UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Física

Medición de la sección eficaz de producción de single top en colisiones protón-antiprotón a 1.96 TeV

Trabajo de Tesis para optar por el título de
Doctor de la Universidad de Buenos Aires en el área Ciencias Físicas

por Lic. Andrés Jorge Tanasijczuk

Director de Tesis: Dr. Ricardo Piegaia
Director Asistente de Tesis: Dr. Gustavo Javier Otero y Garzón
Lugar de Trabajo: Depto. de Física, Facultad de Ciencias Exactas y Naturales, UBA

25 de Marzo de 2010

Resumen

La presente tesis describe la búsqueda de quarks top producidos de manera simple (a diferencia de la producción en pares top-antitop) a través de un vértice de la interacción electrodébil en colisiones protón-antiprotón a una energía de centro de masa de $\sqrt{s} =$ 1.96 TeV. El análisis utiliza un total de 2.3 fb^{-1} de datos colectados con el detector DØ en Fermilab, correspondientes a dos períodos diferentes de corrida del colisionador Tevatron. En el Tevatron son dos los canales que contribuyen a la producción simple de quarks top, el canal s y el canal t. En el canal s, la aniquilación de un quark y un antiquark produce un bosón W virtual, el cual luego decae en un quark top y un quark bottom. El quark top decae casi de manera exclusiva en un bosón W y un quark bottom. Aquí se consideran los estados finales en los que el bosón W decae leptónicamente en un electrón o un muón más un neutrino. Así, a nivel del detector, el estado final que caracteriza al canal s contiene un leptón, energía faltante que da cuenta del neutrino, y dos jets correspondientes a los dos quarks bottom. En el canal t, el estado final contiene un jet adicional proveniente de un quark liviano. Claramente, una reconstrucción precisa de los eventos requiere una medición precisa de la energía de los jets. Dado que la cantidad de eventos de fondo es enorme aún después de aplicar una selección de eventos, se utiliza una técnica de multivariables, redes neuronales bayesianas, para separar la señal del fondo. Para medir la sección eficaz de quarks top simples se computa una probabilidad bayesiana. Asumiendo que el exceso observado se debe a eventos de quarks top simples, la sección eficaz de producción de quarks top simples medida es

$$\sigma\left(p\bar{p} \to tb + X, tqb + X\right) = 4.70^{+1.18}_{-0.93} \text{ pb}.$$

El exceso observado se asocia con un valor-p de $(3.2 \pm 2.3) \times 10^{-8}$, asumiendo la hipótesis de la presencia de únicamente eventos de fondo. Dicho valor-p corresponde a un exceso sobre el fondo de 5.4 desviaciones estándar para una densidad gaussiana. El valor-p computado usando la sección eficaz del modelo estándar de 3.46 pb para la señal es $(22.7 \pm 0.6) \times 10^{-6}$, correspondiente a una significancia esperada de 4.08 desviaciones estándar.

Palabras clave: top quark simple, redes neuronales Bayesianas, calibración de la energía de los jets, Tevatron.

Abstract

Measurement of the single top production cross section in proton-antiproton collisions at 1.96 TeV

This thesis describes a search for singly produced top quarks via an electroweak vertex in head-on proton-antiproton collisions at a center of mass energy of $\sqrt{s} = 1.96$ TeV. The analysis uses a total of 2.3 fb^{-1} of data collected with the DØ detector at Fermilab, corresponding to two different run periods of the Tevatron collider. Two channels contribute to single top quark production at the Tevatron, the s-channel and the t-channel. In the s-channel, a virtual W boson is produced from the aniquilation of a quark and an antiquark and a top and a bottom quarks are produced from the W decay. The top quark decays almost exclusively into a W boson and a bottom quark. Final states are considered in which the W boson decays leptonically into an electron or a muon plus a neutrino. Thus, at the detector level, the final state characterizing the s-channel contains one lepton, missing energy accounting for the neutrino, and two jets from the two bottom quarks. In the t-channel, the final state has an additional jet coming from a light quark. Clearly, a precise reconstruction of the events requires a precise measurement of the energy of the jets. A multivariate technique, Bayesian neural networks, is used to extract the single top signal from the overwhelming background still left after event selection. A Bayesian likelihood probability is then computed to measure the single top cross section. Assuming the observed excess is due to single top events, the measured single top quark production cross section is

$$\sigma\left(p\bar{p} \to tb + X, tqb + X\right) = 4.70^{+1.18}_{-0.93} \text{ pb}.$$

The observed excess is associated with a p-value of $(3.2 \pm 2.3) \times 10^{-8}$, assuming the background-only hypothesis. This p-value corresponds to an excess over background of 5.4 standard deviations for a Gaussian density. The p-value computed using the standard model signal cross section of 3.46 pb is $(22.7 \pm 0.6) \times 10^{-6}$, corresponding to an expected significance of 4.08 standard deviations.

Keywords: single top, Bayesian neural networks, jet energy scale, Tevatron.

Acknowledgments

Working in experimental particle physics has been quite a change for me. During my whole undergraduate studies I knew I wanted to be a theoretical physicist, devoting my life to think about space and time and quantum theory. Maybe I was too much a philosopher rather than a real theoretical physicist, and maybe because of that I also think the life had a different plan for me. After a year and half of trying to define my future career, I decided to visit Dr. Ricardo Piegaia in the Universidad de Buenos Aires. I introduced myself and asked him what is he working on and if there would be a possibility I could join his team and do a doctorate under his supervision. At that time he did not know me, and he trusted me:-) That is the first reason I will always be thankful to Ricardo. Muchas gracias Ricardo por todo el apoyo que me brindaste durante estos años. Y quiero aprovechar la ocasión para felicitarte (ya se que esta es una sección de agradecimientos y no de felicitaciones, pero siento la necesidad de hacerlo) por tu gran dedicación y compromiso para con la educación en nuestra facultad.

Half a year later, in January 2006, I was in Fermilab. Here I should make a pause and express my special thanks to the DØ spokespersons who always supported argentinian (and of course other) students when funds were needed. As soon as I arrived to Fermilab, some people helped me to accommodate myself and make my life easier, but nobody more than Charly with whom I shared a home for the next two years. It always happens like that, you will always find someone who is ready to help you. But what rare times happens is that with the time you become friends. So, this is the time to thank Charly, Jorge, Murilo, Titto, Gustavo, Victor and José, the friends the life gave me in my stay in Fermilab. I will always remember the soccer games in the Fermilab gym, going out to restaurants with them and hanging out in Chicago.

I spent my first year and half in Fermilab working within the DØ Jet Energy Scale group. My closest interaction was with Aurelio Juste, who led the group at that time, and Mikko Voutilainen. I want to express my gratitude to both of them for their explanations about JES, which they knew I could barely understand, but anyways they always tried to make from me a better scientist. Many thanks to Mikko also for his continuous help on technical stuff as ROOT and JES related codes. My thanks must to be extended to the other members of the group too: Christophe Royon, Jiří Kvita, Jeroen Hegeman, Alexander Kupco, Jochen Cammin, David Lam and Zdeněk Hubáček (I hope I am not forgetting anybody).

During the second half of my stay in Fermilab I worked within the DØ single top group. I had the fortune to met wonderful people there too, principally Cecilia Gerber and Gustavo Otero y Garzón who guided me along this time. But since this thesis would not have been possible without all of the single top people working as a group, all of

them deserve my thankfulness: Reinhard Schwienhorst and Ann Heinson (conveners together with Cecilia Gerber), Ernest Aguiló, Jorge Benitez, Dag Gillberg, Ike Hall, Shabnam Jabeen, Supriya Jain, Liang Li, Zhiyi Liu, Gustavo Otero y Garzón, Monica Pangilinan, Harrison Prosper and others.

Finally, the last lines are devoted to my love. As I said, I think I am more a philosopher than a real physicist. I profoundly feel that the reason I went to USA was to find the love of my life, Marta Kulis. I met her in July 2007, and today she is my wife, the best wife I could ever imagen. Dziękuję Ci mója kochana, za całe twoje wsparcie. Za to, ie jesteś osobą petną dobrych uczuć i głębokiego myślenia.

Abbreviations and Acronyms

CDF	Collider Detector at Fermilab
LEP	Large Electron-Positron Collider (at CERN)
LHC	Large Hadron Collider (at CERN)
SLD	Stanford Large Detector (at SLAC)
CC	Central Calorimeter
EC	Endcap Calorimeter
ICD	InterCryostat Detector
ICR	InterCryostat Region
EM	ElectroMagnetic
CH	Coarse Hadronic (calorimeter)
FH	Fine Hadronic (calorimeter)
CFT	Central Fiber Tracker
SMT	Silicon Microstrip Tracker
PV	Primary Vertex
SV	Secondary Vertex
JER	Jet Energy Resolution
JES	Jet Energy Scale
PhES	Photon Energy Scale
MB	Minimum Bias
ZB	Zero Bias
NP	Noise and Pile-up
MI	Multiple Interactions
UE	Underlying Event
MC	Monte Carlo
NN	Neural Network
QCD	Quantum ChromoDynamics
QED	Quantum ElectroDynamics
QFT	Quantum Field Theory
SM	Standard Model

Contents

1 Theoretical Framework **1**
- 1.1 The Standard Model . . . 1
  - 1.1.1 Spinors . . . 2
  - 1.1.2 Matter Content . . . 3
  - 1.1.3 Gauge Symmetries and Interactions . . . 4
  - 1.1.4 The Higgs Mechanism . . . 10
- 1.2 Cross Sections . . . 11
- 1.3 The Top Quark . . . 15
  - 1.3.1 Top Quark Pair Production . . . 15
  - 1.3.2 Single Top Production . . . 16
  - 1.3.3 Top Quark Decay . . . 19
  - 1.3.4 Single Top Decay Channels . . . 19

2 Experimental Setup **21**
- 2.1 Fermilab's Collider Complex . . . 21
- 2.2 The DØ Detector . . . 23
  - 2.2.1 Tracking System . . . 25
  - 2.2.2 Silicon Microstrip Tracker . . . 26
  - 2.2.3 Central Fiber Tracker . . . 28
  - 2.2.4 Solenoidal Magnet . . . 29
  - 2.2.5 The Calorimeters . . . 30
  - 2.2.6 Preshower Detectors . . . 34
  - 2.2.7 Muon System . . . 35
  - 2.2.8 Luminosity Monitor . . . 37
  - 2.2.9 Trigger System . . . 38

3 Event Reconstruction and Object Identification **42**
- 3.1 Tracks . . . 42
- 3.2 Vertices . . . 44
  - 3.2.1 Primary Vertices . . . 45
  - 3.2.2 Secondary Vertices . . . 47
- 3.3 Calorimeter Clusters . . . 48
- 3.4 Leptons and Photons . . . 49
  - 3.4.1 Electromagnetic Clusters . . . 50
  - 3.4.2 Electrons . . . 51
  - 3.4.3 Photons . . . 53
  - 3.4.4 Muons . . . 54
- 3.5 Jets . . . 57
  - 3.5.1 Jets from b-quarks . . . 60
- 3.6 Missing Transverse Energy . . . 63

4 Jet Energy Scale **64**

4.1 Definition of the Corrections Contributing to the JES 65
4.2 Estimators of the JES Subcorrections 66
4.3 Data and Monte Carlo JES Samples 66
4.3.1 Data Samples 67
4.3.2 Monte Carlo Samples 67
4.4 Offset Correction 68
4.4.1 Noise and Pile-up 69
4.4.2 Multiple Interactions 69
4.4.3 Total Jet Offset Energy 71
4.4.4 Effect of Zero-Suppression on Offset 72
4.4.5 Uncertainties 73
4.5 Response Correction 74
4.5.1 Missing E_T Projection Fraction Method 74
4.5.2 Sample Selection 76
4.5.3 Photon Corrections: Dijet Background Contamination and Photon Energy Scale 77
4.6 Central Calorimeter Response 81
4.6.1 Uncertainties 83
4.7 Eta-Dependent Corrections 84
4.7.1 Resolution Bias in Dijet 86
4.7.2 Sample Dependence of the Relative Response 90
4.7.3 Global Fit 92
4.7.4 Summary of Results 93
4.7.5 Uncertainties 94
4.8 Zero-Suppression Bias Correction 96
4.9 Topology Bias Correction 99
4.10 Showering Correction 100
4.10.1 Methodology 101
4.10.2 Results 103
4.10.3 Uncertainties 103
4.11 Mapping the Measured Jet Energy to E' 105
4.12 Combined JES 107

5 Data and Simulation Samples **110**
5.1 Data Samples 110
5.2 Single Top Backgrounds 111
5.3 Simulation Samples 112
5.3.1 Single Top Monte Carlo Samples 113
5.3.2 Background Monte Carlo Samples 113
5.3.3 Monte Carlo Samples Normalization 114
5.4 Corrections to Monte Carlo Events 116
5.5 Treatment of the Multijets Background 118
5.5.1 Modeling 118

5.5.2 Multijets Reshaping . . . 119
5.5.3 Multijets and W+Jets Normalization to Data . . . 122
5.6 Further Normalization of the W+Heavy-Flavor Samples after b-Tagging 125

6 Event Selection **127**
6.1 Event Selection Cuts before b-tagging . . . 127
6.1.1 Generic Cuts . . . 127
6.1.2 Triangle Cuts . . . 129
6.1.3 Total Transverse Energy Cuts . . . 132
6.1.4 Cut Flow Tables for MC Signal and Data . . . 135
6.2 W+Jets Samples Reweightings . . . 138
6.3 Selection of b-Tag Events . . . 144
6.3.1 Taggability Rate Functions . . . 144
6.3.2 b-Tagging Event Weights and b-Jet Assignment Combinations . 146
6.4 Number of Events after Selection . . . 148
6.5 Event Yields . . . 149
6.6 Data versus Background Model Comparison . . . 156
6.7 Cross Check Samples . . . 172
6.8 Systematic Uncertainties . . . 174

7 Bayesian Neural Networks **180**
7.1 Splitting of the Samples . . . 180
7.2 Overview . . . 181
7.3 BNN Structure and Mathematics . . . 182
7.3.1 BNN Structure . . . 182
7.3.2 BNN Mathematics . . . 182
7.4 Analysis Road-Map . . . 186
7.5 Selection of Variables with High Discrimination Power . . . 187
7.6 BNN Training and Verification . . . 188
7.6.1 Training . . . 188
7.6.2 Treatment of Noise in the Training Data . . . 189
7.6.3 Verification . . . 190
7.7 BNN Output Transformation and Histogram Binning . . . 194
7.7.1 Monotone Transformation . . . 194
7.7.2 Binning Choice . . . 194
7.8 BNN Outputs . . . 196

8 Measurements **199**
8.1 Methodology . . . 199
8.1.1 Bayesian Approach . . . 199
8.1.2 Numerical Calculation − Prior Modeling . . . 201
8.2 Definition of Ensemble Tests, and Expected and Observed Measurements 203
8.3 Expected Results . . . 203

8.3.1 Stability Test . . . 206
8.4 Ensemble Tests . . . 208
8.5 Observed Results . . . 210
8.6 Significance . . . 214
8.7 Measurement of $|V_{tb}|$ . . . 216
8.8 Combination with the Boosted Decision Trees and Matrix Elements Analyses . . . 218

9 Summary **221**

Appendices **221**

A Studies on the Instrumental Background in Electron+Jets+$\not\!\!E_T$ Final State Selections **222**
A.1 The Matrix Method . . . 223
A.1.1 The Matrix Method as a Tool to get the Normalization of Backgrounds . . . 223
A.1.2 Using the Matrix Method to improve the Instrumental Background Model . . . 227
A.2 Monte Carlo Samples . . . 228
A.3 Fake Rate . . . 229
A.4 The Triangle Cut . . . 231
A.5 Effect of the Triangle Cut on the Fake Rate . . . 232
A.6 Reshaping of the loose-minus-tight Model . . . 232

B Trigger Efficiency **245**

C Plots from the Cross-Check Samples **250**

D Shape-Changing Systematic Uncertainties **257**

E Selected BNN Input Variables **268**

F Verification Plots **300**

G BNN Discriminant Plots **313**

1 Theoretical Framework

Since the times of the greek philosophers, the idea that all the matter in our universe is composed of fundamental indivisible entities (at that time called atoms) has never been left aside. Moreover, the behavior of the atoms should follow a rational set of laws and all of the world that surrounds us should be explained by these laws. These concepts gave birth to physics as a natural science. Nowadays, the branch of physics dedicated to the study of the fundamental constituents of matter and the forces acting between them, is called particle physics.

All the known particles that have been experimentally observed so far can be explained in terms of a small set of elementary particles (particles that are believed not to have further internal structure). Quarks for example, are part of these fundamental particles. Similarly, all the known forces in nature can be reduced to a set of four interactions: gravity (the weakest of all, but the most familiar and the first to be discovered), electromagnetism, the weak interaction (discovered by being responsible for the decay of radiactive atoms) and the strong interaction (originally proposed to be responsible for keeping together protons and neutrons in the atomic nucleous). Our current best understanding of elementary particles and their interactions is summarized in the so-called Standard Model (SM) of particle physics. The SM includes three of the four interactions: the strong, the electromagnetic and the weak interactions. The fourth interaction, gravity, is well described at macroscopic scales by Einstein's theory of General Relativity, but has resisted along the years to the many intents to be incorporated into the quantum-mechanical framework of the SM.

The role of experimental particle physics is to test the SM in all conceivable ways, seeking to discover whether something more lies beyond it. In this sense, the top quark provides probably the best of the chances to find evidence of new physics beyond the SM by focusing only on properties of particles within the SM (as opposed to a direct search of new particles). The reason for that will become clear in Sect. 1.3. But before going that far, a brief overview of the SM is given in Sect. 1.1. The aim of this chapter is to give the reader an overview of the theoretical background necessary to understand the basics of the top quark and its interactions (especially the weak interaction) in the context of a hadron collider, and motivate the importance of measuring its properties and in particular its production by means of the weak force, which is the topic of this thesis.

1.1 The Standard Model

The theoretical framework for the SM is provided by a relativistic Quantum Field Theory (QFT). In a relativistic QFT, the basic entities are quantum mechanical fields that permeate spacetime and whose dynamics satisfy the laws of special relativity. The

quanta of the fields are interpreted as the particles we observe. Quantum fields are classified as bosonic or fermionic fields, according to the value of their spin. Bosonic fields have integer spin, they satisfy Bose-Einstein statistics, and the corresponding particles are generically called bosons. Fermionic fields have half-integer spin, they satisfy Fermi-Dirac statistics, and the corresponding particles are generically called fermions. Thus, fermions satisfy Pauli's exclusion principle, reason for which they form matter. In the SM, fermionic fields of spin 1/2 (called "spinors" and generically denoted by $\psi(x)$) are used to describe elementary particles of matter, while bosonic fields of spin 1 are responsible for the strong, weak and electromagnetic interactions.

1.1.1 Spinors

Spinors are fields with 4 complex components:

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}, \tag{1.1}$$

whose free-dynamics is described in Dirac's theory by the free-particle action:

$$\mathcal{S}_{\text{free}} = \int \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi \; d^4x, \tag{1.2}$$

where γ^μ ($\mu = 0, 1, 2, 3$) are 4×4 Dirac matrices, m is the mass of the respective particle and $\overline{\psi} = \psi^\dagger \gamma^0$.

Being spin-1/2 fields, their eigenvalues with respect to the projection of the spin operator $\hat{S}$ into any direction of space, are $-\hbar/2$ and $+\hbar/2$. However, this is not generally a conserved quantity unless the projection is in the direction of the particles momentum $\vec{p}$. This defines the *helicity* operator

$$\hat{h} = \frac{2}{\hbar}\hat{S} \cdot \frac{\vec{p}}{|\vec{p}|}, \tag{1.3}$$

under which a spinor may have eigenvalues -1 or $+1$, and in which case it is said to be *left-handed* or *right-handed* respectively. The helicity plays a crucial role in the electroweak interactions as it will be explained in Sect. 1.1.3. A spinor ψ can be divided into its left- and right-handed components as defined in the next equation:

$$\begin{aligned} \psi &= \frac{(1-\gamma^5)}{2}\psi + \frac{(1+\gamma^5)}{2}\psi \\ &= \psi_L + \psi_R, \end{aligned} \tag{1.4}$$

where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. The operators $\mathbb{P}_L = (1-\gamma^5)/2$ and $\mathbb{P}_R = (1+\gamma^5)/2$ form a complete set of projection operators, as they satisfy the relations

$$\mathbb{P}_{L/R}^2 = \mathbb{P}_{L/R}, \qquad \mathbb{P}_L \cdot \mathbb{P}_R = \mathbb{P}_R \cdot \mathbb{P}_L = 0, \qquad \mathbb{P}_L + \mathbb{P}_R = 1. \tag{1.5}$$

In terms of the left- and right-handed components of the spinors, and making use of the set of Eqs. (1.5), the free-particle action from Eq. (1.2) can be rewritten as:

$$\mathcal{S}_{\text{free}} = \int (\overline{\psi_L} i\gamma^\mu \partial_\mu \psi_L + \overline{\psi_R} i\gamma^\mu \partial_\mu \psi_R) - m(\overline{\psi_L}\psi_R + \overline{\psi_R}\psi_L)\ d^4x. \tag{1.6}$$

In the SM all particles are massless, meaning that the terms $m\overline{\psi}\psi$ in the lagrangian are absent. How particles acquire mass is an unsolved problem in particle physics. Sect. 1.1.4 describes the way this is proposed to happen within the SM. This is the only part of the model that has no experimental verification so far.

1.1.2 Matter Content

In the SM, fermions are grouped in three generations (families), with each generation containing a pair of quarks and a pair of leptons. Quarks participate in the strong interaction; leptons do not. Moreover, the weak interaction differentiates between left-handed and right-handed components. Table 1.1 shows the left-handed matter content of the SM with their mass and their quantum numbers relevant to the electroweak interaction: the third component of the weak isospin (I_3), the hypercharge (Y) and the electric charge (Q)[1]. These three charges are related to each other by the following formula:

$$Q = I_3 + \frac{Y}{2}. \tag{1.7}$$

Flavor	Symbol	Weak isospin	Weak hypercharge	Electric charge	Mass
up quark	u	$+1/2$	$+1/3$	$+2/3$	$1.5 - 3.3$ MeV$/c^2$
down quark	d	$-1/2$	$+1/3$	$-1/3$	$3.5 - 6.0$ MeV$/c^2$
electron	e	$-1/2$	-1	-1	0.511 MeV$/c^2$
electron neutrino	ν_e	$+1/2$	-1	0	< 3 eV$/c^2$
charm quark	c	$+1/2$	$+1/3$	$+2/3$	1.27 GeV$/c^2$
strange quark	s	$-1/2$	$+1/3$	$-1/3$	$70 - 130$ MeV$/c^2$
muon	μ	$-1/2$	-1	-1	105.7 MeV$/c^2$
muon neutrino	ν_μ	$+1/2$	-1	0	< 0.19 MeV$/c^2$
top quark	t	$+1/2$	$+1/3$	$+2/3$	171.3 GeV$/c^2$
bottom quark	b	$-1/2$	$+1/3$	$-1/3$	4.20 GeV$/c^2$
tau	τ	$-1/2$	-1	-1	1.777 GeV$/c^2$
tau neutrino	ν_τ	$+1/2$	-1	0	< 18 MeV$/c^2$

Table 1.1 : List of the three generations of left-handed fermions in the Standard Model, showing their $SU(2)_L \times U(1)_Y$ and $U(1)_{em}$ charges and their masses [1].

[1]The electric charge is given in multiples of the absolute value of the electric charge of the electron.

An equivalent table exists for the right-handed components, where the weak isospin is zero and therefore the hypercharge doubles the electric charge. Moreover, since there is no evidence of the existence of right-handed neutrinos, they are not part of the matter content of the SM.

Each particle has its corresponding antiparticle, with opposite quantum numbers and the same mass. The antiparticle of the left-handed top quark is the right-handed antitop quark. In general, an antiparticle is denoted by adding a bar over the symbol that denotes the particle. So, for example, the antitop quark (or top antiquark) is denoted by $\bar{t}$. In the case of the electron, muon and tau leptons, it is also common to use the same symbol for both particle and antiparticle, and add the electric charge next to the right to differentiate between them. For example, the electron is denoted by e^- and the positron (antielectron) by e^+.

1.1.3 Gauge Symmetries and Interactions

The SM accomodates the fundamental fiels of the model such that certain symmetries are satisfied, specifying the lagrangian ($\mathscr{L}$), or the action ($\mathcal{S}$), from where all the field dynamics can be derived by means of the principle of least-action. The (relativistic) lagrangian must satisfy an obvious symmetry; it must be invariant under transformations of the Poincare group (translations$\times SO(3,1)$). But there are also other special kind of symmetries present in the SM called *gauge invariance*. The understanding of the quantum aspects of gauge invariance led to the development of relativistic Quantum Gauge Field Theory. Gauge invariance is a powerful symmetry that tames the uncontrollable infinities that appear in quantum amplitudes, making the theory renormalizable[2]. Gauge invariance also encodes the rich symmetry structure of conserved charges observed in elementary particle physics. The three interactions included in the SM are described as theories of quantum gauge symmetry with unitary Lie groups of different dimensions. Forces are interpreted in perturbative QFT as the product of the exchange of intermediate particles, which are said to be virtual (or to be off-shell), because they do not precisely obey the $p^\mu p_\mu = m^2c^4$ relationship for their short time of existence. Virtual particles with negative $p^\mu p_\mu$ are said to be "space-like", while particles with positive $p^\mu p_\mu$ are said to be "time-like".

The Example of Electromagnetism

Maxwell's equations are invariant under *local* gauge transformations of the photon field A_μ of the form:

$$A_\mu(x) \to A_\mu(x) - \partial_\mu \alpha(x), \tag{1.8}$$

[2]A theory is called renormalizable when expressions leading infinite values can be absorbed in a finite number of other terms of the theory by redefining its parameters. Renormalizability is a condition for the theory to be self-consistent.

with $\alpha(x)$ an arbitrary function. Indeed, these transformations leave the field strength

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{1.9}$$

action

$$\mathcal{S} = -\frac{1}{4}\int F^{\mu\nu}F_{\mu\nu}\ d^4x \tag{1.10}$$

and all physical observables unchanged. These transformations form a *gauge group* known as the unitary Lie group $U(1)$. To make clear that this symmetry group belongs to the electromagnetic interaction, the group is usually written as $U(1)_{em}$.

Spinors are also affected by the transformation. In electromagnetism the transformation rule for spinors is:

$$\psi(x) \to \psi'(x) = e^{i\alpha(x)\mathrm{Q}}\psi(x), \tag{1.11}$$

where Q is said to be the generator of the transformations of the group. From the physical point of view, Q is the associated conserved charge (in the case of electromagnetism, Q is the electric charge operator).

The action in Eq. (1.2) is not gauge invariant under the transformation specified in Eq. (1.11) for the spinor $\psi(x)$, unless the transformation is *global* ($\alpha(x) = \alpha =$ constant). In order to achieve *local* gauge invariance, the partial derivate ∂_μ is replaced by the covariant derivative

$$\mathrm{D}_\mu = \partial_\mu + iA_\mu\mathrm{Q} \tag{1.12}$$

with Eq. (1.8) giving the tranformation rule for the field $A_\mu(x)$. The full electromagnetic action is therefore:

$$\mathcal{S}^{EM} = -\frac{1}{4}\int F^{\mu\nu}F_{\mu\nu}\ d^4x\ + \int i\overline{\psi}\gamma^\mu\partial_\mu\psi\ d^4x\ - \int \overline{\psi}\gamma^\mu A_\mu\mathrm{Q}\psi\ d^4x. \tag{1.13}$$

The last term in Eq. (1.13) specifies the interaction between the electromagnetic field and the electrically-charged particle represented by ψ. This is the way the interaction fields are introduced in the SM, namely as compensating fields that make the theory gauge invariant under local gauge transformations.

Gauge Symmetries of the Standard Model

The gauge symmetry groups of the SM are $SU(3)_C \times SU(2)_L \times U(1)_Y$, with the subindices indicating either the quantum number associated to the symmetry group, or in the case of $SU(2)_L$ that when the transformation acts on spinors it does it on the left-handed components. The $SU(3)_C$ symmetry describes the strong interaction, while the $SU(2)_L \times U(1)_Y$ part describes the electroweak interaction. The electroweak interaction is responsible for the production (and decay) of single top quarks[3], so it

[3]When the phrase "single top" is mentioned throughout this thesis, and although in some cases it may refer to processes in which a top quark is produced, the charge-conjugate (CP) process that produces an antitop quark, is implicitly considered as well.

will be discussed in more detail later in this section. Before going into that, general rules for $SU(N)$ groups are given.

As mentioned, the relevant symmetry groups in the SM are the unitary group $U(1)$ and the special unitary groups of degree N, denoted $SU(N)$, of unitary matrices with unit determinant[4]. The corresponding transformations are represented by unitary matrices:

$$\mathrm{U}(x) = e^{i\sum_j \alpha_j(x)\mathrm{T}_j} = 1 + i\sum_j \alpha_j(x)\mathrm{T}_j + \mathcal{O}(\alpha^2), \tag{1.14}$$

with arbitrary functions $\alpha_j(x)$ and with T_j being the infinitesimal generators of the transformations of the group. $U(1)$ has one generator while $SU(N)$ groups have N^2-1 generators, which satisfy the anticommutation rules $[\mathrm{T}_i, \mathrm{T}_j]_- = i\sum_{k=1}^{N^2-1} f_{ijk}\mathrm{T}_k$, with f_{ijk} a completely antisymmetric tensor whose components are called the structure constants of the group. The T_j are traceless hermitian matrices. For a given gauge symmetry group, there are as many gauge bosons as generators in the group. The dimension of the generator matrices T_j (and therefore of the matrix $\mathrm{U}(x)$) depends on the representation of the symmetry group. Spinors belong to the (irreducible) fundamental representation, where the generators T_j of $SU(N)$ are represented by matrices of dimension $N \times N$ and spinors form an N-component field

$$\psi(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \\ . \\ . \\ . \\ \psi_N(x) \end{pmatrix} \tag{1.15}$$

with each $\psi_i(x)$ being a 4-component spinor as shown in Eq. (1.1). The transformation rule for this multicomponent field is in general

$$\psi(x) \to \psi'(x) = \mathrm{U}(x)\psi(x) = e^{i\sum_j \alpha_j(x)\mathrm{T}_j}\psi(x), \tag{1.16}$$

being Eq. (1.11) the special case of dimension one.

The generalization of the covariant derivative to higher dimensions is given by

$$\mathrm{D}_\mu = \partial_\mu + igW_\mu \tag{1.17}$$

where g is the coupling constant of the interaction associated with the corresponding gauge symmetry group and W_μ is a combination of the N^2-1 compensating fields W_μ^j:

$$W_\mu(x) = \sum_{j=1}^{N^2-1} W_\mu^j(x)\mathrm{T}_j. \tag{1.18}$$

[4]Unitary groups $U(N)$ can be obtained as the product $U(1) \times SU(N)$.

Thus, the coupling is the same for all the compensating fields.

The field strength is generalized to

$$F_{\mu\nu} = \sum_{j=1}^{N^2-1} F^j_{\mu\nu} \mathrm{T}_j = \sum_{j=1}^{N^2-1} (\mathrm{D}_\mu W^j_\nu - \mathrm{D}_\nu W^j_\mu) \mathrm{T}_j, \tag{1.19}$$

and the field strength action is

$$\mathcal{S} = -\frac{1}{4} \int \sum_{j=1}^{N^2-1} F^j_{\mu\nu} F^{\mu\nu\ j}\ d^4x. \tag{1.20}$$

The transformation of the field $W_\mu(x)$ that leaves the action unchanged is

$$W_\mu(x) \to W'_\mu(x) = \mathrm{U}(x) W_\mu(x) \mathrm{U}^\dagger(x) - \frac{i}{g} \mathrm{U}(x)(\partial_\mu \mathrm{U}^\dagger(x)), \tag{1.21}$$

from where Eq. (1.8) is the special case of $U(1)$ and expanded to leading order in $\alpha(x)$. The field W_μ transforms in the adjoint representation of the group and is represented by $(N^2-1) \times (N^2-1)$ matrices.

The Strong Interaction

The gauge group of the theory of the strong interactions, known as Quantum Chromodynamics (QCD), is $SU(3)$ and the corresponding charge is called "color". There are eight generators in the group, thus eight massless gauge bosons, named gluons and denoted by G^j_μ. The color-charged fermions are the quarks (the constituents of the nucleons), from which there are six varieties: up (u), down (d), charm (c), strange (s), top (t) and bottom (b). Quarks come thus in three colors, defined as "red" (R), "green" (G) and "blue" (B),

$$Q(x) = \begin{pmatrix} Q_{red}(x) \\ Q_{green}(x) \\ Q_{blue}(x) \end{pmatrix} \tag{1.22}$$

whose components mix under an $SU(3)_C$ gauge transformation (Eq. (1.16)). Quarks have never been isolated as single free particles[5]. Rather, they form colorless[6] bound states called hadrons. Hadrons can be divided into baryons and mesons. Baryons are bound states of three quarks, each one of a different color (one R, one G, and one B). Examples of baryons are protons (p) and neutrons (n). Mesons are bound states of a quark of a given color and an antiquark with the respective anticolor. Examples of mesons are pions (π^+, π^-, π^0) and kaons (K^+, K^-, K^0). This phenomenon is called

[5]Only at the very first moments of the universe, when the temperature was very high, quarks and gluons existed as free particles.

[6]Colorless means that the corresponding quantum state is neutral under $SU(3)_C$.

color confinement and it is viewed as a consequence of the fact that the strength of the strong force increases as the separation between quarks increases[7]. At some point, the energy stored in the field is so high, that it is energetically more favorable for a quark-antiquark pair to be created out of the vacuum, than for the distance between quarks to increase further. Thus, if quarks inside a hadron separate, new hadrons are formed. This process is called *hadronization* (or fragmentation). Because of hadronization of quarks, single quarks produced at accelerators can not be detected as such, but rather as collimated bundles of (mainly hadronic) particles called *jets*. The jet direction points approximately into the direction of the original quark. Radiated gluons also hadronize producing jets.

Contrary to what happens for an abelian group of transformations like $U(1)$, where the intermediate field is neutral under the charge of the multicomponent fields (e.g., the photon is electrically neutral), in a non-abelian group of transformations like $SU(N)$, the intermediate fields do have charge. So, for example, gluons carry a color and an anticolor (e.g., red and antigreen), giving to the strong interaction its peculiar properties.

The Electroweak Interaction

The weak and electromagnetic forces are different manifestations of a unique force named the electroweak interaction. The unification of these two forces into one interaction is encoded in the $SU(2)_L \times U(1)_Y$ gauge symmetry group. Transformations of $SU(2)_L$ act on the left-handed components of the spinors, while right-handed components are untached[8]. Left-handed components transform under the fundamental representation of $SU(2)_L$ forming doublets, while right-handed components are singlets. The associated quantum number is called "weak isospin". Left-handed fermions have weak isospin $1/2$ with projections $+1/2$ for the "up" component of the doublet and $-1/2$ for the "down" component respectively, as the following general notation intends to show,

$$F_L = \begin{pmatrix} f_L^{+1/2} \\ f_L^{-1/2} \end{pmatrix} = \begin{pmatrix} f_L^u \\ f_L^d \end{pmatrix}, \tag{1.23}$$

while right-handed spinors have weak isopin 0.

Leptons of the same generation (see Table 1.1) are $SU(2)_L$ partners. The same happens with the quarks, except for a very important difference: the quark eigenstates of the weak interaction are not the same as the quark eigenstates of the strong interaction. The matter content of the SM presented as $SU(2)_L$ doublets and singlets in a base where the up-type quark eigenstates are the same in the strong and the weak interactions is:

[7] It has not been mathematically demonstrated that QCD is a confining theory.

[8] The SM is a quiral theory, meaning that the left- and right-handed components of the spinors are treated differently.

$$\begin{array}{lllllllll} \text{Quarks}: & \begin{pmatrix} u_L \\ d'_L \end{pmatrix}, & u_R, & d'_R; & \begin{pmatrix} c_L \\ s'_L \end{pmatrix}, & c_R, & s'_R; & \begin{pmatrix} t_L \\ b'_L \end{pmatrix}, & t_R, \quad b'_R; \\ \text{Leptons}: & \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, & e_R; & & \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, & \mu_R; & & \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}, & \tau_R. \end{array}$$

For the down-type quarks, the two kind of eigenstates are related to each other by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix [2]:

$$\begin{pmatrix} |d'\rangle \\ |s'\rangle \\ |b'\rangle \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} |d\rangle \\ |s\rangle \\ |b\rangle \end{pmatrix}. \tag{1.24}$$

The elements $|V_{ij}|^2$ of the CKM matrix represent the probability for a quark of flavor i to decay into a quark of flavor j. In this thesis, the CKM matrix element V_{tb} plays an important role, as the production cross section for single top quark is proportional to $|V_{tb}|^2$.

The generators of the $SU(2)_L$ group in the fundamental representation are the 2×2 Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.25}$$

Using coupling constants g for $SU(2)_L$ and g' for $U(1)_Y$, and denoting B^0_μ the gauge field associated to $U(1)_Y$ and W^j_μ the ones associated to $SU(2)_L$, the covariant derivatives for left- and right-handed fermion components are respectively

$$\mathrm{D}_\mu F_L = (\partial_\mu + i\frac{g'}{2}B^0_\mu(x)\mathrm{Y} + i\frac{g}{2}\sum_{j=1}^{3} W^j_\mu(x)\sigma_j)F_L, \tag{1.26}$$

$$\mathrm{D}_\mu f_R = (\partial_\mu + i\frac{g'}{2}B^0_\mu(x)\mathrm{Y})f_R, \tag{1.27}$$

and introducing them into the free-particle action (Eq. (1.6)), the interaction terms that arise are

$$\mathscr{L}^{EW}_{\mathrm{int}} = -g\overline{F_L}W^j_\mu\gamma^\mu\frac{\sigma_j}{2}F_L - \frac{g'}{2}B^0_\mu(\overline{F_L}\gamma^\mu\mathrm{Y}F_L + \overline{f_R}\gamma^\mu\mathrm{Y}f_R). \tag{1.28}$$

The $SU(2)_L \times U(1)_Y$ symmetry is however not explicitly manifest in nature, since it would require for example all particles to be massless (see Sect. 1.1.4). The symmetry is said to be "broken". The process that breaks the $SU(2)_L \times U(1)_Y$ symmetry of the SM is generically called Electroweak Symmetry Breaking (EWSB) and how this happens is an unsolved problem in particle physics. Within the SM, the process that breaks the electroweak symmetry is the Higgs Mechanism, which will be briefly discussed in Sect. 1.1.4. From EWSB, the $U(1)_{em}$ symmetry survives and therefore the

electric charge is conserved in all the interactions. Thus, the terms in the electroweak interaction lagrangian must be rewritten in terms of the charged and neutral currents of the weak interaction and electromagnetism. This is accomplished by the set of transformations

$$W_\mu^+ = (W_\mu^1 + iW_\mu^2)/\sqrt{2}, \tag{1.29}$$
$$W_\mu^- = (W_\mu^1 - iW_\mu^2)/\sqrt{2}, \tag{1.30}$$
$$A_\mu = \cos(\theta_\mathrm{w})B_\mu^0 + \sin(\theta_\mathrm{w})W_\mu^3, \tag{1.31}$$
$$Z_\mu^0 = -\sin(\theta_\mathrm{w})B_\mu^0 + \cos(\theta_\mathrm{w})W_\mu^3. \tag{1.32}$$

with $\tan(\theta_\mathrm{w}) = \mathrm{g}/\mathrm{g}'$. The electroweak interaction lagrangian splits in $\mathscr{L}_\mathrm{int}^{EW} = \mathscr{L}_\mathrm{int}^{W^+,W^-} + \mathscr{L}_\mathrm{int}^{Z^0} + \mathscr{L}_\mathrm{int}^{EM}$ with the charged-current weak lagragian given by

$$\mathscr{L}_\mathrm{int}^{W^+,W^-} = -\frac{g}{\sqrt{2}}[\overline{f_L^u}\gamma^\mu f_L^d W_\mu^- + \overline{f_L^d}\gamma^\mu f_L^u W_\mu^+]. \tag{1.33}$$

This is the part we are interested in, since it describes single top quark production (and decay).

1.1.4 The Higgs Mechanism

Mass terms in the lagrangian of the form

$$m\overline{\psi}\psi = m(\overline{\psi_L}\psi_R + \overline{\psi_R}\psi_L) \tag{1.34}$$

violate the $SU(2)_L$ symmetry. Thus, to preserve gauge invariance under $SU(2)_L$, the matter fields in the SM must be massless. Similarly, gauge invariance also requires all gauge bosons to be massless. However, except for the gluons and the photon, all other particles are known to have mass. In the SM, particles acquire their mass through the interaction with a complex massive scalar field, called the Higgs field, which is a doublet under $SU(2)_L$, and has hypercharge 1,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{1.35}$$

The lagrangian of the scalar Higgs field is

$$\mathscr{L}^\phi = (\mathrm{D}^\mu\phi)^\dagger \mathrm{D}_\mu\phi - V(\phi), \tag{1.36}$$

with the covariant derivative given by

$$\mathrm{D}_\mu\phi = (\partial_\mu + i\frac{g'}{2}B_\mu^0(x) + i\frac{g}{2}\sum_{j=1}^{3} W_\mu^j(x)\sigma_j)\phi, \tag{1.37}$$

and with the most general renormalizable self-interaction potential that preserves the $SU(2)_L$ symmetry,

$$V(\phi) = -\frac{m_H}{2}\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2. \tag{1.38}$$

The lowest energy states are described by

$$\phi = e^{i\theta\vec{n}\cdot\frac{\vec{\sigma}}{2}} \begin{pmatrix} 0 \\ \sqrt{\frac{m_H}{4\lambda}} \end{pmatrix}. \tag{1.39}$$

At low energies, the field acquires a vacuum expectation value choosing one of all directions in $SU(2)_L$ (described here as $\theta = 0$), *spontaneously* breaking the electroweak symmetry to a smaller symmetry, the $U(1)$ group of electromagnetism. Performing small perturbations around the ground state, mass terms for the W_μ^+, W_μ^- and Z_μ^0 bosons appear in the lagrangian, but not for the gauge boson of the surviving $U(1)$ symmetry (the photon) since it does not couple to the component of the ϕ field chosen as getting the vacuum expectation value (the ϕ^0 field). Three of the four degrees of freedom of the Higgs field give rise to the mass of the weak gauge bosons. The fourth degree of freedom is a new (neutral and massive) particle predicted by the model, the Higgs particle. The fermions also acquire mass, but thanks to the Yukawa coupling to the Higgs field, which are new introduced terms of the form

$$\mathscr{L}_{\text{int}}^{\text{Yukawa}} = -\kappa_{\psi\phi}\overline{\psi}\phi\psi. \tag{1.40}$$

The Higgs particle has not been found yet. The Higgs mechanism is the only part of the SM that is still waiting for verification. However, from the experimental point of view, still another milestone was missing, the observation of single top quark production, as single top is a background to Higgs processes in which the Higgs particle is produced together with a W boson from a virtual W (this is the cleanest final state used to search for the Higgs at the Tevatron).

As a final consideration aboout the SM, it has to be said that as successful as it is, it is far from being the final story in particle physics. Rather, it is considered to be an effective theory valid up to the energies explored so far, which are below the limit set by the energy scale of EWSB. The SM not only does not incorporate gravity. It cannot explain the abundance of dark matter in the universe neither dark energy. Also, it does not explain neutrino masses (although some simple extensions of the model do). Moreover, there are 19 parameters whose values are not predicted by the theory and have to be measured experimentally.

1.2 Cross Sections

This section gives the basics about cross section calculations in a hadron-hadron collision[9]. The results will then be applied to present partial results on the electroweak

[9]The formulas given in this section are based on Refs. [3] and [4].

production cross section of the top quark.

The probability amplitude for the transition of an initial state $|\Psi_{\text{in}}\rangle$ of N_{in} incoming free particles at time $t_{\text{in}} = -\infty$ to a final state $|\Psi_{\text{out}}\rangle$ of N_{out} outgoing free particles at time $t_{\text{out}} = +\infty$ is given by [3]

$$\mathcal{A}_{\text{in,out}} = -2\pi i \delta^4(p_{\text{in}} - p_{\text{out}})\mathcal{M}_{\text{in,out}}, \tag{1.41}$$

where p_{in} and p_{out} are the total four-momenta of the initial and final state group of particles respectively and $\mathcal{M}_{\text{in,out}}$ is the matrix element $\langle\Psi_{\text{out}}||\Psi_{\text{in}}\rangle$. The transition probability is proportional to the amplitude squared,

$$P_{\text{in}\to\text{out}} = \frac{1}{V_{\text{in}}V_{\text{out}}}|\mathcal{A}_{\text{in,out}}|^2, \tag{1.42}$$

with proportionality constants $1/V_{\text{in/out}}$ defined by the proper normalization of the initial and final states to a finite volume in the multi-particle momentum phase space. Defining the final state interval by $d^3p_{\text{out}} = \prod_{i=1}^{N_{\text{out}}} d^3p_i$ so that the total number of states in that interval is $V_{\text{out}}d^3p_{\text{out}}$, the total probability for the system to wind up in a range d^3p_{out} of final states is

$$dP_{\text{in}\to\text{out}} = \frac{1}{V_{\text{in}}}|\mathcal{A}_{\text{in,out}}|^2 d^3p_{\text{out}}. \tag{1.43}$$

The transition rate is given by $d\Gamma_{\text{in}\to\text{out}} \equiv dP_{\text{in}\to\text{out}}/T$, where T is the time interval during which the interaction is acting. If one is not dealing with a decay process ($N_{\text{in}} = 1$), where the mean lifetime of the particle imposes a maximum time scale for the interaction, one can take the limit $T \to \infty$. So, we finally get to the concept of *cross section* for a process with $N_{\text{in}} > 1$, which is defined as the transition rate per unit of (incoming) flux Φ_{in}. After properly defining the square of the delta function in Eq. (1.41), summing over the spins of the initial state particles, and specializing to $N_{\text{in}} = 2$, the differential cross section is [3]

$$d\hat{\sigma}_{\text{in}\to\text{out}} \equiv d\Gamma_{\text{in}\to\text{out}}/\Phi_{\text{in}} = \sum_{\text{spins}} \frac{(2\pi)^4 E_1 E_2 |\mathcal{M}_{\text{in,out}}|^2}{\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}} \delta^4(p_1 + p_2 - p_{\text{out}}) d^3p_{\text{out}}. \tag{1.44}$$

Eq. (1.44) gives the general expression of the differential cross section for a $2 \to N_{\text{out}}$ process, where the incoming particles are considered as point particles. In a hadron collider, the particles intervening in the collision are the partons from the incoming hadrons; in the case of the Tevatron, a parton from a proton and a parton from an antiproton. The proton is made up from two u quarks and one d quark, and the antiproton from the corresponding antiquarks. These are called the valence quarks from the proton and the antiproton. However, to view a hadron as only composed of their valence quarks is misleading, since hadronic interactions are occurring between the quarks at all times, with a constant creation and aniquilation of gluons and other

quarks. Thus, a hadron should be better considered as formed from its valence quarks immersed on a "sea" of additional partons (gluons and quarks). So, the hard interaction in a hadron-hadron collision may occur between valence quarks of both hadrons, between a valence quark from one of the hadrons and a sea parton from the other hadron, or between sea partons of both hadrons. Having said that, the total hadron-level cross section for a $2 \to N_{\text{out}}$ partonic process can be written as

$$\sigma = \sum_{i,j} \iint dx_i dx_j f_i^{h_1}(x_i, \mu_{\text{F}}^2) f_j^{h_2}(x_j, \mu_{\text{F}}^2) \times \hat{\sigma}_{i,j\to\text{out}}(x_i, x_j, \mu_{\text{F}}^2), \tag{1.45}$$

where $\hat{\sigma}_{i,j\to\text{out}}$ is the parton-level cross section from Eq. (1.44). The $f_{i/j}^{h_1/h_2}(x_{i/j}, \mu_{\text{F}}^2)$ are the parton distribution functions (PDFs), which provide the probability density of finding a parton of flavor i/j inside a hadron h_1/h_2 with a fraction $x_{i/j}$ of the hadron longitudinal momentum. The PDFs are evaluated at a given factorization scale μ_{F}, which should represent a typical energy scale in the process under consideration. Two well-known parametrizations of the PDFs are the ones provided by the CTEQ collaboration [5] and by Martin-Roberts-Stirling-Thorne (MRST) [6]. Fig. 1.1 shows the product $xf(x)$ as a function of x for the MRST2004 NNLO PDF set [7][10]. The indices i, j in Eq. (1.45) run over all incoming parton flavors. The total cross section is the sum over all possible parton flavor combinations, integrated over all the allowed combinations of parton momentum fractions $x_{i/j}$.

It becomes useful for the calculation of the cross sections to change variables from the p_i's to partonic kinematical invariants. The standard kinematical invariants are the Mandelstam variables $s = (p_1 + p_2)^2$ (the partonic center-of-mass energy squared), $t = (p_1 - p_3)^2$, and $u = (p_2 - p_3)^2$. The corresponding hadronic kinematical invariants are $s_h = (p_{h1} + p_{h2})^2$ (the hadronic center-of-mass energy squared)[11], $t_h = (p_{h1} - p_3)^2$, and $u_h = (p_{h2} - p_3)^2$. If the incoming partons are light quarks or gluons, then one can set $m_1 = 0$ and $m_2 = 0$, and thus $p_1 = x_1 p_{h1}$ and $p_2 = x_2 p_{h2}$. The partonic invariants can be rewriten as $s = (x_1 p_{h1} + x_2 p_{h2})^2$, $t = (x_1 p_{h1} - p_3)^2$, and $u = (x_2 p_{h2} - p_3)^2$. The relations between partonic and hadronic quantities are $t - m_3^2 = x_1(t_h - m_3^2)$ and $u - m_3^2 = x_2(u_h - m_3^2)$, which imply $dt/dt_h = x_1$ and $du/du_h = x_2$. Thus, taking the derivatives with respect to t_h and u_h on the cross section of Eq. (1.45) gives

$$\frac{d^2\sigma}{dt_h du_h} = \sum_{i,j} \iint dx_i dx_j \; [x_i f_i^{h_1}(x_i, \mu_{\text{F}}^2)][x_j f_j^{h_2}(x_j, \mu_{\text{F}}^2)] \times \frac{d^2\hat{\sigma}_{\text{in}\to\text{out}}}{dtdu}(x_i, x_j, \mu_{\text{F}}^2). \tag{1.46}$$

Integrating then in t_h and u_h, one finally gets a general expression of the cross section

[10] This set of PDFs was used in the calculation of the single top quark production cross sections in Ref. [4].

[11] Along this chapter, the hadronic center-of-mass energy is denoted by $\sqrt{s_h}$ to distinguish it from the partonic center-of-mass energy $\sqrt{s}$. In all other chapters of this thesis, the partonic variables are not relevant, and $\sqrt{s}$ is used to denote the hadronic center-of-mass energy, as it is commonly done in experimental high-energy physics.

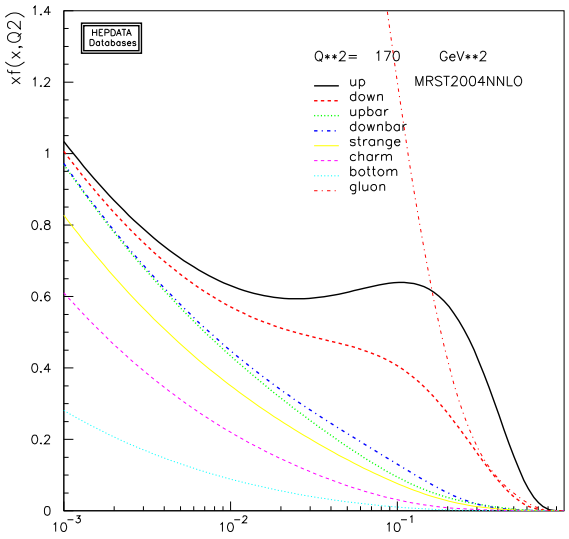

Figure 1.1 : Parton distribution functions according to the MRST2004 NNLO parametrization [7] for a factorization scale $Q^2 \equiv \mu_{\rm F}^2 = (170\ {\rm GeV})^2$. Because of the non-perturbative effects of the strong interactions inside a hadron, the PDFs cannot be calculated from theory. They are instead estimated from global fits of QCD calculations to particle collision data.

for a hadron-hadron collision,

$$\sigma = \sum_{i,j} \int_{t_h^{min}}^{t_h^{max}} \int_{u_h^{min}}^{u_h^{max}} \iint dx_i dx_j \ [x_i f_i^{h_1}(x_i, \mu_{\rm F}^2)][x_j f_j^{h_2}(x_j, \mu_{\rm F}^2)] \times \frac{d^2\hat{\sigma}_{\rm in\to out}}{dtdu}(x_i, x_j, \mu_{\rm F}^2)\ dt_h du_h. \qquad (1.47)$$

In perturbative QFT, the transition amplitude from $|\Psi_{\rm in}\rangle$ to $|\Psi_{\rm out}\rangle$ has an infinite number of contributions, each one pictorically represented by a *Feynman diagram*. Feynman diagrams show paths of particles in spacetime as lines, and interactions between particles as vertices where the lines meet. External lines correspond to the incoming and outgoing particles, while intermediate lines represent exchanged particles. The perturbation expansion and the associated Feynman diagrams are useful to the extent that the strength of the interaction is small, so that the lowest-order terms, or diagrams with the fewest vertices, give the main contribution to the matrix element. Feynman diagrams are thus classified by the order of the power at which the coupling constants appear, or the number of internal loops they contain. Loops are generated by the splitting of an internal line into two lines, which then are merged again into one internal line. Thus, each loop adds two coupling constants to the term. Diagrams with no loops are said to represent the calculation at tree-level or leading-order (LO). If diagrams with one loop are added, the calculation is said to be at next-to-leading-order (NLO), and so on. Once the order of the calculation has been fixed and the Feynman diagrams selected, one writes the contribution of each diagram by using the

Feynman rules, which assign a mathematical expression to each element of the diagram (vertices, incoming/outgoing lines, internal lines, etc). The expressions associated to vertices are deduced from the interaction lagrangian. For example, the charged-current electroweak interaction lagrangian from Eq. (1.33) sets the expression associated to the vertices in the single top production Feynman diagrams. In general, integration on the particle's momentum propagating along a loop diverges. These divergent terms can be recombined together with other terms of similar structure in the series expansion, and by redefining the parameters of the theory (the normalization of the wave function, the mass of the particles, and the coupling constants), cancel the infinities in a process called renormalization[12].

1.3 The Top Quark

The top quark is the $Q = 2/3, I_3 = 1/2$ member of the third generation of quarks. With a mass of $m_t \approx 170$ GeV/c^2, it is by far the heaviest of all known particles. Because its mass is close to the EWSB scale ($v = 246$ GeV), it is believed that the top quark could play a special role in the process of EWSB. Thus, one of the primary motivations for the Tevatron Run II is to accurately determine the top quark properties for further insights into electroweak theory which could provide information about the mechanism of mass generation, as well as to see if any hints of new physics (new quarks, extra weak gauge bosons, flavor changing neutral currents, etc.) may be visible (see Refs. [8, 9]).

In order to explain CP violations observed in kaon decay, Kobayashi and Maskawa predicted in 1973 the existence of a third generation of quarks [10], which where called top and bottom. In 1977, the bottom quark was discovered at Fermilab [11] by the E288 experiment, using di-muon events. This strongly suggested that a sixth quark, the top, should exist to complete the $SU(2)_L$ doublet. It was known that the top quark should be heavier than the bottom quark, but it was thought not to be heavier than a few tens of GeV/c^2, so that it would soon be discovered as well. However, it took 18 years until the top quark was discovered, in 1995, at Fermilab by the CDF [12] and the DØ [13] collaborations, in proton-antiproton collisions from Tevatron's Run I, at a mass of 176 ± 18 GeV/c^2.

1.3.1 Top Quark Pair Production

The top quark was discovered in top-antitop ($t\bar{t}$) events. Top pairs are produced at a hadron collider via the strong interaction, through processes such as $q\bar{q} \to t\bar{t}$ and

[12]The concept of renormalization was already introduced in Sect. 1.1.3 when it was emphasized that quantum gauge field theories are renormalizable.

$gg \to t\bar{t}$. The leading-order feynman diagrams are shown in Fig. 1.2. The rate and kinematic distributions of top quarks produced in this way are a measure of the top's interaction with the gluons.

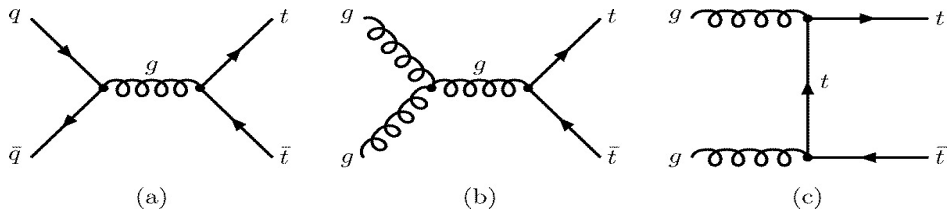

Figure 1.2 : Leading-order Feynman diagrams for the production of top pairs via the strong interaction.

Top quarks can also be produced at a hadron collider in association with a bottom quark, by means of the weak coupling to a W boson. Single top has a production rate of about one-half of that of top pair, which is ~ 7.5 pb at the Tevatron. However, $t\bar{t}$ events have much cleaner signals, with smaller backgrounds, than single top events. In general, a cut-based event selection is enough to get a $t\bar{t}$ sample of events with small backgrounds, while in the case of single top, additional techniques must be applied to separate the signal from the overwhelming background still present after a typical event selection based on cuts on kinematic variables. Since the observation of doubly produced top quarks, a tremendous effort has been put to search for singly produced top quarks. After 14 years, in 2009, both DØ [14] and CDF [15] finally announced observation of single top quarks. This thesis describes the work of one of the analyses that constitute this DØ discovery.

1.3.2 Single Top Production

Single top production occurs via the electroweak interaction as described in Eq. (1.33). One should specialize the f fields to the third generation of quarks t and b':

$$-\frac{g}{\sqrt{2}} \sum_{q=d,s,b} [V_{tq} \overline{t_L} \gamma^\mu q_L W_\mu^- + h.c.]. \tag{1.48}$$

Using that $|V_{td}|^2 + |V_{ts}|^2 \ll |V_{tb}|^2$ (a property described later in Sect. 1.3.3), the relevant interaction vertex is:

$$-\frac{g}{\sqrt{2}} [V_{tb}^* \overline{b_L} \gamma^\mu t_L W_\mu^+ + V_{tb} \overline{t_L} \gamma^\mu b_L W_\mu^-]. \tag{1.49}$$

The leading order feynman diagrams for single top quark production at a hadron collider through a Wtb vertex are shown in Fig. 1.3. The three modes are sensitive to quite different manifestations of physics beyond the SM [9].

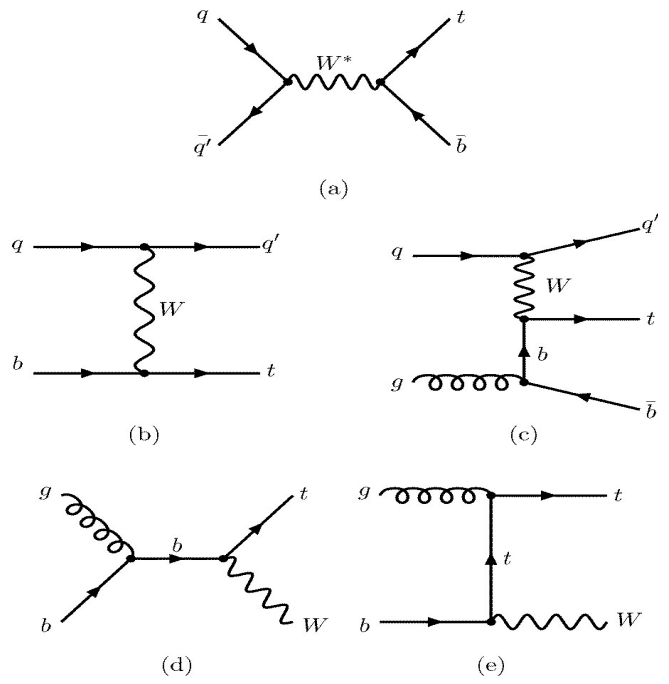

Figure 1.3 : Leading-order Feynman diagrams for the production of single top quark via the electroweak interaction: (a) s-channel, (b, c) t-channel, and (d, e) associated production with a W boson. In the Tevatron, the s-channel and t-channel dominate.

Single top production proceeds through three distinct sub-processes at a hadron collider. The s-channel mode, showed in Fig. 1.3(a), involves the production of an off-shell, time-like W boson, which then decays into a top and a bottom quark. The dominant process is $u\bar{d} \to \bar{b}t$ with a 97.4% contribution to the total s-channel cross section. Additional processes like $c\bar{s} \to \bar{b}t$ and the Cabibbo-suppressed $u\bar{s} \to \bar{b}t$, $c\bar{d} \to \bar{b}t$, and $u\bar{d} \to \bar{s}t$, have contributions of 1% or less. The t-channel W-gluon fusion mode, showed in Fig. 1.3(b), involves the exchange of a space-like W boson between a light quark and a bottom quark inside the incident hadrons, resulting in a light (not from a bottom quark) jet, a single top quark, and eventually a b-jet from the splitting gluon. In the t-channel, the dominant processes are $ub \to dt$ (65.7%) and $\bar{d}b \to \bar{u}t$ (21.4%). Additional processes involving only quarks are $cb \to st$ (2.7%) and the Cabibbo-suppresed $ub \to st$ (3.6%). Additional processes involving antiquarks and quarks are $\bar{s}b \to \bar{c}t$ (4.4%) and the Cabibbo-suppressed $\bar{d}b \to \bar{c}t$ (1.2%). Other Cabibbo-suppresed processes have negligible contributions. Finally, the tW mode of single top production involves an initial state b quark emitting a (close to) on-shell W boson, resulting in a tW final state (see Fig. 1.3(c)). Because of the massive particles in the final state, this mode has an extremely small rate at the Tevatron. Each mode has rather distinct event kinematics, and thus are potentially observable separately from each other. In this thesis, the production cross section for single top (and antitop) through the s- and t-channels together, is measured. These channels are also sometimes denoted as tb and tqb respec-

tively. The kinematics and the color flows of the two channels are quite different. The t-channel resembles deep inelastic scattering while the s-channel resembles the Drell-Yan process.

The cross sections for single top quark production at the Tevatron have identical contributions from diagrams that produce a top quark and the CP diagrams that produce an antitop quark. Considering this, the s-channel SM cross section in $p\bar{p}$ collisions at $\sqrt{s_h} = 1.96$ GeV and for a top mass $m_t = 170$ GeV/c^2, is $\sigma_{tb} = 1.12 \pm 0.04$ pb [4]. For the t-channel, the theoretical cross section in $p\bar{p}$ collisions at $\sqrt{s_h} = 1.96$ GeV and for a top mass $m_t = 170$ GeV/c^2, is $\sigma_{tqb} = 2.34 \pm 0.12$ pb [4][13]. The tW-channel cross section is only ~ 0.3 pb, reason why it is neglected.

For a $p_1 + p_2 \to p_3 + p_4$ process, the Born-level differential partonic cross section is given by [4]

$$\frac{d^2\hat{\sigma}_{p_1p_2\to p_3p_4}}{dtdu} = \frac{|\mathcal{M}_{p_1p_2,p_3p_4}|^2}{16\pi s^2}\delta(s_4), \tag{1.50}$$

where $s_4 = s+t+u-m_3^2-m_4^2$ is a fourth kinematical partonic invariant. This differential partonic cross section must be convoluted with the PDFs as shown in Eq. (1.47) to get the total hadronic cross section. The matrix element $|\mathcal{M}_{p_1p_2,p_3p_4}|^2$ for the dominant process in the s-channel, $u\bar{d} \to \bar{b}t$, is [4]

$$|\mathcal{M}_{u\bar{d}\to\bar{b}t}|^2 = k|V_{tb}|^2|V_{ud}|^2\frac{t(t-m_t^2)}{(s-m_W^2)^2}, \tag{1.51}$$

with $k = (4\pi^2\alpha^2)/\sin^4(\theta_{\rm W}) \sim 4.5 \times 10^{-2}$ where α is the fine-structure constant and $\theta_{\rm W}$ the Cabibbo angle. For the two most relevant processes in the t-channel, $ub \to dt$ and $\bar{d}b \to \bar{u}t$, the matrix elements $|\mathcal{M}_{p_1p_2,p_3p_4}|^2$ are [4]

$$|\mathcal{M}_{ub\to dt}|^2 = k|V_{tb}|^2|V_{ud}|^2\frac{s(s-m_t^2)}{(t-m_W^2)^2}, \tag{1.52}$$

$$|\mathcal{M}_{\bar{d}b\to\bar{u}t}|^2 = k|V_{tb}|^2|V_{ud}|^2\frac{(s+t)^2-(s+t)m_t^2}{(t-m_W^2)^2}. \tag{1.53}$$

Thus, the measurement of the single top production cross section also allows for the measurement of the CKM matrix element $|V_{tb}|$. In its single top observation paper, DØ presented the most stringent limit to date: $|V_{tb}| > 0.78$ at the 95% confidence level [14]. The measurement of $|V_{tb}|$ is one of the ways in which the study of the production of single top quarks at hadron colliders helps to identify physics beyond the SM related to the EWSB. The s-channel mode and the t-channel mode are sensible to different forms of new physics, thus providing complementary information about the top quark

[13]These are "matched" cross sections, i.e. exact NLO cross section plus NNLO and NNNLO corrections from soft-gluon radiation, each order at next-to-leading-logarithmic (NLL) accuracy. The factorization and renormalization scales were set to the top mass value, and the MRST2004 NNLO set of PDFs was used.

properties. Polarization observables also provide potentially useful information about the structure of the interactions of the top quark.

1.3.3 Top Quark Decay

As discussed in Sect. 1.1.3, because of the hadronization process, the signature of the production of a quark in high energy collisions is a jet. This is actually true for all but the top quark. Because of its high mass, the mean lifetime of the top quark is extremely short, of the order of 5×10^{-25} s. This is shorter than the characteristic time scale of the strong interaction $1/\Lambda_{\mathrm{QCD}} \sim 10^{-23}$ s. For this reason, the top quark has no time to hadronize before it decays, providing a unique opportunity to study a "bare" quark. The information about the quantum numbers of the top quark is cleanly transferred to the decay products instead of being lost in the hadronization process.

The only known way that a top quark can decay is through the weak interaction producing a W boson and a down-type quark (down, strange, or bottom). Same as for single top production, Eq. (1.48) provides the interaction vertices responsible for top quark decay. Model-independent measurements of the ratio

$$R = \frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2} \tag{1.54}$$

by CDF [16] and DØ [17], show that $|V_{td}|^2 + |V_{ts}|^2 \ll |V_{tb}|^2$. For instance, the DØ measurement gives $|V_{td}|^2 + |V_{ts}|^2 = (-0.03^{+0.18}_{-0.16})|V_{tb}|^2$. Under the assumption of only three generations and a unitary CKM matrix, the CKM matrix elements V_{tq} are even more constrained [18][14]:

$$|V_{td}| = (8.1 \pm 0.6) \times 10^{-3}, \quad |V_{ts}| = (38.7 \pm 2.3) \times 10^{-3}, \quad |V_{tb}| = 0.999133^{+0.000044}_{-0.000043}.$$

The result $|V_{td}|^2 + |V_{ts}|^2 \ll |V_{tb}|^2$ implies that, within the SM, the branching fraction $B(t \to Wb) \simeq 100\%$. Thus, top quarks produced at the Tevatron decay almost exclusively into a W boson and a b quark. Top quark decay thus provides interesting information about the quiral structure of the Wtb interaction. The W bosons from top quark decays carry polarization from the parent particle, hence pose themselves as a unique probe to top polarization. Models of new physics allow other non-standard decay modes [9].

1.3.4 Single Top Decay Channels

The experimental signatures of single top events can be classified based on the decay products of the Ws. W bosons decay into a lepton and the respective neutrino, or to

[14] In several extensions of the SM involving e.g. a fourth generation of quarks, or an additional heavy quark singlet which mixes with the top quark, $|V_{tb}|$ can be significantly smaller than 1 [19].

an up-type quark and a down-type quark (all of them left-handed as described by the charged-current weak interaction lagrangian, Eq. (1.33)). The W mass is $m_W \approx 80$ GeV/c^2, so that its decay into a top quark is kinematically suppressed by the larger top mass. The decay width of the W boson to a quark-antiquark pair is proportional to the corresponding modulus-squared CKM matrix element and the number of quark colors, $N_C = 3$. The unitarity of the CKM matrix implies that $|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = 1$ and $|V_{cd}|^2 + |V_{cs}|^2 + |V_{cb}|^2 = 1$, and therefore the hadronic branching fraction of the W boson is approximately $2/3$. The other $1/3$ is left to the leptonic decays with an almost even contribution from each generation, $B(e + \nu_e) = B(\mu + \nu_\mu) = B(\tau + \nu_\tau) = 1/9$. The hadronic branching ratio is dominated by the Cabibbo-favored channels, ud and cs. The experimental measures of the branching ratios are [1]: $B(e + \nu e) = (10.75 \pm 0.13)\%$, $B(\mu + \nu\mu) = (10.57 \pm 0.15)\%$, $B(\tau + \nu\tau) = (11.25 \pm 0.20)\%$, and $B(\text{hadrons}) = (67.60 \pm 0.27)\%$.

The signatures of single top events with hadronic W decays are final states containing only jets. This channel is called the *all-jets* or *all-hadronic* channel. Although one or two of the jets are b-jets, these final states are buried in the multijets background. On the other hand, in leptonic W decays, the presence of a final state lepton and missing transverse energy, reduces the multijets background to a manageable ammount. The final states considered in this analysis contain an electron or a muon (or their antiparticles) and missing transverse energy accounting for the neutrino. These are called the *lepton+jets* channels (electron, muon, and tau+jets). Tau leptons decay into electrons, muons or hadrons. The first two channels contribute to the direct leptonic decays. Hadronic tau decays lead to narrow jets, which are hard to distinguish from jets originating from quarks or gluons. Depending on the tau identification efficiency and the missing transverse energy, part of the hadronic tau decays are absorbed in the hadronic W decay channel.

The general criteria we will apply to select single top events is:

- one lepton (electron or muon) with high transverse momentum,
- missing transverse energy,
- 2−4 jets, with one or two of them being tagged as coming from a b-quark.

Events with four jets are included to account for initial or final state radiation. The t-channel (s-channel) dominates the sample with one (two) b-jets. The 2 jets, 1 b-tag channel contains events from the s-channel where one of the two jets originating from a b quark was not tagged (the b-tagging efficiency is ~ 0.6). Because the light quark in the t-channel is emitted preferentially at very forward (parallel to the incidence) directions, it is sometimes not reconstructed. We consider therefore a total of $2 \times 3 \times 2 = 12$ physics final state channels. However, because the data are further separated in two different run periods of the Tevatron (Run IIa and Run IIb) between which the DØ detector has been upgraded, the considered final state channels are actually twenty four.

