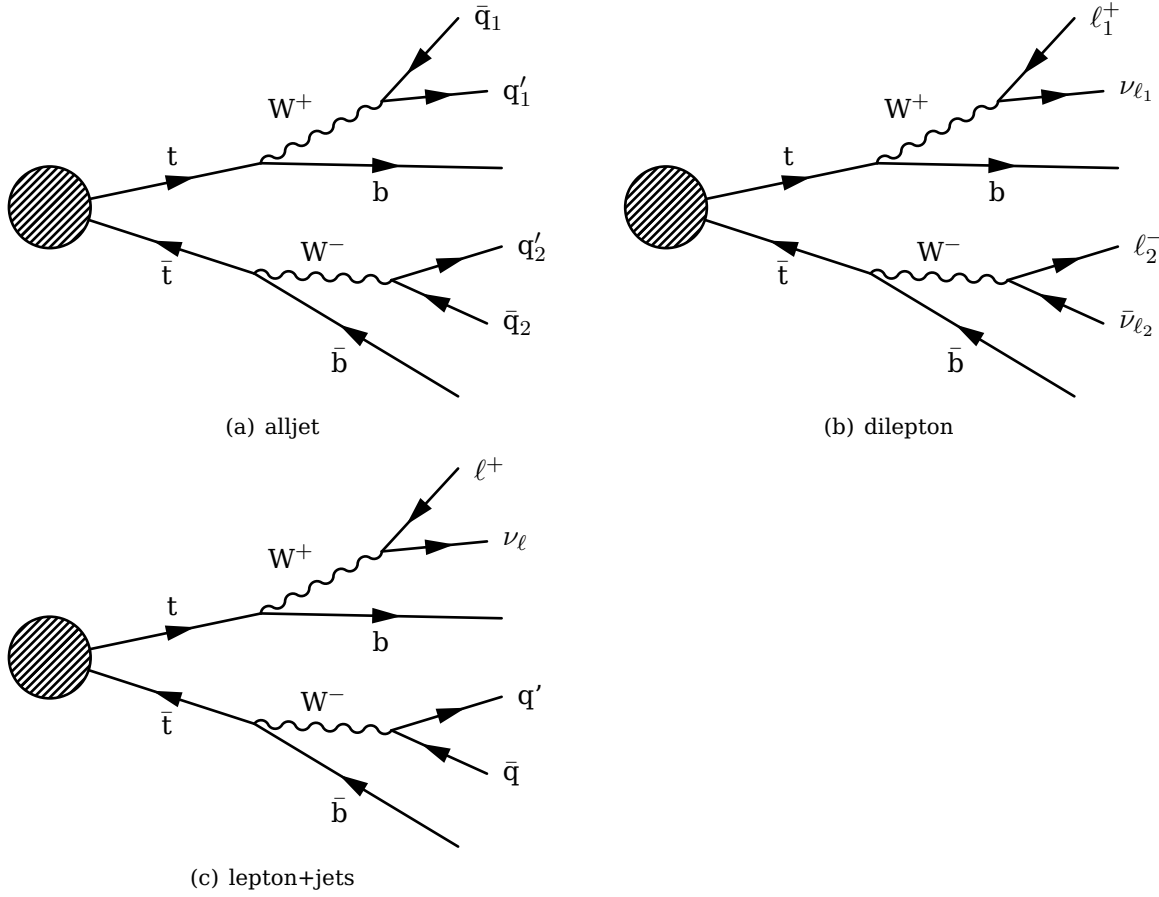


Figure 2.4: The $t\bar{t}$ decay channels are classified according to the decay of the W boson. This leads to three possibilities for the decay of a $t\bar{t}$ pair; (a) the *alljets* channel where both Ws decay hadronically, (b) the *dilepton* channel where both Ws decay leptonically, and (c) the *lepton+jets* channel where one W decays hadronically and the other one decays leptonically.

3. Methods

This section contains a description of the matrix method which is later used to derive the instrumental background as well as a few details about special statistical methods used in the analysis of the data.

3.1. Lepton Fake Rate & Real Lepton Efficiency

One of the main problems of the lepton+jets channel is the misidentification of a jet as an electron or as a muon and thus selecting events that belong to a different decay channel or are just instrumental background. On the other hand not every real lepton passes the electron identification criteria; this real lepton efficiency is encoded in the value ϵ_{sig} .

In order to determine the multijet background two different data samples are required, the so called *loose* sample with comparatively lower requirements for identifying an electron (or muon) and the *tight* sample with much more stringent identification requirements. The *tight* sample needs to be a subset of the *loose* sample and the shape of the multijet background is then estimated by taking the events that passed the loose identification requirements but not the tight ones referred to as the *loose-tight* sample.

Defining N^{QCD} as the number of events with a fake lepton and $N^{\text{W-like}}$ as the number of events with a real isolated lepton the number of events in the *loose* and *tight* samples (n_l , n_t) can be calculated by

$$\begin{aligned} n_l &= N^{\text{QCD}} + N^{\text{W-like}} \\ n_t &= \epsilon_{\text{QCD}} \cdot N^{\text{QCD}} + \epsilon_{\text{sig}} \cdot N^{\text{W-like}} \end{aligned} \quad (3)$$

where the lepton fake rate ϵ_{QCD} is the probability that a fake lepton passes the *tight* selection criteria and the real lepton efficiency ϵ_{sig} is the probability that a real lepton passes the *tight* selection criteria.

The value of ϵ_{sig} is simply determined from Monte Carlo samples containing a real isolated lepton (that fulfilled the *loose* selection criteria) in the final state. Knowing that

$$N^{\text{QCD}} = 0 \quad (4)$$

for Monte Carlo the real lepton efficiency can be found by dividing the number of events from the *tight* and the *loose* samples

$$\epsilon_{\text{sig}} = \frac{n_t}{n_l} \quad (5)$$

Similarly the value of ϵ_{QCD} is determined from a set of special samples that is derived applying exactly the same cuts as the ones used for the analysis, but instead

of cutting on a minimal value of missing transverse energy \cancel{E}_T now the cut is on a maximum value of \cancel{E}_T . This analysis uses $\cancel{E}_T < 10 \text{ GeV}$ (and $\cancel{E}_T < 15 \text{ GeV}$ to estimate one of the systematic uncertainties). Assuming that the sample with low \cancel{E}_T contains almost no signal (the neutrino that accompanies a real lepton is detected via the missing transverse energy)

$$N^{\text{W-like}} \approx 0 \quad (6)$$

the lepton fake rate can again be found by dividing, here for the data sample

$$\epsilon_{\text{QCD}} = \frac{n_t}{n_l} \quad (7)$$

Because according to the processed Monte Carlo samples there are some events in these low \cancel{E}_T regions with real leptons (and perhaps miscalculated \cancel{E}_T) the sum of all Monte Carlo samples is subtracted from the data before dividing. Thus n_t and n_l here are the numbers after those subtractions.

Both of these calculations are not performed by simply using the sum of all events, but by dividing (and subtracting) histograms filled from both samples (for plots see appendix B). The variations of the fake rate and the efficiency over the used variables are thus included in the statistical fitting error while the differences for the values derived from different variables can be used as a systematic uncertainty.

For later normalization of the *loose-tight* sample describing the shape of the instrumental multijet background one resolves the equation system given in (3)

$$N^{\text{QCD}} = \frac{\epsilon_{\text{sig}} \cdot n_l - n_t}{\epsilon_{\text{sig}} - \epsilon_{\text{QCD}}}, \quad N^{\text{W-like}} = \frac{n_t - \epsilon_{\text{QCD}} \cdot n_l}{\epsilon_{\text{sig}} - \epsilon_{\text{QCD}}} \quad (8)$$

and scales the background to an integral of N^{QCD} .

The exact definitions of the *loose* and the *tight* criteria have only very small effects on the final multijet background. For the real lepton efficiency it is just important to define the *loose* point in a way that selects (almost) all of the (real!) leptons in the MC samples. For the lepton fake rate it is important to choose the requirements in a way that leaves enough statistics for the *loose-tight* sample. Selecting for example looser requirements for *loose* will result in a smaller value for ϵ_{QCD} but the total yield N^{QCD} remains almost constant as the numerator increases as well.

3.2. Multivariate Analysis Methods

Because of the very small predicted ratio of detected signal versus background (physical and noise) especially for the single top analyses, it has been proven useful to use statistical methods that are not relying on a single variable but working with multiple variables at the same time. These new methods are therefore called *multivariate analysis* (MVA).

This section outlines two methods used in various parts of the selection.

To implement these MVA methods into the selection and analysis code the TMVA package [16] for ROOT [17] is used.

Typical application of these methods include the identification of jets containing a b quark (b-jets), the new electron identification, and the single top analysis at the Tevatron [18].

3.2.1. Boosted Decision Tree

A decision tree is a multivariate tool where a sample is divided into signal and background by using successive decisions nodes (that usually only do a threshold discrimination based on one variable) organized in a tree-like structure.

The training with a known sample is done iteratively by starting at the root node and first using the variable that gives the best separation between signal and background. This process is repeated for each of the subsamples until the sample size reaches a minimum value or the separation quality (i.e. purity of the subsamples) reaches a maximum quality (one or both defined before as training parameters).

Boosting is a meta-algorithm that can be used to improve the learning capabilities of many MVA methods. In each round of training the different samples get a different weight that is calculated based on if they have been classified correctly in the round before or not.

3.2.2. Neural Networks

Artificial Neural Networks (ANN) are usually implemented on a multi-layer feed-forward approach. The first layer (so called *input layer*) receives the values of different variables and passes them on to every node in the next layer while multiplying it with the weight of the specific connection. The nodes in the intermediate layers take all values they have received, sum them up, and run the result through a so called *transfer function*. The output of this function is then passed on to the next layer, repeating the process until the *output layer* is reached that usually has one (separating the signal and background by setting a threshold on that node) or two (one signal and one background node; the one with the higher value determines the classification) nodes.

The ANN can be trained on a known sample by adapting the weights associated with the different connections (which is usually done using some sort of back-propagation algorithm) depending on how correct the output from the training step before was.

4. Experimental Setup

After a 15 year long planning phase Fermilab was commissioned in 1967 as the *National Accelerator Laboratory*. It is located near Batavia, IL about 50 km west of Chicago. In 1974 it was renamed to *Fermi National Accelerator Laboratory* in honour of Nobel Prize winner Enrico Fermi.

It is home to multiple physics and engineering experiments and research laboratories for various fields including particle physics, accelerator physics, detector physics, neutrino physics, and dark matter and dark energy searches. One of these facilities was the Tevatron which was the most powerful particle accelerator for many years with its two detectors DØ and CDF; they lead to numerous insights into high energy physics. The Tevatron was finally shut down in September 2011 but physics analyses continue to get new results from the full dataset.

4.1. The Tevatron

The Tevatron is a circular particle accelerator with a circumference of 6.3 km and was completed in 1982 with the first record setting particle collisions at 512 GeV occurring in 1983. It is the last component of a whole accelerator chain (see figure 4.1 for details) and accelerates protons in one direction and antiprotons in the other one, colliding them at two bunch crossing points (DØ, CDF) around the circumference of the machine. The particles are accelerated in 36 so called *bunches* of around 10^9 to 10^{12} particles each. As the particles move nearly at the speed of light there is a collision every 396 ns yielding a collision frequency of approximately 2.53 MHz assuming that every bunch crossing leads to a collision.

During the first run period of the Tevatron – called *RunI* – the center of mass energy was $\sqrt{s} = 1.8$ TeV.

In 1996 the accelerator was shut down for five years and upgraded and was then capable of colliding the particles at a center of mass energy of

$$\sqrt{s} = 1.96 \text{ TeV}$$

during the second run period *RunII*. This was the world record energy for over eight years, before the *LHC* at *CERN* broke that record in November 2009 by accelerating its two beams to 1.18 TeV each reaching $\sqrt{s} = 2.36$ TeV [20].

The instantaneous luminosity of the Tevatron has peak values around $4 \cdot 10^{32} \text{ cm}^2 \text{ s}^{-1}$; the time evolution of the integrated luminosity of the whole RunII is depicted in figure 4.2.

4.2. The DØ Detector

The DØ detector [22] is a general purpose detector covering almost the full solid angle (" 4π ") that is built around one of the bunch crossing points of the Tevatron.

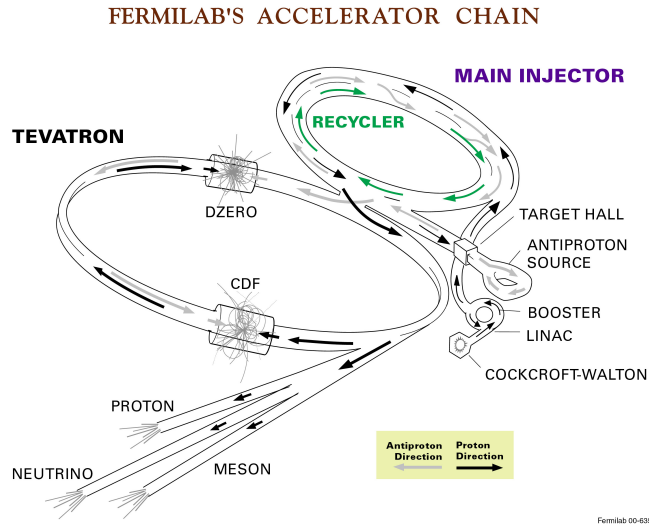


Figure 4.1: The Fermilab accelerator chain [19].

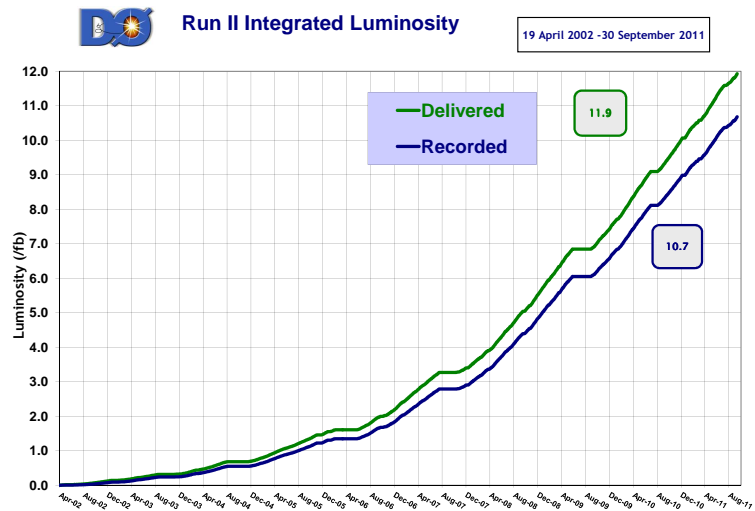


Figure 4.2: Delivered and recorded integrated luminosity for RunIIb [21].

It recorded its first proton antiproton collisions in 1992 running with a tracker, a calorimeter, and a muon system but without a solenoid. The long shutdown period of the Tevatron between RunI and RunII was used to upgrade the detector significantly; its current configuration is shown in figure 4.3. The main components are – from inside out – the tracking system (details in figure 4.4) containing the silicon microstrip detector (SMT) [23] [24], the central fiber tracker (CFT), the superconducting solenoid generating a 2 T magnetic field, and the forward (FPS) and central preshower detectors, the uranium liquid argon calorimeter consisting of the electromagnetic, the fine hadronic, and the coarse hadronic layer and separated into the central calorimeter (CC) and the endcap calorimeters (EC). The outermost components are the muon systems consisting of wire chambers, scintillators, and the toroidal magnet generating a magnetic field of 1.8 T.

Moreover the detector includes some special purpose components as for example the forward proton detector (FPD) or the luminosity monitors (LM) providing additional information for forward physics or diffractive processes and measurement capabilities for the instantaneous luminosity, respectively.

The detector was running in this configuration from 2001 until the Tevatron was shut down in September 2011 (or more precisely three months longer to record zero luminosity measurements) with multiple short shutdowns. One of the longer shutdowns in 2006 was used to upgrade the tracking system of the detector by inserting the so called *Layer 0* into the SMT; a new layer of silicon trackers closer to the beam and thus closer to the collision point than ever before. The data sets recorded by the DØ detector during RunII are therefore grouped into *RunIIa* and *RunIIb* referring to the collision data collected before and after this upgrade, respectively.

The main purpose of the tracking system is to determine the momentum of charged particles by measuring the bending radius due to the magnetic field from the solenoid, while the calorimeter measures the energy of the particles. The muon system is specialised hardware to identify and measure muons and has its own momentum determination capability due to the separate toroidal magnet.

In particular the SMT is very important in locating primary (PV) and secondary interaction vertices (SV). These are later used to identify jets originating from a b quark (b-jets). The CFT is important to gather tracking information over longer distances and the CPS and FPS are used to increase electron identification and background rejection.

The DØ detector uses a right-handed coordinate system with the z -axis along the proton beam, the x -axis pointing to the center of the Tevatron ring, and the y -axis pointing upwards. Due to the symmetries of the detector it is more useful to use a cylindrical coordinate system with (r, ϕ, η) where

$$r = \sqrt{x^2 + y^2}$$

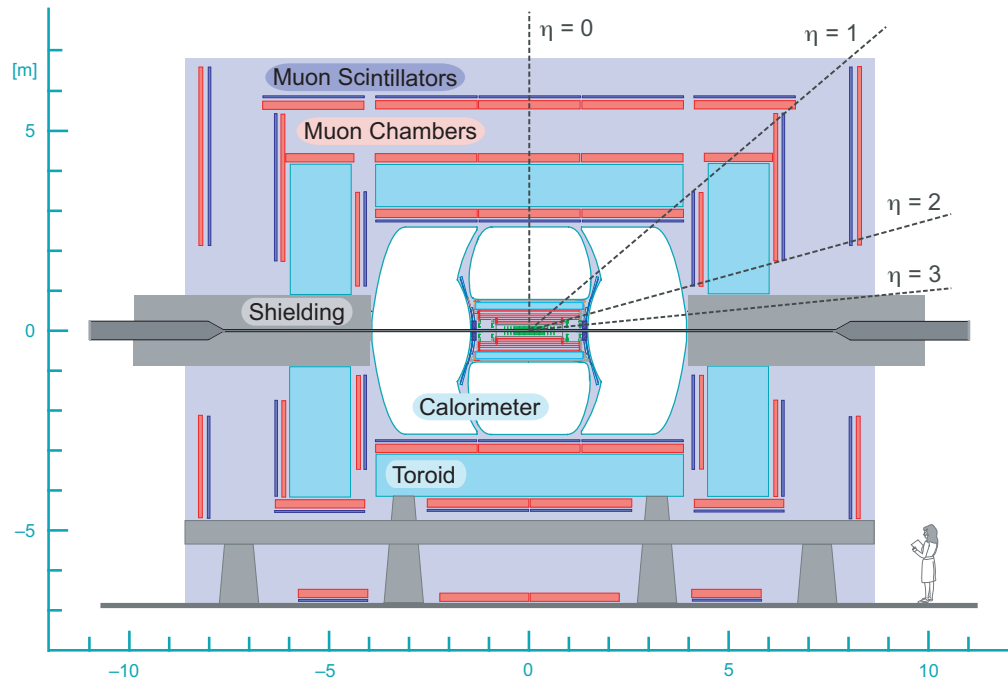


Figure 4.3: Cross section view of the full DØ detector from [25].

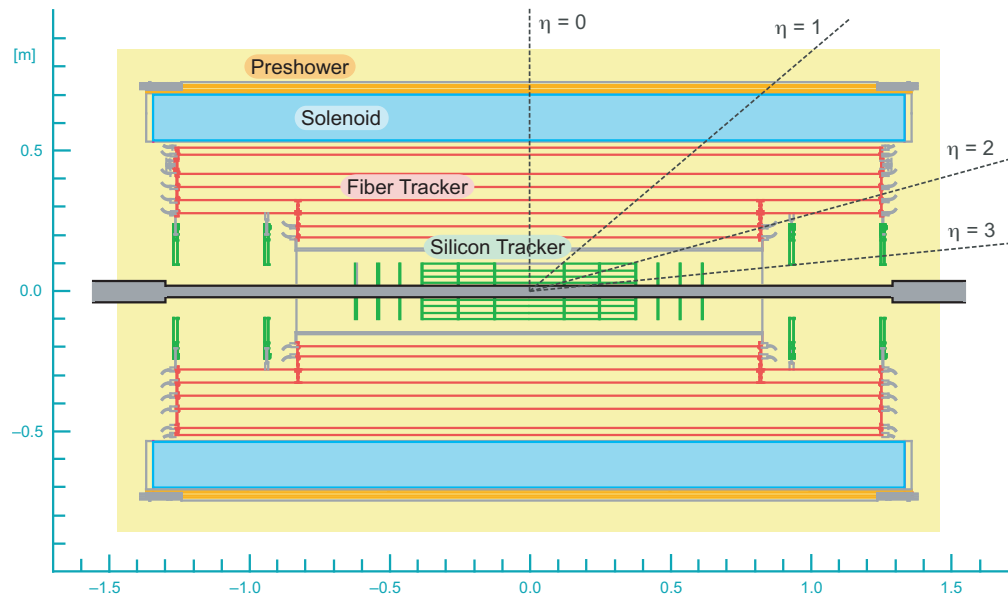


Figure 4.4: Cross section view of the tracking system of the DØ detector from [25].

$$\phi = \arctan \frac{x}{y}$$
$$\eta = -\ln \tan \frac{\theta}{2}$$

with the azimuthal angle θ . η is called pseudo-rapidity and is equivalent to the rapidity

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}$$

for massless particles (p_z is the longitudinal component of the momentum i.e. the one perpendicular to p_T).

The DØ detector “sees” around two million collisions per second. Because only 50 Hz can be reconstructed and written to tape continuously the detector needs a trigger system that decides if an event should be kept or thrown away. The goal is to keep as many interesting physics events (i.e. hard-scattering events where the colliding partons carried a large fraction of the proton momentum) as possible while discarding the soft collisions. For that purpose there is a three tier trigger system that reduces the event rate consecutively down to 50 Hz. The first level – called *L1* trigger – is implemented directly in hardware (including customly programmed FPGA boards) and reduces the rate to about 2 kHz. The *L2 trigger* is a combination of hardware and software tests employing more complicated conditions such as simple object identifications; it brings the rate down to 1 kHz. The *L3 trigger* consists of a farm that runs a simplified version of the reconstruction algorithm and ultimately decides which events should be saved to tape.

Because the instantaneous luminosity can change by about one order of magnitude during a store it is useful to use so-called *prescales* that are defined as a (discrete) function of the instantaneous luminosity. If an event type is said to have a prescale of k this means that only every k th event passing the corresponding triggers is kept.

	Luminosity [pb^{-1}]
RunIIa	1081.21
RunIIb1	1217.67
RunIIb2	3039.84
RunIIb3	1994.27
RunIIb4	2403.91
RunIIa	1081.21
RunIIb	8655.69
RunII	9736.90

Table 5.1: Integrated luminosities for the different run periods after applying data quality requirements.

5. Analysis

This section describes details of the analysis process that is used to produce the selection of events used in top quark analyses, verify the output samples, and calculate the lepton fake rates and the real lepton efficiencies. It only gives some results as examples; the full results sets are listed in the next section starting on page 41 and in appendix A.

5.1. Methodology

5.1.1. Reconstruction

The raw data recorded by the detector (for details see section 4) is ran through offline reconstruction where the read-out from the electronics is “converted” into physics objects like electrons, muons, jets, missing energy, etc. The events are tagged with a data quality flag (determined by setting different quality criteria for all the sub-detectors) during reconstruction and thus the integrated luminosity usable for analyses further decreases leading to the numbers listed in table 5.1.

5.1.2. Monte Carlo Samples

In order to determine what happens when a proton and an antiproton collide and what is then recorded by the detector one uses Monte Carlo simulations of all known or theoretically predicted physics processes.

The events are generated with different event generators (e.g. PYTHIA [26], ALPGEN [27], etc.) and contain all required information for the DØ reconstruction software package. This program tracks all the four-momentum vectors of the particles through the DØ detector whilst the detector response and electromagnetic

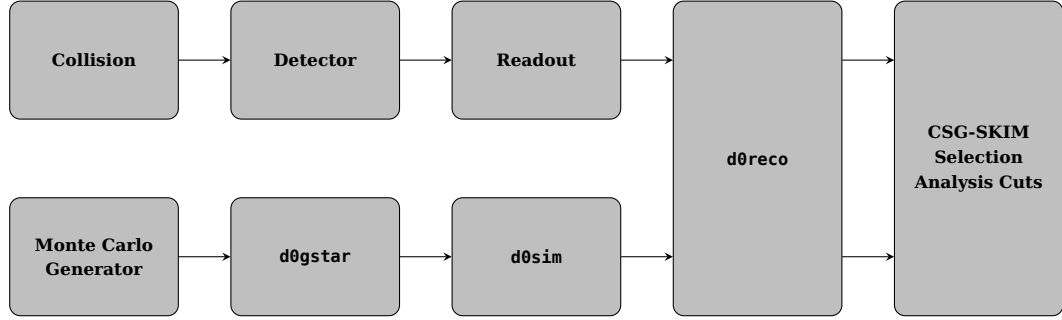


Figure 5.1: Analysis process for data and Monte Carlo samples.

and nuclear interactions are given by a subpackage based on GEANT [28] and another subpackage simulating the electronic read-out. The Monte Carlo event data is digitized and reconstructed running the data reconstruction code of a certain run period. The whole process is shown in detail in figure 5.1.

Table 5.2 lists the relevant Monte Carlo samples used in this selection.

Due to inefficiencies and simulation artifacts multiple corrections (incorporated in the V+jets [29] framework) are necessary for the output of the Monte Carlo simulations to actually match the data in the end.

5.2. Analysis Chain

Starting from preskimmed data and from different Monte Carlo samples the process begins with a (pre)selection of interesting events using the V+jets framework [29]. The output from this selection consists of trees containing all the selected events; the R00T framework [17] is used to save these trees into files. These files can either be used directly for analysis or as input to the Control plots program that produces another set of R00T-files which is then used to generate control plots or as input for the fake rate calculation.

While the initial trees contain information at event-level (i.e. one can look at each variable of every event) files produced later in the process only contain histograms for the different variables for whole run periods.

5.2.1. Initial Selection

At this level only some of the most important cuts are applied to keep the file size of all the produced output files in a reasonable range. The output is stored in the so called TopTree format which is basically a R00T TTree containing over 200 variables per event. It is better not to apply any other cuts at this stage in order to cover as much phase space as possible without requiring time-consuming re-runs of the selection for studies using a somewhat different set of cuts.

(a) Samples used. The X and Y values divide the samples into different energy intervals. The value M is a placeholder for the different masses.

Type	Name
alpgen+pythia W+jets	CSG_alpgenpythia_w_lnu_V
	CSG_alpgenpythia_w+2b_lnu+2b_V
	CSG_alpgenpythia_w+2c_lnu+2c_V
alpgen+pythia gam/Z+jets	CSG_alpgenpythia_gamz_ee_X_Y_V
	CSG_alpgenpythia_gamz+2b_ee+2b_X_Y_V
	CSG_alpgenpythia_gamz+2c_ee+2c_X_Y_V
	CSG_alpgenpythia_gamz_mumu_X_Y_V
	CSG_alpgenpythia_gamz+2b_mumu+2b_X_Y_V
	CSG_alpgenpythia_gamz+2c_mumu+2c_X_Y_V
	CSG_alpgenpythia_gamz_tautau_X_Y_V
	CSG_alpgenpythia_gamz+2b_tautau+2b_X_Y_V
	CSG_alpgenpythia_gamz+2c_tautau+2c_X_Y_V
pythia diboson	CSG_pythia_w+w_incl_V
	CSG_pythia_w+z_incl_V
	CSG_pythia_z+z_incl_V
alpgen+pythia t+tbar	CSG_alpgenpythia_t+t_lnu+2b+2lpc_mM_V
	CSG_alpgenpythia_t+t_2l+2nu+2b_mM_V
comphep+pythia single top	CSG_compheppythia_singletop-tb_m1725_V
	CSG_compheppythia_singletop-tqb_m1725_V

(b) Versions used for each run period.

Version V	
RunIIb1	p211100_v12
RunIIb2	p211800_Run2b2_v6
RunIIb3	p212100_Run2b3_v3

Table 5.2: Relevant Monte Carlo samples used for this selection.

The other requirements for the new cuts were of course given by the previous $t\bar{t}$ lepton+jets selection as well as the previous single top selection [30]. As both groups want to use the TopTree format as input for their future analyses, it needs to be able to reproduce the respective previous selections.