2 Experimental Setup

High energy particle colliders and suitable detectors are the experimental setups physicists use to study the laws that govern physics at the smallest distance scales (highest energy regimes). The top quark is the heaviest known fundamental particle with a mass of the order of the gold atom. To produce such a heavy particle via the interaction of two quarks, initial state hadrons with an energy of about 1 TeV are necessary. The measurement of the production cross section of single top quarks presented in this thesis was done using collision-data from the Tevatron, a 1.96 TeV center of mass (CM) energy hadron collider at the Fermi National Accelerator Laboratory (Fermilab), located in Batavia near Chicago, Illinois, USA. The purpose of this chapter is to shortly describe the Fermilab's collider complex and in more detail the detector used to collect the data. This is done in Sects. 2.1 and 2.2 respectively.

2.1 Fermilab's Collider Complex

The measurement presented in this thesis was done based on data collected by the DØ experiment at Fermilab. Fermilab hosts the Tevatron collider, which at least until the beginning of the year 2010 remains the world's largest and highest energy particle accelerator. The Tevatron is the last piece of a bigger collider complex sketched out in Fig. 2.1, which generates, accelerates and brings to collision the proton and antiproton beams.

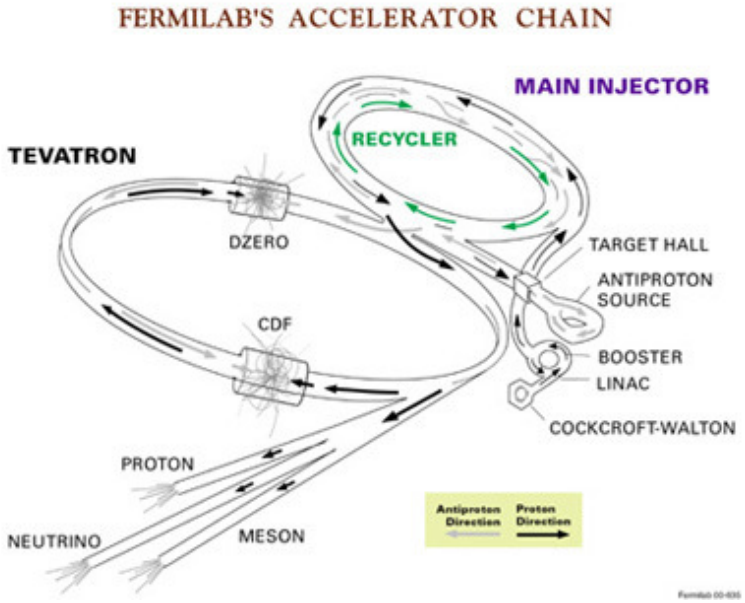

Figure 2.1 : Diagram of Fermilab's accelerator chain.

The first stage in the acceleration chain is the creation (and acceleration) of negative hydrogen ions (H^-) from compressed hydrogen gass (H_2) in a 750 kV electrostatic Cockcroft-Walton voltage multiplier. The so created continuous beam of ions then passes through a 150 m long Alvarez type Linear Accelerator (Linac), which uses radio frequency (RF) power to generate an oscillating electrical field in a repeated gap/drift-tube structure of increasing lenght, to accelerate the ions to 400 MeV. In the Linac, the beam also becomes bunched. The separation between bunches is such that the empty spaces where there is no beam coincide with where the electric field is pointing opposite to the beam direction. At the end of the Linac, the ions pass through a thin carbon foil, which strips off the electrons, leaving a beam of protons that is further injected into the Booster, a circular magnetic accelerator (synchrotron). The Booster is 475 m in circumference and accelerates protons from 400 MeV to 8 GeV in a period of 0.033 seconds. The beam goes around the Booster 16,000 times, seeing about 500 kV every turn. This accelerating voltage is given to the particles by RF cavities. The RF frequency in the cavities must change at a rate of 1 GHz per second to match the changing velocity of the protons as they are accelerated from 400 MeV to 8 GeV.

From the Booster, the protons pass into the Main Injector, which serves multiple purpose. In the Main Injector, protons are accelerated to 150 GeV for injection into the Tevatron or to 120 GeV in the fixed target operation mode for antiproton creation. The antiprotons are created by the Antiproton Source, consisting of three major components: a Target Station, a Debuncher and an Accumulator. In the Target Station, 120 GeV protons are smashed on to a nickel target every 1.5 sec producing a range of particles including antiprotons, which can be separated and stored in the Accumulator ring. The purpose of the Accumulator, as its name implies, is to accumulate antiprotons. This is accomplished by momentum stacking successive pulses of antiprotons from the Debuncher over several hours or days. Both RF and stochastic cooling systems are used in the momentum stacking process. The Accumulator can then pass the antiprotons to the Main Injector. The Main Injector then increases antiproton energy to 120 GeV in the fixed target operation mode or to 150 GeV for injection into the Tevatron. Because RF is used to accelerate the beam in the Main Injector, the 120 GeV protons that arrive at the target station are bunched, and therefore the antiprotons coming off the target are also bunched, having a very large (narrow) spread in energy (time) that is difficult for downstream accelerators to accept. The Debuncher accelerator is used to exchange the large energy (narrow time) spread into a narrow energy (large time) spread.

Placed in the Main Injector tunnel directly above the Main Injector beamline is the Recycler, a fixed-energy high reliability storage ring for antiprotons. The purpose of the Recycler is to further increase the luminosity of the Tevatron Collider over the luminosity goals of the Main Injector by itself. It uses permanent magnets and functions as a post-Accumulator ring. As the stack size in the Accumulator ring increases, there comes a point when the stacking rate starts to decrease. By emptying the contents of

the Accumulator into the Recycler periodically, the Accumulator is always operating in its optimum antiproton intensity regime. The other role of the Recycler, and by far the leading factor in luminosity increase, is to act as a receptacle for antiprotons left over at the end of Tevatron stores.

The Tevatron is the last link in the acceleration chain. It is a superconducting synchrotron that accelerates protons and antiprotons from the Main Injector in opposite directions up to 0.98 TeV in a 6.8 km ring. To hold the particles on track the Tevatron uses superconducting dipole magnets cooled in liquid helium at 4.3 K producing 4.2 T magnetic field. Protons and antiprotons cross paths at two different experimental areas of the Tevatron, named B0 and D0, to collide at a CM energy of 1.96 TeV. Surrounding these two collision points, are the big general-purpose CDF (Central Detector of Fermilab) and DØ detectors respectively, which record the outcome of the collisions.

The Tevatron started to operate in 1983. CDF began to take data in 1985 in what was called the Run 0 of the Tevatron. The DØ detector was build later and started data tacking in 1992 during the Tevatron Run I (1992−1996). During Run I, the Tevatron operated using six bunches each of protons and antiprotons, with 3500 ns between bunch crossings and a CM energy of 1.8 TeV. At this time the Main Injector did not exist and the Main Ring was used to accelerate protons for antiproton production while the Tevatron operated in collider mode. After Run I, the Tevatron collider was upgraded, increasing its CM energy to 1.96 TeV and reducing the bunch crossing spacing to 396 ns[15]. The Main Ring was removed and the Main Injector was incorporated to the accelerator complex. In March 2001, the Run II of the Tevatron started, which is hoped to last until the end of 2010. There was a shutdown period in the spring/summer of 2006, when the DØ detector was upgraded (mainly its tracking system). The Run II period before the detector upgrade is refered to as Run IIa, while the period after the upgrade is refered to as Run IIb. Approximatedly half of the data used in this analysis correspond to Run IIa, while the other half corresponds to the first year of Run IIb.

2.2 The DØ Detector

The DØ detector was proposed in 1983 to study high mass states and large p_T phenomena arising from the $p\bar{p}$ Tevatron collisions, performing very well during Run I, and leading, for example, to the discovery of the top quark in 1995. The detector was then significantly upgraded for Run II to take advantage of the improvements in the Tevatron (higher CM energy, shorter time between bunch crossings, higher instantaneous luminosity) and to enhance the physics reach of the experiment. The original

[15]Bunches for each particle, p or $\bar{p}$, are grouped in three superbunches spaced by 2 μs of twelve bunches each.

DØ detector as built for Run I is described in detail in Ref. [20], while a complete description of the upgraded DØ detector used during Run II is given in Ref. [21]. Since the analysis presented in this thesis was done using data from Run II, the detector description given in this chapter is based on the last reference.

DØ is a fairly typical high energy physics detector. It measures 9 m tall and 15 m long and consists of four major parts: a tracking and vertexing system, the calorimeters, a muon system, and a trigger system. A diagram of the detector is shown in Fig. 2.2. The innermost part of the detector surrounding the beryllium beam pipe close to the interaction point is the central tracking system, which is immersed in a static magnetic field produced by a solenoidal magnet that surrounds the tracking. After the magnet is the central calorimeter. Two forward calorimeters (north and south) are placed as end caps of the central tracking and central calorimeter. There are preshower detectors (not seen in the figure) located just before the calorimeter cryostat walls. Finally, the outermost part of the detector is the muon system, composed of muon chambers, muon scintillators and a toroidal magnet.

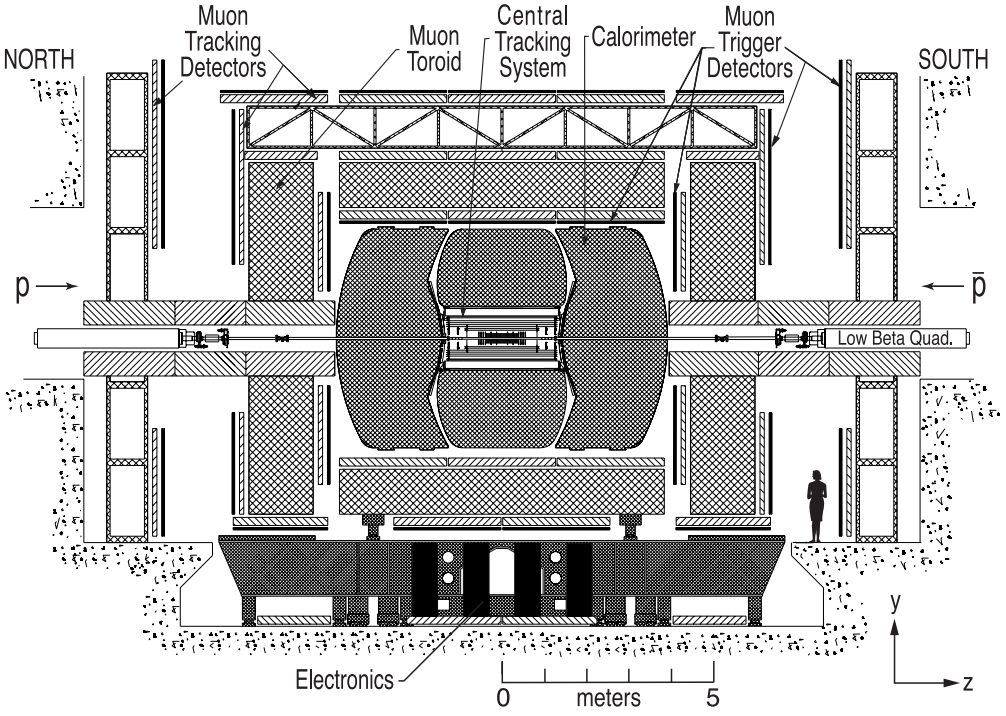

Figure 2.2 : Diagram of the upgraded DØ detector for Tevatron's Run II, as view from inside of the accelerator ring.

In the detector description and data analysis, a right-handed cartesian coordinate system is used in which the z-axis lies parallel to the beam pipe pointing along the proton direction, the y-axis is upwards and the x-axis points outwards from the center of the Tevatron ring. The origin of the coordinate system is at the geometrical center of the detector. The x and y axes define the transverse plane to the interaction, as

the momentum of the incoming particles lies entirely on the z-axis. The perpendicular distance from the z-axis is given by $r = \sqrt{x^2 + y^2}$. Azimuthal angles (ϕ) range from 0 to 2π, polar angles (θ) range from 0 to π, and are defined counter-clockwise with 0 degrees corresponding to the x and z axes directions respectively. Instead of θ, a much more convenient variable is used, namely, the pseudorapidity

$$\eta = -\ln[\tan(\theta/2)], \tag{2.1}$$

which approximates the Lorentz invariant true rapidity

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

for finite angles in the limit $E \gg mc^2$. The region defined by $|\eta| < 1.0$ is referred to as the "central" region, while the so-called "forward" region corresponds to high values of $|\eta|$ (typically $1.5 < |\eta| < 4.0$).

The interaction region, which has an extension of ≈ 40 μm in x and y and tens of cm in z, must have its center within 1 mm of the detector's center in the transverse plane and a few centimeters in the z direction.

2.2.1 Tracking System

With the central tracking system, tracks (trayectories) of charged particles can be reconstructed. The geometry and location of the tracks is then used, together with additional information from other detector parts, to identify final state particles (in this analysis electrons and muons) and measure their transverse momentum. It is also used to locate the hard-interaction (primary) vertex as well as decay (secondary) vertices. Secondary vertex identification is used to tag jets from b quarks, which is crucial to identify top quarks.

A schematic view of the Run IIa central tracking system is shown in Fig. 2.3. It measures 2.37 m long and surrounds the DØ beam pipe. It consist of the silicon microstrip tracker (SMT) and the central fiber tracker (CFT), surrounded by a superconducting solenoidal magnet that produces a 2 T magnetic field (B) pointing in the z direction in the tracking area, that serves to bend the trajectories of charged particles in the transverse plane. The radius of curvature (R) of the trajectory is related to the transverse momentum (p_T) and charge (q) of the particle according to the formula

$$R = \frac{p_T}{qB}. \tag{2.3}$$

The two tracking detectors locate the primary vertex with a resolution of about 35 μm along the beamline. They can tag b quark jets with an impact parameter resolution of better than 15 μm in $r-\phi$ for particles with $p_T > 10$ GeV/c at $|\eta| = 0$. The high resolution of the vertex position allows good measurement of the lepton p_T, jet transverse energy (E_T), and missing transverse energy ($\not\!\!E_T$).

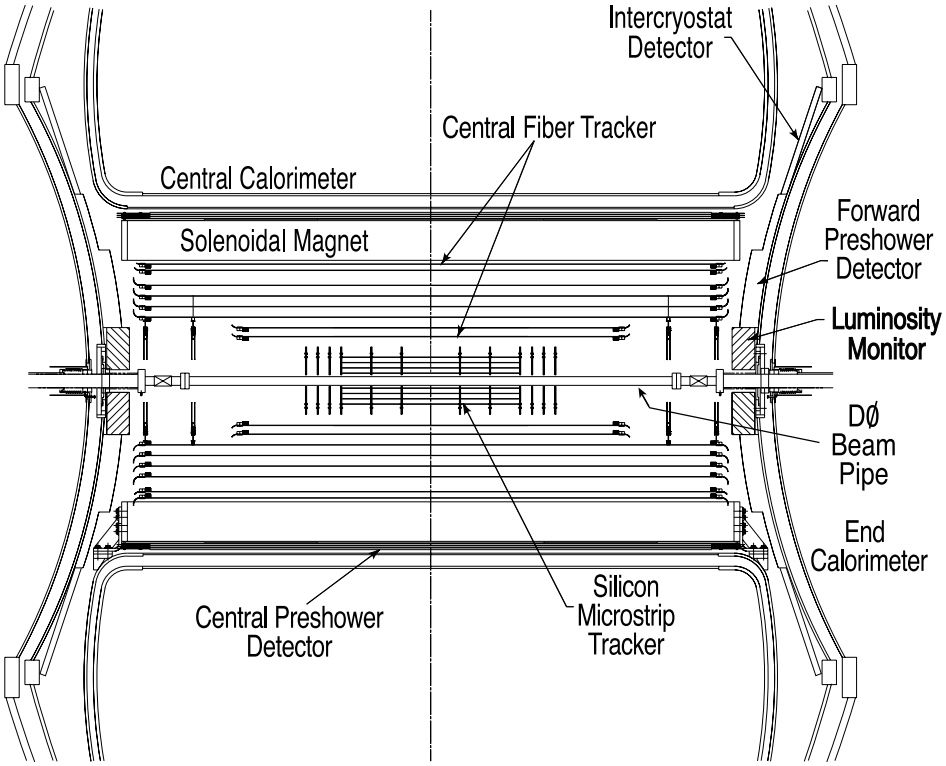

Figure 2.3 : Cross-sectional view of the Run IIa central tracking detector in the $x-z$ plane, including the solenoidal magnet. Also shown are the preshower detectors placed in front of the calorimeter chryostats.

2.2.2 Silicon Microstrip Tracker

The SMT detector is built from thin slices of silicon called wafers, with parallel metal strips along them. By the presence of an electric field across the wafer, excited electrons (and holes) in the silicon, created by a charged particle that passed through the silicon bulk, are drifted up to the closest strip, and a pulse of current is generated in this strip. The SMT is read out using 128-channel chip. It has a total of 912 readout modules, and about 790k readout channels. Silicon detectors give very precise information about tracking and vertexing (the hit resolution is 10 μm), used for example, to identify b quarks. However, since silicon sensors are very expensive, they had to be placed as close as possible to the beam pipe, where with a small instrumented area, a large range in η can be covered. Figure 2.4 shows an isometric view of the silicon tracker. The detector provides tracking and vertexing information over almost all the η coverage of the calorimeter and muon systems.

In the central part of the SMT, six barrel modules with four concentric silicon readout layers each (a fifth layer, named "layer 0", was added closest to the beam pipe in Run IIb), measure the proyection of the trajectories into the transverse plane. The

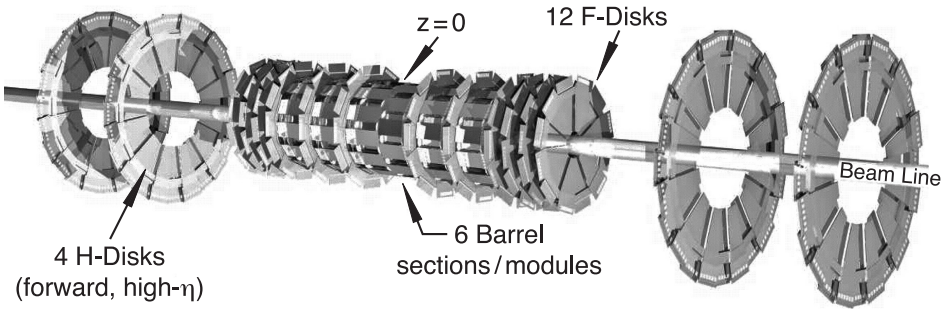

Figure 2.4 : The disk/barrel design of the Run IIa silicon microstrip tracker. For Run IIb the most forward north and south H-disks were removed.

silicon modules installed in the barrels are called "ladders". Figure 2.5 shows a cross section view of the four barrel layers used in the Run IIa SMT, where the position of the ladders can be appreciated. The two inner layers have 12 ladders, while the two outer layers have 24 ladders. The barrels use double-sided sensors in the second and fourth layers, with axial (parallel to the beam) strips on one side and 2° stereo strips on the other. The strip pitch is 50 (62.5) μm in the p-side (n-side). The first and third layers use single-sided sensors with axial strips in the two outer barrels with a strip pitch of 50 μm, while the inner four central barrels use double-sided-double-metal sensors with strips at a stereo angle of 90° and a strip pitch of 50 (153.5) μm in the p-side (n-side). The barrel detectors primarily measure the $r-\phi$ coordinate of the track hits.

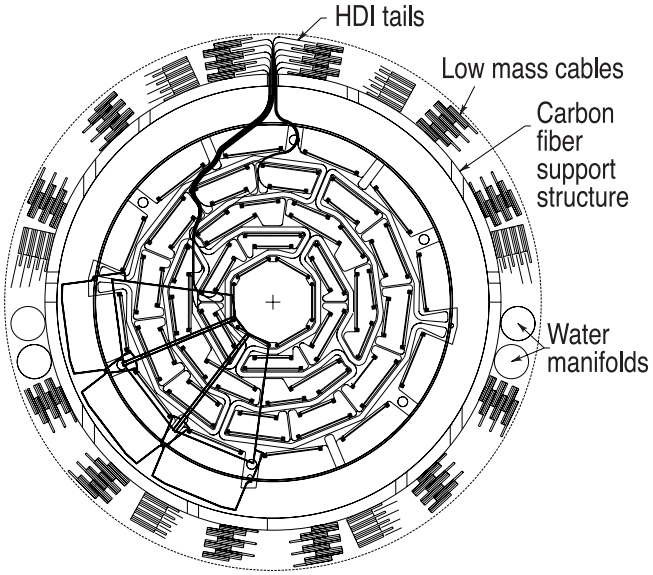

Figure 2.5 : Cross section view of the four layers in a barrel of the silicon microstrip tracker. In each layer, the disposition of the silicon wafers is shown.

The barrels are interspersed with six "F-disks" that provide additional information on

the $r-\phi$ position of the hits plus their position in the z direction. Forward of the three barrel/disk assemblies on each side is a unit consisting of three F-disks. The very forward region of the SMT (high values of η) is covered by four large diameter "H-disks" (for Run IIb, the two most forward H-disks were removed). The silicon sensors used in the disks are trapezoid shaped with readout strips arranged parallel to the long edge of the device. The F-disks have 12 double-sided wedges, providing an effective stereo angle of 30°. The strip pitch is 50 (62.5) μm in the p-side (n-side). The H-disks have 24 full wedges, each consisting of two single-sided wedges mounted back-to-back, giving an effective stereo angle of 15°. The strip pitch is 40 (80) μm in the inner (outer) sensors.

2.2.3 Central Fiber Tracker

Outside the silicon, DØ has a second tracking detector that provides full tracking coverage up to $|\eta| \approx 1.7$. This outer tracker is made using scintillating fibers, whose core material is polystryrene (PS) doped with paraterphenyl (pT) (an organic fluorescent dye) to about 1% by weight. A charged particle passing through the fiber core, may excite PS atoms, the excitations are transfered to the pT via a dipole-dipole interaction, and the pT atoms decay back rapidly emiting a photon of light ($\lambda \approx 340$ nm) that travels through the fiber. However, the emited light would rapidly be re-absorbed by the PS. To avoid this, a secondary wave-shifter dye, 3-hydroxyflavone (3HF), is added at a low concentration, but spectrally matched to the pT. The 3HF absorbs the 340 nm light photons emitted by the pT and re-emites it at 530 nm, which is well transmitted in PS.

The fibers are mounted on eight concentric support cylinders with the six outer cylinders being 2.52 m long, and the two innermost cylinders being shorter, 1.66 m long, to let space for the H-disks of the SMT, as can be seen in Fig. 2.3. Each cylinder supports one doublet layer of axial (z) fibers and a second doublet layer at a stereo angle of $+3°$ (u) or $-3°$ (v) with respect to the axial fibers. The fiber doublet orientation is zu in the odd-positioned (first, third, etc) cylinders and zv in the even-positioned ones. The scintillating fibers are coupled to clear fiber waveguides that carry the scintillation light to visible light photon counter (VLPCs) for read out.

The scintillating fibers are assembled into ribbons consisting of 256 fibers in two layers of 128 fibers each. Grooves made into a flat piece of acetal serve as guides for the first layer of fibers. The spacing between the grooves varies between 928 and 993 μm depending on the radius of the corresponding support cylinder. The fibers itself are 835 μm in diameter. The second layer of fibers is on top of the first layer and is offset by one-half of a fiber spacing as shown in Fig. 2.6. The hit resolution of each doublet layer is of 100 μm, much less than the SMT hit resolution. The acetal is then inserted into a rigid, curved backing plate of the corresponding cylinder radius.

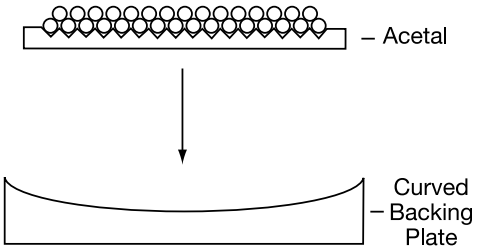

Figure 2.6 : Schematic graph showing the disposition of scintillating fibers into a ribbon. The grooves are made on a flat piece of acetal, which is then curved to the desired curvature radius by inserting it into a curved backing plate.

2.2.4 Solenoidal Magnet

The superconducting solenoidal magnet, also shown in Fig. 2.3, was designed to optimize the momentum resolution, $\delta p_T/p_T$, and tracking pattern recognition. It provides a uniform 2 T magnetic field parallel to the z direction accros all the tracking detector area. It measures 2.73 m long and 1.42 m in diameter and is operated with a current of 4749 A. Figure 2.7 shows a schematic view of the DØ magnetic field (in kG) in the $y-z$ plane including the solenoidal and toroidal magnets. The solenoid is about 0.9 radiation lenght (X_0) thick, which means that particles will initiate showers on it.

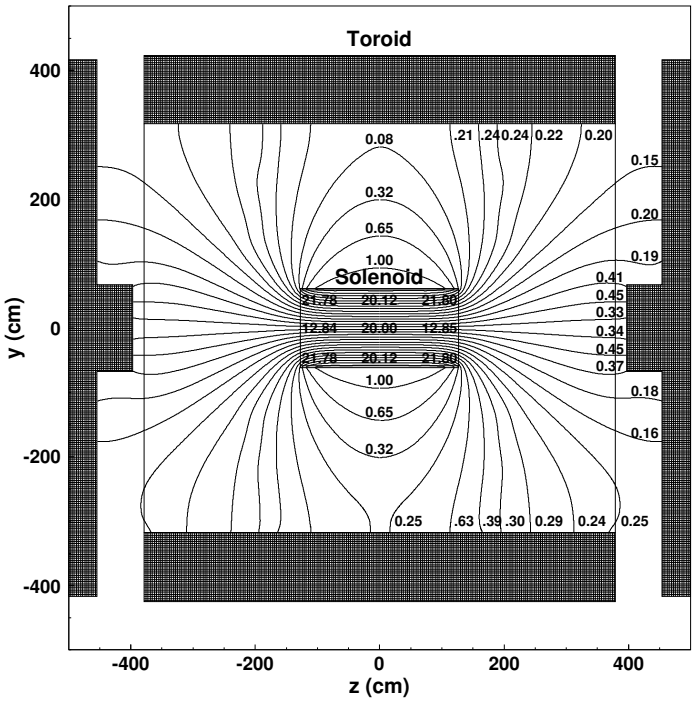

Figure 2.7 : DØ's magnetic field in $y-z$ produced by the solenoidal and toroidal magnets when operated at full current.

2.2.5 The Calorimeters

The calorimeters are the most important subdetector parts used in this thesis. They provide energy measurement for jets, electrons and photons, and assist in identification of jets, electrons, photons and muons. From these objects, only photons are not used in the single top quark cross section measurement. However, they are crucial objects in the jet energy scale determination. Energy depositions in unsampled regions, and elusive neutrinos which pass through the detector without leaving any trace, produce an energy imbalance in the transverse plane ($\not\!\!E_T$), which is also precisely measured in the calorimeters. The $\not\!\!E_T$ is another fundamental object in the jet energy scale and in single top processes with W leptonic decays.

A schematic view of the calorimeters is shown in Fig. 2.8. The pseudorapidity coverage of each of the calorimeters is seen in Fig. 2.9. The central calorimeter (CC) covers up to $|\eta| \approx 1$, and the two end calorimeters, ECN (north) and ECS (south), extend coverage up to $|\eta| \approx 4$. Each calorimeter consists of three sections: closest to the interaction region is the electromagnetic (EM) section, followed by the fine hadronic and the course hadronic modules.

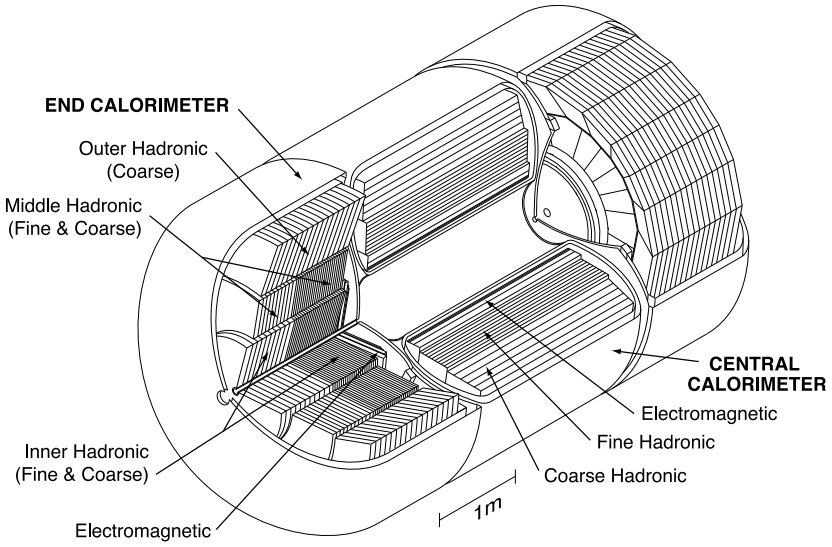

Figure 2.8 : Isometric view of the DØ calorimeter system.

The calorimeter detectors are formed from individual readout cells, which are grouped in towers shown in Fig. 2.9, with each tower subdivided in depth. The centers of the cells of increasing shower depth lie along rays projecting from the center of the interaction region. A typical calorimeter cell is shown in Fig. 2.10. A unit cell consists of an absorber plate and a signal board placed parallel to each other, with liquid argon as sampling material filling the gaps between absorber plates and signal boards. Each of the three calorimeters (CC, ECN and ECS) is located within its own cryostat to keep the detector temperature at approximately 90 K. The absorber plates of the EM

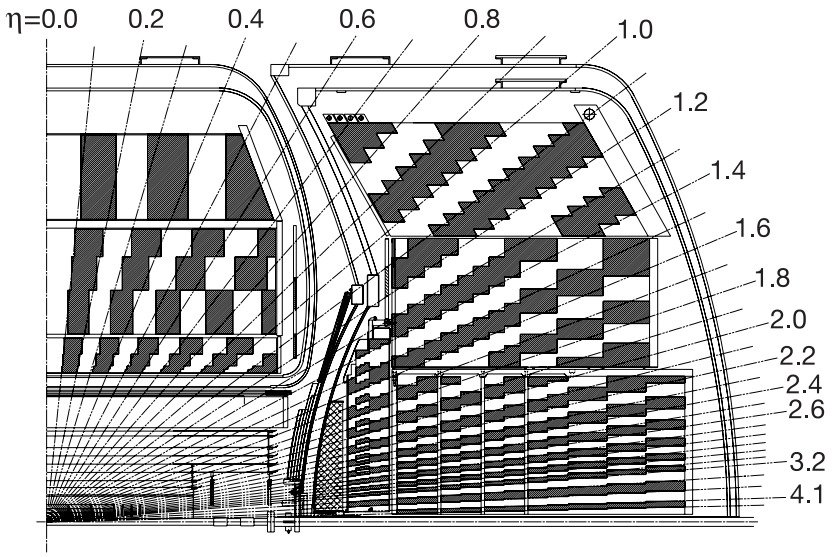

Figure 2.9 : Shematic view of the south portion of the DØ calorimeters showing the transverse and longitudinal segmentation pattern. The shading pattern indicates groups of cells for signal readout. The lines indicate 0.1 pseudorapidity intervals from the center of the detector.

modules are 3 mm thick in the CC and 4 mm in the EC, and are made from uranium (U^{238}). The fine hadronic sections use 6-mm-thick uranium-niobium (2%) alloy, and the course hadronic modules contain relatively thick (46.5 mm) plates of copper (in the CC) or stainless steel (EC). An electric field is established by grounding the metal absorber plates and connecting the resistive surfaces of the signal boards to positive high voltage (typically 2.0 kV). Signal boards are made from two 0.5 mm G-10 sheets for all but the EM and small-angle hadronic modules in the EC, where instead multilayer printed circuit boards are used. Surfaces facing to the liquid argon gap serve as the high voltage electrodes for the gap. The facing inner surfaces are milled into the pattern necessary for segmented readout. Several such pads at approximately the same η and ϕ are ganged together in depth to form a readout cell. The tower widths in both EM and hadronic modules are $\Delta\eta = 0.1$ and $\Delta\phi = 2\pi/64 \approx 0.1$, except for the third EM layer, which is segmented twice as finely.

When a passing through particle interacts with the atoms of an absorber plate, a cascade of secondary particles is initiated. These secondary particles lose part of their energy by ionizing argon atoms in the gap, freeing electrons that drift to the next signal board. Secondary particles that reach a next absorber plate may generate other particles, and the process repeats. The whole process is referred to as a shower development.

The appropriate scale lengths for describing the cascades are the *radiation length*

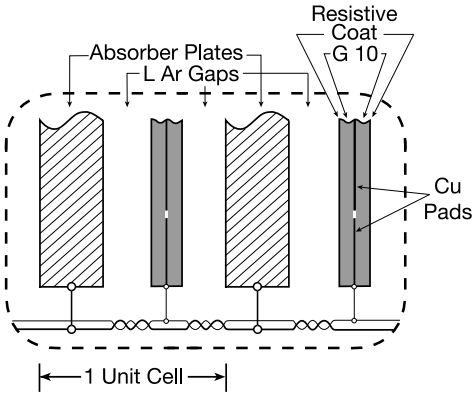

Figure 2.10 : Schematic view of the liquid argon gap and signal board unit cell for the calorimeter.

(X_0) for electromagnetic showers and the *nuclear interaction length* (λ_A) for hadronic showers. When a high-energy electron passes through matter, it loses energy by *bremsstrahlung* (radiation emitted by a charged particle under acceleration), while high-energy photons do it by e^-e^+ *pair production*. The radiation length is the mean free path[16] for high-energy electrons in the corresponding material. In the case of high-energy photons, the radiation length is 7/9 of the mean free path. For hadrons, the nuclear interaction length is the mean free path for energy loses by means of hadronic interactions.

The radiation length of uranium is 3.15 mm, comparable to the thickness of the absorber plates in the EM calorimeters[17]. Thus, electrons and photons entering the EM calorimeter initiate in average a new cascade per plate, multiplying copiously at each step. The cascade process continues until the secondary particles are no longer energetically capable of multiplying. At that point, the maximum number of shower particles exist. The EM modules of the calorimeters are subdivided in four separate layers, which together provide $\sim 20\ X_0$ of material. (The depths of the calorimeters themselves in CC and EC, from the outer warm walls to the first active liquid argon gap, is approximately 1.5 X_0. The components between the interaction region and the first active gap provide about 4.0 (4.5) X_0 of material in CC (EC).) With the maximum of such electromagnetic showers expected at the third layer (for this reason the third EM layer has a finer segmentation, to allow more precise location of electromagnetic shower centroids), they are practically entirely contained in the EM calorimeters. On the other hand, hadrons (charged pions, kaons, lambdas, etc.) mostly penetrate up to the hadronic modules where the hadronic shower then initiates.

The calorimeter signal is decisively determined by the peculiarities of the processes occurring at the last stages of the shower development. Although the energy of the

[16] Average distance over which the energy of a particle is reduced by a factor 1/e.

[17] The liquid argon radiation length is 14 cm, effectively acting only as a sampling material.

particles is very low, the number of particles is so large that a considerable fraction of the total energy is deposited through these processes. Beyond the depth of shower maximum, the number of particles dies away by ionization range out in the case of electrons, and Compton scattering and photoelectric absorption in the case of photons. Hadronic showers/calorimetry are far more complicated (see e.g. Ref. [22]). In the development of a shower generated by a high-energy hadron, lot of pions and ηs are produced (pion production occurs with approximately 1/3 probability for π^0s and 2/3 for charged pions). Since π^0s and ηs instantly decay into photons, a hadron shower has in general an electromagnetic and a non-electromagnetic component.

The following classes of secondary particles can be distinguished in the non-electromagnetic part of a hadronic shower: *a)* ionozing hadrons (charged pions, kaons, protons, etc.); *b)* soft neutrons; *c)* soft photons from nuclear deexcitation. The contribution to the calorimeter signal of pions, kaons, etc. is small ($10-15\%$ of the non-electromagnetic energy). Actually, most of the non-electromagnetic energy is used in nuclear processes to release nucleons and nucleon aggregates (like α-particles) from atomic nuclei and to give kinetic energy to the reaction products. Usually the nuclei also emit photons in this processes. Apart from these photons, protons and neutrons may contribute to the calorimeter signal. The vast majority of the protons will lose their complete kinetic energy by ionization. The neutrons exclusively lose kinetic energy through strong interactions, producing new protons, neutrons, and photons. Soft neutrons are the most crucial shower component for the calorimeter response since they predominate at the shower maximum. The last step in the neutron's lifetime as a free particle is capture. The cross section of this process is only considerable when the neutron is practically thermalized.

The average calorimeter response (signal per unit energy) for electromagnetic and hadronic showers are typically different ($e/h \neq 1$). Because of the loses in breaking nuclear bindings, the later is smaller and $e/h > 1$. The ratio e/h is a crucial variable to characterize the performance of a calorimeter. A calorimeter is said to be compensated when the ratio has the ideal value of 1. The DØ Run II calorimeter is strongly non-compensating. The reason is explained next.

The electron drift time in the 2.3 mm liquid argon gap is approximately 450 ns, a bit more than the 396 ns between bunch crossings. This fact, enhanced by the increased Tevatron's instantaneous luminosity for Run II, which results in up to five number of interactions in average per bunch crossing, results in considerable accumulations of energy in the calorimeter from bunch crossings previous to the current one (this effect is called "pile-up"). To reduce the noise from pile-up and thus keep a good signal-to-noise ratio, the charge integration time of the calorimeter preamplifiers has been reduced to 260 ns in Run II from the original 2.2 μs in Run I. This short integration time affects (lowers) principally the response of the calorimeter to a hadronic shower, because the relatively long time scale of neutron thermalization extends the time needed for its energy deposition.

The readout of the calorimeter involves about 55k channels and is done in three stages. First, signals are transported to charge preamplifiers, then to the analog signal shaping and storage circuits on baseline subtractor (BLS) boards, and finally the precision signals from the BLSs are transmitted to ADCs, from where they enter to the data acquisition system for trigger decision at Level 3 (see Sect. 2.2.9) and storage to tape. The BLSs provide baseline subtraction to remove any low frequency noise, noise from additional interactions, or noise from pile-up from the signal. In addition, faster shaped analog sums of the signals are picked off to provide prompt inputs to the calorimeter trigger system for both the Level 1 and Level 2 calorimeter trigger decisions.

The region $0.8 < |\eta| < 1.4$ in pseudorapidity is not well covered by the calorimeters as it corresponds to the region between the CC and ECs cryostats. In addition, there is substantial unsampled material (cryostat walls, cables, support structures) in there. Two intercryostat detectors (ICD) are attached to the exterior surfaces of the north and south end cryostats, made from sixteen 0.5 inch thick (trapezoid shaped) scintillating plates of $\Delta\eta \times \Delta\phi = 0.3 \times 2\pi/16 \approx 0.3 \times 0.4$ each, placed side-by-side in a ring covering the region $1.1 < |\eta| < 1.4$. The scintillating tiles are read out by PMTs. The PMT signal is stretched into a signal similar to that of the liquid argon readout to be compatible with the calorimeter BLS/ADC system. Additionally, single-cell structures called massless gaps (MG) (calorimeter readout cells with no absorber) are placed between the calorimeters and the cryostat walls, covering $0.8 < |\eta| < 1.2$ in the CC and $1.0 < |\eta| < 1.3$ in the ECs. The MGs measure showers that develop in the otherwise unsampled material between calorimeters.

2.2.6 Preshower Detectors

Preshower detectors are placed just before the cryostat walls of the calorimeters. The central preshower (CPS) covers the region $|\eta| < 1.3$, while the two forward preshowers cover $1.5 < |\eta| < 2.5$ (see Figs. 2.3 and 2.9). Both preshower detectors are made from equilateral triangular strips of scintillators arranged side-by-side with no dead space between strips. A wavelenght-shifting fiber at the center of the triangle strip collects and carries the light to the end of the detector, where they are connected to clear light-guide fibers that transmit the light up to VLPCs for read out.

The preshower detectors function both as calorimeters as well as tracking detectors. The CPS consists of three concentric cylindrical layers (axial x and stereo u and v) of triangular strips. Between the solenoid and the CPS, there is a lead radiator, that together with the solenoid, provide approximately 2 X_0 of material for particles at $|\eta| = 0$, and up to about 4 X_0 for the largest angles. The FPS are made from two layers (at different z) of two planes of scintillator strips each, with a 2 X_0 thick lead-stainless-steel absorber between the two layers. Both the first layer and the absorber cover $1.65 < |\eta| < 2.5$, while the second layer covers $1.5 < |\eta| < 2.5$. In the outer region of the

FPS ($1.5 < |\eta| < 1.65$), where there is no absorber nor first layer, the solenoidal magnet provides 3 X_0 of material in front of the preshower. Thus, electron and photon showers start already before the preshower detectors for $|\eta| < 1.65$, and in the absorber layer for $1.65 < |\eta| < 2.5$. The first layer of the FPS is then used to measure the location of tracks of charged particles; the second layer on the contrary is used for showering purposes, except in the case of heavier charged particles, which are less likely to shower in the absorber, and in this case only tracking information is obtained. The preshowers thus serve to enhance spatial match between tracks and calorimeter showers. They are used in electron and photon identification and for background rejection.

2.2.7 Muon System

The muon detector system lies in the outermost part of the DØ detector as seen in Fig. 2.2. High energy muons are a good sign of interesting collisions. Unlike electrons, photons and hadronic particles, muons do not get absorbed in the calorimeter, and their lifetime is long enough for them to leave the detector, what makes them easy to identify.

Exploded views of the muon system are shown in Figs. 2.11 and 2.12. The central muon system provides coverage for $|\eta| < 1.0$. It consists of three square layers of proportional drift tubes (PDTs), a central toroidal magnet located between the first inner layer (A) and the two outer layers (B and C), the cosmic cap and bottom scintillation counters, and the $A\phi$ scintillation counters installed on the PDTs mounted between the calorimeter and the toroidal magnet. The forward muon system provides coverage for $1.0 < |\eta| < 2.0$ and uses smaller mini drift tubes (MDTs). Three layers of scintillation counters, the end toroidal magnets and a shielding of the beam pipe, also form part of the forward muon detector. The forward muon system is particularly important for tracks with $1.6 < |\eta| < 2.0$, which do not go through all layers of the CFT.

The primary measurement of the muon momentum is made using the central tracking system. The toroidal magnet allows for stand-alone momentum measurement in the muon system, which improves the momentum resolution for high energetic muons. Having a stand-alone muon momentum measurement has additional benefits, like enabling low-p_T cutoff in the Level 1 muon trigger, and rejection of muons from pion and kaon decays.

The scintillation counters are fast and therefore used for triggering; the wire chambers are used for precise coordinate measurement and triggering. Both types of detectors contribute to background rejection. The scintillation counters provide timing information, which is used to reject cosmic ray background, while the wire chambers help with tracking. The cosmic cap and bottom counters provide a fast timing signal to associate a muon in a PDT with the appropriate bunch crossing and discriminate against the cosmic ray background. $A\phi$ scintillation counters provide fast detector for triggering

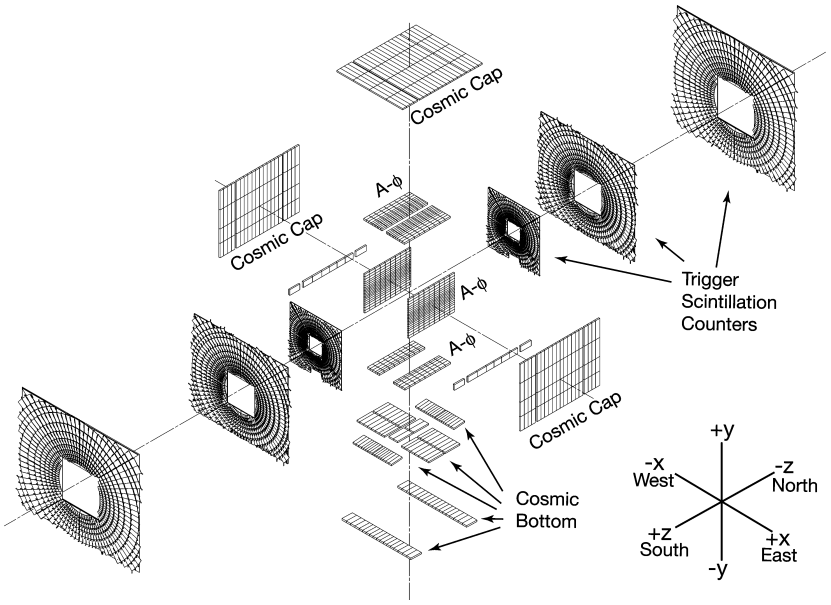

Figure 2.11 : Exploded view of the muon scintillator detectors.

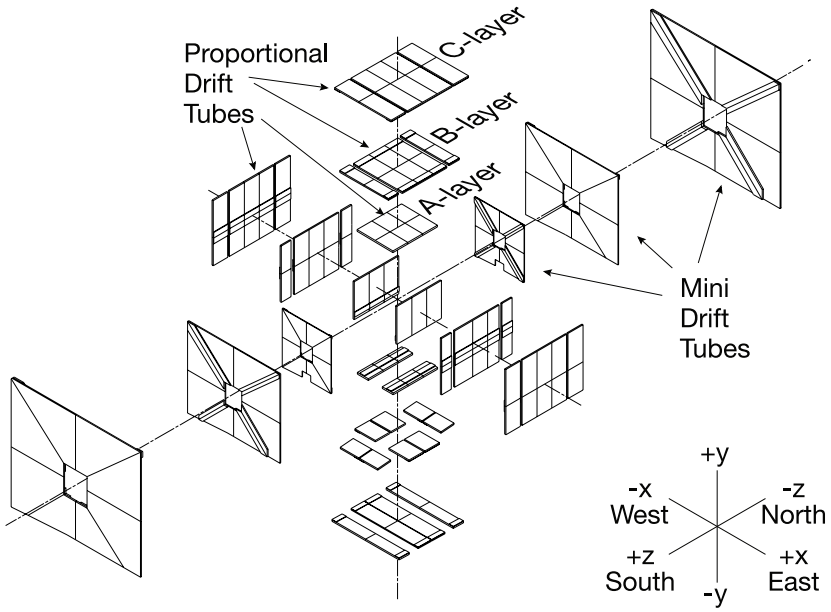

Figure 2.12 : Exploded view of the muon wire chambers.

on and identifying muons and for rejecting out-of-time backscatter from the forward direction. They also provide the time stamp for low-p_T muons which do not penetrate the toroid and thus do not reach the cosmic cap or bottom counters. The shielding in the accelerator tunnel reduces the non-muon background from beam halo. Shielding of the beam pipe between the end calorimeter cryostat and the wall of the collision hall, prevents non-muon background arising from the interaction of proton and antiproton remnants with the end of the calorimeter, the beam pipe and the low-beta quadrupole magnets.

2.2.8 Luminosity Monitor

In accelerator and particle physics, the *instantaneous luminosity* is defined as the number of particle collisions that occur per second per unit of transversal area. Increasing the number of collisions gives more probabilities for discoveries as more statistics becomes available. Collisions are recorded during the so-called "physics runs", which are time periods of well detector and data acquisition (DAQ) system functioning within accelerator stores. The instantaneous luminosity within a store decays exponentially with time. The *integrated luminosity* (L) is, as its name suggests, the time integral of the instantaneous luminosity, and is usually measured in pb^{-1} (or fb^{-1})[18]. So, for example, the integrated luminosity of Run I was 120 pb^{-1} and the Run IIa integrated luminosity was $\sim$ 1.1 fb^{-1}. This analysis uses 2.3 fb^{-1} of data divided in two independent sets, the 1.1 fb^{-1} from Run IIa and the earliest 1.2 fb^{-1} from Run IIb.

To measure the instantaneous luminosity, a dedicated detector −the luminosity monitor (LM)− is used. The LM consists of two rings of plastic scintillation counters located in front of the end calorimeter cryostats, surrounding the beam pipe below the forward preshower detectors (see Fig. 2.3). The scintillation counters are read out by PMTs connected to preamplifiers and ultimately to ADCs. The LMs pseudorapidity acceptance is $2.7 < |\eta| < 4.4$. The detector measures the average number of inelastic collisions per beam crossing, $\bar{N}_{LM}$, which multiplied by the beam crossing frequency f, gives the number of inelastic collisions per second (as measured by the LM system). Dividing this quantity by the effective cross section of the LM system, the instantaneous luminosity of the accelerator is obtained:

$$L = \frac{f\bar{N}_{LM}}{\sigma_{LM}}. \tag{2.4}$$

σ_{LM} is given by the product of the $p\bar{p}$ inelastic cross section (σ_{inel}) at the Tevatron, the LM detector acceptance (a_{LM}) and the LM detection efficiency (ϵ_{LM}):

$$\sigma_{LM} = \sigma_{inel} \times a_{LM} \times \epsilon_{LM}. \tag{2.5}$$

[18] 1 barn = 10^{-28} m^2, an area comparable to the cross sectional area of an uranium atom.

Since $\bar{N}_{LM}$ is typically greater than one, it becomes a difficult task to directly measure it from the activity signals in the LM. So, to account for multiple $p\bar{p}$ collisions in a single beam crossing, the fraction of beam crossings with no collisions is measured in the LM and then $\bar{N}_{LM}$ is determined using Poisson statistics.

A potential background that would degrade the instantaneous luminosity measurement in the LM, is beam halo. To separate this processes, the difference in time-of-flight as measured by the two LM detectors at $z = \pm 140$ cm is calculated, and translated to a distance measure by approximating the particles velocity by the speed of light, $z_v = \frac{c}{2}(t_- - t_+)$. This quantity gives a fast measure of the z position of the primary vertex v in case of real beam-beam collisions, while for beam halo particles it gives $\approx \pm 140$ cm. Beam halo background is eliminated by requiring $|z_v| < 100$ cm, range that encompasses almost all $p\bar{p}$ Tevatron collisions occurring at DØ ($\Delta_{z_v} \approx 30$ cm).

Equations (2.4) and (2.5) are used in general to determine the cross section of a process X measured by the DØ experiment. One has to replace $f\bar{N}_{LM}$ by the measured number of events N_X, the cross section σ_{inel} by the cross section σ_X of process X, and use the proper acceptance a and efficiency ϵ to correctly define the effective cross section.

2.2.9 Trigger System

Every second, there are 1.7 million Tevatron $p\bar{p}$ collisions occurring at the center of the DØ detector. From these collisions, only those containing "interesting" physics events are selected to store on computer tape. The trigger is a system of fast electronics and computers that decides, in real time, whether a physics event is interesting enough to be worth keeping. The trigger system has three levels at which detector data are analyzed in series. Only if the event is accepted at the trigger level i ($i = 1, 2, 3$), it is further analyzed at level $i + 1$ (or recorded if $i = 3$). So, each succeeding level in the trigger system analyzes fewer number of events per second (having more available time per event), but does it in more detail and at a deeper level of complexity. An overview of the DØ trigger system and data acquisition system is shown in Fig. 2.13. The first stage −the Level 1 (L1) trigger− uses hardware circuits and reduces the event rate from 1.7 MHz to 2 KHz, based on a rough reconstruction of tracks in the central tracking and energy deposition in the calorimeter. The Level 2 (L2) stage also uses hardware engines, including preprocessors at specific subdetectors that provide information to a global L2 processor. The L2 trigger reduces the event rate by a factor of two, from 2 KHz to 1 KHz. Finally, the Level 3 (L3) trigger uses sophisticated software algorithms running on a computer farm to reconstruct more complicated event characteristics. Events passing the L3 stage do it at a rate of 50 Hz and are sent to tape. Figure 2.14 shows the L1 and L2 DØ trigger systems. The trigger framework (TFW) takes the L1 decision based on information from the specific L1 trigger devices (L1CAL, L1CTT and L1FPD). In addition, it manages the prescaling of triggers (predefined

ratio of rate reduction of a specific trigger, by randomly selecting events that passed the corresponding trigger).

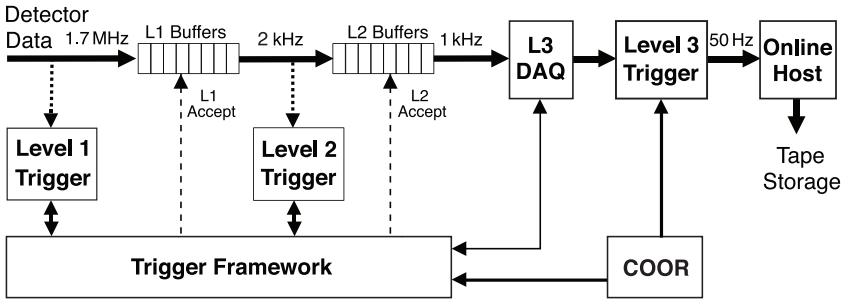

Figure 2.13 : Schematic overview of the DØ trigger and data acquisition systems.

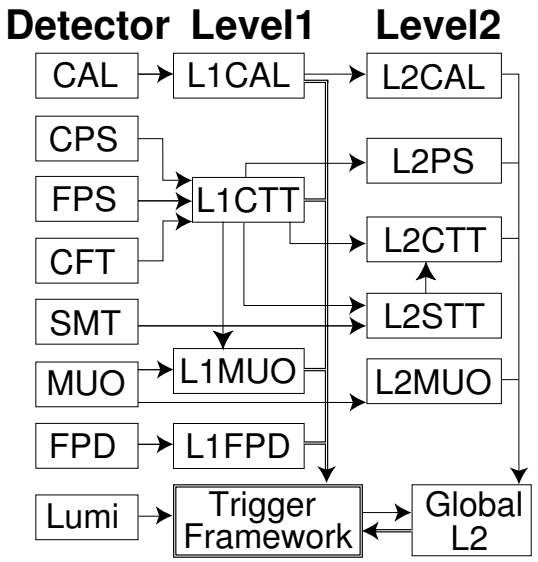

Figure 2.14 : Block diagram of the DØ L1 and L2 trigger systems. The arrows show the flow of trigger-related data.

In the following, a brief description of the three trigger levels is given.

The Level 1 Trigger

As seen in Fig. 2.14, the L1 trigger system uses four subdetector triggers. The calorimeter trigger (L1CAL) searches for patterns of energy deposition exceeding predefined limits on transverse energy deposits. Electromagnetic (EM) and hadronic (H) trigger towers are 0.2×0.2 in $\Delta\eta \times \Delta\phi$. The variables used in trigger calculations are the EM transverse energy and EM+H transverse energy, formed by adding the corresponding EM and H trigger towers energies. The triggers available for use in the experiment include global variables ($\sum E_T$ −the sum of all tower E_Ts− and $\not\!\!E_T$)

and local variables (each EM tower and each EM+H tower is compared to many programmable E_T threshold sets). The central track trigger (L1CTT) and muon system trigger (L1MUO) compare tracks, separately and together, to see if they exceed preset thresholds in transverse momentum. L1CTT uses fast discriminator data provided by three scintillator-based detectors: CFT, CPS and FPS. L1MUO looks for patterns consistent with muons using hits from muon wire chambers, muon scintillation counters, and tracks from the L1CTT. The L1 forward proton detector trigger (L1FPD) is used for diffractive physics.

The Level 2 Trigger

L2 includes preprocessors for tracking, calorimeter, preshower and muon systems, working in parallel, that build its own physics objects. Data are then passed to a global processor (L2Global) where the trigger decision is made. At this stage data can be combined to form higher quality physics objects to examine event-wide correlations in all L2 physics objects. The calorimeter preprocessor system identifies jets and EM candidates and calculates event $\not\!\!E_T$ for the global processor. Both the central and the forward preshower detectors are designed to provide high electron detection efficiency, electron-photon separation and high background (charged hadron) rejection at the trigger level. This is accomplished by providing evidence for early shower development and by giving a good spatial point for comparison with calorimeter clusters for tracks. Clusters centroids in the three CPS layers are compared to produce η and ϕ coordinates for clusters that match in the three layers. The presence or absence of CFT trigger tracks associated with CPS clusters is also provided and output clusters are flagged as electrons (when there is a track associated with the cluster) or photons (when no track is associated). The η and ϕ coordinates are binned to correspond to the calorimeter trigger towers geometry of $\eta \times \phi = 0.2 \times 0.2$. The L2STT performs online pattern recognition in the data from the SMT. It reconstructs charged particle tracks found in the CFT at L1 with increased precision by utilizing the much finer spacial resolution of the SMT. The L2STT improves the momentum measurement of charged particles tracks at the trigger level. The primary physics justification of the L2STT is its ability to measure the impact parameter of tracks precisely enough to tag the decays of long-lived particles, specifically B hadrons, helping to select events with an enhanced heavy-flavor content. Finally, L2Muon uses calibration and more precise timing information to improve the quality of the muon candidates.

The Level 3 Trigger

A high level, fully programmable software trigger, performs a limited reconstruction of events. L3 decisions are based on complete physics objects as well as on the relationships between such objects, like for example, the separation in rapidity or azimuthal angle between physics objects, or the invariant mass. Candidate physics objects, or relations between them, are generated by object-specific software algorithms

called *filter tools*. Tools perform unpacking of raw data, locating hits, forming clusters, applying calibration, and reconstructing electrons, muons, taus, jets, vertices and $\not\!\!E_T$. Tools use *reference sets* (*refsets*) of programmable algorithm parameters as inputs to define the physics objects precisely (for example, jet refsets specify the cone size, electron refsets specify the electromagnetic energy fraction, etc). Individual calls to the tools are made by *filters* that define the specific selection criteria. These criteria include the refset used by the tool, as well as thresholds and other cuts applied by the filter on the results of a tool (for example, the requirement of two jets within a given pseudorapidity range above a fixed E_T threshold).

Jet triggers were extensively used for the calibration of the jet energy with the calorimeter. On the other hand, for the single top analysis, there was no specific trigger requirement in the event selection. Events were required to fire any of a long list of "reasonable" triggers. This is to increase statistics and because disagreement was seen between data and simulation in physics kinematic distributions when requiring for example a single lepton+jet trigger, which was believed to come from mismodeling of the available trigger turn-on curves.

3 Event Reconstruction and Object Identification

The previous chapter described the different subdetector parts of the DØ detector and the signals they produce when particles pass through. The next steps in the data processing chain are implemented within the DØ Offline Reconstruction Program (d0reco), which is responsible for reconstructing the objects that are used to perform all DØ physics analyses. The program processes either collider events recorded during online data taking or simulated events produced with the DØ Monte Carlo (MC) program. The executable is run offline and the results are placed into the central data storage system for further analysis. The first step in d0reco involves decoding the raw information of individual detectors to associate electronics channels with physical detector elements, including the application of detector specific calibration constants. This information is then used in a second step to reconstruct clusters of energy (for example, in the calorimeter and the preshowers) and/or hits (in the tracking detectors). Geometry constants are used to associate detector elements with physical positions in space. The third step in d0reco is to build global tracks out of the hits in the SMT and CFT detectors. The results are used as input, in the fourth step, by the vertexing algorithms to search for primary vertex candidates indicating the locations of $p\bar{p}$ interactions. Then, displaced secondary vertex candidates associated with the decays of long-lived particles are identified. The results of the above algorithms are then available for the final step in d0reco, the identification of physics objects for analysis usage. Information from each of the preceding reconstruction steps are combined to create standard physics object candidates (electron, photon, muon, neutrino (missing E_T) and jet candidates). Finally, using all previous results, candidates for heavy-quark and tau decays are identified. The identification of physics objects depends on selection criteria that define the "quality" of the physics objects. This chapter describes the algorithms and procedures used to reconstruct and identify the physics objects used along this thesis.

3.1 Tracks

A very important part in the event reconstruction is the determination of the trajectories followed by (charged) particles as they pass through the detector. These trajectories are represented by reconstructed *tracks*. In modern high-energy high-luminosity colliders, like the Tevatron, many events may happen in a single beam crossing, and in each event many charged particles can be produced even in a small region in η. As a result, the tracking system can have a high density of hits, which makes the finding of tracks a very complicated task. Tracks are reconstructed with the tracking system (SMT and CFT) and are used in the reconstruction of interaction vertices and in the identification of leptons and photons.

For track reconstruction purposes, the tracking detector is described by a collection of surfaces, which are cylinders coaxial with the z axis for the CFT, and x-y and r-z planes for the SMT. When charged particles pass through a tracking detector, they interact with the active material creating clusters of energy deposition. The clusters are used to generate hits, which are defined as the position in space of the intersection between a track and a surface describing the detector. Hits consist of one or more parameters with associated errors, and each track consists of a surface, a set of track parameters at that surface and an error matrix for these parameters. The track reconstruction involves two processes: track-finding and track-fitting. The goal of track-finding is to assign ordered lists of hits to reconstructed tracks, while the aim of track-fitting is to determine the optimal kinematic track parameters.

In the DØ experiment, a "smart" combination of three different algorithms is used to reconstruct tracks in two general steps. First, two pattern recognition algorithms, the histogramming track finder (HTF) [23] and the alternative algorithm for tracking (AATrack) [24, 25], produce a complete set of candidate tracks. The DØ Kalman track fit algorithm [26] is then used in the second step to provide the final tracks in the event. The combination of the best features of all three algorithms improves the track reconstruction performance. The Kalman algorithm has actually no built-in way to start a track. Thus, the Kalman fitter needs as input an initial guess or a partially reconstructed track, and it then extends the track by the inclusion of additional measurements. The Kalman fit determines, from a set of measurements, the optimal track parameters, with errors, on any surface.

The HTF algorithm is a histogramming approach based on a Hough transform. The algorithm assumes that the trajectory of a charged particle in the transverse plane (x, y), in the fairly homogeneous magnetic field present in the tracking area and assuming absence of any material, can be approximated by a circle, and is therefore characterized by three parameters, e.g. (ρ, d_0, ϕ), where $\rho = qB/p_T$ is the inverse of the curvature, d_0 is the impact parameter (distance of closest approach to the origin), and $\phi \in [0, 2\pi)$ is the direction of the track at the point of closest approach to (0, 0). Assuming $d_0 \sim 0$, each trajectory is univocally determined giving three points (the origin plus two hits). Thus, for any pair of hits in the x-y plane, there is a unique value in the ρ-ϕ plane associated to that pair. Hits corresponding to a real trajectory result in a peak in that plane, while background hits appear randomly disseminated. Building of tracks is then reduced to finding maxima in the ρ-ϕ plane. To make the algorithm more efficient, a Hough transform is used, where instead of considering all possible pair of hits, maxima are found as intersections of lines, each line corresponding to the family of circles that pass through the origin and a hit. The z information is added by another histogramming, which considers the coordinate space (r, z) and the parameter space (z_0, C), where z_0 is the position of the track origin along the z-axis, and $C = dz/dr$ is the track inclination in (r, z). A 3D Kalman filter is used to build the final tracks. This procedure applies either to the SMT or the CFT. Tracks are then extrapolated

to the CFT or SMT respectively to build complete tracks.

The AATrack is a road-following with a Kalman filter update. An initial track hypothesis is constructed using 3 SMT hits. The selection of hits starts from the innermost layer of the SMT and goes to the outer-most layer of the tracking system. To incorporate further hits to the track hypothesis, the track parameters are first evaluated at a following surface using a propagator [27], which in general will account for the magnetic field and for interactions with detector material. Hits are created from a window of "nearby" clusters in the surface of the propagated track, and a hit is then added to create a new track using χ^2 minimization. Because of detectors gaps and inefficiencies, a track may cross a surface without picking up a hit. These are called missed surfaces. The construction of a track hypothesis continues until the end of the detector is reached or three consecutive missed surfaces are found after the last incorporated hit. The fitting is done with a Kalman filter, which allows to update the track parameters and their error matrix by adding the information from a hit in the surface of the propagated track. In this way, a pool of track hypotheses is constructed. The tracks in the pool are then ordered according to the following criteria: 1) hypotheses with larger number of hits are placed first; 2) for hypotheses with the same number of hits, those with smaller number of missed surfaces are placed first; 3) for hypotheses with the same number of hits and missed surfaces, those with smaller χ^2 of track fit are placed first. Hypotheses are considered in the previously defined order and a hypothesis is declared an AATrack if it satisfies a set of so-called "number of shared hits" criteria (see Ref. [25]). The fake rate can be reduced by determining the primary vertices (PVs) position using this AATrack sample: AATracks with small impact parameter with respect to any PV receive two additional "hits", track hypotheses are reordered and the final AATracks are determined in the same way as explained above. So far, this method applies to tracks with at least 3 SMT hits. Tracks with smaller number of SMT hits are instead reconstructed starting from 3 hits in the CFT, requiring the track hypothesis to pass near some PV position found with SMT tracks, continuing the track reconstruction in the CFT layers as described above for the SMT, and finally extrapolating into the SMT and collect all possible SMT hits.

3.2 Vertices

An accurate determination of the position of the interaction points in each beam crossing becomes relevant, for example, in the calculation of the transverse momentum of electrons and muons and the transverse energy of jets, since the direction in space where an object propagates must be measured with respect to the interaction point where this object originates. Also, the identification of jets coming from b-quarks relies on the ability to discriminate between primary interaction points, which principally correspond to the hadron-hadron collisions, and displaced interaction points corresponding to the decay of the B-hadrons. In the previous chapters, the interaction

points were also called vertices. For the purposes of this chapter is better to reserve this terminology to denote only an interaction point at the reconstructed level, in a similar way as we differentiate in Sect. 3.1 between 'trajectories' and 'tracks'. So, with the current resolution of the tracking detectors, the event PVs contain the hadron-hadron collisions interaction points plus secondary (and maybe other) interaction points corresponding to the decay of short-lived particles (resonances), while secondary vertices (SVs) are possible due to the "long" lifetime of some particles (B- and D-hadrons, Λs, K_s^0, etc.).

At the Tevatron instantaneous luminosities, the number of $p\bar{p}$ interactions follows a Poisson distribution with a mean of 3.5. Only one of these interactions will eventually be triggered as a hard-scatter, while the others are soft elastic interactions (called minimum bias interactions). However, one PV will in principle be reconstructed for each $p\bar{p}$ interaction with the one corresponding to the hard-scatter usually referred to as "the PV". As mentioned at the end of Sect. 2.2, the position of the PVs in the x-y plane is very constrained, but in the z direction PVs can be anywhere in a region of ~ 1 m along the detector center. As for tracks, vertex reconstruction follows two main steps: vertex finding and vertex fitting. Vertex finding consists in the association of particle tracks to vertices, and vertex fitting in the accurate estimation of the vertex position.

3.2.1 Primary Vertices

One of the main challenges of PV reconstruction, is the ability to discriminate tracks from PVs and tracks from SVs with small decay length. In DØ, the PV reconstruction is performed using a robust Adaptive primary vertex algorithm [28, 29], which was specially designed to deal with misassociated and misreconstructed tracks in a much more efficient way as the previously used standard Kalman Filter vertex fitting algorithm [30]. The Adaptive vertex fitter algorithm is implemented iteratively. In each iteration, all primary vertex track candidates are fitted using the Kalman filter algorithm, but assigning to each track a weight that depends on its χ^2 distance to the vertex position determined in the previous iteration. If the weight is less than 10^{-6}, the track is eliminated from the fit. For the first iteration, all track weights are set to 1. The weighting function used is a sigmoidal,

$$w_i = \frac{1}{1 + e^{(\chi_i^2 - \chi_{cutoff}^2)/2T}}, \tag{3.1}$$

where χ_i^2 is the χ^2 contribution of track i to the vertex, and χ_{cutoff}^2 and T are parameters that control the width and sharpness of the function. The algorithm iterates until the error, defined as the maximum change in track weights ($\max|w_i - w_{i-1}|$), is less than 10^{-4} and the total number of iterations is smaller than 100. If the error stays greater than 10^{-4} until iteration 100 the algorithm returns non-convergence. The standard

Kalman vertex fitting algorithm is similar to the Adaptive technique with $T = 0$ in Eq. (3.1), in which case tracks have only two possible weights, 0 or 1. In the Adaptive fitting method, $T > 0$, and all tracks can potentially contribute to the PVs with different weights. The closer the track is to the PV, the higher is its weight and its consequent contribution to the vertex position estimation. The Adaptive algorithm reduces PV biases due to the presence of close SVs and significantly improves the positioning resolution.

The reconstruction and identification of primary vertices using Adaptive fitting consist of the following steps:

Track selection In order to remove poorly measured and fake tracks, tracks are required to have $p_T > 0.5$ GeV/c, and at least 2 SMT hits if the track η-z is within the SMT fiducial region, defined by

$$\begin{aligned} -36 \text{ cm} < z < {} & 36 \text{ cm} + 44(-\eta - 1) \text{ cm} && \text{if} \quad -2 < \eta < -1, \\ -36 \text{ cm} < z < {} & 36 \text{ cm} && \text{if} \quad -1 < \eta < 1, \\ -36 \text{ cm} - 44(\eta - 1) \text{ cm} < z < {} & 36 \text{ cm} && \text{if} \quad 1 < \eta < 2. \end{aligned} \tag{3.2}$$

Otherwise, no SMT hit requirement is imposed, as tracks outside this region mostly have 0 SMT hits. Tracks are then clustered along the z direction to identify tracks belonging to different interactions. Starting from the highest-p_T selected track, tracks are added to the cluster if they are within 2 cm of the seed track.

Vertex fitting Vertex fitting is a three-step process. In the first step, all selected tracks within each z-cluster are fitted into a common vertex using the Kalman Filter vertex fitting algorithm, to get an estimate of the beam position and width. In this fit, the track(s) with the highest χ^2 contribution to the vertex are removed in turn, until the total vertex χ^2 per degree of freedom is smaller than 10. In the second step, in each of the z-clusters, tracks with a distance of closest approach (dca) to the previously determined beam spot smaller than 5σ ($dca/\sigma(dca) < 5$), are preselected. Finally, the preselected tracks are fitted into a common vertex using the Adaptive vertex fitter algorithm explained above.

Vertex selection Vertex selection consists in the identification of the hard-scatter and additional minimum bias (MB) vertices in the event. The hard-scatter vertex is selected using a minimum bias probability selection algorithm [31], where each vertex is assigned a probability to be consistent with a MB interaction. The vertex with the smallest MB probability is defined as the hard-scatter PV. The algorithm is based on the fact that tracks from hard interactions have a harder p_T spectrum than tracks from MB interactions. The first step of the algorithm is to cluster all reconstructed vertices in the z direction within 2 cm of each other, and select the highest multiplicity vertex in each cluster, so that split primary vertices[19] are removed. For every selected

[19]These are low multiplicity vertices, close to the real high multiplicity PV, formed from poorly

vertex, all tracks within some distance around the vertex are used to compute the MB probability.

The primary vertex reconstruction and selection efficiencies are very close to 100% in the whole z range entering in the definition of the SMT fiducial region ($|z| < 80$ cm). The PV resolutions in the transverse plane and the z direction are around 10 μm and 20 μm respectively.

3.2.2 Secondary Vertices

Secondary and terciary vertices are important to determine cascade decays of long-lived particles, and for the identification of b-quark jets. In DØ, the secondary vertex reconstruction algorithm is based on a Kalman filter technique [32]. The algorithm to find secondary vertices consists of the following steps:

Track clustering First, tracks associated to MB vertices are removed. From the remaining tracks, those with $p_T > 2$ GeV/c are used as seeds to build clusters of tracks using a simple cone of size 0.5 in $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. Starting from the highest-p_T seed track, the closest track is added in turn to the cluster if $\Delta R < 0.5$, and the cluster $\vec{p}$ is recomputed to redefine the direction of the cluster. Initially, the seed track $\vec{p}$ defines the cluster $\vec{p}$. The clustering process finishes when there are no more seed tracks. A cluster is kept if $\sum_{\text{tracks}} p_T^{\text{track}} > 5$ GeV/c (the sum is over the tracks that form the cluster). Note that clusters may share tracks. The minimum p_T of a seed track and the minimum $|\vec{p}_T|$ to keep a cluster were tuned to efficiently remove fake clusters[20] while keeping the purity of selecting matched (not fake) clusters very close to 100%.

Vertex finding Within each cluster, all possible 2-tracks combinations are fitted using a Kalman Filter technique and those pairs with $\chi^2 < 10$ are selected as seeds for vertex finding. There is a high number of fakes coming from the casual crossing of 2 tracks, where mostly all cases have at least one track associated to the PV. To reduce this background, tracks previously associated to the PV are rejected before forming the seeds. For each seed, any track (not only those belonging to the cluster) is attached if its χ^2 contribution to the vertex is smaller than 10.

Preliminary vertex selection Vertices are preselected with the following cuts on kinematic variables:

– Decay length: the transverse (longitudinal) decay length is required to be less than 2.5 cm (3.0 cm). These cuts remove Λ and K_s^0 decays, as well as B/D-hadron decays for events where the selected PV is really a MB vertex far away

reconstructed tracks that do not get attached to (although they may come from) the real PV.

[20]Fake clusters are defined in MC as those with zero or only one track matched to a MC vertex.

from the true PV.

- Opening angle: the opening angle is defined as $A = 1 - \cos(\alpha)$, where α is the angle between the cluster's direction (defined by its momentum) and the vector that goes from the PV to the SV. The opening angle is required to be less than 0.1, meaning that the cluster of tracks originated from the SV (the tracks of the decay products) should point almost in the same direction as the incident particle decaying in the SV.

- Track multiplicity and track p_T: 2-track vertices are required to have at least one of the tracks with $p_T > 2$ GeV/c; vertices with a higher multiplicity should have at least one track with $p_T > 1.5$ GeV/c.

Final vertex selection Remove multiple associations of a same track to more than one vertices. Iteratively select the best vertex based on the opening angle and remove all remaining vertices that have at least one shared track with the selected vertex until no more shared-track vertices are left.

The global efficiency after final vertex selection is around 0.45. The secondary vertex resolution is around 150 μm.

3.3 Calorimeter Clusters

Calorimeter objects are intensively used along this thesis. The jet energy calibration procedure, which will be explained later in chapter 4, entirely relies on photons, jets and missing transverse energy. The single top analysis also makes use of jets and missing E_T, as well as electrons. All calorimeter objects are reconstructed starting from the individual calorimeter cells.

Each cell in the DØ calorimeter is subjected to a gaussian noise from the electronic readout components, as well as a non-gaussian noise from the decay of the uranium absorber. The amount of this noise is called the pedestal and it is determined for each calorimeter channel in periods without beam between subsequent Tevatron stores. To reduce the contribution of calorimeter noise to the measured cell energy and the amount of information to be read by the reconstruction algorithms, the average of the pedestal distribution is subtracted in the hardware. This implies that the measured energy of a given individual cell can have positive or negative values. A cell is then read out only if the absolute value of its energy is larger than a certain threshold, set in the hardware to $1.5\sigma_{\text{ped}}$, where σ_{ped} is the root-mean-squared (RMS) of the pedestal distribution. This cut on the calorimeter cell energy is referred to as "online zero-suppression".

Extremely noisy ("hot") calorimeter cells, characterized by a large value of their energy (typically > 1 GeV), are usually related to detector failures. Hot cells are searched for

at the online level first, then at the Level 3 trigger, and finally at an early stage of the reconstruction with the NADA algorithm [33, 34]. Once identified, they are also removed.

In order to get a finer and uniform treatment of the calorimeter noise in the reconstruction of calorimeter objects (electrons, photons, jets, missing E_T), DØ implemented the so-called T42 algorithm [35–37], which builds calorimeter clusters that serve as the starting points for the different calorimeter objects reconstruction algorithms. The idea behind the algorithm is that noise cells are most likely to appear isolated. The algorithm starts by rejecting cells with negative energy, since they are not able to compensate the positive noise when grouping cells into towers as was initially argued [35]. This is because the different layers of the DØ calorimeter do not have the same noise level (e.g. the CH layers have a higher noise level than the EM and FH layers). Cells having an energy above $4\sigma_{\text{ped}}$ are believed to originate from signal[21] and therefore they are always kept. These are the seed cells around which calorimeter clusters are build. Cells with $2.5\sigma_{\text{ped}} < E_{\text{cell}} < 4\sigma_{\text{ped}}$ are only considered if they are 3D-neighbor to a seed cell[22]. The T42 algorithm leads to a better rejection of noise cells, and hence to better energy resolution. In the rest of this thesis, "zero-suppression" will refer to the combined effect of the hardware and T42 thresholds.

In addition to the energy, each calorimeter cell has a direction in space defined by the positions of the cell's geometric center and the reconstructed hard-scatter PV ($\vec{d}_{\text{cell}} = \vec{r}_{\text{cell}} - \vec{r}_{\text{PV}}$, $\hat{d}_{\text{cell}} = \vec{d}_{\text{cell}}/|\vec{d}_{\text{cell}}|$). Using these two variables, a momentum vector $\vec{p}_{\text{cell}}$ is computed for each cell assuming it is a massless object ($\vec{p}_{\text{cell}} = |E_{\text{cell}}|\hat{d}_{\text{cell}}$).

3.4 Leptons and Photons

In this analysis, isolated electrons and muons are used to identify decays of W bosons, while photons play a central role in the calibration of the energy of the jets. Electrons and photons are initially reconstructed from electromagnetic clusters in the calorimeter consistent in shape with an electromagnetic shower [38]. To discriminate electrons from photons, information from the tracking detectors can be used by combining an isolated track to the EM cluster. The central preshower detector (CPS) may also be used to confirm the presence of electrons and photons in the central calorimeter region.

[21] Only 0.003% of the noisy cells, i.e. less than 1.5 noisy cells per event in average, survive a positive $4\sigma_{\text{ped}}$ cut.

[22] The algorithm does not considers the first layer of the EM calorimeter nor the layers 8, 9 and 10 of the intercryostat region, so all their cells with positive energy are kept in the event and not used as neighbors.

3.4.1 Electromagnetic Clusters

The standard electromagnetic clustering algorithm at DØ is the simple cone algorithm based on towers. An EM tower is defined by adding the energy measured by the calorimeter in all four EM layers plus the first hadronic (FH1) layer. Electromagnetic clusters are formed from seed EM towers which have $E_T > 500$ MeV. Starting with the highest-E_T seed EM tower, neighboring towers are added if they have $E_T > 50$ MeV and if they are within $\Delta R < 0.3$ of the seed tower in the central region of the detector, or within a cone radius of 10 cm in the third layer of the EM calorimeter in the end caps. Such preclusters are used as starting points for the final clusters if their energy exceeds 1 GeV. Any EM tower within $\Delta R < 0.4$ is added to the cluster, and the centroid of the final cluster is computed by weighting cell positions with the logarithm of the cell energies in the EM3 layer of the calorimeter (EM3 is the most finely segmented layer in $\Delta\eta \times \Delta\phi$). The shower centroid together with the primary vertex is used to calculate the direction of the particle four-momentum.

Additional requirements are imposed to the electromagnetic clusters to define them as isolated electron/photon candidates. The final cluster must have $E_T > 1.5$ GeV, and should be relatively narrow, which is enforced by requiring that more than 40% of the cluster energy be contained in the central most energetic tower. Since a true EM shower deposits most of its energy in the first few layers (EM layers) of the calorimeter, the cluster must have a large *EM fraction*, $f_{\rm EM} = E_{\rm EM}/E_{tot} > 0.9$, where $E_{\rm EM}$ is the cluster energy deposited in the EM layers and E_{tot} is the total energy of the cluster. Electron/photon candidates are required to be isolated in the calorimeter by imposing a higher cut on the *isolation fraction*

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

where $E_{\rm total}(R < 0.4)$ is the total energy in the towers within a cone (in the η-ϕ space) of radius $R = 0.4$ around the direction of the cluster, summed over the entire depths of the calorimeter except the CH layers, and $E_{\rm EM}(R < 0.2)$ is the energy in the towers in a cone of $\Delta R = 0.2$ summed over the EM layers only. The cut is $f_{\rm iso} < 0.2$.

Electrons and high-p_T photons produce EM showers in the preshower detectors. The 3D clusters reconstructed in the preshower detectors are matched to the EM clusters passing the above cuts by requiring them to be in a window $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$ around the EM cluster. If a preshower 3D cluster is matched, its position together with the primary vertex is used to recalculate the direction of the particle momentum. The distinction between electrons and photons is made at this point by finding matching tracks in the central tracking system (see Sect. 3.4.2). If one or more tracks are found, the candidate is considered as an electron and assigned an ID of ± 11 ($+11$ for electrons and -11 for positrons). The momentum direction is recalculated using the direction of the matched track. In the absence of the associated track, the candidate becomes a photon and assigned an ID of 10.

3.4.2 Electrons

The criteria used in forming electron/photon candidates (Sect. 3.4.1) are purposely quite loose. So, after the reconstruction of electrons, there remains a considerable amount of background, which contaminates the reconstructed sample. Thus, additional constraints must be applied to increase the discriminating power of the electron selection over the backgrounds. The main sources of instrumental electron backgrounds ("fake" electrons) are believed to be *i)* π^0 showers which overlap a track from a nearby charged particle, *ii)* photons which convert to e^+e^- pairs, and to a lesser extent *iii)* charged pions that undergo charge exchange in the detector material and *iv)* fluctuations of QCD shower shapes. The following is a description of the standard quantities employed for the electron identification, in addition to the previously introduced electromagnetic fraction ($f_{\rm EM}$) and isolation fraction ($f_{\rm iso}$) variables.

- *H-matrix* χ^2
  EM showers should develop with a logitudinal and lateral shape comparable to that of an electron. Each cluster is assigned a 7×7 covariance matrix based on 7 parameters which compare the values of the energy deposited in each layer of the EM calorimeter and the total shower energy with the average distributions obtained in Monte Carlo. The H matrix is the inverse of this covariance matrix. To measure the consistency in the shower shape between a certain cluster and an electromagnetic cluster, a χ^2 is computed from the H matrix. A shower that closely resembles an electromagnetic shower will have a low H-matrix χ^2 value, of O(15), while pions for example produce showers with an associated H-matrix χ^2 of O(10^3).

- *Track match* $Prob(\chi^2_{spacial})$
  The track matching procedure consists in defining a "road", 0.05×0.05 in $\Delta\eta \times \Delta\phi$, between the calorimeter cluster and the primary vertex positions, and search for a track with $p_T > 1.5$ GeV/c within this road. Each found track is extrapolated into the EM3 layer of the calorimeter and a $Prob(\chi^2_{spatial})$ is computed based on the spatial distance between the cluster at the EM3 floor of the calorimeter and the track extrapolation. The track with the highest $Prob(\chi^2_{spatial})$ is considered as track matched to the EM object.

- *Electromagnetic likelihood* ($\mathcal{L}_{EM}$)
  The definition of the electromagnetic likelihood [39–41] depends on seven variables defined by combining information from the calorimeter and the tracking system that offer discrimination between jets with a large EM content and real electrons or photons:

  - *Ntrks* = number of tracks in $\Delta R < 0.4$ (tracks are required to have $p_T > 0.5$ GeV/c, $|dca| < 1$ cm and $\Delta z(z_{\rm track}, {\rm PV}) < 2$ cm);

– *SumPt* = scalar sum of track p_T in $0.05 < \Delta R < 0.4$ with same track quality requirements as for the *Ntrks* variable;

– $E_T/p_T = E_T^{\text{EM cluster}}/p_T^{\text{track}}$;

– *dca*: measures the shortest distance of the selected track to the line parallel to the z-axis which passes through the primary vertex position;

– f_{EM}: (defined in Sect. 3.4.1);

– *H-matrix* χ^2: (defined above);

– *Track match* $Prob(\chi^2_{spacial})$: (defined above).

Based on Monte Carlo distributions for each variable from the signal and background samples, probability distributions for each variable are produced. These distributions are used to assign a probability for a given EM object to be signal, $P_{\text{sig}}(\mathbf{x})$, or background, $P_{\text{bkg}}(\mathbf{x})$, where $\mathbf{x}$ is a vector of likelihood variables. (Each likelihood variable for the object is given a probability to be signal or background from the probability distributions, and assuming no correlations, these individual probabilities are multipled together to give an overall probability: $P_{\text{sig/bkg}}(\mathbf{x}) = \prod_i P_{\text{sig/bkg}}(x_i)$.) The likelihood discriminant is computed as

$$\mathcal{L}_{EM}(\mathbf{x}) = \frac{P_{\text{sig}}(\mathbf{x})}{P_{\text{sig}}(\mathbf{x}) + P_{\text{bkg}}(\mathbf{x})}. \tag{3.4}$$

For electrons $\mathcal{L}_{EM}(\mathbf{x})$ tends toward one, whereas for background objects it tends toward zero.

Based on the previously defined variables, the electron definitions used in this analysis are now given.

Loose isolated electron
A loose isolated electron is required to have ID = 10, ±11, $f_{\text{EM}} > 0.9$ and H-matrix $\chi^2_{Cal7} < 50$. The isolation condition is given by $f_{\text{iso}} < 0.15$. The EM cluster must be matched to a track from the tracking detectors with a $Prob(\chi^2_{spacial}) > 0.001$. The track p_T must be greater than 5 GeV/c and its distance of closest approach in z to the primary vertex be less than 1 cm.

Tight isolated electron
A tight isolated electron must pass all the loose isolated electron requirements and have a value of the seven-variable EM-likelihood $\mathcal{L}_{EM} > 0.85$.

Ultraloose electron
The ultraloose electron definition is the same as the loose isolated electron definition, except that no track match is required. Ultraloose electrons are used in the analysis to define a data-driven sample to model the multijet background (see Sect. 6.1.1).

Table 3.1 shows the efficiencies and fake rates for the three electron ID definitions given above. They were measured in $Z \to e^- e^+$ and QCD data samples respectively, for reconstructed electrons in CC ($|\eta_{\rm det}^{\rm electron}| < 1.1$) and with $p_T^{\rm electron} \sim 40$ GeV/c.

Electron ID Efficiencies and Fake Rates

Electron definition	Run IIa		Run IIb	
	Efficiency	Fake rate	Efficiency	Fake rate
ultraloose	~ 0.993	~ 0.56	~ 0.99	~ 0.6
loose isolated	0.860 ± 0.002	0.0250 ± 0.0008	0.844 ± 0.002	0.0210 ± 0.0005
tight isolated	0.764 ± 0.003	0.001 ± 0.00014	0.735 ± 0.003	0.0013 ± 0.0002

Table 3.1 : Average electron efficiencies and fake rates for the three electron ID's used in the analysis, corresponding to electrons in the CC region and with a transverse momentum around 40 GeV/c.

3.4.3 Photons

Photons are used in this thesis to calibrate the energy of the jets by using γ+jet events (see Sect. 4). Since photons have very similar shower shapes in the calorimeter as electrons, their reconstruction is also based on the formation of electromagnetic clusters that define electron/photon candidates. In Sect. 3.4.2 the requirements applied to the EM clusters to identify them as electrons were given. In the present section, the selection criteria that are applied to the EM clusters to identify them as photons are enumerated.

- The electromagnetic cluster must be isolated with or without an associated track (ID = 10 or $\pm$11)[23].
- It must be reconstructed either in the central region ($|\eta_{\rm det}| < 1$) or in the end cap regions ($1.5 < |\eta_{\rm det}| < 2.5$), and must be in the fiducial regions of the detector (objects near the poorly understood intermodule boundaries are excluded).
- The electromagnetic energy of the cluster must represent more than 96% of the cluster's total energy: $f_{\rm EM} > 0.96$ (see Sect. 3.4.1).
- The cluster has to be isolated in the η-ϕ space: $f_{\rm iso} < 0.07$ (see Sect. 3.4.1).
- The cluster should not be matched to a track from the tracking detector with a probability higher than 0.001: $Prob(\chi^2_{spatial}) < 0.001$ (see Sect. 3.4.2).
- The energy weighted width squared of the cluster in $r \times \phi$ in the third layer of the EM calorimeter must be less than 14 cm^2.

[23] Although ID $\pm$11 are supposed to be electrons, they are included because there is a significant fraction of photons with a track close enough to be called electrons.

- The scalar sum of the transverse momenta of all tracks in a hollow cone with $\Delta R \in [0.05, 0.7]$ around the photon candidate must be less than 1 GeV/c. Tracks are considered if their transverse momentum exceeds 0.4 GeV/c, and if their distance of closest approach in z to the vertex is less than 1 cm.

- Two variables constructed from energy depositions in the preshower detector are used to discriminate against wide clusters and photons from neutral pions: the squared difference between the preshower position in ϕ and the position of the cluster in the third layer of the EM calorimeter, weighted by the energy depositions (the energy depositions squared) in strips of the preshower detector, must be less than 0.003 (0.0015).

This set of criteria define the "tight" photon selection. For the purpose of background studies, namely the contamination from dijet events where one of the jets is misidentified as a photon, two additional sets with less stringent criteria are considered. "Loose" photon selection is the same as the tight one, but no cut on the scalar sum of transverse momenta of associated tracks is applied, as well as no information from the preshower detector is used. "Medium" selection is also based on the tight one, but the cut on the scalar sum of transverse momenta of associated tracks is released to 2 GeV/c and outer radius of the hollow cone set to 0.4.

3.4.4 Muons

The muon system with its toroid magnet provides unambiguous muon identification with momentum measurement, covering $|\eta_{\text{det}}| < 2$. Muon reconstruction in DØ [42–44] starts with the formation of track segments, one in each of the three layers (A, B and C) of the muon detector, which can then be combined to form a complete track in the muon system. A muon identified using only information from the muon detector is called a "local muon". Information from the central tracking system (SMT and CFT) can also be used to match a local muon with a central track. The central tracking is highly efficient at finding tracks in the whole angular acceptance of the muon detector, and has a much better momentum resolution. In some cases, minimal interacting particle (MIP) signatures in the calorimeter are also used to help in muon identification.

There are two basic parameters that characterize reconstructed muon candidates, the *type* and the *quality*. The *type* of a muon is represented by the parameter $nseg$, whose possible values are 0, ± 1, ± 2 and ± 3. Positive (and 0) values of $nseg$ indicate that the local muon matches to a track in the central tracking system, while negative values indicate matching failure[24]. The absolute values $|nseg| = 1$, 2 or 3 indicate respectively that the local muon is made up of only an A-layer segment, only a BC-layer segment, or

[24]The matching of central tracks to muon segments, involves extrapolating the central tracks out to the muon system, then matching.

both an A- and a BC-layer segment. $nseg = 0$ muons do not have any segment. In the case where a central track could be matched, then the default kinematic parameters of the muon (4-vector, charge, η, ϕ, etc.) are those of the central track. From the negative values of $nseg$, only $nseg = -3$ is widely used. Contrary to $nseg = 3$, were a local muon track is build out from the A- and BC-segments, for $nseg = -3$ the fit of the A- and BC-segments into a local muon track may or may not be successful. If the fit is successful the momentum of the fitted local muon track becomes the momentum of the muon. Otherwise the momentum is estimated from the curvature between A- and BC-segments.

The muon *quality* can be "Loose", "Medium" or "Tight". The quality definitions are given below.

- Tight
  Only $|nseg| = 3$ muons can be Tight. A muon is Tight if it has: at least two A layer wire hits, at least one A layer scintillator hit, at least three BC layer wire hits, at least one BC scintillator hit and a converged local fit ($\chi^2_{loc} > 0$).

- $|nseg| = 3$ Medium/Loose
  An $|nseg| = 3$ muon is Medium if it has at least two A layer wire hits, at least one A layer scintillator hit, and at least two BC layer wire hits. In p20 (p17) at least one BC scintillator hit is also required, except for central muons (with less than four BC wire hits). An $|nseg| = 3$ Loose muon is defined as a Medium muon, but allowing one of the above tests to fail, with the A wire and scintillator requirement treated as one test and requiring always at least one scintillator.

- $nseg = +2$ Loose/Medium
  An $nseg = 2$ muon is Loose if it has: at least one BC layer scintillator hit and at least two BC layer wire hits. An $nseg = 2$ muon is defined as Medium if it fulfills the Loose requirements and if it is located in the bottom part of the detector (octant 5 and 6 with $|\eta_{\rm det}| < 1.6$).

- $nseg = +1$ Loose/Medium
  An $nseg = 1$ muon is Loose if it has: at least one scintillator hit and at least two A layer wire hits. An $nseg = 1$ muon is defined as Medium if it fulfills the Loose requirements and if it is located in the bottom part of the detector (octant 5 and 6 with $|\eta_{\rm det}| < 1.6$).

To control the purity of muons matched to a central track, three qualities of track have been defined:

- *track_loose*: A track is loose if its distance of closest approach in the x-y plane with respect to the PV of the event is smaller than 0.2 cm ($|dca_{xy}| < 0.2$ cm). If the track has an SMT hit, the requirement is tightened to $|dca_{xy}| < 0.02$ cm[25].

[25]The typical resolutions observed in the data for muons from Z decays are 500 μm for tracks

- *track_ medium*: A track is medium if it fulfills the loose requirements and if the central track fit has $\chi^2/ndf < 4$.

- *track_ tight*: A track is tight if it fulfills the medium requirements and if it has at least one SMT hit.

The dca_{xy} cuts in the above tracking criteria help in the suppression of cosmic ray muons. For a more efficient veto on cosmic muons, additional cuts on the time interval between the beam crossing time (t_0) and the muon system scintillator hit time, are typically applied.

Another important concept used in muon identification is *muon isolation*. Muon isolation variables are devised to separate the signal muons originated from the decay of W bosons from the heavy flavor background muons (B-habron $\rightarrow \mu + X$). Because muons from heavy flavor decays tend to be embedded inside a jet, the isolation variables are either defined in terms of the tracks near the muon track or the calorimeter energy surrounding the muon $\vec{p}$. Five variables are defined as shown below. Different cuts on these variables are used to define 10 different muon isolation criteria [42].

- *TrackHalo* $= |\sum_{\text{tracks}} p_T^{\text{track}}|$ in $\Delta R < 0.5$ around the muon track; the sum excludes the muon track.

- *ScaledTrackHalo* $= |\sum_{\text{tracks}} p_T^{\text{track}}/p_T^{\mu}|$ in $\Delta R < 0.5$ around the muon track; the sum excludes the muon track.

- *CalorimeterHalo* $= |\sum_{\text{cells}} E_T^{\text{cell}}|$ in $0.1 < \Delta R$(cell, muon cal-track) < 0.4.

- *ScaledCalorimeterHalo* $= |\sum_{\text{cells}} E_T^{\text{cell}}/p_T^{\mu}|$ in $0.1 < \Delta R$(cell, muon cal-track) < 0.4.

- $\Delta R(\mu, jet)$ = distance to the closest jet in the η-ϕ space.

Based on the previously defined variables, the muon definitions used in this analysis are now given.

Loose isolated muon
The muon quality must be Medium with type $|nseg| = 3$. The track reconstructed in the muon system must match a central *track_ medium* track, which has be close to the primary vertex (Δz(track, PV) < 1 cm). The muon must be isolated from a jet ($\Delta R(\mu, jet) > 0.5$) and it is required to pass the loose cosmic ray rejection timing requirements ($|\Delta t$(A-layer scint, $t_0)| < 10$ ns and $|\Delta t$(BC-layer scints, $t_0)| <$ 10 ns).

Tight isolated muon
Tight isolated muons are loose isolated muons with the additional requirement to

without SMT hits and 20 μm for tracks with SMT hits.

pass the "TopScaledLoose" isolation criterion, which means *ScaledTrackHalo* < 0.2 and *ScaledCalorimeterHalo* < 0.2.

The average muon reconstruction efficiency for Medium quality ($nseg = 1, 2, 3$) is $\sim 82\%$, compared to $\sim 92\%$ for the Loose quality [44]. The efficiency for only Medium $nseg = 3$ muons is $\sim 73\%$. There is little dependence of the efficiency with the instantaneous luminosity and momentum ($< 2\%$), but a strong dependence with $\eta_{\rm det}$ ($\sim 70\%$ for $|\eta_{\rm det}| < 1.1$ and $\sim 85\%$ for $1.4 < |\eta_{\rm det}| < 1.8$). The efficiencies shown above are for matched muons with a central *track_ medium* track.

3.5 Jets

Jets resulting from the hard interaction usually involve a large number of particles, which in turn deposit energy in numerous calorimeter cells. Jets are reconstructed from calorimeter clusters (see Sect. 3.3) using the Run II mid-point cone algorithm [45] (which belongs to the class of fixed-cone algorithms). The jet centroid is defined as $(y_{\rm jet},\ \phi_{\rm jet})$, and objects are clustered if their distance relative to the jet axis, $\Delta R = \sqrt{(y - y_{\rm jet})^2 + (\phi - \phi_{\rm jet})^2} < R_{\rm cone}$, where $R_{\rm cone}$ is the cone radius. Jet energy scale corrections and uncertainties have been determined for $R_{\rm cone} = 0.5$ and 0.7. In the single top analysis, jets reconstructed with a radius $R_{\rm cone} = 0.5$ are used.

The jet reconstruction procedure involves a number of steps. First, pseudo-projective calorimeter towers ($\Delta\eta \times \Delta\phi = 0.1 \times 0.1$) are reconstructed by adding the four-momentum of the calorimeter cells above threshold they contain, treating each cell as massless. The momentum of each cell is defined as explained at the end of Sect. 3.3. As a result, calorimeter towers are massive. In a second step, the calorimeter towers with $p_T \geq 1$ GeV/c are used as seeds to find pre-clusters, which are formed by adding neighboring towers within $\Delta R < 0.3$ of the seed towers. The pre-clustering step is used to reduce the number of seeds passed to the main algorithm, in order to keep the analysis computationally feasible. A cone of radius $R_{\rm cone}$ is formed around each pre-cluster, centered at its centroid, and a new jet center is computed using the E-scheme:

$$\mathrm{p}^{\rm jet} = \left(E^{\rm jet}, \vec{p}^{\rm jet}\right) = \sum_i \left(E_i, \vec{p}_i\right), \tag{3.5}$$

$$p_T^{\rm jet} = \sqrt{(p_x^{\rm jet})^2 + (p_y^{\rm jet})^2}, \tag{3.6}$$

$$y^{\rm jet} = \frac{1}{2}\ln\left(\frac{E^{\rm jet} + p_z^{\rm jet}}{E^{\rm jet} - p_z^{\rm jet}}\right), \tag{3.7}$$

$$\phi^{\rm jet} = \tan^{-1}\left(\frac{p_y^{\rm jet}}{p_x^{\rm jet}}\right), \tag{3.8}$$

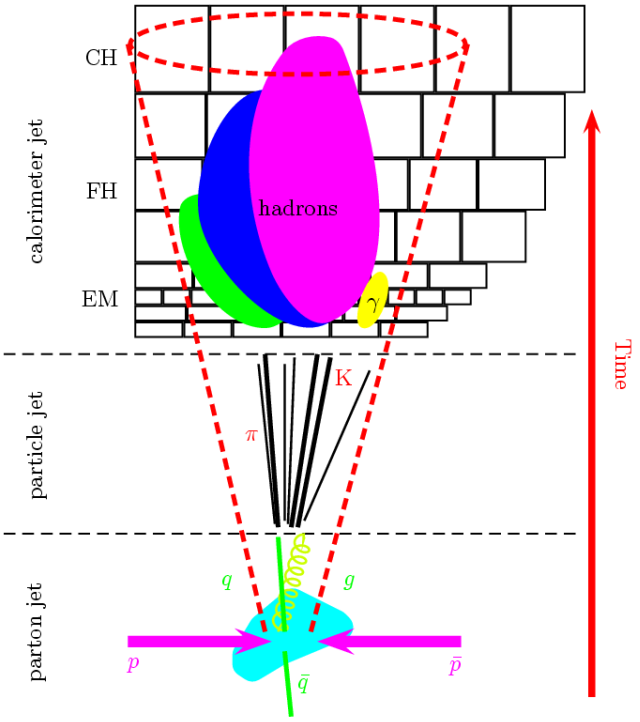

Figure 3.1 : Diagram illustrating a reconstructed jet at the parton, particle and detector (calorimeter) levels.

where the sums are over all towers contained in the cone. This procedure is repeated iteratively for each of the seeds, always using the E-scheme, until the jet center is stable. Stable solutions are called proto-jets. The sensitivity to soft radiation is reduced by the addition of mid-points between pairs of proto-jets and repeating the iterative procedure for these mid-point seeds. The last step of the algorithm involves splitting and merging to treat overlapping proto-jets, i.e. proto-jets separated by a distance $\Delta R < 2R_{\text{cone}}$. Overlapping proto-jets are merged into a single jet if more than 50% of the p_T of the lower-energy jet is contained in the overlap region. Otherwise, the energy of each cell in the overlap region is assigned to the nearest jet. Finally, the jet four-momentum is recomputed using the E-scheme and jets with $p_T < 6$ GeV/c are discarded.

The jet algorithm described above can also be applied to stable particles in MC events. Stable particles are defined as those having a lifetime long enough to not decay within the DØ detector volume. All stable particles produced in the $p\bar{p}$ interaction are considered, including not only the ones from the hard-scatter process, but also from the interactions between the spectator partons of the incident hadrons. The exception are muons and neutrinos, which are not included. Jets clustered from the list of considered stable particles (particle jets) are used to define the particle level jet energy (see Fig. 3.1).

The selection criteria used to identify jets are enumerated bellow.

- The fraction of energy deposited in the electromagnetic part of the calorimeter (EMF) must be greater than 0.05 and less than 0.95. Jets in the forward region ($|\eta_{\rm det}^{\rm jet}| > 2.5$) must satisfy EMF > 0.04.

- The fraction of energy deposited in the coarse hadronic part of the calorimeter (CHF) must be less than 0.44 for jets in $|\eta_{\rm det}| < 0.8$, less than 0.46 for jets in the endcap region $1.5 < |\eta_{\rm det}| < 2.5$, and less than 0.4 for all other jets. Jets in the region $0.85 < |\eta_{\rm det}| < 1.25$ are allowed to have CHF < 0.6 if at the same time the number of cells that contain 90% of the jet energy is less than 20. This cut is aimed at removing jets dominated by noise originating in the coarse hadronic calorimeter.

- The jet must be confirmed by level 1 trigger information. This cut is defined by the ratio

$$L1_{\rm ratio} = \frac{p_T^{\text{from L1 readout}}}{p_T^{\text{from precision readout}}}, \tag{3.9}$$

  where $p_T^{\text{from precision readout}}$ is the vector sum from a jet's tower p_Ts from the precision readout, excluding the coarse hadronic layers, and $p_T^{\text{from L1 readout}}$ is the scalar sum of p_Ts in a cone of radius $\Delta R = 0.5$ from the highest-energy L1 trigger towers in the event. A jet must satisfy

  - $L1_{\rm ratio} > 0.50$, or
  - $L1_{\rm ratio} > 0.35$ and $p_T < 15$ GeV/c and $1.4 < |\eta_{\rm det}|$ (end cap), or
  - $L1_{\rm ratio} > 0.10$ and $p_T < 15$ GeV/c and $3.0 < |\eta_{\rm det}|$ (forward), or
  - $L1_{\rm ratio} > 0.20$ and $p_T \geq 15$ GeV/c and $3.0 < |\eta_{\rm det}|$ (forward).

The purpose of this Level 1 confirmation is to remove jets dominated by noise in the precision readout electronics.

Once jets are identified, certified jet energy scale corrections [46] are applied to them in the data and MC to convert jet energies from reconstructed to particle-level energies. (In chapter 4 the determination of the jet energy scale corrections is explained in detail.) Jets containing a "medium-nseg3" muon within $\Delta R(\mu$,jet-axis)< 0.5 are considered to originate from a semileptonic b-quark decay and are corrected for the momentum carried away by the muon and the neutrino. For this correction, it is assumed that the neutrino carries the same momentum as the muon.

3.5.1 Jets from b-quarks

A jet originated from the decay of a B-hadron is called a b-jet. B-hadrons are formed from the hadronization of (final state) b-quarks. Identifying jets coming from b-quarks is a powerful tool to separate events with final state b-quarks from the multijets background in hadron collisions. With a lifetime of $\sim 10^{-12}$ s, B-hadrons can travel several millimeters before decaying into other mesons and baryons. Thus, the presence of a b-quark in the final state can be inferred from the reconstruction of a SV from charged particle tracks (see Fig. 3.2), as well as from the fact that the tracks in the resulting jet should have a large impact parameter significance (dca/σ_{dca}) with respect to the PV. Another characteristic that can be used to identify a b-jet is the identification of a muon inside the jet, as B-hadrons decay 10% of the time into $\mu\nu_\mu + X$.

In DØ, a handful of algorithms were developed to identify b-jets, based on the above characteristics. In this analysis, the B-ID Group's recommendations are followed and the latest of these algorithms used, the "neural network (NN) b-jet tagger" [47], which combines variables calculated by the following three previously developed tagging algorithms:

- Secondary Vertex Tagging (SVT): It uses tracks which are significantly displaced from the PV to reconstruct SVs. A jet is considered tagged if a SV is located within $\Delta R(\eta, \phi) < 0.5$ of the jet. In Fig. 3.2, the jet pointing up-right, which contains the SV within the jet cone, would be tagged by the algorithm.

- Jet Lifetime Impact Parameter (JLIP): It combines the impact parameter information from all the tracks identified in a jet into one variable, the jet lifetime probability (JLIP Prob). JLIP Prob is the probability that all tracks originate from the PV. The closer to zero the more likely that the jet originated from a b-quark. The operating points for the tagger range from 0.002 to 0.04.

- Counting Signed Impact Parameter (CSIP): It counts the number of tracks identified in a jet (tracks are matched to jets if they have a $\Delta R(\eta, \phi) < 0.5$) which have a large impact parameter (IP) significance with respect to the PV. Events must have at least 3 tracks with an IP significance greater than 2 or 2 tracks with an IP significance greater than 3 to be considered tagged.

Each of the taggers listed above calculates several variables containing valuable information on how likely it is that the jet originated from a b-quark. Although each of these variables provides powerful discrimination between u, d, s and b-jets, a combination of them by means of a multivariate technique could provide much more discrimination. Actually, compared to the individual b-taggers, the NN tagger reduces (increases) the fake rate (signal efficiency) by up to 80% (50%) for a fixed signal efficiency (fake rate).

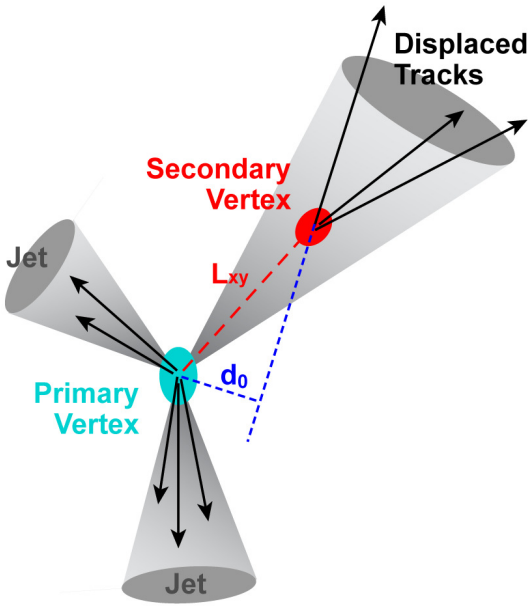

Figure 3.2 : Diagram illustrating an event with three reconstructed jets, one of them containing a SV within the jet cone. The parameter L_{xy} represents the decay length.

The NN tagger uses the following variables, ranked in order of separation power, to discriminate b-jets from other jets:

- SVT_{SL} DLS: the decay length significance of the SV with respect to the PV (when multiple SVs are found in a jet, the one with the highest DLS is selected);
- CSIP Comb: a weighted combination of the impact parameter significances of all tracks in the jet as determined by the CSIP tagger;
- JLIP Prob: the probability that the jet originated from the PV;
- SVT_{SL} χ^2/ndf: the χ^2/ndf of a fit constraining all tracks in the jet to the selected SV;
- SVT_L N_{Tracks}: the number of tracks used to reconstruct the selected SV;
- SVT_{SL} Mass: the p_T corrected mass of the selected SV[26];
- SVT_{SL} Num: the number of SVs reconstructed within a cone of $\Delta R(\eta, \phi) < 0.5$ around the jet.

[26]This is calculated from the combined rest mass of the tracks assuming they are all pions, with the mass corrected for neutral particles.

Before applying the NN tagger, some quality conditions are imposed on the jets to be tagged. For example, the calorimeter jet must be a 'good' jet matched to a track jet within a cone of $\Delta R(\eta, \phi) < 0.5$ and the track jet must have at least two 'good' quality tracks in order to be able to construct a SV within it. This quality selection on jets is called *taggability*. The b-tagging algorithm is then applied only on *taggable* jets. For data samples, the algorithm is applied directly, while on simulated samples it is applied by means of parametrizations called tag rate functions (TRFs). The TRFs are derived on data in order to realistically estimate the b-tag efficiency in the simulation (the simulated samples have a higher b-tag efficiency than data). A more detailed explanation about taggability and about how the b-tagging algorithm is applied on simulated events is given in Sect. 6.3.

To optimize the performance of the NN, the most powerful variable from each of the constituent taggers is used to select the input jets to the NN. The optimal cuts were selected trying to ensure the maximal number of b-jets whilst minimizing the number of fakes. Jets are input into the NN if the have either SVT_{SL} DLS > 2.5 or CSIP Comb > 8 or JLIP Prob < 0.02 (failure of all three conditions results in a NN output of 0).

The NN tagger retrieves an output value between 0 and 1; the higher the NN output the more probable it is the jet comes from a b-quark. Twelve different operating points have been defined (outlined in Table 3.2), each characterized by a given lower cut on the NN output.

Name	MegaTight	UltraTight	VeryTight	Tight	Medium	oldLoose
NN output	> 0.925	> 0.9	> 0.85	> 0.775	> 0.65	> 0.5
Name	Loose	L2	L3	L4	L5	L6
NN output	> 0.45	> 0.325	> 0.25	> 0.2	> 0.15	> 0.1

Table 3.2 : The NN tagger's operating points.

In this analysis, the NN's Tight and oldLoose operating points are used. The data samples are divided into single-tagged and double-tagged analysis channels. In the single-tagged channels, the Tight operating point is used to identify the b-jet, and a veto is applied on a second tag using the oldLoose operating point. In the double-tagged channels, the oldLoose operating point is used to identify b-jets. Notice that these subsamples are orthogonal between each other. The average b-tagging efficiencies for the Tight and oldLoose operating points, measured in data for jets with $|\eta_{\text{det}}| < 2.5$, are 49% and 61% respectively. Correspondingly, the average fake rates, measured in the central calorimeter, are 0.82% and 2.5%.

A complete description of the NN tagger, its training and testing, etc, is given in Ref. [47]. Its performance in p17 and p20 data are described in detail in Refs. [48] and [49] respectively.

3.6 Missing Transverse Energy

The single top events considered in this analysis have a neutrino in its final state. Neutrinos cannot be detected in DØ; however, its presence can be inferred from momentum conservation in the transverse plane. Since the total transverse momentum of the event must be zero, the sum of the $\vec{p}_T$'s of the (undetected) neutrinos equals the negative sum of the $\vec{p}_T$'s of all other (detected) particles in the event. In practice, what is used is the so-called "missing transverse energy"[27] ($\vec{\not\!\!E}_T$) computed as

$$\not\!\!E_{x,y}^{\text{meas}} = -\sum_{i \in \text{cells}} p_{x,y}^{i}, \tag{3.10}$$

where the sum runs over the cells in the EM and FH calorimeters, ICD and massless gaps, that pass the T42 selection. Cells in the CH calorimeter are only added if they are inside a good jet.

In general, this raw quantity must be adjusted for energy scale corrections that are applied to the reconstructed event calorimeter objects (jets, electrons and photons). Also, for each muon in the event, the muon energy not detected by the calorimeter must be subtracted from the measured missing E_T:

$$\not\!\!E_{x,y}^{\text{corr}} = \not\!\!E_{x,y}^{\text{meas}} - \sum_{i \in \text{calo-objects}} (E_{x,y\ i}^{\text{corrected}} - E_{x,y\ i}^{\text{uncorrected}}) - \sum_{i \in \text{muons}} (p_{x,y\ i} - E_{x,y\ i}^{\text{calorimeter}}). \tag{3.11}$$

[27] The name "missing E_T" refers to the definition of the p_T of a cell (see Sect. 3.3).

4 Jet Energy Scale

In any $p\bar{p}$ collision, final state partons (quarks and gluons) undergo hadronization processes which result in a more or less collimated jet of particles whose core is, in general, aligned in the direction of the original parton. These particles deposit most of their energy in the calorimeter, except for loses in the dead material lying before the device as well as uninstrumented inter-cryostat regions (ICR). The calorimeter electronics is calibrated online, in periods without beam, using pulse charges injected into the preamplifiers. In addition, an offline inter-calibration makes the calorimeter as uniform as possible in ϕ and η.

This chapter gives an overview of the jet energy scale (JES) determination for the Run IIa period. A complete description is given in Ref. [46]. The JES applied to the Run IIb data used in the single top analysis presented in this thesis corresponds to a preliminary version of the Run IIb JES [50], which is the same as the Run IIa JES except that the offset subtraction was rederived because of the higher instantaneous luminosity in Run IIb compared to Run IIa, and the response scaled up or down when necessary in 0.4-wide bins in jet detector pseudorapidity $\eta_{\text{jet}}^{\text{det}}$.

The JES uses jets reconstructed out of calorimeter clusters or stable MC particles, using in both cases a fixed-cone algorithm as discussed in Sect. 3.5. In the first case the jet is referred to as a (reconstructed) calorimeter jet, while in the second case as a (reconstructed) particle jet. A particle jet is therefore a reconstructed object and is not the same as the group of particles produced in the parton shower evolution (see Sect. 4.10 for more details). The JES is derived separately for $R_{\text{cone}} = 0.5$ and $R_{\text{cone}} = 0.7$ jets. Since the single top analysis uses jets with $R_{\text{cone}} = 0.5$, the majority of the plots presented in this chapter correspond to this cone size. The calorimeter measures lower jet energies compared to that of both the parton that gave origin to the jet and the particle jet. The goal of the JES is to relate, on average, the measured jet energy to the energy of the corresponding particle level jet, using three basic correction factors as indicated by the following formulae:

$$E_{\text{jet}}^{\text{ptcl}} = \frac{E_{\text{jet}}^{\text{meas}} - E_{\text{O}}}{R_{\text{jet}} S_{\text{jet}}}, \tag{4.1}$$

where

- E_{O} represents measured energy which does not belong to the high-p_T $p\bar{p}$ interaction. It includes energy from calorimeter electronic noise and radioactive decay of the uranium absorber, energy left from previous crossings (pile-up), and energy from additional $p\bar{p}$ interactions in the actual crossing. The energy contributed by spectator partons (the underlying event) is taken as part of the hard interaction.

- R_{jet} represents the energy response of the calorimeter to particle jets, which is smaller than unity, due to the losses explained above and the non-compensating

nature of the calorimeter (lower response to hadrons than to electromagnetic particles). The response is also affected by module-to-module inhomogeneities. For a given cone size, the calorimeter jet response depends on the jet energy and the jet detector pseudorapidity.

- S_{jet} corrects for the fraction of energy deposited outside (inside) the jet cone from particles belonging (not belonging) to the particle jet, as a result of the development of showers in the calorimeter and the finite calorimeter cell size.

4.1 Definition of the Corrections Contributing to the JES

The aim of this section is to give a precise expression for each of the correction factors involved in Eq. (4.1).

The energy of a reconstructed particle jet is given by the sum of energies of all stable particles belonging to that jet:

$$E_{\text{jet}}^{\text{ptcl}} = \sum_{i \in \text{ptcljet}} E_i \,. \tag{4.2}$$

The measured energy from a calorimeter jet has contributions from particles inside and outside the particle jet, plus a small amount of offset energy:

$$E_{\text{jet}}^{\text{meas}} = \sum_{i \in \text{ptcljet}} E_i^{\text{meas}} S_i + \sum_{i \notin \text{ptcljet}} E_i^{\text{meas}} S_i + E_{\text{O}}. \tag{4.3}$$

Here E_i^{meas} is the visible energy of particle i, and S_i represents the fraction of that energy that is inside the calorimeter jet cone. Visible energy means that all detector and reconstruction effects are taken into account. From Eq. (4.3), the definition of the *true offset* energy is obtained. It is the energy that needs to be subtracted to the measured jet energy such that

$$E_{\text{jet}}^{\text{meas}} - E_{\text{O}} = \sum_{i \in \text{ptcljet}} E_i^{\text{meas}} S_i + \sum_{i \notin \text{ptcljet}} E_i^{\text{meas}} S_i. \tag{4.4}$$

Since particles from the underlying event may contribute to the particle jet, the offset correction does not subtract the underlying event energy contribution from $E_{\text{jet}}^{\text{meas}}$. The offset energy includes noise, pile-up, and additional $p\bar{p}$ interactions since the right hand side in Eq. (4.4) refers only to particles from the hard-scatter primary vertex. Notice also that E_{O} includes the extra energy that gets above the cell zero-suppression threshold, and thus becomes visible, as a result of the combined effect of the offset and jet energy added to each cell. In other words, the right hand side in Eq. (4.4) is the jet energy measured as if there would be no offset.

The *true response* is conveniently defined as the ratio of visible energy for particles from the particle jet to the energy of the particle jet:

$$R_{\text{jet}} = \frac{\sum_{i \in \text{ptcljet}} E_i^{\text{meas}}}{E_{\text{jet}}^{\text{ptcl}}}. \tag{4.5}$$

Finally, inserting all the previous definitions in Eq. (4.1), the *true showering* gets automatically defined as

$$S_{\text{jet}} = \frac{\sum_{i \in \text{ptcljet}} E_i^{\text{meas}} S_i + \sum_{i \notin \text{ptcljet}} E_i^{\text{meas}} S_i}{\sum_{i \in \text{ptcljet}} E_i^{\text{meas}}}, \tag{4.6}$$

representing a correction for the fact that the *true response* ignores that a calorimeter jet has boundaries through which energy of the particle jet can leak out and energy not from the particle jet can leak in. A showering factor greater (smaller) than 1 means a net flow of energy towards (outwards) the jet.

4.2 Estimators of the JES Subcorrections

At DØ, data-driven methods are used to estimate the offset and the response corrections, since the MC simulation of the DØ detector is not precise enough to achieve a JES with an uncertainty level of 1-2% as it is desired. The resulting offset correction and response correction estimators, $\hat{E}_{\text{O}}$ and $\hat{R}_{\text{jet}}$, suffer from a number of biases. Suitable correction factors, k_{O} and k_{R}, are derived in MC to bring them, in average, as close as possible to the true corrections. On the other hand, the showering correction is derived using MC samples, and the estimator $\hat{S}_{\text{jet}}$ is a-priori unbiased. The final corrected jet energy is then given by the following expression:

$$E_{\text{jet}}^{\text{corr}} = \frac{(E_{\text{jet}}^{\text{meas}} - \hat{E}_{\text{O}}) k_{\text{O}}}{\hat{R}_{\text{jet}} k_{\text{R}} \hat{S}_{\text{jet}}}\,, \tag{4.7}$$

where $E_{\text{jet}}^{\text{corr}}$ should be, in average, equal to $E_{\text{jet}}^{\text{ptcl}}$, if all biases are taken into account and properly corrected for.

4.3 Data and Monte Carlo JES Samples

This section gives an overview of the different data and MC samples used in the JES determination.

4.3.1 Data Samples

The following data samples are selected from the complete Run IIa data set to either determine, or validate, the JES corrections:

- **Photon+jet (γ+jet)**: This sample is collected using triggers requiring an isolated electromagnetic cluster with different p_T thresholds. It is used to measure the calorimeter jet response, inter-calibrate the calorimeter jet response in η, and determine the showering correction.

- **Dijet**: This sample is collected using triggers that require at least one jet with $p_T > 15, 25, 45, 65, 95, 125$ GeV/c. It is used, together with the γ+jet sample described above, to determine the η-dependent part of the calorimeter jet response. Dijet events produce jets with higher p_T than γ+jet events. Thus, while the later are used in the low and medium jet energy regions, the former serve for the high-E_{jet} extrapolation.

- **Minimum bias (MB)**: This sample is collected using the so-called "minimum bias trigger", which requires hits in the north and south luminosity counters, signaling the presence of a $p\bar{p}$ inelastic collision. This sample is dominated by soft interactions and is used to estimate the contribution from multiple interactions to the offset energy.

- **Zero bias (ZB)**: This sample is collected during beam crossings but without any trigger requirement, and thus represents a truly unbiased measurement (therefore its name) of the energy in the calorimeter regardless of the nature of the $p\bar{p}$ interaction. This sample is used to estimate the contribution from noise and pile-up to the offset energy.

4.3.2 Monte Carlo Samples

Given that physics analyses rely on comparing data against MC simulations, the JES must be determined separately for MC as well. The following MC samples are generated using PYTHIA 6.323 [51] with CTEQ6L1 [52] parton distribution functions:

- **Photon+jet**: This sample includes the $2 \to 2$ direct photon production processes $p\bar{p} \to q\gamma + X$ and $p\bar{p} \to g\gamma + X$.

- **Dijet**: This sample contains the inclusive $2 \to 2$ parton processes $p\bar{p} \to q\bar{q} + X$, $p\bar{p} \to qg + X$ and $p\bar{p} \to gg + X$.

- **Dijet (γ-like)**: This sample contains the same inclusive $2 \to 2$ parton processes as the dijet sample above. Several generator level cuts are applied on the generated

particles in order to enrich the sample in jets faking photons[28]. The purpose of this sample is to study, and to correct for, the dijet background contamination in the γ+jet sample in data.

To simulate the underlying event, the so-called "PYTHIA TUNE A" [54] is used, which was optimized to describe CDF Run I data. Tune A was later seen to have too much initial state radiation compared to DØ Run II data [55]. Thus, a so-called Tune DWT [56] was developed that improved the description of radiation. Tune DWT is used to compare results against Tune A and assign a systematic uncertainty when both differ. On the other hand, the offset energy components (noise, pile-up and multiple interactions) are not simulated but obtained from data: digitized signals from zero bias (ZB) data events are overlaid on the MC. The default MC production at DØ uses overlaid ZB events with the symmetric $1.5\sigma_{ped}$ zero-suppression cut (see Sect. 3.3) applied at the calorimeter cell level ("suppressed ZB overlay"). In order to study the impact of this approximation, additional γ+jet and dijet samples are generated without ZB overlay ("no ZB overlay"), as well as with ZB overlay without the $1.5\sigma_{ped}$ zero-suppression cut ("unsuppressed ZB overlay").

4.4 Offset Correction

The goal of the offset correction is to subtract from the measured energy of a reconstructed calorimeter jet, all contributions that are not related with the hard-scatter interaction. This includes energy from the following three different sources: noise, pile-up, and multiple interactions. One more source could eventually also be considered for subtraction, which is the energy resulting from the (soft) interactions involving the beam remnants (spectator partons that constituted the colliding proton and antiproton), and the initial state radiation. If this energy would be subtracted as well, then the JES would correct the jet energy up to the parton level instead to the particle level. This is the approach followed, for example, by CDF [57]. However, a consistent subtraction of the underlying event is actually not possible, since the spectator partons and radiated gluons are color-connected to the hard-scattered partons, meaning that the physics of the underlying event and the hard-scatter process are inherently related. Because of this reason, DØ's JES approach is to correct the jet energy up to the particle level only[29].

The average offset energy is estimated for each calorimeter ring in η (summed over all towers in ϕ, and denoted by $i\eta$), and parameterized as a function of the number of primary vertices (N_{PV}) and the instantaneous luminosity (L), by adding the estimated

[28]The selection cuts follow the procedure discussed in Sect. 4.1 of Ref. [53].

[29]Comparisons with theory at the particle level involve non-perturbative calculations of the hadronization process and the underlying event, in addition to the perturbative calculations of the Feynman diagrams.

contributions from noise and pile-up (NP), and multiple interactions (MI):

$$\hat{E}_{\rm O}^{\rm ring}(i\eta, N_{\rm PV}, L) = \hat{E}_{\rm NP}^{\rm ring}(i\eta, L) + \hat{E}_{\rm MI}^{\rm ring}(i\eta, N_{\rm PV}, L). \quad (4.8)$$

4.4.1 Noise and Pile-up

The main sources of calorimeter noise are the noise in its read-out electronics and the radioactivity of the uranium in the absorber plates. The uranium radiation makes the noise energy distribution non-symmetric with a larger tail on the positive side. The pedestal subtraction brings the average cell energy due to noise to zero. However, because of the particularities of the online zero-suppression and the T42 algorithm (see Sect. 3.3), a net positive contribution of offset energy due to noise is expected.

Given the short time between consecutive bunch crossings (396 ns) compared to the shaping time of the calorimeter preamplifiers ($\approx$ 15 μs), it often happens that the signal of interest appears on top of the tails of signals from previous bunch crossings. This effect is called pile-up, and it mainly depends on the instantaneous luminosity; the higher L, the higher the pile-up energy.

The average energy per $i\eta$ ring due to noise and pile-up, $\hat{E}_{\rm NP}^{\rm ring}$, is measured in a ZB data sample. To not include $p\bar{p}$ interactions in the sample, ZB events that fired the MB trigger or have any reconstructed primary vertex, are not considered. Because the size of the towers is not the same for all $i\eta$ rings (they are larger in the forward region) the energy in each $i\eta$ ring is then divided by the ring area to better compare the NP contribution across the whole $i\eta$ range. Figure 4.1 shows the average energy density per $i\eta$ ring for three different values of the instantaneous luminosity L. The bump in the region $8 < |i\eta| < 15$ corresponds to the poorly instrumented ICR, where large weight factors are applied in the conversion of ADC counts to energy measure to compensate the low gain, thus amplifying also the noise fluctuations.

4.4.2 Multiple Interactions

As mentioned in Sect. 3.2, at the Tevatron instantaneous luminosities there are in average ~ 3.5 $p\bar{p}$ interactions per bunch crossing, with one of them being may be a hard-scatter and the others just soft interactions. The average energy of these additional soft $p\bar{p}$ interactions per $i\eta$ ring, $\hat{E}_{\rm MI}^{\rm ring}$, is estimated from the average energy per ring measured in MB events, $\hat{E}_{\rm MB}^{\rm ring}$. Figure 4.2 (left) shows the measured average MB energy in four evenly separated positive $i\eta$ rings as a function of the $N_{\rm PV}$, up to $N_{\rm PV} = 10$ where the statistics runs out. As it can be observed, the energy scales linearly (within errors) with the $N_{\rm PV}$. Assuming each additional interaction gives one reconstructed PV, one concludes from Fig. 4.2 that each additional interaction contributes equally to the offset energy. Figure 4.2 (right) shows the average energy density per $i\eta$ ring for

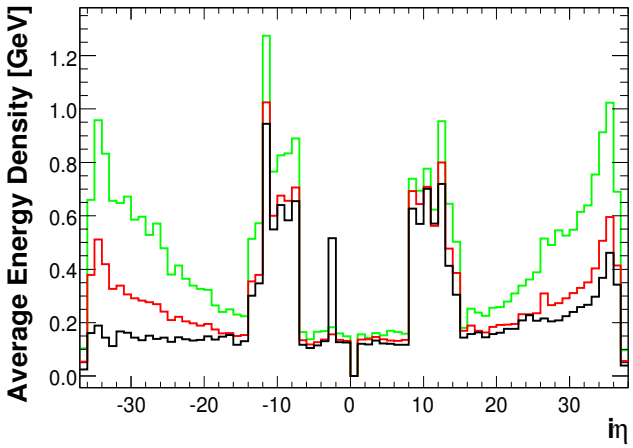

Figure 4.1 : Average energy density per $i\eta$ ring in ZB events with LM veto and zero primary vertices for different instantaneous luminosities: $L = 0.1 \times 10^{32}$ cm^{2}s^{-1} (black), $L = 0.5 \times 10^{32}$ cm^{2}s^{-1} (red) and $L = 1.2 \times 10^{32}$ cm^{2}s^{-1} (green). The empty bin at $i\eta = 0$ is just because by definition $i\eta = \{\pm 1, \pm 2, ...\}$.

MB events with different $N_{\rm PV}$. The increase of the MB energy density with $|i\eta|$ reflects the fact that the p_T spectrum of the products of a soft interaction is approximately flat in η. Thus, particles emitted at small angles with respect to the beam line have more energy than particles escaping in a perpendicular direction.

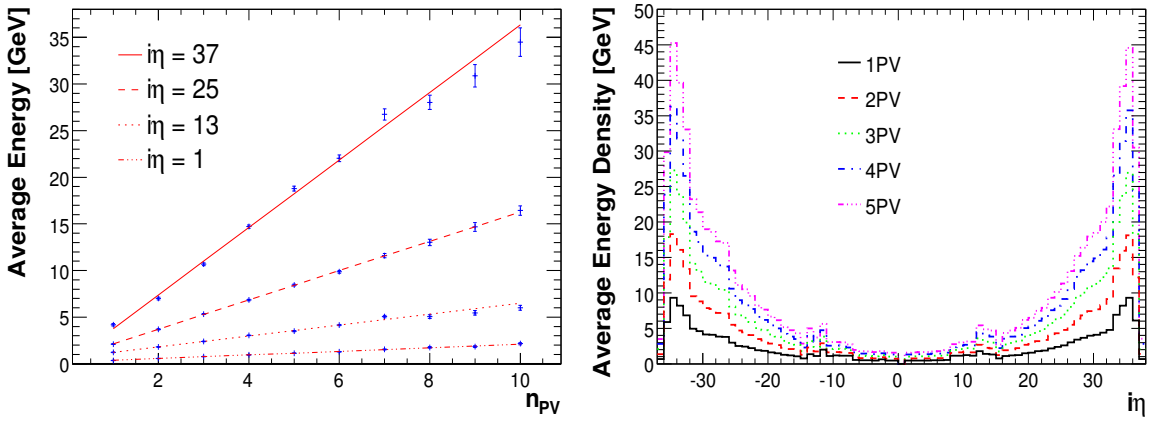

Figure 4.2 : Left: Average MB energy as a function of $N_{\rm PV}$ for different $i\eta$ rings for $L = 1.6 \times 10^{32}$ cm^{2}s^{-1}. Fits are done using linear functions and the results of the fits are used to provide the MB energy at any value of $N_{\rm PV}$. Right: Average MB energy density per $i\eta$ ring as a function of $N_{\rm PV}$ for $L = 0.2 \times 10^{32}$ cm^{2}s^{-1}.

The average contribution per $i\eta$ ring from MI to the offset energy in an event with n reconstructed PVs (i.e. one hard interaction plus $n-1$ additional interactions), is

defined as the average energy in MB events with $N_{\rm PV} = n$ minus the average energy in MB events with exactly one PV:

$$\hat{E}^{\rm ring}_{\rm MI}(i\eta, N_{\rm PV} = n, L) = \hat{E}^{\rm ring}_{\rm MB}(i\eta, N_{\rm PV} = n, L) - \hat{E}^{\rm ring}_{\rm MB}(i\eta, N_{\rm PV} = 1, L). \quad (4.9)$$

In this way, all the energy not associated with additional soft $p\bar{p}$ interactions is canceled out.

4.4.3 Total Jet Offset Energy

The total offset energy for a reconstructed calorimeter jet is estimated by summing the average offset energy per $i\eta$ ring, Eq. (4.8), over all the jet area. Denoting by $f^{\rm tower}$ the fraction of towers in a given $i\eta$ ring that are within the jet cone whose axis points in $\eta^{\rm jet}_{\rm det}$, the estimated jet offset energy is given by

$$\hat{E}_{\rm O}(\eta^{\rm jet}_{\rm det}, N_{\rm PV}, L) = \sum_{i\eta \in R_{\rm cone}} \hat{E}^{\rm ring}_{\rm O}(i\eta, N_{\rm PV}, L) f^{\rm tower}(i\eta, \eta^{\rm jet}_{\rm det}). \quad (4.10)$$

Figure 4.3 shows the estimated jet offset energy as a function of $\eta^{\rm jet}_{\rm det}$, for events with up to five reconstructed PVs. The left plot corresponds to $R_{\rm cone} = 0.7$ and the right plot to $R_{\rm cone} = 0.5$. Notice that the scales in the y-axis are different. The offset energy for $R_{\rm cone} = 0.5$ jets is roughly a factor of two smaller than for $R_{\rm cone} = 0.7$ jets, in good agreement with the naive expectation based on the ratio of areas.

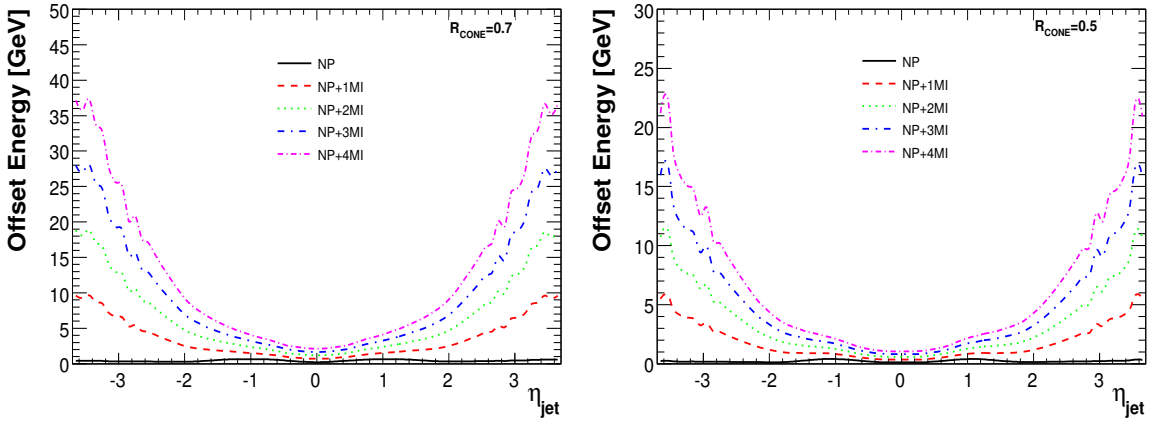

Figure 4.3 : Estimated total offset energy (in GeV) within a jet of radius $R_{\rm cone} = 0.7$ (left) and $R_{\rm cone} = 0.5$ (right) as a function of $\eta^{\rm jet}_{\rm det}$ in events with $L = 0.3 \times 10^{32}$ cm^{2}s^{-1}, which represents the average instantaneous luminosity of the MB sample. The different lines show the prediction for NP only ($N_{\rm PV} = 1$), as well as NP and MI ($N_{\rm PV} > 1$).

4.4.4 Effect of Zero-Suppression on Offset

As already mentioned in Sect. 4.1, the goal of the offset subtraction is to correct the energy of a calorimeter jet to what would be measured if no offset would be present. In absence of zero-suppression, the total offset energy estimated from ZB and MB events would provide a good correction, but when zero-suppression is applied, this estimate may differ substantially from the true offset energy inside the jet. This is because calorimeter cells inside a jet have a much higher energy content than cells in an average ZB or MB event, and therefore it is more likely for them to become above the zero-suppression threshold with the consequent increase in visible offset energy. Consider for instance a situation without MI and recall that the pedestal subtraction brings the noise average to zero. The offset energy measured in ZB events corresponds to the small fraction of large positive fluctuations of the noise that pass the zero-suppression threshold, while within a jet environment almost all positive fluctuations become visible and only in the most energetic cells within the jet core the negative fluctuations would cancel the positive ones. The net effect is that the offset energy measured in ZB events underestimates the true offset energy from noise inside a jet, and although the bias increases with the jet energy, given that it is of the order of a few GeV it becomes less relevant as the jet energy increases (see Fig. 4.4).

To correct for this bias, a γ+jet MC sample is considered and processed in three different ways (see Sect. 4.3.2): 1) without ZB overlay, 2) with (zero-)suppressed ZB overlay, and 3) with (zero-)unsuppressed ZB overlay. The bias correction is defined as a multiplicative factor:

$$k_{\mathrm{O}}^{\mathrm{ZS}} = \frac{E_{\mathrm{jet}}^{\mathrm{meas,noZB}}}{(E_{\mathrm{jet}}^{\mathrm{meas,ZB}} - \hat{E}_{\mathrm{O}})}, \tag{4.11}$$

with the average jet energy in the sample without ZB overlay (the numerator) being the reference level to correct to. Three sets of events (one in each sample) with same partonic origin, but different overlay, are selected by requiring first, in the sample with no overlay, exactly one photon and one reconstructed jet satisfying $\Delta\phi(\gamma, \mathrm{jet}) > 3$, and then, in the corresponding event with overlay, a reconstructed jet that matches within $\Delta R(\phi, \eta) < R_{\mathrm{cone}}/2$ with the reconstructed jet in the case of no ZB overlay. The denominator in Eq. (4.11) is calculated using the matched jets in the sample with ZB overlay. In the JES derived for MC, the sample with (zero-)suppressed ZB overlay is used, since this is what DØ uses in its standard MC production. In the JES derived for data, the sample with (zero-)unsuppressed ZB overlay is used, since that provides the most realistic description of the per-cell energy spectrum arising from noise, pile-up and multiple interactions.

The $k_{\mathrm{O}}^{\mathrm{ZS}}$ correction is measured in 0.4-wide bins of $|\eta_{\mathrm{det}}^{\mathrm{jet}}|$. Figure 4.4 shows the correction factor for $R_{\mathrm{cone}} = 0.5$ jets with $|\eta_{\mathrm{det}}^{\mathrm{jet}}| < 0.4$ (left) and $2.0 < |\eta_{\mathrm{det}}^{\mathrm{jet}}| < 2.4$ (right) in the unsuppressed ZB overlay case as a function of p'_T, an estimator of the *true* jet p_T (see Sect. 4.6). As expected, the correction factor increases when multiple interactions

are added (i.e. when going from $N_{\rm PV} = 1$ to $N_{\rm PV} \geq 2$); the offset energy from MI measured in MB events also underestimates the real contribution of MI inside a jet. The dependence with $N_{\rm PV}$ is more important in the forward region because the MI energy itself is higher. The central value of the correction is estimated for the average $N_{\rm PV}$ of the sample, which corresponds to 1.5 (1.8) in (un)suppressed ZB overlay, and the difference between the cases with $N_{\rm PV} = 1$ and $N_{\rm PV} \geq 2$ is used to assign a systematic uncertainty. For $R_{\rm cone} = 0.7$, jets the correction factor is $\sim 0.5\%$ larger, which is understandable from the fact that a bigger cone includes more active cells.

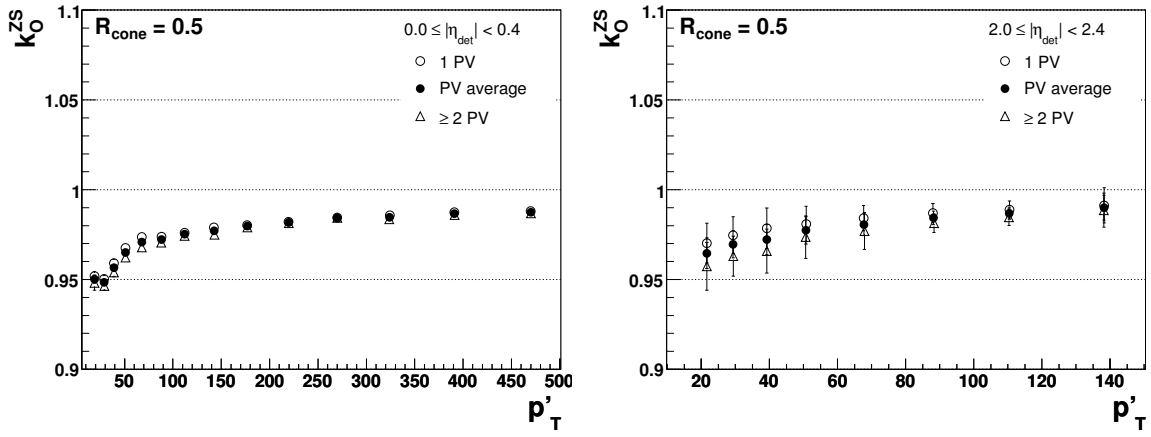

Figure 4.4 : $k_{\rm O}^{\rm ZS}$ correction factor for $R_{\rm cone} = 0.5$ jets with $|\eta_{\rm det}^{\rm jet}| < 0.4$ (left) and $2.0 < |\eta_{\rm det}^{\rm jet}| < 2.4$ (right) in the unsuppressed ZB overlay case. The different symbols correspond to different primary vertex multiplicity in the event.

4.4.5 Uncertainties

The statistical uncertainty in the measurement of the offset energy correction $\hat{E}_{\rm O}$ is negligible, as are the systematic uncertainties from the parameterization in terms of $\eta_{\rm det}^{\rm jet}$, $N_{\rm PV}$ and L. Only a conservative 5% systematic uncertainty is assigned to the $\hat{E}_{\rm O}$ correction to cover for its time dependence. The uncertainties in the offset correction are therefore completely dominated by the uncertainties in the $k_{\rm O}^{\rm ZS}$ correction factor, which come from two sources: the dependence with $N_{\rm PV}$, and the choice of ΔR matching criteria between jets in the samples with and without ZB overlay. The first uncertainty is estimated as half the difference between the value of $k_{\rm O}^{\rm ZS}$ for the cases of $N_{\rm PV} = 1$ and $N_{\rm PV} \geq 2$. The second uncertainty is estimated as half the difference between the $k_{\rm O}^{\rm ZS}$ correction factors derived with the matching criteria between jets varied to $\Delta R < R_{\rm cone}/2 \pm 0.1$. Both contributions are illustrated in Fig. 4.19, Sect. 4.8, for central jets reconstructed with $R_{\rm cone} = 0.5$ in the unsuppressed ZB overlay case.

4.5 Response Correction

The response correction is by far the largest JES subcorrection, being of order 20-35%. The measurement of the calorimeter response of a jet is based on the Missing E_T Projection Fraction (MPF) method, which will be explained in Sect. 4.5.1, and makes use of two samples, one of γ+jet and one of dijet events (defined in Sect. 4.3). This is needed because the jet response in these two samples is different, with the reason for it being the different relative composition in gluon- and quark-jets and the different response of jets with different partonic origin (see Sect. 4.7.2).

The response is factorized into two terms, the response in the central calorimeter, and a term that inter-calibrates the response in the different calorimeter regions to the central region:

$$R_{\text{jet}}(E, \eta^{\text{jet}}_{\text{det}}) = R_{\text{CC}}(E)\, F_{\eta}(E, \eta^{\text{jet}}_{\text{det}}). \tag{4.12}$$

This splitting is basically imposed by the type of samples and method used to measure the response. Dijet events are richer in forward and high energy jets than γ+jet events, and the MPF method applied on dijet events can only measure the relative response of one jet respect to the other. The JES uses γ+jet events to determine the CC-response (R_{CC}) and the two samples for the η-dependent corrections (F_{η}). The factorization of the response also facilitates the derivation of a dijet-specific JES (see Refs. [46, 58]), which is used e.g. in the inclusive jet cross section analysis [58, 59]. Top quark analyses use the γ+jet JES.