All these considerations including some quick studies evaluating the effects of dropping different cuts on the disk usage of the output files led to the following selection cuts:

- one lepton with $p_T > 10$ GeV
- at least one jet with corrected $p_T > 15$ GeV
- $\cancel{E}_T > 15$ GeV
- $d\phi(\cancel{E}_T, lep) > 1.5 - 0.0375 \cdot \cancel{E}_T$ (for e+jets)
 $d\phi(\cancel{E}_T, lep) > 1.2 - 0.0300 \cdot \cancel{E}_T$ (for μ +jets)

The first cut on the transverse momentum of the lepton is based on the cuts used in the preskims produced for analysis with the V+jets framework, going lower would require to run over much more data while – most probably – not increasing the number of interesting physics events. The cut on the transverse momentum of the jets is based on a limitation set by the jet energy scale corrections that are only available for jets with $p_T > 15$ GeV. The missing transverse energy cut was lowered from previous analyses by 5 GeV or more; requiring an even looser condition would lead to a dramatic increase in total file size. The triangular cuts are based on the fact, that QCD background events are concentrated in this specific region while the $t\bar{t}$ and single top signal events are expected to be more evenly distributed over this phase space plane; the detailed distributions from data and Monte Carlo samples are plotted in figure 5.2.

The initial selection also includes vetoes on a second lepton to ensure orthogonality to the dilepton selection as well as other cuts and corrections defined by the default V+jets configuration (e.g. a cut on $|z| < 60$ cm where z is the z -position of the primary vertex or a correction of the jet energy scale (JES) [31]).

5.2.2. Analysis Cuts

As one of the goals of this selection was to combine the previous $t\bar{t}$ lepton+jets selection with the previous single top selection the control plots and fake rate determination presented here is done for two different sets of cuts:

The $t\bar{t}$ cuts were taken from the previous $t\bar{t}$ selection [32] and contain the following cuts (in addition to the initial selection cuts)

- event accepted by superOR trigger

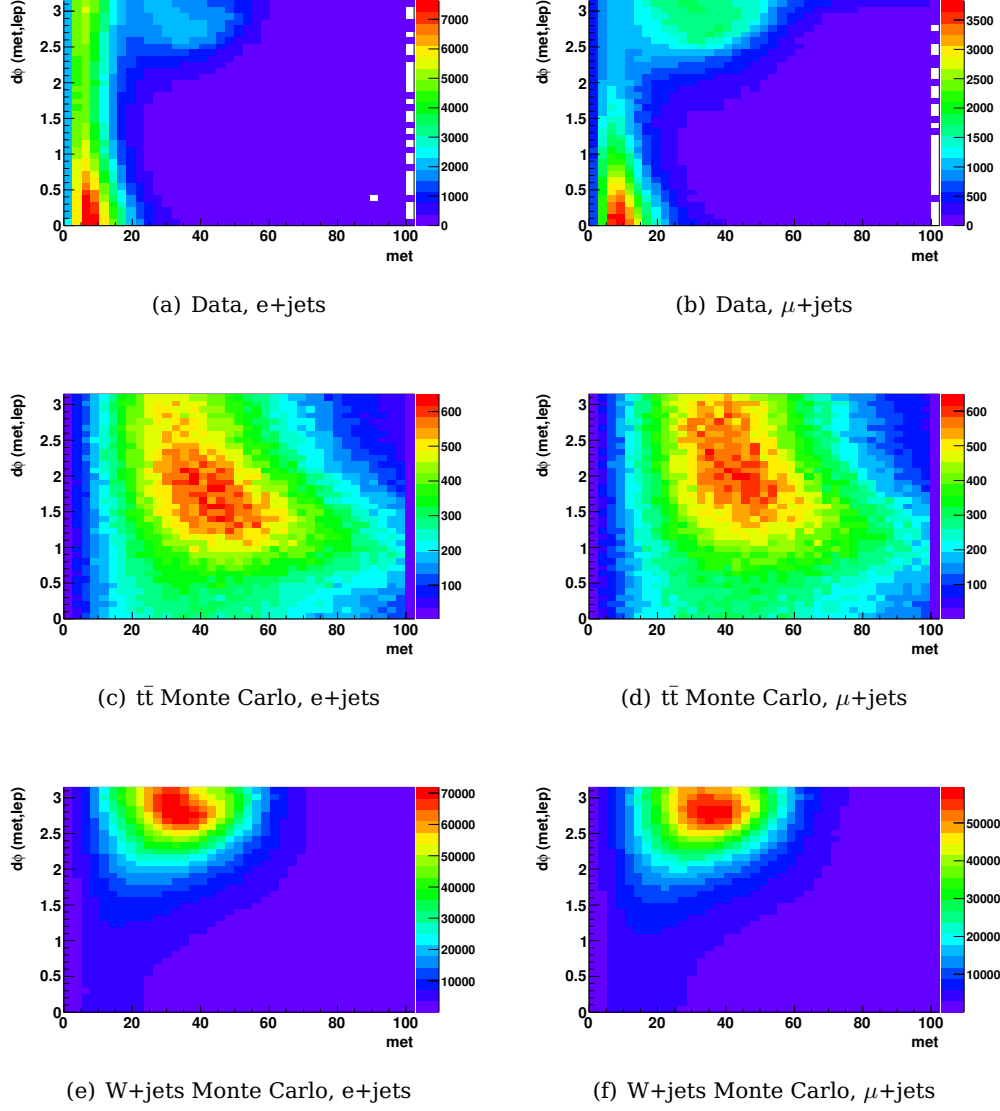


Figure 5.2: $d\phi(\ell, \cancel{E}_T)$ vs \cancel{E}_T distribution for data and Monte Carlo signal for RunIIb1. The distribution in the data plots motivates the triangular cuts used in this selection. All six plots are jet bin inclusive.

- one isolated lepton ($dR_{min} > 0.5$) with $p_T > 20$ GeV
- at least one vertex confirmed² jet with $p_T > 20$ GeV and $|\eta| < 2.5$
- $\cancel{E}_T > 20$ GeV
- leading jet with $p_T > 40$ GeV
- for electrons
 - $d\phi(\cancel{E}_T, lep) > 2.2 - 0.0450 \cdot \cancel{E}_T$
 - $|\eta| < 1.1$ (detector eta)
- for muons
 - $d\phi(\cancel{E}_T, lep) > 2.1 - 0.0350 \cdot \cancel{E}_T$
 - $M_T^W < 250$ GeV and $\cancel{E}_T < 250$ GeV
 - $|\eta| < 2.0$ (detector eta)

The single top cuts are taken from the previous single top selection and already include some improvements only possible with the new software versions and additions used for this selection ³

- event accepted by superOR trigger
- one isolated lepton ($dR_{min} > 0.5$) with $p_T > 20$ GeV
- at least one vertex confirmed jet with $p_T > 20$ GeV and $|\eta| < 3.4$
- $200 \text{ GeV} > \cancel{E}_T > \begin{cases} 20 \text{ GeV} & 1, 2 \text{ jets} \\ 25 \text{ GeV} & 3, 4 \text{ jets} \end{cases}$
- $H_T(all) > \begin{cases} 120 \text{ GeV} & 1, 2 \text{ jets} \\ 140 \text{ GeV} & 3 \text{ jets} \\ 160 \text{ GeV} & 4 \text{ jets} \end{cases}$
- leading jet with $p_T > 25$ GeV and $d\phi(\cancel{E}_T, leading - jet) < 1.5 + 0.0469 \cdot \cancel{E}_T$
- for electrons
 - $d\phi(\cancel{E}_T, lep) > 2.0 - 0.05 \cdot \cancel{E}_T$
 - $d\phi(\cancel{E}_T, lep) > 1.5 - 0.03 \cdot \cancel{E}_T$
 - $d\phi(\cancel{E}_T, lep) < 2.0 + 0.0476 \cdot \cancel{E}_T$
 - $|\eta| < 1.1$ (detector eta)

²one vertex originating from the hard interaction

³The exact cuts for the single top selection are being optimized while this thesis is written, the upcoming DØ note might contain the values for slightly different cuts.

- for muons
 - $d\phi(\cancel{E}_T, lep) > 1.2 - 0.0141 \cdot \cancel{E}_T$
 $d\phi(\cancel{E}_T, lep) < 2.5 + 0.0214 \cdot \cancel{E}_T$
 - $|\eta| < 2.0$ (detector eta)
 - $-\frac{b-a}{\pi-a} + \frac{b}{\pi-a} \cdot d\phi(\cancel{E}_T, lep) < |trkcurvsig|$ with $(a, b) = \left\{ \left(\frac{7}{8}\pi, cut \right), \left(2, \frac{cut}{2} \right) \right\}$,
 $cut = \begin{cases} 6 & \text{for RunIIa} \\ 4 & \text{for RunIIb} \end{cases}$

The different triangle cuts on $d\phi(\ell, \cancel{E}_T)$ and \cancel{E}_T are plotted in figure 5.3; the cut on track curvature significance used for the μ +jets channel in the single top selection cuts is plotted in figure 5.4.

5.2.3. Control Plots

To check for the agreement of data with Monte Carlo plots showing all Monte Carlo samples contributing to the final state that is looked at and the data for different variables are made. The single Monte Carlo samples are scaled using the theoretical cross-section and the integrated luminosity.

In addition all the W+jets samples are rescaled by the same factor S_W to match normalization of all Monte Carlo samples to data.

5.2.4. Fake Rate

The lepton fake rate and the real lepton efficiency is determined for every set of cuts separately using the method described in detail in section 3.

5.3. Variables

All of the over 200 variables are used to check for inconsistencies in the data Monte Carlo agreement. The following variables are used in the sample plots shown in the next section and in appendix B.

H_T the scalar sum of the p_T of all jets.

lepton p_T p_T of the reconstructed lepton

lepton η η of the reconstructed lepton. (Actually there are two definitions of lepton η ; the *physics* lepton η measured in relation to the position of the primary vertex and the *detector* lepton η measured in relation to the center of the detector. Whenever not explicitly mentioned otherwise lepton η refers to the physics lepton η .)

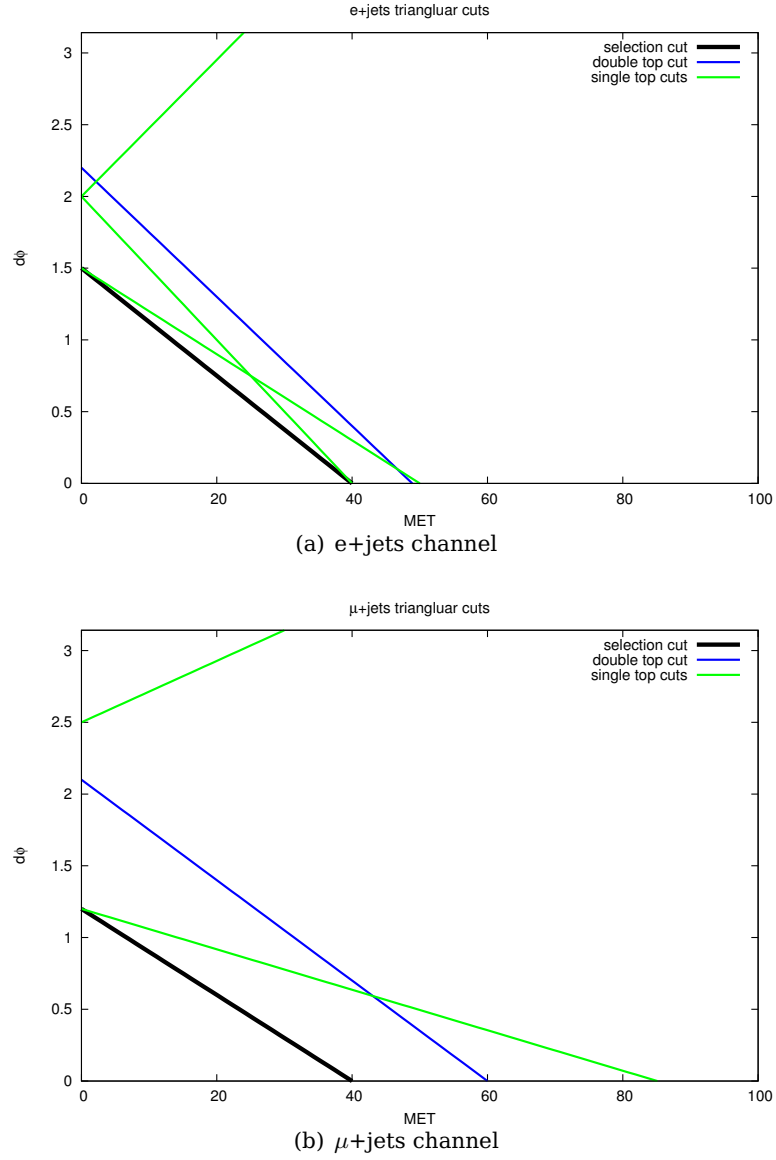


Figure 5.3: The different triangle cuts for the (a) e+jets and (b) μ +jets channel.

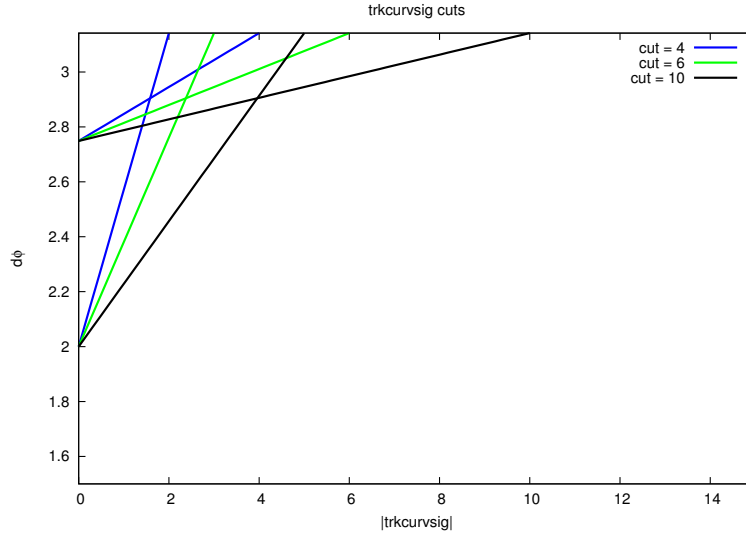


Figure 5.4: The trkcurvsig cut for different values of the parameter *cut*. Note the zero-suppressed scale.

***n*th jet p_T** p_T of the *n*th reconstructed jet (the jets are ordered according to p_T).

***n*th jet η** η of the *n*th reconstructed jet (the jets are ordered according to p_T).

\cancel{E}_T the missing transverse energy is the transverse component of the energy not measured by the detector but implied by energy and momentum conservation.

$d\phi(\ell, \cancel{E}_T)$ the angle between the lepton and the missing transverse energy (describing the isolation of the lepton).

trkcurvsig the track curvature is usually called κ and is a direct measurement of the bending extracted from the track fit. p_T can then be calculated by

$$p_T [\text{GeV}] = Q \cdot \frac{B_z [\text{T}]}{\kappa [\text{m}^{-1}]} \quad .$$

Here it is just

$$\kappa = \frac{Q}{p_T}$$

and the significance is given by dividing κ by the p_T resolution σ_{p_T} .

6. Results

One part of the result of this thesis are the complete TopTree samples produced with the new setup and various new or updated software tools which of course cannot be included in the thesis. The other part of the results mainly consists of all the values for ϵ_{QCD} and ϵ_{sig} as well as of control plots (most of them) showing a good agreement between data and Monte Carlo. In this section only a few important plots and some values for fake rate and efficiency are listed for the sake of readability. More plots and numbers are included in the appendix A and B and the whole set of important control plots (for all possible combinations of run period, analysis cuts, b-tagging, jet bins, and so on) and numbers can be found in the upcoming DØ selection note⁴.

6.1. Control Plots

Most of the control plots show a very good agreement between data and Monte Carlo. The jet bin exclusive plots for H_T of the whole RunIIb for the previous $t\bar{t}$ selection can be found in figure 6.1 for the e+jets and in figure 6.2 for the μ +jets channel.

The same plots for the analysis cuts used for single top can be found in figure 6.3 for the e+jets and in figure 6.4 for the μ +jets channel.

The total yields (including statistical errors) of RunIIb per jet bin for the standard $t\bar{t}$ selection cuts can be found in table 6.1 for e+jets and in table 6.2 for μ +jets. The Monte Carlo contributions are listed by the different input samples used. Dibosons are the events with two decaying bosons, either WW , WZ or ZZ . Wlp are events with a W boson and light partons while Wbb and Wcc contain events with a W boson and a $b\bar{b}$ (or $c\bar{c}$) pair; all three are W+jets events. The Z+jets events are subdivided in a similar way and even differentiated by lepton flavour. tb and tbq are the two single top production modes, $t\bar{t}bar$ the events with $t\bar{t}$ decaying through the lepton+jets channel, and $t\bar{t}ll$ the events with $t\bar{t}$ decaying through the dilepton channel. The tables for the single top selection cuts can be found in the appendix A.

6.1.1. Muon Track Curvature Significance

One of the largest discrepancies seen for the $t\bar{t}$ control plots are found in the higher jet bins of the $d\phi(\ell, \cancel{E}_T)$ distribution (i.e. at high $d\phi$) for the μ +jets channel (for the e+jets channel there are no such discrepancies). Figure 6.5 shows the jet bin exclusive histograms for RunIIb4 only, but the behavior is exactly the same in all the other run periods.

Adopting the cut on the muon track curvature significance used in the single top selection helps to get a much better agreement for these bins also for the $t\bar{t}$ selection

⁴DØ note 6308. Selection of $t\bar{t}bar$ Events in the Lepton+jets Channel with 9.7 fb⁻¹.

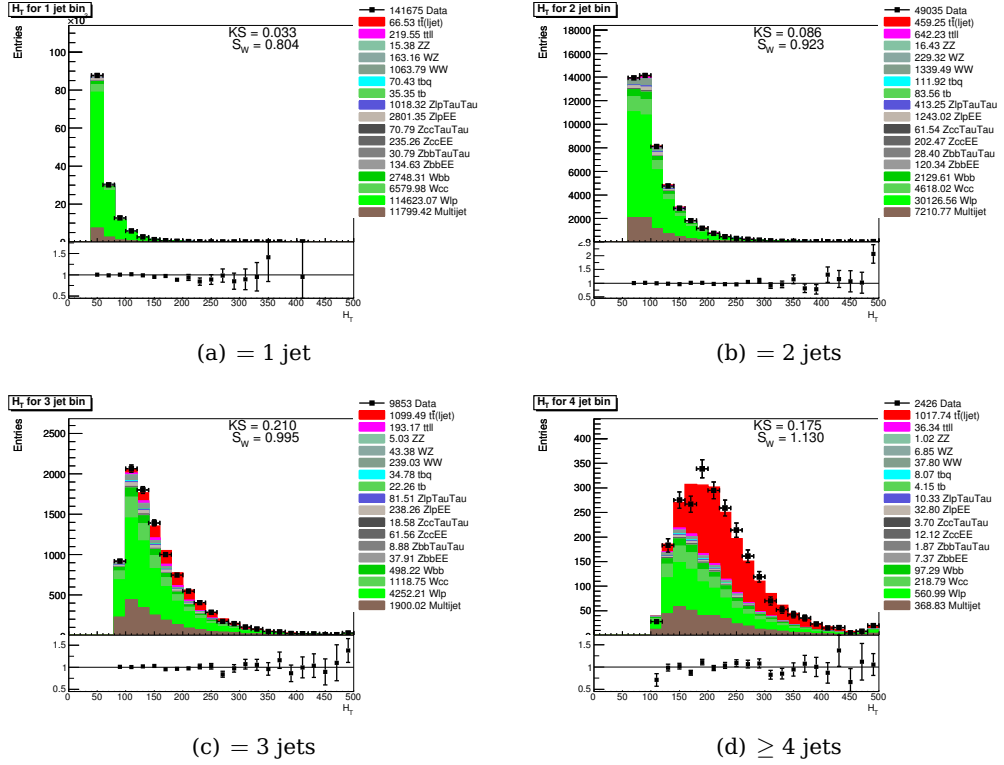


Figure 6.1: H_T plots for RunIIb e+jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets. The numbers on the right-hand side of the plot list the yields of the different samples.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

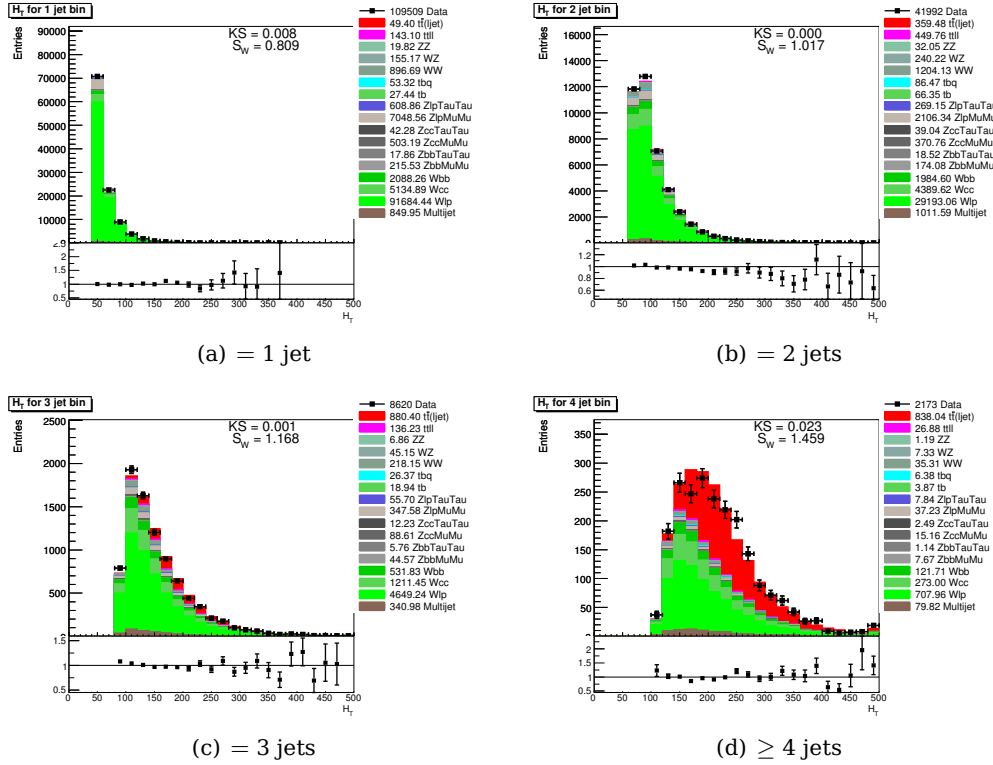


Figure 6.2: H_T plots for RunIIb μ +jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets.

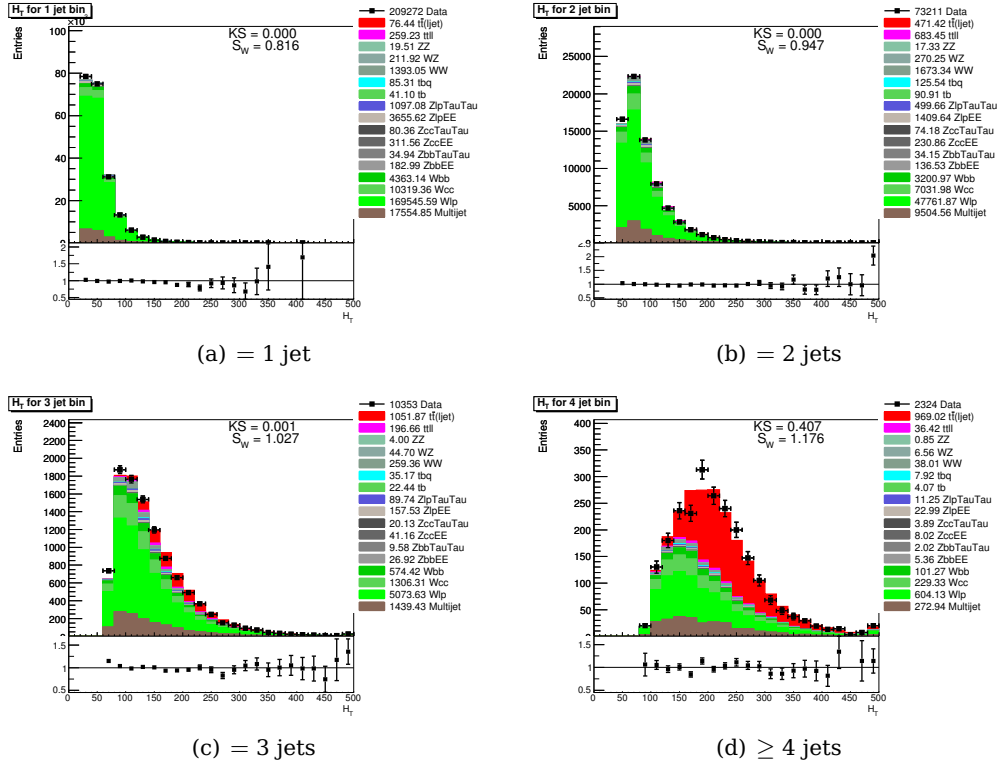


Figure 6.3: H_T plots for RunIIb e+jets using the single top cuts, before applying any b-tag for jets.

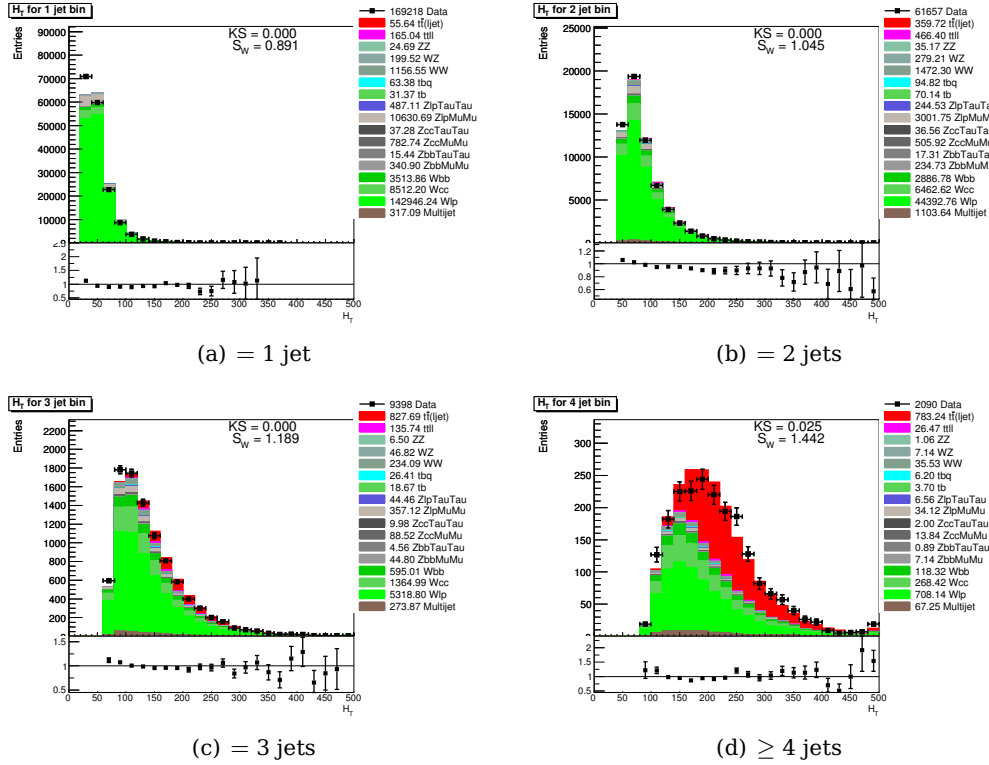


Figure 6.4: H_T plots for RunIIb μ +jets using the single top cuts, before applying any b-tag for jets.