4.5.1 Missing E_T Projection Fraction Method

The MPF method aims to estimate the calorimeter jet response by looking at the event missing transverse energy ($\vec{\not{E}}_T$) in e.g. γ+(1)jet candidate events. Since the energy of a photon is well calibrated, the missing transverse energy would be mostly due to the reduced calorimeter jet response. More formally, consider a general X+jet process, in which X is referred to as the "tag object" and the jet as the "probe object". At the particle level, before any detector effects enter into play, the tag object and the probe jet, or rather the full hadronic recoil, present momentum balance in the transverse plane:

$$\vec{p}_{T\text{tag}}^{\,\text{ptcl}} + \vec{p}_{T\text{recoil}}^{\,\text{ptcl}} = 0. \tag{4.13}$$

The balance between $\vec{p}_T$s is lost when measured quantities are considered, due to the a priori different calorimeter responses to the tag object and the hadronic recoil, and a missing transverse energy appears instead:

$$\vec{p}_{T\text{tag}}^{\,\text{meas}} + \vec{p}_{T\text{recoil}}^{\,\text{meas}} + \vec{\not{E}}_T^{\,\text{meas}} = 0, \tag{4.14}$$

where

$$\vec{p}_{T\text{tag}}^{\,\text{meas}} = R_{\text{tag}}\,\vec{p}_{T\text{tag}}^{\,\text{ptcl}}, \tag{4.15}$$

$$\vec{p}_{T\text{recoil}}^{\,\text{meas}} = R_{\text{recoil}}\,\vec{p}_{T\text{recoil}}^{\,\text{ptcl}}, \tag{4.16}$$

and $\vec{\not{E}}_T^{\,\text{meas}}$ is the measured missing transverse energy in the event. Introducing the relations (4.15) and (4.16) into Eq. (4.14) and projecting it onto the direction of the tag object in the transverse plane ($\hat{n}_{T\text{tag}}$), one gets the following expression that defines the MPF method:

$$\frac{R_{\text{recoil}}}{R_{\text{tag}}} = 1 + \frac{\vec{\not{E}}_T^{\,\text{meas}} \cdot \hat{n}_{T\text{tag}}}{p_{T\text{tag}}^{\text{meas}}}. \tag{4.17}$$

The right hand side in Eq. (4.17) defines the "MPF response", R_{MPF}, while the left hand side provides the interpretation of what is being measured. Notice that the definition of the jet does not enter into this relation, and only matters for the event selection; the MPF method measures the response of the whole recoil (relative to that to the tag object) and is thus insensitive to detector showering. Furthermore, by performing a suitable event selection, as for example asking for only one reconstructed jet back-to-back with the tag object, it is possible to approximate R_{recoil} by R_{jet} in Eq. (4.17). Finally, up to this point, the tag object was not yet specified. As mentioned in Sect. 4.5, in this JES two different samples of tag and probe back-to-back objects are used. First, to determine the absolute response of a jet in the central calorimeter, a γ+(1)jet sample is selected, with the probe jet constrained to be in $|\eta_{\text{det}}^{\text{jet}}| < 0.4$. The tag object in this case is a photon with $R_{\text{tag}} = R_\gamma \simeq 1$. Second, to inter-calibrate the response of different calorimeter regions, a dijet sample is also selected with one of the jets, the tag object, constrained to $|\eta_{\text{det}}^{\text{jet}}| < 0.4$.

The MPF method suffers from a small bias related to the approximation $R_{\text{recoil}} \simeq R_{\text{jet}}$ which can indeed never be achieved. A suitable factor is defined and estimated in MC to correct for this topology bias.

The jet response highly depends on the jet energy, as higher energy jets develop more secondary showers increasing the jet electromagnetic component and so the jet response (since $e/h > 1$). The energy dependence of the jet response is obtained in terms of the E' energy estimator, defined as

$$E' = p_{T\gamma}^{\text{meas}} \cosh(\eta_{\text{jet}}) \quad \text{in } \gamma + \text{jet events}, \tag{4.18}$$

where η_{jet} is the jet pseudorapidity with respect to the reconstructed hard-scatter PV in the event. Note that the E' variable is strongly correlated with the particle level jet energy. Using the E' estimator, Eq. (4.18), instead of the measured probe jet energy, avoids biasing the response due to the poor resolution with which the jet energy is measured. The photons transverse momentum and the η position of jets are both measured much more precisely than the jet energy itself. A derived quantity, the p'_T

transverse momentum estimator, defined as

$$p'_T = E' / \cosh(\eta^{\text{jet}}_{\text{det}}), \tag{4.19}$$

is also extensively used.

4.5.2 Sample Selection

As already mentioned in Sect. 4.5, the determination of the absolute response in CC only utilizes the γ+jet sample, while the measurement of F_η involves both, the γ+jet and the dijet samples. In data, γ+jet (dijet) candidate events containing at least one EM (jet) candidate are collected using a set of single EM (jet) triggers with different $p^{\text{meas}}_{T\gamma}$ ($p^{\text{meas}}_{T\text{jet}}$) thresholds. In order to avoid trigger biases, the measured transverse momentum of the tag object (photon or jet respectively) is required to be in the high efficiency range of the fired trigger, and the photon candidate must match within $\Delta R < 0.4$ the L1 trigger EM cluster.

The offline event selection applied to each of these samples is summarized below. Events with only one hard-scatter PV (selected jets and photons must be associated to it) and at most one additional PV are accepted. The selection of dijet events closely follows the γ+jet selection, with one of the jets effectively playing the role of the photon.

Photon+jet

- Events are required to have only one photon with $p^{\text{meas}}_{T\gamma} > 7$ GeV/c, satisfying the tight photon identification (ID) criteria (see Sect. 3.4.3) and being within $|\eta^{\gamma}_{\text{det}}| < 1$.
- Events with only one reconstructed jet are kept. For the measurement of the absolute response in CC, the jet must be within $|\eta^{\text{jet}}_{\text{det}}| < 0.4$ such that it is well contained in the central calorimeter. For the measurement of the relative response, jets up to $|\eta^{\text{jet}}_{\text{det}}| < 3.6$ are accepted.
- The jet and the photon are required to be back-to-back: $\Delta\phi(\gamma, \text{jet}) > 3$.
- To eliminate $W(\to e\nu)$+jet background events, upper cuts proportional to $p^{\text{meas}}_{T\gamma}$ are applied on $\not\!\!E_T$, depending on the range of $p^{\text{meas}}_{T\gamma}$.

Dijet

- Events must have exactly two reconstructed jets, with at least one of the jets (the "tag" jet) within $|\eta^{\text{jet}}_{\text{det}}| < 0.4$, so that its core is well contained inside the central calorimeter. The other jet (the "probe" jet) can be in any $|\eta^{\text{jet}}_{\text{det}}| < 3.6$. If

both jets have $|\eta_{\rm det}^{\rm jet}| < 0.4$, both possibilities for "tag" and "probe" assignments are considered.

- The two jets are required to be back-to-back: $\Delta\phi(\text{tag}, \text{probe}) > 3$.

4.5.3 Photon Corrections: Dijet Background Contamination and Photon Energy Scale

Physics backgrounds in the γ+jet sample in data (e.g. $Z/\gamma^* \to e^+e^-$, $W(\to e\nu)$+jet and di-photons) are reduced to negligible amounts with the stringent cuts on $\not\!\!E_T$ and $\Delta\phi(\gamma, \text{jet})$ and by using the tight photon ID criteria. However, and despite these cuts, the sample suffers from the contamination of dijet instrumental background, where one of the jets fluctuates to a leading π_0 ($\pi_0 \to \gamma\gamma$) and the two photons of the π_0 decay are parallel enough to be misidentified as one photon. Although the fake rate of this process is very small for the tight photon ID, the $\mathrm{O}(10^3)$ larger dijet production rate compared to direct γ+jet production keeps this background at a considerable level. Figure 4.5 shows the estimated purity of the selected γ+jet sample as a function of E' for probe jets in the central and the forward regions. The purity is estimated using the γ+jet and dijet(γ-like) MC samples described in Sect. 4.3.2. The shape dependence of the purity with the "photon" p_T and the probe jet pseudorapidity are obtained from direct MC leading-order cross section and selection efficiency predictions for signal and background, while the overall scale is determined doing a template fit to data in the scalar sum of the transverse momenta of all tracks in a hollow cone of $0.05 < \Delta R < 0.7$ within the direction of photon candidate. As expected, the purity depends significantly on the photon ID, since the criteria are defined to reject photons from neutral pion decays with different efficiencies. The background is also reduced for higher energies, since the associated tracks transverse momentum increases, lowering the probability to pass the hollow cone cut (see Sect. 3.4.3).

The hadronic activity surrounding the "photon" diminishes the amount of $\not\!\!E_T$ in the probe jet direction leading to a positive bias in the measured MPF response. This bias is estimated in MC and is explicitly corrected for by means of a multiplicative factor on the MPF response:

$$k_{R,\eta}^{\gamma,\rm bkg} = \frac{R_{\rm MPF,\eta}^{\gamma^{\rm meas}+\rm jet}}{R_{\rm MPF,\eta}^{\rm mixture}} \equiv \frac{1}{1 + c_{\rm bkg,\eta}} \approx 1 - c_{\rm bkg,\eta}, \tag{4.20}$$

since $c_{\rm bkg,\eta} \ll 1$, and where both MPF responses from MC are calculated using the measured p_T of the photon candidate ($p_{T\gamma}^{\rm meas}$). This is relevant, because as it will be shown later in this section, $p_{T\gamma}^{\rm meas}$ is not at particle level (as needed) and a separate factor will be applied to correct for it. $R_{\rm MPF,\eta}^{\rm mixture}$ is an estimate of the measured MPF response in the selected γ+jet sample in data. It can be expressed as a linear combination of the MPF responses for pure γ+jet signal and for dijet background, weighted by the

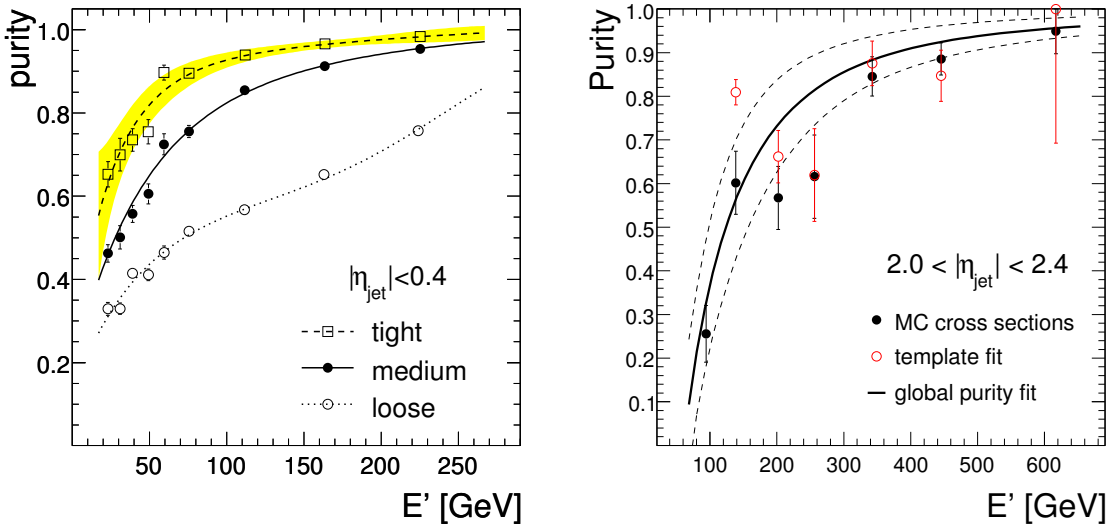

Figure 4.5 : Left: Estimated purity of the selected γ+jet sample in data for central probe jets. The different curves correspond to the different photon ID criteria: tight (dashed), medium (solid) and loose (dotted). The band around the dashed curve represents the uncertainty on the purity for the tight sample. Right: Estimated purity of the selected γ+jet sample in data for forward probe jets and for the tight photon ID. Results from MC cross sections (full circles) and the template fit method (open circles) are shown. The fit (solid line) represents the global result and the dashed lines the uncertainty band.

respective fractional sample composition:

$$R^{\text{mixture}}_{\text{MPF},\eta} = \rho_\eta R^{\gamma^{\text{meas}}+\text{jet}}_{\text{MPF},\eta} + (1-\rho_\eta) R^{\text{dijetbkg}}_{\text{MPF},\eta}, \tag{4.21}$$

where ρ_η is the sample purity (see Fig. 4.5). The signal γ+jet and dijet background MPF responses appearing in Eq. (4.21) also depend on the photon ID. Figure 4.6 compares them in the central calorimeter as predicted by the MC, for the medium and tight photon ID criteria. The fits of the relative differences in response are shown with a solid line in the bottom insets. The estimated relative correction to the MPF response for $|\eta^{\text{jet}}_{\text{det}}| = 0$ due to the dijet background contamination, $c_{\text{bkg},\eta=0}$, is shown as dashed lines in Fig. 4.7 for the medium and the tight photon ID criteria.

Another bias correction to the MPF response defined by the right hand side in Eq. (4.17) in the γ+jet sample, relates to the fact that $p^{\text{meas}}_{T\gamma} \neq p^{\text{ptcl}}_{T\gamma}$, and therefore $R_\gamma \neq 1$. The energy calibration of the electromagnetic calorimeter is obtained using electrons from $Z \to e^+e^-$ decays, which ensures perfect calibration for electrons/positrons of $E \approx 45$ GeV. Dependencies with η_{det} and p_T are derived from MC using the standard version of d0gstar (the GEANT [60] based simulation of the DØ detector) to simulate interactions with the detector, yielding an electron energy scale (EES) with 0.5% accuracy. This EES is part of the reconstruction and is applied to any reconstructed

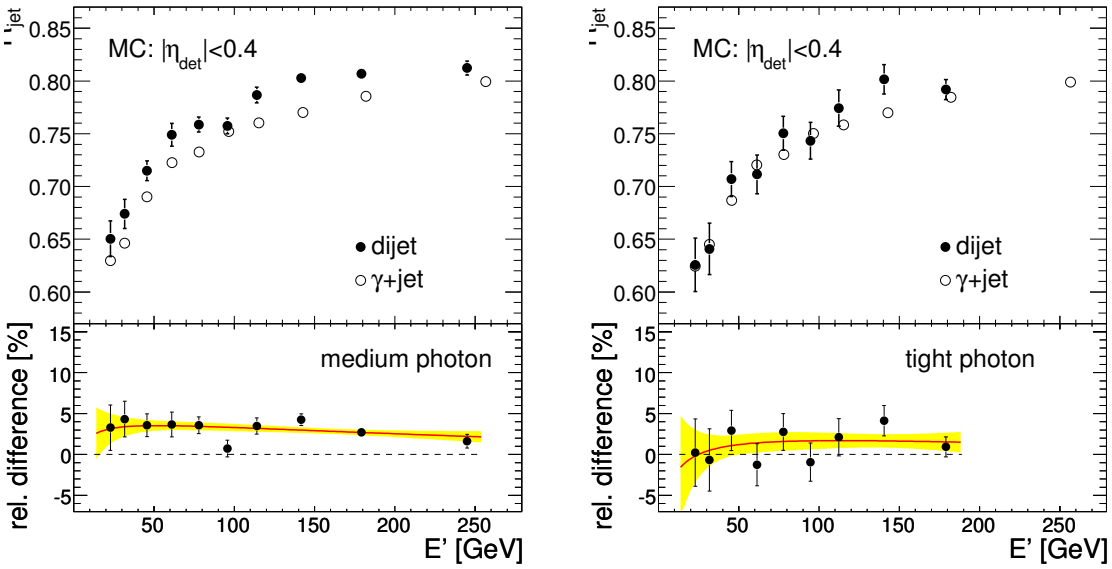

Figure 4.6 : Comparison of MPF response between signal γ+jet and dijet MC samples, for the medium (left) and the tight (right) photon identification criteria.

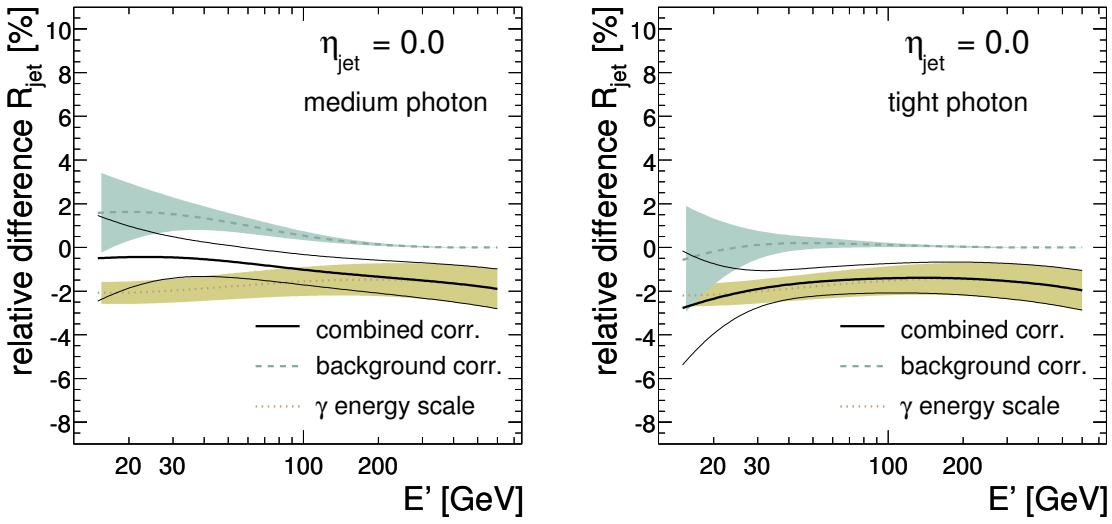

Figure 4.7 : Estimated corrections to the photon candidate for the MPF response measurement in data: dijet background (dashed), photon energy scale (dotted) and the sum of both (solid). Different plots correspond to different photon identification criteria: medium (left) and tight (right).

electromagnetic object ($e^{\pm}, \gamma$). However, since photons interact less with the material in front of the calorimeter, this EES is not directly applicable to photons. The photon energy scale (PhES) is defined as the additional correction, after EES, needed to bring

the energy of the photons to the particle level.

The correction factor due to the PhES can only be estimated in MC and is defined as

$$k_{\mathrm{R}}^{\gamma,\mathrm{PhES}} = \frac{R_{\mathrm{MPF,CC}}^{\gamma+\mathrm{jet}}}{R_{\mathrm{MPF,CC}}^{\gamma^{\mathrm{meas}}+\mathrm{jet}}} = \frac{p_{T\gamma}^{\mathrm{meas}}}{p_{T\gamma}^{\mathrm{ptcl}}} \equiv 1 - c_{\mathrm{PhES}}, \tag{4.22}$$

where c_{PhES} is the PhES. This correction highly relies on how well the MC describes the interactions of electromagnetic particles with the detector. In fact, the standard simulation in d0gstar is known not to describe with sufficient precision neither the electromagnetic showers nor the amount of material in front of the calorimeter. Additional corrections to the EES and the PhES in MC, are derived using a more detailed MC simulation with a special version of d0gstar, which was shown to reproduce the longitudinal energy depositions in the calorimeter for electrons from Z and J/Ψ decays with high precision. These additional corrections are therefore the ones needed to match both energy scales in MC to data. Among other things, the improved simulation adds fudge material before the solenoid; the preferred value is $\sim 0.28X_0$. Figure 4.8 shows the difference in the PhES between the improved and the standard MC for two extremes of fudge material, $0.17X_0$ (full circles) and $0.36X_0$ (open circles), and the interpolation to $0.28X_0$, where the fit to it (dotted line) is used as the central value of the additional correction. The full difference with respect to the extreme cases ($0.17X_0$ and $0.36X_0$) is assigned as an uncertainty.

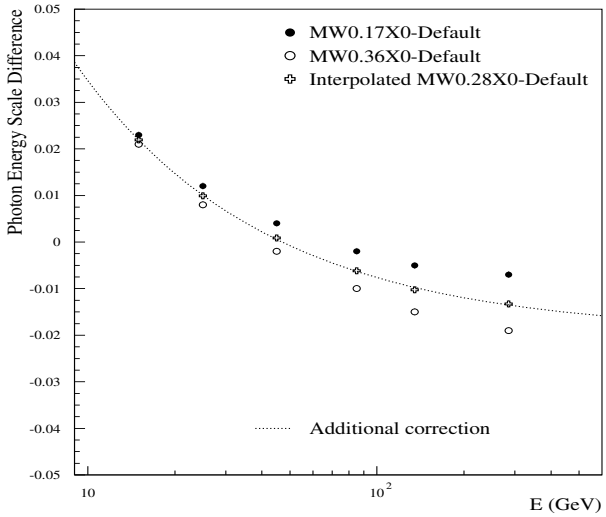

Figure 4.8 : Difference in the PhES between the special and the default versions of d0gstar. The corresponding difference in the EES is -6% (-5.5% and -6.7%) for 45 GeV electrons/positrons in the case of $0.28X_0$ ($0.17X_0$ and $0.36X_0$) of fudge material in front of the solenoid; these differences in EES are also applied to photons before deriving the PhES.

The dotted lines in Fig. 4.7 show the estimated total relative correction to the MPF response due to the PhES, $c_{\rm PhES}$, for the medium and tight photon ID criteria. As it can be appreciated, the dependence of this correction with E' and with the photon ID criteria is very small.

The total correction to the MPF response in γ+jet data, due to the combined effect of dijet background and PhES, is:

$$k_{\rm R}^{\gamma} = k_{{\rm R},\eta}^{\gamma,{\rm bkg}} \times k_{\rm R}^{\gamma,{\rm PhES}} = \frac{(1 - c_{\rm PhES})}{(1 + c_{{\rm bkg},\eta})} \approx 1 - (c_{{\rm bkg},\eta} + c_{\rm PhES}), \tag{4.23}$$

since $c_{{\rm bkg},\eta}, c_{\rm PhES} \ll 1$. The sum $(c_{{\rm bkg},\eta} + c_{\rm PhES})$ for $\eta_{\rm det}^{\rm jet} = 0$ is shown in Fig. 4.7 as a thick solid line. The thin solid lines represent the total uncertainty, which is calculated by propagation of the corresponding uncertainties on the background and photon energy scale corrections also shown in the figure. For the medium photon ID at $E' < 50$ GeV, the background correction almost compensates the PhES, and the total correction is -0.5%. For higher energies as well as for the tight photon ID the background contamination is negligible and the PhES correction dominates.

4.6 Central Calorimeter Response

The (photon corrected) MPF response in data $R_{\rm MPF,CC}^{\gamma+{\rm jet}}$, using the tight photon ID, is shown in Fig. 4.9 for $R_{\rm cone} = 0.7$ (left) and $R_{\rm cone} = 0.5$ (right) jets in the central calorimeter, as a function of the true jet energy estimator E' given by Eq. (4.18). As expected, the difference between them is very small; at $E' = 100$ GeV, the first is only 0.5% lower than the second. On the other hand, the measured response in MC at $E' = 100$ GeV is $\approx 3\%$ higher than in data. The energy dependence of the CC-response is well described using a quadratic logarithmic function:

$$R_{\rm CC}(E') = p_0 + p_1 \log(E'/E_0) + p_2 \log^2(E'/E_0), \tag{4.24}$$

where $E_0 = 100$ GeV and p_i $(i = 0, 1, 2)$ are free parameters to be determined. In Fig. 4.9 left (right), the dotted (solid) line represents the result of a χ^2 fit to the points using Eq. (4.24).

As it can be appreciated, the sample statistics is concentrated at low to moderate values of E' with the last point at $E' \approx 300$ GeV. As a consequence, the statistical uncertainty of the fit, which is less than 0.2% in the $E' \sim 30{-}100$ GeV range, rapidly increases up to $\sim 2\%$ at $E' = 600$ GeV, as it can be seen in the bottom inset in Fig. 4.9 for the case $R_{\rm cone} = 0.5$. In the case of $R_{\rm cone} = 0.7$ jets, the uncertainty at high energies is reduced by means of a dedicated MC study in which the cell level energy deposited by hadrons is scaled down in order to reproduce the jet response measured in data. This is a topic that goes beyond the scope of this thesis, since the single top analysis uses $R_{\rm cone} = 0.5$

jets; the interested reader should refer to Ref. [46]. The solid line in Fig. 4.9 (left) represents the result of the MPF response fit based on scaled-MC, and the dashed lines in the bottom inset represent the statistical uncertainty. The result of the MC-based fit agrees very well with the quadratic logarithmic fit, but it has significantly smaller uncertainty at high energies. The improvement in the uncertainty is due to the more solid physics foundation of the MC-based prediction (e.g. the energy-dependence of the quark and gluon fractions in γ+jet events, the fragmentation differences between quark and gluons, etc.).

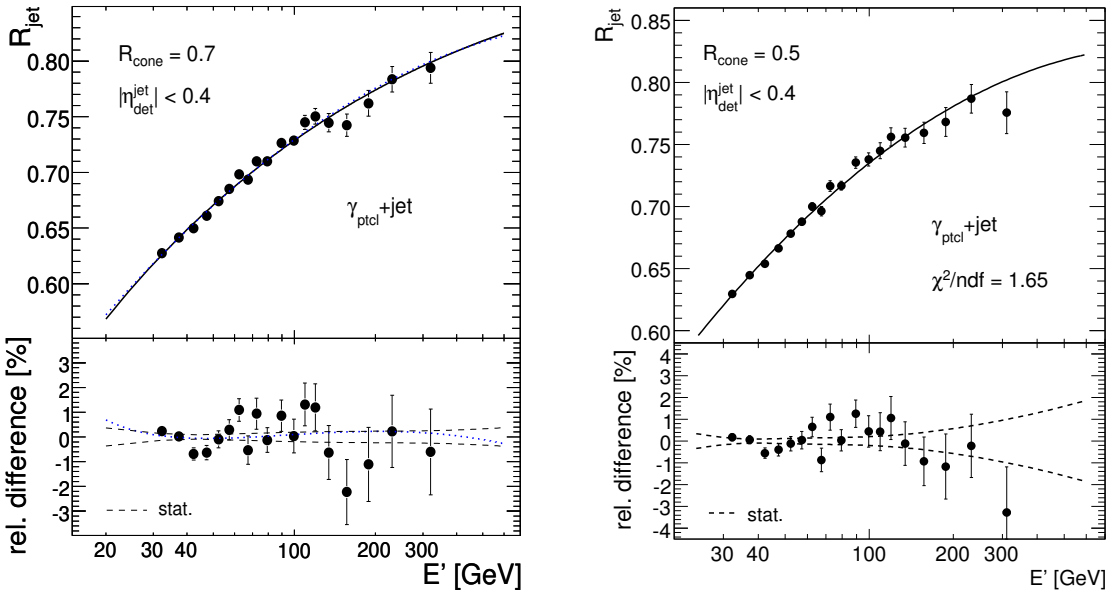

Figure 4.9 : Absolute MPF response measured in the γ+jet data sample, for $R_{\rm cone} = 0.7$ (left) and $R_{\rm cone} = 0.5$ (right) jets. The response is corrected for the effects of the dijet background contamination and the photon energy scale, as described in Sec. 4.5.3. See main text for more details.

Finally, it will be useful for the η-dependent corrections to have a parameterization of the CC-response in terms of the measured energy of the jet instead of E'. This is obtained by mapping the E' variable to the measured probe jet energy as well as the MPF response measurements according to:

$$E' \longrightarrow E_{\rm meas} = R_{\rm CC}(E')E', \tag{4.25}$$

$$R^{\gamma+{\rm jet}}_{\rm MPF,CC}(E') \longrightarrow R^{\gamma+{\rm jet}}_{\rm MPF,CC}(E_{\rm meas}) = R^{\gamma+{\rm jet}}_{\rm MPF,CC}(E'), \tag{4.26}$$

and fitting the mapped MPF response $R^{\gamma+{\rm jet}}_{\rm MPF,CC}(E_{\rm meas})$ with a quadratic logarithmic function:

$$R^{{\rm vs}E_{\rm meas}}_{\rm CC}(E_{\rm meas}) = \tilde{p}_0 + \tilde{p}_1 \log(E_{\rm meas}/E_0) + \tilde{p}_2 \log^2(E_{\rm meas}/E_0). \tag{4.27}$$

4.6.1 Uncertainties

In the case of the MPF response measurement in MC, the only uncertainty results from the statistical error of the fit. In data, there are additional systematic uncertainties related to the photon corrections and to the high energy extrapolation in the case of $R_{\rm cone} = 0.7$ jets. Figure 4.10 shows the main sources of uncertainty in the absolute MPF response measurement in data for $R_{\rm cone} = 0.7$ jets. The uncertainties for $R_{\rm cone} = 0.5$ jets up to an energy of ~ 150 GeV[30] are identical except that there is no such high energy extrapolation error (labeled "frag.+PDF" in the figure). As it can be appreciated, the dominant sources of systematic uncertainty are the ones associated to the dijet background correction at $E' < 30$ GeV and to the PhES at $E' > 30$ GeV.

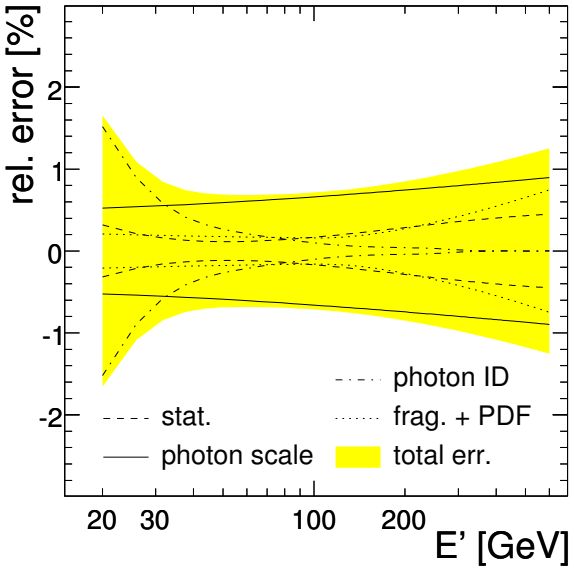

Figure 4.10 : Relative uncertainties on the absolute MPF response measurement in data.

The uncertainty on the dijet background correction, labeled as "photon ID" in Fig. 4.10 is related to the uncertainty on the $c_{{\rm bkg},\eta}$ correction factor (see blue band in Fig. 4.7), which has two components added in quadrature: purity (see yellow band in Fig. 4.5) and relative response between pure γ+jet and dijet MC events (see yellow band in Fig. 4.6).

The uncertainty on the PhES correction has two main contributions, which are added in quadrature. The first contribution is the uncertainty on the determination of the absolute electron energy scale, which has been estimated to be 0.5% using the $Z \to e^+e^-$ mass peak. The second contribution is related to uncertainties in the MC description of the relative energy scale between photons and electrons, as a result of

[30]Most of the jets in the single top analysis have $p_T < 150$ GeV/c.

the different interaction with matter. It consists of two parts: one is related with the knowledge of material in front of calorimeter and the other with the uncertainties in the simulation of photon-initiated showers in the material. The estimation of the first uncertainty was explained in Sect. 4.5.3. The second uncertainty is difficult to quantify; it is conservatively estimated as 40% larger than the uncertainty due to the material description.

4.7 Eta-Dependent Corrections

To the purpose of the jet energy scale, the DØ calorimeter can be assumed to be uniform in ϕ. On the other hand, the non-uniformity of the DØ detector in η (e.g. different amounts of dead material in front of the solenoid, the poorly instrumented ICR, etc.) makes the calibration in η insufficient, reflected in the observed large variations in the response as a function of $\eta_{\rm det}^{\rm jet}$. The exception is the CC cryostat where the jet response is rather uniform. Taking the well measured response in CC (see Sect. 4.6) as the reference, the goal of the η-dependent corrections is to make the MPF response uniform versus $\eta_{\rm det}^{\rm jet}$, as expressed by Eq. (4.12). The non-uniformities of the detector also result in a different energy dependence of the calorimeter response across $\eta_{\rm det}^{\rm jet}$; therefore F_η is also a function of the jet energy.

The relative MPF response correction is estimated using samples of γ+jet and dijet events. The former allows a direct and consistent derivation of the MPF response relative to the central calorimeter. The latter brings the additional statistics required to measure this correction in fine bins of $\eta_{\rm det}^{\rm jet}$ and up to much higher energies than the γ+jet sample can reach. By contrasting and combining these two different physics samples, the uncertainties are significantly reduced.

Although the response is a function of the energy, for the purpose of the η-dependent corrections the binning of the MPF response in γ+jet events is conveniently done in the variable p'_T, defined by Eq. (4.19), and then mapped to E', Eq. (4.18), before fitting. Similarly, the binning variable in dijet events is an estimator of the transverse momentum of the probe jet using the tag jet:

$$p'^{\,\rm meas}_T = p^{\rm meas,Ocorr}_{T\rm tag} \frac{\cosh(\eta_{\rm jet})}{\cosh(\eta_{\rm det}^{\rm jet})}, \tag{4.28}$$

where $p^{\rm meas,Ocorr}_{T\rm tag}$ is the measured transverse momentum of the tag jet after applying to it the offset correction. This last variable is the one used to calculate the relative MPF response in dijet events:

$$R^{\rm dijet}_{{\rm relMPF},\eta} = 1 + \frac{\vec{\not\!\!E}_T^{\,\rm meas} \cdot \hat{n}_{T\rm tag}}{p^{\rm meas,Ocorr}_{T\rm tag}}. \tag{4.29}$$

Note that the offset correction in the transverse momentum of the tag jet is not propagated to the missing E_T.

Several steps are involved in the measurement of the η-dependent corrections. This steps are briefly enumerated below before going into more details in the following sections.

1. For each p'_T ($p'_T{}^{\text{meas}}$) bin, the (relative) MPF response, Eqs. (4.17) and (4.29), is estimated in the γ+jet (dijet) sample in fine bins of $\eta^{\text{jet}}_{\text{det}}$ (0.1 in CC, 0.05 in ICR, and $0.2 - 0.4$ in EC).

2. In the case of dijet, the binning in the variable $p'_T{}^{\text{meas}}$ introduces a bias in the measured average of $p'_T{}^{\text{meas}}$ and therefore in $R^{\text{dijet}}_{\text{relMPF},\eta}$, related to the poor resolution in the measured transverse momentum of the tag jet. These biases are corrected for according to the procedure described in Sect. 4.7.1, leading to resolution bias corrected measurements of the relative MPF response, $R^{\text{dijet,corr}}_{\text{relMPF},\eta}$, and tag jet transverse momentum, $p^{\text{meas,corr}}_{T\text{tag}}$. The variable E' in dijet is then defined as
$$E' = \frac{p^{\text{meas,corr}}_{T\text{tag}}}{R^{\text{vs}E_{\text{meas}}}_{\text{CC}}(p^{\text{meas,corr}}_{T\text{tag}})}\cosh(\eta_{\text{jet}}), \tag{4.30}$$
and the variable p'_T as in Eq. (4.19). Note that since the tag jet is within $|\eta^{\text{jet}}_{\text{det}}| < 0.4$, $p^{\text{meas,corr}}_{T\text{tag}} \approx E^{\text{meas}}_{\text{tag}}$.

3. Restricted to the $\eta^{\text{jet}}_{\text{det}}$ bins within $|\eta^{\text{jet}}_{\text{det}}| < 0.4$, and for each p'_T ($p'_T{}^{\text{meas}}$) bin, the (relative) MPF response measurements in γ+jet (dijet) are fitted with the expected dependence of the CC-response with $\eta^{\text{jet}}_{\text{det}}$ using the following expressions:
$$\hat{p}_{3,p'_T} R_{\text{CC}}(\eta^{\text{jet}}_{\text{det}}; p'_T) \qquad \text{in } \gamma + \text{jet}, \tag{4.31}$$
$$\hat{p}_{3,p'_T} \frac{R_{\text{CC}}(\eta^{\text{jet}}_{\text{det}}; p'_T)}{R_{\text{CC}}(\eta^{\text{jet}}_{\text{det}} = 0; p'_T)} \quad \text{in dijet}, \tag{4.32}$$
where, from Eq. (4.24),
$$R_{\text{CC}}(\eta^{\text{jet}}_{\text{det}}; p'_T) = \hat{p}_0(p'_T) + \hat{p}_1(p'_T)\log(\cosh(\eta^{\text{jet}}_{\text{det}})) + \hat{p}_2(p'_T)\log^2(\cosh(\eta^{\text{jet}}_{\text{det}})), \tag{4.33}$$
with
$$\begin{aligned} \hat{p}_0(p'_T) &= p_0 + p_1\log(p'_T/E_0) + p_2\log^2(p'_T/E_0), \\ \hat{p}_1(p'_T) &= p_1 + 2p_2\log(p'_T/E_0), \\ \hat{p}_2(p'_T) &= p_2. \end{aligned} \tag{4.34}$$
The p_i ($i = 0, 1, 2$) inside the R_{CC} parameterization are the already determined coefficients for the CC-response (from Sect. 4.6), and $\hat{p}_{3,p'_T}$ is the free parameter to be determined. In the case of data, the photon corrections, $k^{\gamma}_{\text{R},\eta=0}$, are applied

to the MPF response measurements before the fit. The purpose of this step is to determine any residual correction (i.e. $\hat{p}_{3,p'_T}$) required to ensure that the relative MPF response measurements, that will be constructed in a later step, are on average one within $|\eta^{\text{jet}}_{\text{det}}| < 0.4$. In the case of dijet, $\hat{p}_{3,p'_T}$ represents a residual to the estimated resolution bias correction in CC, and it is found to be consistent with one within $< 0.5\%$. In the case of γ+jet, $\hat{p}_{3,p'_T}$ is found to be one on average, as expected.

4. Using the MPF response measurements from steps 1 and 2, and the $\hat{p}_{3,p'_T}$ determined in step 3, the η-dependent correction measurements in a given $(p'_T, \eta^{\text{jet}}_{\text{det}})$ bin are computed as:

$$F_\eta^{\gamma+\text{jet}}|_{p'_T} = R^{\text{mixture}}_{\text{MPF},\eta}|_{p'_T} \times \frac{1}{\hat{p}_{3,p'_T} R_{\text{CC}}(\eta^{\text{jet}}_{\text{det}}; p'_T)} \quad \text{in } \gamma + \text{jet}, \tag{4.35}$$

$$F_\eta^{\text{dijet}}|_{p'_T} = R^{\text{dijet,corr}}_{\text{relMPF},\eta}|_{p'_T} \times \frac{R_{\text{CC}}(\eta^{\text{jet}}_{\text{det}} = 0; p'_T)}{\hat{p}_{3,p'_T} R_{\text{CC}}(\eta^{\text{jet}}_{\text{det}}; p'_T)} \quad \text{in dijet}, \tag{4.36}$$

Note that the numerator in Eq. (4.35) is corrected neither for the dijet background nor for the miscalibration of the PhES, and therefore, even in CC, this ratio is slightly smaller than one. The photon corrections to the numerator are instead introduced into the fitting function to the $F_\eta^{\gamma+\text{jet}}$ measurements.

5. The F_η corrections from step 4 in each (p'_T, $\eta^{\text{jet}}_{\text{det}}$) bin are mapped to the corresponding (E', $\eta^{\text{jet}}_{\text{det}}$) bin. The solid (open) circles in Fig. 4.11 illustrate the measured correction for γ+jet (dijet) as a function of E' in two different $\eta^{\text{jet}}_{\text{det}}$ bins.

6. A simultaneous 2-dimensional ("global") fit to the full set of $F_\eta^{\gamma+\text{jet}}$ and F_η^{dijet} measurements is performed in the $(E', \eta^{\text{jet}}_{\text{det}})$ space. The result of this fit gives the final η-dependent corrections for all $(E', \eta^{\text{jet}}_{\text{det}})$.

4.7.1 Resolution Bias in Dijet

Response measurements in dijet events are binned in terms of the poorly measured tag-jet p_T. Given the steeply falling nature of the jet p_T spectrum, upward fluctuations (due to resolution) from lower p_T jets into a given p_T bin dominate against downward fluctuations from higher p_T jets, resulting in a flattened observed jet p_T spectrum compared to the true spectrum. Consequently, the average value of the measured (smeared) jet p_T in a given bin ($\langle p_T^{\text{smear}} \rangle$) is shifted up compared to the average value of the true (unsmeared) jet p_T ($\langle p_T^{\text{true}} \rangle$). This is the so-called "resolution bias". Denoting with $f(p_T^{\text{ptcl}})$ the particle level jet p_T differential cross section and with $g(p_T^{\text{smear}} - p_T^{\text{true}})$ the smearing function, the average value of p_T^{true} for a given value of p_T^{smear} can be

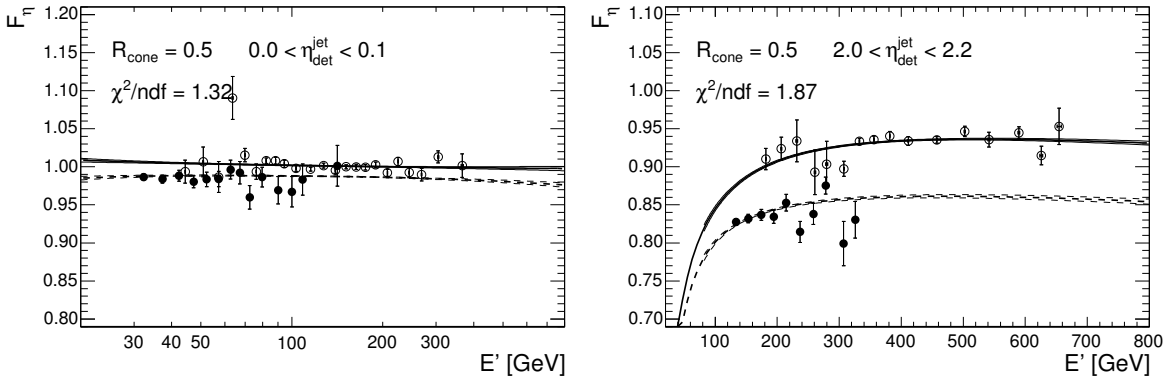

Figure 4.11 : Relative MPF response correction for $R_{\text{cone}} = 0.5$ jets in data as a function of E' and for two different $\eta_{\text{det}}^{\text{jet}}$ bins: $0 < \eta_{\text{det}}^{\text{jet}} < 0.1$ (left) and $2.0 < \eta_{\text{det}}^{\text{jet}} < 2.2$ (right). The solid (open) circles represent the measurements in the γ+jet (dijet) sample. The lines shown represent the result from the global fit discussed in Sect. 4.7.3.

calculated according to the following formula:

$$\left\langle p_T^{\text{true}} \right\rangle |_{p_T^{\text{smear}}} = \frac{\int_0^\infty f(p_T^{\text{ptcl}}) g(p_T^{\text{smear}} - p_T^{\text{true}}(p_T^{\text{ptcl}})) p_T^{\text{true}}(p_T^{\text{ptcl}})\, dp_T^{\text{ptcl}}}{\int_0^\infty f(p_T^{\text{ptcl}}) g(p_T^{\text{smear}} - p_T^{\text{true}}(p_T^{\text{ptcl}}))\, dp_T^{\text{ptcl}}}. \tag{4.37}$$

Note that p_T^{smear} and p_T^{true} are assumed to only differ because of resolution effects, i.e. they are at the same level (e.g. if p_T^{true} is thought of as the particle level p_T, then p_T^{smear} should be thought of as the measured p_T corrected by JES). Equation (4.37) can be solved analytically at particle level in a simplified case where the p_T cross section is exponentially falling, $f(p_T^{\text{ptcl}}) = N_0 \exp(-\alpha p_T^{\text{ptcl}})$, and the smearing is gaussian with constant resolution, $g(p_T^{\text{smear}} - p_T^{\text{ptcl}}) = \text{Gauss}(p_T^{\text{smear}}; p_T^{\text{ptcl}}, \sigma = \text{const})$:

$$\begin{aligned} \left\langle p_T^{\text{ptcl}} \right\rangle |_{p_T^{\text{smear}}} &= \frac{\int_0^\infty N_0 \exp(-\alpha p_T^{\text{ptcl}}) \text{Gauss}\left(p_T^{\text{smear}}; p_T^{\text{ptcl}}, \sigma\right) p_T^{\text{ptcl}}\, dp_T^{\text{ptcl}}}{\int_0^\infty N_0 \exp(-\alpha p_T^{\text{ptcl}}) \text{Gauss}\left(p_T^{\text{smear}}; p_T^{\text{ptcl}}, \sigma\right) dp_T^{\text{ptcl}}} \\ &\approx p_T^{\text{smear}} - \alpha\sigma^2, \end{aligned} \tag{4.38}$$

where the only approximation made above is the extension of the lower limit of integration from 0 to $-\infty$. This example clearly shows that the bias is due to both the steeply falling p_T cross section and the finite p_T resolution, and as intuitively predicted it goes in the positive direction ($\alpha\sigma^2 > 0$).

In general, if p_T^{smear} is the binning variable, its resolution bias corrected value is given by

$$p_T^{\text{smear,corr}} = p_T^{\text{smear}} \left(1 - \frac{\delta p_T}{p_T^{\text{smear}}}\right), \tag{4.39}$$

with

$$\delta p_T = p_T^{\text{smear}} - \left\langle p_T^{\text{true}} \right\rangle |_{p_T^{\text{smear}}}, \tag{4.40}$$

where $\langle p_T^{\text{true}} \rangle |_{p_T^{\text{smear}}}$ is given by Eq. (4.37).

In the case of the relative MPF response measurements in dijet, $p_T^{\text{smear}} = p_{T\text{tag}}^{\text{meas,Ocorr}}$. Since this variable appears in the definition of the MPF response (Eq. (4.17)), the last will be also biased. According to the left hand side in Eq. (4.17) and noting that $R_{\text{tag}} = p_{T\text{tag}}^{\text{meas,Ocorr}}/p_{T\text{tag}}^{\text{ptcl}}$, the resolution bias corrected relative MPF response is given by:

$$R_{\text{relMPF}}^{\text{dijet,corr}} = R_{\text{relMPF}}^{\text{dijet}} \left(1 - \frac{\delta p_T}{p_T^{\text{smear}}}\right)^{-1}. \tag{4.41}$$

The resolution bias correction is calibrated in CC imposing the constraint that, when the probe jet is also in CC, the relative MPF response should be identically one, $F_\eta^{\text{dijet}} = 1$; any residuals are explicitly corrected for by means of the $\hat{p}_{3,p'_T}$ parameter (see Eq. (4.36)).

The expected bias in $p_{T\text{tag}}^{\text{meas,Ocorr}}$ is computed using Eqs. (4.40) and (4.37) with p_T^{smear} at the measured level:

$$\begin{aligned}\delta p_T(p_{T\text{tag}}^{\text{meas,Ocorr}}, \eta_{\text{det}}^{\text{probe}}) = p_{T\text{tag}}^{\text{meas,Ocorr}} \\ - \frac{\int f(x, \eta_{\text{det}}^{\text{probe}}) \text{Gauss}(p_{T\text{tag}}^{\text{meas,Ocorr}}; xR_{\text{CC}}(x), \sigma(x)) xR_{\text{CC}}(x)\, dx}{\int f(x, \eta_{\text{det}}^{\text{probe}}) \text{Gauss}(p_{T\text{tag}}^{\text{meas,Ocorr}}; xR_{\text{CC}}(x), \sigma(x))\, dx},\end{aligned} \tag{4.42}$$

with f representing the particle level spectrum of the tag jet for the dijet topologies used in the η-dependent corrections and σ the resolution at $p_{T\text{tag}}^{\text{meas,Ocorr}}$ level for central jets. Figure 4.12 illustrates the magnitude of the resolution bias correction in data. The determination of f and σ are briefly described next.

Dijet p_T Spectrum

In order to obtain the most realistic description of the dijet p_T spectrum, the same data as for the response measurements are used. The smeared cross section is parameterized with the ansatz typically used for the inclusive jet cross section unsmearing [59], but with a modified kinematic limit term that includes both tag and probe jets:

$$\begin{aligned}f(p_T, \eta) = N_0(\eta) \left(\frac{p_T}{p_{T,0}}\right)^{-\alpha(\eta)} \left[\left(1 - \frac{2p_T}{\sqrt{s}}\right)\left(1 - \frac{2p_T \cosh(\eta)}{\sqrt{s}}\right)\right]^{\beta(\eta)/2} \\ \cdot \exp\left(-\gamma(\eta)\frac{p_T}{p_{T,0}}\right),\end{aligned} \tag{4.43}$$

where $p_{T,0} = 100$ GeV/c, and N_0, α, β and γ are the parameters to be determined. Typical values for the parameters are $N_0 = 0.6{-}1.0$, $\alpha = 4.0{-}4.5$ and $\beta = 7.0{-}9.0$, while γ is fixed to 0. The smeared cross section is fitted to data in 0.1-wide bins of $\eta_{\text{det}}^{\text{probe}}$,

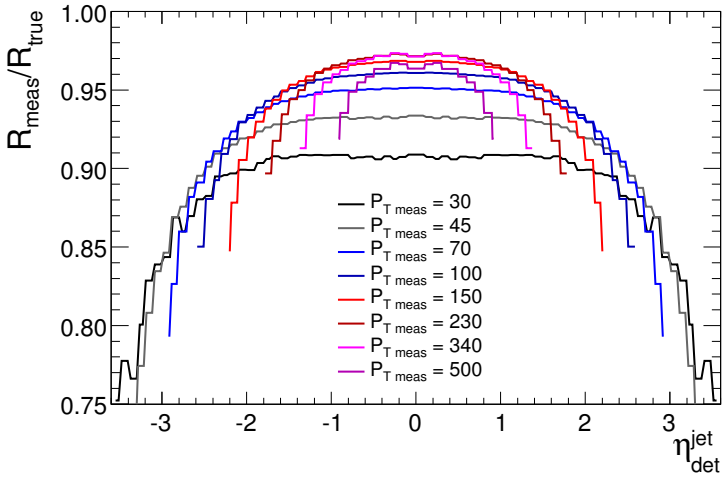

Figure 4.12 : Resolution bias correction in data for different values of $p_{T\text{tag}}^{\text{meas}}$ as a function of $\eta_{\text{det}}^{\text{probe}}$.

following the measurements to within a few percent. The procedure is validated in MC by observing good agreement between the particle level spectrum and the unsmearing of the measured tag jet p_T spectrum.

Tag Jet p_T Resolution

The so-called "raw" jet p_T resolution, which includes both detector and physics smearing effects, is measured using the Asymmetry Method [61] with both jets in CC:

$$\sigma_{\text{raw}} = \sqrt{2}\,\text{RMS}\left(\frac{p_{T,2}-p_{T,1}}{p_{T,2}+p_{T,1}}\right), \quad |\eta_{1,2}| < 0.4, \tag{4.44}$$

where the jet p_T's are corrected by offset and CC-response, and is parameterized as

$$\sigma = \sqrt{N^2 + S^2 p_T + C^2 p_T^2}\,. \tag{4.45}$$

The so-called "corrected" jet p_T resolution, containing only detector smearing effects, is obtained by unfolding the physics smearing resulting from soft radiation below the jet p_T reconstruction threshold ($p_{T,\text{soft}}^{\text{cut}} = 6$ GeV/c) and particle level imbalance (underlying event, physics showering, etc.):

$$\sigma_{\text{corr}} = \sqrt{(\sigma_{\text{raw}} K_{\text{soft}})^2 - \sigma_{\text{MC}}^2}. \tag{4.46}$$

The soft radiation correction K_{soft} is obtained by extrapolating to zero the fit to the raw resolutions measured with different reconstruction thresholds between 6 GeV/c and 20 GeV/c, and is parameterized as $K_{\text{soft}} = 1 - \exp(-a_0 - a_1 p_T)$. The particle level

imbalance σ_{MC} is estimated in particle-level MC using the same methods as used for data, and correcting it for soft radiation to avoid double-counting.

However, MC studies show that the least unbiased relative MPF response is obtained when considering an intermediate "partly-corrected" resolution, where the particle level imbalance σ_{MC} component is not subtracted:

$$\sigma_{\mathrm{part}} = \sqrt{(\sigma_{\mathrm{raw}} K_{\mathrm{soft}})^2}. \tag{4.47}$$

The fact of not including the soft radiation part in the resolution can be understood by the fact that the relative MPF response in dijet is not sensible to (and therefore not biased by) soft radiation, since the MPF method looks at the p_T balance between the tag jet and the whole recoil instead of the two reconstructed jets. On the other hand, although the MPF response is also not affected by the particle level imbalance, Eq. (4.42) does assume that the tag and probe jets correspond to the same particle level p_T, which may explain why σ_{MC} needs to be included in the resolution.

The rising jet energy response improves the relative jet p_T resolution by $\sim$ 10% compared to the resolution at the measured jet p_T level. Since the last is what is needed in Eq. (4.42), the following correction is applied to the relative resolution:

$$\frac{\sigma(p_T)}{p_T} \to \frac{\sigma(p_T)}{p_T} \left(1 + \frac{p_T R'_{\mathrm{CC}}(p_T)}{R_{\mathrm{CC}}(p_T)}\right). \tag{4.48}$$

As already mentioned, the "$\hat{p}_3$-term" provides any residual correction to the resolution bias in CC. To estimate an uncertainty for the resolution bias correction outside CC, the η-dependent corrections in dijet are compared in Fig. 4.13 for the three different scenarios of possible resolutions and using the p_T spectrum derived from data, and a fourth scenario using the MC-derived spectrum and the default partly-corrected resolution. The bottom insets in the plots show the relative difference with respect to the nominal case (data-derived spectrum and partly-corrected resolution). The dashed lines represent the total systematic uncertainty assigned to the resolution bias correction. The uncertainty increases linearly from 0.1% in CC to 0.75% in the very forward region.

4.7.2 Sample Dependence of the Relative Response

As shown in Fig. 4.11, the measured F_η corrections are significantly different between the γ+jet and dijet samples. Part of the observed difference is artificial, as it comes from the fact that the γ+jet points are corrected neither for the presence of dijet background nor for the miscalibration of the photon energy. This bias is taken into account in the fitting procedure as described in Sect. 4.7.3.

The main physical contribution to the discrepancy originates from the different parton flavor composition of the γ+jet and dijet samples. In γ+jet events with central jets,

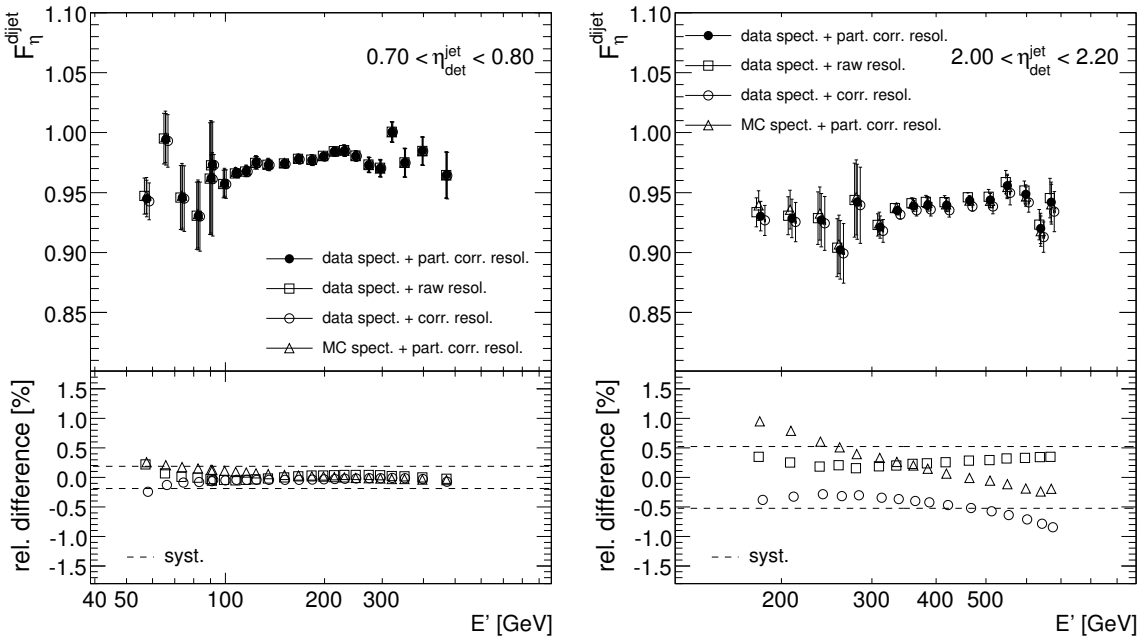

Figure 4.13 : Comparison of the relative response in dijets data for different scenarios of the resolution bias correction in two different $\eta_{\rm det}^{\rm jet}$ bins. The bottom insets show the relative difference with respect to the nominal one, i.e. data spectrum+partly-corrected resolution. The dashed lines represent the assigned systematic uncertainty to the resolution bias correction in data.

the fraction of gluon-initiated jets is ~ 0.1 at 20 GeV, growing to ~ 0.5 at 120 GeV and ~ 0.95 at 500 GeV. Jets from dijet events present exactly the opposite behavior. Moreover, since gluon jets have on average 9/4 more particles than quark jets, the energy spectrum of the particles within a gluon jet is softer than the spectrum of the particles within a quark jet of the same energy, resulting in a lower jet response. The ratio of the dijet to γ+jet corrections is found to be independent of E' over the range where both samples overlap ($50 \lesssim p'_T \lesssim 200$ GeV/c), and the $\eta_{\rm det}^{\rm jet}$ dependence is well described by a quadratic logarithmic function in $\cosh(\eta_{\rm det}^{\rm jet})$ (as the absolute response parameterization in Eq. (4.33)). Therefore, a "dijet-to-γ+jet scale factor" SF_η is defined in the whole E' range and parameterized as

$$SF_\eta \equiv \frac{F_\eta^{\rm dijet}(E')}{F_\eta^{\gamma+{\rm jet}}(E')} = 1 + b\log(\cosh(\eta_{\rm det}^{\rm jet})) + c\log^2(\cosh(\eta_{\rm det}^{\rm jet})), \tag{4.49}$$

with b and c being free parameters to be determined by the fit. This scale factor is essential to combine both measurements consistently in a unique global fit in the 2-dimensional plane of $(E', \eta_{\rm det}^{\rm jet})$, as explained in Sect. 4.7.3.

4.7.3 Global Fit

Given that the $F_\eta(E')$ corrections represent the relative jet response between a jet in a given η bin and a jet in CC, both at the same energy E', the chosen parameterization is a ratio of quadratic logarithmic functions versus E' with the denominator being the fixed CC-response parameterization, Eq. (4.24), determined in Sect. 4.6:

$$F_\eta^{\gamma+\text{jet}}(E';\{P_{i,\eta}\}) = \frac{P_{0,\eta} + P_{1,\eta}\log(E'/E_0) + P_{2,\eta}\log^2(E'/E_0)}{R_{\text{CC}}(E')}. \tag{4.50}$$

Furthermore, by means of the scale factor SF_η defined in Eq. (4.49), the corresponding parameterization for dijet events is given by

$$F_\eta^{\text{dijet}}(E';\{P_{i,\eta}\}, SF_\eta) = SF_\eta\, F_\eta^{\gamma+\text{jet}}(E';\{P_{i,\eta}\}), \tag{4.51}$$

with $F_\eta^{\gamma+\text{jet}}(E';\{P_{i,\eta}\})$ given by Eq. (4.50).

In the case of the γ+jet sample, the quadratic logarithmic parameterization corresponds to the response of pure γ+jet events with the photon at the particle level, and therefore it must be divided by k_{R}^{γ}, Eq. (4.23), when fitting the $F_\eta^{\gamma+\text{jet}}$ measurements. Assuming the response of the recoil in a dijet event is independent of whether the tag object is a jet or a misidentified photon, a useful expression can be obtained for k_{R}^{γ}:

$$\frac{1}{k_{\text{R}}^{\gamma}(E';SF_\eta)} = \frac{1}{k_{\text{R}}^{\gamma,\text{PhES}}(p'_T)}\left(\rho_\eta(p'_T) + (1-\rho_\eta(p'_T))SF_\eta\frac{R_{\text{MPF,CC}}^{\text{dijetbkg}}(p'_T)}{R_{\text{MPF,CC}}^{\gamma^{\text{meas}}+\text{jet}}(p'_T)}\right), \tag{4.52}$$

where the correction for the PhES, $k_{\text{R}}^{\gamma,\text{PhES}}$, the purity ρ_η and the ratio of responses $R_{\text{MPF,CC}}^{\text{dijetbkg}}/R_{\text{MPF,CC}}^{\gamma^{\text{meas}}+\text{jet}}$, are all estimated in MC as already discussed in Sect. 4.6 for the measurement of the absolute MPF response. The crucial step to obtain Eq. (4.52) is the validity of the following approximation:

$$\frac{R_{\text{MPF},\eta}^{\text{dijetbkg}}(E')}{R_{\text{MPF},\eta}^{\gamma^{\text{meas}}+\text{jet}}(E')} = SF_\eta\frac{R_{\text{MPF,CC}}^{\text{dijetbkg}}(p'_T)}{R_{\text{MPF,CC}}^{\gamma^{\text{meas}}+\text{jet}}(p'_T)}, \tag{4.53}$$

which has been verified in MC. Using the right hand side of Eq. (4.53) with SF_η being part of the fit to the data, makes the photon corrections versus $\eta_{\text{det}}^{\text{jet}}$ less dependent on the MC simulations.

In order to obtain η-dependent corrections that vary smoothly across the detector, the fit to the relative MPF response corrections is performed in the 2-dimensional $(E', \eta_{\text{det}}^{\text{jet}})$ plane, with the parameters $(P_{0,\eta}, P_{1,\eta}, P_{2,\eta}, SF_\eta)$ expressed as suitable functions of $\eta_{\text{det}}^{\text{jet}}$ (see e.g. Fig. 4.14), and whose coefficients become the actual parameters to be determined. In general, good quality fits are obtained through this procedure, with $\chi^2/ndf \approx 1-2$. The 2-dimensional fitting function is defined as

$$F_\eta(E';\{P_i(\eta_{\text{det}}^{\text{jet}})\}, SF(\eta_{\text{det}}^{\text{jet}})) = \begin{cases} F_\eta^{\gamma+\text{jet}}(E';\{P_i(\eta_{\text{det}}^{\text{jet}})\})/k_{\text{R}}^{\gamma}(E';SF(\eta_{\text{det}}^{\text{jet}})) & \text{if } \gamma\text{+jet,} \\ F_\eta^{\text{dijet}}(E';\{P_i(\eta_{\text{det}}^{\text{jet}})\}, SF(\eta_{\text{det}}^{\text{jet}})) & \text{if dijet,} \end{cases} \tag{4.54}$$

where $F_\eta^{\gamma+\text{jet}}(E';\{P_i(\eta_{\text{det}}^{\text{jet}})\})$, $F_\eta^{\text{dijet}}(E';\{P_i(\eta_{\text{det}}^{\text{jet}})\},SF(\eta_{\text{det}}^{\text{jet}}))$ and $k_{\text{R}}^{\gamma}(E';SF(\eta_{\text{det}}^{\text{jet}}))$ are given by Eqs. (4.50), (4.51) and (4.52) respectively, but replacing $P_{i,\eta} \to P_i(\eta_{\text{det}}^{\text{jet}})$ and $SF_\eta \to SF(\eta_{\text{det}}^{\text{jet}})$. Another advantage of combining the two samples in a unique 2-dimensional global fit is the reduction of the uncertainty from extrapolation of the relative response correction to the highest energies. The global fit procedure has been verified in MC to give a reliable high energy extrapolation. In Fig. 4.11, the solid and dashed lines represent the result of the global fit for dijet and γ+jet respectively in two individual $\eta_{\text{det}}^{\text{jet}}$ bins, where it can be clearly seen how the dijet measurements impose the high energy behavior of the η-dependent corrections also for γ+jets. The small bands around the lines represent the statistical uncertainty calculated from the covariance matrix, which has been reduced enormously by combining both measurements.

Figure 4.14 shows the parameterizations of (P_0, P_1, P_2, SF) resulting from the global fit, in bins of $\eta_{\text{det}}^{\text{jet}}$ and evaluated at the mean value of $\eta_{\text{det}}^{\text{jet}}$ in each bin. The SF term parameterization has a solid physical foundation with only two free parameters (as explained in Sect. 4.7.2) and therefore its observed smoothness. The P_1 and P_2 terms have four free parameters each, two governing the behavior in the central region ($|\eta_{\text{det}}^{\text{jet}}| \lesssim 1.1$) and two for the forward region ($|\eta_{\text{det}}^{\text{jet}}| \gtrsim 1.4$); in the intermediate ICR only continuity is imposed. The P_1 term tells the rate at which the response increases with the jet energy, and as such it can be thought of as being proportional to the amount of sampling material the jet traverses at a given angle of incidence with respect to $\eta_{\text{det}}^{\text{jet}} = 0$ ($|\eta_{\text{det}}^{\text{jet}}| \to +\infty$) in the central (forward) region, which is given by $\sin^{-1}(\theta_{\text{det}}^{\text{jet}}) = \cosh(\eta_{\text{det}}^{\text{jet}})$ and $|\cos^{-1}(\theta_{\text{det}}^{\text{jet}})| = |\tanh^{-1}(\eta_{\text{det}}^{\text{jet}})|$ respectively. The P_2 term is thought of as a second order effect to P_1. Thus, the parameterizations used are:

$$P_{1/2}(\eta_{\text{det}}^{\text{jet}}) = \begin{cases} d_{1/2} + e_{1/2}(\cosh^{2/4}(\eta_{\text{det}}^{\text{jet}}) - 1) & \text{in the central region,} \\ f_{1/2} + g_{1/2}(\tanh^{-2/4}(\eta_{\text{det}}^{\text{jet}}) - 1) & \text{in the forward region.} \end{cases} \tag{4.55}$$

On the other hand, the P_0 term is the one left with more freedom, representing a residual to the calibration of the calorimeter in fine $\eta_{\text{det}}^{\text{jet}}$ bins.

4.7.4 Summary of Results

The final relative response correction in γ+jet (see Eq. (4.50)) obtained from the simultaneous global fit procedure described in Sect. 4.7.3, is shown in Fig. 4.15 for $R_{\text{cone}} = 0.5$ jets as a function of $\eta_{\text{det}}^{\text{jet}}$ and for different values of E'. The corresponding correction for dijet (see Eq. (4.51)) is the same but scaled by the scale factor shown in the bottom right plot in Fig. 4.14. The measured correction for $R_{\text{cone}} = 0.7$ jets is very similar. The figure illustrates the non-uniformity of the calorimeter, with a jet energy response that (for a fixed energy) decreases with higher rapidities. At low energies, this effect is primarily due to the lowering of the jet p_T with increasing $\eta_{\text{det}}^{\text{jet}}$. Indeed, the η-dependent corrections for fixed values of $p_{T\text{jet}}$ are much more flat, as observed in

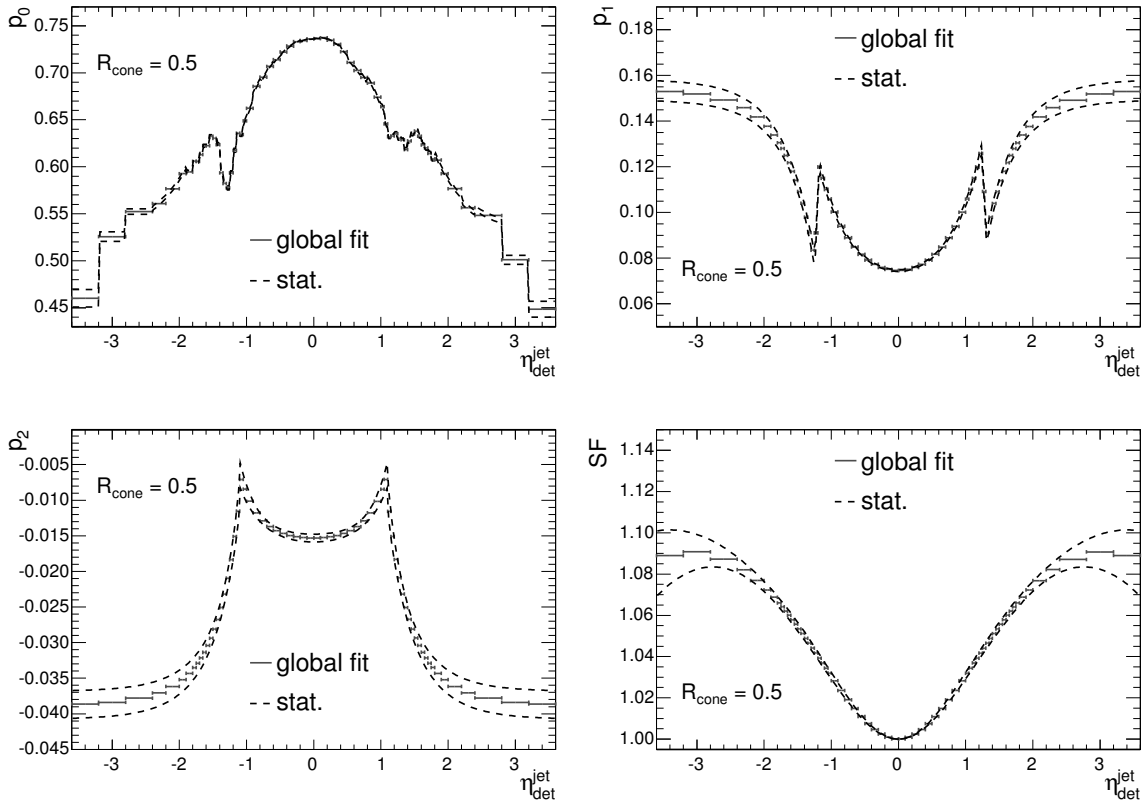

Figure 4.14 : Parameterizations for (P_0, P_1, P_2, SF) versus $\eta^{\rm jet}_{\rm det}$ resulting from the global fit to the relative MPF response measurements in data for $R_{\rm cone} = 0.5$ jets. The dashed lines illustrate the statistical uncertainty band.

Fig. 4.27. While in CC the relative response is, by definition, one on average, in the EC (and in the ICR) a large energy dependence is observed for low to moderate energy jets. The deeps around $1.1 \lesssim |\eta^{\rm jet}_{\rm det}| \lesssim 1.5$ correspond to the poorly instrumented ICR, where the observed asymmetry between negative and positive values of $\eta^{\rm jet}_{\rm det}$ is a consequence of the passage of the calorimeter cabling in the negative η hemisphere.

4.7.5 Uncertainties

Figure 4.16 illustrates the relative uncertainty on the relative MPF response corrections for $R_{\rm cone} = 0.5$ jets in 0.4-wide bins of $|\eta^{\rm jet}_{\rm det}|$. The uncertainties corresponding to $R_{\rm cone} = 0.7$ jets are rather similar. There are three sources of uncertainty shown in the figures. One is the statistical uncertainty, which is computed by error propagation on Eq. (4.54) using the full covariance matrix from the global fit. To cover for possible imperfections/arbitrariness in the global fit, the statistical uncertainty is increased by $\sqrt{\chi^2/ndf}$ in those $\eta^{\rm jet}_{\rm det}$ bins where $\chi^2/ndf > 1$. As it can be appreciated in Fig. 4.11,

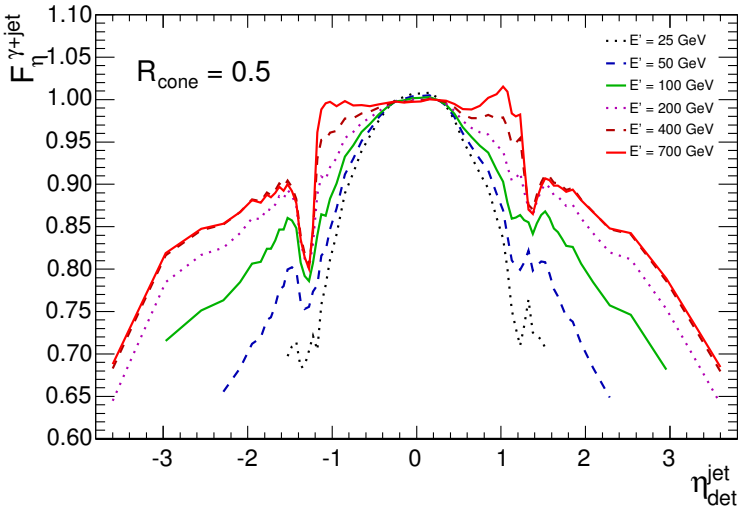

Figure 4.15 : Final relative MPF response correction for $R_{\rm cone} = 0.5$ jets in γ+jet events as a function of $\eta_{\rm det}^{\rm jet}$. The different lines correspond to particular values of E': for example, the green line corresponds to $E' = 100$ GeV and therefore has the same shape as the P_0 term in the upper left plot in Fig. 4.14.

the energy dependence of the statistical uncertainty is very weak, being constrained at low p_T by the γ+jet measurements and at high p_T by the dijet measurements. In Fig. 4.16 it is shown for $p'_T = 50$ GeV/c. The statistical uncertainty becomes important in the very forward region ($|\eta_{\rm det}^{\rm jet}| > 2.8$) where there is very limited statistics.