Contribution	= 1 jet	= 2 jets	= 3 jets	≥ 4 jets
Data	141675.00 \pm 376.40	49035.00 \pm 221.44	9853.00 \pm 99.26	2426.00 \pm 49.25
Multijet	11799.42 \pm 219.39	7210.77 \pm 131.97	1900.02 \pm 45.84	368.83 \pm 17.29
WW	1063.80 \pm 3.13	1339.49 \pm 3.40	239.03 \pm 1.46	37.80 \pm 0.60
WZ	163.16 \pm 0.72	229.32 \pm 0.82	43.38 \pm 0.36	6.85 \pm 0.15
Wbb	2748.29 \pm 7.60	2129.59 \pm 7.01	498.22 \pm 3.31	97.29 \pm 1.55
Wcc	6579.96 \pm 11.46	4618.05 \pm 10.28	1118.75 \pm 5.04	218.79 \pm 2.35
Wlp	114623.73 \pm 84.52	30126.62 \pm 36.75	4252.19 \pm 12.88	560.99 \pm 4.35
ZZ	15.38 \pm 0.16	16.43 \pm 0.17	5.03 \pm 0.09	1.02 \pm 0.04
ZbEE	134.63 \pm 0.68	120.34 \pm 0.64	37.91 \pm 0.89	7.37 \pm 0.15
ZbbTauTau	30.79 \pm 0.36	28.40 \pm 0.35	8.88 \pm 0.19	1.87 \pm 0.09
ZcEE	235.26 \pm 1.26	202.47 \pm 1.16	61.56 \pm 0.62	12.12 \pm 0.28
ZccTauTau	70.79 \pm 0.74	61.54 \pm 0.68	18.58 \pm 0.37	3.70 \pm 0.16
ZlpEE	2801.35 \pm 10.64	1243.03 \pm 6.15	238.26 \pm 2.24	32.80 \pm 0.76
ZlpTauTau	1018.32 \pm 6.82	413.25 \pm 3.54	81.51 \pm 1.39	10.33 \pm 0.47
tb	35.35 \pm 0.24	83.56 \pm 0.36	22.26 \pm 0.19	4.15 \pm 0.08
tbq	70.43 \pm 0.50	111.92 \pm 0.65	34.78 \pm 0.38	8.07 \pm 0.20
ttbar	66.53 \pm 1.18	459.25 \pm 2.35	1099.48 \pm 3.36	1017.73 \pm 3.01
ttl	219.55 \pm 0.78	642.21 \pm 1.30	193.17 \pm 0.62	36.34 \pm 0.25
$\sum_i MC_i$	141676.74 \pm 235.88	49036.23 \pm 137.82	9853.02 \pm 48.23	2426.03 \pm 18.34

Table 6.1: Total yields for the $t\bar{t}$ selection cuts in RunIIb for the e+jets channel, before applying any b-tag for jets; the W+jets processes are scaled in order to match Monte Carlo to data.

Contribution	= 1 jet	= 2 jets	= 3 jets	≥ 4 jets
Data	109509.00 ± 330.92	41992.00 ± 204.92	8620.00 ± 92.84	2173.00 ± 46.62
Multijet	849.95 ± 114.60	1011.59 ± 129.03	340.98 ± 44.25	79.82 ± 11.50
WW	896.69 ± 2.68	1204.13 ± 3.01	218.15 ± 1.31	35.31 ± 0.54
WZ	155.17 ± 0.65	240.22 ± 0.78	45.15 ± 0.35	7.33 ± 0.14
Wbb	2088.23 ± 5.88	1984.55 ± 6.44	531.83 ± 3.46	121.71 ± 1.82
Wcc	5134.87 ± 9.05	4389.61 ± 9.49	1211.45 ± 5.31	273.00 ± 2.84
Wlp	91676.95 ± 68.64	29192.28 ± 35.58	4649.23 ± 13.35	707.96 ± 5.12
ZZ	19.82 ± 0.16	32.05 ± 0.20	6.86 ± 0.09	1.19 ± 0.04
ZbbMuMu	215.52 ± 0.74	174.08 ± 0.71	44.57 ± 0.37	7.67 ± 0.16
ZbbTauTau	17.86 ± 0.25	18.52 ± 0.27	5.76 ± 0.15	1.14 ± 0.06
ZccMuMu	503.19 ± 1.57	370.76 ± 1.43	88.61 ± 0.70	15.16 ± 0.30
ZccTauTau	42.28 ± 0.52	39.04 ± 0.51	12.23 ± 0.28	2.49 ± 0.13
ZlpMuMu	7048.27 ± 14.46	2106.34 ± 6.46	347.58 ± 2.40	37.23 ± 0.77
ZlpTauTau	608.86 ± 4.70	269.15 ± 2.62	55.70 ± 1.05	7.84 ± 0.39
tb	27.44 ± 0.27	66.35 ± 0.29	18.94 ± 0.16	3.87 ± 0.08
tbq	53.32 ± 0.37	86.47 ± 0.51	26.37 ± 0.29	6.38 ± 0.15
ttbar	49.40 ± 1.05	359.48 ± 1.83	880.39 ± 2.73	838.03 ± 2.61
tll	143.10 ± 0.59	449.75 ± 1.00	136.23 ± 0.48	26.88 ± 0.20
$\sum_i MC_i$	109530.90 ± 134.93	41994.39 ± 134.58	8620.02 ± 46.84	2173.02 ± 13.33

Table 6.2: Total yields for the $t\bar{t}$ selection cuts in RunIIb for the μ +jets channel, before applying any b-tag for jets; the W+jets processes are scaled in order to match Monte Carlo to data.

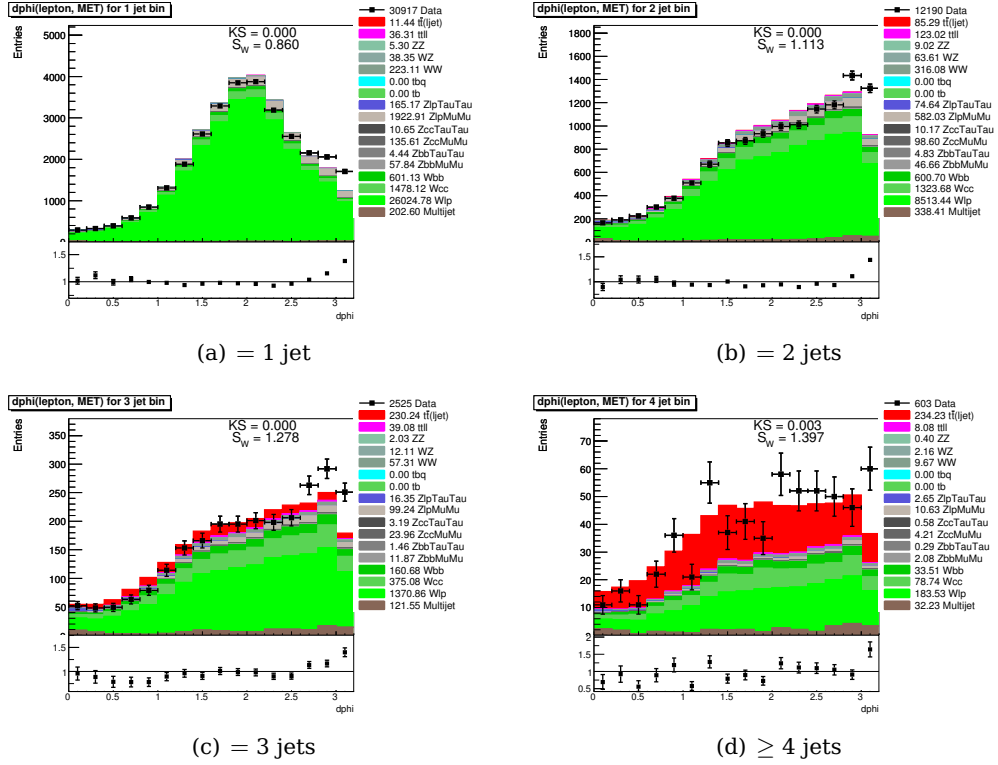


Figure 6.5: $d\phi(\ell, \cancel{E}_T)$ (isolation or “distance” between the lepton and the neutrino) plots for RunIIb4 μ +jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets.

Fake Rate for the lepton+jets Selection using the Full RunIIB Data Set

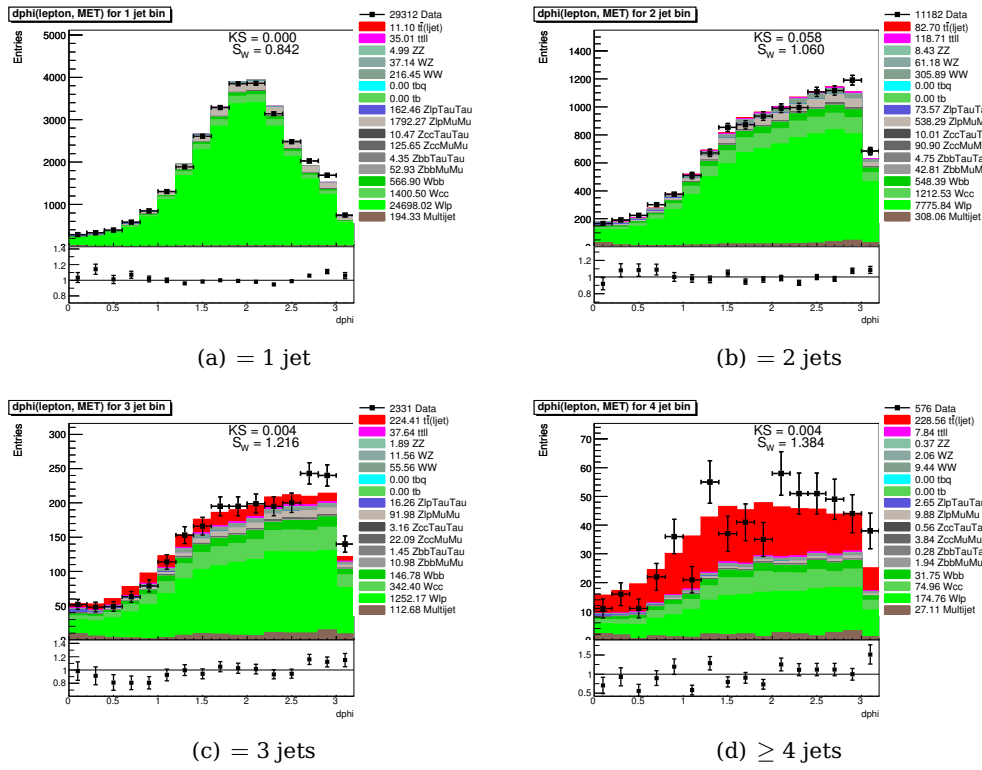


Figure 6.6: $d\phi(\ell, \cancel{E}_T)$ plots for RunIIB4 μ +jets using the $t\bar{t}$ selection cuts and the cut on track curvature significance (see equation 9) using $cut = 10$, before applying any b-tag for jets.

cuts by throwing away misreconstructed muons. The cut is defined as

$$-\frac{b \cdot a}{\pi - a} + \frac{b}{\pi - a} \cdot d\phi(\ell, \cancel{E}_T) < |\text{trkcurvsig}| \quad , \quad (9)$$

with $(a, b) = \left\{ \left(\frac{7}{8}\pi, cut \right), \left(2, \frac{cut}{2} \right) \right\}$ and $cut = 10$ (for a graphical representation of the cut also look at figure 5.4).

The jet bin exclusive plots for the $t\bar{t}$ selection cuts combined with this additional cut can be found in figure 6.6.

Looking at other plots produced including this new cut there is for example also a better description of the tails of the lepton η distribution (see figure 6.7). This clearly shows that the cut is useful not only by giving a better description in the obvious phase space (also see figure 6.8 for plots of the trkcurvsig itself) that is cut but also for other distributions. In addition the cut has a $t\bar{t}$ signal (generated with ALPGEN and PYTHIA) efficiency of about 97% in the inclusive 4 jet bin for RunIIB4 (used for most $t\bar{t}$ analyses) so that it can be added without losing too much statistics.

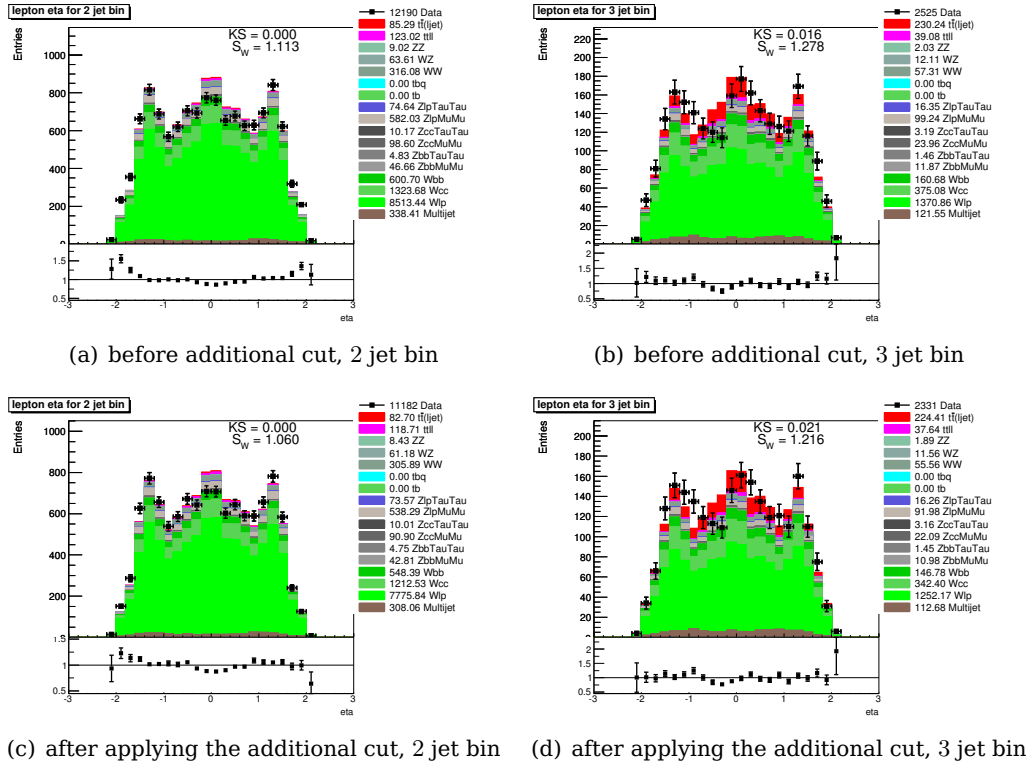


Figure 6.7: Lepton η plots for RunIIb4 μ +jets using the $t\bar{t}$ selection cuts. (a) and (b) show the plots before applying the additional cut for the 2 and 3 jet bin, respectively. (c) and (d) show the same plots with a cut on track curvature significance with $cut = 10$. All plots are before applying any b-tag for jets.

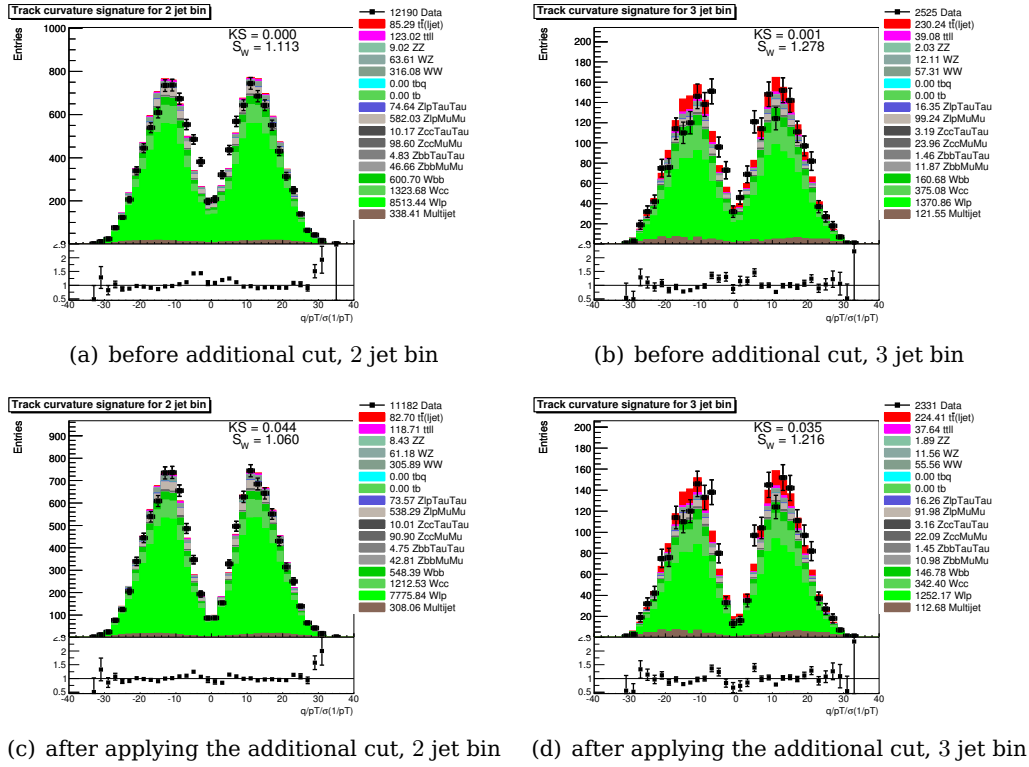


Figure 6.8: trkcurvsig plots for RunIIb4 μ +jets using the $t\bar{t}$ selection cuts. (a) and (b) show the plots before applying the additional cut for the 2 and 3 jet bin, respectively. (c) and (d) show the same plots with a cut on track curvature significance with $cut = 10$. All plots are before applying any b-tag for jets.

(a) e+jets				
	= 1 jet	= 2 jets	= 3 jets	≥ 4 jets
RunI Ib1	0.836	0.930	1.032	1.187
RunI Ib2	0.787	0.887	0.999	1.075
RunI Ib3	0.806	0.948	0.976	0.964
RunI Ib4	0.822	0.978	1.054	1.394

(b) μ +jets				
	= 1 jet	= 2 jets	= 3 jets	≥ 4 jets
RunI Ib1	0.670	0.841	0.990	1.357
RunI Ib2	0.816	1.034	1.165	1.716
RunI Ib3	0.855	1.027	1.184	1.269
RunI Ib4	0.860	1.113	1.278	1.397

Table 6.3: S_W scale factors for the $t\bar{t}$ selection cuts per run period and jet bin.

6.2. W Scale Factors S_W

The W scale factors applied to scale the sum of all W+jets Monte Carlo samples to data are listed in table 6.3.

The scale factors look reasonable for the e+jets channel and agree well with values derived in previous selections. For the μ +jets channel the values are higher by about 1.4 than in previous selections [33]. No conclusive reason could be found causing this discrepancy. The W_{p_T} reweighting is correctly applied and in general the same software is used to calculate those factors in both channels so that every severe problem should also show up in the e+jets channel. The highest value for the 4 jet bin in the RunI Ib2 μ +jets case is a well known and thoroughly investigated excess. It is consistent with previous selections. The only explanation left is that it was caused by a statistical fluctuation, which is confirmed by the fact that later run periods do not show that excess anymore.

6.3. Lepton Fake Rate & Real Lepton Efficiency

The final jet bin exclusive and inclusive values for ϵ_{QCD} and ϵ_{sig} can be found in table 6.4 (ϵ_{QCD}) and table 6.5 (ϵ_{sig}). The 3 and 4 jet bin are statistically very limited and thus show large fluctuations. The corresponding ratio plots can be found in appendix B. Table 6.6 summarizes the inclusive 2 jet bin values that are used as final values for ϵ_{QCD} and ϵ_{sig} in analyses. In general, the values are consistent for the different run periods with the exception of the RunI Ib3 e+jets value that is lower than the other ones. No conclusive reason was found for that fluctuation. Two plots were created in order to check if this lower value may be there due to an

error in the software or in the calculations. Both control plots (see figure 6.9) show RunIIb3 data and Monte Carlo. Plot (a) uses the correct value for ϵ_{QCD} as derived for RunIIb3 to scale the multijet background; plot (b) uses the value derived for RunIIb4 (RunIIb4 was chosen because it is most similar to RunIIb3 and because it had the second lowest ϵ_{QCD} value of all the run periods). The plot with the correct ϵ_{QCD} value clearly shows a better agreement between data and Monte Carlo; therefore this issue is likely caused by a statistical fluctuation or by a differently trained decision tree used for electron identification (these trees are trained separately for every run period) and it was not investigated further⁵. For the μ +jets channel no such check was necessary as all the differences can be explained by the statistical and systematical errors.

6.4. Fake Rate and Efficiency Systematic Errors

This study uses the same method to determine the errors for the lepton fake rates and the real lepton efficiencies that was used before in the 7.3 fb^{-1} lepton+jets selection.

6.4.1. Statistical Errors

The statistical errors of the calculated fake rates and efficiencies are taken from the fitting error obtained when fitting the ratio histogram to a constant. Of course this error strongly depends on how large the chosen range is where the function is fitted to data. Usually this range is chosen so that the fit only covers bins with large enough statistics (i.e. a reasonable amount of data in it even for the tight selection) and omits regions where there are known problems (e.g. due to detector acceptance).

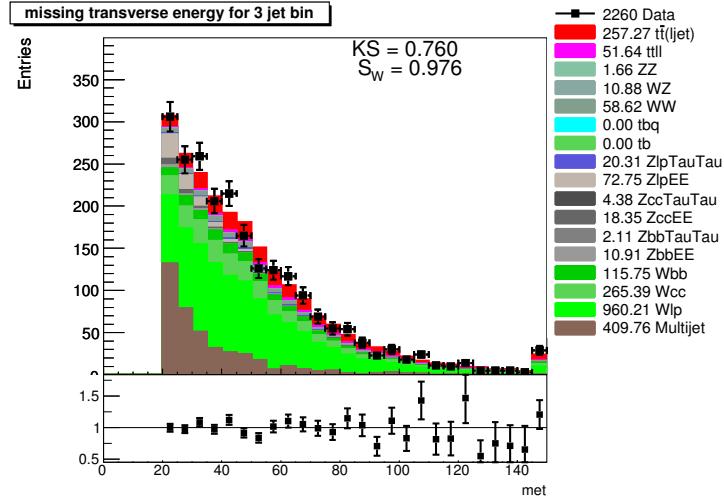
For the central values of ϵ_{QCD} lepton η is fitted over the range $[-1.4, 1.4]$ for the e+jets channel and over $[-2, 2]$ for the μ +jets channel due to known problems describing electrons in the intercyrostat (ICR) region. Deriving the central value for ϵ_{QCD} from \cancel{E}_T could be unstable because of the small fitting range caused by using the special samples.

The samples to fit ϵ_{sig} are taken from Monte Carlo so statistics is not an issue and therefore it is possible to fit over the whole available range for each variable in order not to introduce an additional manual tuning.

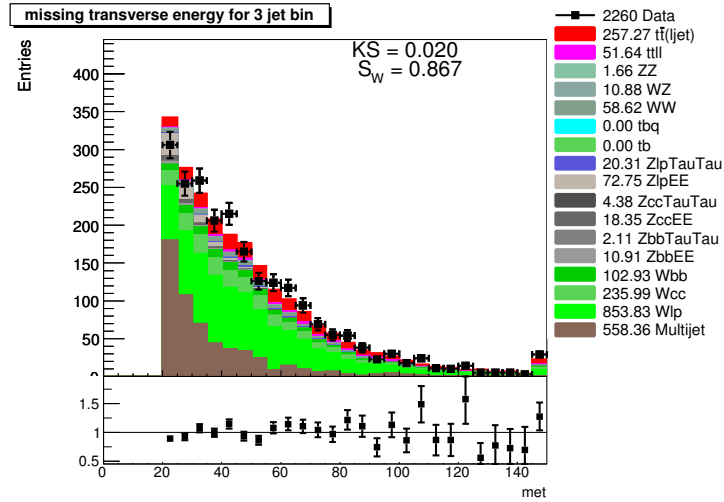
6.4.2. Systematic Errors

The systematic errors are estimated using different approaches. For ϵ_{QCD} the lepton fake rate values from three other variables (lepton p_T , \cancel{E}_T , M_T) are calculated in

⁵After the completion of the calculations for this thesis a bug was discovered in the jetID for RunIIb3 Monte Carlo that could affect the corresponding scale factor.



(a) using ϵ_{QCD} value derived for RunIIb3



(b) using ϵ_{QCD} value derived for RunIIb4

Figure 6.9: E_T plots for RunIIb3 e+jets channel. (a) uses the correct value of ϵ_{QCD} to scale the multijet background, (b) uses the ϵ_{QCD} value derived for RunIIb4.

(a) RunIIb1, e+jets				(b) RunIIb1, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.306	± 0.007	$+0.021$ -0.022	= 1 jet	0.019	± 0.037	$+0.133$ -0.077
= 2 jets	0.327	± 0.012	$+0.012$ -0.012	= 2 jets	0.121	± 0.028	$+0.042$ -0.043
= 3 jets	0.343	± 0.033	$+0.042$ -0.042	= 3 jets	0.089	± 0.066	$+0.357$ -0.357
≥ 4 jets	0.038	± 0.054	$+0.500$ -0.500	≥ 4 jets	0.544	± 0.178	$+0.505$ -0.505
≥ 2 jets	0.331	± 0.011	$+0.014$ -0.014	≥ 2 jets	0.128	± 0.026	$+0.036$ -0.038
(c) RunIIb2, e+jets				(d) RunIIb2, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.297	± 0.004	$+0.012$ -0.012	= 1 jet	0.138	± 0.017	$+0.079$ -0.086
= 2 jets	0.322	± 0.007	$+0.006$ -0.006	= 2 jets	0.175	± 0.017	$+0.033$ -0.034
= 3 jets	0.333	± 0.017	$+0.011$ -0.011	= 3 jets	0.136	± 0.039	$+0.060$ -0.061
≥ 4 jets	0.328	± 0.049	$+0.077$ -0.077	≥ 4 jets	0.169	± 0.081	$+0.232$ -0.231
≥ 2 jets	0.324	± 0.006	$+0.007$ -0.007	≥ 2 jets	0.175	± 0.016	$+0.030$ -0.031
(e) RunIIb3, e+jets				(f) RunIIb3, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.221	± 0.004	$+0.021$ -0.022	= 1 jet	0.092	± 0.019	$+0.064$ -0.071
= 2 jets	0.251	± 0.007	$+0.010$ -0.010	= 2 jets	0.182	± 0.020	$+0.041$ -0.041
= 3 jets	0.271	± 0.018	$+0.008$ -0.008	= 3 jets	0.033	± 0.028	$+0.123$ -0.123
≥ 4 jets	0.214	± 0.046	$+0.030$ -0.030	≥ 4 jets	0.070	± 0.106	$+0.302$ -0.302
≥ 2 jets	0.254	± 0.006	$+0.009$ -0.009	≥ 2 jets	0.173	± 0.017	$+0.039$ -0.040
(g) RunIIb4, e+jets				(h) RunIIb4, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.302	± 0.004	$+0.017$ -0.017	= 1 jet	0.200	± 0.022	$+0.083$ -0.091
= 2 jets	0.309	± 0.007	$+0.011$ -0.011	= 2 jets	0.220	± 0.020	$+0.056$ -0.057
= 3 jets	0.347	± 0.018	$+0.018$ -0.018	= 3 jets	0.188	± 0.040	$+0.029$ -0.030
≥ 4 jets	0.263	± 0.043	$+0.020$ -0.020	≥ 4 jets	0.055	± 0.072	$+0.211$ -0.211
≥ 2 jets	0.313	± 0.006	$+0.010$ -0.010	≥ 2 jets	0.218	± 0.018	$+0.054$ -0.054

Table 6.4: Jet bin dependent ϵ_{QCD} values for the $t\bar{t}$ selection cuts for all four run periods. The 4 jet bin usually suffers from low statistics and thus the fits do not converge.