In the case of $F_\eta^{\gamma+{\rm jet}}$, a systematic uncertainty is assigned for the η-dependence of the dijet background correction $k_{{\rm R},\eta}^{\gamma,{\rm bkg}}$ (the term between parenthesis in the right hand side of Eq. (4.52)) by propagating the uncertainties in the purity (see e.g. the error band in the right plot in Fig 4.5) and in the ratio between the MPF response for dijet background and pure γ+jet events. To avoid double counting when propagating uncertainties using Eq. (4.12), the uncertainty in $k_{\rm R}^{\gamma,{\rm PhES}}$ is not included and the uncertainty in $k_{\rm R,CC}^{\gamma,{\rm bkg}}$ is subtracted in quadrature. The remaining assigned uncertainty is displayed as the dotted line in Fig. 4.16.

In the case of dijet, two other systematic uncertainties need to be considered. The first one is due to the need of the resolution bias correction. The derivation of this systematic uncertainty and its magnitude were explained in Sect. 4.7.1. The second uncertainty is related to the high energy extrapolation of the η-dependent corrections in dijet. To evaluate the accuracy of such extrapolation, a test is performed in MC in which the global fit is performed using only the measurements contained in a similar energy range as available in data. The result of such fit is then compared to the actual MC measurements available at higher energies. Although some deviations between the prediction and the high E' measurements are observed in the test, they are thought

to be associated to instabilities in the global fit due to the limited MC statistics and large MC weight fluctuations rather than to an inaccurate prediction. However, to stay conservative, an energy dependent systematic uncertainty is assigned: it is 1% at $E' \sim 800$ GeV and decreases linearly to zero at the E' value where sufficiently precise measurements are available in data in each $|\eta^{\text{jet}}_{\text{det}}|$ bin.

The last of the uncertainties assigned to the η-dependent corrections is purely related to how well the global fit method works. An internal closure test is performed, defining suitable observables for γ+jet and for dijet, which compare the corresponding F_η corrections determined by the global fit as a function of E' against the MPF response measurements in 0.4-wide bins of $|\eta^{\text{jet}}_{\text{det}}|$. In case the corrections work perfectly, the estimated closure observables should be identical to one and independent of energy. Any systematic deviations from this ideal behavior are considered as a residual in the F_η corrections. The corresponding uncertainty is determined by fitting a constant versus E' in each 0.4-wide $|\eta^{\text{jet}}_{\text{det}}|$ bin to the estimated observables. In general, residuals smaller than 1% over the available energy range are found, owing to the physical energy parameterization used during the global fit. The assigned systematic uncertainty for $R_{\text{cone}} = 0.5$ jets as a function of $|\eta^{\text{jet}}_{\text{det}}|$ is represented by the dashed line in Fig. 4.16. In the case of dijet, the high energy extrapolation uncertainty explained in the previous paragraph is added in quadrature to, and displayed together with, the residual uncertainty.

4.8 Zero-Suppression Bias Correction

As already discussed in Sect. 4.4.4 in the context of the offset correction, the application of zero-suppression has the "undesired" effect that regions in the calorimeter with different level of occupancy present different amount of visible offset energy: the higher the occupancy, the more offset energy becomes visible, as it is more likely for the cells to pass the zero-suppression threshold. In the case of a γ+jet event, the different effect of zero-suppression in the jet and photon hemispheres generates a small amount of E_T imbalance that biases (increases) the MPF response.

As done for the offset correction, a correction factor to the MPF response in data (MC) is estimated in MC making use of γ+jet samples where the same events are processed without ZB overlay and with unsuppressed (suppressed) ZB overlay. Events are selected exactly the same as for the $F_\eta^{\gamma+\text{jet}}$ measurement in MC, with the photon at the particle level. The correction factor is defined as

$$k_{\text{R}}^{\text{ZS}} = \frac{R_{\text{MPF},\eta}^{\gamma+\text{jet,noZB}}}{R_{\text{MPF},\eta}^{\gamma+\text{jet,ZB}}}. \tag{4.56}$$

Figure 4.17 shows the k_{R}^{ZS} correction factor in the central and forward regions for $R_{\text{cone}} = 0.5$ jets in the case of unsuppressed ZB overlay. The correction for $R_{\text{cone}} = 0.7$

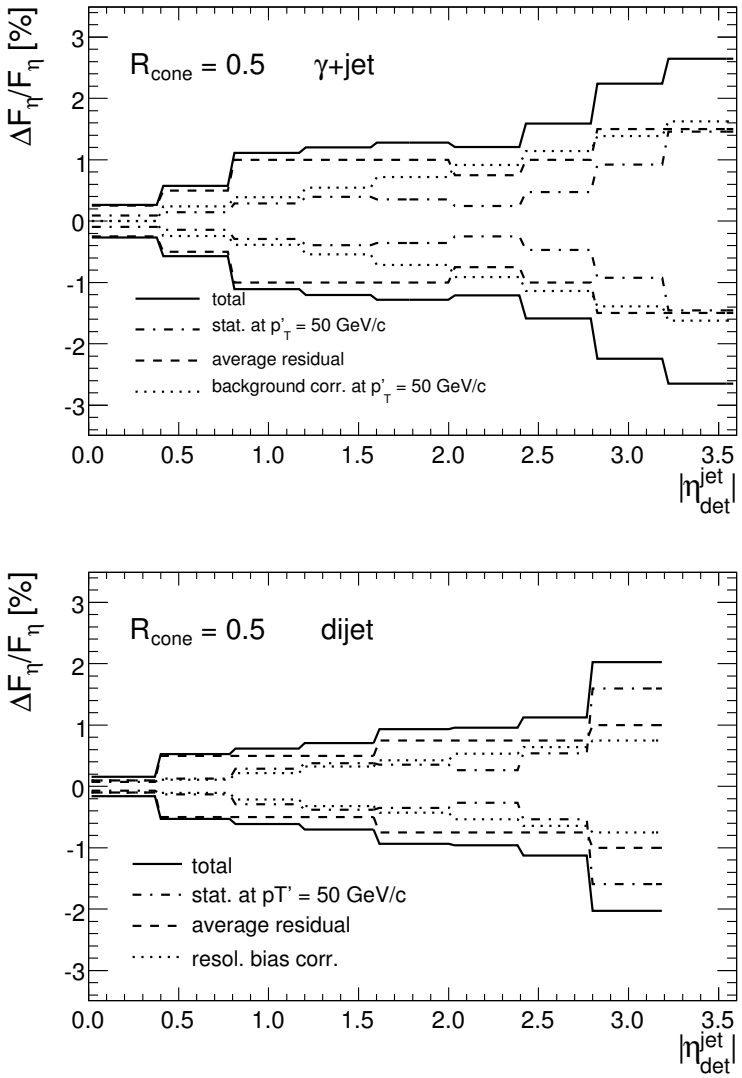

Figure 4.16 : Relative uncertainties on the γ+jet (up) and dijet (down) relative MPF response correction for $R_{\rm cone} = 0.5$ jets, as a function of $|\eta^{\rm jet}_{\rm det}|$. The solid line shows the total uncertainty, resulting from the sum in quadrature of the individual contributions: statistical (dashed-dotted), average residual (dashed) and background/resolution-bias correction (dotted).

jets is almost identical, as expected from the fact that the MPF method is not sensitive to the jet cone size.

The $k^{\rm ZS}_{\rm R}$ and $k^{\rm ZS}_{\rm O}$ correction factors account for biases in the MPF response and offset respectively that have the same origin, namely the effect of zero-suppression, and as such they are strongly correlated. Moreover, since both factors are similar in size, there is a large cancelation in the ratio $k^{\rm ZS}_{\rm O}/k^{\rm ZS}_{\rm R}$ as shown in Fig. 4.18, and in some of the

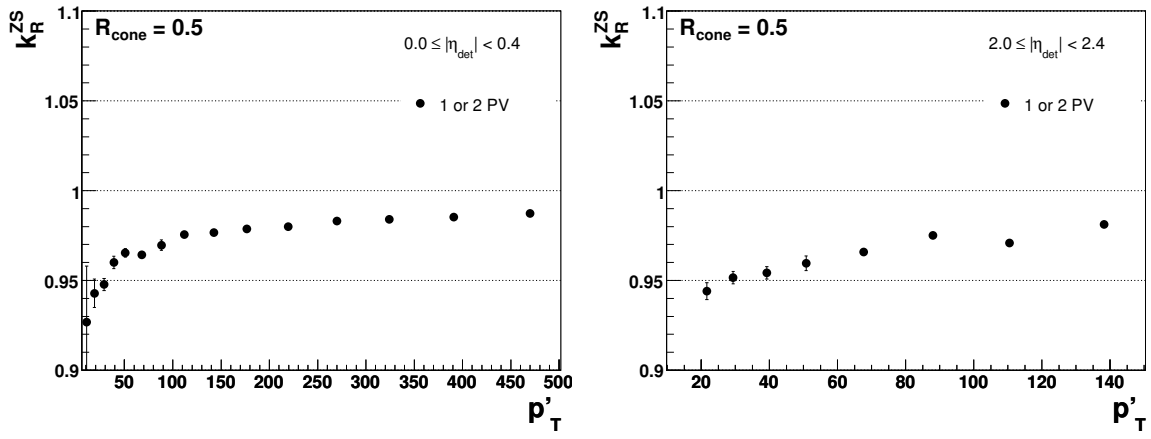

Figure 4.17 : $k_{\mathrm{R}}^{\mathrm{ZS}}$ correction factor as a function of p'_T for $R_{\mathrm{cone}} = 0.5$ jets with $|\eta_{\mathrm{det}}^{\mathrm{jet}}| < 0.4$ (left) and $2.0 < |\eta_{\mathrm{det}}^{\mathrm{jet}}| < 2.4$ (right) in unsuppressed ZB overlay γ+jet MC.

systematic uncertainties. Therefore, it is convenient to parameterize the ratio instead of both factors separately.

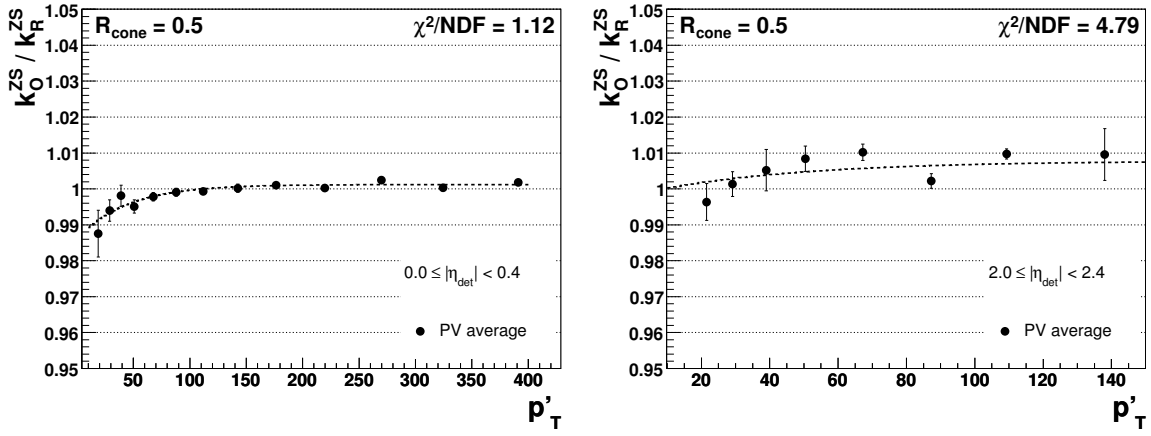

Figure 4.18 : $k_{\mathrm{O}}^{\mathrm{ZS}}/k_{\mathrm{R}}^{\mathrm{ZS}}$ correction factor as a function of p'_T for $R_{\mathrm{cone}} = 0.5$ jets in unsuppressed ZB overlay γ+jet MC.

The dominant uncertainties in the $k_{\mathrm{O}}^{\mathrm{ZS}}/k_{\mathrm{R}}^{\mathrm{ZS}}$ ratio are illustrated in Fig. 4.19 for very central jets, and correspond to those already described in Sect. 4.4.5 for the $k_{\mathrm{O}}^{\mathrm{ZS}}$ factor, plus the statistical uncertainty. The uncertainties for more forward jets are similar except that the narrowing of the N_{PV}, and therefore the total, error band occurs at higher values of p'_T (e.g. from $\sim$ 40 GeV/c at $|\eta_{\mathrm{det}}| = 0.0$ to $\sim$ 140 GeV/c at $|\eta_{\mathrm{det}}| = 2.0$).

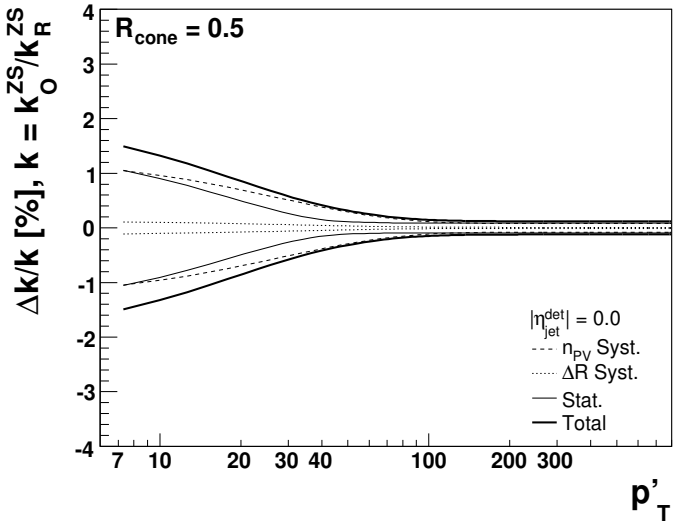

Figure 4.19 : Relative uncertainties on the $k_{\mathrm{O}}^{\mathrm{ZS}}/k_{\mathrm{R}}^{\mathrm{ZS}}$ correction factor as a function of p'_T for very central $R_{\mathrm{cone}} = 0.5$ jets in unsuppressed ZB overlay γ+jet MC. The two sources of systematic uncertainties, N_{PV} and ΔR matching, come from the $k_{\mathrm{O}}^{\mathrm{ZS}}$ factor exclusively. The statistical uncertainty is on the ratio $k_{\mathrm{O}}^{\mathrm{ZS}}/k_{\mathrm{R}}^{\mathrm{ZS}}$.

4.9 Topology Bias Correction

The last of the corrections applied to the MPF response has to do with the fact that the MPF method calibrates the energy, or more strictly speaking the p_T, of the whole hadronic recoil instead of the particle jet alone. The recoil may contain soft radiated jets that do not pass the 6 GeV/c reconstruction threshold, which could be of course in either the jet or photon hemisphere, lowering or raising the estimated MPF response respectively. Also, the difference between energy and p_T calibration is more important for low energy jets where the jet mass becomes non-negligible. Finally, there is also a rapidity bias that the MPF method could absorb, which tends to move reconstructed low (high) energy jets towards more forward (central) values of rapidity compared to the rapidity of the particle jet. All these effects that deviate the MPF response from the true response are accounted for in the following correction factor:

$$k_{\mathrm{R}}^{\mathrm{topo}} = \frac{R_{\mathrm{true},\eta}^{\gamma+\mathrm{jet,noZB}}}{R_{\mathrm{MPF},\eta}^{\gamma+\mathrm{jet,noZB}}}, \tag{4.57}$$

estimated in γ+jet MC without ZB overlay, selecting events with the same criteria as for the relative MPF response measurement in γ+jets, and with the photon at particle-level. The true jet response is estimated as the ratio of the average visible energy in the calorimeter from particles belonging to the particle jet matching the reconstructed jet within $\Delta R < R_{\mathrm{cone}}/2$, to the average particle jet energy. Figure 4.20 shows the

$k_{\mathrm{R}}^{\mathrm{topo}}$ correction for $R_{\mathrm{cone}} = 0.5$ jets. Such correction is also derived from MC using a tuned single pion response, to be applied specifically for data jets. The successive combination of the k_{R}^{γ}, $k_{\mathrm{R}}^{\mathrm{ZS}}$ and $k_{\mathrm{R}}^{\mathrm{topo}}$ correction factors brings the MPF response as close as possible to the true response correction.

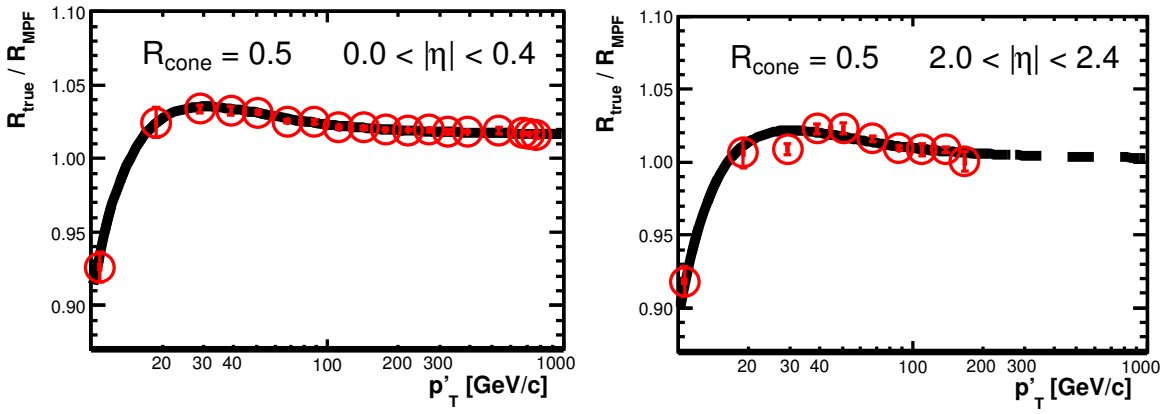

Figure 4.20 : Topology bias correction for $R_{\mathrm{cone}} = 0.5$ jets with $|\eta_{\mathrm{det}}^{\mathrm{jet}}| < 0.4$ (left) and $2.0 < |\eta_{\mathrm{det}}^{\mathrm{jet}}| < 2.4$ (right).

4.10 Showering Correction

Before going into the showering correction, it is important to distinguish between two different kinds of showering. A hard $p\bar{p}$ collision is modeled as a hard parton-parton collision, plus the interactions between the spectator partons (the underlying event) and initial-state radiation. A parton shower is then added to each outgoing parton from the $p\bar{p}$ collision, and the hadronization process is modeled. A reconstructed particle jet (using the cone algorithm) contains most of the particles originated from the parton, since some of the particles may escape from the jet cone. Moreover, particles originated from other sources may leak into the cone. A typical reason for this is large-angle gluon radiation producing particles that escape the reach of the reconstructed particle jet. This type of effect is called *physics showering*. On the other hand, particles originated from the underlying event may enter into the particle jet. The physics showering and the underlying event are responsible for the difference between the parton level and the particle level jet energies.

A second kind of showering due to only detector level effects (e.g. the bending of low-momentum particles in the magnetic field, the development of showers in the calorimeter and the finite granularity of the calorimeter) causes part of the energy of particles from inside (outside) the particle jet to be deposited outside (inside) the calorimeter jet cone boundaries. This is called *detector showering*, and it is taken into account neither by the *true offset* nor by the *true response* corrections. It is the goal

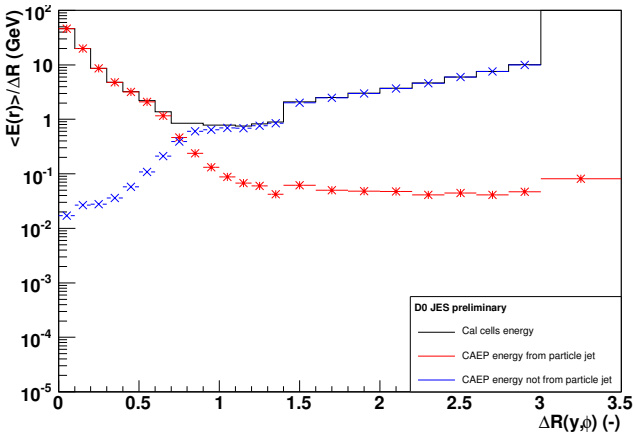

Figure 4.21 : Jet energy profile for cone 0.7 jets with $100 < p'_T < 130$ GeV/c and $|\eta^{\rm jet}_{\rm det}| < 0.4$ in γ+jet MC without ZB overlay. The red (blue) points show the contributions from particles (not) belonging to the particle jet matching the reconstructed jet. The discontinuities in $\Delta R = 1.4$ and $\Delta R = 3.0$ are due to changes in the bin size.

of the true showering correction, Eq. (4.6), to cancel the net energy flow across the jet boundaries produced by detector showering.

4.10.1 Methodology

The showering correction is determined in both data and MC, using back-to-back γ+(1)jet event candidates with $|\eta^{\gamma}_{\rm det}| < 1$ and $|\eta^{\rm jet}_{\rm det}| < 3.6$. The observable defined to measure the showering correction is the so-called "jet energy profile", which is the distribution of visible energy in the calorimeter in annuli of a given width at successive distances $\Delta R(y, \phi)$ with respect to the jet axis. Figure 4.21 shows as an example the average jet energy profile (black line) for central $R_{\rm cone} = 0.7$ jets in MC without ZB overlay. Shown also in the figure are the contributions from particles from inside (red points) and outside (blue points) the matching particle jet. This information can be obtained by tracking all the MC particles in their way through the detector.

As it can be observed, the energy profile from the particle jet falls steeply as one moves away from the jet axis; depositions at large ΔR are due to the bending of low-momentum jet particles in the magnetic field of the detector. Notice that the area covered by the rings grows with ΔR. The combination of this effect with a uniform spatial distribution of energy for not-jet particles, causes the increase of the not-particle-jet profile with ΔR seen in the figure. The not-particle-jet profile decreases towards the jet center, is non-zero everywhere, and has a non-negligible contribution near the jet boundary. The photon appears at $\Delta R \gtrsim 3.0$.

Given the distinguishing shapes of the particle-jet and not-particle-jet profiles, they can be used as templates to estimate the contributions from particles inside and outside the particle jet to the measured profile in data. This method assumes that the MC does a good job in modeling the jet profiles, including both the physics and the instrumental effects. In the case of data, a third profile is needed to model the offset. To obtain it, the jet profile measured in γ+jet MC without ZB overlay is subtracted from the equivalent jet profile measured in MC with unsuppressed ZB overlay. Moreover, to account for differences in the response between data and MC, two global factors, α and β, are allowed to rescale the particle-jet and not-particle-jet profiles respectively. Since the offset profile comes from ZB data events, no similar scale factor is used for it. A χ^2 fit is then performed (in the range $\Delta R < 2.5$) to estimate the scale factors α and β:

$$E_{\text{meas}}^{\text{data}} \equiv \alpha \times E_{\text{ptcl}}^{\text{MC-template}} + \beta \times E_{\text{not-ptcl}}^{\text{MC-template}} + E_{\text{offset}}^{\text{unsupZB-template}}. \tag{4.58}$$

Figure 4.22 compares the measured jet profile in γ+jet events in data to the fitted jet profile, for central (left) and forward (right) jets. The pink points represent the offset profile; the fact of seeing it grow outside the jet is a consequence of the MI part of the offset, which grows with $|\eta_{\text{det}}|$ (see Fig. 4.2 (right) and recall that events are selected with $1 \leq N_{\text{PV}} \leq 2$). The drop at large ΔR in the forward jet profile compared to the central one, reflects the reach of the detector limits at high rapidity. The predicted and measured profiles are in good overall agreement, except near the jet boundaries, $\Delta R = R_{\text{cone}}$. The main sources of discrepancy are the dijet background contamination and zero-suppression effects. A uniform uncertainty in $|\eta_{\text{det}}^{\text{jet}}|$ is assigned to cover for these discrepancies. Once α and β are determined, the estimator of the showering correction is given by:

$$\hat{S}_{\text{jet}} = \frac{\int_0^{R_{\text{cone}}} E_{\text{ptcl}}^{\text{MC-template}}}{\int_0^{\infty} E_{\text{ptcl}}^{\text{MC-template}}} + \frac{\beta \times \int_0^{R_{\text{cone}}} E_{\text{not-ptcl}}^{\text{MC-template}}}{\alpha \times \int_0^{\infty} E_{\text{ptcl}}^{\text{MC-template}}}. \tag{4.59}$$

The template-based method is only used for the JES in data. In the case of MC, the true showering correction, Eq. (4.6), is directly estimated using the γ+jet MC sample without ZB overlay. The numerator in Eq. (4.6) represents the measured energy of the reconstructed jet as given by the cone algorithm in absence of offset. For each jet, the denominator is the sum over particles from the matched particle jet of the visible energy they deposit in all calorimeter cells, also in absence of offset. The method used in MC also allows for a calibration of the estimator $\hat{S}_{\text{jet}}$ in data. The calibration factor is given by the ratio of the true showering correction calculated as explained above and the showering estimator $\hat{S}_{\text{jet}}$ calculated by replacing the jet profile from data by the jet profile from γ+jet MC with unsuppressed ZB overlay (in which case the fit returns $\alpha = \beta = 1$). The calibration factor is greater than (but very close to) 1 and typically less than 0.5%.

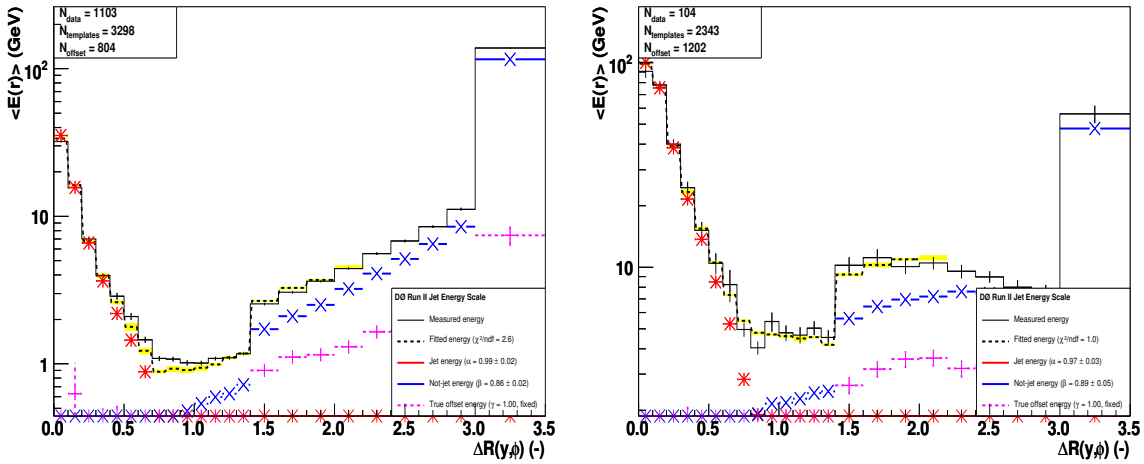

Figure 4.22 : Comparison of the measured jet profile in γ+jet data (solid line) to the fitted jet profile (dashed line) for jets with $80 < p'_T < 100$ GeV/c and two different $|\eta^{\rm jet}_{\rm det}|$ regions: $|\eta^{\rm jet}_{\rm det}| < 0.4$ (left) and $2.0 < |\eta^{\rm jet}_{\rm det}| < 2.4$ (right) in γ+jet data. The yellow band represents the statistical error from the fit.

4.10.2 Results

The showering corrections are fitted with a smooth parametrization as a function of p'_T and $|\eta^{\rm jet}_{\rm det}|$. Figure 4.23 shows the size of the corrections for $R_{\rm cone} = 0.5$ jets. For a given pseudorapidity bin, the size of the correction becomes smaller for higher energy jets, consequence of the narrowing of the jet bulk which implies less leaking out energy. The increase in the showering correction (more energy leaking out of the cone) for forward jets compared to central ones with same p'_T, is due to the smaller physical size of the forward jets compared to e.g. the radius of curvature of particles trajectories in the magnetic field. The results for $R_{\rm cone} = 0.7$ jets are closer to 1, because wider jets encompass more of the particles from the particle jet.

4.10.3 Uncertainties

Figure 4.24 shows the size of the relative uncertainties affecting the showering determination in data, for central $R_{\rm cone} = 0.5$ jets. The sources of each of these uncertainties and how they are estimated, are explained below:

- The statistical uncertainty from the fit of the showering correction as a function of $(p'_T, |\eta^{\rm jet}_{\rm det}|)$.
- A uniform uncertainty of 0.4% [46] is assigned to cover for the discrepancies seen in the template fits near the cone edges.

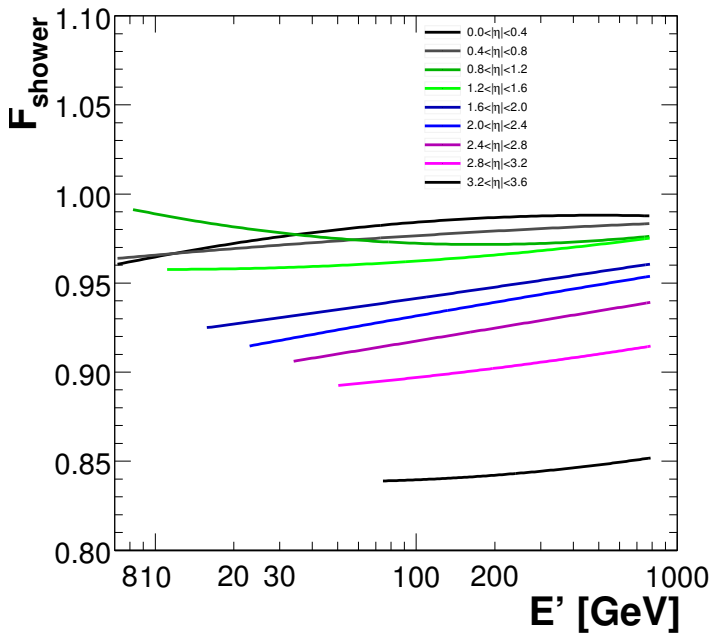

Figure 4.23 : Showering correction vs E' for $R_{\text{cone}} = 0.5$ jets in data in different 0.4-wide pseudorapidity bins.

- The presence of dijet background in the γ+jet data sample (predominantly at low p_T), effectively changes the average parton flavor composition of the jets, distorting the jet profile. To estimate an uncertainty that would account for this effect, two template fits are performed in MC, one in which the jet profile is derived from pure γ+jet MC and another one deriving it from γ+jet MC plus dijet (γ-like) MC mixed in the expected sample composition in data. In both cases, the templates are the same as those used for data, i.e. from pure γ+jet MC. The relative difference with respect to the pure γ+jet MC case is taken as the systematic uncertainty, labeled as "purity" in Fig. 4.24. This uncertainty dominates at very low p_T and goes quickly down to 0.1% at $p_T \approx 70$ GeV/c.

- Variations in the chosen ΔR matching criterion between the reconstructed and the particle jet may affect the MC template profiles. The particle-jet and not-particle-jet profiles are rederived varying the ΔR matching criterion to $\Delta R = R_{\text{cone}}/2 \pm 0.1$ and the corresponding template fits are performed again. The relative difference divided by the nominal case $\Delta R = R_{\text{cone}}/2$ is taken as the systematic uncertainty. The uncertainty is $\sim 0.5\%$ at low p_T and decreases down to 0.1% at $p_T \approx 80$ GeV/c.

- The single pion response is overestimated in MC compared to data, potentially biasing the showering correction. Since a scaling of the single pion response is

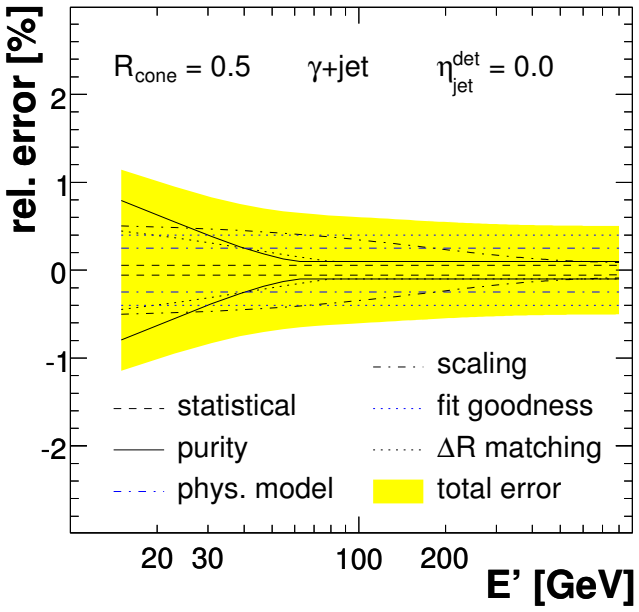

Figure 4.24 : Relative uncertainties on the showering correction vs E' for central $R_{\rm cone} = 0.5$ jets in data. The uncertainty labeled "purity" corresponds to the dijet background in data, the one labeled "phys. model" to the modeling of the underlying event, and the one labeled "scaling" to differences in the single pion response between data and MC.

expected to only affect the tails of the MC profiles not changing the result of the template fit, the systematic uncertainty is in this case estimated comparing the true showering correction, with and without scaling the single pion response down. The uncertainty is 0.5% at low p_T and decreases as a function of p_T, reaching 0.1% at $p_T \approx 500$ GeV/c.

- The modeling of the underlying event, affecting mainly the not-particle-jet profile, is based on Tune A of PYTHIA. A systematic uncertainty to account for possible dependences on the physics modeling of the underlying event, is estimated as the difference in the true showering correction between PYTHIA Tune DW and Tune A. The assigned uncertainty is 0.25% in $|\eta^{\rm jet}_{\rm det}| < 0.4$ increasing linearly up to 0.5% in $3.2 < |\eta^{\rm jet}_{\rm det}| < 3.6$.

4.11 Mapping the Measured Jet Energy to E'

The JES corrections are all derived as functions of the variable E', Eq. (4.18), which being directly related to the photon p_T, it represents an estimator of the jet energy at

parton level. This means that before applying the JES corrections to a measured jet in data, the E' value for that jet needs to be estimated. The $E_{\rm meas} \to E'$ mapping is performed by numerically solving the following equation:

$$E_{\rm jet}^{\rm meas} - \hat{E}_{\rm O} = E' S_{\rm jet}^{\rm phys}(E') \left(\frac{R_{{\rm MPF},\eta}^{\gamma+{\rm jet}}(E') k_{\rm R}^{\rm topo}(E') \hat{S}_{\rm jet}(E')}{k_{\rm O}^{\rm ZS}/k_{\rm R}^{\rm ZS}(E')} \right), \tag{4.60}$$

where the term $S_{\rm jet}^{\rm phys}$, defined as $E_{\rm jet}^{\rm ptcl}/E'$, is introduced to account for the fact that the JES corrects the energy up to the particle level only, while E' is strictly speaking a parton-level energy. According to the discussion at the beginning of Sect. 4.10, this factor represents then the physics showering. It is measured in γ+(1)jet MC events in 0.2-wide bins of $|\eta_{\rm det}^{\rm jet}|$. Figure 4.25 illustrates the physics showering for $R_{\rm cone} = 0.5$ jets. As expected, $S_{\rm jet}^{\rm phys} < 1$ (the raise at low p_T is supposed to be caused by biases due to the event selection and/or the 6 GeV reconstruction threshold for particle jets and therefore it is not parameterized). The correction decreases as the jet energy increases, because the jet becomes more collimated, and grows with $|\eta_{\rm det}^{\rm jet}|$ where it could be as large as 15%, because of the narrowing of the jet cone. Since the response depends only logarithmically with the energy, a precision of $\leq 5\%$ in the mapping is generally enough.

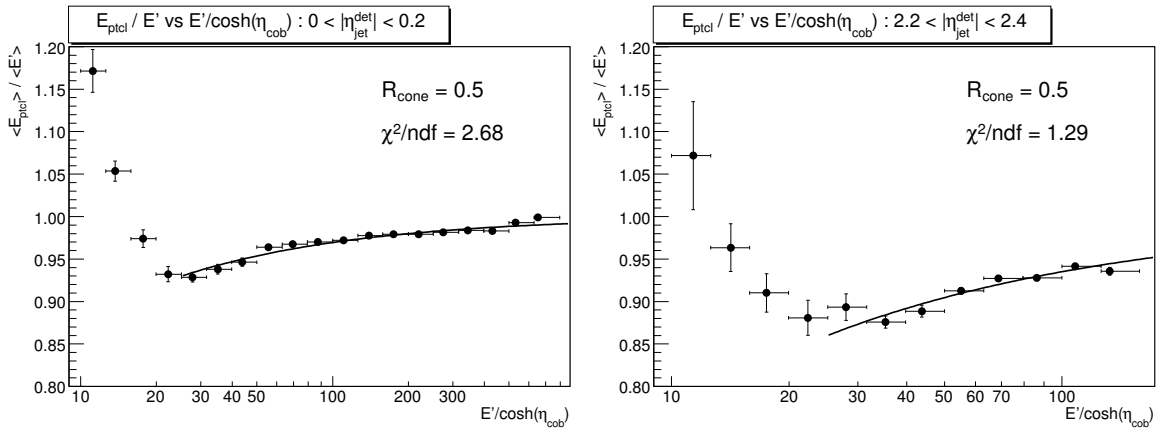

Figure 4.25 : Physics showering term $S_{\rm jet}^{\rm phys}$ as a function of $E'/\cosh(\eta_{\rm cob}) \approx p'_T$ ($\eta_{\rm cob}$ is the value of $\eta_{\rm det}$ at the center of the considered $\eta_{\rm det}$ bin) for $R_{\rm cone} = 0.5$ jets with $|\eta_{\rm det}^{\rm jet}| < 0.2$ (left) and $2.2 < |\eta_{\rm det}^{\rm jet}| < 2.4$ (right).

The offset energy is first subtracted to the measured jet energy to define a "detector level jet energy" ($E_{\rm jet}^{\rm det} \equiv E_{\rm jet}^{\rm meas} - \hat{E}_{\rm O}$). Equation (4.60) is then solved using the Newton method:

$$E_{n+1} = E_n - \frac{f(E_n)}{f'(E_n)}, \tag{4.61}$$

with

$$f(E) \equiv E \times \frac{R^{\gamma+\mathrm{jet}}_{\mathrm{MPF},\eta}(E) k^{\mathrm{topo}}_{\mathrm{R}}(E) \hat{S}_{\mathrm{jet}}(E)}{k^{\mathrm{ZS}}_{\mathrm{O}}/k^{\mathrm{ZS}}_{\mathrm{R}}(E)} \times S^{\mathrm{phys}}_{\mathrm{jet}}(E) - E^{\mathrm{det}}_{\mathrm{jet}}, \tag{4.62}$$

$E_{n=0} = E^{\mathrm{det}}_{\mathrm{jet}}$, and $f'(E)$ is obtained by means of the chain-rule and defining the derivative of each of the JES corrections as the relative variation of the correction between $(E + 0.01)$ and $(E - 0.01)$. The precision of the mapping is observed to be $< 3\%$ for jets with $p'_T > 30$ GeV/c.

4.12 Combined JES

This section summarizes the γ+jet JES corrections and uncertainties for $R_{\mathrm{cone}} = 0.5$, which is the one used in the single top analysis. Figure 4.26 (left) illustrates the magnitude of the JES subcorrections (offset, total response and showering) and the combination of them, as a function of the measured jet p_T for three different values of $\eta^{\mathrm{jet}}_{\mathrm{det}}$ in the range of interest for the single top analysis. Figure 4.26 (right) shows the corresponding uncertainties. Similarly, Figure 4.27 presents the JES corrections and associated uncertainties as a function of $\eta^{\mathrm{jet}}_{\mathrm{det}}$ for three fixed values of the measured jet p_T in the range of interest for the single top analysis (25, 50 and 100 GeV/c).

As already mentioned along this chapter, and as it can be observed in the figures, the response is by far the largest of the three JES subcorrections and thus it also dominates in the systematics. The dominant source of systematics at low p_T is the dijet background correction, while at high energies the statistical uncertainty on the CC-response becomes the largest one (this is for cone 0.5 jets; recall that for cone 0.7 jets this uncertainty has been enormously reduced to the same level as the second dominant source, the PhES). Outside the CC, the total JES uncertainty is dominated by the uncertainties on the η-dependent corrections residuals and statistical errors.

Within the central region ($|\eta^{\mathrm{jet}}_{\mathrm{det}}| < 0.8$) and for $p_T \lesssim 200$ GeV/c, the response has the same level of precision as the offset and showering corrections, bringing the total JES uncertainty to a level of $\sim 1\%$. In general, in a wide kinematic range the uncertainty in the (response and therefore in the) total JES stays below 2%. This has been a major achievement. In fact, it is the most precise JES attained in a hadron collider experiment, and physics analyses in DØ have largely benefited from it, the most relevant case being the inclusive jet cross section measurement [58].

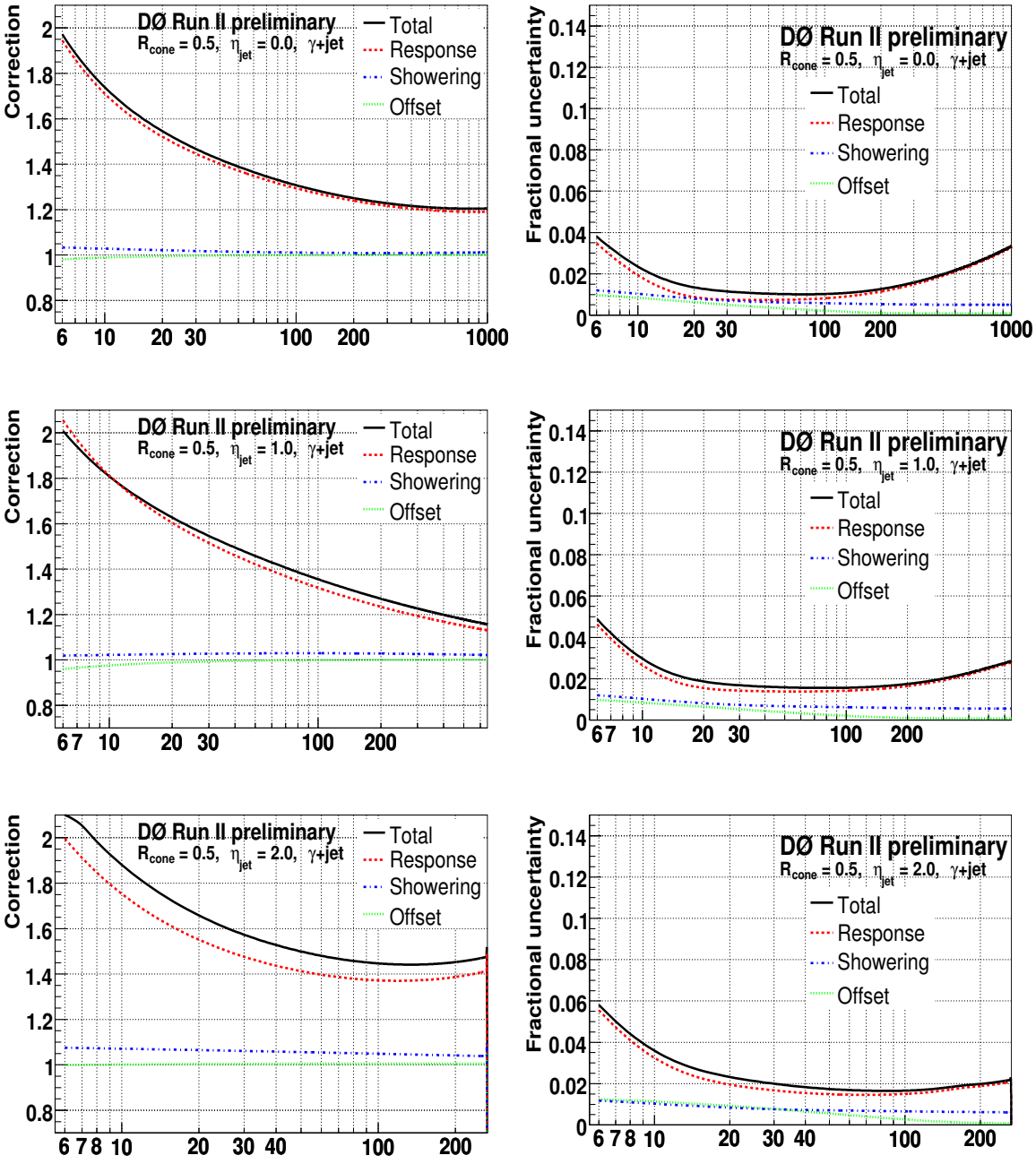

Figure 4.26 : Magnitude of the γ+jet JES subcorrections and their combination (left column), and the corresponding fractional uncertainties (right column), for $R_{\rm cone} = 0.5$ jets as a function of the uncorrected jet p_T for $\eta^{\rm jet}_{\rm det} = 0$ (first row), $\eta^{\rm jet}_{\rm det} = 1.0$ (second row) and $\eta^{\rm jet}_{\rm det} = 2.0$ (third row).

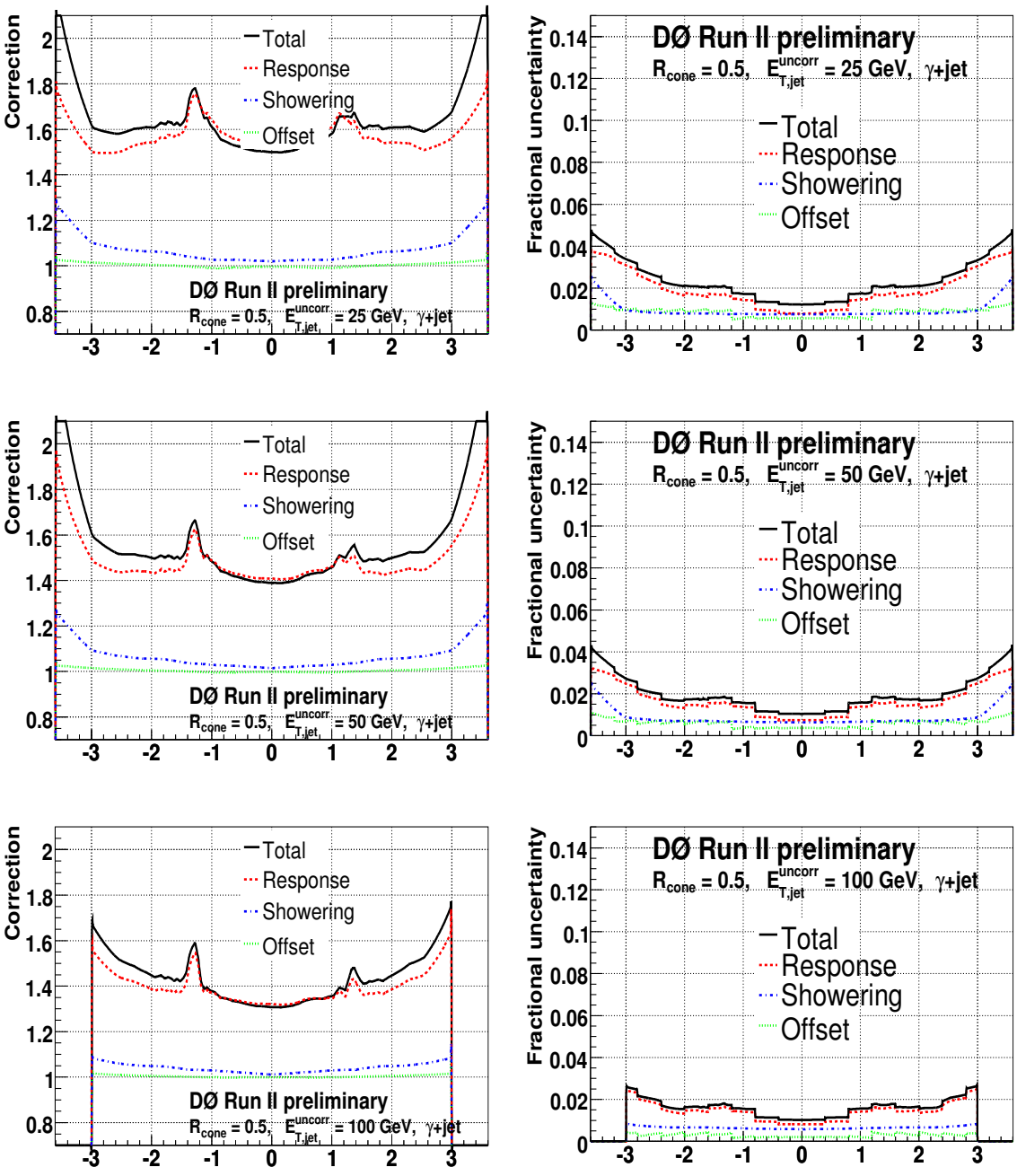

Figure 4.27 : Magnitude of the γ+jet JES subcorrections and their combination (left column), and the corresponding fractional uncertainties (right column), for $R_{\rm cone} = 0.5$ jets as a function of jet $\eta_{\rm det}$ for $p_{T\rm jet}^{\rm meas} = 25$ GeV/c (first row), $p_{T\rm jet}^{\rm meas} = 50$ GeV/c (second row) and $p_{T\rm jet}^{\rm meas} = 100$ GeV/c (third row).

5 Data and Simulation Samples

The signature of a single top event is unfortunately similar to other physics processes signatures like production of a W boson with a pair of b quarks or the production of a pair of top quarks with one of the W bosons from the tops decaying into $\ell\nu$ and the other one into jets. Physics processes that have the same final state (in what particle content refers) as the one from the signal, are called physics backgrounds. Processes that do not have the same final state as the signal, but mimic it at the detector level, are called instrumental backgrounds. To separate single top events from their backgrounds is not a task that can be done just with an appropriate event selection. More complicated techniques are needed to exploit the event kinematic differences between single top events and their backgrounds; in this thesis Bayesian Neural Networks are used. Clearly, a good model of all the backgrounds becomes absolutely relevant, which is essentially achieved by comparing kinematic distributions between data and the sum of all background models (and predicted signal when not negligible) at different stages of event selection.

This chapter presents the data and simulation samples used to perform the single top analysis, as well as the standard corrections applied to the simulated events to account for differences in data and MC efficiencies. It also explains the data-driven method used to model the multijets (instrumental) background.

5.1 Data Samples

The data samples used in this analysis were collected between August 2002 and August 2007, during the Run IIa and beginning of Run IIb Tevatron periods. After applying generic, standard good quality requirements to the data, the combined samples size in the electron and muon channels is reduced to 2.3 fb^{-1} of integrated luminosity, as shown in Table 5.1 [62]. The data have been reconstructed with production code versions p17.09.03 and p17.09.06 for Run IIa data, and p20.07.01 and p20.08.xx for Run IIb data (for this reason, data belonging to the Run IIa and Run IIb periods are also sometimes called p17 and p20 data respectively). They have been obtained from the Common Samples Group's "EMinclusive" and "MUinclusive" skims [63].

A general first step in the selection of a data sample is to require a trigger condition. For the single top analysis, a reasonable trigger requirement is an OR condition among the lepton+jets triggers (as actually done in the previous single top analysis [64]). Since triggers cannot be applied to simulated samples, their efficiency (turn-on) curves have to be first measured in data and then used to correct the simulated events. Moreover, the turn-on curves are selection-specific, since the trigger efficiency depends on the event kinematics. In order to increase the signal acceptance compared to the previous analysis, and to avoid possible biases from the mismodeling of the turn-on curves, the

		Integrated Luminosity [fb^{-1}]		
Channel	Trigger Version	Delivered	Recorded	Good Quality
Run IIa electron	v8.00 – v14.93	1.312	1.206	1.043
Run IIa muon	v8.00 – v14.93	1.349	1.240	1.055
Run IIb e and μ	v15.00 – v15.80	1.497	1.343	1.216
Total Run II Integrated Luminosity				**2.3**

Table 5.1 : Integrated luminosities for the datasets used in this analysis.

data samples used in this analysis are required to pass an OR condition among all DØ triggers related to photons, leptons and jets. More specifically, events in the electron (muon) channel are required to pass at least one of the EM, jet or EM+jets (muon, jet or μ+jets) DØ triggers. Only DØ triggers related to b-tagging, gap, and forward proton triggers (as well as EM triggers in the case of the muon channel, and muon triggers in the case of the electron channel) are not included in this "Mega-OR". The efficiency of the trigger Mega-OR is consistent, within a 5-10% uncertainty, with 100% for the single top selection (see Appendix B), and no correction is therefore needed for MC events.

5.2 Single Top Backgrounds

Single top backgrounds are listed below in order of their importance for the analysis. Only background processes that after event selection are not negligible in size compared to the statistical uncertainty on the data are considered.

- ***W*+jets**
  The W+jets backgrounds include among others $Wb\bar{b}$, $Wc\bar{c}$ and Wjj events, where j refers to a light jet (gluon, u, d or s). $Wb\bar{b}$ events ($p\bar{p} \rightarrow Wg + X \rightarrow \ell\nu b\bar{b} + X$) have exactly the same final state as single top, while the other W+jets background components enter into the data when a c jet or light jet is misidentified as a b jet. Although the b-tagging fake rate is small, the large production rate of W+light-jets compensates for it, and a non-negligible fraction of these events contaminates the data after event selection. W+jets are the dominant backgrounds at low jet multiplicity and largely dominate the preselected data sample before the identification of b jets.

- **Top pairs**
  $t\bar{t}$ processes with one of the W's decaying semileptonically and the other one into quarks ($p\bar{p} \rightarrow t\bar{t} + X \rightarrow WWb\bar{b} + X \rightarrow \ell\nu jjb\bar{b} + X$) have the same final state as single top with high jet multiplicity. $t\bar{t}$ processes with both W's decaying semileptonically are also a background due to lepton reconstruction (and identification) inefficiencies. The $t\bar{t}$ background dominates at high jet multiplicities.

- **Multijets**
  This is an instrumental background with a different origin in the electron and in the muon channels. In the electron channel, it is caused by a jet with a high electromagnetic content (π_0s) faking an isolated electron, and missing E_T artificially produced mainly by the non-calibration of the energy of the jet faking the lepton, plus the fluctuations from jet energy resolution. This multijets background also includes a small fraction of γ+jets events where the photon is the one that fakes an electron due to a track matching by chance with the EM cluster. On the other hand, in the muon channel, the multijets background arises from jets containing a real muon inside, coming from the decay of mainly B-hadrons and in a lesser extent charged pions, kaons and other particles. When the hadronic activity around the muon fluctuates down, the jet may not be reconstructed and the muon is reconstructed as isolated. The missing E_T has a similar origin as in the electron channel: the non-calibration of the energy of the jet that has not been reconstructed, plus the fluctuations from jet energy resolution.

- ***Z*+jets**
  Z+jets events with the Z boson decaying leptonically into $\ell^-\ell^+$ and one of the leptons not being reconstructed may fake single top signatures when in addition fake missing E_T gets created as it happens in a multijets background event. Z+jets are a relatively small background to single top.

- **Dibosons**
  Events producing two weak bosons (WW, WZ or ZZ), where one of them decays (semi)leptonically and the other one into quarks, conform the smallest of all considered single top backgrounds.

5.3 Simulation Samples

All the backgrounds except multijets are modeled using MC simulations. The ability of the detector simulation to correctly model a jet being reconstructed as an isolated electron, or in the case of the muon channel being not reconstructed at all, and especially to model the generation of a large amount of fake missing E_T, depends on the simulation of extreme fluctuations in the shower development. Moreover, since the multijets events that pass the analysis event selection represent just a tiny fraction of the total multijets events produced, a MC simulation of this background would require the generation of a huge sample. Consequently, multijets are modeled using a data-driven method that will be explained in Sect. 5.5. In this section, details about the simulated samples are presented.

5.3.1 Single Top Monte Carlo Samples

Single top events for this analysis are generated with the CompHEP-SingleTop [65] Monte Carlo event generator (version 4.2p1), which produces events whose kinematic distributions match those from NLO calculations. The top quark mass is set to 170 GeV/c^2, the PDF set is CTEQ6M [66], and the factorization scales are m_t^2 for the s-channel and $(m_t/2)^2$ for the t-channel. The top quarks and the W bosons from top quarks are decayed in CompHEP-SingleTop to ensure the spins are properly transferred. PYTHIA [67] (version 6.409) is used to add initial-state radiation and hadronization. TAUOLA [68] (version 2.5) is used to decay tau leptons, and EvtGen [69] (version 00-14-05) to decay b hadrons. Cross sections and branching fractions for both s- and t-channels are shown in Table 5.3 together with the number of events in the samples.

5.3.2 Background Monte Carlo Samples

The W+jets, Z+jets, and $t\bar{t}$ samples for this analysis are generated using ALPGEN [70] (version 2.11), which is a parton level event generator that uses exact LO matrix element calculations of the differential cross sections. The evolution of the partonic final state through a parton shower and hadronization is done by interfacing ALPGEN with PYTHIA, which also adds initial-state radiated jets. To avoid double counting of equivalent configurations, the partons generated by ALPGEN at Feynman diagram level are then matched with parton level jets from PYTHIA following the MLM procedure [71]. For the W+jets and Z+jets sets, events with heavy flavor jets added by PYTHIA are removed [72] so as not to duplicate the phase space of those generated already by ALPGEN.

The W+jets sample comprises three subsamples, $Wb\bar{b}$, $Wc\bar{c}$ and Wlp, where lp stands for "light partons". In this context, light parton means a gluon or a massless u, d, s or c quark. The Wlp sample includes Wjj and Wcj events ($j = g, u, d, s$). The Wcj events are separated from the Wjj ones when calculating number of predicted events and presenting plots. Each of these three W+jets samples are divided into subsamples according to the number of (additional) ALPGEN light partons in the final state. So, the Wlp sample is divided in $W + \mathrm{N}lp$ (N = 0, 1, 2, 3, 4, 5), and the $Wc\bar{c}$ sample is divided in $Wc\bar{c} + \mathrm{N}lp$ (N = 0, 1, 2, 3) as is the $Wb\bar{b}$ sample. In the subsamples with a parton multiplicity that is not the highest one, the exclusive MLM matching criteria is used, meaning that a one-to-one matching must be established between the Feynman diagram level partons in ALPGEN and the parton jets in PYTHIA. On the other hand, in the subsamples with the highest multiplicity, the inclusive MLM matching criteria is used, which allows more parton jets than Feynman diagram level partons.

For the W+jets events, the PDF set is CTEQ6L1, and the factorization scale is

$m_W^2 + \sum m_T^2$, where m_T is the transverse mass defined as $m_T^2 = m^2 + p_T^2$ and the sum $\sum m_T^2$ extends to all final state partons (including the heavy quarks, excluding the W decay products) [70]. The Wlp samples need parton level cuts on the light partons to avoid divergences in the cross section. These cuts are $p_T(lp) > 8$ GeV/c and $\Delta R(lp, lp) > 0.4$ for all massless partons (including the charm partons in these samples). For the W+heavy-flavor samples, there are no p_T nor ΔR cuts on the b or c partons, but additional light partons have the $p_T(lp) > 8$ GeV/c and $\Delta R(lp, lp) > 0.4$ cuts applied.

For the Z+jets samples, everything is similar as for W+jets, but the factorization scale is $m_Z^2 + \sum m_T^2$. Since Z+jets represents a much smaller background than W+jets, the light parton multiplicity does not need to be as high as in W+jets. The Z+jets subsamples are $Z + \mathrm{N}lp$ (N = 0, 1, 2, 3), $Zc\bar{c} + \mathrm{N}lp$ and $Zb\bar{b} + \mathrm{N}lp$ (N = 0, 1, 2).

Two $t\bar{t}$ samples are used, $t\bar{t} \to \ell\nu b\bar{b}jj$ and $t\bar{t} \to \ell\ell\nu\nu b\bar{b}$ ($t\bar{t} \to \ell$+jets and $t\bar{t} \to \ell\ell$+jets for short). As with W+jets and Z+jets, these two samples are divided in three subsamples each, with 0, 1, and 2 additional light partons: $t\bar{t} + \mathrm{N}lp \to \ell\nu b\bar{b} + (\mathrm{N}+2)lp$ and $t\bar{t} + \mathrm{N}lp \to \ell\ell\nu\nu b\bar{b} + \mathrm{N}lp$ (N = 0, 1 or 2). The top quark mass is set to 170 GeV/c^2 (as for the signal samples), the factorization scale is $m_t^2 + \sum p_T^2$(jets), and the PDF set is CTEQ6L1.

The single top and Z+jets samples have three separate subsamples for the electron, muon, and tau decay channels. For the W+jets and $t\bar{t}$ samples all lepton decay channels are generated together with approximately equal probabilities (the actual fractions follow the branching ratios).

The diboson samples (WW, WZ, and ZZ) are generated using PYTHIA, with inclusive decays.

All the MC events have the underlying event added using a PYTHIA simulation with DØ's version of Tune A. They are then processed through d0gstar (the simulation of the DØ detector) which determines how much energy is deposited in the active areas of the detector. The electronic response of the detector is simulated by a program called d0sim. Detector noise, pile-up and additional $p\bar{p}$ interactions are introduced in the simulation by overlaying ZB events through d0sim. This ZB sample takes into account the hardware baseline subtraction done by the DØ calorimeter electronics. Finally, MC events are reconstructed by d0reco as if they were real data.

5.3.3 Monte Carlo Samples Normalization

Each generated MC sample is normalized to the total integrated luminosity of the data ($\mathcal{L}_{\mathrm{int}}$) before the selection process is started using a sample dependent normalization

constant calculated as follows:

$$C_{\text{norm}} = \frac{\mathcal{L}_{\text{int}} \times (\sigma_{\text{theory}} B)}{N_{\text{events}}} \quad \text{for single top and diboson samples,} \tag{5.1}$$

$$= \frac{\mathcal{L}_{\text{int}} \times (\sigma_{\text{MC}} k B)}{N_{\text{events}}} \quad \text{for all other samples,} \tag{5.2}$$

where σ_{MC} is the total MC cross section of the subprocess in question (e.g. $p\bar{p} \to Wb\bar{b} + 2lp$), k is the NLO k-factor (the ratio between the NLO and the MC cross sections) for the encompassing process (e.g. $p\bar{p} \to Wb\bar{b}$), B is the branching ratio of the considered decay (e.g. $W \to \ell\nu$) and N_{events} is the total number of events in the corresponding MC sample (e.g. the $Wb\bar{b} + 2lp \to \ell\nu b\bar{b} + 2lp$ sample). Since single top and dibosons are generated inclusively (there are no subprocesses with different number of extra light partons), the theoretically calculated higher-order cross section (σ_{theory}) is directly used instead of calculating first a k-factor and using it then to multiply the MC cross section for each subprocess. Table 5.3 shows the NLO cross sections and branching ratios used in the analysis, and the number of events for the corresponding union of MC samples. The k-factors for the W+jets, Z+jets and $t\bar{t}$ samples are shown in Table 5.2. The cross sections for single top and top pairs are calculated in Refs. [4] and [73] respectively, and correspond to a top mass of 170 GeV/c^2. The diboson cross sections are calculated using the NLO event generator MCFM [74].

NLO K-Factors

	k-factors	
Sample	Run IIa	Run IIb
Wjj	1.30	1.30
Wcj	1.80	1.80
$Wc\bar{c}$	1.911	1.911
$Wb\bar{b}$	1.911	1.911
Zjj	1.346	1.30
$Zc\bar{c}$	1.6825	2.171
$Zb\bar{b}$	1.6825	1.976
$t\bar{t} \to \ell$+jets	1.42	1.42
$t\bar{t} \to \ell\ell$	1.36	1.36

Table 5.2 : K-factors used to scale the boson+jets and $t\bar{t}$ cross sections to NLO. For Z+jets the numbers are different for the two data-taking periods, because the version of ALPGEN used is different.

Because cross section calculations for W+jets production have sizeable uncertainties and are very sensitive to the renormalization and factorization scales, additional scaling is needed for this samples to equal the number of events observed in data. Thus, in addition to the normalization constants from Eq. (5.2) applied to the $Wjj + \text{N}lp$, $Wcj + \text{N}lp$, $Wc\bar{c} + \text{N}lp$ and $Wb\bar{b} + \text{N}lp$ subsamples separately, the W+jet sample as a

The Monte Carlo Event Sets

Event Type	Cross Section [pb]	Branching Ratio	N° of p17 Events	N° of p20 Events
Signals				
$tb \to \ell$+jets	$1.12^{+0.05}_{-0.12}$	0.3240 ± 0.0032	0.6M	0.8M
$tqb \to \ell$+jets	$2.34^{+0.13}_{-0.17}$	0.3240 ± 0.0032	0.5M	0.8M
Signal total	$\mathbf{3.46^{+0.18}_{-0.29}}$	$\mathbf{0.3240 \pm 0.0032}$	**1.1M**	**1.6M**
Backgrounds				
$t\bar{t} \to \ell$+jets	$7.91^{+0.61}_{-1.01}$	0.4380 ± 0.0044	2.6M	1.3M
$t\bar{t} \to \ell\ell$	$7.91^{+0.61}_{-1.01}$	0.1050 ± 0.0010	1.3M	0.9M
Top pairs total	$\mathbf{7.91^{+0.61}_{-1.01}}$	$\mathbf{0.5430 \pm 0.0054}$	**3.9M**	**2.2M**
$Wb\bar{b} \to \ell\nu b\bar{b}$	93.8	0.3240 ± 0.0032	2.3M	2.5M
$Wc\bar{c} \to \ell\nu c\bar{c}$	266	0.3240 ± 0.0032	2.3M	3.0M
$Wjj \to \ell\nu jj$	24,844	0.3240 ± 0.0032	21.0M	18.3M
***W*+jets total**	**25,205**	$\mathbf{0.3240 \pm 0.0032}$	**25.6M**	**23.8M**
$Zb\bar{b} \to \ell\ell b\bar{b}$	43.0	0.10098 ± 0.00006	1.0M	1.0M
$Zc\bar{c} \to \ell\ell c\bar{c}$	114	0.10098 ± 0.00006	0.2M	1.0M
$Zjj \to \ell\ell jj$	7,466	0.10098 ± 0.00006	3.9M	7.0M
***Z*+jets total**	**7,624**	$\mathbf{0.10098 \pm 0.00006}$	**5.1M**	**9.0M**
$WW \to X$	12.0 ± 0.7	1.0 ± 0.0	2.9M	0.7M
$WZ \to X$	3.68 ± 0.25	1.0 ± 0.0	0.9M	0.6M
$ZZ \to X$	1.42 ± 0.08	1.0 ± 0.0	0.9M	0.5M
Diboson total	$\mathbf{17.1 \pm 1.0}$	$\mathbf{1.0 \pm 0.0}$	**4.7M**	**1.8M**

Table 5.3 : The (NLO) cross sections, branching ratios, and initial numbers of events in the Monte Carlo event samples. The symbol ℓ stands for electron plus muon plus tau decays.

whole is normalized to data (together with multijets) at the pretag level of the selection, as it will be described in Sect. 5.5.3.

5.4 Corrections to Monte Carlo Events

Because of limitations in the modeling of the DØ detector, it is necessary to correct the default simulation output in order to match the performance observed in real data. The efficiencies to reconstruct, identify and select objects in the MC are higher than in data, so the following scale factors are used to correct for those differences:

- **Beam longitudinal shape**
  Because the beam shape in z in MC events differs from the shape measured in data, each MC event is reweighted by a factor that depends on the z position of the true MC primary vertex, the instantaneous luminosity of the overlaid data,

and the run period (Run IIa or Run IIb). The weight is derived by fitting the beam longitudinal shape in ZB events with the predicted shape of the luminous region in a collider with gaussian bunched beams [75]. The weight is a ratio between this fit function with parameters chosen according to the above input event parameters and a gaussian with $\sigma = 25$ cm.

- **Instantaneous luminosity**
  The instantaneous luminosity in MC events is read from the ZB overlay event. Since ZB data samples contain mostly low luminosity events[31], MC events are reweighted such that the instantaneous luminosity spectrum in MC matches those in Run IIa and Run IIb data.

- ***Z* p_T distribution**
  The Z p_T spectrum at the generator level given by ALPGEN is reweighted [76, 77] to match the observed spectrum in data as measured in [78].

- **EM ID efficiency**
  Monte Carlo events in the electron channel are corrected for efficiency differences between data and MC in electromagnetic cluster finding and electron identification. The scale factor is divided into two parts: preselection and post-preselection. The preselection part is related to the basic electron criteria that are common among many electron quality definitions, like ID and electromagnetic and isolation fractions. The preselection scale factor is parameterized in $\eta_{\rm det}$. The post-preselection factor refers to criteria that are unique for the particular electron quality definition used in this analysis, as the H-matrix cut, the track-matching requirements, and the likelihood cut. The post-preselection scale factor is parameterized in the two-dimentional plane of ($\eta_{\rm det}$, ϕ). Both factors are derived using $Z \to e^+e^-$ data and simulated events, as described in Ref. [79, 80]:

$$\varepsilon_{e-\rm ID} = \frac{\epsilon^{\rm Data}_{\rm Presel}}{\epsilon^{\rm MC}_{\rm Presel}} \times \frac{\epsilon^{\rm Data}_{\rm PostPresel}}{\epsilon^{\rm MC}_{\rm PostPresel}}.$$

- **Muon ID and isolation efficiency**
  MC events in the muon channel are corrected for differences in the MC and data in the muon ID efficiency, the track match, and the isolation. The ID scale factor is parameterized in ($\eta_{\rm det}$, ϕ); track match is parameterized in track-z and $\eta_{\rm CFT}$; and the isolation in η. Parametrizations of the correction factors are derived in $Z \to \mu^+\mu^-$ data and MC, as described in Ref. [42]:

$$\varepsilon_{\mu-\rm ID} = \frac{\epsilon^{\rm Data}_{\rm MediumID}}{\epsilon^{\rm MC}_{\rm MediumID}} \times \frac{\epsilon^{\rm Data}_{\rm TrackMatch}}{\epsilon^{\rm MC}_{\rm TrackMatch}} \times \frac{\epsilon^{\rm Data}_{\rm TightIso}}{\epsilon^{\rm MC}_{\rm TightIso}}.$$

[31] In Run IIa for example, zero-(un)suppressed ZB events were collected during early (late) runs of the run period, where the instantaneous luminosity is lower (higher). The default MC production overlays zero-suppressed ZB data; hence the lower instantaneous luminosity observed in MC compared to the data.