(a) RunIIb1, e+jets				(b) RunIIb1, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.887	± 0.002	± 0.067	= 1 jet	0.932	± 0.002	± 0.043
= 2 jets	0.876	± 0.002	± 0.023	= 2 jets	0.916	± 0.002	± 0.023
= 3 jets	0.868	± 0.004	± 0.005	= 3 jets	0.901	± 0.004	± 0.004
≥ 4 jets	0.870	± 0.005	± 0.009	≥ 4 jets	0.888	± 0.005	± 0.005
≥ 2 jets	0.874	± 0.002	± 0.004	≥ 2 jets	0.913	± 0.002	± 0.018
(c) RunIIb2, e+jets				(d) RunIIb2, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.893	± 0.002	± 0.099	= 1 jet	0.917	± 0.002	± 0.042
= 2 jets	0.886	± 0.003	± 0.034	= 2 jets	0.896	± 0.003	± 0.017
= 3 jets	0.878	± 0.006	± 0.002	= 3 jets	0.878	± 0.006	± 0.009
≥ 4 jets	0.880	± 0.009	± 0.009	≥ 4 jets	0.874	± 0.010	± 0.012
≥ 2 jets	0.884	± 0.003	± 0.009	≥ 2 jets	0.893	± 0.003	± 0.011
(e) RunIIb3, e+jets				(f) RunIIb3, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.888	± 0.002	± 0.091	= 1 jet	0.919	± 0.002	± 0.067
= 2 jets	0.882	± 0.003	± 0.008	= 2 jets	0.900	± 0.003	± 0.022
= 3 jets	0.880	± 0.006	± 0.010	= 3 jets	0.886	± 0.006	± 0.007
≥ 4 jets	0.884	± 0.007	± 0.011	≥ 4 jets	0.876	± 0.007	± 0.009
≥ 2 jets	0.882	± 0.003	± 0.006	≥ 2 jets	0.897	± 0.003	± 0.013
(g) RunIIb4, e+jets				(h) RunIIb4, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.889	± 0.002	± 0.095	= 1 jet	0.915	± 0.002	± 0.073
= 2 jets	0.883	± 0.003	± 0.011	= 2 jets	0.896	± 0.003	± 0.023
= 3 jets	0.881	± 0.006	± 0.009	= 3 jets	0.884	± 0.006	± 0.008
≥ 4 jets	0.885	± 0.007	± 0.012	≥ 4 jets	0.874	± 0.008	± 0.011
≥ 2 jets	0.883	± 0.003	± 0.005	≥ 2 jets	0.893	± 0.003	± 0.012

Table 6.5: Jet bin dependent ϵ_{sig} values for the $t\bar{t}$ selection cuts for all four run periods.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

(a) ϵ_{QCD} , inclusive 2 jet bin, $t\bar{t}$ selection cuts

	e+jets	μ+jets
RunIIb1	$0.331 \pm 0.011(\text{stat.}) \pm_{0.014}^{0.014}(\text{syst.})$	$0.128 \pm 0.026(\text{stat.}) \pm_{0.038}^{0.036}(\text{syst.})$
RunIIb2	$0.324 \pm 0.006(\text{stat.}) \pm_{0.007}^{0.007}(\text{syst.})$	$0.175 \pm 0.016(\text{stat.}) \pm_{0.031}^{0.030}(\text{syst.})$
RunIIb3	$0.254 \pm 0.006(\text{stat.}) \pm_{0.009}^{0.009}(\text{syst.})$	$0.173 \pm 0.017(\text{stat.}) \pm_{0.040}^{0.039}(\text{syst.})$
RunIIb4	$0.313 \pm 0.006(\text{stat.}) \pm_{0.010}^{0.010}(\text{syst.})$	$0.218 \pm 0.018(\text{stat.}) \pm_{0.054}^{0.054}(\text{syst.})$

(b) ϵ_{sig} , inclusive 2 jet bin, $t\bar{t}$ selection cuts

	e+jets	μ+jets
RunIIb1	$0.874 \pm 0.002(\text{stat.}) \pm 0.004(\text{syst.})$	$0.913 \pm 0.002(\text{stat.}) \pm 0.018(\text{syst.})$
RunIIb2	$0.884 \pm 0.003(\text{stat.}) \pm 0.009(\text{syst.})$	$0.893 \pm 0.003(\text{stat.}) \pm 0.011(\text{syst.})$
RunIIb3	$0.882 \pm 0.003(\text{stat.}) \pm 0.006(\text{syst.})$	$0.897 \pm 0.003(\text{stat.}) \pm 0.013(\text{syst.})$
RunIIb4	$0.883 \pm 0.003(\text{stat.}) \pm 0.005(\text{syst.})$	$0.893 \pm 0.003(\text{stat.}) \pm 0.012(\text{syst.})$

Table 6.6: Lepton fake rate ϵ_{QCD} and real lepton efficiency ϵ_{sig} for the $t\bar{t}$ selection cuts derived from the inclusive 2 jet bin.

addition to the central value derived from lepton η . The largest difference from any of those values to the central one is the first part of the systematic uncertainty. The second part of the error is derived by varying different parameters. The $t\bar{t}$ measured production cross section $\sigma_{t\bar{t}} = 7.78 \text{ pb}$ [34] is varied by 10%, the light flavour k -factors by 20% and the heavy flavour k -factors by 15% (for W+jets and Z+jets separately) and all the differences are summed up in quadrature. The last part is taken from the difference between the same central value but derived from a sample with $\cancel{E}_T < 15 \text{ GeV}$ (rather than the $\cancel{E}_T < 10 \text{ GeV}$ used normally). The list of the other applied cuts can be found on page 34.

So, in summary there are the following components contributing to the systematic uncertainty:

δ_{var} : biggest difference to value derived from other variable (lepton p_T , \cancel{E}_T , M_T)

δ_{par} : combination from results obtained by varying the k -factors and $\sigma_{t\bar{t}}$

$\delta_{\cancel{E}_T}$: value derived for $\cancel{E}_T < 15 \text{ GeV}$

These three terms are again added in quadrature and yield the final systematic error of the ϵ_{QCD} value.

$$\delta_{\text{syst}} = \sqrt{\delta_{\text{var}}^2 + \delta_{\text{par}}^2 + \delta_{\cancel{E}_T}^2} \quad .$$

This is not perfect from the point of statistics as the values derived for the different \cancel{E}_T cuts are slightly correlated and the difference should not just be added in

quadrature. However, the correlation is small and the resulting systematic error is in a reasonable range where it can explain for example the fluctuations between the values derived for different jet bins. Without adding $\delta_{\cancel{E}_T}$ these values often showed a statistically significant difference.

The resulting systematic uncertainty is asymmetric because the values derived by varying the k -factors and the production cross section do not differ symmetrically from the central value for positive and negative variations. All the different results can be found in tables 6.7 and 6.8 while tables 6.9 and 6.10 list the values derived from other variables.

The systematic error for the ϵ_{sig} is derived in a similar way. Here the central value is derived from \cancel{E}_T by fitting the whole range. The systematic error is calculated by adding

δ_{var} : biggest difference to value derived from other variable (lepton p_T , lepton η)

δ_{par} : biggest difference to value derived with only W+jets or only $t\bar{t}$ MC samples

in quadrature

$$\delta_{\text{syst}} = \sqrt{\delta_{\text{var}}^2 + \delta_{\text{par}}^2} \quad .$$

Here of course the systematic uncertainty is symmetric.

(a) RunIIb1, e+jets

	Cen	\cancel{E}_T	$-k_{lp, var}^W$	$-k_{hf, var}^W$	$-k_{lp, var}^Z$	$-k_{hf, var}^Z$	$-\sigma_{\bar{t}t, var}$	$+k_{lp, var}^W$	$+k_{hf, var}^W$	$+k_{lp, var}^Z$	$+k_{hf, var}^Z$	$+\sigma_{\bar{t}t, var}$	δ_{var}	δ_{par}
= 1 jet	0.306	0.290	0.315	0.306	0.312	0.306	0.306	0.296	0.305	0.299	0.305	0.306	+0.019	0.009
= 2 jets	0.327	0.327	0.332	0.328	0.331	0.328	0.327	0.323	0.327	0.323	0.327	0.327	+0.019	0.011
= 3 jets	0.343	0.321	0.347	0.344	0.347	0.344	0.343	0.339	0.342	0.339	0.342	0.343	+0.006	0.035
≥ 4 jets	0.038	0.376	0.035	0.036	0.029	0.037	0.034	0.041	0.040	0.047	0.039	0.041	+0.023	0.368
≥ 2 jets	0.331	0.328	0.335	0.331	0.335	0.331	0.331	0.326	0.330	0.327	0.330	0.331	+0.338	0.012

(b) RunIIb1, μ +jets

	Cen	\cancel{E}_T	$-k_{lp, var}^W$	$-k_{hf, var}^W$	$-k_{lp, var}^Z$	$-k_{hf, var}^Z$	$-\sigma_{\bar{t}t, var}$	$+k_{lp, var}^W$	$+k_{hf, var}^W$	$+k_{lp, var}^Z$	$+k_{hf, var}^Z$	$+\sigma_{\bar{t}t, var}$	δ_{var}	δ_{par}
= 1 jet	0.019	< 0	0.124	0.026	0.050	0.022	0.019	< 0	0.013	< 0	0.017	0.019	+0.109	0.077
= 2 jets	0.121	0.110	0.150	0.126	0.133	0.124	0.122	0.091	0.117	0.109	0.119	0.121	+0.007	0.025
= 3 jets	0.089	0.099	0.111	0.095	0.099	0.092	0.091	0.068	0.084	0.079	0.086	0.088	+0.035	0.356
≥ 4 jets	0.544	0.542	0.536	0.542	0.536	0.540	0.537	0.547	0.545	0.552	0.547	0.551	+0.026	0.505
≥ 2 jets	0.128	0.117	0.157	0.133	0.141	0.131	0.129	0.097	0.123	0.115	0.125	0.127	+0.011	0.012

(c) RunIIb2, e+jets

	Cen	\cancel{E}_T	$-k_{lp, var}^W$	$-k_{hf, var}^W$	$-k_{lp, var}^Z$	$-k_{hf, var}^Z$	$-\sigma_{\bar{t}t, var}$	$+k_{lp, var}^W$	$+k_{hf, var}^W$	$+k_{lp, var}^Z$	$+k_{hf, var}^Z$	$+\sigma_{\bar{t}t, var}$	δ_{var}	δ_{par}
= 1 jet	0.297	0.289	0.303	0.297	0.301	0.297	0.297	0.290	0.297	0.292	0.297	0.297	+0.011	0.005
= 2 jets	0.322	0.319	0.326	0.323	0.325	0.323	0.322	0.319	0.322	0.319	0.322	0.322	+0.011	0.003
= 3 jets	0.333	0.326	0.336	0.334	0.336	0.334	0.334	0.330	0.333	0.331	0.333	0.333	+0.006	0.008
≥ 4 jets	0.328	0.327	0.330	0.329	0.330	0.329	0.330	0.326	0.327	0.326	0.327	0.326	+0.008	0.076
≥ 2 jets	0.324	0.321	0.328	0.325	0.327	0.325	0.324	0.321	0.324	0.321	0.324	0.324	+0.004	0.004

(d) RunIIb2, μ +jets

	Cen	\cancel{E}_T	$-k_{lp, var}^W$	$-k_{hf, var}^W$	$-k_{lp, var}^Z$	$-k_{hf, var}^Z$	$-\sigma_{\bar{t}t, var}$	$+k_{lp, var}^W$	$+k_{hf, var}^W$	$+k_{lp, var}^Z$	$+k_{hf, var}^Z$	$+\sigma_{\bar{t}t, var}$	δ_{var}	δ_{par}
= 1 jet	0.138	0.104	0.200	0.142	0.156	0.140	0.138	0.068	0.135	0.120	0.137	0.138	+0.073	0.031
= 2 jets	0.175	0.189	0.196	0.178	0.184	0.177	0.175	0.152	0.171	0.165	0.173	0.175	+0.081	0.018
= 3 jets	0.136	0.143	0.153	0.140	0.144	0.138	0.137	0.118	0.131	0.127	0.133	0.134	+0.028	0.057
≥ 4 jets	0.169	0.139	0.156	0.162	0.158	0.160	0.158	0.179	0.175	0.179	0.178	0.180	+0.021	0.229
≥ 2 jets	0.175	0.186	0.195	0.178	0.184	0.176	0.175	0.153	0.171	0.165	0.173	0.174	+0.036	0.016

Table 6.7: Results of systematic error analysis for ϵ_{QCD} using $\bar{t}\bar{t}$ selection cuts - part 1.

(a) RunIIb3, e+jets

	Cen	\mathcal{E}_T	$-k_{lp,var}^W$	$-k_{hf,var}^W$	$-k_{lp,var}^Z$	$-k_{hf,var}^Z$	$-\sigma_{\bar{t}t,var}$	$+k_{lp,var}^W$	$+k_{hf,var}^W$	$+k_{lp,var}^Z$	$+k_{hf,var}^Z$	$+\sigma_{\bar{t}t,var}$	δ_{var}	δ_{par}
= 1 jet	0.221	0.208	0.228	0.222	0.236	0.222	0.221	0.215	0.221	0.206	0.221	0.221	+0.021	0.005
= 2 jets	0.251	0.254	0.255	0.252	0.259	0.252	0.251	0.248	0.251	0.243	0.250	0.251	+0.022	0.001
= 3 jets	0.271	0.274	0.273	0.271	0.277	0.272	0.271	0.268	0.270	0.264	0.269	0.270	+0.010	0.003
≥ 4 jets	0.214	0.217	0.216	0.214	0.218	0.215	0.215	0.211	0.213	0.209	0.212	0.212	+0.008	0.029
≥ 2 jets	0.254	0.258	0.257	0.255	0.262	0.255	0.254	0.251	0.254	0.246	0.253	0.254	+0.007	0.001
													-0.009	

(b) RunIIb3, μ +jets

	Cen	\mathcal{E}_T	$-k_{lp,var}^W$	$-k_{hf,var}^W$	$-k_{lp,var}^Z$	$-k_{hf,var}^Z$	$-\sigma_{\bar{t}t,var}$	$+k_{lp,var}^W$	$+k_{hf,var}^W$	$+k_{lp,var}^Z$	$+k_{hf,var}^Z$	$+\sigma_{\bar{t}t,var}$	δ_{var}	δ_{par}
= 1 jet	0.092	0.085	0.151	0.095	0.110	0.093	0.092	0.025	0.088	0.072	0.090	0.092	+0.063	0.014
= 2 jets	0.182	0.196	0.202	0.185	0.191	0.184	0.182	0.161	0.179	0.173	0.180	0.182	+0.027	0.030
= 3 jets	0.033	0.115	0.034	0.032	0.034	0.033	0.033	0.030	0.034	0.031	0.033	0.032	+0.028	0.092
≥ 4 jets	0.070	0.330	0.053	0.064	0.069	0.069	0.069	0.086	0.077	0.071	0.072	0.072	+0.082	0.153
≥ 2 jets	0.173	0.186	0.192	0.177	0.182	0.175	0.174	0.153	0.170	0.164	0.171	0.173	+0.261	0.030
													-0.026	

(c) RunIIb4, e+jets

	Cen	\mathcal{E}_T	$-k_{lp,var}^W$	$-k_{hf,var}^W$	$-k_{lp,var}^Z$	$-k_{hf,var}^Z$	$-\sigma_{\bar{t}t,var}$	$+k_{lp,var}^W$	$+k_{hf,var}^W$	$+k_{lp,var}^Z$	$+k_{hf,var}^Z$	$+\sigma_{\bar{t}t,var}$	δ_{var}	δ_{par}
= 1 jet	0.302	0.292	0.307	0.302	0.313	0.302	0.302	0.297	0.302	0.290	0.301	0.302	+0.016	0.005
= 2 jets	0.309	0.308	0.311	0.309	0.315	0.309	0.309	0.306	0.308	0.302	0.308	0.308	+0.007	0.007
= 3 jets	0.347	0.333	0.349	0.347	0.352	0.348	0.347	0.345	0.346	0.342	0.346	0.346	+0.015	0.011
≥ 4 jets	0.263	0.270	0.264	0.263	0.267	0.264	0.264	0.261	0.262	0.258	0.261	0.261	+0.009	0.018
≥ 2 jets	0.313	0.311	0.316	0.314	0.320	0.314	0.313	0.311	0.313	0.307	0.312	0.313	+0.007	0.007
													-0.008	

(d) RunIIb4, μ +jets

	Cen	\mathcal{E}_T	$-k_{lp,var}^W$	$-k_{hf,var}^W$	$-k_{lp,var}^Z$	$-k_{hf,var}^Z$	$-\sigma_{\bar{t}t,var}$	$+k_{lp,var}^W$	$+k_{hf,var}^W$	$+k_{lp,var}^Z$	$+k_{hf,var}^Z$	$+\sigma_{\bar{t}t,var}$	δ_{var}	δ_{par}
= 1 jet	0.200	< 0	0.262	0.204	0.220	0.202	0.200	0.128	0.196	0.179	0.199	0.200	+0.065	0.051
= 2 jets	0.220	< 0	0.240	0.223	0.229	0.221	0.220	0.198	0.216	0.211	0.218	0.220	+0.023	0.052
= 3 jets	0.188	< 0	0.201	0.192	0.195	0.190	0.189	0.174	0.184	0.181	0.186	0.187	+0.024	0.025
≥ 4 jets	0.055	< 0	0.054	0.053	0.055	0.054	0.049	0.056	0.057	0.055	0.056	0.061	+0.016	0.211
≥ 2 jets	0.218	< 0	0.237	0.222	0.227	0.220	0.219	0.198	0.214	0.209	0.216	0.218	+0.007	0.049
													+0.096	
													-0.021	
													-0.022	

Table 6.8: Results of systematic error analysis for ϵ_{QCD} using tt selection cuts – part 2.

(a) RunIIb1, e+jets				
	Cen	\cancel{E}_T	M_T	lepton p_T
= 1 jet	0.306 ± 0.007	0.306 ± 0.007	0.304 ± 0.007	0.296 ± 0.007
= 2 jets	0.327 ± 0.012	0.328 ± 0.012	0.325 ± 0.012	0.317 ± 0.012
= 3 jets	0.343 ± 0.033	0.345 ± 0.034	0.334 ± 0.033	0.308 ± 0.033
≥ 4 jets	0.038 ± 0.054	0.406 ± 0.100	0.323 ± 0.091	0.234 ± 0.088
≥ 2 jets	0.331 ± 0.011	0.331 ± 0.011	0.327 ± 0.011	0.318 ± 0.011
(b) RunIIb1, μ +jets				
	Cen	\cancel{E}_T	M_T	lepton p_T
= 1 jet	0.019 ± 0.037	0.019 ± 0.036	0.096 ± 0.025	0.092 ± 0.016
= 2 jets	0.121 ± 0.028	0.138 ± 0.028	0.146 ± 0.023	0.144 ± 0.021
= 3 jets	0.089 ± 0.066	0.102 ± 0.067	0.162 ± 0.049	0.445 ± 0.038
≥ 4 jets	0.544 ± 0.178	0.746 ± 0.112	0.329 ± 0.133	0.039 ± 0.017
≥ 2 jets	0.128 ± 0.026	0.136 ± 0.025	0.139 ± 0.021	0.140 ± 0.019
(c) RunIIb2, e+jets				
	Cen	\cancel{E}_T	M_T	lepton p_T
= 1 jet	0.297 ± 0.004	0.297 ± 0.004	0.296 ± 0.004	0.292 ± 0.004
= 2 jets	0.322 ± 0.007	0.323 ± 0.007	0.322 ± 0.007	0.320 ± 0.007
= 3 jets	0.333 ± 0.017	0.333 ± 0.017	0.331 ± 0.017	0.326 ± 0.018
≥ 4 jets	0.328 ± 0.049	0.342 ± 0.050	0.347 ± 0.049	0.251 ± 0.049
≥ 2 jets	0.324 ± 0.006	0.325 ± 0.006	0.324 ± 0.006	0.321 ± 0.006
(d) RunIIb2, μ +jets				
	Cen	\cancel{E}_T	M_T	lepton p_T
= 1 jet	0.138 ± 0.017	0.149 ± 0.018	0.147 ± 0.014	0.107 ± 0.009
= 2 jets	0.175 ± 0.017	0.180 ± 0.017	0.175 ± 0.015	0.156 ± 0.013
= 3 jets	0.136 ± 0.039	0.154 ± 0.039	0.192 ± 0.036	0.167 ± 0.032
≥ 4 jets	0.169 ± 0.081	0.140 ± 0.108	0.397 ± 0.094	0.285 ± 0.075
≥ 2 jets	0.175 ± 0.016	0.179 ± 0.015	0.178 ± 0.014	0.158 ± 0.012

Table 6.9: Values for ϵ_{QCD} derived from different variables using $t\bar{t}$ selection cuts – part 1. The value with the largest absolute difference from the central value and thus yielding δ_{var} is set in bold.

(a) RunIIb3, e+jets				
	Cen	\cancel{E}_T	M_T	lepton p_T
= 1 jet	0.221 ± 0.004	0.222 ± 0.004	0.221 ± 0.004	0.217 ± 0.004
= 2 jets	0.251 ± 0.007	0.252 ± 0.007	0.253 ± 0.007	0.250 ± 0.007
= 3 jets	0.271 ± 0.018	0.272 ± 0.018	0.270 ± 0.018	0.268 ± 0.019
≥ 4 jets	0.214 ± 0.046	0.242 ± 0.048	0.237 ± 0.047	0.235 ± 0.049
≥ 2 jets	0.254 ± 0.006	0.254 ± 0.006	0.255 ± 0.006	0.254 ± 0.007
(b) RunIIb3, μ +jets				
	Cen	\cancel{E}_T	M_T	lepton p_T
= 1 jet	0.092 ± 0.019	0.106 ± 0.019	0.098 ± 0.014	0.083 ± 0.010
= 2 jets	0.182 ± 0.020	0.189 ± 0.020	0.173 ± 0.017	0.151 ± 0.015
= 3 jets	0.033 ± 0.028	0.110 ± 0.034	0.124 ± 0.029	0.110 ± 0.027
≥ 4 jets	0.070 ± 0.106	0.100 ± 0.121	0.071 ± 0.078	0.224 ± 0.054
≥ 2 jets	0.173 ± 0.017	0.178 ± 0.017	0.162 ± 0.015	0.143 ± 0.013
(c) RunIIb4, e+jets				
	Cen	\cancel{E}_T	M_T	lepton p_T
= 1 jet	0.302 ± 0.004	0.302 ± 0.004	0.301 ± 0.004	0.296 ± 0.004
= 2 jets	0.309 ± 0.007	0.309 ± 0.007	0.306 ± 0.007	0.301 ± 0.007
= 3 jets	0.347 ± 0.018	0.349 ± 0.018	0.344 ± 0.018	0.336 ± 0.018
≥ 4 jets	0.263 ± 0.043	0.253 ± 0.043	0.272 ± 0.044	0.281 ± 0.047
≥ 2 jets	0.313 ± 0.006	0.314 ± 0.006	0.311 ± 0.006	0.306 ± 0.006
(d) RunIIb4, μ +jets				
	Cen	\cancel{E}_T	M_T	lepton p_T
= 1 jet	0.200 ± 0.022	0.201 ± 0.022	0.185 ± 0.017	0.149 ± 0.013
= 2 jets	0.220 ± 0.020	0.220 ± 0.020	0.199 ± 0.017	0.168 ± 0.015
= 3 jets	0.188 ± 0.040	0.179 ± 0.039	0.164 ± 0.034	0.163 ± 0.031
≥ 4 jets	0.055 ± 0.072	0.198 ± 0.088	0.210 ± 0.075	0.266 ± 0.069
≥ 2 jets	0.218 ± 0.018	0.218 ± 0.018	0.195 ± 0.015	0.169 ± 0.013

Table 6.10: Values for ϵ_{QCD} derived from different variables using $t\bar{t}$ selection cuts – part 2. The value with the largest absolute difference from the central value and thus yielding δ_{var} is set in bold.

7. Conclusion

In conclusion the project was successful in combining the single top and $t\bar{t}$ lepton+jets selection and thus in saving future manpower both in doing the selection as well as in developing the tools needed to further process the data and prepare it for final analysis and last but not least in the effort it takes to determine the correct shape and scale of the multijet background. Even more manpower can be saved using the new scripts to produce the selection that now have automated error-recovery capabilities. By loosening the selection cuts both of the old selections can be reproduced with the full RunII data set thus allowing for repeating every analysis based on one of these previous selections with all data available. Also, there is the chance to do completely new analyses based on this selection because it covers more phase space than the previous ones.

In addition the extended storage format used for this initial selection allows to save various additional information as for example the inclusion of two tight working points, correction factors for multiple trigger definitions, additional events with only EC electrons or GoodJets (including the necessary corrections) and so on. Hereby the new format makes it possible that the selection can be used for many different kinds of analyses without rerunning over all the data and Monte Carlo samples.

The new software versions used performed as expected and especially the new multivariate electron identification performs very well.

The matrix method used to determine the scale of the multijet background by calculating the lepton fake rate and the real lepton efficiency also works with the new electron identification and gives reasonable results that are consistent when comparing over different run periods or different analysis cuts. Also the statistical and systematical errors for the values are in a reasonable range as they explain the fluctuations for values derived for different run periods or jet bins. As in the previous selection it is only feasible to derive one value for all the jet bins because the higher jet bins suffer from very low statistics.

The values for ϵ_{sig} and ϵ_{QCD} presented here were derived for the cuts needed to produce the new single top selection or for the cuts to reproduce the previous $t\bar{t}$ selection. The software tools were developed and set up in a way that makes it very easy, albeit a little time consuming due to the needed computing time, to calculate the lepton fake rate and the real lepton efficiency for almost every possible set of cuts.

The control plots for the different variables and run periods look ok in general showing a good agreement between the data and the summed up Monte Carlo samples after a correct scaling of the multijet background by using ϵ_{sig} and ϵ_{QCD} as well as rescaling of all the W+jets Monte Carlo. Some known problem for different “hard-to-model” variables like jet η remain, but these descriptions can be improved a lot with reweighting the instantaneous luminosity and then jet η itself.

One of the most interesting disagreements is the instantaneous luminosity because

many other variables are dependent on this luminosity value. In principle the luminosity profiles available in the central framework should help to give an almost perfect description. Clearly, this issue needs to be investigated further but for the moment it is possible to restore all the agreement lost here by just reweighting according to instantaneous luminosity on the analysis level.

All in all the selection is now ready to start with the first analyses and additional samples for the many different systematic uncertainties should be ready within a few weeks.

But as always there is of course room for improvement from the technical as well as from the physics and statistics point of view. Some of these ideas for further upgrades are described in the next section.

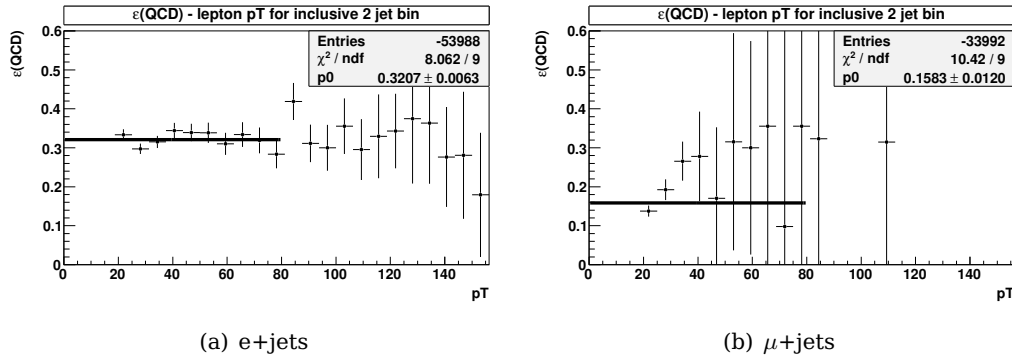


Figure 8.1: ϵ_{QCD} plots for the inclusive 2 jet bin in RunIIb2 derived from lepton p_T .

8. Summary & Outlook

In summary the simple matrix method yields consistent results and is successful in scaling the multijet background in a way to reach a good data Monte Carlo agreement. The estimation of all the fake rates and efficiencies was done run period dependent and for the two different sets of cuts corresponding to the previous $t\bar{t}$ lepton+jets selection and to the previous single top selection, respectively. The developed software can be run easily for other sets of cuts or for a combination of run periods.

The following subsections describe different possible additions to the matrix method and ideas to improve the combined selection to reach an even better description.

8.1. Fake Rate and Efficiency Dependencies

The matrix method used in this analysis fits the ϵ_{QCD} ratio with a polynomial of degree 0. The ratio plots especially for lepton p_T show variations that look – especially for μ +jets – significantly different from this straight-line approach (see figure 8.1). In a first step one could try to fit the lepton p_T ratio using higher order polynomials and calculate the overall normalization of the background with this p_T dependent ϵ_{QCD} (or ϵ_{sig}). However, a first test with a simple (discrete) p_T dependence only showed a small improvement in terms of description. Additionally, it would be possible to reweight the QCD background using this lepton p_T dependent values in a second step in order to get a more accurate shape for the multijet background.

The ϵ_{sig} ratio derived from lepton p_T shows a strong dependency on lepton p_T for both channels. The value grows (linear or exponential) with and levels off towards higher jet bins (see figure 8.2). Because the higher bins have much lower statistics we could underestimate the ϵ_{sig} for high p_T by around 2 to 3% for e+jets (5 to 10% for μ +jets) while overestimating it for low bins by about the same amount. Also, the ϵ_{sig} derived from lepton η shows some interesting variations (see figure 8.3).