- **Jet shifting, smearing and removal**
  Corrections have to be applied to simulated jets to account for differences in their reconstruction and identification efficiencies compared to data, deviations of the relative jet energy scale from unity and the improved jet energy resolution in the simulation. Jets in the MC are recalibrated and oversmeared by modifying their transverse momentum as follows:

$$p_T^{\text{new}} = p_T^{\text{orig.}} + p'_T[C_{\text{shift}}(p'_T) + \text{Gauss}(0, C_{\text{smear}}(p'_T))], \tag{5.3}$$

  where p'_T is an estimate of the particle level jet p_T, C_{shift} is the relative jet energy scale, C_{smear} is the oversmearing factor given by the subtraction in quadrature of the jet p_T-resolutions in data and MC, and $\text{Gauss}(0, C_{\text{smear}})$ is a random number drawn from a Gaussian with zero mean and with standard deviation C_{smear}. These corrections are derived using real and simulated γ+jet events [81], using the true p_T^{γ} in place of p'_T. After shifting and smearing, a cut is applied on the jet p_T at 15 GeV/c, value at which full (reconstruction and identification) efficiency is reached both in data and simulation.

- **Taggability and *b*-tagging efficiency**
  A detailed explanation of the taggability derivation and how the tagging efficiencies are measured and then implemented as event weights is given in Sect. 6.3.

- ***W*+jets ALPGEN mismodeling**
  The leading log ALPGEN MC provides a good model of the W+jets background in general. However, discrepancies are observed in the jet η variables with respect to data. In order to correct this problem, the W+jets events are reweighted such that the sum of them and the other backgrounds matches the distributions in the pretag data. See Sect. 6.2 for more details.

5.5 Treatment of the Multijets Background

5.5.1 Modeling

The multijets background is modeled using the "orthogonal" datasets, where the events pass all selection cuts before b-tagging (see Sect. 6.1), but the lepton identification criteria are heavily loosened so that the sample becomes dominated by multijets events. More details are given below separately for the electron and muon channels.

Electron Channel

In the electron-channel orthogonal dataset, the EM object is not required to have a matching track as it is in the signal data. Moreover, to keep the multijets sample

orthogonal to the signal data, the EM-likelihood cut is inverted for EM objects with a track match.

Muon Channel

In the muon-channel orthogonal sample, no isolation requirements are made on the muon. In order to make the jets in these events match those in the signal data, any jets close to the muon are removed from the event, and missing E_T is recalculated to regain momentum balance.

5.5.2 Multijets Reshaping

Electron Channel

The change in the electron quality selection introduces a difference (bias) in the shape for some kinematic variables distributions compared to those of the multijets background in the signal data. Studies about the instrumental background in a generic electron+jets+$\not\!\!E_T$ final state analysis are presented in Appendix A, where a method to correct for these biases is proposed and in turn followed in this particular single top analysis. Here, a summary of the method is given and the effect of the correction shown.

One starts by considering a tight data sample (events with a "tight" electron) and a loose data sample (events with a "loose" electron) from which the tight sample is a subset. The number of events in each of these samples is denoted as N_{tight} and N_{loose} respectively. Two subsets are considered in both the loose and tight samples: events containing real electrons (mostly W+jets and $t\bar{t}$ events) and events containing fake electrons (mostly multijets events), with $N^{\text{real}\,e}_{\text{loose/tight}}$ and $N^{\text{fake}\,e}_{\text{loose/tight}}$ number of events respectively. The efficiencies for a real and a fake loose electron to pass the tight electron criteria are $\epsilon_{\text{real}\,e}$ and $\epsilon_{\text{fake}\,e}$. With these definitions one can write the following equations:

$$N_{\text{loose}} = N^{\text{fake}\,e}_{\text{loose}} + N^{\text{real}\,e}_{\text{loose}}, \tag{5.4}$$

$$N_{\text{tight}} = N^{\text{fake}\,e}_{\text{tight}} + N^{\text{real}\,e}_{\text{tight}} = \epsilon_{\text{fake}\,e}\, N^{\text{fake}\,e}_{\text{loose}} + \epsilon_{\text{real}\,e}\, N^{\text{real}\,e}_{\text{loose}}. \tag{5.5}$$

Taking the loose electron criterion as the one used for the multijets sample, one can use these two equations to estimate the multijets (and W+jets) contributions in the signal data sample. This is the standard application of these so-called "matrix method" equations [82]. The results are:

$$N^{\text{fake}\,e}_{\text{tight}} = \epsilon_{\text{fake}\,e}\, \frac{(N_{\text{loose}}\, \epsilon_{\text{real}\,e} - N_{\text{tight}})}{(\epsilon_{\text{real}\,e} - \epsilon_{\text{fake}\,e})}, \tag{5.6}$$

$$N^{\text{real}\,e}_{\text{tight}} = \epsilon_{\text{real}\,e}\, \frac{(N_{\text{loose}}\, \epsilon_{\text{fake}\,e} - N_{\text{tight}})}{(\epsilon_{\text{fake}\,e} - \epsilon_{\text{real}\,e})}. \tag{5.7}$$

A closer look into Eqs. (5.4) and (5.5) reveals the biases previously mentioned of the sample used to model the multijets background, which is represented by the "loose-but-not-tight" sample:

$$N_{\text{tight}}^{\text{fake}\,e} = \frac{\epsilon_{\text{fake}\,e}}{(1-\epsilon_{\text{fake}\,e})}\, N_{\text{loose-tight}} - \frac{\epsilon_{\text{fake}\,e}}{(1-\epsilon_{\text{fake}\,e})}\, \frac{(1-\epsilon_{\text{real}\,e})}{\epsilon_{\text{real}\,e}}\, N_{\text{tight}}^{\text{real}\,e}. \tag{5.8}$$

It turns out that the second term in (5.8) is less than 1% of the size of the first one and can be neglected. This difference arises in part because $N_{\text{loose-tight}}$ is approximately a factor of 20 larger than N_{tight} ($\sim$20M vs. $\sim$1M) and in part from almost a factor of 10 difference in the coefficients of the two terms. Hence,

$$N_{\text{tight}}^{\text{fake}\,e} = \frac{\epsilon_{\text{fake}\,e}}{(1-\epsilon_{\text{fake}\,e})}\, N_{\text{loose-tight}}. \tag{5.9}$$

The probability for a fake electron to pass the tight electron likelihood cut, $\epsilon_{\text{fake}\,e}$, is measured with data that pass all the selection cuts (before b-tagging) except the $\not\!\!E_T$ cut. Assuming that the low-$\not\!\!E_T$ region ($\not\!\!E_T < 10$ GeV) is largely dominated by multijet events, and that $\epsilon_{\text{fake}\,e}$ is independent of $\not\!\!E_T$, then $\epsilon_{\text{fake}\,e}$ is determined as the ratio of "tight" over "loose" events in the low-$\not\!\!E_T$ region. Figure 5.1 shows the fake rate as a function of the p_T of the reconstructed lepton, the variable where the most relevant kinematic dependence is found. The data sample used to measure the fake rate contains however a small amount of events with real electrons (principally W+jets and Z+jets) that becomes relatively more important for higher p_T values, causing the observed raise of the measured fake rate with increasing lepton p_T. This bias is small and is not corrected for, since it is covered by the $\geq 30\%$ uncertainty assigned to the multijets normalization (see Sect. 5.5.3). The sharp raise of the fake rate below 30 GeV/c responds to the fact that the tracking efficiency increases for lower p_T tracks and therefore it becomes more probable to find a track matching with the EM cluster.

As a consequence of the strong p_T dependence of $\epsilon_{\text{fake}\,e}$, the factor in front of $N_{\text{loose-tight}}$ in Eq. (5.9) is far from being flat, implying that the multijets model differs in shape from the multijets contribution in the signal data. To correct for this kinematic bias, events in the orthogonal data sample are reweighted with the factor $\epsilon_{\text{fake}\,e}/(1-\epsilon_{\text{fake}\,e})$ evaluated at the p_T of the reconstructed lepton, before using the dataset to model the multijets background in the analysis. This results in a shape changing that is very pronounced at low momentum and its effect is a much better description of the data in this region, as shown in Fig. 5.2.

Muon Channel

A comparison of kinematical distributions between the multijets background models in the previous single top analysis[32] and the current one is presented in Fig. 5.3. When

[32]In the previous analysis, the muon had to pass the $\Delta R(\text{muon}, \text{jet}) > 0.5$ cut and only fail the track isolation and energy isolation criteria.

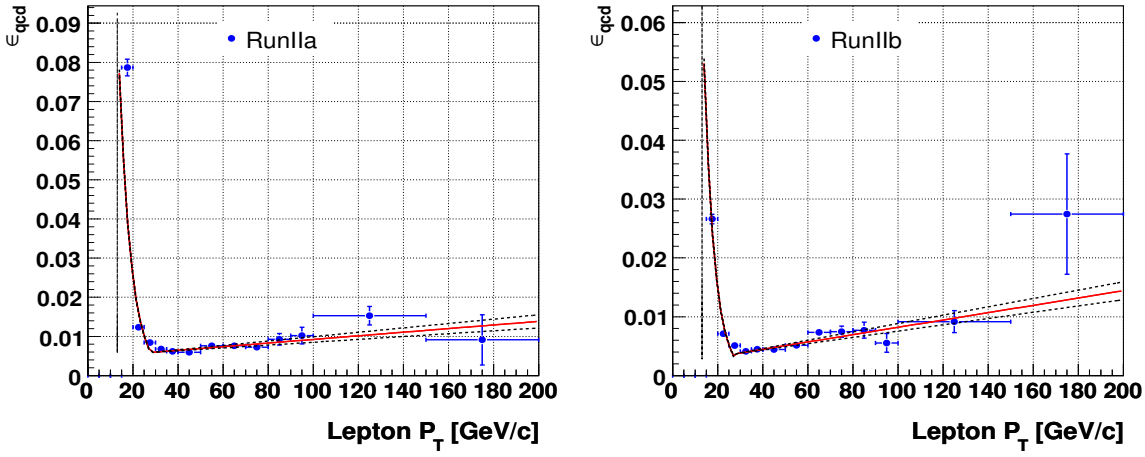

Figure 5.1 : $\epsilon_{\text{fake}\,e}$ as a function of electron p_T for Run IIa (left) and Run IIb (right) data. Since no significant differences have been observed for different jet multiplicities, $\epsilon_{\text{fake}\,e}$ is measured in events with at least two jets.

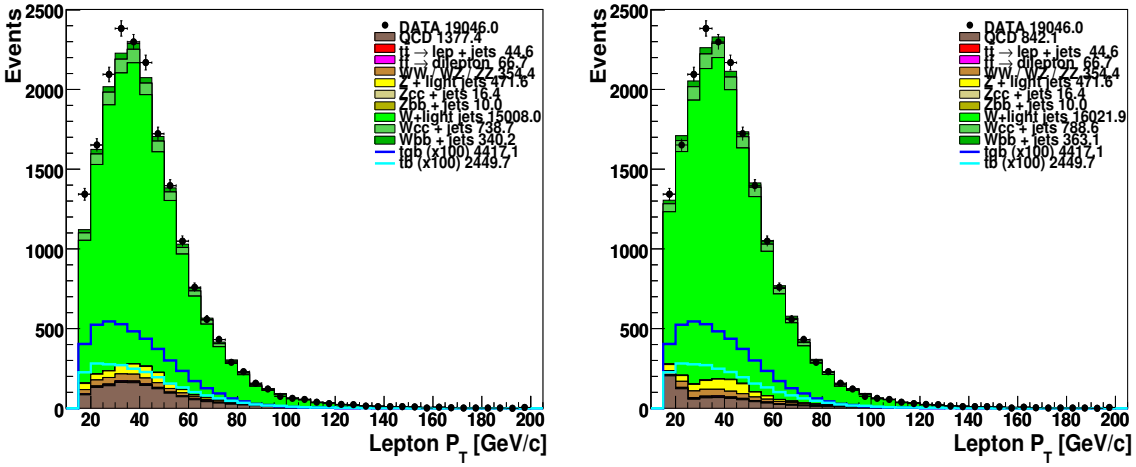

Figure 5.2 : Electron transverse momentum distribution before (left) and after (right) multijet background reshaping in events with 2 jets for the Run IIa period. The events entering in these plots are those that pass the selection described in Sects. 6.1.1 and 6.1.2 for the electron channel.

looking at these plots one has to keep in mind that both the old sample (labeled "Reverse isolation"), and the new sample (labeled "Large QCD") are only models of the true multijets background present in the signal sample. Given the large uncertainties on the multijets background normalization (∼30%), and the fact that the selection cuts are chosen so that the multijets background is a very small fraction of the data (∼7%), the similarities between the "Large QCD" and "Reverse Isolation" samples give confidence that the new multijets sample is as good a model for the true multijets

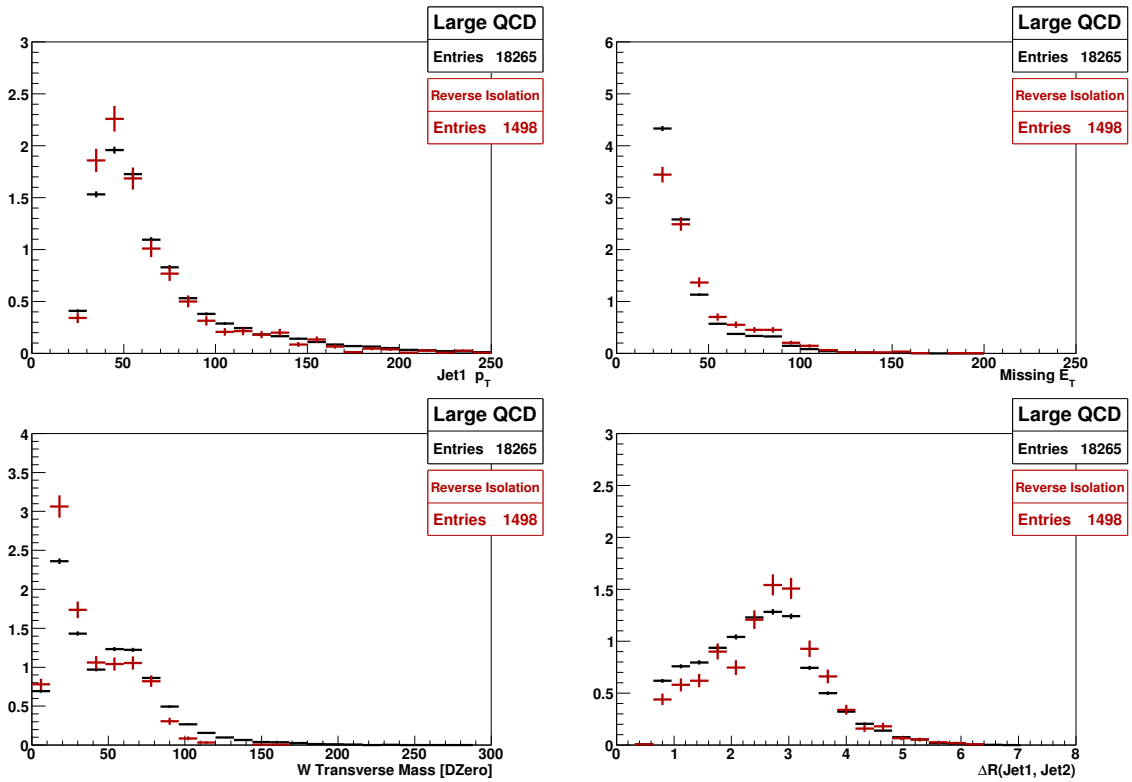

Figure 5.3 : Kinematic distributions comparing the old (red points) and new (black points) multijets background models. Upper row, p_T(jet1) and $\not\!\!E_T$, lower row, $M_T(W)$ and $\Delta R(\text{jet1}, \text{jet2})$.

contamination in the signal sample as the old sample.

5.5.3 Multijets and W+Jets Normalization to Data

As explained in Sect. 5.3.2, the W+jets MC samples are normalized using NLO calculations which have large variations with the factorization scale. Because of this, and together with the fact that W+jets is the largest single top background, this sample cannot be normalized using only theoretical cross sections (as are all other MC background components) and needs to be normalized by other means. Also, the instrumental background whose shape is derived from data, needs to be normalized. In the analysis described here, the sum of the W+jets and multijets backgrounds is normalized to data at the "pretag" stage of the event selection (before starting the b-tagging part described in Sect. 6.3), when lot of statistics is still available, the signal contribution is irrelevant, and the first data-background agreement check is performed to unfold any problem that could arise from the always complex b-tagging procedure.

The normalization condition is given by the following equation:

$$Y_{\text{pretag data}} = S_{W+\text{jets}} Y_{W+\text{jets}} + S_{\text{multijets}} Y_{\text{multijets}} + Y_{\text{all other bkgs}}, \tag{5.10}$$

where Y_{sample} is the "event yield" (sum of total event weights) in that sample (sample = pretag data, W+jets, multijets and all other backgrounds). The W+jets event yield in Eq. (5.10) includes all MC efficiency correction factors described in Sect. 5.4 and the scaling by the normalization constant from Eq. (5.2). The normalization factor $S_{W+\text{jets}}$ accounts therefore for higher order effects that are not covered by the k-factors. On the other hand, the muon channel multijets yield is just the number of pretag events, while the electron channel multijets yield includes the weight from the reshaping described later in Sect. 5.5.2.

The normalization factors $S_{W+\text{jets}}$ and $S_{\text{multijets}}$ are the quantities that need to be determined, and they are anticorrelated as can be seen in Eq. (5.10). The normalization method used in this analysis consists in choosing the $S_{W+\text{jets}}$ and $S_{\text{multijets}}$ factors that, while constrained to satisfy Eq. (5.10), give the maximum Kolmogorov-Smirnov (KS) test p-value (when comparing data against the sum of backgrounds) on a kinematic variable that has different shape distributions for W+jets and multijets events and is thus sensitive to their relative normalization. The KS test is not sensible to the normalizations of the two distributions being compared (data and sum of backgrounds), but only to their shape. To further stabilize the results, three such variables are used: p_T^{lepton}, $\not\!\!E_T$, and $M_T(W)$. The final pair $(S_{W+\text{jets}}, S_{\text{multijets}})$ is computed as the weighted average of the individual pairs $(S^i_{W+\text{jets}}, S^i_{\text{multijets}})$ that arise from each of the three variables, with the weights being the corresponding maximum KS test p-values (KS^i_{max}):

$$S_{W+\text{jets}} = \frac{\sum_{i=1}^{3} S^i_{W+\text{jets}} \times \text{KS}^i_{\text{max}}}{\sum_{i=1}^{3} \text{KS}^i_{\text{max}}}, \tag{5.11}$$

$$S_{\text{multijets}} = \frac{\sum_{i=1}^{3} S^i_{\text{multijets}} \times \text{KS}^i_{\text{max}}}{\sum_{i=1}^{3} \text{KS}^i_{\text{max}}}, \tag{5.12}$$

The scale factors are determined for each jet multiplicity bin (2, 3, 4 jets) separately. The final scale factors for W+jets and multijets are listed in Table 5.4. A clear trend is observed in the case of W+jets where the normalization scale factor increases with the jet multiplicity. This is reasonable since higher order corrections to the NLO cross sections become more important as more jets are present in the event. In the case of multijets, the bigger size of the electron channel sample and the reweighting (see

Sect. 5.5.2) of $\sim$ O(10^{-2}) applied to it, explain the one order of magnitude difference between electron and muon channel normalization factors.

Normalization Scale Factors from the KS Test Method

	$S_{W+\text{jets}}$				$S_{\text{multijets}}$			
Jet	Run IIa		Run IIb		Run IIa		Run IIb	
bin	e	μ	e	μ	e	μ	e	μ
2 jets	1.51	1.30	1.41	1.23	0.348	0.0490	0.388	0.0639
3 jets	1.92	1.79	1.75	1.57	0.291	0.0291	0.308	0.0410
4 jets	2.29	2.06	1.81	1.92	0.189	0.0244	0.424	0.0333

Table 5.4 : W+jets and multijets normalization scale factors from the KS test method.

The systematic uncertainties on the normalization scale factors, $\delta S_{W+\text{jets}}$ and $\delta S_{\text{multijets}}$, are estimated by taking the maximum variation between the scale factors from each variable separately and the final scale factors:

$$\delta S_{W+\text{jets}} = \max_{i=\{p_T^{lepton}, \not\!\!E_T, M_T(W)\}} |S^i_{W+\text{jets}} - S_{W+\text{jets}}|, \tag{5.13}$$

$$\delta S_{\text{multijets}} = \max_{i=\{p_T^{lepton}, \not\!\!E_T, M_T(W)\}} |S^i_{\text{multijets}} - S_{\text{multijets}}|. \tag{5.14}$$

The reason to define the uncertainties in this way, is that the width of the $S^i_{\text{multijets}}$ distribution obtained from the range of acceptable KS test p-values is smaller than the variations in $S^i_{\text{multijets}}$ between the three variables. A minimum value for the uncertainty is set separately for the W+jets and the multijets background to protect against low values due to fluctuations. This raises five of the multijets uncertainties to 30% and eight of the W+jets ones to 1.8%. The resulting uncertainties on the normalization scale factors from KS test method are shown in Table 5.5.

Normalization Scale Factor Uncertainties
from the KS Test Method

	$S_{W+\text{jets}}$				$S_{\text{multijets}}$			
Jet	Run IIa		Run IIb		Run IIa		Run IIb	
bin	e	μ	e	μ	e	μ	e	μ
2 jets	2.3%	1.8%	2.2%	1.8%	42%	40%	38%	53%
3 jets	1.8%	1.8%	1.8%	1.8%	30%	30%	30%	51%
4 jets	1.8%	1.8%	5.0%	3.9%	30%	30%	30%	54%

Table 5.5 : Uncertainties on the W+jets and multijets normalization scale factors from the KS test method.

5.6 Further Normalization of the W+Heavy-Flavor Samples after b-Tagging

A further heavy-flavor scale factor correction, S_{HF}, is needed to slightly adjust the ALPGEN W+heavy-flavor MC samples to achieve good data-background agreement *after* b-tagging (thus, the S_{HF} normalization factor is applied after the event selection described in Sect. 6.3). A method to calculate S_{HF} is explained in Ref. [83]. It consists of looking for the heavy-flavor scale factor correction that keeps good data-background agreement in a sample with a certain b-tagging requirement ($'$), as well as in a sample orthogonal to the former in the b-tagging requirement ($''$). The normalization condition for each of these two samples are:

$$Y'_{\rm data} = S \times (Y'_{W+{\rm lj}} + S_{HF} Y'_{W+{\rm hf}}) + Y'_{\rm all\,other\,bkgs} + Y'_{\rm sig}, \tag{5.15}$$

$$Y''_{\rm data} = S \times (Y''_{W+{\rm lj}} + S_{HF} Y''_{W+{\rm hf}}) + Y''_{\rm all\,other\,bkgs} + Y''_{\rm sig}, \tag{5.16}$$

where $Y_{\rm sample}$ denotes the event yield of the corresponding sample (with sample = data, W+light-jets (including W+1c-jet), W+heavy-flavor ($Wc\bar{c}$ and $Wb\bar{b}$), all other backgrounds, and signal). The yields $Y_{W+{\rm hf}}$ and $Y_{W+{\rm lj}}$ are the ones obtained after applying the ($'$) or ($''$) b-tagging requirement to the pretag samples with the reweightings (see Sect. 6.2) and pretag normalizations (see Sects. 5.3.3 and 5.5.3) included. Similarly, the multijets component in the yields $Y_{\rm all\,other\,bkgs}$ is obtained by applying the ($'$) or ($''$) b-tagging requirement to the pretag samples with the pretag normalization to data (see Sect. 5.5.3) included, as well as the reweighting (see Sect. 5.5.2) in the case of the electron channel. The two unknown variables are S_{HF} and S. The former is the one looked for and the latter is a necessary scale factor to keep the normalization agreement achieved before b-tagging. Solving for S_{HF} one gets:

$$S_{HF} = \frac{Y'_{W+{\rm lj}}(Y''_{\rm data} - Y''_{\rm all\,other\,bkgs} - Y''_{\rm sig}) - Y''_{W+{\rm lj}}(Y'_{\rm data} - Y'_{\rm all\,other\,bkgs} - Y'_{\rm sig})}{Y''_{W+{\rm hf}}(Y'_{\rm data} - Y'_{\rm all\,other\,bkgs} - Y'_{\rm sig}) - Y'_{W+{\rm hf}}(Y''_{\rm data} - Y''_{\rm all\,other\,bkgs} - Y''_{\rm sig})}. \tag{5.17}$$

The samples used to calculate this number are the ones containing two jets, since they are dominated by the W+jets background. Table 5.6 shows the derived heavy-flavor scale factor corrections. The weighted average of the two entries in the last column of Table 5.6 is used as the final S_{HF} value.

Systematic uncertainty contributions to S_{HF} have been evaluated. The only two relevant components are the systematic uncertainty due to the choice of the signal cross section value and the uncertainty due to the Wcj k-factor. Including them, the value for the heavy-flavor scale factor correction is (note that the values in Table 5.6 are rounded in the second decimal, and therefore there is no inconsistency with the final S_{HF} value):

$$\begin{aligned} S_{HF} &= 0.946 \pm 0.082\ ({\rm stat}) \pm 0.066\ ({\rm signal\text{-}xsec}) \pm 0.076\ (Wcj\ k\text{-factor}) \\ &= 0.95 \pm 0.13\ ({\rm stat} \oplus {\rm syst})\ (13.7\%). \end{aligned}$$

Heavy-Flavor Scale Factor Corrections S_{HF} for W+Jets Events

	Run IIa		Run IIb		Run IIa+b
	e	μ	e	μ	$e+\mu$
1 tag	1.04 ± 0.19	1.15 ± 0.20	0.95 ± 0.20	0.65 ± 0.18	0.95 ± 0.10
2 tags	0.84 ± 0.30	0.70 ± 0.29	1.23 ± 0.40	1.17 ± 0.40	0.97 ± 0.17

Table 5.6 : Heavy-flavor scale factor corrections for the two run periods and lepton types and the combination of them, calculated using two-jet events. Note that these values are not used directly in the analysis, but are used to calculate one combined scale factor, as explained in the text.

Systematic Uncertainty on $Wb\bar{b}$ and $Wc\bar{c}$ Yields from their K-Factor Uncertainties

$k_{Wc\bar{c}}/k_{Wb\bar{b}}$	S_{HF}	Deviation from Nominal
0.9	0.911 ± 0.079	-0.035 (-4%)
1.1	1.000 ± 0.087	$+0.055$ ($+6\%$)

Table 5.7 : Heavy-flavor scale factor corrections and their deviation from the value used in this analysis for different $Wc\bar{c}/Wb\bar{b}$ k-factor ratios.

The ratio of the $Wc\bar{c}$ to $Wb\bar{b}$ NLO k-factors is expected to be close to one. Table 5.7 shows the effect on the heavy-flavor scale factor correction S_{HF} when varying this assumption by $\pm 10\%$. The resulting variation in S_{HF} is not taken into account as an uncertainty on it, but instead it is applied as a 5% flat systematic anticorrelated on each of the $Wb\bar{b}$ and $Wc\bar{c}$ yields. The 5% anticorrelated systematic error is the one needed to get a 10% variation in the $Wc\bar{c}$ to $Wb\bar{b}$ NLO k-factors ratio (if one k-factor is decreased by 5%, the other is increased by 5%, since they are anticorrelated, so the variation in the ratio is $\sim 10\%$).

6 Event Selection

6.1 Event Selection Cuts before b-tagging

This section describes the selection applied to the data and MC samples to find W-like events containing an isolated lepton and missing transverse energy, plus two to four jets with high transverse momentum. In Sect. 6.1.1 the generic selection cuts are enumerated. Sections 6.1.2 and 6.1.3 describe additional cuts that are needed to reduce the multijets background in order to obtain a satisfactory agreement between the data and the background model. At this point there is no identification of heavy flavor (b) jets, so the samples after this selection are called "pretag". These samples are dominated by W+jets events, with some $t\bar{t}$ contribution that becomes more significant for higher jet multiplicities, and smaller contributions from multijets, Z+jets and diboson events. The b-tagging part of the selection is described later in Sect. 6.3.

6.1.1 Generic Cuts

This section enumerates the selection cuts applied to the data and MC samples to select W-like events with leptonic decay.

Common for both electron and muon channels

- Good quality (for data).
- Pass trigger requirement: at least one of the selected triggers has to fire (Sect. 5.1).
- Good primary vertex: $|z_{\mathrm{PV}}| < 60$ cm with at least three tracks attached.
- Exactly 2, 3 or 4 good jets with $p_T > 15$ GeV/c and $|\eta_{\mathrm{det}}| < 3.4$.
- The leading jet (jet1) is required to have $p_T > 25$ GeV/c.
- Missing transverse energy:
  - $20 < \not\!\!E_T < 200$ GeV in events with exactly 2 good jets;
  - $25 < \not\!\!E_T < 200$ GeV in events with 3 or 4 good jets.

Electron channel

- Only one 'tight isolated' electron with $|\eta_{\mathrm{det}}| < 1.1$ and $p_T > 15$ (20) GeV/c in events with 2 (3 or 4) good jets.
- No additional 'loose isolated' electron with $p_T > 15$ GeV/c and within $|\eta_{\mathrm{det}}| < 2.5$.

- No 'tight isolated' muon with $p_T > 15$ GeV/c and within $|\eta_{\rm det}| < 2.0$.
- Electron coming from the primary vertex: $|\Delta z(e, \mathrm{PV})| < 1$ cm.

Muon channel

- Only one 'tight isolated' muon with $p_T > 15$ GeV/c and $|\eta_{\rm det}| < 2.0$.
- No additional 'loose isolated' muon with $p_T > 4$ GeV/c and within $|\eta_{\rm det}| < 2.0$.
- No 'loose isolated' electron with $p_T > 15$ GeV/c and within $|\eta_{\rm det}| < 2.5$.
- Muon coming from the primary vertex: $|\Delta z(\mu, \mathrm{PV})| < 1$ cm.
- Transverse momentum of the leading jet within the ICD region of the detector: leading jet $p_T > 30$ GeV/c when $1.0 < |\eta_{\rm det}^{\rm jet1}| < 1.5$.

Selection of the orthogonal samples

These are data-driven samples for the modeling of the instrumental (multijets) backgrounds. All the selection cuts listed above are kept the same except for the lepton requirements, which are replaced with:

Electron channel

- Only one 'ultraloose' electron that fails the seven-variable EM-likelihood cut (i.e. it satisfies $\mathcal{L}_{EM} < 0.85$).
- No second electron of this type is allowed in any $|\eta_{\rm det}|$ region.

Muon channel

- The $\Delta R(\mu, \mathrm{jet}) > 0.5$ isolation requirement is dropped.
- Jets near the muon are removed and the $\not\!\!E_T$ of the event is corrected accordingly.
- The events must not be in the signal data sample.

The name "orthogonal" comes from the fact that these samples are selected such that they share no events with the signal data samples. The definitions of the orthogonal samples have been loosened with respect to previous single top analyses (see Ref. [64]) to gain in multijets statistics, given that after b-tagging there remained (in previous analyses) only a tiny amount of events, and this introduced a large statistical uncertainty. With the new loosened definitions, the statistics increases in a factor of ~ 25 (20) before (after) b-tagging.

6.1.2 Triangle Cuts

To reduce the huge amount of instrumental background left after the generic event selection described in Sect. 6.1.1, further cuts are applied in the two-dimentional planes $|\Delta\phi(\text{jet1},\not\!\!E_T)|$ vs. $\not\!\!E_T$ and $|\Delta\phi(\text{lepton},\not\!\!E_T)|$ vs. $\not\!\!E_T$ at low values of $\not\!\!E_T$. The dominant topology in QCD 3-jet events corresponds to the leading jet opposite in ϕ to the other two. In this case, if one of the jets is reconstructed as an isolated lepton, the missing transverse energy this generates results parallel to the lepton (as explained in Sect. 5.2) and antiparallel to the remaining highest p_T jet (and which is said to be the leading jet of the so reconstructed lepton+2jet event). Figure 6.1 shows the distributions of data, the multijets model and MC $tb + tqb$ events in the two-dimensional planes of $\Delta\phi(\text{jet1},\not\!\!E_T)$ vs. $\not\!\!E_T$ and $\Delta\phi(\text{lepton},\not\!\!E_T)$ vs. $\not\!\!E_T$ in the electron+2jet channel in Run IIa data, before the $\not\!\!E_T$ and the triangle cuts are applied. Equivalent plots are shown in Fig. 6.2 for the muon+2jet channel. The plots for Run IIb look similar. The triangle cuts are listed below and shown as white lines in the plots. Events inside the triangles are removed.

Common to both electron and muon channels

- $|\Delta\phi(\text{jet1},\not\!\!E_T)| < 1.5 + (\pi - 1.5) \times \not\!\!E_T/(35\text{ GeV})$ with $\not\!\!E_T \in [0, 35]$ GeV.

Electron channel

- $|\Delta\phi(e,\not\!\!E_T)| > 2.0 - 2.0 \times \not\!\!E_T/(40\text{ GeV})$ with $\not\!\!E_T \in [0, 40]$ GeV.
- $|\Delta\phi(e,\not\!\!E_T)| > 1.5 - 1.5 \times \not\!\!E_T/(50\text{ GeV})$ with $\not\!\!E_T \in [0, 50]$ GeV.
- $|\Delta\phi(e,\not\!\!E_T)| < 2.0 + (\pi - 2.0) \times \not\!\!E_T/(24\text{ GeV})$ with $\not\!\!E_T \in [0, 24]$ GeV.

Muon channel

- $|\Delta\phi(\mu,\not\!\!E_T)| > 1.2 - 1.2 \times \not\!\!E_T/(85\text{ GeV})$ with $\not\!\!E_T \in [0, 85]$ GeV.
- $|\Delta\phi(\mu,\not\!\!E_T)| < 2.5 + (\pi - 2.5) \times \not\!\!E_T/(30\text{ GeV})$ with $\not\!\!E_T \in [0, 30]$ GeV.

It can clearly be seen how the cuts surround the regions where the multijets background is mostly concentrated. In the third rows of plots, one can also see that the impact on the signal is very low. The upper left cut in the $\Delta\phi(\text{lepton},\not\!\!E_T)$ vs. $\not\!\!E_T$ plane does not seem to remove much of the multijets background, but it does remove the necessary amount that is needed to get data-background agreement at high values of $\Delta\phi(\text{lepton},\not\!\!E_T)$ in its one-dimensional distribution.

Further triangle cuts are necessary in the muon channel to reject events with poorly measured muons in the data that might mimic $\not\!\!E_T$, and that are difficult to reproduce in the background model. Cuts are applied to the muon track curvature significance for events where the muon and missing transverse energy are back-to-back in ϕ. Figure 6.3 shows the distribution of the absolute value of the track curvature significance vs.

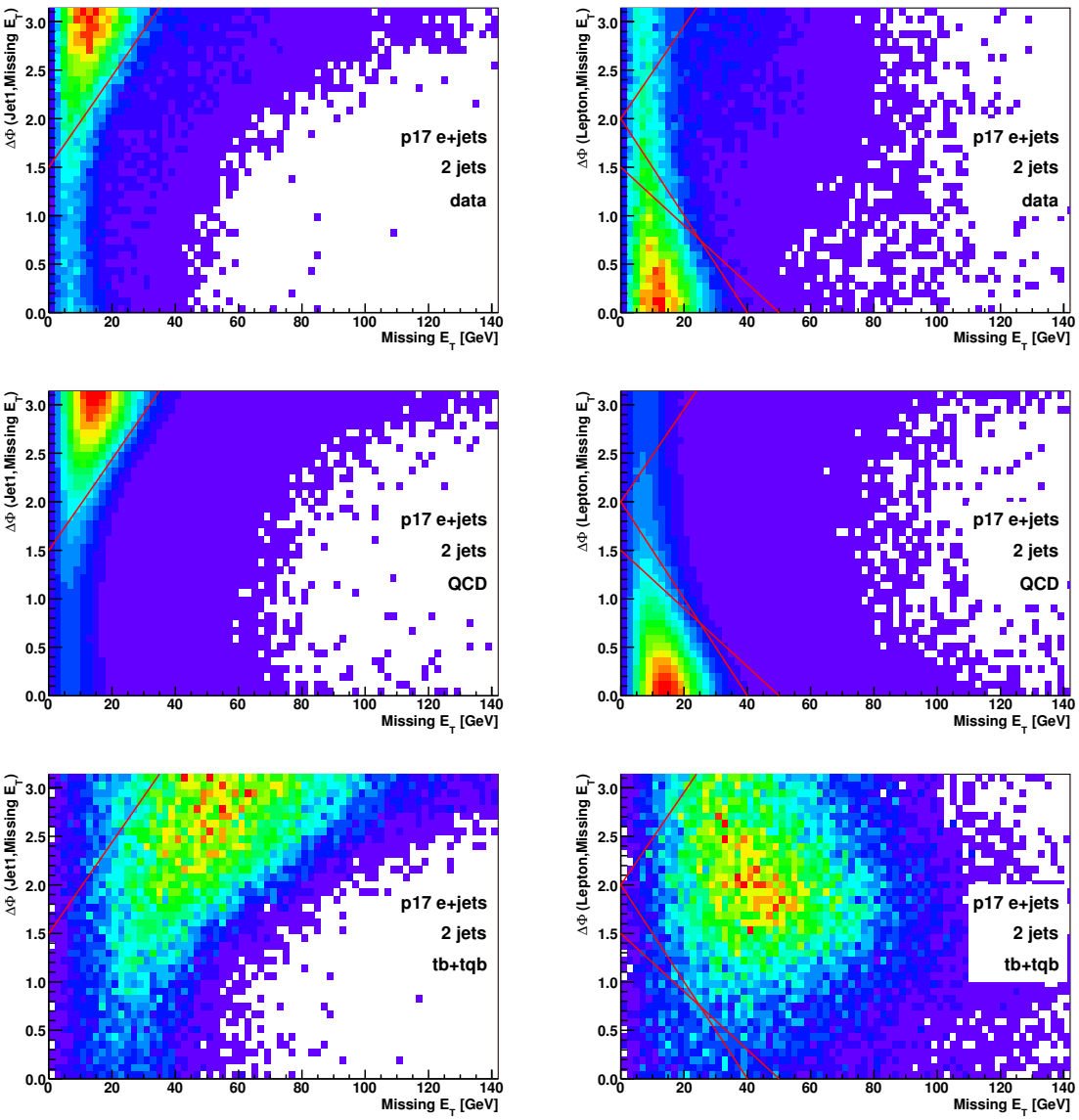

Figure 6.1 : $\Delta\phi$(jet1,$\not\!\!E_T$) vs. $\not\!\!E_T$ (left) and $\Delta\phi$(lepton,$\not\!\!E_T$) vs. $\not\!\!E_T$ (right) two-dimensional distributions for data (up), multijets (middle) and signal $tb + tqb$ (down), in the p17 electron channel with exactly two jets, before the $\not\!\!E_T$ cut. The triangle cuts are shown with red lines. Events inside the triangles (mostly multijets) are rejected.

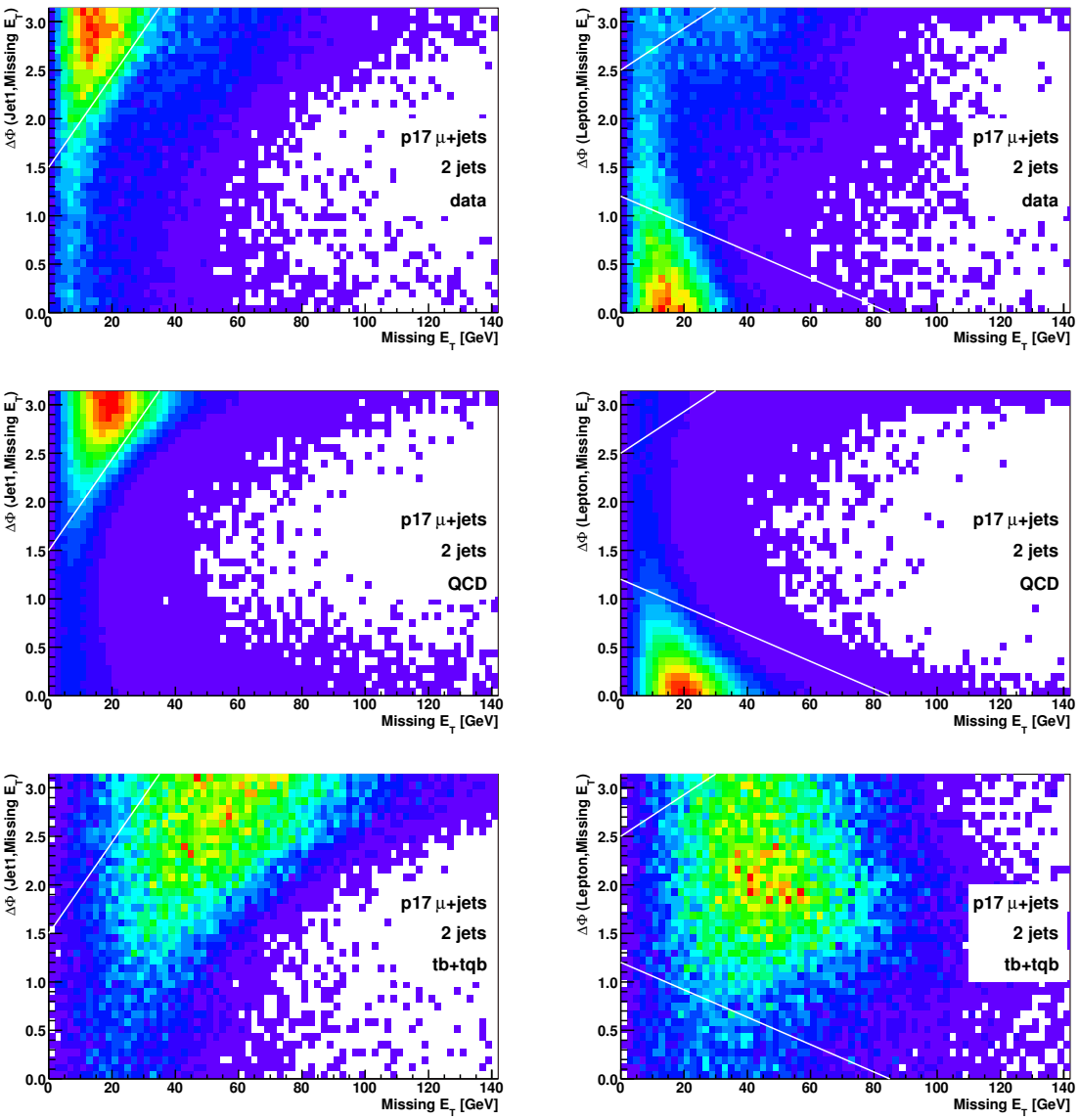

Figure 6.2 : $\Delta\phi$(jet1,$\not\!\!E_T$) vs. $\not\!\!E_T$ (left) and $\Delta\phi$(lepton,$\not\!\!E_T$) vs. $\not\!\!E_T$ (right) two-dimensional distributions for data (up), multijets (middle) and signal $tb + tqb$ (down), in the p17 muon channel with exactly two jets, before the $\not\!\!E_T$ cut. The triangle cuts are shown with red lines. Events inside the triangles (mostly multijets) are rejected.

$\Delta\phi(\mu,\not\!\!E_T)$ and the position of the cuts, which are also listed below. Figure 6.4 shows the distributions of $\Delta\phi(\mu,\not\!\!E_T)$ and $\Delta\phi(\text{jet1},\not\!\!E_T)$. It can be seen that the excess of data at high $\Delta\phi(\mu,\not\!\!E_T)$ and at low $\Delta\phi(\text{jet1},\not\!\!E_T)$ is removed by these triangle cuts.

Muon channel

Muon track curvature significance cuts |TrackCurvSig(μ)| vs. $|\Delta\phi(\mu,\not\!\!E_T)|$, where |TrackCurvSig($\mu$)| = $|\frac{q/p_T}{\sigma(1/p_T)}|$, and q and p_T are the charge and transverse momentum of the charged track associated with the muon:

- $|\Delta\phi(\mu,\not\!\!E_T)| < 0.875\pi + 0.125\pi \times |\text{TrackCurvSig}(\mu)|/(4(6))$

  with |TrackCurvSig(μ)| $\in [0, 4(6)]$ for Run IIa (Run IIb).

- $|\Delta\phi(\mu,\not\!\!E_T)| < 2.0 + (\pi - 2.0) \times |\text{TrackCurvSig}(\mu)|/(2(3))$

  with |TrackCurvSig(μ)| $\in [0, 2(3)]$ for Run IIa (Run IIb).

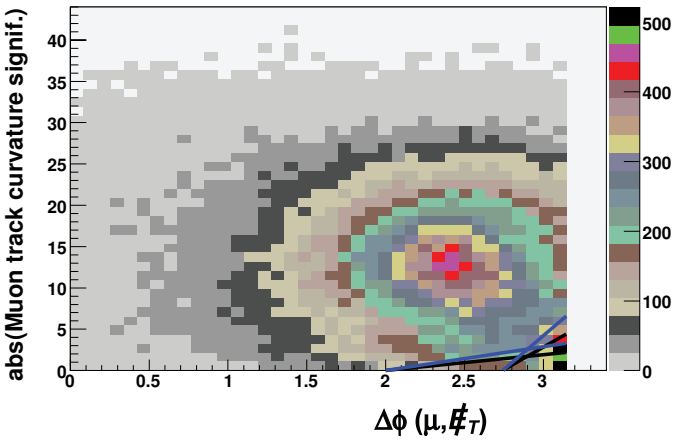

Figure 6.3 : The absolute value of the muon track curvature significance vs. $\Delta\phi(\mu,\not\!\!E_T)$. The cuts are shown as heavy black (blue) lines for Run IIa (Run IIb).

6.1.3 Total Transverse Energy Cuts

Even after applying the triangle cuts described in Sect. 6.1.2 and the kinematic reweightings that will be described in Sect. 6.2 to the W+jets MC samples, the agreement between the data and the background model at the pretag level is still not successful. The reason for that is again the presence of multijets events in kinematic regions that are difficult to model. To reject more multijets background, cuts on the total transverse energy of the event (H_T) are applied as listed below.

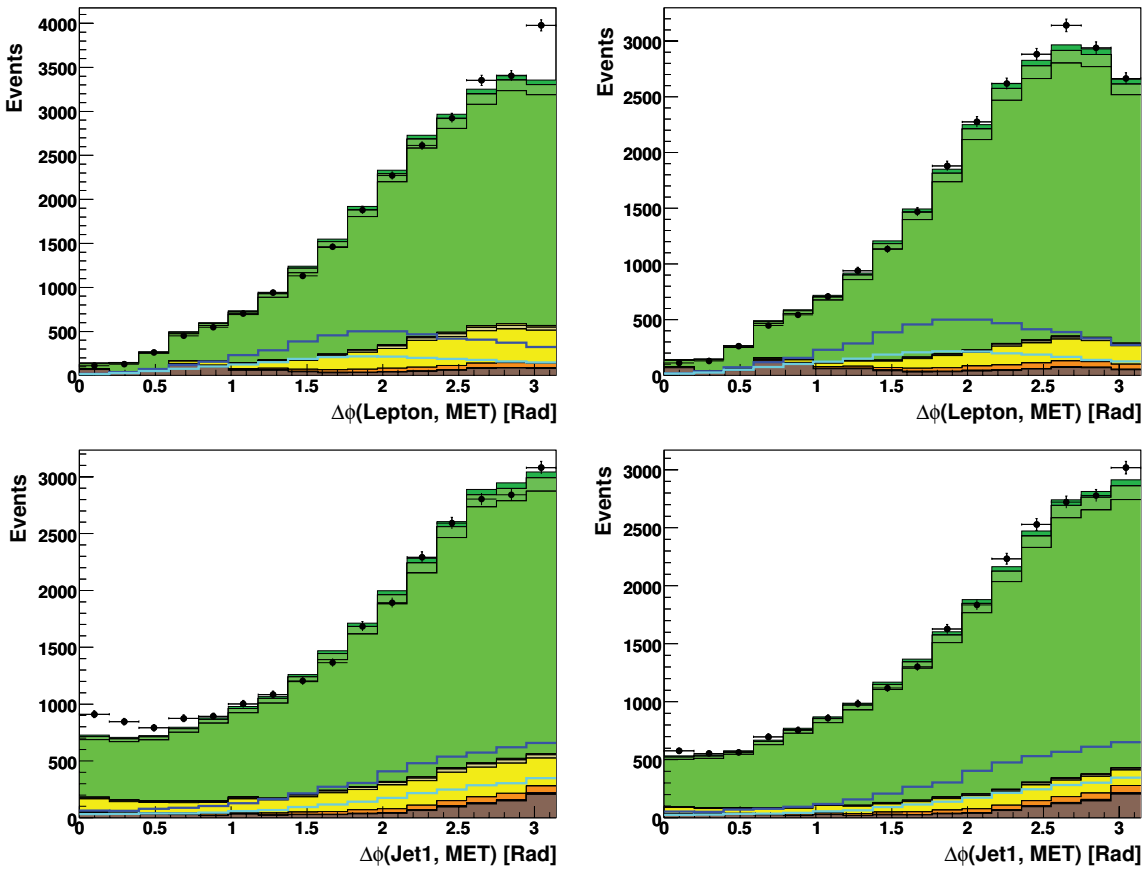

Figure 6.4 : Distributions of $\Delta\phi(\mu, \not\!\!E_T)$ (upper plots) and $\Delta\phi(\text{jet1}, \not\!\!E_T)$ (lower plots) in the p17 muon channel with exactly two jets before (left plots) and after (right plots) the muon track curvature significance triangle cuts.

Electron channel

$H_T(\text{alljets}, e, \not\!\!E_T)$ = scalar sum of the transverse energies of all good jets, the electron transverse momentum, and the missing transverse energy.

- $H_T(\text{alljets}, e, \not\!\!E_T) > 120$ GeV in events with exactly 2 good jets.
- $H_T(\text{alljets}, e, \not\!\!E_T) > 140$ GeV in events with exactly 3 good jets.
- $H_T(\text{alljets}, e, \not\!\!E_T) > 160$ GeV in events with exactly 4 good jets.

Muon channel

$H_T(\text{alljets}, \mu, \not\!\!E_T)$ = scalar sum of the transverse energies of all good jets, the muon transverse momentum, and the missing transverse energy.

- $H_T(\text{alljets}, \mu, \not\!\!E_T) > 110$ GeV in events with exactly 2 good jets.

- H_T(alljets, μ,$\not\!\!E_T$) > 130 GeV in events with exactly 3 good jets.
- H_T(alljets, μ,$\not\!\!E_T$) > 160 GeV in events with exactly 4 good jets.

Figure 6.5 shows the H_T distributions before the cuts are applied, together with the W boson transverse mass ($M_T(W)$) distributions before and after the H_T cuts, for the Run IIa electron channel (this is the channel where the H_T cuts have the biggest impact). The H_T cuts remove less than 2% of the single top signal events in each analysis channel, but almost half of the multijet background keeping it below $\sim 5\%$ in each channel.

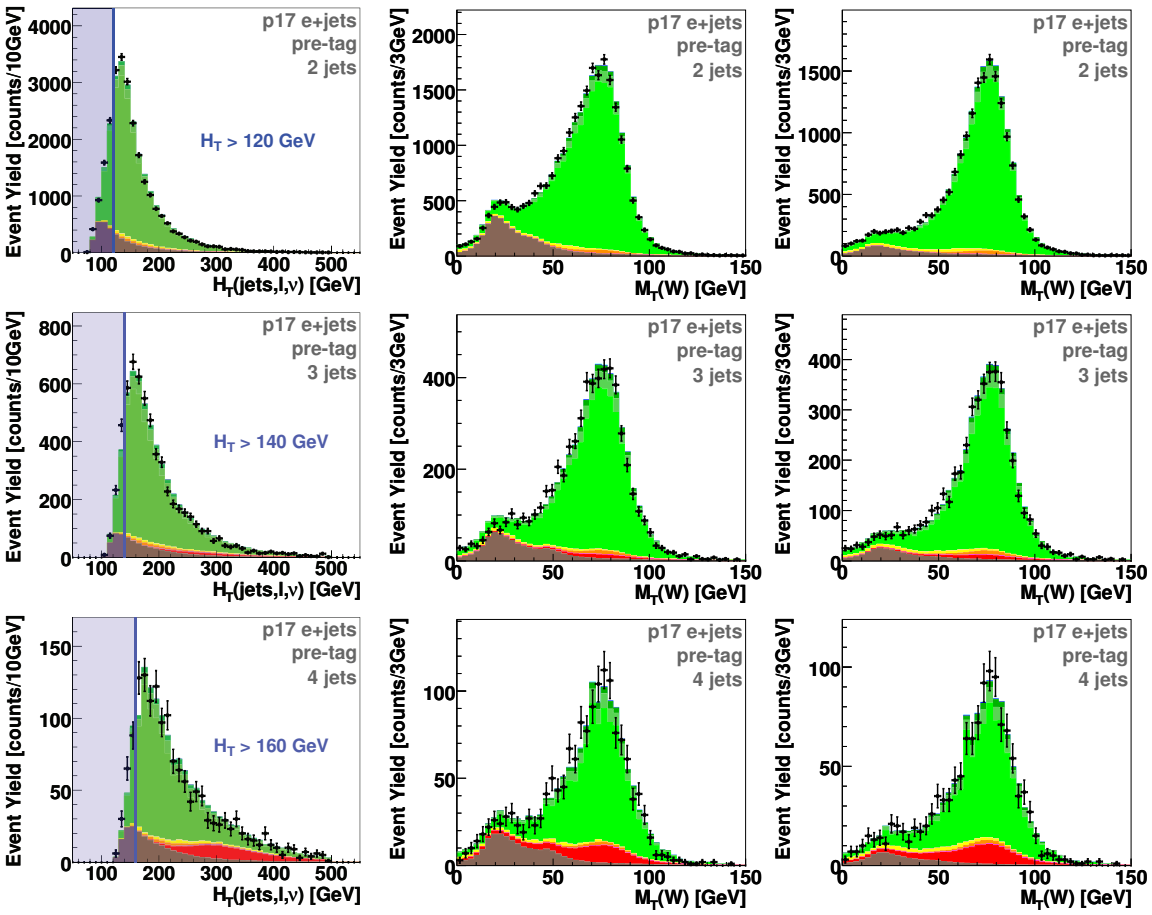

Figure 6.5 : Pretag distributions for the Run IIa electron channel in events with exactly two jets (first row), three jets (middle row) and four jets (bottom row). Left column: H_T(lepton, $\not\!\!E_T$,alljets) before the H_T cuts. Middle column: W boson transverse mass before the H_T cuts. Right column: W boson transverse mass after the H_T cuts.

6.1.4 Cut Flow Tables for MC Signal and Data

Tables 6.1, 6.2 and 6.3 show the number of events for MC signal samples and for data after some of the selection steps explained in Sect. 6.1 above.

Cut Flow for tb Monte Carlo

Selection cut − e channel	Run IIa tb	(%)	Run IIb tb	(%)
Initial	400,000	(100)	600,000	(100)
ZeroLumi, duplicate, DQ	376,482	(94.1)	538,312	(89.7)
Jet selection	371,476	(92.9)	534,079	(89.0)
$p_T^{\text{jet1}} > 25$ GeV/c	360,830	(90.2)	519,723	(86.6)
Tight lepton ($p_T > 15$ GeV/c)	96,728	(24.2)	119,434	(19.9)
Second lepton veto	96,352	(24.1)	119,031	(19.8)
Veto muon	96,315	(24.1)	118,985	(19.8)
Vertex selection	95,093	(23.8)	118,028	(19.7)
$15 < \not\!\!E_T < 200$ GeV	89,039	(22.3)	110,229	(18.4)
First electron triangle cut	85,463	(21.4)	105,926	(17.7)
Second electron triangle cut	84,617	(21.2)	104,905	(17.5)
Third electron triangle cut	83,892	(21.0)	103,910	(17.3)
Jet triangle cut	81,906	(20.5)	101,291	(16.9)
Number of good jets cut	66,412	(16.6)	81,330	(13.6)
H_T(alljets, e,$\not\!\!E_T$) cut	65,276	(16.3)	79,997	(13.3)
$\not\!\!E_T > 20/25/25$ GeV	62,425	(15.6)	76,371	(12.7)
$p_T^{\text{electron}} > 15/20/20$ GeV/c	60,439	(15.1)	73,914	(12.3)
Selection cut − μ channel	Run IIa tb	(%)	Run IIb tb	(%)
Initial	400,000	(100)	600,000	(100)
ZeroLumi, duplicate, DQ	374,407	(93.6)	535,883	(89.3)
Jet selection	370,388	(92.6)	531,898	(88.7)
Tight lepton ($p_T > 15$ GeV/c)	103,850	(26.0)	142,162	(23.7)
Second lepton veto	102,579	(25.6)	140,502	(23.4)
Veto electron	101,835	(25.5)	139,890	(23.3)
$p_T^{\text{jet1}} > 25$ GeV/c	98,410	(24.6)	135,491	(22.6)
Vertex selection	96,858	(24.2)	133,046	(22.2)
$15 < \not\!\!E_T < 200$ GeV	91,124	(22.8)	124,976	(20.8)
First muon triangle cut (loose)	88,967	(22.2)	121,997	(20.3)
Second muon triangle cut	87,994	(22.0)	120,668	(20.1)
Jet triangle cut	85,024	(21.3)	116,697	(19.5)
Number of good jets cut	70,870	(17.7)	97,375	(16.2)
H_T(alljets, e,$\not\!\!E_T$) cut	70,342	(17.6)	96,713	(16.1)
$\not\!\!E_T > 20/25/25$ GeV	67,365	(16.8)	92,553	(15.4)
$p_T^{\text{jet1}} > 30$ GeV/c if in ICD	67,036	(16.8)	92,095	(15.4)
First muon triangle cut	63,592	(15.9)	87,395	(14.6)
TrackCurvSig(μ) cut	62,466	(15.6)	86,426	(14.4)

Table 6.1 : Number of MC tb events after each selection step.

Cut Flow for *tqb* Monte Carlo

Selection cut − e channel	Run IIa *tqb*	(%)	Run IIb *tqb*	(%)
Initial	375,000	(100)	600,000	(100)
ZeroLumi, duplicate, DQ	351,768	(93.8)	539,371	(89.9)
Jet selection	345,854	(92.2)	534,328	(89.1)
$p_T^{\text{jet1}} > 25$ GeV/c	332,850	(88.8)	516,360	(86.1)
Tight lepton ($p_T > 15$ GeV/c)	88,776	(23.7)	113,510	(18.9)
Second lepton veto	88,468	(23.6)	113,153	(18.9)
Veto muon	88,439	(23.6)	113,119	(18.9)
Vertex selection	87,274	(23.3)	112,246	(18.7)
$15 < \not\!\!E_T < 200$ GeV	81,709	(21.8)	104,923	(17.5)
First electron triangle cut	78,791	(21.0)	101,204	(16.9)
Second electron triangle cut	78,230	(20.9)	100,441	(16.7)
Third electron triangle cut	77,485	(20.7)	99,439	(16.6)
Jet triangle cut	75,669	(20.2)	97,002	(16.2)
Number of good jets cut	61,491	(16.4)	78,709	(13.1)
H_T(alljets, e,$\not\!\!E_T$) cut	60,162	(16.0)	77,086	(12.9)
$\not\!\!E_T > 20/25/25$ GeV	57,284	(15.3)	73,163	(12.2)
$p_T^{\text{electron}} > 15/20/20$ GeV/c	55,325	(14.8)	70,511	(11.8)
Selection cut − μ channel	Run IIa *tqb*	(%)	Run IIb *tqb*	(%)
Initial	350,000	(100)	599,250	(100)
ZeroLumi, duplicate, DQ	327,973	(93.7)	540,256	(90.2)
Jet selection	323,583	(92.5)	535,067	(89.3)
Tight lepton ($p_T > 15$ GeV/c)	89,649	(25.6)	141,100	(23.6)
Second lepton veto	88,757	(25.4)	139,636	(23.3)
Veto electron	88,129	(25.2)	138,793	(23.2)
$p_T^{\text{jet1}} > 25$ GeV/c	84,270	(24.1)	132,962	(22.2)
Vertex selection	83,019	(23.7)	130,531	(21.8)
$15 < \not\!\!E_T < 200$ GeV	77,994	(22.3)	122,623	(20.5)
First muon triangle cut (loose)	76,268	(21.8)	119,893	(20.0)
Second muon triangle cut	75,467	(21.6)	118,504	(19.8)
Jet triangle cut	72,972	(20.9)	114,569	(19.1)
Number of good jets cut	59,778	(17.1)	93,745	(15.6)
H_T(alljets, e,$\not\!\!E_T$) cut	59,146	(16.9)	92,903	(15.5)
$\not\!\!E_T > 20/25/25$ GeV	56,428	(16.1)	88,522	(14.8)
$p_T^{\text{jet1}} > 30$ GeV/c if in ICD	56,049	(16.0)	87,923	(14.7)
First muon triangle cut	53,573	(15.3)	84,103	(14.0)
TrackCurvSig(μ) cut	52,637	(15.0)	83,411	(13.9)

Table 6.2 : Numbers of MC *tqb* events after each selection step.

Cut Flow for Data

Selection cut − e channel	Run IIa Data	(%)	Run IIb Data	(%)
Initial	335,220,569	(100)	207,299,315	(100)
ZeroLumi, duplicate, DQ	266,288,632	(79)	178,635,715	(86)
Trigger	239,165,711	(71)	169,252,679	(82)
Tight lepton ($p_T > 15$ GeV/c)	2,536,008	(0.76)	2,082,467	(1.0)
Second lepton veto	2,483,784	(0.74)	2,039,680	(0.98)
Jet selection	1,266,682	(0.38)	1,024,900	(0.49)
$p_T^{\text{jet1}} > 25$ GeV/c	568,856	(0.17)	468,752	(0.23)
Veto muon	568,619	(0.17)	468,488	(0.23)
Vertex selection	536,645	(0.16)	429,777	(0.22)
$15 < \not\!\!E_T < 200$ GeV	200,025	(0.059)	169,597	(0.081)
First electron triangle cut	119,261	(0.035)	107,602	(0.051)
Second electron triangle cut	116,993	(0.034)	105,635	(0.050)
Third electron triangle cut	113,133	(0.033)	101,961	(0.049)
Jet triangle cut	100,936	(0.030)	92,973	(0.044)
Number of good jets cut	41,785	(0.012)	39,642	(0.019)
H_T(alljets, e,$\not\!\!E_T$) cut	30,310	(0.0090)	29,070	(0.014)
$\not\!\!E_T > 20/25/25$ GeV	27,567	(0.0082)	26,291	(0.013)
$p_T^{\text{electron}} > 15/20/20$ GeV/c	26,847	(0.0080)	25,595	(0.012)
Additional duplicate removal	24,662	(0.0074)	25,595	(0.012)
Selection cut − μ channel	Run IIa Data	(%)	Run IIb Data	(%)
Initial	330,306,915	(100)	352,185,449	(100)
ZeroLumi, duplicate, DQ	270,187,161	(82)	303,612,626	(86)
Trigger	247,589,750	(75)	265,892,437	(76)
Jet selection	170,537,062	(52)	187,509,219	(53)
Tight lepton ($p_T > 15$ GeV/c)	919,420	(0.28)	774,158	(0.22)
Second lepton veto	906,444	(0.27)	761,494	(0.22)
Veto electron	903,054	(0.27)	758,603	(0.22)
$p_T^{\text{jet1}} > 25$ GeV/c	386,770	(0.13)	339,038	(0.096)
Vertex selection	357,389	(0.12)	300,569	(0.085)
$15 < \not\!\!E_T < 200$ GeV	192,933	(0.058)	169,964	(0.048)
First muon triangle cut (loose)	142,940	(0.043)	131,555	(0.037)
Second muon triangle cut	138,612	(0.041)	127,700	(0.036)
Jet triangle cut	118,931	(0.036)	112,164	(0.031)
Number of good jets cut	49,649	(0.015)	48,963	(0.015)
H_T(alljets, μ,$\not\!\!E_T$) cut	41,219	(0.013)	41,027	(0.012)
$\not\!\!E_T > 20/25/25$ GeV	38,291	(0.012)	38,176	(0.012)
$p_T^{\text{jet1}} > 30$ GeV/c if in ICD	36,962	(0.011)	36,960	(0.011)
First muon triangle cut	34,297	(0.010)	34,357	(0.010)
TrackCurvSig(μ) cut	31,957	(0.0097)	32,939	(0.010)
Additional duplicate removal	31,581	(0.0096)	32,939	(0.010)

Table 6.3 : Number of data events after each selection step.

6.2 W+Jets Samples Reweightings

Although the ALPGEN MC samples provide a good shape model of the W+jets background in general, certain variables exhibit some discrepancies with data. In particular, the W+jets ALPGEN jet η distributions tend to be narrower. In order to correct for it, the W+jets events are reweighted such that its sum to the other backgrounds matches the distributions in pretag data. The distributions that are reweighted are (in this order): $\eta_{\rm det}(\mu)$ (muon channel only), η(jet1), η(jet2), $\Delta\phi$(jet1,jet2), $\Delta\eta$(jet1,jet2), η(jet3) and η(jet4). Figure 6.6 shows the derivation of the reweighting for η(jet2) in the p17 and p20 electron channels with exactly 2-jets as an example. The first two plots show the data distributions (black points), the sum of all backgrounds (light blue points), the W+jets distributions before reweighting (green points), and the multijets (brown line) and other MC (yellow line) distributions for p17 (left) and p20 (right). The sum of the backgrounds is normalized to the data following the procedure described in Sect. 5.5.3, and it can be seen that the shapes do not agree well. The second two plots show the data (black) after subtraction of non-W backgrounds, and the W+jets model (green). The value of the black point in each bin is divided by that of the green point to obtain the red points shown in the last two plots. These red points are the reweighting factors needed to make the W+jets model match the W+jets background in the data.

A priori there should be no difference between the reweighting factors in the electron and muon channels, since this procedure is intended to address a mismodeling of jet quantities in the ALPGEN W+jets samples. This is in fact observed in Fig. 6.7. There are also no significant differences between the reweighting functions determined for different jet multiplicities separately, and so events with two, three, and four jets are combined when calculating the reweighting factors. This is illustrated in Fig. 6.8. In summary, the reweighting factors are calculated on a bin by bin basis as a weighted average over all different lepton and jet multiplicity channels with weights $1/\sigma^2_{i,\rm channel}$ where $\sigma_{i,\rm channel}$ is the uncertainty of the reweighting histogram for bin i and a particular channel. The uncertainty of the combination is calculated by adding in quadrature the statistical uncertainty $(\sum_{\rm channels}(1/\sigma^2_{i,\rm channel}))^{-1/2}$ and a systematic uncertainty taken as half the difference between the final weight and 1.

Given that the corrections are empirical and based on reconstructed quantities, the reweighting factors are treated separately for the Run IIa and Run IIb datasets. The final reweighting sets of factors for η(jet1), η(jet2), $\Delta\phi$(jet1,jet2) and $\Delta\eta$(jet,jet2) are shown in Fig. 6.9 for Run IIa in the first row, Run IIb in the second row, and the comparison between Run IIa and Run IIb and the combination of these in the third row. The largest difference between them occurs at very large pseudorapidity where the data statistics are low to measure the weights and where there are few jets that need to be corrected.

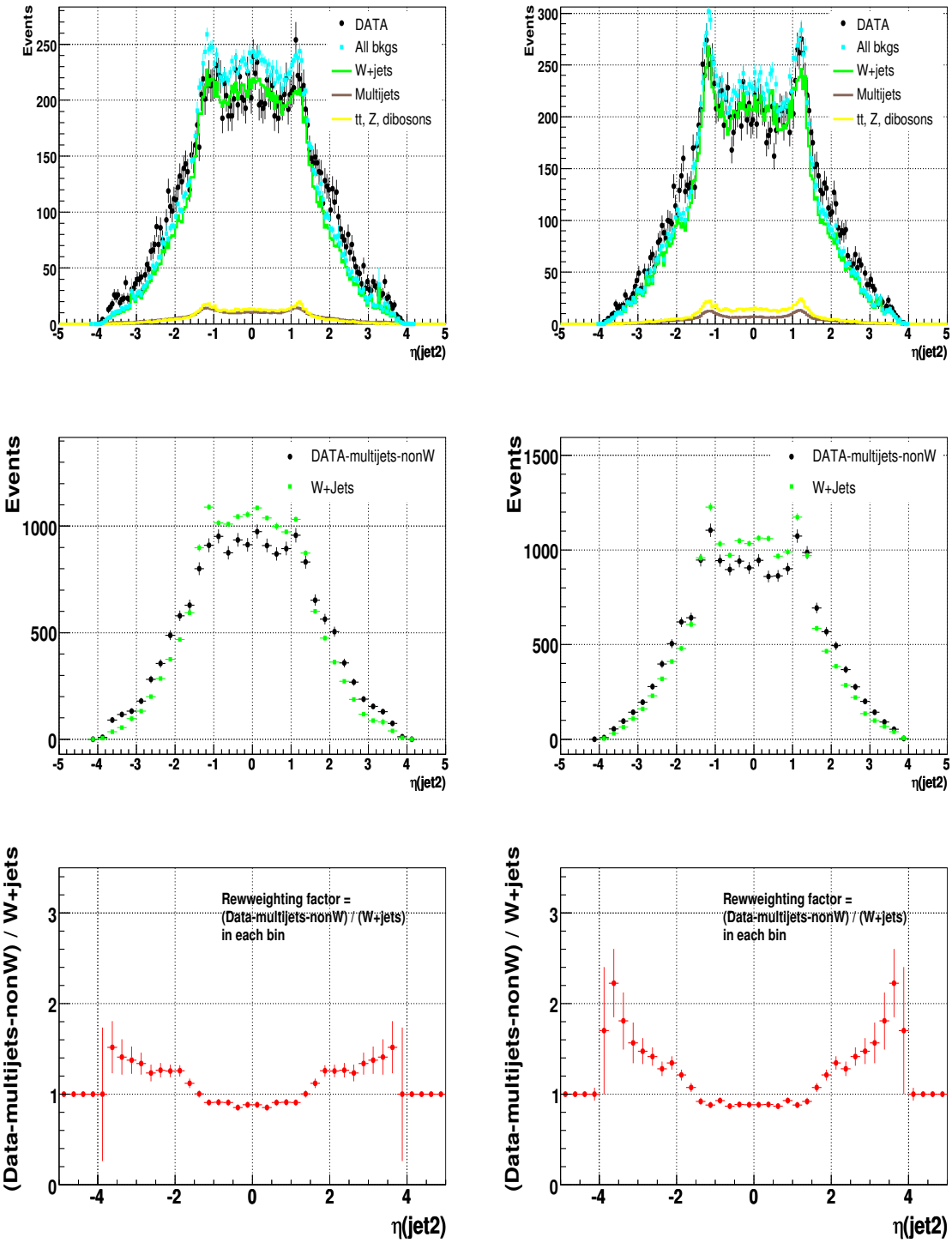

Figure 6.6 : An example of how the ALPGEN W+jets reweighting factors for jets are calculated (p17 left, p20 right). See text for details.