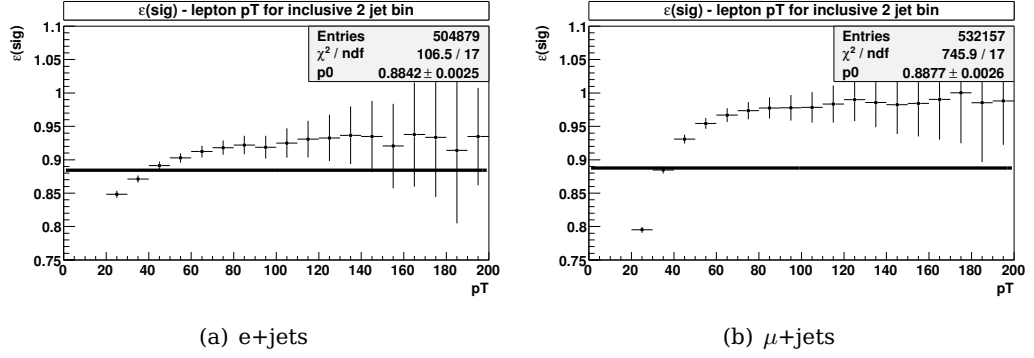


Figure 8.2: ϵ_{sig} plots for the inclusive 2 jet bin in RunIIb2 derived from lepton p_T .

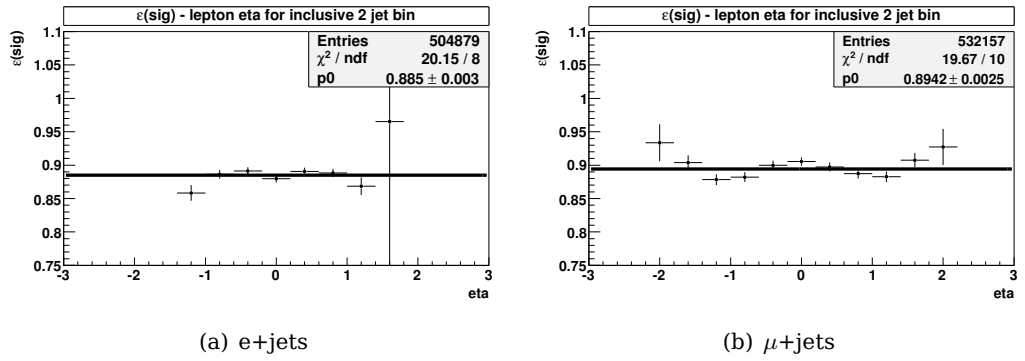


Figure 8.3: ϵ_{sig} plots for the inclusive 2 jet bin in RunIIb2 derived from lepton η .

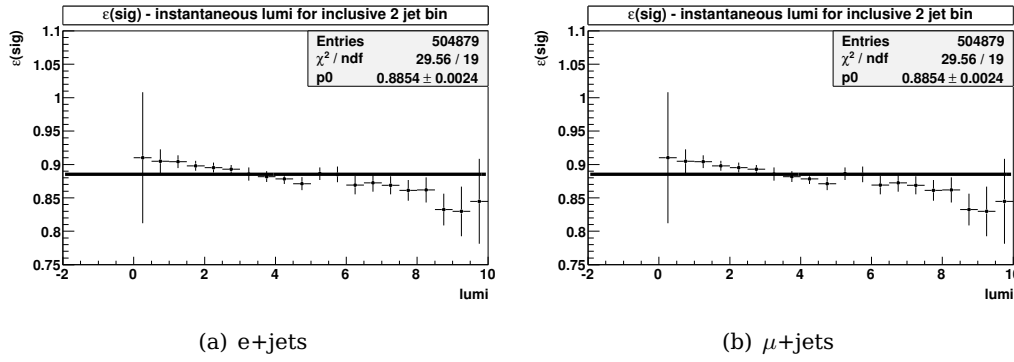


Figure 8.4: ϵ_{sig} plots for the inclusive 2 jet bin in RunIIb2 derived from the instantaneous luminosity.

In addition the ϵ_{sig} value derived from the instantaneous luminosity (not used in any calculation) was checked and clearly it should not be fitted with a straight line. The ratio plot shows that it may be worthwhile to try deriving the real lepton efficiency also dependent on instantaneous luminosity (see figure 8.4) in addition to the already considered dependencies on run period and cuts. Before considering such a dependency it would of course be necessary to reweight the Monte Carlo samples so that the instantaneous luminosity exactly matches the data.

Because the most compelling dependencies (a clearly linear increase in the first few bins, exponential growth and level off over a significant range of ϵ_{sig} , linear decrease over all the bins) is seen in for μ +jets where the multijet background is smaller by a factor of about ten compared to the e+jets channel these improvements have not been pursued yet with highest priority.

8.2. Using the full Dataset

In addition to the RunIIb dataset used in the analysis presented here there is one additional fb^{-1} from RunIIa that should be added to get as much data as possible and further limit the statistical uncertainties. At the time of the completion of this thesis there were still some crucial input files missing, that prevented us from processing the RunIIa data with the new multivariate electron identification.

If these trigger turn-on files are provided within the next few weeks the RunIIa data should be processed with exactly the same configuration as RunIIb, otherwise it would be possible to add the RunIIa data processed with previous software versions in a special merging procedure.⁶

⁶While completing the work on the thesis these files became available. The upcoming DØ note will thus include lepton fake rate and real lepton efficiency values as well as control plots for the full RunII data set.

8.3. General Code for Reweighting

The current tools that create the histograms from the tree by copying and calculating the different variables and filling the histograms lack the ability to do a general reweighting for one or more distributions. Earlier studies clearly showed that for example a reweighting of the instantaneous luminosity also drastically improves the description of other distributions like jet η .

An easy way to do these reweightings (even for multiple variables) would allow for more quick studies which of them give significant overall improvements for the important variables.

8.4. Better Automatic Evaluation of Data / Monte Carlo Agreement

Currently the estimation of how well a data distribution matches the respective Monte Carlo sum is done using the Kolmogorov-Smirnov test. As the lepton+jets channel is not pure enough there are always some slight discrepancies (sometimes only in one bin) that completely obliterate any kind of agreement in all (or most) of the other bins. The Kolmogorov-Smirnov value basically becomes more and more an indicator of how big the error bars on the data distribution are. This fact leads to a lot of manual work in comparing all the plots “by hand”. To facilitate this task the histograms have been supplemented with the ratio plots just below the distributions during the previous selection. These ratios allow for a quick evaluation on how good the data Monte Carlo agreement is.

For more detailed studies (for example with different reweightings) it would be very helpful to have a more robust statistical test to estimate the data Monte Carlo agreement. For example it could be useful to calculate the overall χ^2 .

Literature

- [1] S.W. Herb, D.C. Hom, L.M. Lederman, et al. Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, Aug 1977. doi:10.1103/PhysRevLett.39.252.
- [2] S. Abachi et al. Observation of the Top Quark. *Phys. Rev. Lett.*, 74:2632–2637, Apr 1995. doi:10.1103/PhysRevLett.74.2632.
- [3] M.L. Perl et al. Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation. *Phys. Rev. Lett.*, 35:1489–1492, Dec 1975. doi:10.1103/PhysRevLett.35.1489.
- [4] Fermilab Inquiring Minds. What is the world made of? <http://www.fnal.gov/pub/inquiring/matter/madeof/index.html>.
- [5] W.N. Cottingham and D.A. Greenwood. *An introduction to the standard model of particle physics*. Cambridge University Press, 2007. ISBN 9780521852494.
- [6] K. Nakamura et al. (Particle Data Group). 2011 Review of Particle Physics. *J. Phys. G*, 37, 075021, 2010 and 2011 partial update for the 2012 edition.
- [7] M. Gell-Mann. A schematic model of baryons and mesons. *Physics Letters*, 8(3):214 – 215, 1964. ISSN 0031-9163. doi:10.1016/S0031-9163(64)92001-3.
- [8] J.I. Friedman and H.W. Kendall. Deep inelastic electron scattering. *Annu. Rev. Nucl. Sci.*, 22:203 – 254, 1972.
- [9] R.P. Feynman. *QED: the strange theory of light and matter*. Penguin Press Science Series. Penguin, 1990. ISBN 9780140125054.
- [10] G. Dissertori, I.G. Knowles, and M. Schmelling. *Quantum chromodynamics: high energy experiments and theory*. International series of monographs on physics. Clarendon Press, 2003. ISBN 9780198505723.
- [11] Emmy Noether. Invariante Variationsprobleme. *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Math.-phys. Klasse*, pages 235–257, 1918.
- [12] S. Moch and P. Uwer, *Phys. Rev. D* 78, 034003 (2008); U. Langenfeld, S. Moch, and P. Uwer, *Phys. Rev. D* 80, 054009 (2009); M. Aliev et al., *Comput. Phys. Commun.* 182, 1034 (2011).
- [13] V. Ahrens, A. Ferroglia, M. Neubert, B.D. Pecjak, and L.L. Yang, *J. High Energy Phys.* 09, 097 (2010); V. Ahrens, A. Ferroglia, M. Neubert, B.D. Pecjak, and L.L. Yang, *Nucl. Phys. Proc. Suppl.* 205-206, 48 (2010).

- [14] Nikolaos Kidonakis and Ramona Vogt. Theoretical top quark cross section at the Fermilab Tevatron and the CERN LHC. *Phys. Rev. D*, 78:074005, Oct 2008. doi:10.1103/PhysRevD.78.074005.
- [15] DØ Top Group. Useful Diagrams of Top Signals and Backgrounds. http://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html.
- [16] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, et al. TMVA: Toolkit for Multivariate Data Analysis. *PoS, ACAT:040*, 2007.
- [17] Rene Brun and Fons Rademakers, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also <http://root.cern.ch/>.
- [18] V.M. Abazov et al. Observation of Single Top-Quark Production. *Phys. Rev. Lett.*, 103:092001, Aug 2009. doi:10.1103/PhysRevLett.103.092001.
- [19] Fermilab Visual Media Services. Fermilab Accelerator Chain - Schematic drawing of Accelerator Complex. http://www-visualmedia.fnal.gov/VMS_Site/gallery/stillphotos/2000/0600/00-0635D.hr.jpg, June 2000.
- [20] CERN Press Office. LHC sets new world record. <http://press.web.cern.ch/press/PressReleases/Releases2009/PR18.09E.html>.
- [21] DØ Operations Group. DØ RunII Integrated Luminosity. http://d0server1.fnal.gov/Projects/Operations/D0RunII_DataTaking.htm, September 2011.
- [22] V.M. Abazov et al. The upgraded DØ detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 565(2):463 – 537, 2006. ISSN 0168-9002. doi: 10.1016/j.nima.2006.05.248.
- [23] S.N. Ahmed et al. The D0 Silicon Microstrip Tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 634(1):8 – 46, 2011. ISSN 0168-9002. doi: 10.1016/j.nima.2010.11.121.
- [24] R. Angstadt et al. The layer 0 inner silicon detector of the D0 experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 622(1):298 – 310, 2010. ISSN 0168-9002. doi:10.1016/j.nima.2010.04.148.

- [25] Ann Heinson. Detector Drawings. <http://www-d0.fnal.gov/Run2Physics/WWW/drawings.htm>.
- [26] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.
- [27] Michelangelo L. Mangano, Fulvio Piccinini, Antonio D. Polosa, Mauro Moretti, and Roberto Pittau. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *Journal of High Energy Physics*, 2003(07):001, 2003.
- [28] S. Agostinelli et al. GEANT4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003. doi:10.1016/S0168-9002(03)01368-8.
- [29] DØ V+jets group. V+jets analysis framework. <https://plone4.fnal.gov/P1/DØWiki/physics/VplusJets/>.
- [30] Suman Bala et al. Single Top Quark Production in 5.4/fb of Data - Signal and Background Modeling and Event Selection, 2010. DØ Note 6099.
- [31] Jean-Laurent Agram et al. Jet Energy Scale at D0 RunII, 2005. DØ Note 4720.
- [32] Andreas Jung et al. Unfolding of differential ttbar cross section distributions in the ttbar lepton+jets channel., 2011. Supplemental material of DØ Note 6237.
- [33] Oleg Brandt et al. Measurement of the ttbar Cross Section in the lepton+jets Channel with 5.3 fb⁻¹, 2010. DØ Note 6028.
- [34] V.M. Abazov et al. Measurement of the top quark pair production cross section in the lepton + jets channel in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. D*, 84:012008, Jul 2011. doi:10.1103/PhysRevD.84.012008.

Acknowledgement

I am grateful to Prof. Cecilia Gerber and Prof. Günther Dissertori for giving me the opportunity to complete my Masters degree with a very interesting thesis project at Fermilab working with the DØ collaboration. I would like to thank Dr. Andreas Jung who never got tired of answering my questions and provided invaluable help in bringing me up to speed on the necessary software and computing infrastructure, in explaining the subtleties of the underlying detector and particle physics, and in suggesting corrections and giving feedback for my thesis paper. In addition I would like to thank the whole DØ Double Top and Single Top groups for their help and their feedback to my work and my presentations.

(a) e+jets				
	= 1 jet	= 2 jets	= 3 jets	≥ 4 jets
RunIIb1	0.838	0.948	1.057	1.133
RunIIb2	0.836	0.935	1.023	1.109
RunIIb3	0.785	0.961	1.023	1.025
RunIIb4	0.825	0.986	1.073	1.481

(b) μ +jets				
	= 1 jet	= 2 jets	= 3 jets	≥ 4 jets
RunIIb1	0.693	0.823	0.965	1.215
RunIIb2	0.901	1.061	1.181	1.632
RunIIb3	0.954	1.080	1.208	1.336
RunIIb4	0.966	1.151	1.330	1.460

Table A.1: S_W scale factors for the single top cuts per run period and jet bin.

A. More Result Values

This appendix contains the following tables (for more results, including b-tagging, look at the upcoming DØ note).

- Table A.1: S_W scale factors for the single top cuts.
- Table A.2: Yield table for RunIIb e+jets using the single top cuts.
- Table A.3: Yield table for RunIIb μ +jets using the single top cuts.
- Table A.4: Jet bin dependent ϵ_{QCD} values for the single top cuts.
- Table A.5: Jet bin dependent ϵ_{sig} values for the single top cuts.

Contribution	= 1 jet	= 2 jets	= 3 jets	≥ 4 jets
Data	209272.00 \pm 457.46	73211.00 \pm 270.58	10353.00 \pm 101.75	2324.00 \pm 48.21
Multijet	17554.85 \pm 554.48	9504.56 \pm 300.21	1439.43 \pm 54.05	272.94 \pm 16.55
WW	1393.06 \pm 3.74	1673.34 \pm 3.96	259.36 \pm 1.61	38.01 \pm 0.64
WZ	211.92 \pm 0.83	270.25 \pm 0.90	44.70 \pm 0.36	6.56 \pm 0.15
Wbb	4363.05 \pm 9.71	3200.88 \pm 8.74	574.42 \pm 3.63	101.27 \pm 1.60
Wcc	10319.31 \pm 14.57	7031.95 \pm 12.82	1306.31 \pm 5.56	229.33 \pm 2.46
Wlp	169546.70 \pm 104.41	47761.30 \pm 47.65	5073.64 \pm 14.50	604.13 \pm 4.68
ZZ	19.51 \pm 0.18	17.33 \pm 0.17	4.00 \pm 0.08	0.85 \pm 0.04
ZbEE	182.98 \pm 0.81	136.53 \pm 0.69	26.92 \pm 0.88	5.36 \pm 0.14
ZbbTauTau	34.94 \pm 0.48	34.15 \pm 0.52	9.58 \pm 0.28	2.02 \pm 0.14
ZcEE	311.56 \pm 1.50	230.86 \pm 1.26	41.16 \pm 0.51	8.02 \pm 0.23
ZccTauTau	80.36 \pm 0.87	74.18 \pm 0.89	20.13 \pm 0.48	3.89 \pm 0.23
ZlpEE	3655.67 \pm 12.87	1409.64 \pm 7.06	157.53 \pm 1.93	22.99 \pm 0.71
ZlpTauTau	1097.08 \pm 7.14	499.66 \pm 3.90	89.74 \pm 1.46	11.25 \pm 0.49
tb	41.10 \pm 0.26	90.91 \pm 0.37	22.44 \pm 0.19	4.07 \pm 0.08
tbq	85.31 \pm 0.56	125.54 \pm 0.69	35.17 \pm 0.38	7.92 \pm 0.19
ttbar	76.44 \pm 1.28	471.42 \pm 2.39	1051.86 \pm 3.28	969.01 \pm 2.93
ttl	259.23 \pm 0.85	683.43 \pm 1.34	196.66 \pm 0.63	36.42 \pm 0.25
$\Sigma_i MC_i$	209233.06 \pm 564.71	73215.92 \pm 304.52	10353.05 \pm 56.54	2324.04 \pm 17.73

Table A.2: Total yields for the single top cuts in RunIIb for the e+jets channel, before applying any b-tag for jets; the W+jets processes are scaled in order to match Monte Carlo to data.

Contribution	= 1 jet	= 2 jets	= 3 jets	≥ 4 jets
Data	169218.00 ± 411.36	61657.00 ± 248.31	9398.00 ± 96.94	2090.00 ± 45.72
Multijet	317.09 ± 82.21	1103.64 ± 187.23	273.87 ± 47.31	67.25 ± 12.60
WW	1156.55 ± 3.11	1472.30 ± 3.32	234.09 ± 1.35	35.53 ± 0.54
WZ	199.52 ± 0.75	279.21 ± 0.84	46.82 ± 0.35	7.14 ± 0.14
Wbb	3513.75 ± 8.04	2886.68 ± 7.86	595.01 ± 3.67	118.32 ± 1.78
Wcc	8512.06 ± 12.30	6462.53 ± 11.63	1364.99 ± 5.66	268.42 ± 2.78
Wlp	142918.21 ± 89.71	44389.01 ± 43.66	5318.76 ± 14.44	708.14 ± 5.08
ZZ	24.69 ± 0.18	35.17 ± 0.21	6.50 ± 0.09	1.06 ± 0.04
ZbbMuMu	340.86 ± 0.92	234.71 ± 0.81	44.80 ± 0.37	7.14 ± 0.15
ZbbTauTau	15.44 ± 0.23	17.31 ± 0.26	4.56 ± 0.13	0.89 ± 0.05
ZccMuMu	782.75 ± 1.92	505.92 ± 1.63	88.52 ± 0.69	13.84 ± 0.29
ZccTauTau	37.28 ± 0.50	36.56 ± 0.50	9.98 ± 0.26	2.00 ± 0.12
ZlpMuMu	10629.80 ± 17.74	3001.75 ± 7.79	357.12 ± 2.43	34.12 ± 0.73
ZlpTauTau	487.11 ± 4.19	244.53 ± 2.49	44.46 ± 0.94	6.56 ± 0.35
tb	31.37 ± 0.28	70.14 ± 0.30	18.67 ± 0.16	3.70 ± 0.07
tbq	63.38 ± 0.41	94.82 ± 0.53	26.41 ± 0.30	6.20 ± 0.15
ttbar	55.64 ± 1.16	359.72 ± 1.83	827.68 ± 2.64	783.23 ± 2.54
tll	165.04 ± 0.63	466.39 ± 1.02	135.74 ± 0.48	26.47 ± 0.20
$\sum_i MC_i$	169250.56 ± 123.99	61660.41 ± 192.99	9397.98 ± 50.09	2090.01 ± 14.25

Table A.3: Total yields for the single top cuts in RunIIb for the μ +jets channel, before applying any b-tag for jets; the W+jets processes are scaled in order to match Monte Carlo to data.

(a) RunIIb1, e+jets				(b) RunIIb1, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.289	± 0.019	$+0.042$ -0.042	= 1 jet	< 0	± 0.172	$+0.302$ -0.274
= 2 jets	0.308	± 0.020	$+0.014$ -0.014	= 2 jets	0.103	± 0.040	$+0.047$ -0.049
= 3 jets	0.275	± 0.047	$+0.279$ -0.279	= 3 jets	0.031	± 0.066	$+0.226$ -0.226
≥ 4 jets	< 0	± 0.042	$+0.467$ -0.467	≥ 4 jets	0.181	± 0.042	$+0.291$ -0.291
≥ 2 jets	0.306	± 0.019	$+0.013$ -0.013	≥ 2 jets	0.093	± 0.034	$+0.044$ -0.046
(c) RunIIb2, e+jets				(d) RunIIb2, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.301	± 0.009	$+0.025$ -0.025	= 1 jet	0.180	± 0.057	$+0.114$ -0.109
= 2 jets	0.293	± 0.010	$+0.010$ -0.010	= 2 jets	0.186	± 0.026	$+0.053$ -0.055
= 3 jets	0.332	± 0.028	$+0.016$ -0.016	= 3 jets	0.233	± 0.048	$+0.056$ -0.056
≥ 4 jets	< 0	± 0.007	$+0.402$ -0.402	≥ 4 jets	0.040	± 0.066	$+0.292$ -0.292
≥ 2 jets	0.298	± 0.010	$+0.009$ -0.009	≥ 2 jets	0.201	± 0.023	$+0.055$ -0.057
(e) RunIIb3, e+jets				(f) RunIIb3, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.196	± 0.010	$+0.032$ -0.033	= 1 jet	0.209	± 0.056	$+0.129$ -0.146
= 2 jets	0.252	± 0.012	$+0.017$ -0.017	= 2 jets	0.185	± 0.028	$+0.057$ -0.058
= 3 jets	0.265	± 0.030	$+0.026$ -0.026	= 3 jets	0.095	± 0.044	$+0.080$ -0.080
≥ 4 jets	< 0	± 0.022	$+0.330$ -0.330	≥ 4 jets	< 0	± 0.076	$+0.170$ -0.170
≥ 2 jets	0.254	± 0.011	$+0.016$ -0.016	≥ 2 jets	0.174	± 0.024	$+0.051$ -0.052
(g) RunIIb4, e+jets				(h) RunIIb4, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.271	± 0.009	$+0.035$ -0.036	= 1 jet	0.228	± 0.060	$+0.118$ -0.125
= 2 jets	0.302	± 0.011	$+0.019$ -0.019	= 2 jets	0.232	± 0.029	$+0.082$ -0.083
= 3 jets	0.340	± 0.028	$+0.199$ -0.199	= 3 jets	0.223	± 0.054	$+0.050$ -0.050
≥ 4 jets	0.258	± 0.077	$+0.220$ -0.220	≥ 4 jets	< 0	± 0.077	$+0.416$ -0.416
≥ 2 jets	0.308	± 0.010	$+0.021$ -0.021	≥ 2 jets	0.239	± 0.026	$+0.077$ -0.078

Table A.4: Jet bin dependent ϵ_{QCD} values for the single top cuts for all four run periods. The 4 jet bin usually suffers from low statistics and thus the fits do not converge.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

(a) RunIIb1, e+jets				(b) RunIIb1, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.886	± 0.001	± 0.064	= 1 jet	0.933	± 0.001	± 0.065
= 2 jets	0.876	± 0.002	± 0.021	= 2 jets	0.913	± 0.002	± 0.020
= 3 jets	0.866	± 0.004	± 0.007	= 3 jets	0.897	± 0.004	± 0.003
≥ 4 jets	0.868	± 0.006	± 0.672	≥ 4 jets	0.885	± 0.006	± 0.005
≥ 2 jets	0.874	± 0.002	± 0.004	≥ 2 jets	0.911	± 0.002	± 0.018
(c) RunIIb2, e+jets				(d) RunIIb2, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.890	± 0.001	± 0.114	= 1 jet	0.917	± 0.001	± 0.048
= 2 jets	0.885	± 0.002	± 0.035	= 2 jets	0.893	± 0.002	± 0.016
= 3 jets	0.877	± 0.006	± 0.002	= 3 jets	0.873	± 0.006	± 0.012
≥ 4 jets	0.879	± 0.009	± 0.010	≥ 4 jets	0.872	± 0.010	± 0.013
≥ 2 jets	0.884	± 0.002	± 0.010	≥ 2 jets	0.890	± 0.002	± 0.011
(e) RunIIb3, e+jets				(f) RunIIb3, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.879	± 0.002	± 0.099	= 1 jet	0.917	± 0.002	± 0.080
= 2 jets	0.882	± 0.002	± 0.009	= 2 jets	0.897	± 0.002	± 0.019
= 3 jets	0.877	± 0.005	± 0.011	= 3 jets	0.880	± 0.005	± 0.010
≥ 4 jets	0.882	± 0.007	± 0.012	≥ 4 jets	0.873	± 0.008	± 0.010
≥ 2 jets	0.881	± 0.002	± 0.006	≥ 2 jets	0.894	± 0.002	± 0.012
(g) RunIIb4, e+jets				(h) RunIIb4, μ +jets			
	Cen	δ_{stat}	δ_{syst}		Cen	δ_{stat}	δ_{syst}
= 1 jet	0.879	± 0.002	± 0.101	= 1 jet	0.912	± 0.002	± 0.083
= 2 jets	0.882	± 0.002	± 0.012	= 2 jets	0.893	± 0.002	± 0.019
= 3 jets	0.879	± 0.006	± 0.010	= 3 jets	0.877	± 0.006	± 0.011
≥ 4 jets	0.884	± 0.008	± 0.012	≥ 4 jets	0.871	± 0.008	± 0.011
≥ 2 jets	0.882	± 0.002	± 0.005	≥ 2 jets	0.890	± 0.002	± 0.010

Table A.5: Jet bin dependent ϵ_{sig} values for the single top cuts for all four run periods.

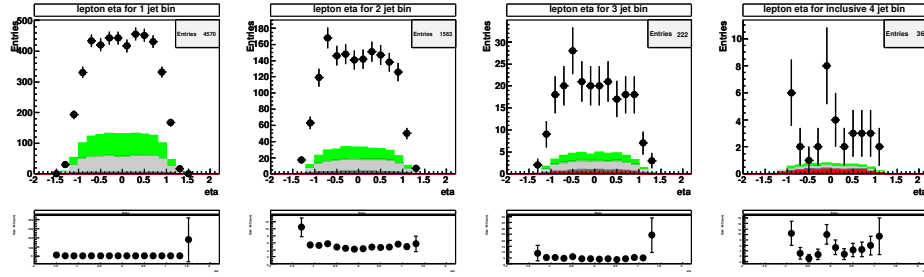
B. More Control Plots

This appendix contains the following plots (for more plots, including b-tagging, look at the upcoming $D\bar{D}$ note).

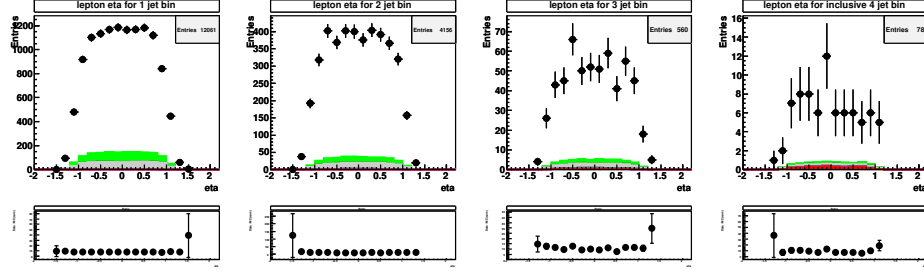
- Figure B.1: ϵ_{QCD} histograms for RunIIb1 e+jets using $t\bar{t}$ selection cuts.
- Figure B.2: ϵ_{QCD} histograms for RunIIb1 μ +jets using $t\bar{t}$ selection cuts.
- Figure B.3: ϵ_{QCD} plots for RunIIb1 e+jets using $t\bar{t}$ selection cuts.
- Figure B.4: ϵ_{QCD} plots for RunIIb1 μ +jets using $t\bar{t}$ selection cuts.
- Figure B.5: ϵ_{QCD} histograms for RunIIb2 e+jets using $t\bar{t}$ selection cuts.
- Figure B.6: ϵ_{QCD} histograms for RunIIb2 μ +jets using $t\bar{t}$ selection cuts.
- Figure B.7: ϵ_{QCD} plots for RunIIb2 e+jets using $t\bar{t}$ selection cuts.
- Figure B.8: ϵ_{QCD} plots for RunIIb2 μ +jets using $t\bar{t}$ selection cuts.
- Figure B.9: ϵ_{QCD} histograms for RunIIb3 e+jets using $t\bar{t}$ selection cuts.
- Figure B.10: ϵ_{QCD} histograms for RunIIb3 μ +jets using $t\bar{t}$ selection cuts.
- Figure B.11: ϵ_{QCD} plots for RunIIb3 e+jets using $t\bar{t}$ selection cuts.
- Figure B.12: ϵ_{QCD} plots for RunIIb3 μ +jets using $t\bar{t}$ selection cuts.
- Figure B.13: ϵ_{QCD} histograms for RunIIb4 e+jets using $t\bar{t}$ selection cuts.
- Figure B.14: ϵ_{QCD} histograms for RunIIb4 μ +jets using $t\bar{t}$ selection cuts.
- Figure B.15: ϵ_{QCD} plots for RunIIb4 e+jets using $t\bar{t}$ selection cuts.
- Figure B.16: ϵ_{QCD} plots for RunIIb4 μ +jets using $t\bar{t}$ selection cuts.
- Figure B.17: H_T plots for RunIIb1 e+jets using the $t\bar{t}$ selection cuts.
- Figure B.18: H_T plots for RunIIb1 μ +jets using the $t\bar{t}$ selection cuts.
- Figure B.19: H_T plots for RunIIb2 e+jets using the $t\bar{t}$ selection cuts.
- Figure B.20: H_T plots for RunIIb2 μ +jets using the $t\bar{t}$ selection cuts.
- Figure B.21: H_T plots for RunIIb3 e+jets using the $t\bar{t}$ selection cuts.
- Figure B.22: H_T plots for RunIIb3 μ +jets using the $t\bar{t}$ selection cuts.
- Figure B.23: H_T plots for RunIIb4 e+jets using the $t\bar{t}$ selection cuts.