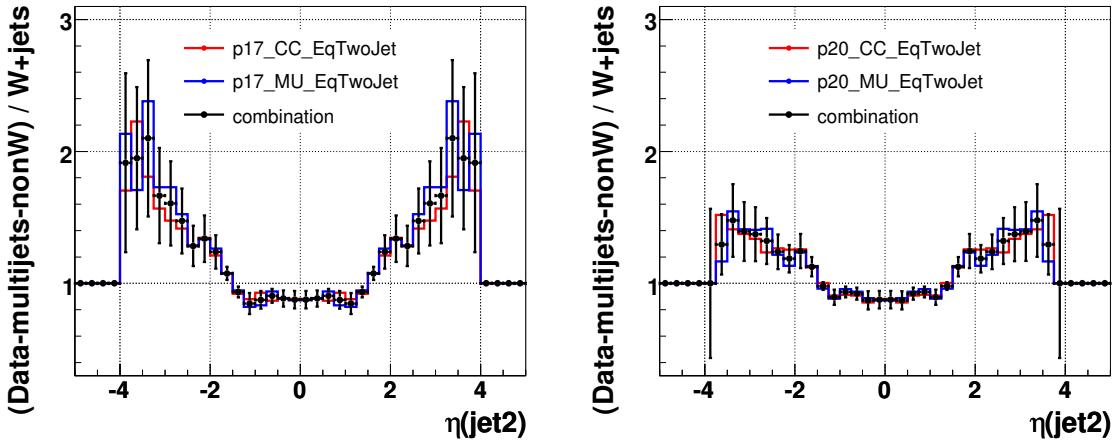

Figure 6.7 : Reweighting factors for η(jet2) from electron and muon channels and their combination for p17 (left) and p20 (right).

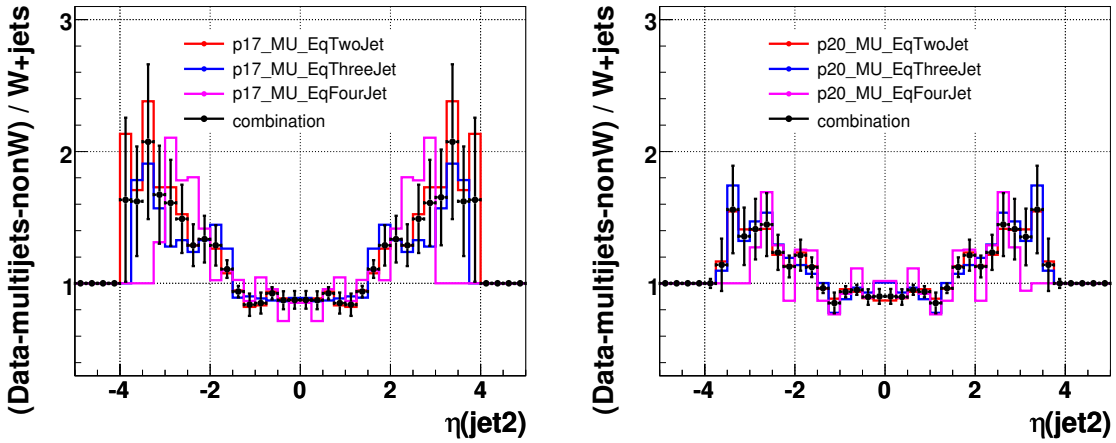

Figure 6.8 : Reweighting factors for η(jet2) from two-jet, three-jet, and four-jet channels and their combination for p17 (left) and p20 (right).

Figure 6.10 shows the muon pseudorapidity distribution before and after reweighting on this variable, while Figs. 6.11 and 6.12 show the equivalent for the η(jet1) and η(jet2) distributions. The effects of reweighting are mostly small, but critical to get good backround-data agreement. Note that after applying each of the ALPGEN W+jets reweighting factors, the normalizations of the W+jets and multijets samples are redetermined using the procedure described in Sect. 5.5.3 (the normalization factors presented in Table 5.4 are the final ones, after all reweightings and corrections).

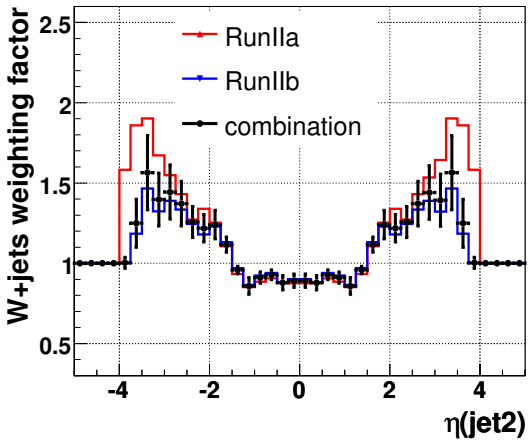

Figure 6.9 : The η(jet2) ALPGEN W+jets reweighting factors for Run IIa, Run IIb and their combination.

MUON DETECTOR ETA

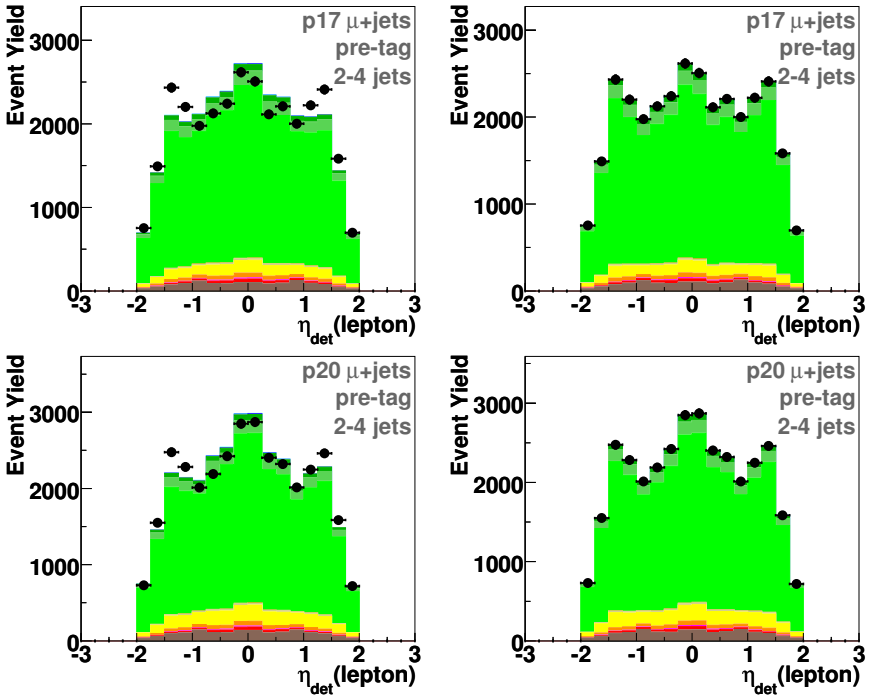

Figure 6.10 : Muon pseudorapidity distributions for Run IIa (up) and Run IIb (down) before reweighting in this variable (left column) and after (right column).

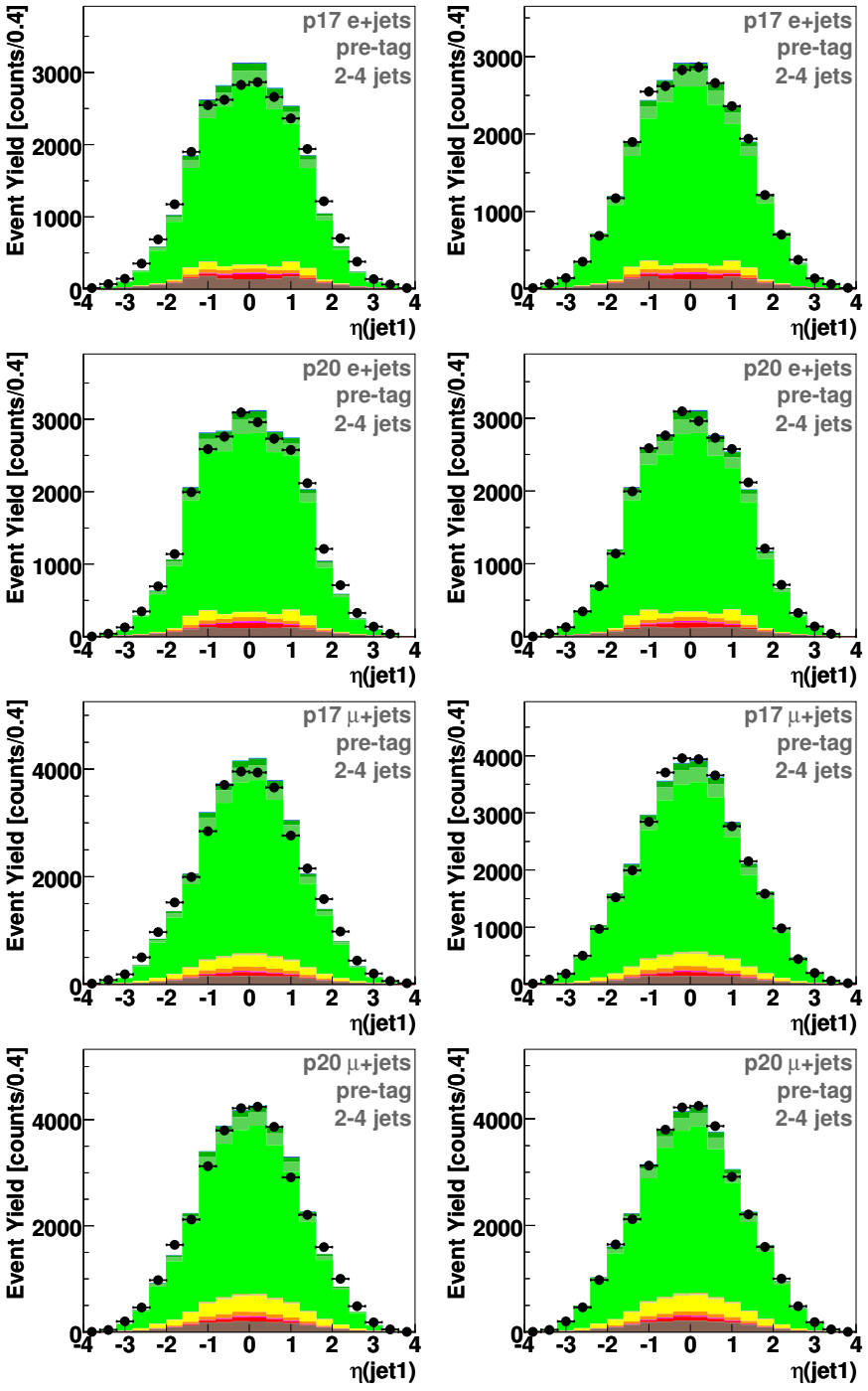

Figure 6.11 : Jet 1 pseudorapidity distributions for the Run IIa and Run IIb electron channels (first two rows) and for the Run IIa and Run IIb muon channels (last two rows) before reweighting in this variable (left column) and after (right column).

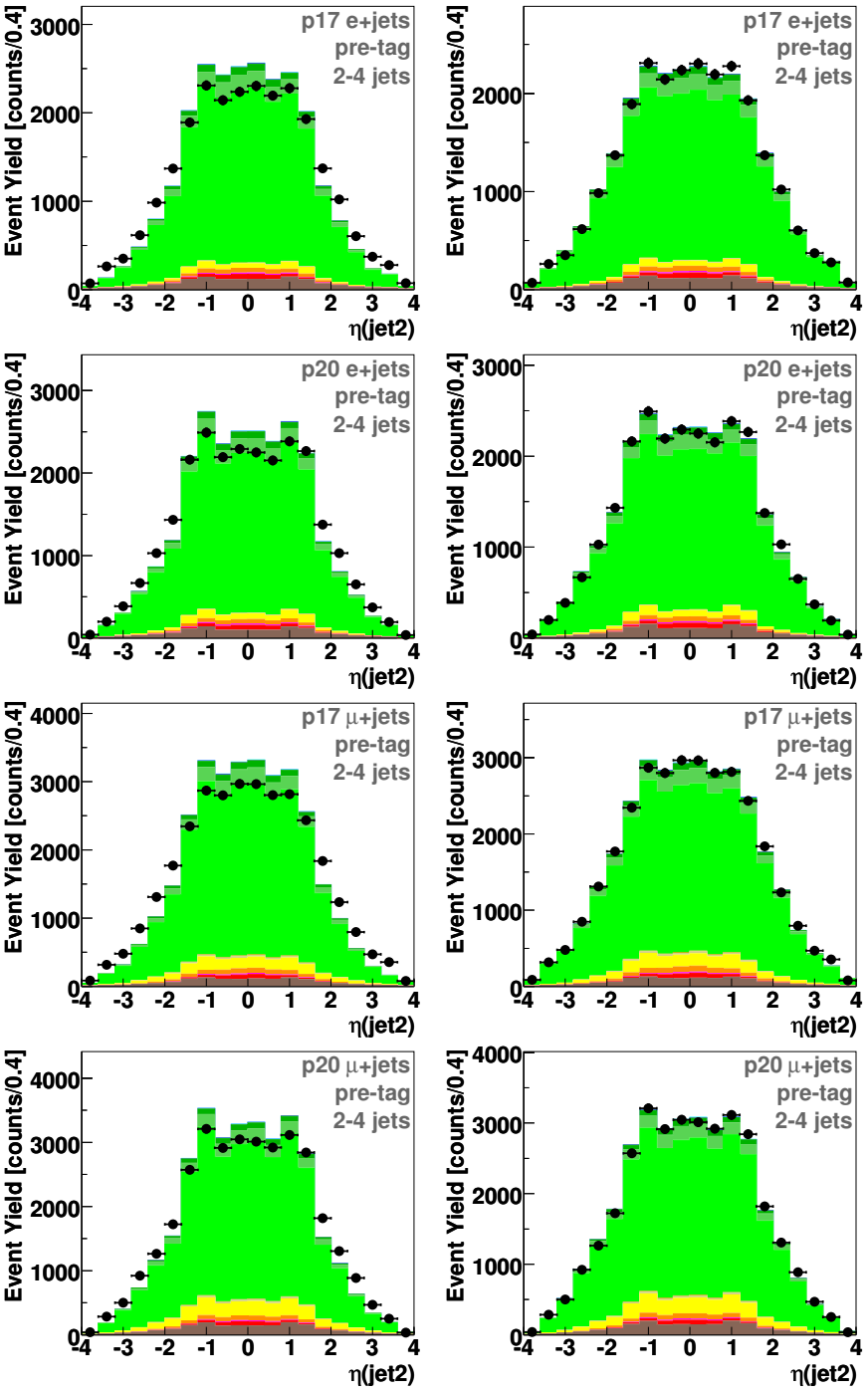

Figure 6.12 : Jet 2 pseudorapidity distributions for the Run IIa and Run IIb electron channels (first two rows) and for the Run IIa and Run IIb muon channels (last two rows) before reweighting in this variable (left column) and after (right column).

6.3 Selection of b-Tag Events

Only taggable jets are considered for b-tagging. A taggable jet is a calorimeter jet matched within $\Delta R < 0.5$ to a track jet. The calorimeter jet must therefore be within the region covered by the tracking system. This is imposed by requiring $|\eta_{\text{jet}}| < 2.5$. The track jet must consist of at least two tracks, with $\Delta R < 0.5$ between them. Each track in the track jet must have at least one SMT hit, and at least one of the tracks must have $p_T > 1$ GeV/c.

Two independent b-tag jet multiplicity bins are defined in this analysis, events with one b-tagged jet and events with two b-tagged jets:

- *One tag*: one jet must pass the Tight b-tagging cut (NNoutput > 0.775) and no other jet can pass the oldLoose cut (NNoutput > 0.5).
- *Two tags*: exactly two jets must pass the oldLoose b-tagging cut.

The veto of a second oldLoose candidate in the one tag definition ensures that there is no overlap between the one-tag and two-tag samples. Finally, the leading b-tagged jet is required to have $p_T > 20$ GeV/c.

As mentioned in Sect. 3.5.1, the NN tagger is applied directly to select b jets in data samples, whereas in simulated events parametrizations are used to define probabilities for a jet to be tagged. The following subsections present an explanation of how these probabilities are derived and how they are used.

6.3.1 Taggability Rate Functions

Taggability is defined as the probability that a jet is taggable. Because the detector is not perfectly modeled in the simulations, one cannot apply the taggability requirements to MC jets directly and get the same answer as for data. Instead, so-called taggability rate functions (P^{taggable}) are applied to the MC events. These functions are parameterized in terms of the jet p_T and jet η variables ($P^{\text{taggable}}(p_T^{\text{jet}}, \eta^{\text{jet}})$). They are derived using "loose" data samples (that is, samples with all selection cuts applied, except for the lepton criteria which is relaxed from tight to loose isolation) to have more statistics for the measurement. To validate the applicability of these functions to the analysis samples, cross checks are made comparing the observed taggability (this means, requiring the jet to be taggable) measured in the analysis data sample with the taggability rate function value in the same analysis data events. The prescription for taggability determination described in Ref. [84] is followed, measuring the taggability in six z_{PV} zones. Figure 6.13 shows the taggability determination and corresponding closure tests for the electron channel in the Run IIb period in the so-called "central minus" zone ($|z_{\text{PV}}| < 38$ cm with $z_{\text{PV}} \times \eta^{\text{jet}} < 0$). The plots for the "central plus" zone

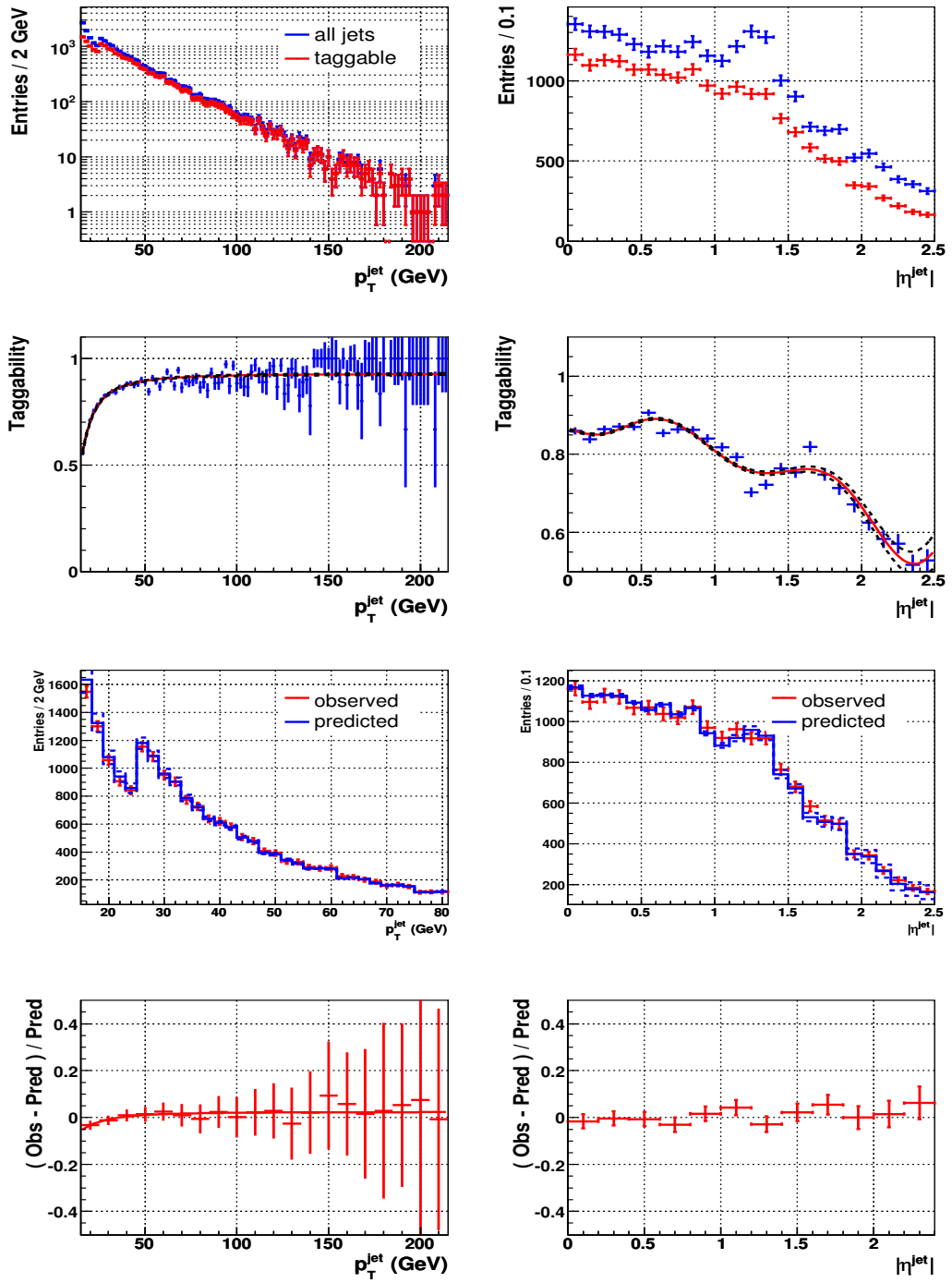

Figure 6.13 : The first row shows the jet p_T (left) and jet η (right) distributions for all jets (blue) and for taggable jets (red) in the "loose" data sample for the Run IIb electron channel and the "central minus" $z_{\rm PV}$ zone. The second row shows the ratio of taggable to all jets and the corresponding fits for the derivation of the taggability. The third row shows in red again the distribution of taggable jets (observed taggability) and in blue the (predicted taggability), while the ratio is shown in the last row.

($|z_{\rm PV}| < 20$ cm with $z_{\rm PV} \times \eta^{\rm jet} > 0$) look similar. Equivalent plots for other $z_{\rm PV}$ zones, for the Run IIa period, and for the muon channel, can be found in Ref. [85].

The tracking acceptance in η is higher for jets in the "negative" zones than for jets in the "positive" zones, and the difference in acceptance increases with increasing $|z_{\rm PV}|$. In general the taggability increases as the jet traverses more layers of the tracking system and tends to zero when $\eta_{\rm jet}$ is such that jets can not reach the SMT as it happens in the right plot of the second row in Fig. 6.13 for high η values. On the other hand, the taggability increases with the p_T of the jet, because the particle content increases and therefore the number of tracks within it, as well as tracks with higher p_T curve less in the magnetic field and are more likely to reach the tracking detector. The observed jump in the jet p_T spectrum at 25 GeV/c is because of the $p_T > 25$ GeV/c cut on the leading jet (see Sect. 6.1.1). The observed taggability and the predicted taggability give the same result within uncertainties.

6.3.2 b-Tagging Event Weights and b-Jet Assignment Combinations

The probability to tag a jet of flavor α, or tag rate function (TRF), can be expressed as the product of the taggability rate function (see Sect. 6.3.1) and the tagging efficiency:

$$\begin{aligned} \mathrm{TRF}_\alpha(p_T^{\rm jet}, \eta^{\rm jet}, z_{\rm PV}) &\equiv \mathcal{P}_\alpha(p_T^{\rm jet}, \eta^{\rm jet}, z_{\rm PV}) \\ &= P^{\rm taggable}(p_T^{\rm jet}, \eta^{\rm jet}, z_{PV}) \times \varepsilon_\alpha(p_T^{\rm jet}, \eta^{\rm jet}). \end{aligned} \tag{6.1}$$

Average b-tagging efficiencies for the analysis data samples are shown in Table 6.4.

b-Tagging Efficiencies

	oldLoose			Tight		
Taggability	b jets	c jets	light jets	b jets	c jets	light jets
	Run IIa					
Without	58%	17%	1.5%	48%	10%	0.42%
With	50%	14%	1.2%	41%	8.6%	0.35%
	Run IIb					
Without	57%	18%	2.1%	46%	11%	0.63%
With	48%	15%	1.7%	39%	8.9%	0.52%

Table 6.4 : b-tagging efficiencies for the oldLoose and Tight operating points measured by averaging the tag rate function values in the analysis data samples. The upper rows show the values without taking into account taggability and the lower rows show values including it.

Using the per jet probability, one can deduce the probability for the event to contain exactly a given number of tags. One has to take into account all possible permutations of tagged/untagged jets, combine per jet probabilities to build the probability for each

permutation, and sum over the permutations. The probability equations for the one-tag and two-tag cases are the following:

$$P_{\text{event}}(1\text{ tag}) = \sum_j [\mathcal{P}^{\text{Tight}}_{\alpha^j}(p_T^j, \eta^j) \prod_{i \neq j} (1 - \mathcal{P}^{\text{oldLoose}}_{\alpha^i}(p_T^i, \eta^i))], \tag{6.2}$$

$$P_{\text{event}}(2\text{ tags}) = \sum_{j,i \neq j} [\mathcal{P}^{\text{oldLoose}}_{\alpha^j}(p_T^j, \eta^j) \mathcal{P}^{\text{oldLoose}}_{\alpha^i}(p_T^i, \eta^i) \prod_{k \neq j \neq i} (1 - \mathcal{P}^{\text{oldLoose}}_{\alpha^k}(p_T^k, \eta^k))], \tag{6.3}$$

where the sums run over all jets in the event.

By using the TRFs, one can estimate the number of events in any b-tag channel, but one cannot say whether an individual jet is tagged or not. Therefore, when using kinematic variables separated by the number of tags or variables that rely on using b-tagging information on the jets, each event is split based on the number of jets and the number of possible tags, assigning for each permutation of tagged/untagged jets the corresponding weight. Thus, each event is taken into account several times, with the sum of weights for all possible combinations in each event being the original probability for the event to be not-tagged, tagged once or tagged twice. For example, an event with two jets can have zero, one or two b-tagged jets. The probability of having both jets tagged has only one possible permutation and is given by

$$P_{\text{2-jet event}}(2\text{ tags}) = \mathcal{P}^{\text{oldLoose}}_{\text{jet}_1} \mathcal{P}^{\text{oldLoose}}_{\text{jet}_2}; \tag{6.4}$$

the probability of having only one of the jets tagged has two combinations,

$$P_{\text{2-jet event}}(1\text{ tag}) = \mathcal{P}^{\text{Tight}}_{\text{jet}_1}(1 - \mathcal{P}^{\text{oldLoose}}_{\text{jet}_2}) + \mathcal{P}^{\text{Tight}}_{\text{jet}_2}(1 - \mathcal{P}^{\text{oldLoose}}_{\text{jet}_1}), \tag{6.5}$$

and the probability of having none of the jets tagged has three permutations,

$$\begin{aligned} P_{\text{2-jet event}}(0\text{ tag}) &= 1 - P_{\text{2-jet event}}(1\text{ tag}) - P_{\text{2-jet event}}(2\text{ tags}) \\ &= (\mathcal{P}^{\text{oldLoose}}_{\text{jet}_1} - \mathcal{P}^{\text{Tight}}_{\text{jet}_1})(1 - \mathcal{P}^{\text{oldLoose}}_{\text{jet}_2}) \\ &+ (\mathcal{P}^{\text{oldLoose}}_{\text{jet}_2} - \mathcal{P}^{\text{Tight}}_{\text{jet}_2})(1 - \mathcal{P}^{\text{oldLoose}}_{\text{jet}_1}) \\ &+ (1 - \mathcal{P}^{\text{oldLoose}}_{\text{jet}_1})(1 - \mathcal{P}^{\text{oldLoose}}_{\text{jet}_2}). \end{aligned} \tag{6.6}$$

Consider for instance the distribution of lepton p_T in the one-tag channel. This 2-jet event will appear twice, once weighted by $\mathcal{P}^{\text{Tight}}_{\text{jet}_1}(1 - \mathcal{P}^{\text{oldLoose}}_{\text{jet}_2})$ and once weighted by $\mathcal{P}^{\text{Tight}}_{\text{jet}_2}(1 - \mathcal{P}^{\text{oldLoose}}_{\text{jet}_1})$, but in both cases with the same value of lepton p_T. On the other hand, in the distribution of, say, the leading b-tagged jet p_T in the one-tag channel, the same 2-jet event appears again twice, but now at the values corresponding to the p_T of the jet_1 with a weight $\mathcal{P}^{\text{Tight}}_{\text{jet}_1}(1 - \mathcal{P}^{\text{oldLoose}}_{\text{jet}_2})$ and the p_T of the jet_2 with a weight $\mathcal{P}^{\text{Tight}}_{\text{jet}_2}(1 - \mathcal{P}^{\text{oldLoose}}_{\text{jet}_1})$.

6.4 Number of Events after Selection

Tables 6.5 and 6.6 show the "raw" number of events (without any weights) in the signal and background samples and in the data after applying the selection criteria discussed in Sects. 6.1 and 6.3. The number of events are used in calculating statistical uncertainties on the backgrounds and signals. They also become important when training a discriminant (e.g. a Neural Network), as they give the statistics available for the training.

Numbers of Events after Selection in Run IIa Data

	Electron channel			Muon channel		
	2 jets	3 jets	4 jets	2 jets	3 jets	4 jets
Signal MC						
tb	42,407	14,339	3,693	42,295	16,181	3,990
tqb	36,568	14,273	4,484	33,744	14,563	4,330
tb+tqb	78,975	28,612	8,177	76,039	30,744	8,320
Background MC						
$t\bar{t} \to \ell\ell$	105,957	67,291	22,008	88,999	66,241	21,622
$t\bar{t} \to \ell$+jets	34,183	110,291	128,273	23,066	103,952	140,513
$Wb\bar{b}$	33,958	11,589	3,439	38,522	14,757	4,362
$Wc\bar{c}$	29,781	11,165	3,202	34,861	14,112	3,925
Wcj	15,518	5,427	1,252	17,462	6,346	1,509
Wjj	214,395	74,244	18,417	260,622	98,177	24,348
$Zb\bar{b}$	2,827	1,202	432	7,371	2,077	591
$Zc\bar{c}$	722	348	125	3,406	1,280	364
Zjj	9,192	3,516	1,353	35,374	11,060	2,729
Dibosons	92,882	25,726	5,897	113,327	33,989	7,637
Pretag data						
Multijets	241,173	105,002	46,458	18,281	8,083	2,827
Signal data	18,582	4,834	1,246	23,243	6,675	1,663
One-Tag data						
Multijets	6,938	3,773	1,845	1,050	591	300
Signal data	508	202	103	627	259	131
Two-Tags data						
Multijets	451	417	285	67	68	38
Signal data	67	61	37	71	62	56

Table 6.5 : Numbers of events in data and MC for the electron and muon channels after selection in the Run IIa sample.

Numbers of Events after Selection in Run IIb Data

	Electron channel			Muon channel		
	2 jets	3 jets	4 jets	2 jets	3 jets	4 jets
Signal MC						
tb	51,228	17,890	4,796	55,557	24,016	6,853
tqb	44,863	19,223	6,425	50,861	24,416	8,134
tb+tqb	96,091	37,113	11,221	106,418	48,432	14,987
Background MC						
$t\bar{t} \to \ell\ell$	61,642	47,435	15,473	57,435	51,705	17,168
$t\bar{t} \to \ell$+jets	15,461	49,628	57,644	11,787	51,564	69,033
$Wb\bar{b}$	50,850	19,131	5,572	63,506	26,499	7,736
$Wc\bar{c}$	37,524	13,860	4,061	48,043	20,081	5,588
Wcj	6,596	2,298	567	8,016	3,082	724
Wjj	91,261	32,106	8,335	120,751	47,562	11,711
$Zb\bar{b}$	3,266	1,418	609	12,124	4,499	1,280
$Zc\bar{c}$	2,185	883	347	9,454	3,308	864
Zjj	9,193	3,516	1,353	35,387	11,067	2,730
Diboson	25,373	7,642	2,001	36,486	11,865	3,001
Pretag data						
Multijets	229,208	101,466	47,660	21,416	9,188	3,233
Signal data	19,048	5,087	1,460	23,972	7,040	1,927
One-Tag data						
Multijets	6,838	3,417	1,773	1,150	689	271
Signal data	547	207	124	595	290	142
Two-Tags data						
Multijets	505	444	287	77	76	29
Signal data	79	56	51	85	79	80

Table 6.6 : Numbers of events in data and MC for the electron and muon channels after selection in the Run IIb sample.

6.5 Event Yields

This section presents the number of signal and background events after selection predicted to be in the nearly 2.3 fb^{-1} of data analyzed. These numbers are called "event yields" and they are calculated after all corrections, reweightings and normalizations have been applied[33]. For data, the event yields are just the number of selected events. Table 6.7 shows the event yields for the Run IIa and Run IIb data taking periods, for all signals and backgrounds before b-tagging, separated by lepton flavor

[33]Note that, contrary to the raw number of events which are integer numbers, the event yields are in general non-integer numbers.

and jet multiplicity. The same information appears in Table 6.8, but for combinations of channels. In this table, the percentages are of the total background for each component. Because of the normalization to data of the W+jets and multijets backgrounds before b-tagging (see Sect. 5.5.3), the background sum yield is defined to be equal to the number of events observed in the data. Note also that event yield values shown in these and subsequent tables have been rounded for clarity, so that the sums of the components will not always equal exactly the values given for these sums. In the normalization procedure described in Sect. 5.5.3, the single top signal has been neglected. Observing the yields in Table 6.7, one can get convinced now that this is in fact a valid approximation, since for example the 2% uncertainty in the W+jets normalization factor implies already a variation in the W+jets yield that is at least two times larger than the single top contribution.

Table 6.9 shows the event yields for the Run IIa and Run IIb data taking periods in events with one b-tagged jet, for all signals and backgrounds, separated by lepton flavor and jet multiplicity. Similarly, Table 6.10 shows the corresponding event yields for events with two b-tagged jets and Table 6.11 shows the yields summed over each jet multiplicity separately and combined, for all b-tag channels. In this last table, the percentages are of the total background for each component.

The proportions of the individual signal and background components with respect to the total backgrounds+signals are illustrated in Fig. 6.14 for the two data taking periods and the two lepton flavors combined, but divided by number of jets and number of b-tagged jets. It can nicely be observed how the W+jets ($t\bar{t}$) background decreases (increases) with the jet multiplicity, and how the signal appears mostly in the 2-jet 2-tag bin with the main contribution coming from the s-channel.

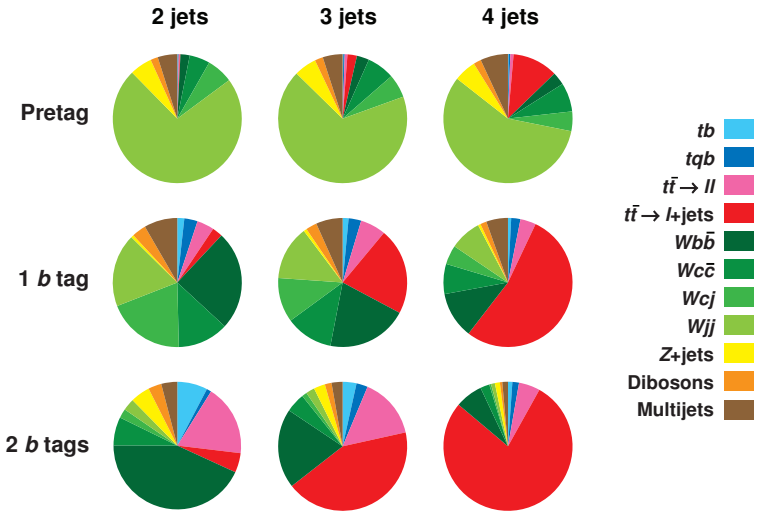

Figure 6.14 : Illustration of the composition of the data sets as a function of number of jets and number of b-tags.

Pretag Event Yields

	Electron channel			Muon channel		
	2 jets	3 jets	4 jets	2 jets	3 jets	4 jets
	Run IIa					
Signals						
tb	23	8.2	2.2	27	11	2.8
tqb	43	17	5.8	51	23	7.0
tb+tqb	67	26	8.0	77	33	10
Backgrounds						
$t\bar{t} \to \ell\ell$	60	37	12	57	41	13
$t\bar{t} \to \ell$+jets	41	136	158	32	143	196
$Wb\bar{b}$	479	160	45	530	211	57
$Wc\bar{c}$	1,041	356	101	1,196	485	125
Wcj	1,338	315	65	1,514	389	81
Wjj	13,847	3,309	722	17,028	4,612	984
$Zb\bar{b}$	18	7.1	3.5	70	22	6.6
$Zc\bar{c}$	33	12	4.2	151	46	13
Zjj	461	125	40	1,309	348	84
Dibosons	339	98	24	457	142	34
Multijets	923	278	74	896	235	69
Background sum	18,582	4,834	1,246	23,243	6,675	1,663
Data	18,582	4,834	1,246	23,243	6,675	1,663
	Run IIb					
Signals						
tb	24	8.4	2.3	26	11	3.3
tqb	43	19	6.3	49	24	7.9
tb+tqb	67	27	8.6	75	35	11
Backgrounds						
$t\bar{t} \to \ell\ell$	65	42	13	61	46	14
$t\bar{t} \to \ell$+jets	43	141	168	33	145	198
$Wb\bar{b}$	458	161	42	499	200	61
$Wc\bar{c}$	1,006	351	94	1,126	453	137
Wcj	1,327	316	70	1,442	377	96
Wjj	14,166	3,489	795	16,941	4,710	1,137
$Zb\bar{b}$	19	8.2	3.4	70	26	7.4
$Zc\bar{c}$	35	15	5.9	152	54	14
Zjj	596	167	55	1,833	507	118
Dibosons	343	103	26	445	145	37
Multijets	987	294	188	1,369	377	108
Background sum	19,048	5,087	1,460	23,972	7,040	1,927
Data	19,048	5,087	1,460	23,972	7,040	1,927

Table 6.7 : Yields after selection but before b-tagging.

Pretag Event Yields with Channels Combined
and Percentages of Total Background

	Run IIa+IIb, electron+muon							
	2 jets		3 jets		4 jets		All channels	
Signal								
tb	100		38.6		10.6		149	
tqb	186		83		27		296	
tb+tqb	286		121		37		444	
Backgrounds								
$t\bar{t} \to \ell\ell$	243	(0.3%)	166	(0.7%)	52	(0.8%)	461	(0.4%)
$t\bar{t} \to \ell$+jets	149	(0.2%)	565	(2.4%)	720	(11.4%)	1,434	(1.2%)
$Wb\bar{b}$	1,966	(2.3%)	732	(3.1%)	205	(3.2%)	2,903	(2.5%)
$Wc\bar{c}$	4,369	(5.1%)	1,645	(7.0%)	457	(7.2%)	6,471	(5.6%)
Wcj	5,621	(6.6%)	1,397	(5.9%)	312	(5.0%)	7,330	(6.4%)
Wjj	61,982	(73.1%)	16,120	(68.2%)	3,638	(57.8%)	81,740	(71.2%)
$Zb\bar{b}$	177	(0.2%)	63	(0.2%)	21	(0.3%)	261	(0.2%)
$Zc\bar{c}$	371	(0.4%)	127	(0.5%)	37	(0.6%)	535	(0.5%)
Zjj	4,199	(4.9%)	1,147	(4.9%)	297	(4.7%)	5,643	(4.9%)
Dibosons	1,584	(1.9%)	488	(2.1%)	121	(1.9%)	2,193	(1.9%)
Multijets	4,175	(4.9%)	1,184	(5.0%)	439	(7.0%)	5,798	(5.1%)
Backgrounds combined								
$t\bar{t}$	392	(0.5%)	731	(3.1%)	772	(12.3%)	1,895	(1.7%)
W+jets	73,938	(87.1%)	19,894	(84.2%)	4,612	(73.3%)	98,444	(85.8%)
Z+jets & dibosons	6,331	(7.5%)	1,825	(7.7%)	476	(7.6%)	8,632	(7.5%)
Multijets	4,175	(4.9%)	1,184	(5.0%)	439	(7.0%)	5,798	(5.1%)
Background sum	84,845		23,636		6,296		114,777	
Backgrounds +signals	85,129		23,759		6,335		115,223	
Data	84,845		23,636		6,296		114,777	
$S : B$	1:297		1:195		1:170		1:259	

Table 6.8 : Yields after selection but before b-tagging for each jet multiplicity and for all analysis channels combined. The percentages are of the total background for each component.

Single-Tag Event Yields

	Electron channel			Muon channel		
	2 jets	3 jets	4 jets	2 jets	3 jets	4 jets
	Run IIa					
Signals						
tb	9.1	3.1	0.82	10.2	3.9	1.0
tqb	17.4	6.6	2.1	20.5	8.6	3.6
tb+tqb	26.4	9.7	3.0	30.7	12.5	3.6
Backgrounds						
$t\bar{t} \to \ell\ell$	23.6	14.2	4.3	22.1	15.6	4.8
$t\bar{t} \to \ell$+jets	16.3	52.0	57.0	12.5	54.3	69.8
$Wb\bar{b}$	135.4	44.2	12.0	146.4	57.1	16.0
$Wc\bar{c}$	66.0	24.8	8.1	73.9	33.4	9.9
Wcj	98.3	24.0	5.0	112.1	30.2	6.3
Wjj	73.6	21.9	6.1	87.0	30.1	8.1
$Zb\bar{b}$	6.5	2.8	0.89	26.8	7.9	2.5
$Zc\bar{c}$	2.7	1.2	0.55	13.3	4.6	1.5
Zjj	5.4	1.8	0.63	12.7	4.3	1.1
Dibosons	16.2	5.3	1.4	22.3	7.8	2.1
Multijets	28.0	10.3	3.0	51.5	17.2	7.3
Background sum	472.1	202.4	99.0	580.6	262.5	129.4
Data	508	202	103	627	259	131
	Run IIb					
Signals						
tb	9.5	3.3	0.91	9.9	4.3	1.2
tqb	16.9	7.2	2.5	18.5	8.7	3.0
tb+tqb	26.4	10.5	3.4	28.4	13.0	4.2
Backgrounds						
$t\bar{t} \to \ell\ell$	25.6	16.1	4.9	23.5	17.0	5.0
$t\bar{t} \to \ell$+jets	16.4	53.8	61.5	12.2	53.6	70.6
$Wb\bar{b}$	129.5	44.3	11.6	136.0	53.4	16.8
$Wc\bar{c}$	68.6	26.2	7.6	72.4	32.6	10.9
Wcj	106.1	25.8	5.4	111.9	29.7	6.6
Wjj	114.4	35.4	9.6	128.3	46.5	14.5
$Zb\bar{b}$	5.0	2.4	1.0	20.1	7.7	2.2
$Zc\bar{c}$	2.1	1.1	0.57	10.7	4.3	1.2
Zjj	6.0	2.1	0.79	13.9	5.0	1.3
Dibosons	17.4	5.8	1.7	22.7	8.4	2.4
Multijets	31.0	10.1	7.1	73.5	28.2	9.0
Background sum	522.1	223.2	111.6	625.3	286.5	140.5
Data	547	207	124	595	290	142

Table 6.9 : Yields after selection for events with exactly one b-tagged jet.

Double-Tag Event Yields

	Electron channel			Muon channel		
	2 jets	3 jets	4 jets	2 jets	3 jets	4 jets
	Run IIa					
Signals						
tb	5.69	2.12	0.58	6.66	2.75	0.75
tqb	0.81	1.61	0.90	0.92	2.15	1.07
tb+tqb	6.50	3.72	1.49	7.59	4.90	1.82
Backgrounds						
$t\bar{t} \to \ell\ell$	13.91	9.70	3.12	14.09	11.22	3.58
$t\bar{t} \to \ell$+jets	4.32	28.63	43.16	3.51	32.18	55.38
$Wb\bar{b}$	33.96	12.34	3.69	35.64	15.71	4.77
$Wc\bar{c}$	5.12	2.75	1.28	5.62	3.56	1.50
Wcj	1.44	0.68	0.19	1.62	0.83	0.26
Wjj	1.45	0.86	0.34	1.70	1.20	0.46
$Zb\bar{b}$	0.88	0.74	0.31	6.14	2.60	0.91
$Zc\bar{c}$	0.15	0.13	0.10	1.05	0.54	0.25
Zjj	0.14	0.09	0.05	0.31	0.20	0.07
Dibosons	2.05	0.85	0.28	3.06	1.37	0.46
Multijets	1.90	1.10	0.48	3.28	1.98	0.93
Background sum	65.33	57.88	53.00	76.03	71.40	68.57
Data	67	61	37	71	62	56
	Run IIb					
Signals						
tb	5.26	2.00	0.58	5.61	2.59	0.78
tqb	0.94	1.89	1.01	0.99	2.22	1.21
tb+tqb	6.20	3.89	1.58	6.60	4.80	1.99
Backgrounds						
$t\bar{t} \to \ell\ell$	13.58	9.99	3.17	12.95	10.79	3.33
$t\bar{t} \to \ell$+jets	4.07	27.71	43.44	3.11	29.00	51.06
$Wb\bar{b}$	30.54	12.19	3.43	30.84	14.42	5.07
$Wc\bar{c}$	5.55	3.15	1.17	5.60	3.72	1.67
Wcj	2.04	0.96	0.28	2.07	1.04	0.33
Wjj	2.81	1.66	0.64	3.21	2.20	0.98
$Zb\bar{b}$	0.69	0.60	0.34	4.34	2.07	0.70
$Zc\bar{c}$	0.14	0.14	0.10	0.86	0.53	0.19
Zjj	0.16	0.11	0.06	0.34	0.24	0.09
Dibosons	1.96	0.91	0.30	2.98	1.38	0.46
Multijets	2.25	1.37	1.13	4.92	3.12	0.97
Background sum	63.78	58.80	54.06	71.22	68.50	64.85
Data	79	56	51	85	79	80

Table 6.10 : Yields after selection for events with exactly two b-tagged jets.

b-Tag Event Yields with Channels Combined and Percentages of Total Background

	Run IIa+IIb, electron+muon							
	2 jets		3 jets		4 jets		All channels	
Signal								
tb	62		24		6.9		93	
tqb	76		39		15		130	
tb+tqb	138		63		22		223	
Backgrounds								
$t\bar{t} \to \ell\ell$	149	(6.0%)	105	(8.5%)	32	(4.4%)	286	(6.5%)
$t\bar{t} \to \ell$+jets	72	(2.9%)	331	(26.9%)	452	(62.7%)	856	(19.3%)
$Wb\bar{b}$	678	(27.4%)	254	(20.6%)	73	(10.1%)	1,005	(22.7%)
$Wc\bar{c}$	303	(12.2%)	130	(10.6%)	42	(5.8%)	475	(10.7%)
Wcj	435	(17.6%)	113	(9.2%)	24	(3.3%)	573	(12.9%)
Wjj	413	(16.7%)	140	(11.4%)	41	(5.7%)	593	(13.4%)
$Zb\bar{b}$	71	(2.9%)	27	(2.2%)	8.8	(1.2%)	106	(2.4%)
$Zc\bar{c}$	31	(1.3%)	13	(1.1%)	4.4	(0.6%)	48	(1.1%)
Zjj	39	(1.6%)	14	(1.1%)	4.1	(0.6%)	57	(1.3%)
Dibosons	89	(3.6%)	32	(2.6%)	9.1	(1.3%)	130	(2.9%)
Multijets	196	(7.9%)	73	(5.9%)	30	(4.2%)	300	(6.8%)
Backgrounds combined								
$t\bar{t}$	221	(8.9%)	436	(35.4%)	484	(67.1%)	1,142	(25.8%)
W+jets	1,829	(73.9%)	637	(51.7%)	181	(25.1%)	2,647	(59.8%)
Z+jets & dibosons	229	(9.2%)	85	(6.9%)	27	(3.7%)	341	(7.7%)
Multijets	196	(7.9%)	73	(5.9%)	30	(4.2%)	300	(6.8%)
Background sum	2,476		1,231		721		4,429	
Backgrounds +signals	2,614		1,294		743		4,652	
Data	2,579		1,216		724		4,519	
$S : B$	1:18		1:20		1:34		1:20	

Table 6.11 : Yields after selection for each jet multiplicity and for all analysis channels combined. The percentages are of the total background for each component.

Events with three or more b-tagged jets are not included in the analysis, because the probability to get both b jets tagged in signal events and a further light jet with a fake tag is very low, while other sources of three-tag events are much higher. The most probable source is $t\bar{t} \to \ell +$ jets events with both b jets tagged and also a charm jet from the decay of one of the W bosons. Table 6.12 presents the numbers of triple-tag events in the data to show they are negligible as expected, with the data yields for the one-tag and two-tag channels given for comparison.

Triple-Tag Events in Data

	Three Jets				Four Jets			
	Electron		Muon		Electron		Muon	
	p17	p20	p17	p20	p17	p20	p17	p20
Pretag	4,834	5,087	6,675	7,040	1,246	1,460	1,663	1,927
1-tag	202	207	259	290	103	124	131	142
2-tags	61	56	62	79	37	51	56	80
3-tags	2	4	2	4	3	8	4	7

Table 6.12 : Numbers of data events in each channel after selection, with triple-tag events shown to check they are as expected. (Only events with exactly one or two b-tagged jets are used in the analysis.)

Finally, Table 6.13 shows the differences between the data and the background model plus SM signal prediction for each analysis channel and combination of channels as a factor times the background+signal model uncertainty. These numbers demonstrate the consistency of the background model with the data. Differences for pretag samples are not shown given that they are all equal to zero because the normalization of the background sum to data.

6.6 Data versus Background Model Comparison

Initially, a set of approximately 600 kinematic variables is defined in the analysis, in which the agreement between the data and the background plus signal model is checked. Among them, there is a set of basic kinematic variables where the agreement must be satisfactory; otherwise the data are not well understood. Examples of these variables are the p_T of the different objects, $\not\!\!E_T$, H_T's, $\Delta\phi$'s, ΔR's, and probably the most relevant, the W transverse mass ($M_T(W)$). In this section, the agreement between data and the model is checked on some of these basic variables.

Variables definitions: Variables are defined and named using the following conventions. Jets are sorted in descending order by their p_T, with index 1 refering to the leading jet in any given jet category: "jet$n$" ($n = 1, 2, 3, 4$) corresponds to every jet in the event; "best$n$" refers to the best jet and "notbest$n$" to all but the best jet. The best

Data Excess (+) or Deficit (−) over SM-Signal+Background Model

Tag bin	Electron Channel 2 jets	3 jets	4 jets	Muon Channel 2 jets	3 jets	4 jets
	Run IIa					
1 b-tag	0.26σ	-0.60σ	0.10σ	0.35σ	-0.78σ	-0.16σ
2 b-tags	-0.51σ	-0.07σ	-2.04σ	-1.18σ	-1.35σ	-1.33σ
	Run IIb					
1 b-tag	-0.03σ	-1.39σ	0.79σ	-1.11σ	-0.40σ	-0.20σ
2 b-tags	0.94σ	-0.74σ	-0.53σ	0.70σ	0.56σ	1.29σ

e+μ, tags & jets comb. Run IIa	Run IIb	Run IIa+IIb, tags & jets comb. e	μ	Run IIa+IIb, e+μ, tags comb. 2 jets	3 jets	4 jets	All channels
-0.37σ	-0.31σ	-0.25σ	-0.42σ	-0.17σ	-0.67σ	-0.22σ	-0.34σ

Table 6.13 : Differences between the data and the predicted background (including SM signals) shown as a factor times the uncertainty on the background+signal predictions.

jet is defined as the one for which the invariant mass $M(W,\text{jet})$ (by W it is meant the system formed by the reconstructed lepton and neutrino) is closest to $m_t = 170$ GeV/c^2. Finally, "tag$n$" refers to b-tagged jets, "untag$n$" to non b-tagged jets, and "light$n$" to all jets except the leading b-tagged jet. This naming convention has no implication on the definition of the 1 b-tag or 2 b-tag samples. In particular, the 2 b-tag samples have two jets that are tagged with the b-tag NN algorithm as b-jets. However, the jet with the highest p_T value is called "tag1", and all the other jets in the event are called "light$n$". For this sample, the second leading b-tagged jet is considered both as "tag2" and as "light$n$", with n being determined by the jet's relative ranking in p_T among all light jets in the event.

Figure 6.15 shows the color scheme used in this dissertation to represent each of the signal and background contributions. In the cases when some background and/or signal sources are merged together, the new color scheme is provided with the corresponding graph.

As said above, one of the most important variables one is interested in, is the W transverse mass, which is defined by means of the reconstructed lepton and neutrino ($\not\!\!E_T$) as

$$M_T(W) \equiv \sqrt{E_T^2(W) - \vec{p}_T^{\,2}(W)} = \sqrt{2\, p_T^{\text{lepton}} \not\!\!E_T \cdot (1 - \cos(\Delta\phi(\text{lepton}, \not\!\!E_T))}\,. \qquad (6.7)$$

For events containing real W bosons, this gives a Jakobian peak with an edge at the W mass (≈ 80 GeV/c^2), but with the bulk of the distribution below the W mass from events with neutrino momentum along the z direction. There is also a long tail above

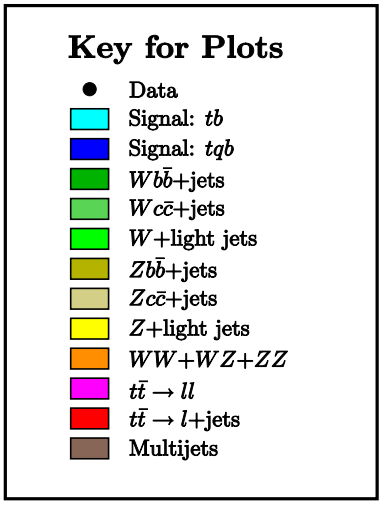

Figure 6.15 : Illustration of the color scheme used in plots of signal and backgrounds in the single top analysis.

the W mass, which arises from the width of the W boson. Events with non-real W's (e.g. multijets) mostly populate the low $M_T(W)$ region (since $\Delta\phi(\text{lepton}, \not\!\!E_T)$ is close to zero −see the right 2-dimensional distribution in the second row of plots in Fig. 6.1−). Figures 6.17 and 6.18 show distributions of the W transverse mass for data and the background (plus signal) model in the electron and muon channels respectively. Plots are shown separately for the pretag and b-tag stages (1 or 2 b-tag jets). They are separated also by the number of jets (2, 3 or 4) and the run period (Run IIa (p17) or Run IIb (p20)). In Fig. 6.19 the electron and muon channels and the Run IIa and Run IIb periods are combined.

The Bayesian neural network technique will be used as a next step to further separate single top-like events from the backgrounds. But to be able to do so, good agreement between the data and the background model is required in the distributions that are used as inputs for the neural networks. Therefore, starting from the initial $\sim$ 600 variables, a couple of steps are taken to reduce them in number and keep only those that are well modeled in all the twenty four channels. The first step is performed using the Kolmogorov-Smirnov (KS) test[34]. Each variable is required to have a KS test p-value of at least 0.1 for the majority of the channels. The second step involves looking at the variables distributions by eye (especially those with the lowest KS test p-values) in order to ensure that the data-background agreement is satisfactory. After this selection, a set of 165 variables is retained. However, some of these variables are not defined for events with only 1 (instead of 2) b-tagged jet or only 2 or 3 (instad of 4) jets. Therefore, the actual number of variables left is 114 (120), 146 (152) and 159

[34]For the KS test input, histograms with a large number of bins ($\sim 1,000$) are used to reduce the number of entries in each histograming bin, a necessary condition for the method to give a meaningful answer.

(165) in the 2jet-1tag (2tag), 3jet-1tag (2tag) and 4jet-1tag (2tag) channels.

Neural networks are a machine learning technique. As such, their ability to discriminate between signal and background events depends on the models fed during the learning process. The BNN used in this analysis takes as an input a set of $\sim$ 15−25 1-dimensional kinematic distributions, which of course must be well modeled, but in addition are required to have the highest possible signal versus background discriminating power. In Sect. 7.5 it is explained how these variables are selected from the set of $\sim$ 165 well modeled variables, and in Appendix E tables are provided with the final selected variables for each of the twenty four channels. Here, the definitions of some of these variables are presented and their distributions shown for all the twenty four channels combined.

The variables are classified into the following categories:

- **Object kinematics**
  These are the individual objects transverse energy (E_T) or momentum (p_T) or pseudorapidity (η). In addition, for each object its η coordinate is multiplied by the charge of the reconstructed lepton to take advantage of the CP symmetry in the t-channel production (see Fig. 6.16 and following explanation).

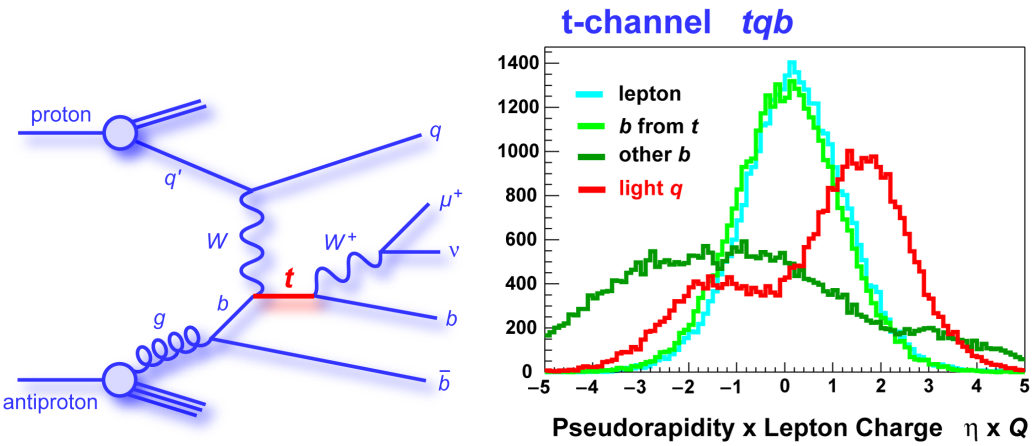

Figure 6.16 : Left: Feynman diagram of the t-channel single top production in a $p\bar{p}$ collision. In this diagram a top quark is produced giving a positively charged lepton (μ^+). Right: Distribution of Q(lepton) $\times$ η(object) for the diagram shown on the left and for its CP conjugate, and where object $\in$ {lepton, b from top decay, b from gluon splitting, light quark (q)}. The asymmetric distribution for the light quark (red curve) is characteristic from this process and serves to separate single top events from background events.

Before the decay of the top quark, the single top t-channel final state consists of a quark pair $t\bar{b}$ ($\bar{t}b$) and a light quark, which in about 80% of the cases corresponds to a d ($\bar{d}$) quark from the weak interaction $u \to W^+d$ ($\bar{u} \to W^-\bar{d}$) with the u ($\bar{u}$)

quark coming from the proton (antiproton). Contrary to what happens with the t ($\bar{t}$) quark, which is produced with a high p_T and very central, the $\bar{b}$ (b) and d ($\bar{d}$) quarks are produced at large angles following the directions of incidence of the antiproton (proton) and proton (antiproton) respectively. Therefore, although the η distribution of the light quark is symmetric, when multiplied by the lepton charge (this tells whether a top or an antitop was produced, and therefore whether the light quark comes from the proton or the antiproton), the distribution presents an asymmetry (see Fig. 6.16 (right)) that characterizes the t-channel and serves as a powerful variable to discriminate the signal against the backgrounds.

- **Event kinematics**
  These are total transverse energy or invariant mass variables.

- **Top quark reconstruction**
  These are variables related to the reconstruction of the top quark, identifying which jet to use, and which neutrino solution to pick:

  - **Neutrino p_z solutions**
    The W boson mass constraint is used on the reconstructed lepton and $\not\!\!E_T$ in order to estimate p_z^ν. The W boson mass m_W is reconstructed as

    $$m_W^2 = (E_{\text{lep}} + E_\nu)^2 - (\vec{p}_{\text{lep}} + \vec{p}_\nu)^2. \tag{6.8}$$

    This quadratic equation gives two possible solutions for p_z^ν. By convention, the solution with the smaller absolute value, denoted as "S1", is usually chosen. However, some variables use the second solution, which is indicated by "S2".

  - **Top mass difference ΔM_{top}**
    The top quark and its mass are reconstructed using all possible combinations of the lepton, the neutrino (including both p_z^ν solutions) and each jet. The invariant mass for each system of (lepton, neutrino, jet) is reconstructed as

    $$M_{\text{top}} = \sqrt{(E_{\text{lep}} + E_\nu + E_{\text{jet}})^2 - (\vec{p}_{\text{lep}} + \vec{p}_\nu + \vec{p}_{\text{jet}})^2}. \tag{6.9}$$

    The difference between M_{top} and 170 GeV/c^2 (which is the value used in the Monte Carlo simulations) is called ΔM_{top}. The combination of (lepton, neutrino, jet) that gives the smallest difference in each event defines the corresponding variables $M_{\text{top}}^{\Delta M^{\text{min}}}$ and $\Delta M_{\text{top}}^{\text{min}}$.

  - **Significance of top quark candidate**
    Assuming the top quark mass resolution function is a Gaussian distribution, the significance of the reconstructed top mass is calculated for each (lepton,

neutrino, jet) combination as

$$\text{Significance}(M_{\text{top}}) = \ln\left(\frac{\text{Gauss}^{\text{Reco}}(M_{\text{top}} = M_{\text{top}}^{\text{Reco}})}{\text{Gauss}^{\text{Reco}}(M_{\text{top}} = 170\ \text{GeV})}\right), \qquad (6.10)$$

where $\text{Gauss}^{\text{Reco}}$ is a Gaussian distribution centered at the reconstructed top quark mass and with standard deviation given by the reconstructed top mass resolution uncertainty, δM_{top}, derived from Eq. (6.9) in terms of the $\not\!\!E_T$ and jet energy resolutions [85]. The lepton energy resolution is ignored as the lepton energy is well measured compared to that of the jets and $\not\!\!E_T$, and no correlations between jet energy resolution and $\not\!\!E_T$ resolution are taken into account. The reconstructed top quark mass that gives the smallest significance in each event is called $M_{\text{top}}^{\text{sig}}$, and the corresponding significance $\text{Significance}_{\text{min}}(M_{\text{top}})$.

- **Angular correlations**
  These are either the ΔR, $\Delta\phi$ angles between jets and leptons, or top quark spin correlation variables. The $V-A$ nature of the weak interactions results in the final state lepton and neutrino having their spins predominantly aligned with that of the top quark and so being emitted in opposite and same directions respectively to that of the top quark. Thus, variables as the cosine of the angle between the lepton or neutrino and the jets momentum provide discrimination between single top signal and background [86].

- **Jet reconstruction**
  This category includes the jet widths in η and ϕ, as well as jet mass (reconstructed from the distribution of energy within the jet) and the relative p_T between the leading jet and the tagging muon that this jet may contain.

Table 6.14 lists the discriminating variables used in the BNN analysis. The variables are shown in Figs. 6.20 to 6.29 for all analysis channels combined. Good data-background agreement is seen in all samples.

Discriminating Variables

Object kinematics	**Top quark reconstruction**	**Jet reconstruction**
p_T(lepton)	$M(W$,jet1) (leading jet top mass)	Width$_\eta$(jet1)
$\not\!\!E_T$	$M(W$,jet1,$S2)$	Width$_\eta$(jet2)
p_T(jet1)	$M(W$,jet2)	Width$_\eta$(jet4)
p_T(jet2)	$M(W$,jet2,$S2)$	Width$_\eta$(tag1)
p_T(jet3)	$M(W$,jet3,$S2)$	Width$_\eta$(tag2)
p_T(jet4)	$M(W$,jet4)	Width$_\eta$(best1)
p_T(tag2)	$M(W$,tag1) ("b-tagged" top mass)	Width$_\eta$(light2)
p_T(light2)	$M(W$,tag1,$S2)$	Width$_\eta$(notbest2)
p_T(best1)	$M(W$,tag2)	Width$_\phi$(jet1)
p_T(notbest2)	$M(W$,light1)	Width$_\phi$(jet2)
E(jet2)	$M(W$,light1,$S2)$	Width$_\phi$(jet4)
E(light1)	$M(W$,best1) ("best" top mass)	Width$_\phi$(tag2)
Q(lepton)$\times\eta$(jet1)	$M(W$,best1,$S2)$	Width$_\phi$(light1)
Q(lepton)$\times\eta$(tag2)	$M(W$,notbest1,$S2)$	Width$_\phi$(light2)
Q(lepton)$\times\eta$(light1)	$M_{\rm top}^{\Delta M^{\rm min}}$	Width$_\phi$(best1)
Q(lepton)$\times\eta$(light2)	$\Delta M_{\rm top}^{\rm min}$	Width$_\phi$(notbest1)
Q(lepton)$\times\eta$(best1)	Significance$_{\rm min}(M_{\rm top})$	$p_T^{\rm rel}$(jet1, μ)
Q(lepton)$\times\eta$(notbest1)	$p_z(\nu, S2)$	M(jet1)
		M(tag1)
Event kinematics	**Angular correlations**	M(best1)
Aplanarity(W,alljets)	ΔR(jet1,jet2)	
Centrality(alljets)	ΔR(lepton,jet1)	
Sphericity(W,alljets)	ΔR(lepton,tag1)	
H_T(alljets)	ΔR(lepton,best1)	
H_T(alljets$-$tag1)	$\Delta R^{\rm min}$(alljets)	
H_T(alljets$-$best1)	$\Delta R^{\rm min}$(lepton,alljets)	
H_T(jet1,jet2)	$\Delta\phi$(lepton,$\not\!\!E_T$)	
H_T(lepton,$\not\!\!E_T$)	$\Delta\phi$(lepton,tag1)	
H_T(lepton,$\not\!\!E_T$,alljets)	$\Delta\phi$(jet1,$\not\!\!E_T$)	
H_T(lepton,$\not\!\!E_T$,jet1,jet2)	cos(lepton,jet1)$_{\rm btaggedtop}$	
M(alljets)	cos(lepton,tag1)$_{\rm btaggedtop}$	
M(alljets$-$tag1)	cos(lepton$_{\rm btaggedtop}$,btaggedtop$_{\rm CMframe}$)	
M(alljets$-$best1)	cos(lepton,light1)$_{\rm btaggedtop}$	
M(jet1,jet2)	cos(lepton,best1)$_{\rm besttop}$	
M(jet3,jet4)	cos(best1,notbest1)$_{\rm besttop}$	
M(light1,light2)	cos(lepton$_{\rm besttop}$,besttop$_{\rm CMframe}$)	
M_T(jet1,jet2)	cos(lepton,Q(lepton)$\times z)_{\rm besttop}$	
p_T(jet1,jet2)		
$\sqrt{\hat{s}}$		
$M_T(W)$		

Table 6.14 : Variables used in the BNN analysis, in five categories: object kinematics, event kinematics, jet reconstruction, top quark reconstruction and angular correlations (where the subscript indicates the reference frame).

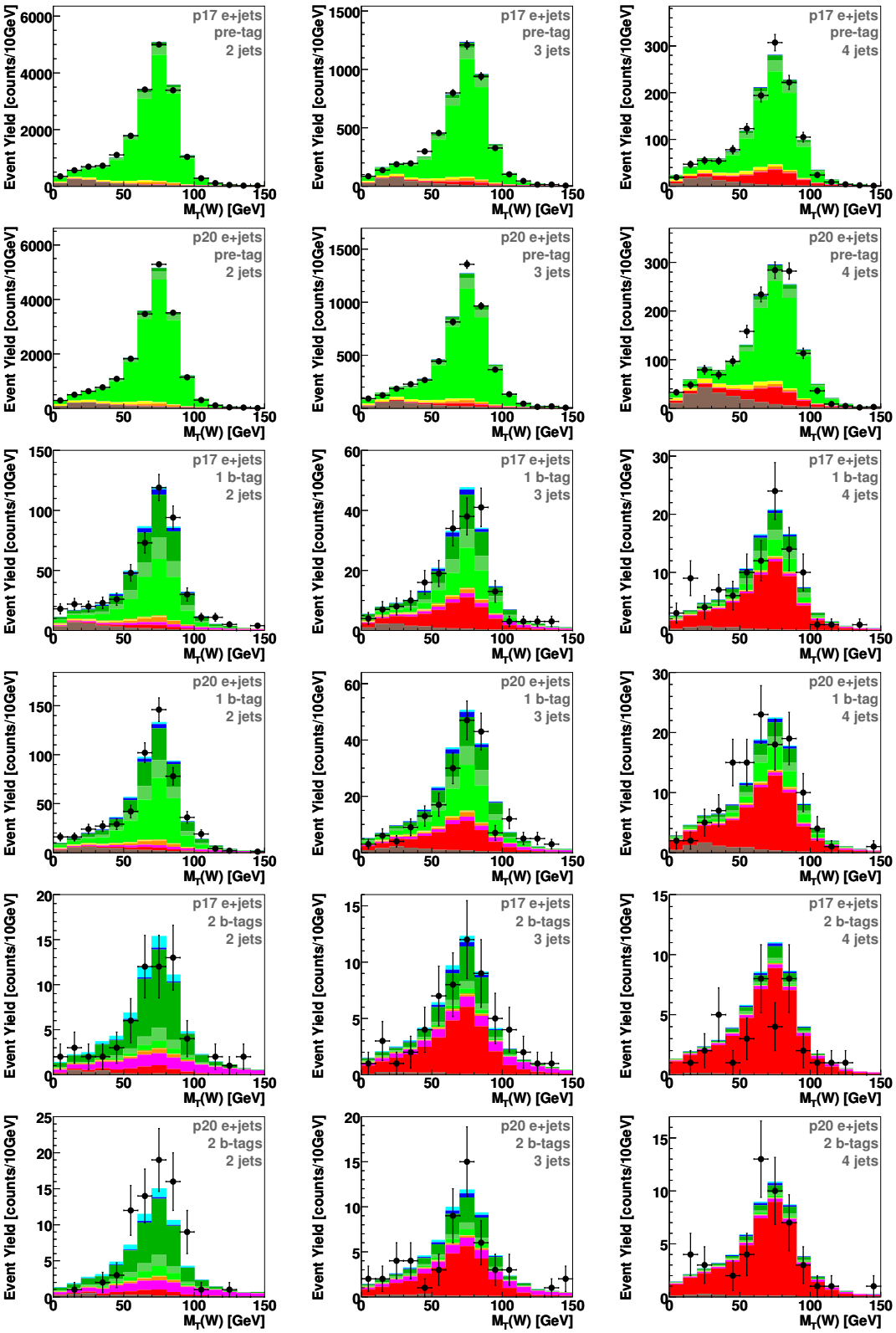

Figure 6.17 : The W boson transverse mass distributions in the electron channel for 2-jet (left column), 3-jet (middle), and 4-jet events (right), for p17 and p20 (alternate rows), pretag (top two rows), single-tag (middle two rows), and double-tag (bottom two rows) events.