- Figure B.24: H_T plots for RunIIb4 μ +jets using the $t\bar{t}$ selection cuts.
- Figure B.25: H_T plots for RunIIb1 e+jets using the single top cuts.
- Figure B.26: H_T plots for RunIIb1 μ +jets using the single top cuts.
- Figure B.27: H_T plots for RunIIb2 e+jets using the single top cuts.
- Figure B.28: H_T plots for RunIIb2 μ +jets using the single top cuts.
- Figure B.29: H_T plots for RunIIb3 e+jets using the single top cuts.
- Figure B.30: H_T plots for RunIIb3 μ +jets using the single top cuts.
- Figure B.31: H_T plots for RunIIb4 e+jets using the single top cuts.
- Figure B.32: H_T plots for RunIIb4 μ +jets using the single top cuts.

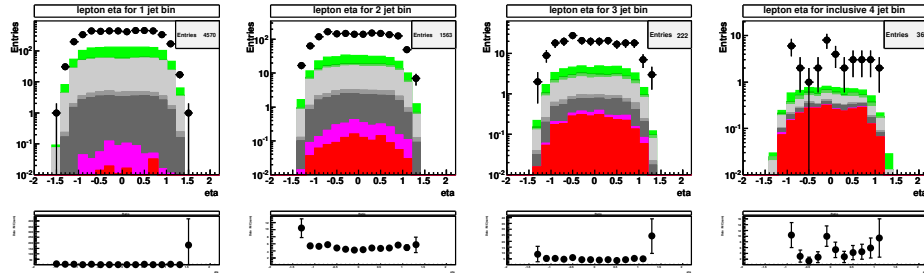
Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set



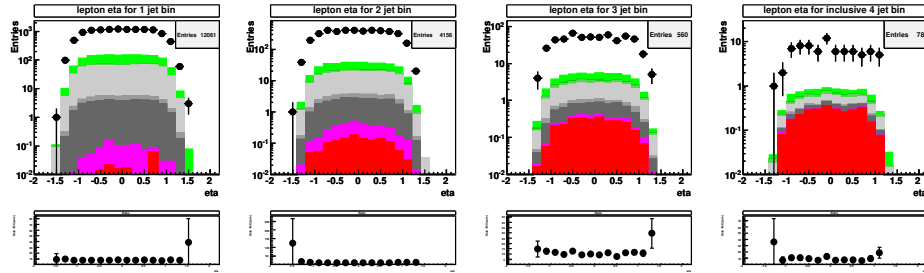
(a) tight samples, normal scale



(b) loose samples, normal scale



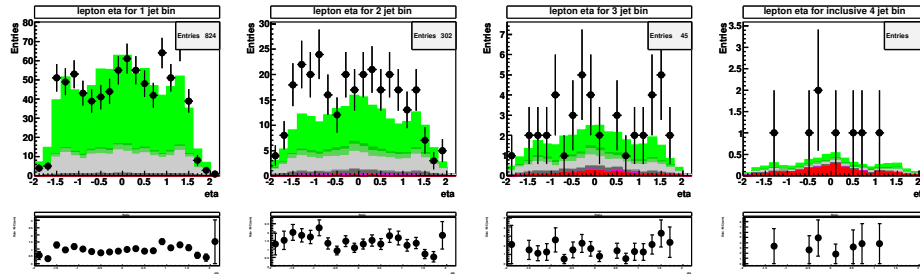
(c) tight samples, log scale



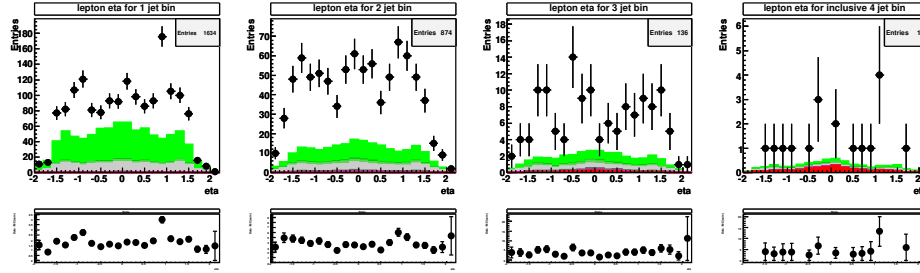
(d) loose samples, log scale

Figure B.1: Histograms for ϵ_{QCD} calculation for RunIIb1 e+jets using $t\bar{t}$ selection cuts. These are all lepton η histograms used to calculate the central value for ϵ_{QCD} . From left to right the histograms show the exclusive 1, 2, 3 jet bin, and the inclusive 4 jet bin.

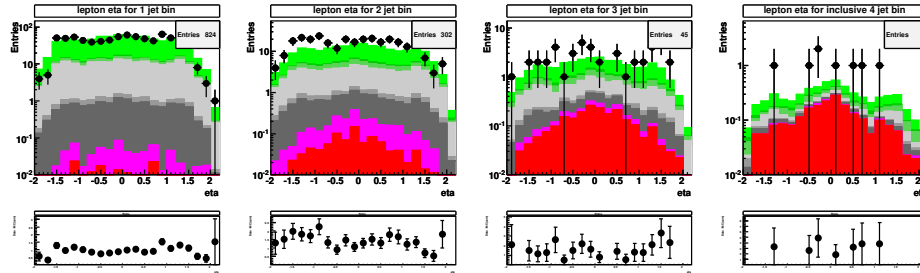
Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set



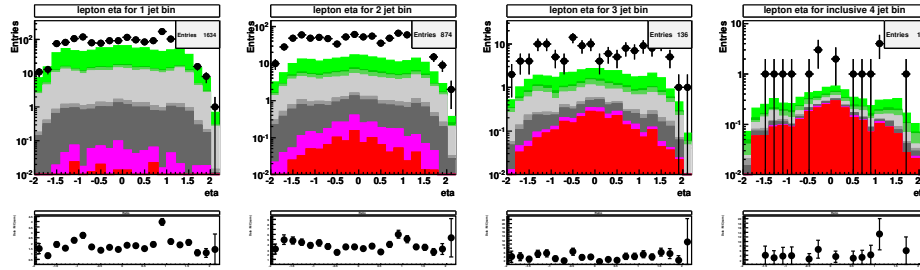
(a) tight samples, normal scale



(b) loose samples, normal scale



(c) tight samples, log scale



(d) loose samples, log scale

Figure B.2: Histograms for ϵ_{QCD} calculation for RunIIb1 μ +jets using $t\bar{t}$ selection cuts. These are all lepton η histograms used to calculate the central value for ϵ_{QCD} . From left to right the histograms show the exclusive 1, 2, 3 jet bin, and the inclusive 4 jet bin.

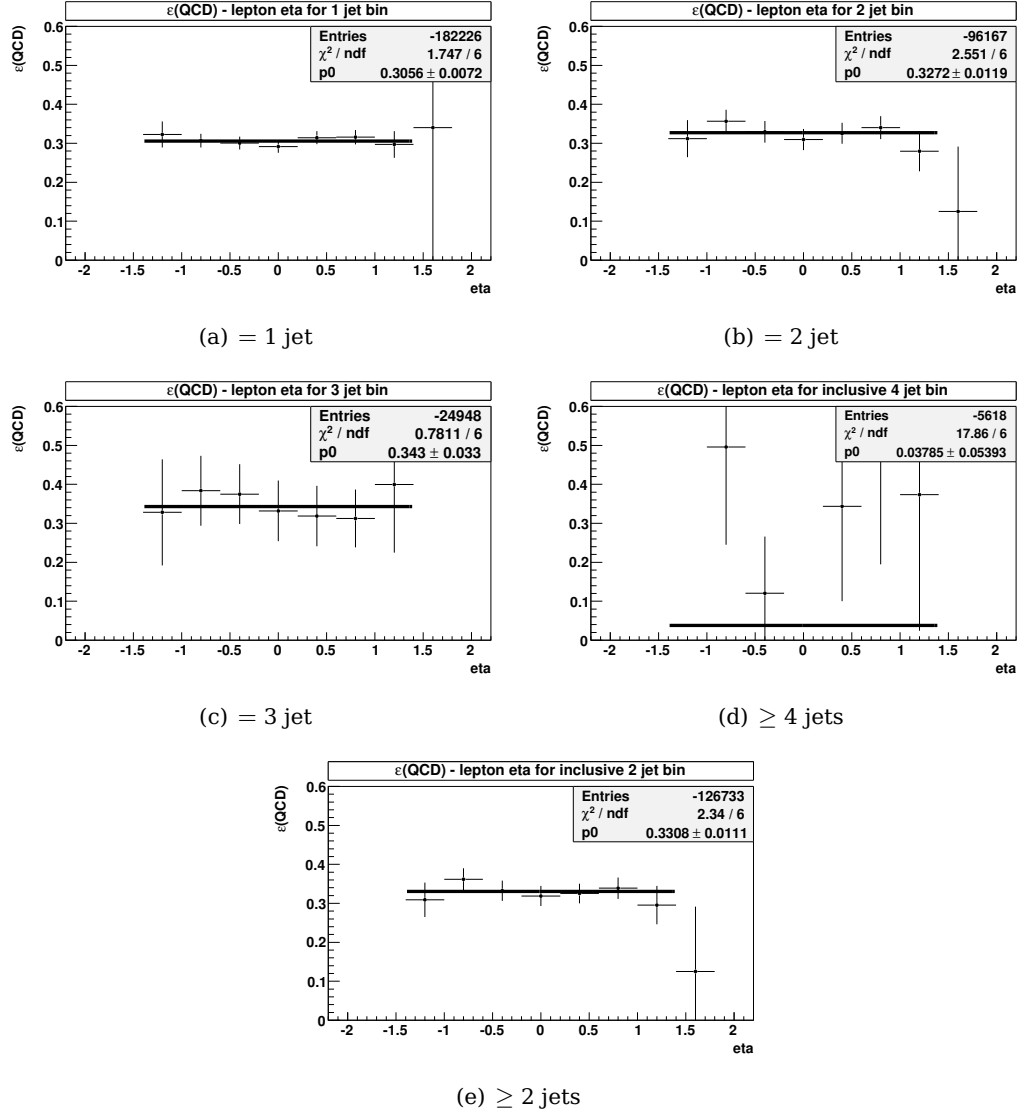


Figure B.3: ϵ_{QCD} plots for RunIIb1 e+jets using $t\bar{t}$ selection cuts. These ratio plots are used to determine the central value of ϵ_{QCD} .

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

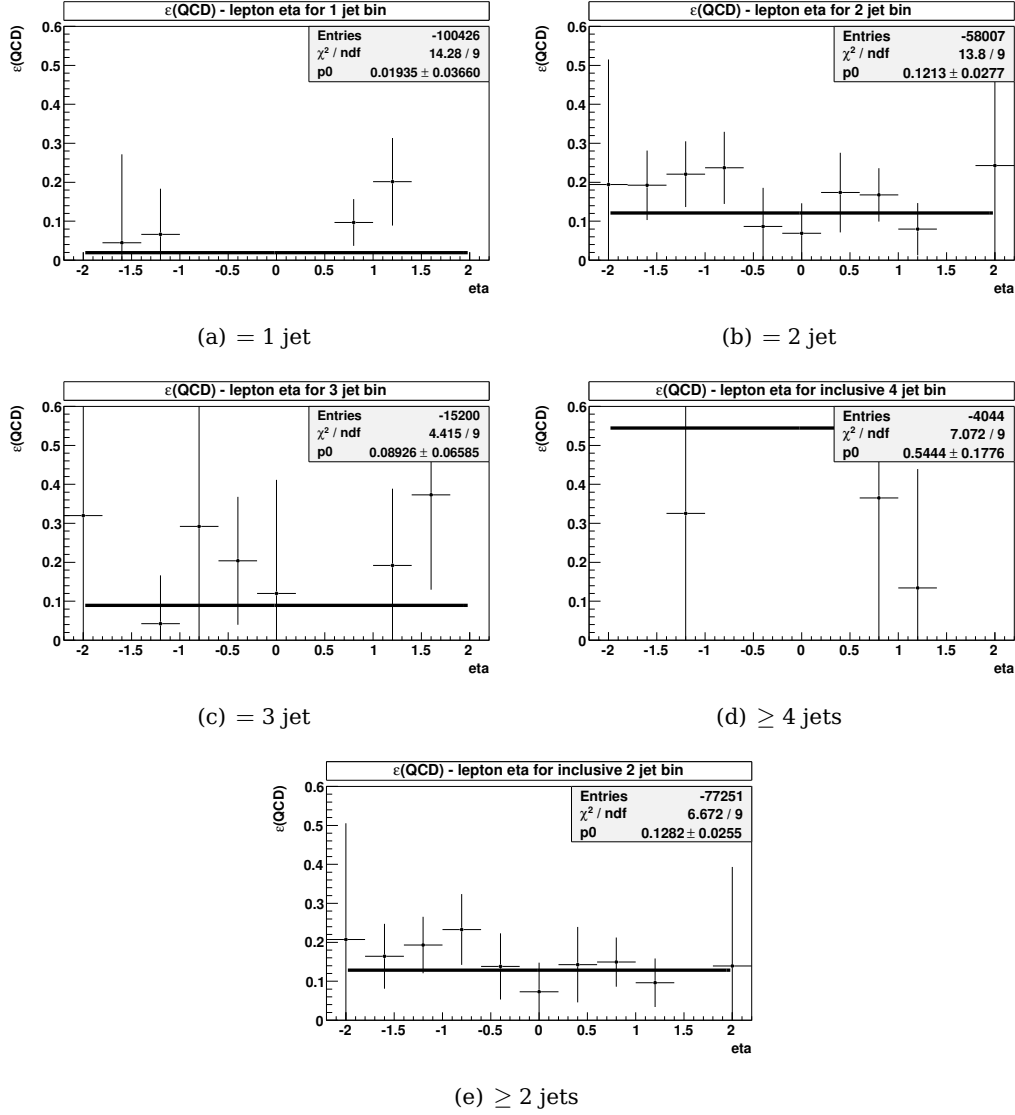
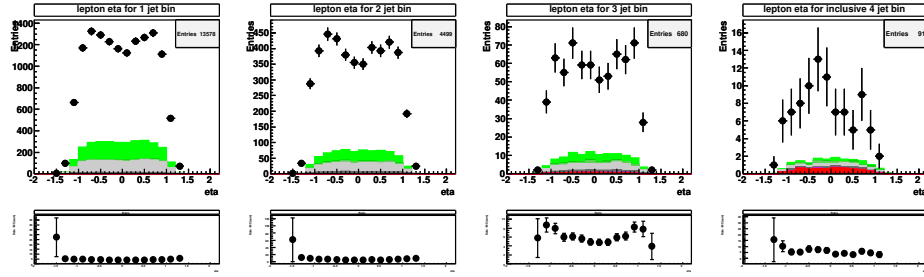
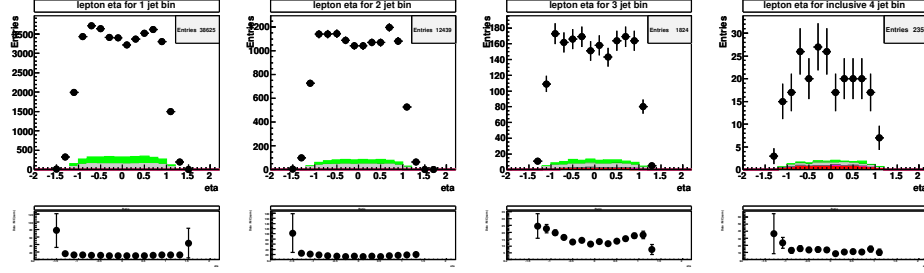


Figure B.4: ϵ_{QCD} plots for RunIIb1 μ +jets using $t\bar{t}$ selection cuts. These ratio plots are used to determine the central value of ϵ_{QCD} .

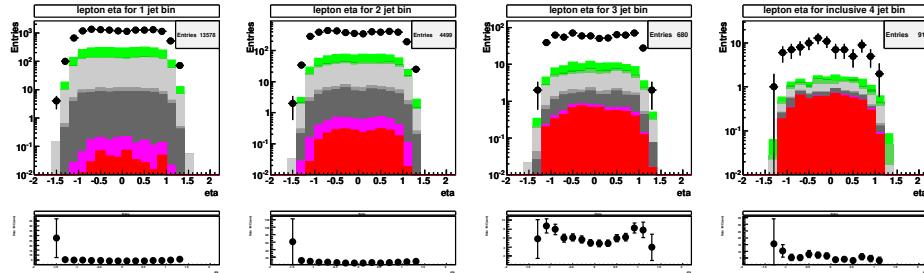
Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set



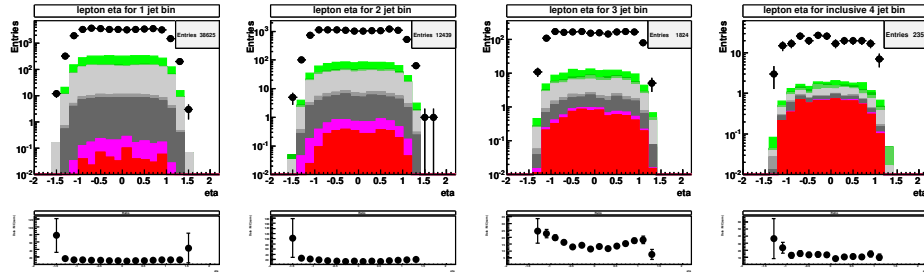
(a) tight samples, normal scale



(b) loose samples, normal scale



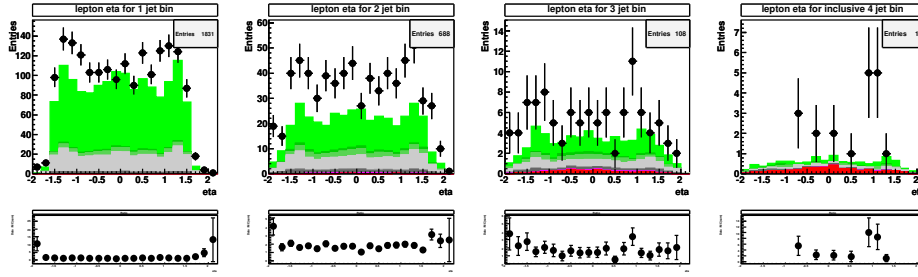
(c) tight samples, log scale



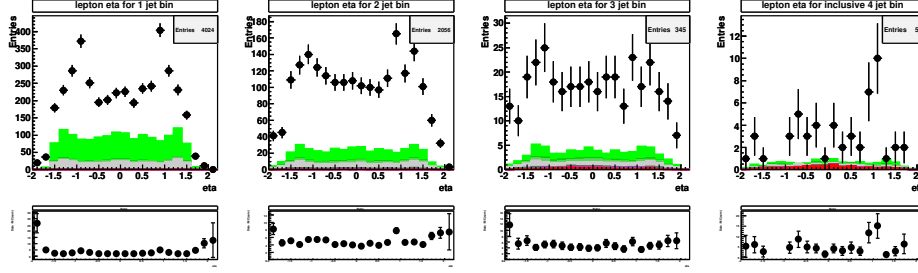
(d) loose samples, log scale

Figure B.5: Histograms for ϵ_{QCD} calculation for RunIIb2 e+jets using $t\bar{t}$ selection cuts. These are all lepton η histograms used to calculate the central value for ϵ_{QCD} . From left to right the histograms show the exclusive 1, 2, 3 jet bin, and the inclusive 4 jet bin.

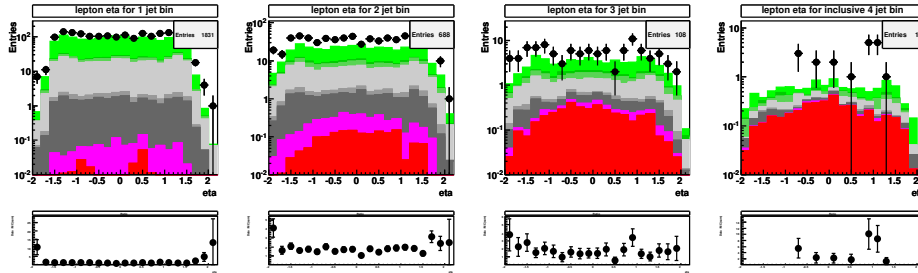
Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set



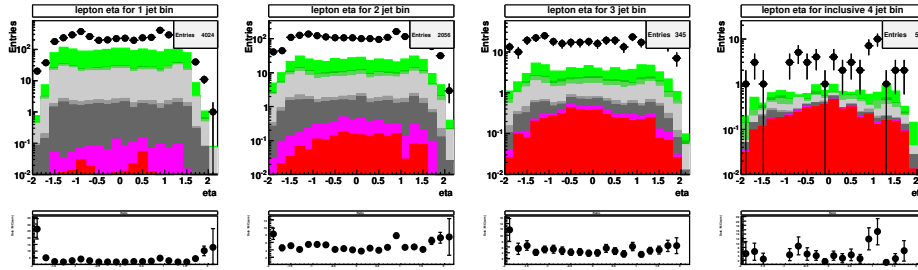
(a) tight samples, normal scale



(b) loose samples, normal scale



(c) tight samples, log scale



(d) loose samples, log scale

Figure B.6: Histograms for ϵ_{QCD} calculation for RunIIb2 μ +jets using $t\bar{t}$ selection cuts. These are all lepton η histograms used to calculate the central value for ϵ_{QCD} . From left to right the histograms show the exclusive 1, 2, 3 jet bin, and the inclusive 4 jet bin.

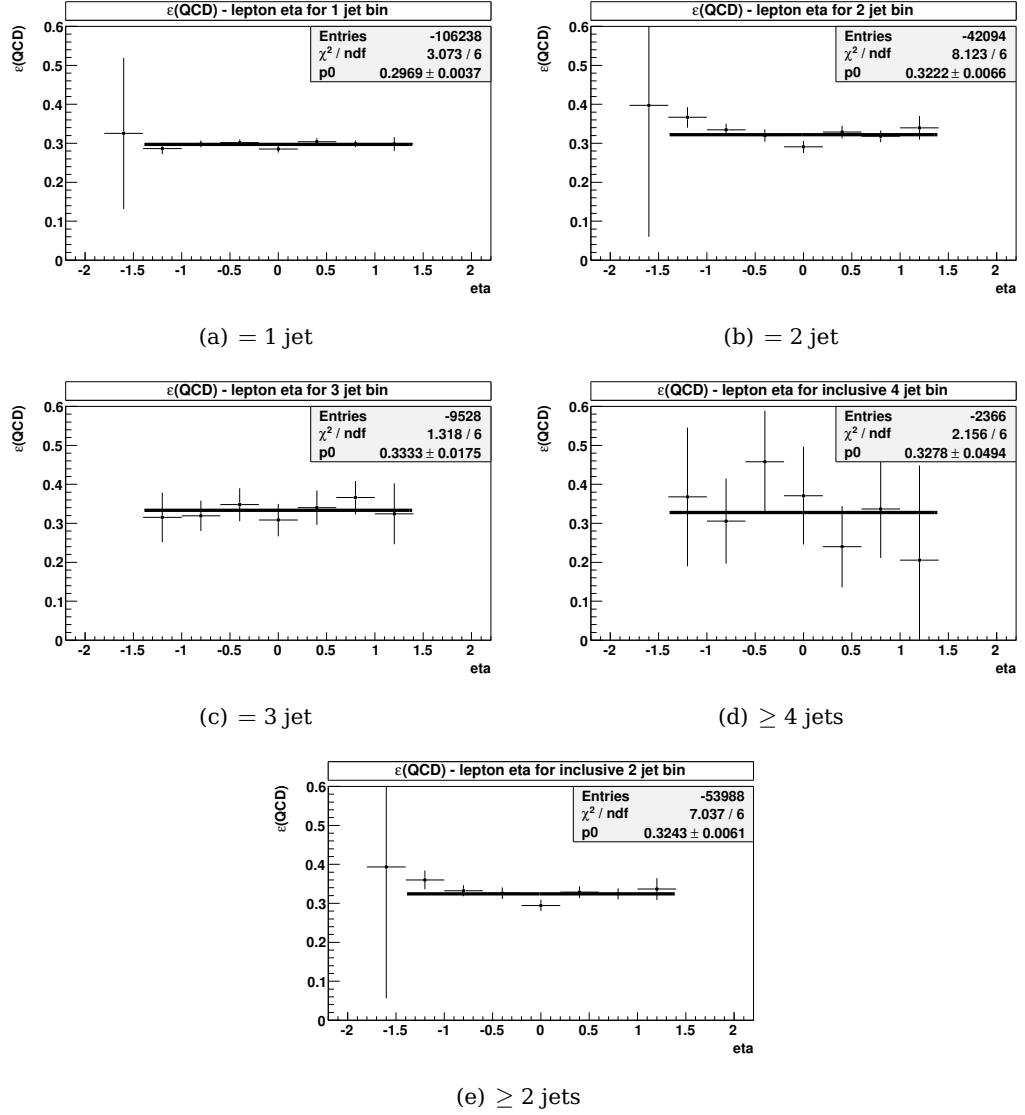
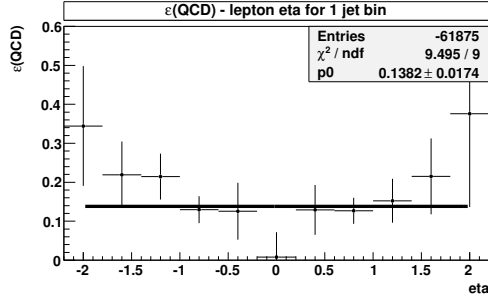
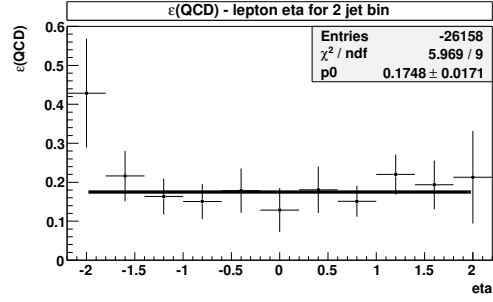


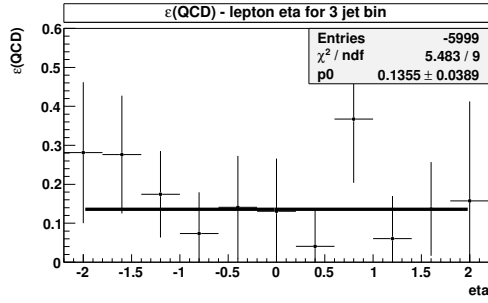
Figure B.7: ϵ_{QCD} plots for RunIIb2 e+jets using $t\bar{t}$ selection cuts. These ratio plots are used to determine the central value of ϵ_{QCD} .



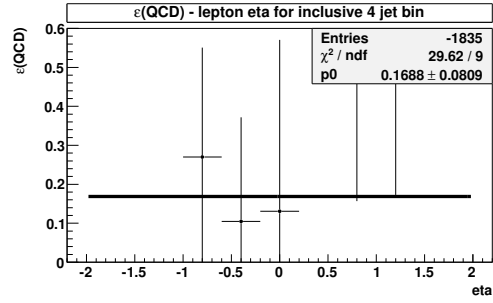
(a) = 1 jet



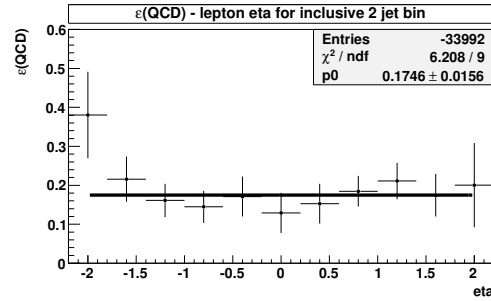
(b) = 2 jet



(c) = 3 jet



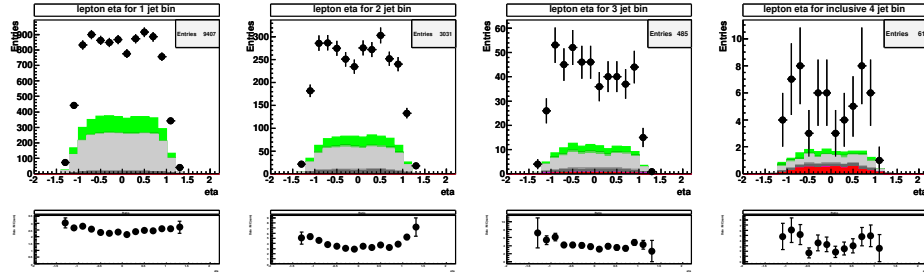
(d) ≥ 4 jets



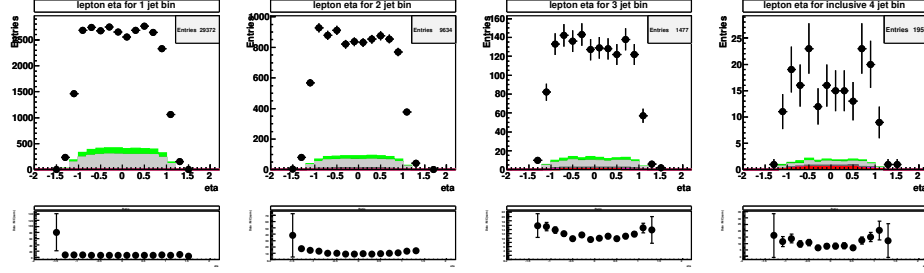
(e) ≥ 2 jets

Figure B.8: ϵ_{QCD} plots for RunIIb2 μ +jets using $t\bar{t}$ selection cuts. These ratio plots are used to determine the central value of ϵ_{QCD} .

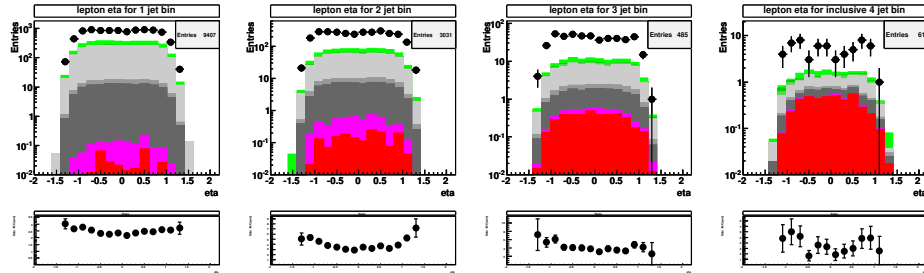
Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set



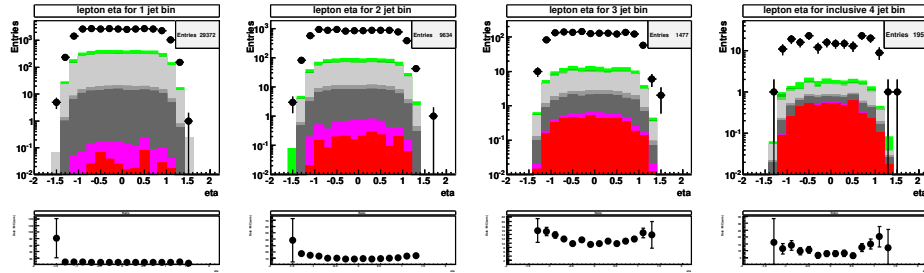
(a) tight samples, normal scale



(b) loose samples, normal scale



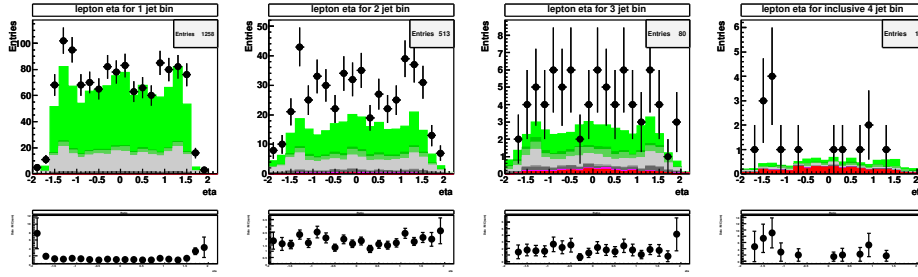
(c) tight samples, log scale



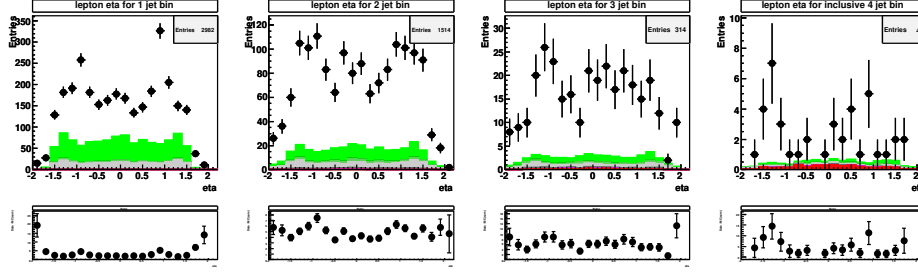
(d) loose samples, log scale

Figure B.9: Histograms for ϵ_{QCD} calculation for RunIIb3 e+jets using $t\bar{t}$ selection cuts. These are all lepton η histograms used to calculate the central value for ϵ_{QCD} . From left to right the histograms show the exclusive 1, 2, 3 jet bin, and the inclusive 4 jet bin.

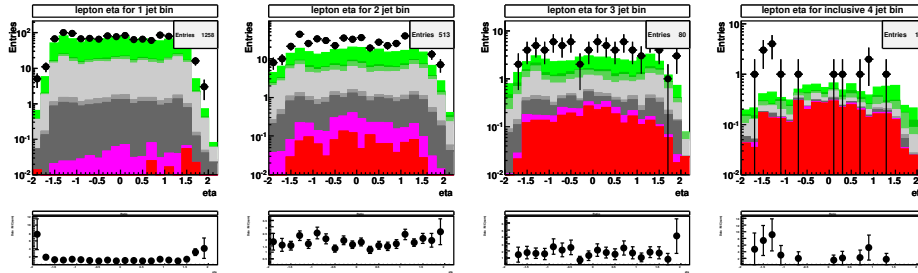
Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set



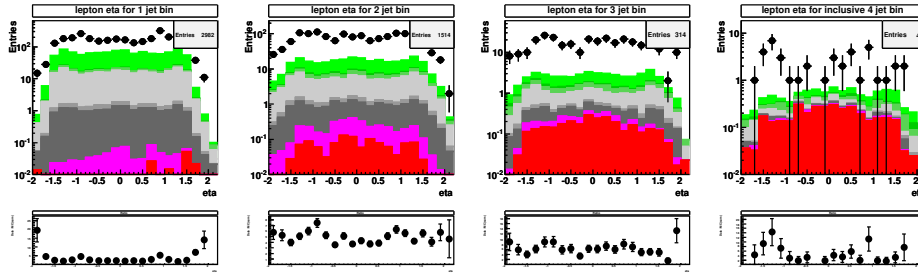
(a) tight samples, normal scale



(b) loose samples, normal scale



(c) tight samples, log scale



(d) loose samples, log scale

Figure B.10: Histograms for ϵ_{QCD} calculation for RunIIb3 μ +jets using $t\bar{t}$ selection cuts. These are all lepton η histograms used to calculate the central value for ϵ_{QCD} . From left to right the histograms show the exclusive 1, 2, 3 jet bin, and the inclusive 4 jet bin.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

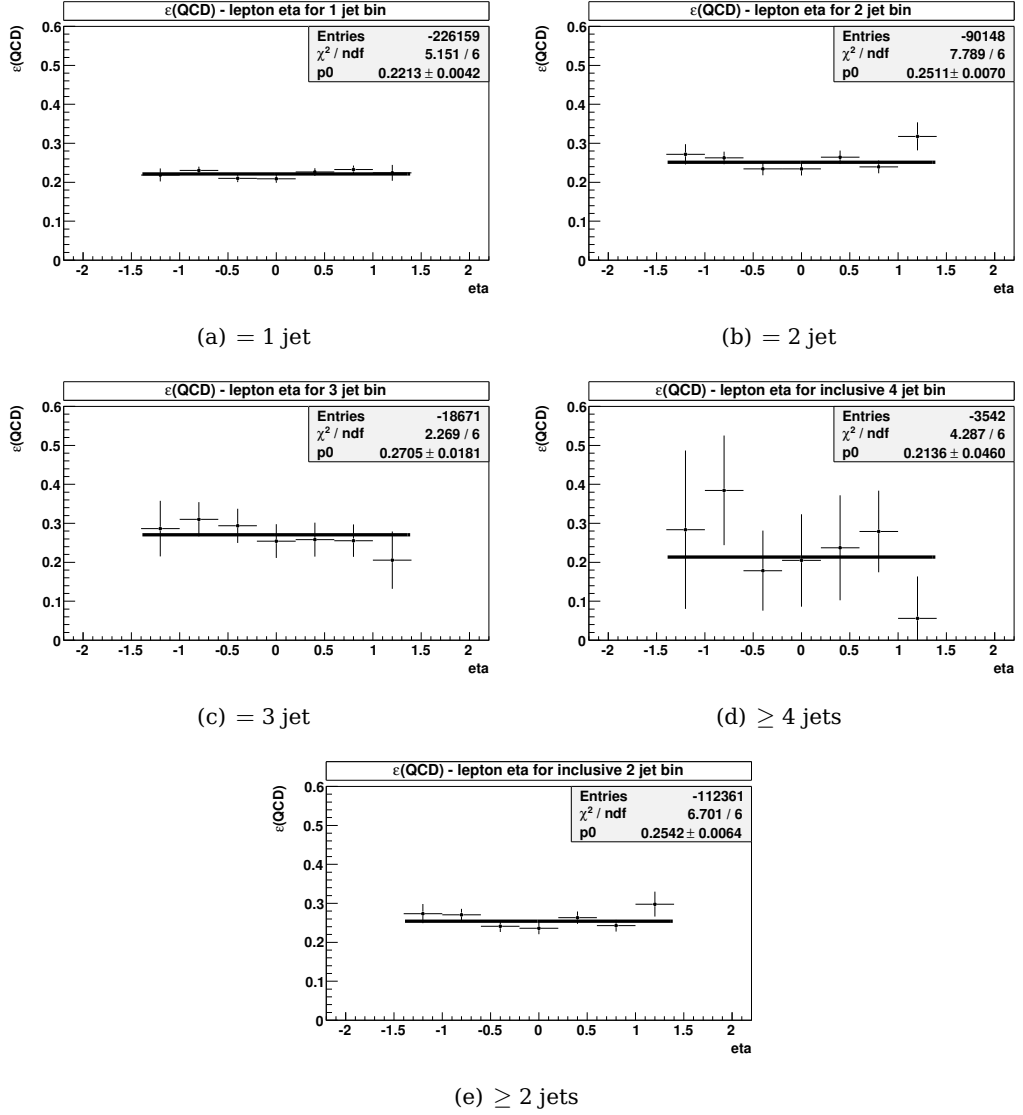


Figure B.11: ϵ_{QCD} plots for RunIIb3 e+jets using $t\bar{t}$ selection cuts. These ratio plots are used to determine the central value of ϵ_{QCD} .

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

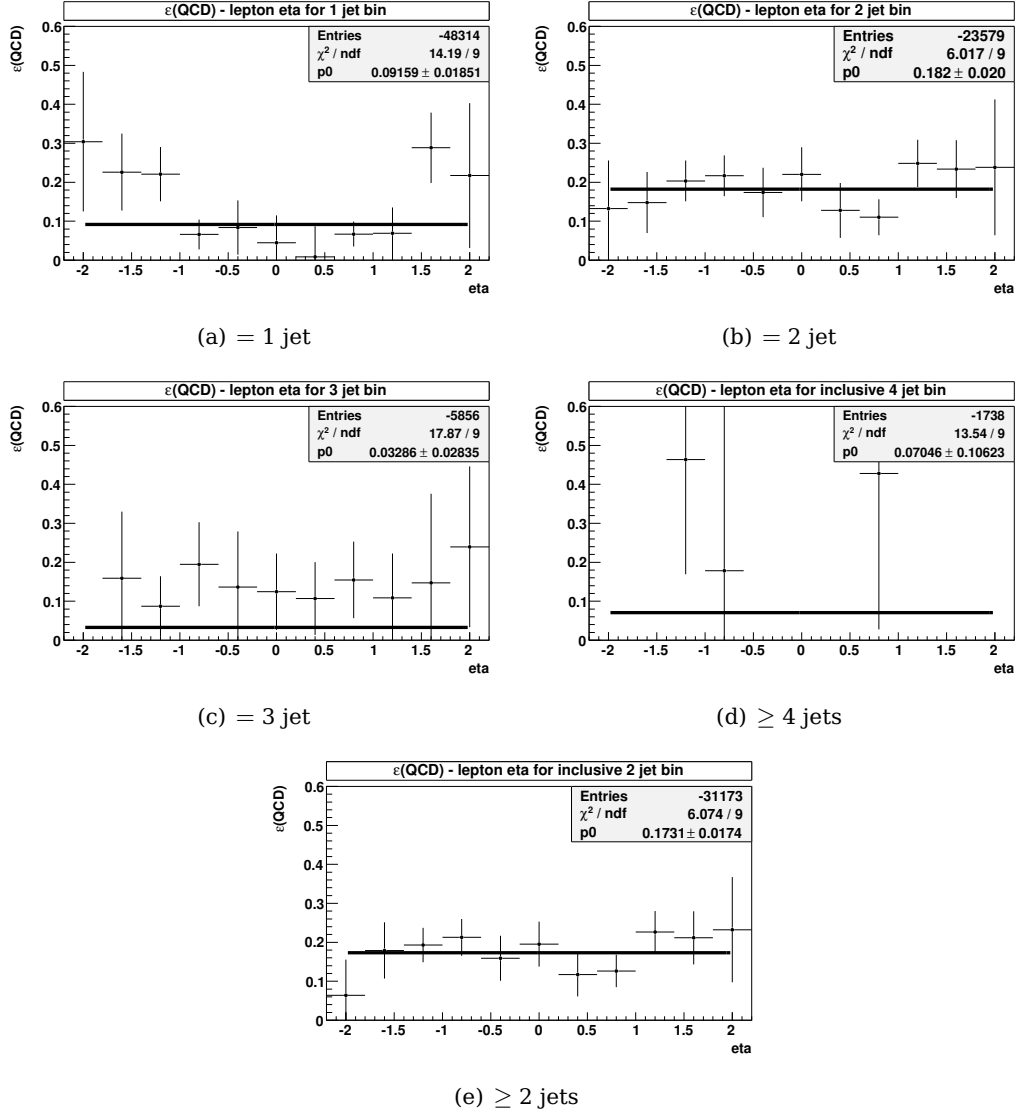
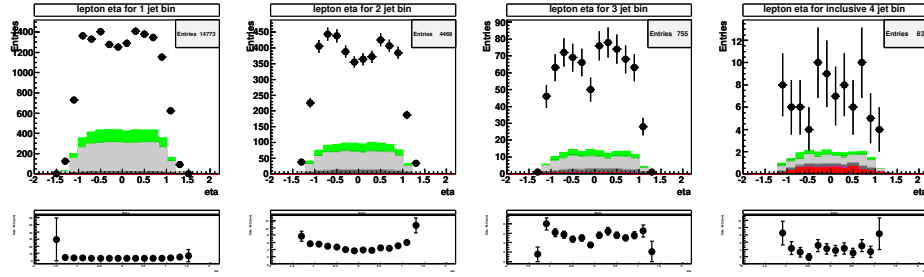
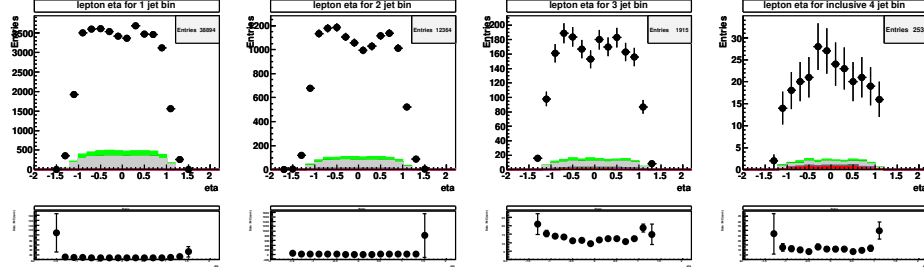


Figure B.12: ϵ_{QCD} plots for RunIIb3 μ +jets using $t\bar{t}$ selection cuts. These ratio plots are used to determine the central value of ϵ_{QCD} .

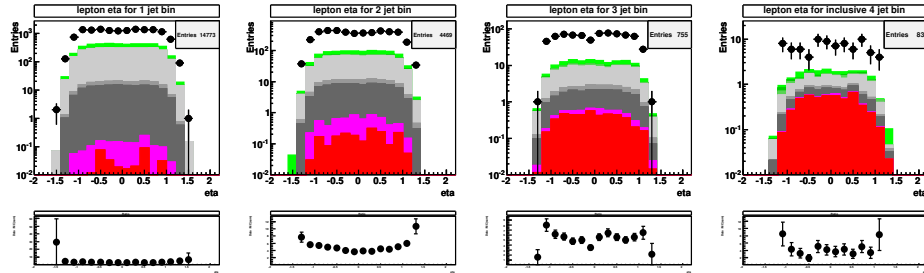
Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set



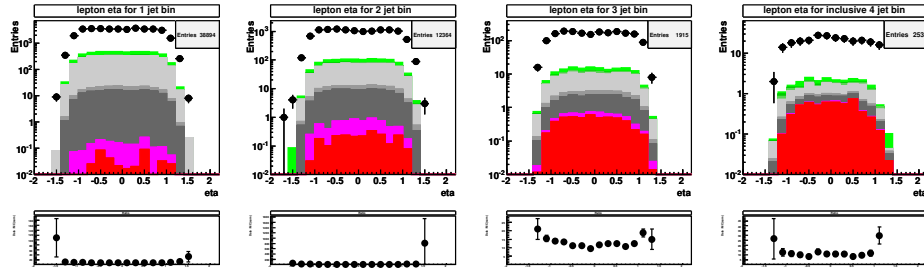
(a) tight samples, normal scale



(b) loose samples, normal scale



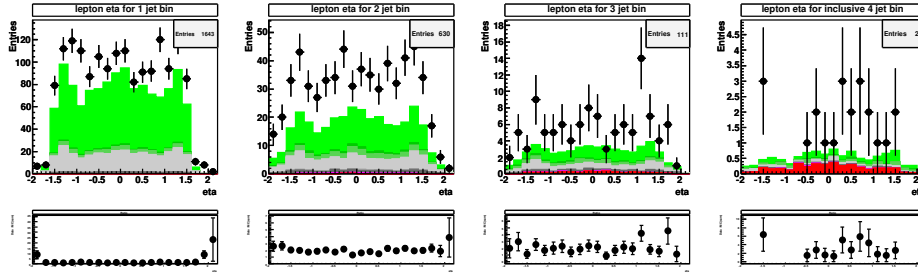
(c) tight samples, log scale



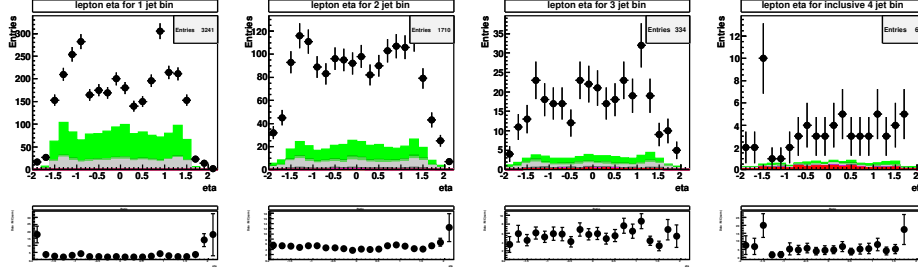
(d) loose samples, log scale

Figure B.13: Histograms for ϵ_{QCD} calculation for RunIIb4 e+jets using $t\bar{t}$ selection cuts. These are all lepton η histograms used to calculate the central value for ϵ_{QCD} . From left to right the histograms show the exclusive 1, 2, 3 jet bin, and the inclusive 4 jet bin.

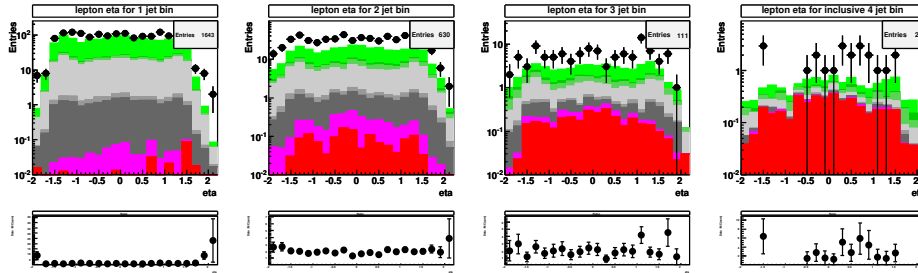
Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set



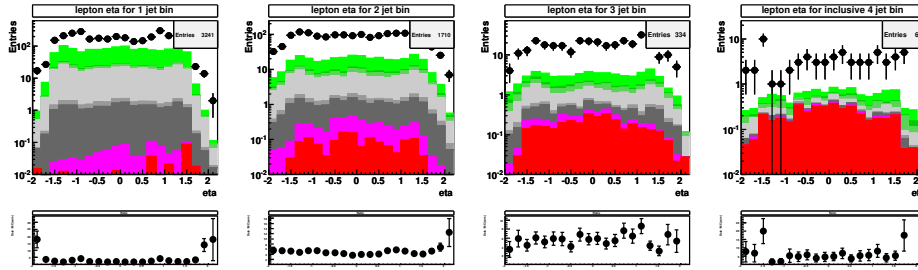
(a) tight samples, normal scale



(b) loose samples, normal scale



(c) tight samples, log scale



(d) loose samples, log scale

Figure B.14: Histograms for ϵ_{QCD} calculation for RunIIb4 μ +jets using $t\bar{t}$ selection cuts. These are all lepton η histograms used to calculate the central value for ϵ_{QCD} . From left to right the histograms show the exclusive 1, 2, 3 jet bin, and the inclusive 4 jet bin.

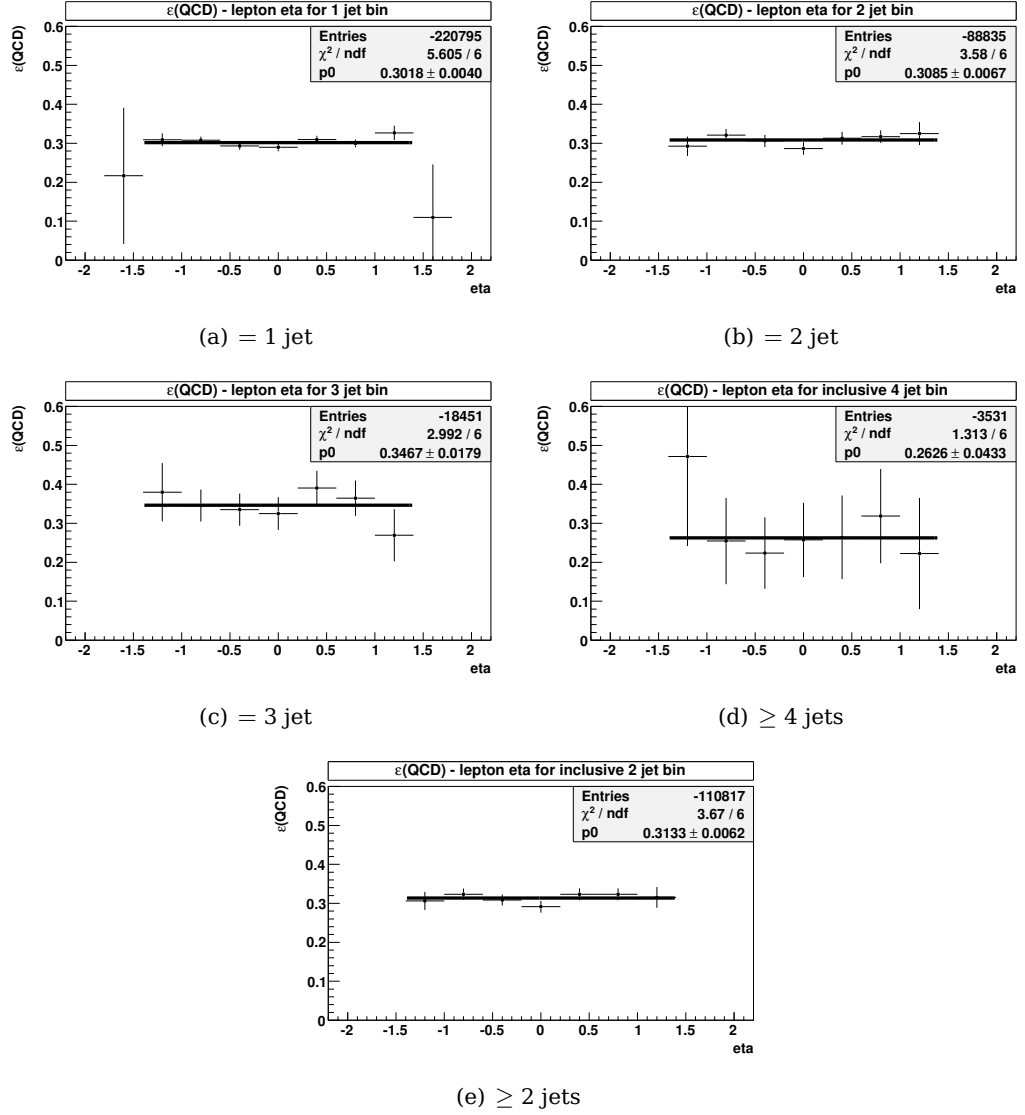
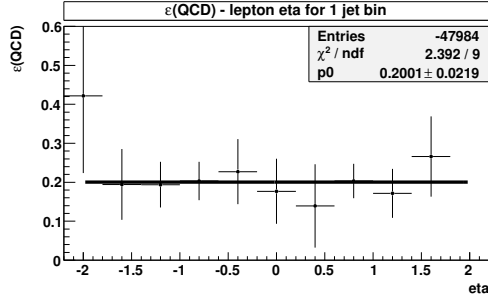
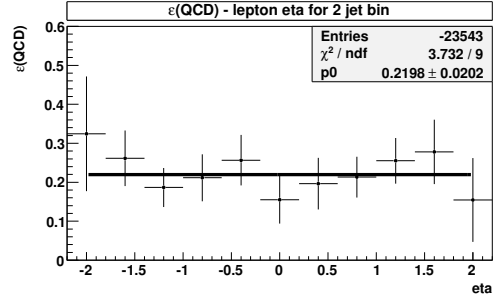


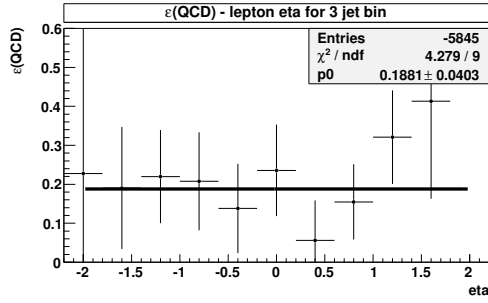
Figure B.15: ϵ_{QCD} plots for RunIIb4 e+jets using $t\bar{t}$ selection cuts. These ratio plots are used to determine the central value of ϵ_{QCD} .



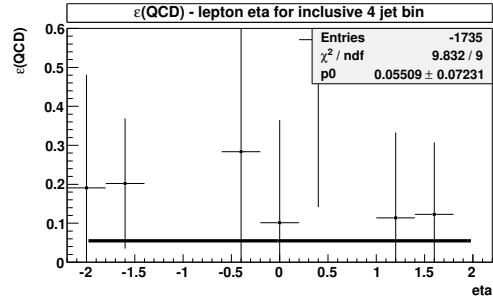
(a) = 1 jet



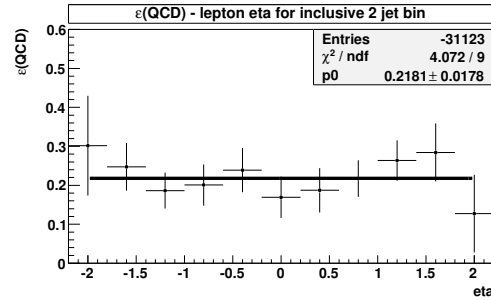
(b) = 2 jet



(c) = 3 jet



(d) ≥ 4 jets



(e) ≥ 2 jets

Figure B.16: ϵ_{QCD} plots for RunIIb4 μ +jets using $t\bar{t}$ selection cuts. These ratio plots are used to determine the central value of ϵ_{QCD} .

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

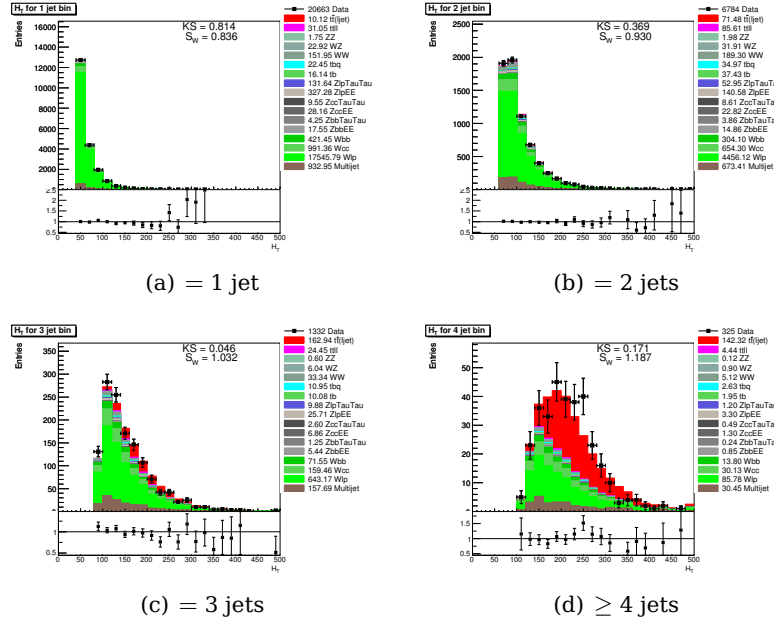


Figure B.17: H_T plots for RunIIb1 e+jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets.

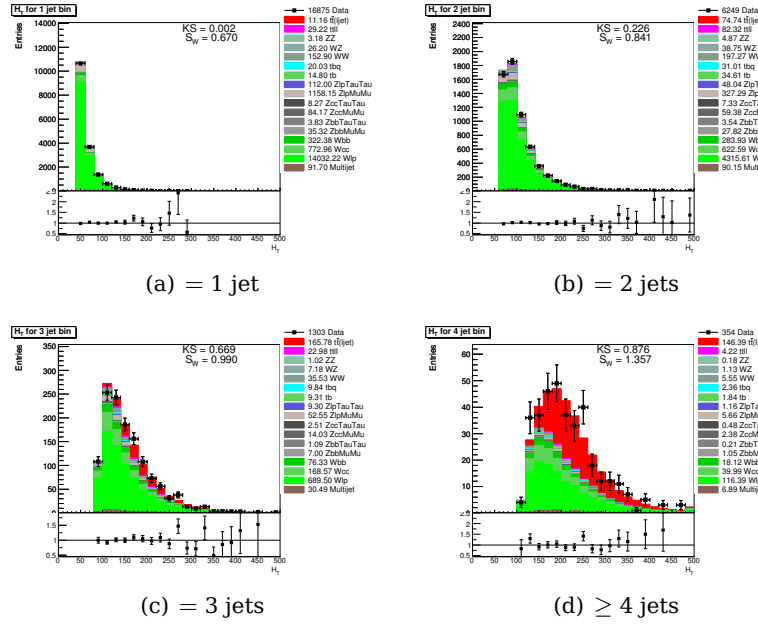


Figure B.18: H_T plots for RunIIb1 μ +jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

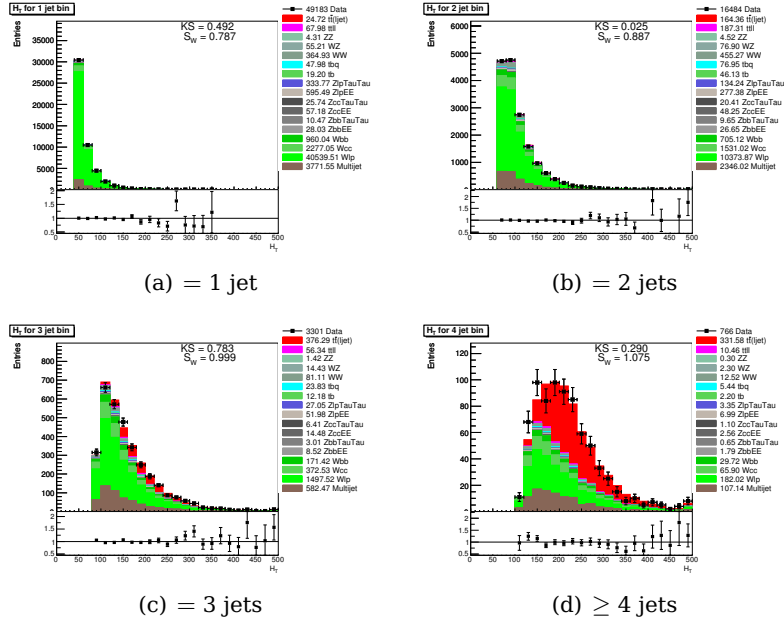


Figure B.19: H_T plots for RunIIb2 e+jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets.

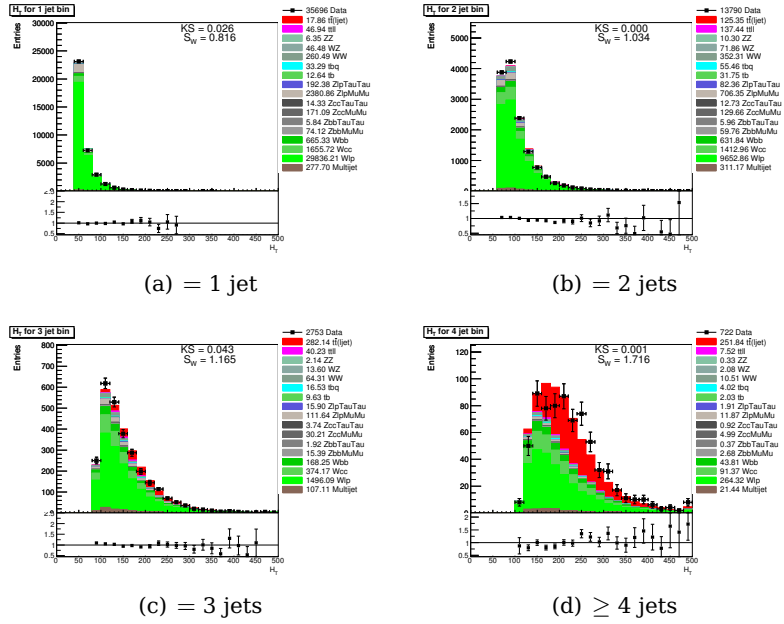


Figure B.20: H_T plots for RunIIb2 μ +jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

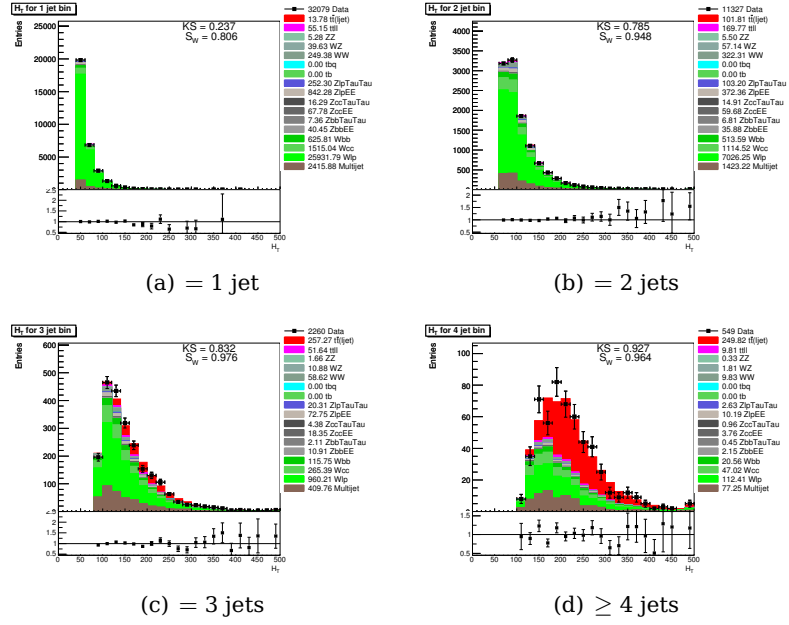


Figure B.21: H_T plots for RunIIb3 e+jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets.

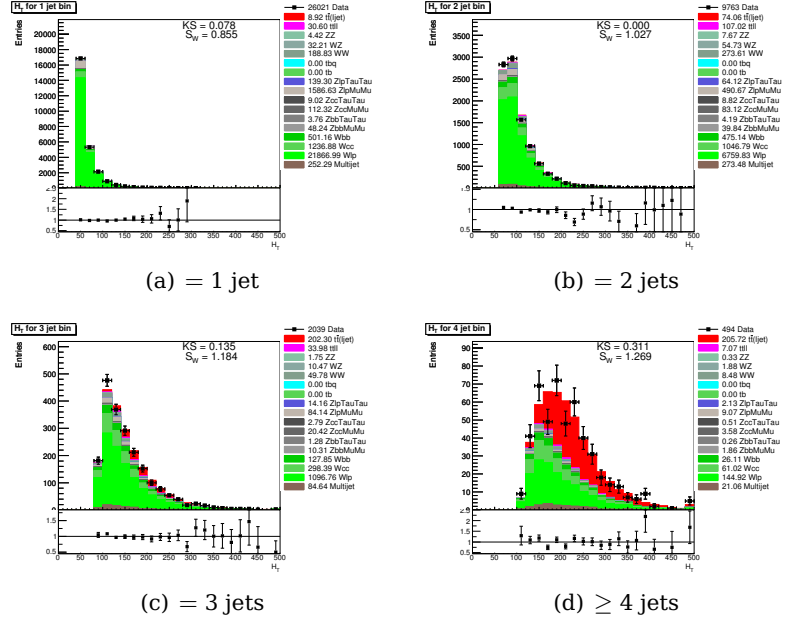


Figure B.22: H_T plots for RunIIb3 μ +jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

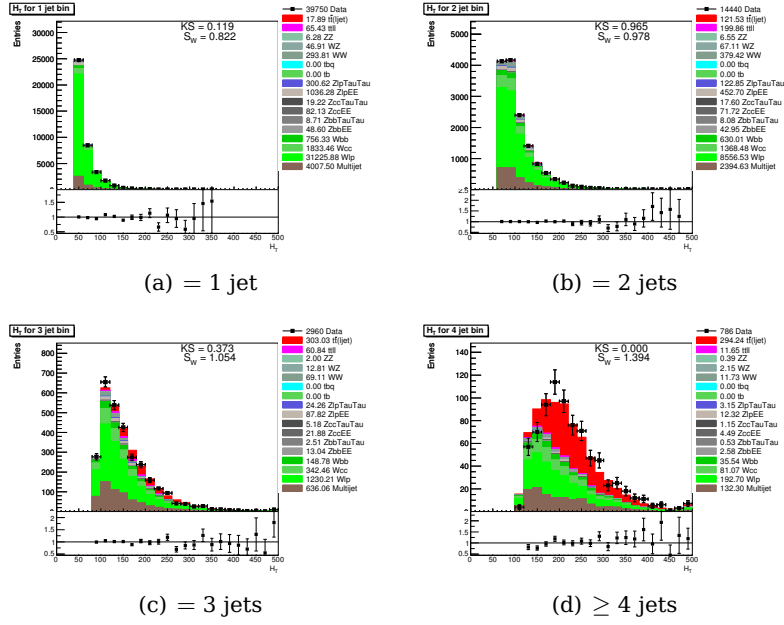


Figure B.23: H_T plots for RunIIb4 e+jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets.

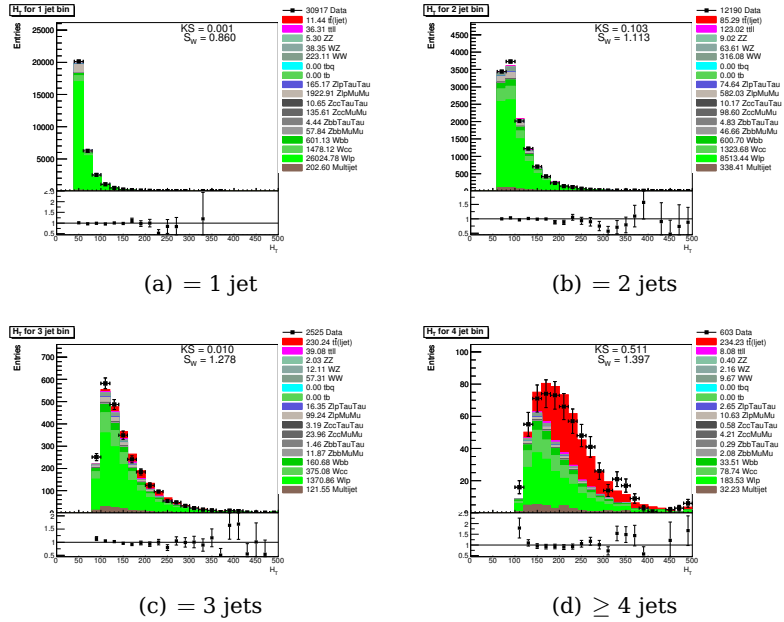


Figure B.24: H_T plots for RunIIb4 μ +jets using the $t\bar{t}$ selection cuts, before applying any b-tag for jets.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

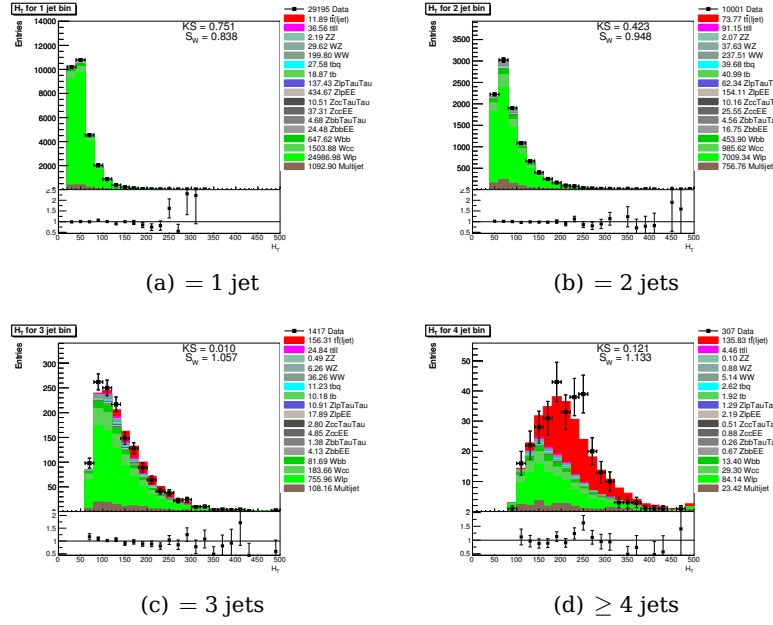


Figure B.25: H_T plots for RunIIb1 e+jets using the single top cuts, before applying any b-tag for jets.

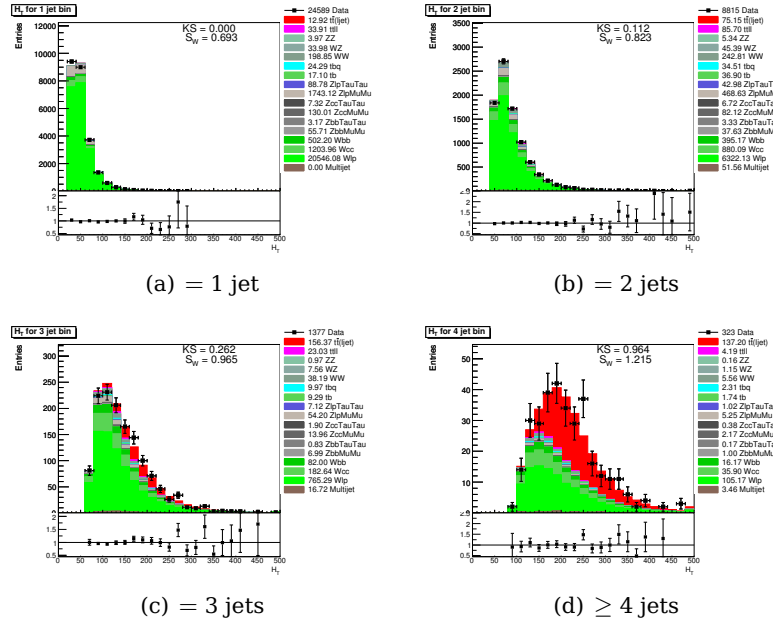


Figure B.26: H_T plots for RunIIb1 μ +jets using the single top cuts, before applying any b-tag for jets.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

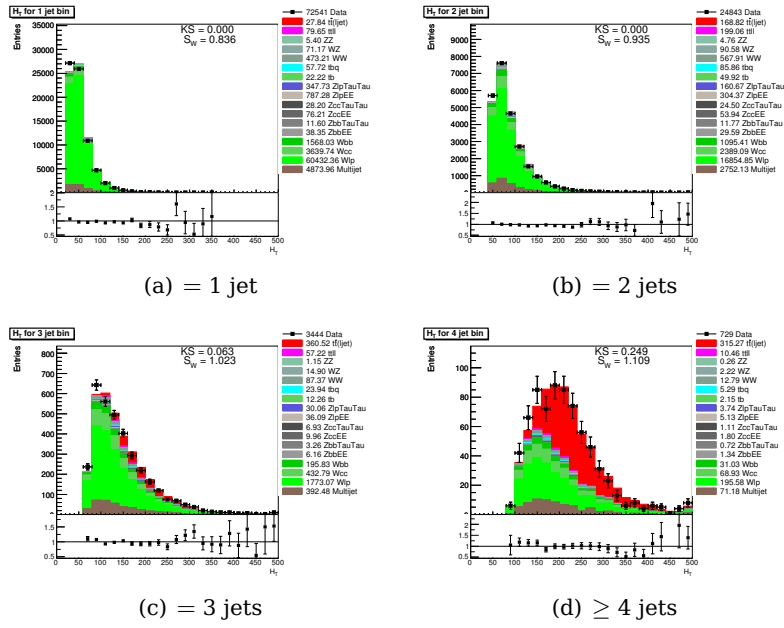


Figure B.27: H_T plots for RunIIb2 e+jets using the single top cuts, before applying any b-tag for jets.

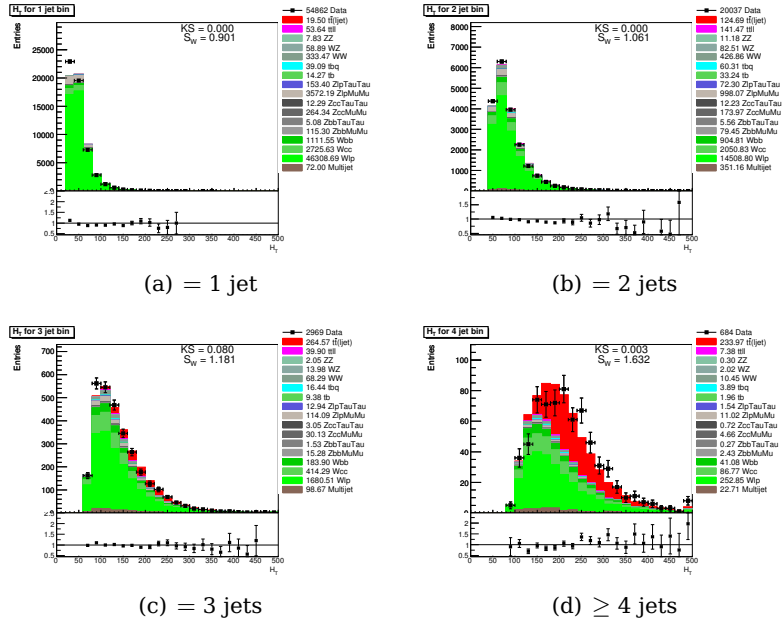


Figure B.28: H_T plots for RunIIb2 μ +jets using the single top cuts, before applying any b-tag for jets.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

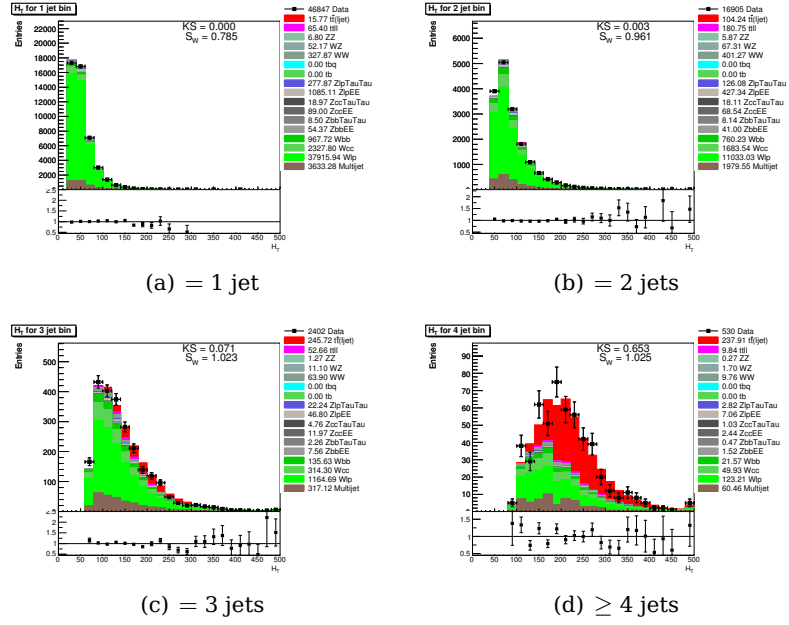


Figure B.29: H_T plots for RunIIb3 e+jets using the single top cuts, before applying any b-tag for jets.

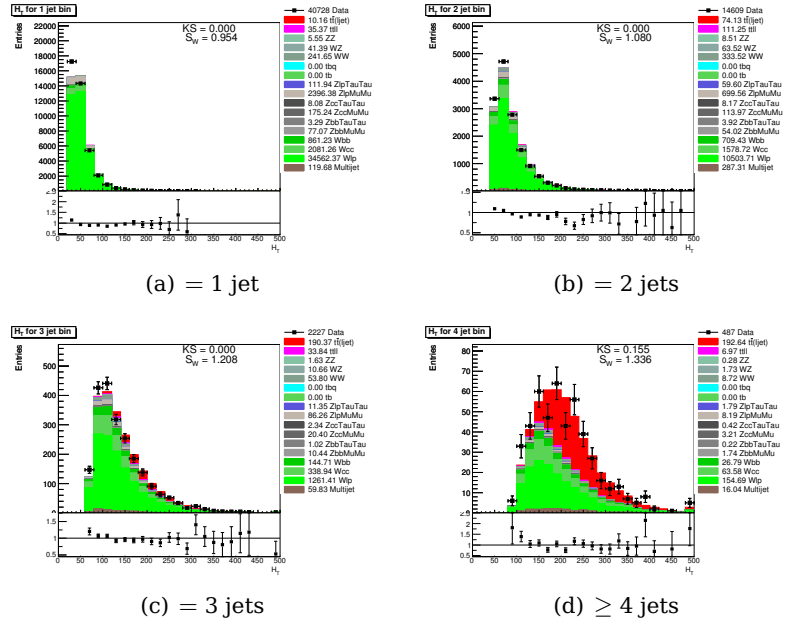


Figure B.30: H_T plots for RunIIb3 μ +jets using the single top cuts, before applying any b-tag for jets.

Fake Rate for the lepton+jets Selection using the Full RunIIb Data Set

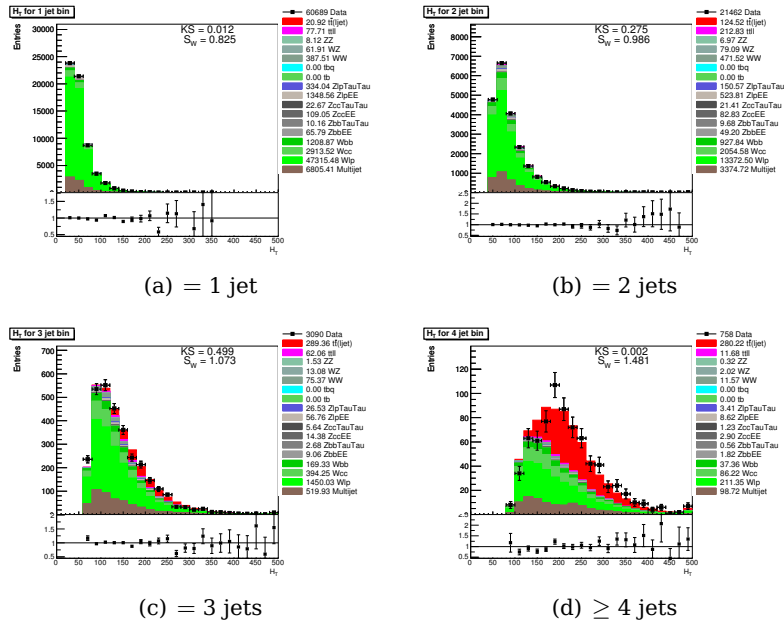


Figure B.31: H_T plots for RunIIb4 e+jets using the single top cuts, before applying any b-tag for jets.

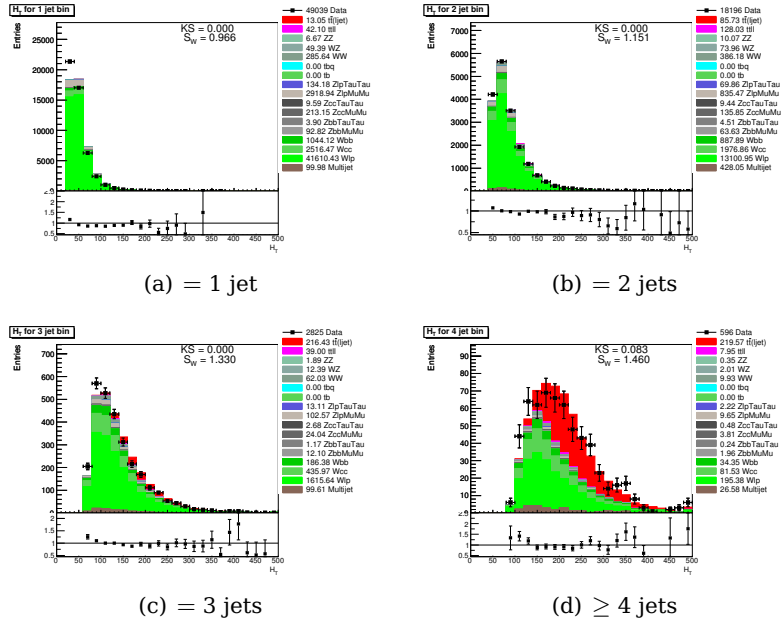


Figure B.32: H_T plots for RunIIb4 μ +jets using the single top cuts, before applying any b-tag for jets.