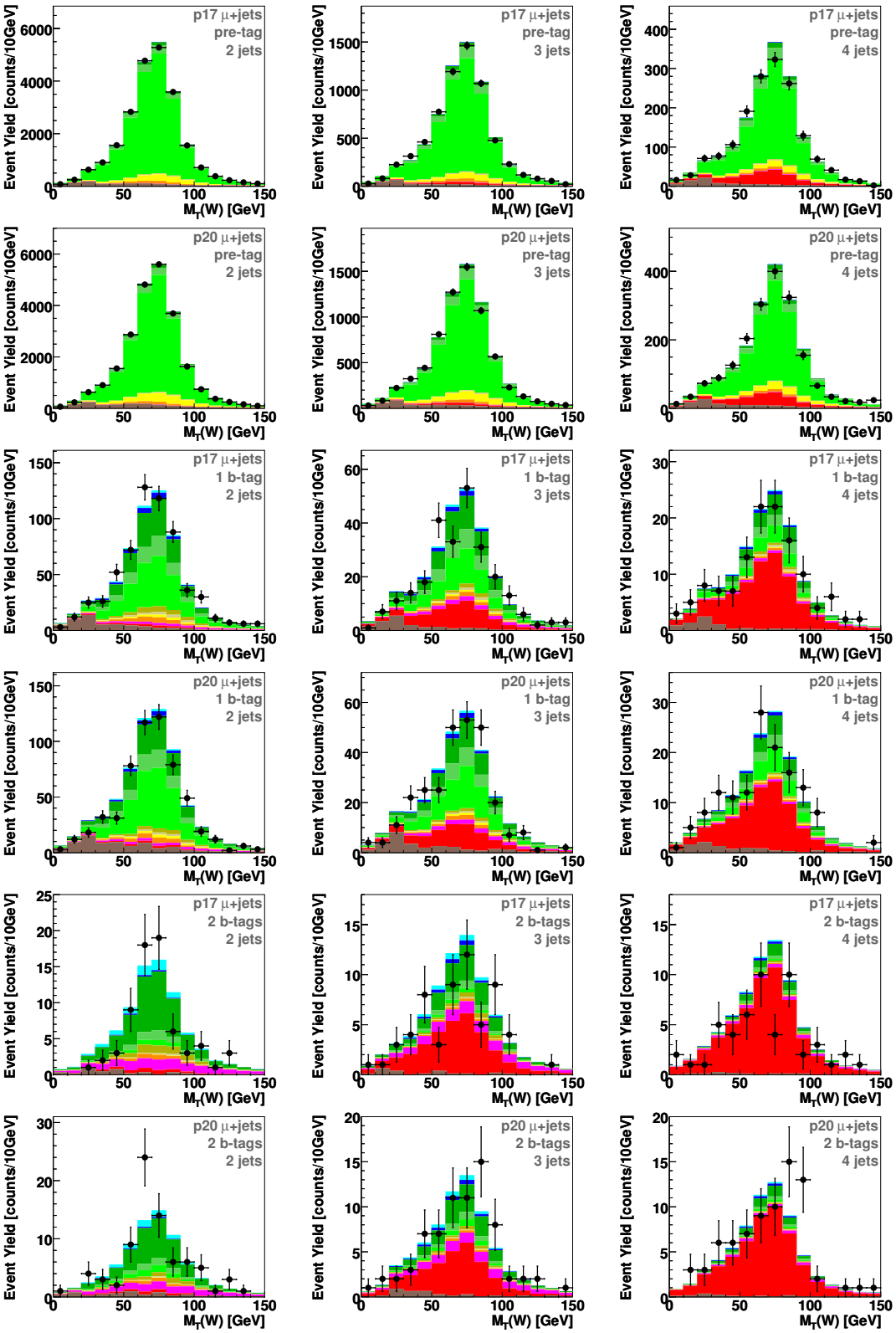

Figure 6.18 : The W boson transverse mass distributions in the muon channel for 2-jet (left column), 3-jet (middle), and 4-jet events (right), for p17 and p20 (alternate rows), pretag (top two rows), single-tag (middle two rows), and double-tag (bottom two rows) events.

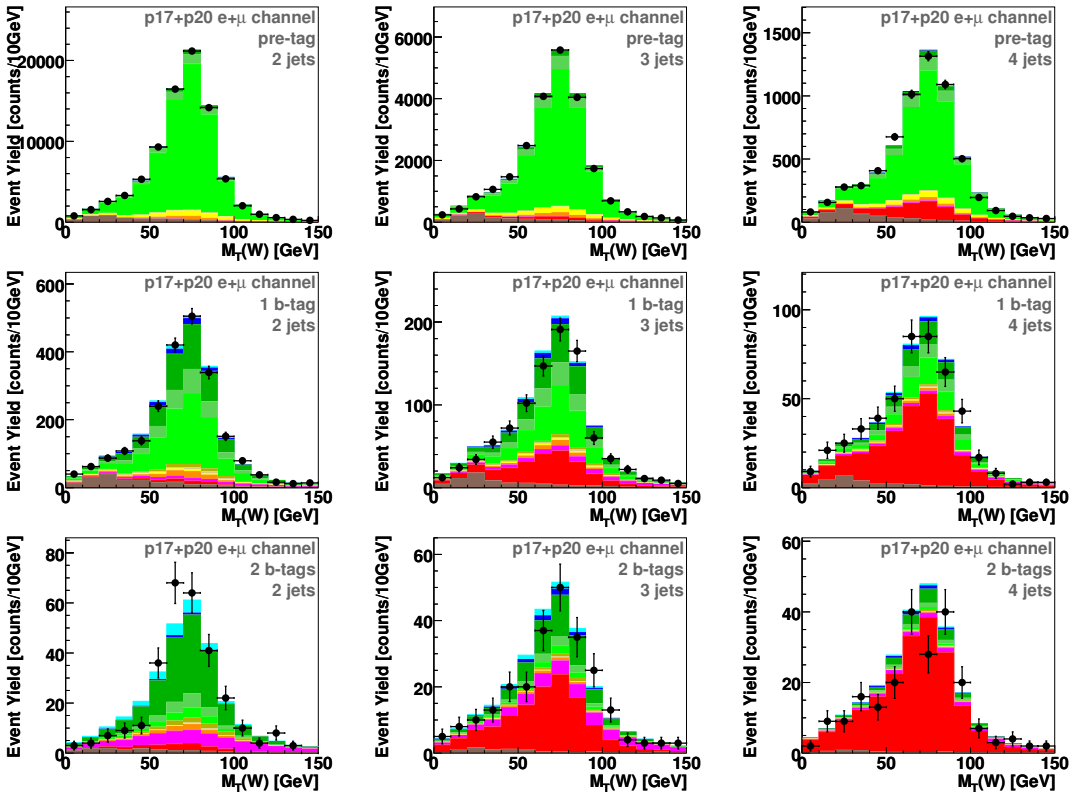

Figure 6.19 : The W boson transverse mass distributions in the electron and muon channel and Run IIa and Run IIb dataset combined, for 2-jet (left column), 3-jet (center), and 4-jet events (right), pretag (top row), single-tag (middle row), and double-tag (bottom row) events.

OBJECT KINEMATICS (I)

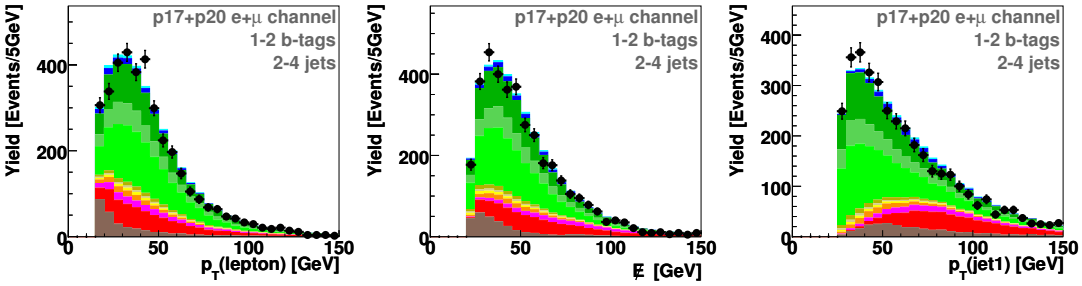

Figure 6.20 : Individual object kinematic variables used in the BNN analysis (part 1), for all channels combined.

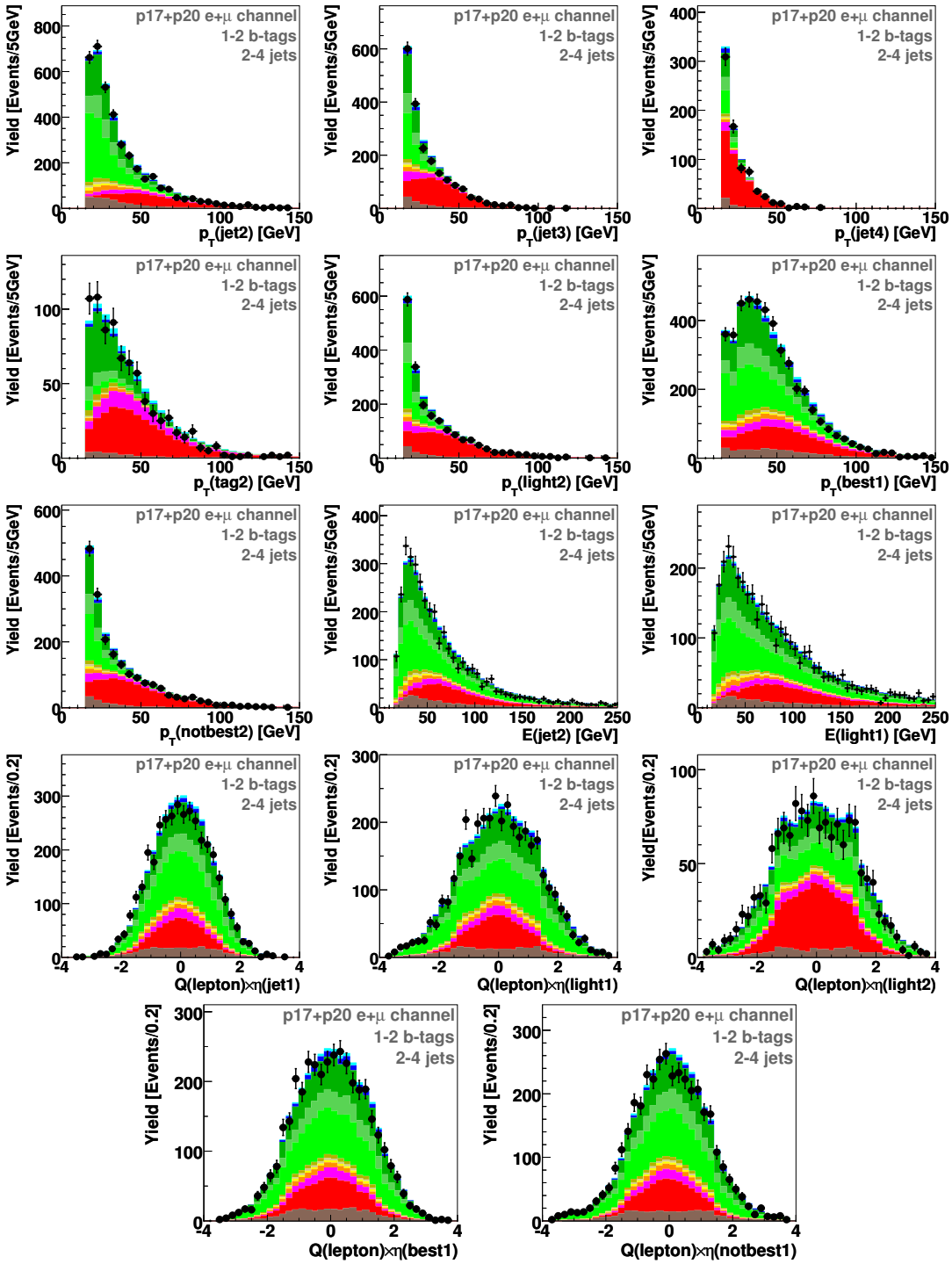

Figure 6.21 : Individual object kinematic variables used in the BNN analysis (part 2), for all channels combined.

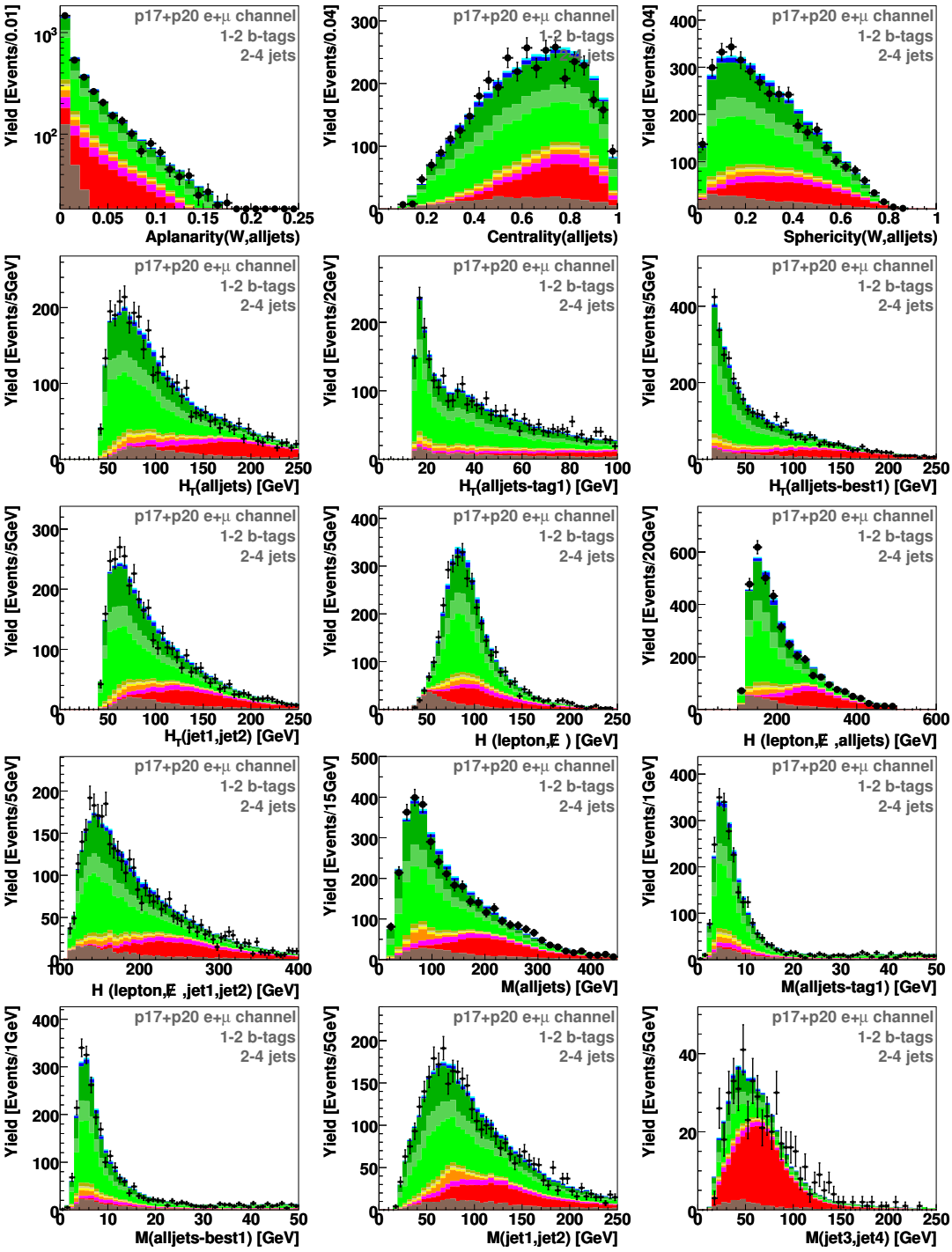

Figure 6.22 : Event kinematic variables used in the BNN analysis (part 1), for all channels combined.

EVENT KINEMATICS (II)

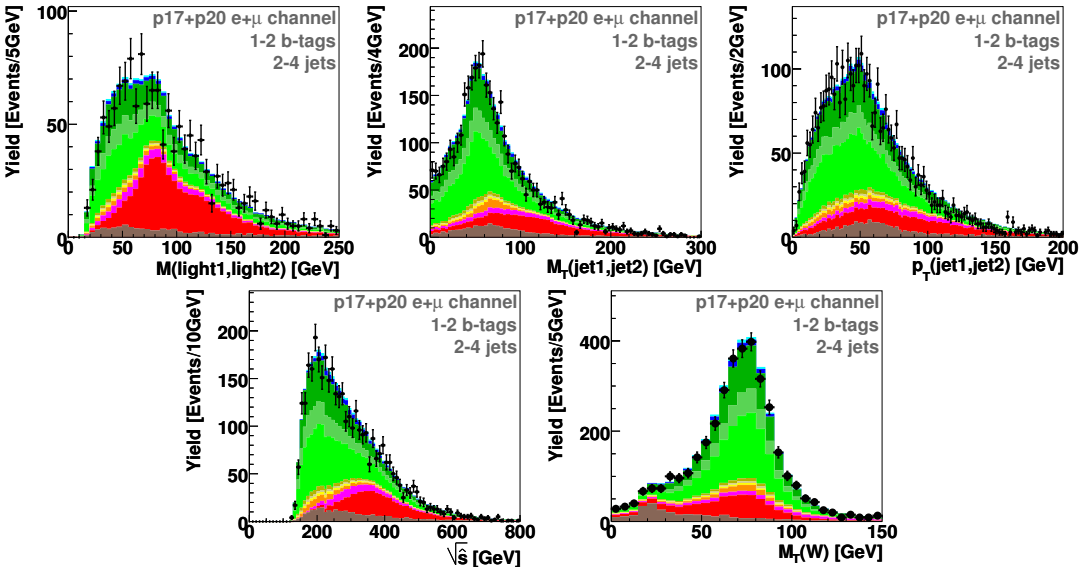

Figure 6.23 : Event kinematic variables used in the BNN analysis (part 2), for all channels combined.

TOP QUARK RECONSTRUCTION (I)

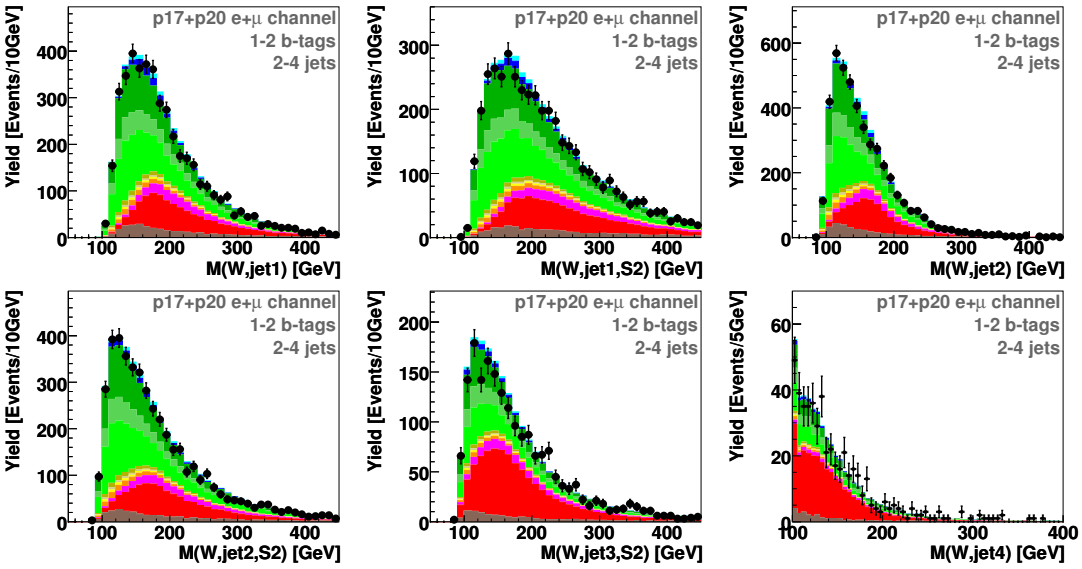

Figure 6.24 : Top quark reconstruction related variables used in the BNN analysis (part 1), for all channels combined.

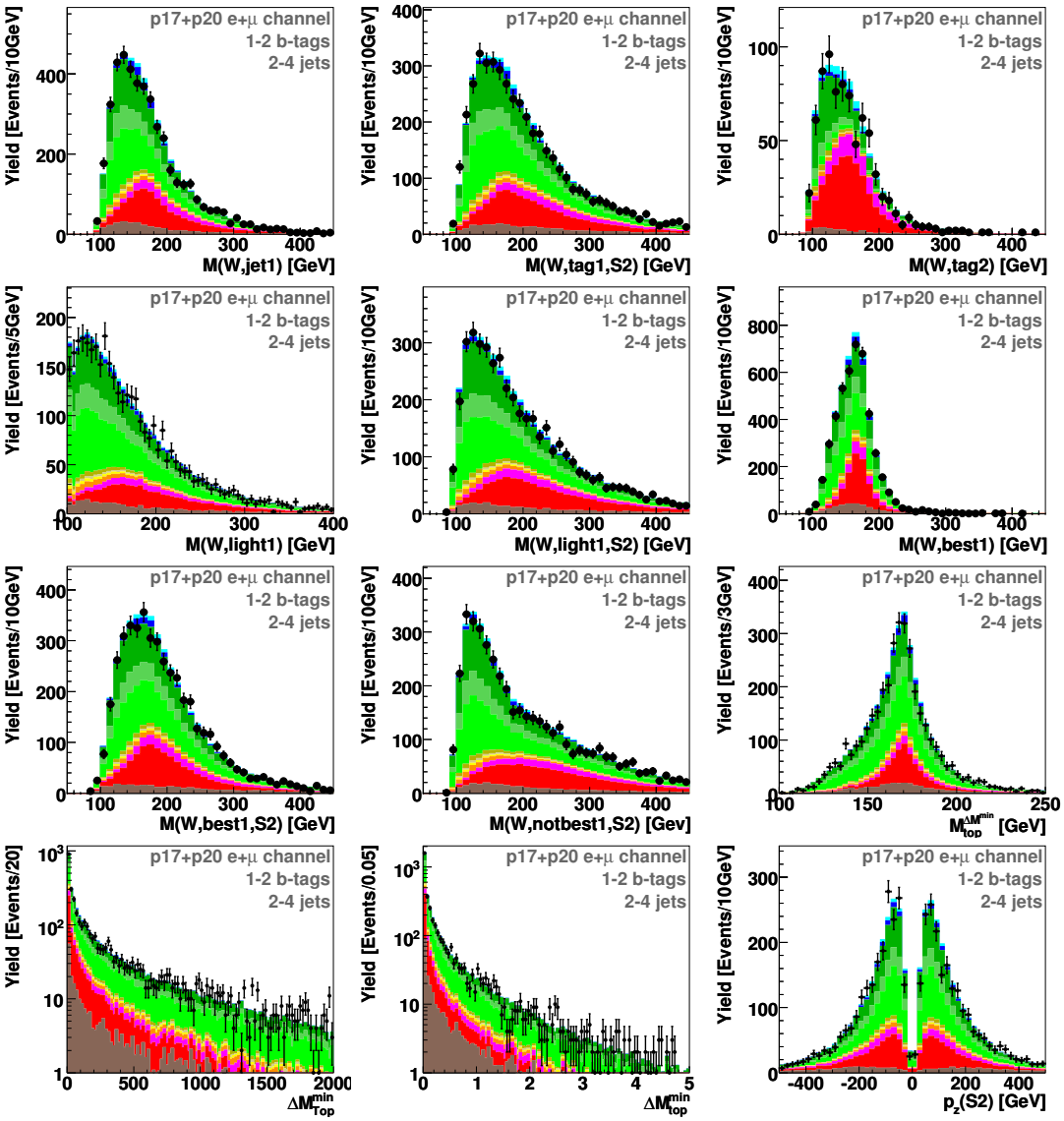

Figure 6.25 : Top quark reconstruction related variables used in the BNN analysis (part 2), for all channels combined.

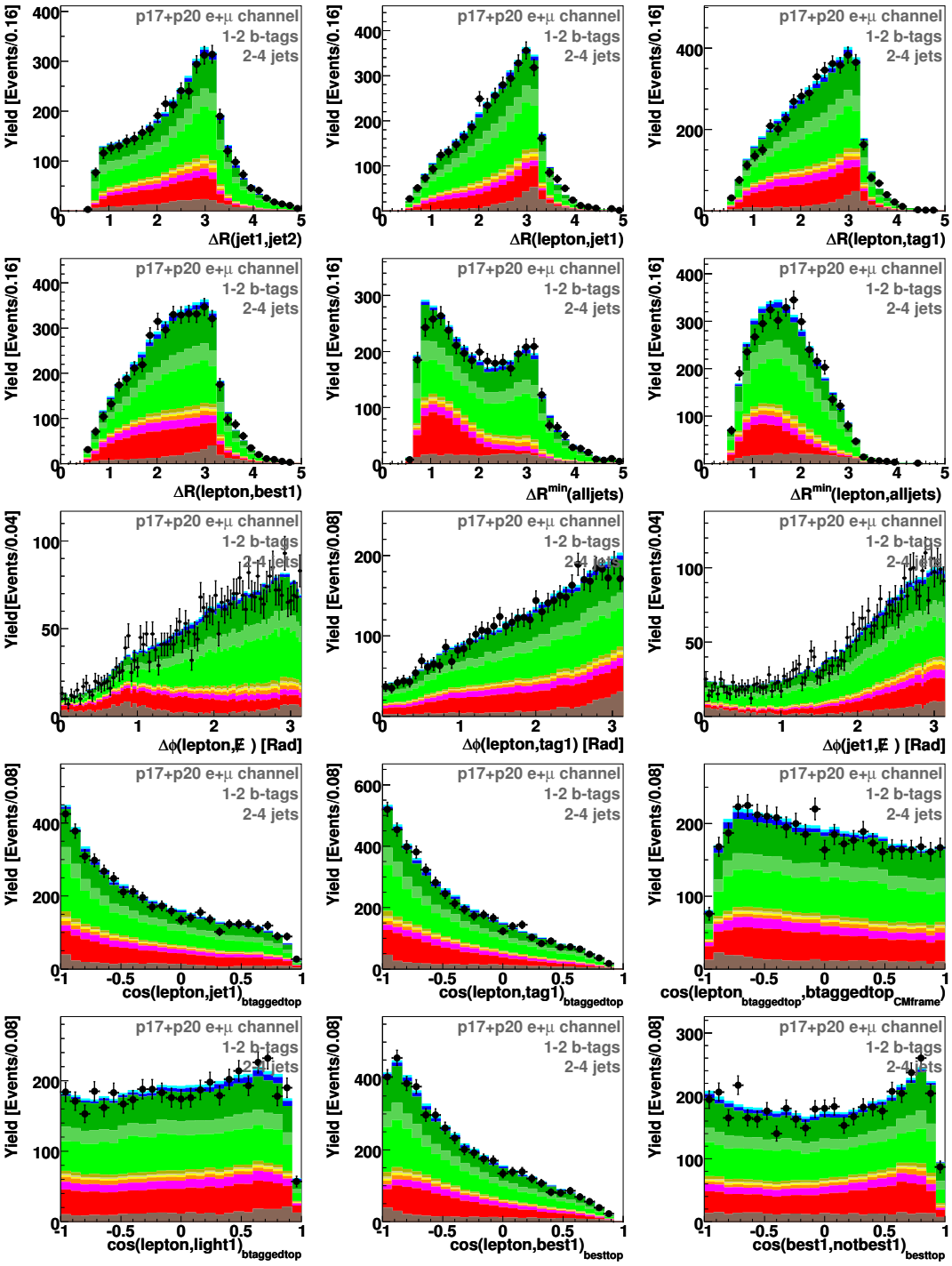

Figure 6.26 : Angular correlation variables used in the BNN analysis (part 1), for all channels combined.

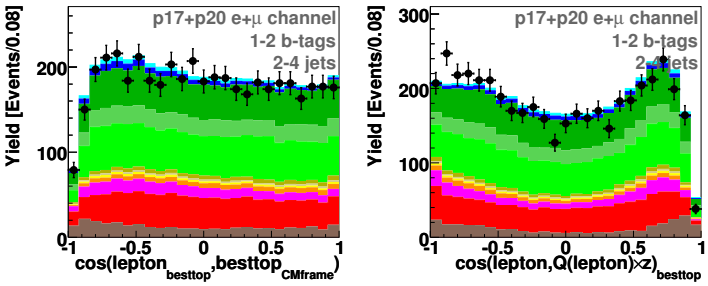

Figure 6.27 : Angular correlation variables used in the BNN analysis (part 2), for all channels combined.

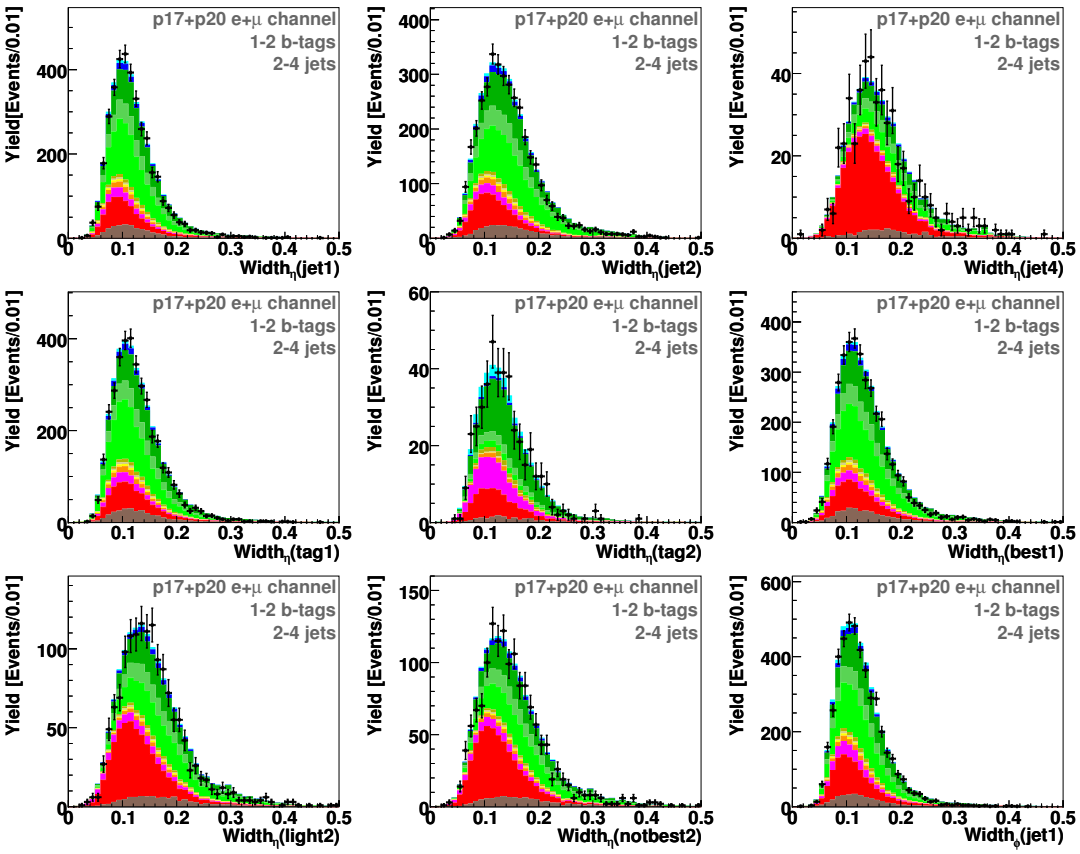

Figure 6.28 : Jet reconstruction variables used in the BNN analysis (part 1), for all channels combined.

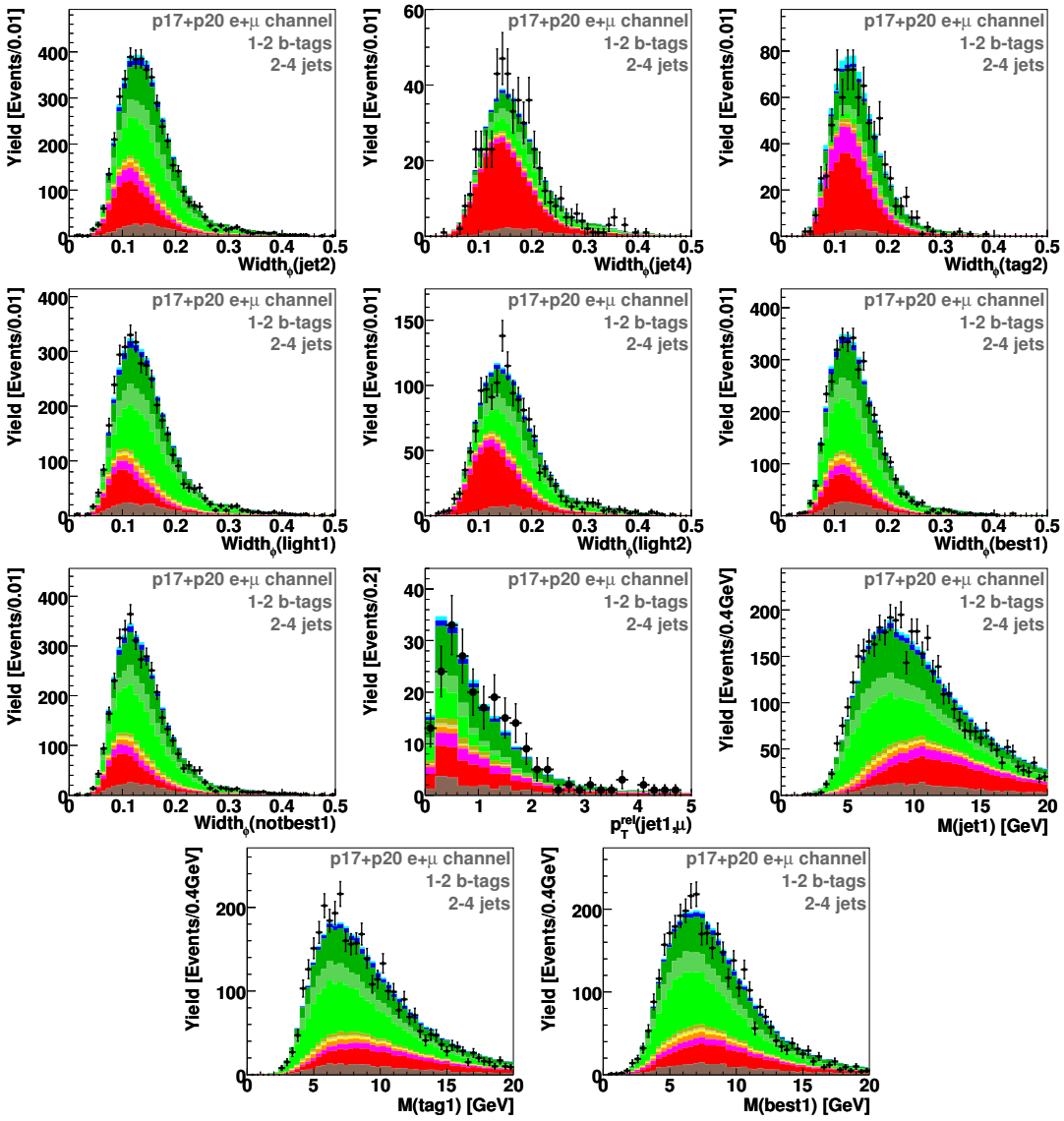

Figure 6.29 : Jet reconstruction variables used in the BNN analysis (part 2), for all channels combined.

6.7 Cross Check Samples

Two cross-check samples are selected for each of the electron and muon channels to test whether the background model reproduces the data in regions dominated by one type of background. The selection criteria are the same as for the main event samples, but then the following additional requirements are imposed:

"W+jets" sample

- Exactly two jets
- H_T(lepton,$\not\!\!E_T$,alljets) < 175 GeV
- One b-tagged jet

"$t\bar{t}$" sample

- Exactly four jets
- H_T(lepton,$\not\!\!E_T$,alljets) > 300 GeV
- One or two b-tagged jets

Table 6.15 shows the event yields for these samples, and Table 6.16 shows the yields with backgrounds and channels combined. The sum of the backgrounds agrees reasonably well with the observed data in each channel. For the W+jets samples, W+jets events form 84% (electron channel) and 74% (muon channel) of the samples, and the $t\bar{t}$ component corresponds to only 1%. For the $t\bar{t}$ samples, $t\bar{t}$ events form 85% (electron) and 82% (muon) of the samples, and the W+jets events correspond to only 10% (electron) and 11% (muon).

Yields for the Cross-Check Samples

	W+Jets Samples				$t\bar{t}$ Samples			
	Run IIa		Run IIb		Run IIa		Run IIb	
	e	μ	e	μ	e	μ	e	μ
Signals								
tb	3	3	3	3	1	1	1	1
tqb	6	7	6	7	1	1	1	1
Backgrounds								
$t\bar{t} \to \ell\ell$	3	3	3	3	**5**	**5**	**5**	**5**
$t\bar{t} \to \ell$+jets	1	1	1	1	**66**	**80**	**68**	**77**
$Wb\bar{b}$	**78**	**88**	**76**	**80**	3	5	4	5
$Wc\bar{c}$	**37**	**44**	**40**	**44**	3	3	2	3
Wjj	**106**	**125**	**133**	**148**	2	3	3	4
Z+jets	7	30	6	25	1	2	1	1
Dibosons	8	12	9	12	0	1	0	1
Multijets	16	31	18	48	1	2	1	3
Backgrounds+Signals	265	344	295	371	83	103	86	101
Data	289	358	320	344	74	97	84	113

Table 6.15 : Yields after selection in the cross-check samples. Shown in bold is the W+jets ($t\bar{t}$) contribution to the "W+jets" ("$t\bar{t}$") cross-check sample.

Figure 6.30 shows the W transverse mass distribution for the cross-check samples, for all channels combined (electron and muon, Run IIa and Run IIb). Plots for separate channels, for this variable and 12 others, are shown in Appendix C.

Yields for the Cross-Check Samples
with Channels and Backgrounds Combined

	W+Jets Samples		$t\bar{t}$ Samples	
Signal				
tb+tqb	38	(3%)	8	(2%)
Backgrounds				
$t\bar{t}$	16	(1%)	311	(83%)
W+jets	999	(78%)	40	(11%)
Z+jets & dibosons	109	(9%)	7	(2%)
Multijets	113	(9%)	7	(2%)
Backgrounds+Signals	1,275		373	
Data	1,311		368	

Table 6.16 : Yields after selection in the cross-check samples, for channels and backgrounds combined. The numbers in parentheses are the percentages of the total background+signal for each source.

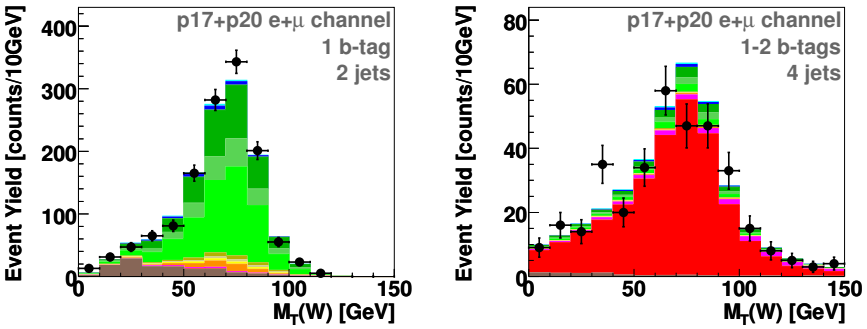

Figure 6.30 : The W transverse mass distribution for the "W+jets" and "$t\bar{t}$" cross-check samples, for all channels combined.

6.8 Systematic Uncertainties

Systematic uncertainties may affect the analysis in two different ways: *i*) as uncertainties on the normalization of the background samples, *ii*) as effects that change the shape of the distributions of the background samples and of the expected signal. This section describes all the uncertainties considered and summarizes them in Table 6.17.

Because W+jets and multijets are normalized to data before b-tagging, their yield estimates are not affected by any of the systematic uncertainties that affect the overall yield. The exception to this is b-tagging, which is applied after normalization. There is still an effect on the shapes of distributions from the uncertainty components that depend on event kinematics. The three systematic sources that affect the shape of the distributions are jet energy scale, tag-rate functions, and ALPGEN W+jets reweighting factors. These uncertainties are discussed in more detail in Appendix D. They are

treated differently in the calculations, including them in each histogram bin of the final analysis discriminant.

The following systematic uncertainties are considered in the analysis:

- **Integrated luminosity**
  There is a 6.1% uncertainty on the luminosity estimate that affects the signal, $t\bar{t}$, Z+jets, and diboson yields.

- **Theory cross sections**
  The uncertainties on the single top and $t\bar{t}$ cross sections come from Refs. [4] and [87] respectively. Components for renormalization and factorization scales, PDF, kinematics, and top quark mass choice are combined in quadrature. The mass component of the uncertainty is taken as the difference between the cross section at 170 GeV/c^2 (the value the analysis is performed at) and at the world average value of 172.4 GeV/c^2 [88] plus one standard deviation of 1.2 GeV/c^2. The resulting uncertainties are $+4.3\%, -11.2\%$ for s-channel tb, $+5.5\%, -7.4\%$ for t-channel tqb, and $+7.7\%, -12.7\%$ for $t\bar{t}$. For tb+tqb combined, the uncertainty is $+5.2\%, -8.4\%$. (The uncertainty on the signal cross section does not of course enter the experimental measurement. Components of it are used in the measurement of $|V_{tb}|$.) The uncertainties on the diboson cross sections from the renormalization and factorization scales and PDF were calculated at NLO using the MCFM generator [74]: for WW the uncertainty is 5.6%, for WZ it is 6.8%, and for ZZ it is 5.5%, and on the sum of the three processes it is 5.8%. The average value of 5.8% is used for the Z+jets background, since there are no direct measurements for this small background component.

- **Branching fractions**
  The branching fractions for a W boson to decay to an electron, muon, or tau lepton, have an average uncertainty of 1.5% (Ref. [1]) and this is included in the MC normalization uncertainties.

- **Parton distribution functions**
  The effect of changing the parton distribution functions on the signals is evaluated by reweighting single top MC events according to the 40 sets of CTEQ error PDFs and measuring the signal acceptance for each of them. A systematic uncertainty of 3% is assigned to the signal acceptances from this source. The PDF uncertainty on the $t\bar{t}$, Z+jets, and diboson background yields is included in the theory cross section uncertainties listed above.

- **Trigger efficiency**
  This analysis uses an OR of many trigger conditions which gives a trigger efficiency close to 100%. A flat uncertainty of 5% is assigned to the trigger efficiency in all channels, except Run IIb muon where it is raised to 10%, based on the plots shown in Appendix B. This uncertainty is treated as uncorrelated between Run IIa and

Run IIb and between electrons and muons. The trigger uncertainty does not have a shape dependence in the final multivariate outputs (see Appendix 1 in Ref. [85]).

- **Instantaneous luminosity reweighting**
  The instantaneous luminosity distributions of all MC samples are reweighted to make them match Run IIa or Run IIb data distributions as appropriate. The initial distributions are from the ZB data overlaid on the MC events to simulate multiple interactions, and are generally at too low values for later data-taking conditions. The uncertainty on this reweighting is 1.0%.

- **Primary vertex modeling and selection**
  The distributions of the z position of the PV in all MC samples are reweighted to match those in data [75]. The uncertainty on this reweighting is 0.05% (negligible). The uncertainty on the difference in PV selection efficiency between data and MC is 1.4%.

- **Electron reconstruction and identification efficiency**
  The uncertainty on the MC scale factor for electron reconstruction and identification includes the dependence of the electron identification scale factor on the variables ignored in the parametrization (jet multiplicity, track match, likelihood, electron p_T and ϕ coordinate), and the limited statistics in each bin of the parametrization. The assigned total uncertainty is 2.5%.

- **Muon reconstruction and identification efficiency**
  The MC scale factor uncertainties for muon reconstruction and identification, including isolation requirements, are estimated by the muon ID group as coming from the tag/probe method, background subtraction, and limited statistics in the parametrization. The assigned total uncertainty is 2.5%.

- **Jet fragmentation**
  This sytematic uncertainty is estimated by comparing the acceptance of $t\bar{t}$ events generated with ALPGEN+PYTHIA (as used in this analysis) to those generated with ALPGEN+HERWIG, with all other generation parameters unchanged. The uncertainty in the $t\bar{t}$ (diboson and signal) MC samples is 0.7% in the two-jet channels, 0.1% (3.7%) in the three-jet channels and 0.7% (4.7%) in the four-jet channels. For the Z+jets MC samples the assigned uncertainty is 4% in all channels.

- **Initial-state and final-state radiation**
  This uncertainty is evaluated in $t\bar{t}$ samples generated with variations of these effects within expectations, and is found to be 3% (0.6%) in the $t\bar{t}$ (diboson and signal) MC samples in the two-jet channels, 2.8% (5.2%) in the three-jet channels and 0.6% (12.6%) in the four-jet channels. In the Z+jets MC samples, the assigned uncertainty is 8% in all channels.

- **b-jet fragmentation**
  The size of the uncertainty from the b jet modeling was evaluated in the $t\bar{t}$ pairs cross section analysis following the method described in Ref. [89]. The uncertainty arises from the difference between the fragmentation parametrizations preferred by SLD vs. LEP data. A 2.0% value is used which affects the $t\bar{t}$ and $Z\bar{b}b$ samples.

- **Jet reconstruction and identification**
  The efficiency to reconstruct and identify jets has an uncertainty of 1%.

- **Jet energy scale**
  The JES correction is raised and lowered by one standard deviation on each MC sample and the whole analysis repeated, which produces a shape-changing uncertainty, and an overall normalization uncertainty (shown for each channel in Appendix D). The normalization part ranges from 1.1% to 13.1% on the signal acceptance and from 0.1% to 2.1% on the combined background.

- **Jet energy resolution**
  The JER uncertainty is estimated in the same manner as the JES uncertainty. A flat 4% uncertainty is assigned, which covers the shape variations for all signals and backgrounds.

- **ALPGEN reweighting**
  The ALPGEN W+jets backgrounds are reweighted based on several pretag data distributions (see Sect. 6.2). The uncertainty from these reweightings affects the shapes of the W+jets background components, and does not change the uncertainty on their normalization. Appendix D presents the BNN discriminant for each analysis channel when the size of the reweighting is raised and lowered by one standard deviation on the W+jets MC samples.

- **Multijets and W+jets normalization**
  The multijets and W+jets backgrounds are normalized to the pretag data following the procedure described in Sect. 5.5.3, where the determination of the corresponding uncertainties is also explained. The normalization uncertainties range from 30% to 54% for the multijets backgrounds and from 1.8% to 5.0% for the W+jets backgrounds, depending on the analysis channel (see Table 5.5 for the uncertainties in each channel).

- **Taggability and tag-rate functions for MC events**
  The uncertainty associated with b-tagging in MC events is evaluated by adding the taggability and the tag rate components of the uncertainty in quadrature. The TRF values are raised and lowered by one standard deviation on each MC sample and the whole analysis repeated (plots are shown in Appendix D). This is done simultaneously for heavy flavor and light jets. This leads to the uncertainties for heavy flavor TRFs and light jets TRFs to be correlated, while they should be uncorrelated. Since treating uncertainties as correlated is equivalent to adding

them lineary and not quadratically, in this analysis the total TRF uncertainties are overestimated. The TRF uncertainties originate from several sources: statistical errors of MC event sets, the assumed fraction of heavy flavor in the multijets MC events for the mistag rate determination, and the parametrizations. These uncertainties affect both shape and normalization of the MC samples. The normalization part of the uncertainty ranges from 2.3% (9.9%) to 4.7% (10.8%) for single-tag (double-tag) signal acceptances, and from 2.1% (9.0%) to 7.0% (11.4%) for single-tag (double-tag) combined backgrounds.

- **W+jets heavy-flavor scale factor correction**
  The heavy-flavor scale factor correction, S_{HF}, for $Wb\bar{b}$ and $Wc\bar{c}$ is measured in data in several channels, as described in Sect. 5.6. The MC tag rate function uncertainty induces fluctuations in the effective scale factor that are at least as large as the channel-to-channel variations in the measurement. Therefore, it can be argued that any additional systematic is double-counting. However, a conservative uncertainty of 13.7% is assigned on the scale factor. Accordingly, an anticorrelated uncertainty of 0.7-1.5% (depending on the analysis channel) is assigned to Wlp to keep the W+jets yield constant. Finally, an additional 5% anticorrelated uncertainty is assigned to the $Wb\bar{b}$ and $Wc\bar{c}$ yields to account for the assumption that $Wc\bar{c}$ and $Wb\bar{b}$ have the same k-factors.

- **Z+jets heavy-flavor scale factor correction**
  The k-factors for $Zb\bar{b}$ and $Zc\bar{c}$ are determined from NLO calculations, and the correction to them, S^{Z}_{HF}, is set to be 1.0 (since they cannot be normalized to data to check it) with an uncertainty of 13.7%, taken from the S_{HF} for W+jets measurement. In the muon channel the uncertainty increases to 20%.

- **Sample statistics**
  The MC and data samples used to estimate the signal and background shapes are limited in size. In particular the multijets background samples have low statistics after b-tagging, as seen in Tables 6.5 and 6.6. The statistical uncertainty on the background is taken into account for each sample in each bin of the final discriminant distribution.

All systematic uncertainties are assumed to be fully correlated between all analysis channels and the corresponding signal and/or background samples. The exceptions are the systematic uncertainties on the theoretical cross sections used to normalize the MC backgrounds (except for the two $t\bar{t}$ backgrounds), and the statistical uncertainty from the size of each MC sample. Also, the lepton ID uncertainty is independent between electrons and muons and Run IIa and Run IIb, and the trigger uncertainty is independent between electrons and muons, Run IIa and Run IIb, and jet multiplicities.

Relative Systematic Uncertainties

Components for Normalization	
Integrated luminosity	6.1%
$t\bar{t}$ cross section	12.7%
Z+jets cross section	5.8%
Diboson cross sections	5.8%
Branching fractions	1.5%
Parton distribution functions	3.0%
(signal acceptances only)	
Triggers	5.0%
Instantaneous luminosity reweighting	1.0%
Primary vertex selection	1.4%
Lepton identification	2.5%
Jet fragmentation	(0.7-4.0)%
Initial-and final-state radiation	(0.6-12.6)%
b jet fragmentation	2.0%
Jet reconstruction and identification	1.0%
Jet energy resolution	4.0%
W+jets heavy-flavor correction	13.7%
Z+jets heavy-flavor correction	13.7%
W+jets normalization to data	(1.8-5.0)%
multijet normalization to data	(30-54)%
MC and multijets statistics	(0.5–16)%
Components for Normalization and Shape	
Jet energy scale for signal	(1.1-13.1)%
Jet energy scale for total background	(0.1-2.1)%
(not shape for Z+jets or dibosons)	
b-tagging, single-tag	(2.1-7.0)%
b-tagging, double-tag	(9.0-11.4)%
Component for Shape Only	
ALPGEN reweighting	—

Table 6.17 : A summary of the relative systematic uncertainties for each of the correction factors or normalizations. The uncertainty shown is the error on the correction or the efficiency, before it has been applied to the MC or data samples.

7 Bayesian Neural Networks

After the cut-based event selection described in the previous chapters is applied, the expected single top signal is still smaller than the uncertainty on the background. Therefore, a more sophisticated method is required to enhance the signal relative to the background. In this thesis, the method chosen is a multivariate machine learning technique, namely, a Bayesian neural network (BNN) [90–92]. Two other different methods are used independently by parallel analyses carried out within the "single top effort" in DØ. These are boosted decision trees (BDTs) [93–95] and matrix element calculations (ME) [96, 97]. The final result is a BNN combination of the three. This chapter starts by explaining how independent subsamples of signal and background events are generated to train and test the BNN and measure the cross section. Next, in Sect. 7.2, an overview is given about what any multivariate technique attempts to achieve. Then a description of the BNN used in this analysis is presented, and finally the results (outputs) it yields are shown.

7.1 Splitting of the Samples

The signal and background samples are divided into three independent subsets. The first one, called the "training" sample, is used to create the multivariate filters (e.g., train the BNN and BDTs). The second subset, called the "testing" sample, is used to train the BNN used in the combination [98] of the three multivariate methods used in the single top DØ analysis. The third subset, called the "yield" sample, is used for the final measurements.

When splitting the samples, permutations from the b-tagging of MC events have to be considered. The permutations of each event are highly correlated, and must hence all go into the same subset to avoid correlation between the subsets. To get around this problem, the samples are divided based on the modulus of the event number, which is assigned to the MC events during the generation process, being thus independent of b-tagging permutations.

The splitting procedure results in three subsets of very similar sizes (slightly different number of events per subset arise due to the fact that the event number is not perfectly evenly distributed). Table 7.1 shows as an example the number of signal and background events in the training subset for each of the twenty four analysis channels. For each signal and background component, a normalization factor is applied to all events in each subset such that the total sum of weights becomes the same as before splitting. This is especially important for the yield subset, since this is the one used to extract the single top cross section by comparing the predicted yields in it to data. As expected, this normalization factor is in all cases very close to 3.0.

Number of Events in the Training Subset

	Electron Channel		Muon Channel	
	tb+tqb	All Bkgrs	tb+tqb	All Bkgrs
	Run IIa			
1tag / 2jets	46,366	302,898	44,799	341,228
1tag / 3jets	23,051	258,371	24,721	284,938
1tag / 4jets	8,270	207,966	8,490	233,693
2tags / 2jets	23,313	160,135	22,411	182,577
2tags / 3jets	23,850	272,907	25,781	303,647
2tags / 4jets	12,923	328,837	13,240	369,986
	Run IIb			
1tag / 2jets	57,082	172,665	63,773	218,302
1tag / 3jets	29,699	146,837	38,124	183,891
1tag / 4jets	11,553	106,754	14,818	130,225
2tags / 2jets	28,658	90,837	31,918	116,420
2tags / 3jets	30,766	155,450	39,147	195,153
2tags / 4jets	18,051	168,953	22,837	205,723

Table 7.1 : Number of signal and background events in the training sample for each of the 24 analysis channels.

7.2 Overview

The analysis attempts to approximate a one-to-one function of the probability $\Pr(S|\underline{x})$ that an event characterized by the variables $\underline{x}$ belongs to the signal class S. This probability can be written, using Bayes' theorem, in terms of the probability density functions for the signal and background, $f(\underline{x}|S)$ and $f(\underline{x}|B)$, and the corresponding class prior probabilities, $\text{prior}(S)$ and $\text{prior}(B)$:

$$\Pr(S|\underline{x}) = \frac{f(\underline{x}|S)\,\text{prior}(S)}{f(\underline{x}|S)\,\text{prior}(S) + f(\underline{x}|B)\,\text{prior}(B)}. \tag{7.1}$$

The probability $\Pr(S|\underline{x})$ has the useful property that using it to cut or weight events, minimizes the probability to misclassify events [99], the principal goal when extracting a weak signal from data. Therefore, each single top analysis attempts to approximate the discriminant

$$D(\underline{x}) = \frac{f(\underline{x}|S)}{f(\underline{x}|S) + f(\underline{x}|B)}, \tag{7.2}$$

built using equal number of signal and background events, in which case $\text{prior}(S) = \text{prior}(B) = 0.5$ and the priors cancel out. The single top analyses differ by the choice of variables $\underline{x}$ and by the numerical method by which the discriminant $D(\underline{x})$ is approximated. The BNN analysis uses non-linear functions to approximate $D(\underline{x})$. The BNN

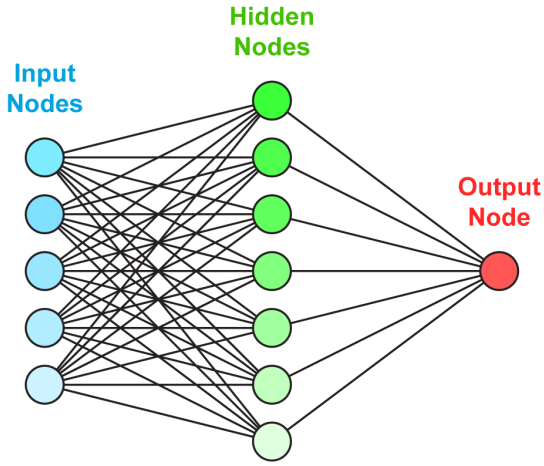

Figure 7.1 : Illustration of the structure of a neural network with three layers: one layer with five input nodes, one layer with seven hidden nodes, and one layer with one output node.

algorithm and the mathematics on how it separates signals from backgrounds are described in Sect. 7.3.

7.3 BNN Structure and Mathematics

7.3.1 BNN Structure

From the structural point of view, a neural network is an interconnected group of nodes (see Fig. 7.1). The simpler neural networks are organized in three subgroups of nodes. One subgroup consists of one layer of N_{in} input nodes that collect all the information to be processed. Another subgroup consists of one layer of N_{out} output nodes that will give the results of the processing. The other subgroup consists of all the intermediate nodes that are between the input and the output nodes. They are called hidden nodes, because the user has no direct interaction with them. Hidden nodes are usually also organized in layers with a varying number of nodes in each layer.

The structure of the neural networks used in each analysis channel is fixed to one layer of $N_{\text{in}} = N_{\text{var}}^{ich}$ input nodes, one layer of $H = 20$ hidden nodes, and one layer of $N_{\text{out}} = 1$ output node. N_{var}^{ich} is the number of selected input variables used for the BNN in channel ich, where $ich \in [1, 2, ..., 24]$. These variables are listed in Appendix E, Tables E.1 to E.24.

7.3.2 BNN Mathematics

From the mathematical point of view, a neural network $n(\underline{x}, \underline{w})$ is a special kind of non-linear function with parameters $\underline{w}$, that is capable of approximating a large class

of functions from $[0,1]^{N_{\rm var}} \to [0,1]$ [100]. The functions that define the neural networks that are used in this analysis are sigmoids:

$$n(\underline{x},\underline{w}) = \frac{1}{1+\exp[-f(\underline{x},\underline{w})]}, \quad \text{where} \tag{7.3}$$

$$f(\underline{x},\underline{w}) = b + \sum_{j=1}^{H} v_j \, \tanh(a_j + \sum_{i=1}^{N_{\rm var}} u_{ji}\, x_i). \tag{7.4}$$

The parameters of the network are $\underline{w} = (u_{ji}, a_j, v_j, b)$; u_{ji} and v_j are generally referred to as *weights*, while a_j and b are called *biases*. Both sets of parameters are collectively referred to as weights. The parameter u_{ji} is the relative weight from input variable i on hidden node j, whereas parameter v_j is the relative weight of hidden node j on the output node. The parameters a_j and b are shifts added by the hidden nodes and the output node respectively to their total input. The numbers $N_{\rm var}$ and H define thus a $1 + (N_{\rm var}+2)H$-dimensional parameter space where each point $\underline{w}$ corresponds to an instance of a neural network function. Typically, one finds a single set of these parameters by minimizing an error function such as

$$E(\underline{w}) = \frac{1}{N}\sum_{i=1}^{N}[t_i - n(\underline{x}_i,\underline{w})]^2, \tag{7.5}$$

using variations of gradient descent [101] with a set $T = \{(t_1,\underline{x}_1),...,(t_N,\underline{x}_N)\}$ of training data. The parameter t is called target and is equal to 1 for signal events and 0 for backgrounds. Given equal number of signal and background events, it can be shown (see for example Refs. [102]) that the minimum of $E(\underline{w})$, as $N \to \infty$, occurs at $n(\underline{x},\underline{w}) = D(\underline{x})$, provided that the number of training events is large enough (and the function $n(\underline{x},\underline{w})$ flexible enough). Because of the nature of the function $n(\underline{x},\underline{w})$, the error function $E(\underline{w})$ induces an exceedingly complex "landscape" in the parameter space of $n(\underline{x},\underline{w})$. Consequently, the best that can be hoped for using any of the standard minimization techniques is to find a local minimum of $E(\underline{w})$. Unfortunately, it is rather easy to find a local minimum that causes the function $n(\underline{x},\underline{w})$ to be fit too well to the training data, that is, to *over-train*; such a network typically performs poorly on an independent set of events. Moreover, only heuristic rules are known about how complex a network needs to be to achieve the best approximation to the discriminant $D(\underline{x})$ given a finite training data-set and a finite number of training iterations. It is these problems and others that the Bayesian approach to neural networks tries to overcome.

A Bayesian neural network is formed by doing an average over the parameter space $\{\underline{w}\}$ of the neural networks $n(\underline{x},\underline{w})$, where each of the networks is weighted by the posterior probability density of the network weights $\underline{w}$ given a training set of data T:

$$\overline{n}(\underline{x}) = \int_{\{\underline{w}\}} n(\underline{x},\underline{w})\, p(\underline{w}|T)\, d\underline{w}. \tag{7.6}$$

The integral (7.6) is in practice approximated by the average of a finite number of networks sampled from the posterior $p(\underline{w}|T)$:

$$\overline{n}(\underline{x}) \approx \frac{1}{K}\sum_{k}^{K} n(\underline{x}, \underline{w}_k). \tag{7.7}$$

This average should approximate the discriminant $D(\underline{x})$, Eq. (7.2), better than a single network obtained by finding a single best point $\underline{w}_0$, and certainly better than any one of the networks $n(\underline{x}, \underline{w}_k)$ in the sum. Moreover, by averaging over many networks, with appropriate Bayesian weighting, the resulting Bayesian network is, for example, less likely to be over-trained and less sensitive to variations in the networks structure. Note also that low correlation among the networks is necessary in order to render this average useful. If the networks were highly correlated, the effective number of terms in the sum of Eq. (7.7) would be far fewer than K.

The Posterior Probability Density of the BNN Weights

Each event in the training data set is characterized by the variables $\underline{x}$ and the target t. By means of Bayes' theorem, the posterior probability density of the network weights $\underline{w}$, $p(\underline{w}|T) \equiv p(\underline{w}|\boldsymbol{t}, \underline{\boldsymbol{x}})$, where the notations $\boldsymbol{t} = \{t_1, t_2, ..., t_N\}$ and $\underline{\boldsymbol{x}} = \{\underline{x}_1, \underline{x}_2, ..., \underline{x}_N\}$ are used, is given by

$$\begin{aligned} p(\underline{w}|\boldsymbol{t}, \underline{\boldsymbol{x}}) &= \frac{p(\boldsymbol{t}|\underline{\boldsymbol{x}}, \underline{w})p(\underline{\boldsymbol{x}}, \underline{w})}{p(\boldsymbol{t}, \underline{\boldsymbol{x}})}, \\ &= \frac{p(\boldsymbol{t}|\underline{\boldsymbol{x}}, \underline{w})p(\underline{w}|\underline{\boldsymbol{x}})p(\underline{\boldsymbol{x}})}{p(\boldsymbol{t}|\underline{\boldsymbol{x}})p(\underline{\boldsymbol{x}})}, \\ &= \frac{p(\boldsymbol{t}|\underline{\boldsymbol{x}}, \underline{w})p(\underline{w})}{p(\boldsymbol{t}|\underline{\boldsymbol{x}})}, \end{aligned} \tag{7.8}$$

where $p(\underline{w}|\underline{\boldsymbol{x}}) = p(\underline{w})$, the prior over weights, is assumed not to depend upon the training data $\underline{\boldsymbol{x}}$. Moreover, the probability that an event characterized by $\underline{x}$ is a signal, with target $t = 1$, given the network parameters $\underline{w}$, is precisely $n(\underline{x}, \underline{w})$, while the probability that it is a background, with $t = 0$, is its complement $(1 - n(\underline{x}, \underline{w}))$. Therefore, the probability density of the targets $\boldsymbol{t}$, given the data $\underline{\boldsymbol{x}}$ and network parameters $\underline{w}$, is given by

$$p(\boldsymbol{t}|\underline{\boldsymbol{x}}, \underline{w}) = \prod_{i=1}^{N} n(\underline{x}_i, \underline{w})^{t_i} \left[1 - n(\underline{x}_i, \underline{w})\right]^{1-t_i}, \tag{7.9}$$

assuming that the events are independent. For weighted training events, each with weight ω_i, one would replace the exponents t_i and $(1 - t_i)$ by $\omega_i\, t_i$ and $\omega_i\,(1 - t_i)$, respectively. However, the BNN software used for this analysis [90] does not provide the option to weight events. To obtain a set T of unweighted events, events are randomly

sampled, in this case from the training samples (see Sect. 7.1), with a probability proportional to the event weight. Note that the denominator in Eq. (7.8) does not depend on the weights $\underline{w}$ and can thus be ignored to the purpose of sampling the posterior $p(\underline{w}|T)$.

The BNN Prior

To complete the Bayesian calculation it is necessary to specify a prior density over the network parameter space $\{\underline{w}\}$. In the Flexible Bayesian Modeling (FBM) package of Radford Neal [90], used in the BNN analysis, a Gaussian prior centered at zero is specified for each weight (smaller weights yield smoother sigmoid functions with a consequent better approximation to the discriminant), and the variance for weights belonging to a given group (either *input-to-hidden weights* (u_{ij}), *hidden-biases* (a_j), *hidden-to-output weights* (v_j) and *output-bias* (b)) is chosen to be the same: σ_u^2, σ_a^2, σ_v^2, or σ_b^2, respectively. Moreover, to allow for a large range of possible variances, while favoring small values, a gamma prior

$$\pi(1/\sigma^2;\mu,\alpha) = \left(\frac{\alpha/2}{\mu}\right)^{\alpha/2} \left(\frac{1}{\sigma^2}\right)^{(\alpha/2)-1} \exp\left(-\frac{\alpha/2}{\mu\sigma^2}\right) /\Gamma(\alpha/2), \tag{7.10}$$

is assigned to each of these four variances, with the mean μ and shape parameter α set to some fixed plausible values specified by the user. Thus,

$$p(w_i) = \text{Gaussian}(w_i|0,\sigma_\beta)\, \pi(1/\sigma_\beta^2;\mu_\beta,\alpha_\beta), \tag{7.11}$$

where β denotes the class (*input-to-hidden* (u), *hidden-biases* (a), *hidden-to-output* (v) or *output-bias* (b)) to which the weight w_i corresponds, and $\pi(1/\sigma_\beta^2;\mu_\beta,\alpha_\beta)$ is given by Eq. (7.10). However, the use of an adaptive prior can cause the network parameter values to become too large if the training data are excessively noisy; meaning, if within the training data there are events with a very large weight compared to the average (or keeping the idea of events with unit weight, if the training data contain events that are repeated many times −because they originally had large weights−). Figure 7.2 shows the evolution of the widths σ_u, σ_a and σ_v, when sampling the probability density $p(\underline{w}|\boldsymbol{t},\underline{\boldsymbol{x}})$, Eq. (7.8), using an adaptive prior (7.10) in (7.11), as a function of the number of sampling iterations. The case shown in the left (right) column uses training data from the Run IIa muon (electron) 1-tag 2-jet channel. While in the case of the muon channel the widths stay stable with progressing number of iterations, in the case of the electron channel the widths σ_u (top right plot) and σ_a (middle right plot) show a stable behavior only up to iteration ~ 250 and then they grow taking larger and larger values. Large values of the (adaptive) widths mean that some of the network parameters prefer large values. A corresponding BNN formed out of the last K iterations (that is how a BNN is typically built), will have a bad performance, i.e. poor signal-to-background separation.

To overcome this problem, the prior widths (σ_β) are kept fixed during the training:

$$p(w_i) = \text{Gaussian}(w_i|0, \sigma_\beta = \text{fixed}). \tag{7.12}$$

This is seen to lead to an improved BNN performance (with respect to what one would get using the gamma priors). The determination of the width values is explained in Sect. 7.6.2.

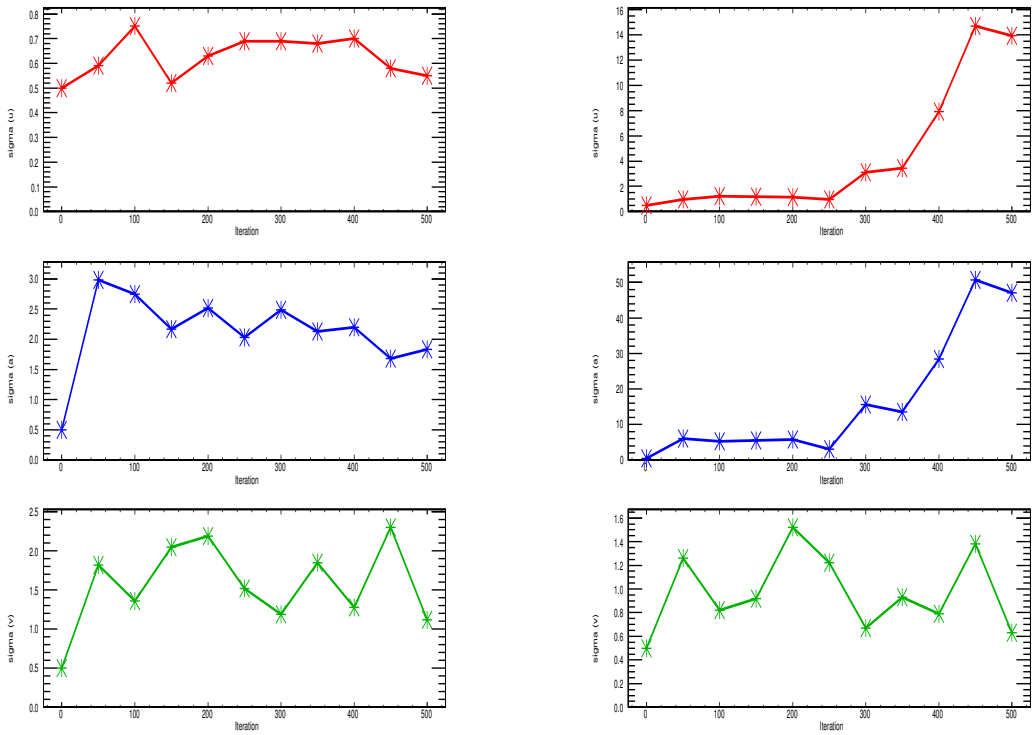

Figure 7.2 : The widths of the Gaussian priors for the parameters u_{ij} (first row), a_j (second row) and v_j (third row) of a BNN as a function of the number of iterations (left column) and Run IIa muon 1 b-tag 2 jets channel (right column). These are plots for the case in which the widths are not set to fixed values as it is ultimately done.

7.4 Analysis Road-Map

With the split signal and background samples as a starting point, the BNN analysis proceeds as follows:

1. Start from a set of approximately 165 variables deemed well modeled based on the Kolmogorov-Smirnov(K-S) test (see Sect. 6.6). For each of the twenty four analysis channels, select a subset of variables for further analysis following the procedure below:

- rank the variables according to their discrimination importance using the RuleFitJF algorithm [103] available in the Toolkit for Multivariate Analysis (TMVA) [104];
- select a subset of variables with importance higher than a predefined importance threshold;
- iterate (approx. 15 times) until the subset of selected variables becomes "stable enough", ending up with typically 25 variables.

2. Having selected in each channel a subset of N_{var}^{ich} variables, perform the training of the networks and build the discriminant $D(\underline{x})$. Check if the performance of the BNN is adequate.
3. Compute the posterior densities of the single top cross section for each of the twenty four analysis channels and its combinations using binned likelihoods formed from the BNN-output distributions of the twenty four analysis channels, and determine the cross sections.
4. Compute a p-value that quantifies the significance of the result using a background-only ensemble.

The rest of this chapter discusses the steps 1 and 2. The extraction of the cross section and its significance, steps 3 and 4, are the subject of Chapter 8.

7.5 Selection of Variables with High Discrimination Power

For many multivariate methods, the computational burden increases rapidly with the dimensionality of the input data. Moreover, keeping the dimensionality low relative to the size of the training data sample reduces the tendency to over-fit a discriminant to the training data. To that end, variables judged to provide insufficient signal-to-background discrimination are eliminated. This is done separately in each of the twenty four channels by means of the RuleFitJF [103] algorithm.

The RuleFitJF algorithm is used to rank the variables according to their "importance". The importance of a variable is a number between 0 and 1. The higher the importance of a given variable, the better its discrimination power. The importance value that a variable gets also depends on what other variables are in the set being ranked. Moreover, although the highest ranked variable always gets an importance equal to 1, the lowest ranked variable does not necessarily get an importance equal to 0. This allows to make a variable selection in which those with importance below a given predefined threshold are eliminated. Thus, if one starts with a set of N_0^{var} variables, after removing the lower ranked ones, one ends up with a reduced set of N_1^{var} ($< N_0^{\text{var}}$) variables. But since the importance is not absolute, the importance of the remaining

variables is in principle subject to change. This means that, in an eventual second iteration, where the RuleFitJF algorithm is applied only to the selected set of N_1^{var} variables, a new ranking is obtained, where each of the N_1^{var} variables gets a new importance value and some of them can thus fall below the chosen threshold. In this way, one could iterate until the number of variables does not get further reduced ($N_{n+1}^{\text{var}} = N_n^{\text{var}}$). This simple procedure would work very straighforward if, as one would in principle expect, the relative ranking of the variables does not change after each iteration, although the importance does. But small changes in the relative ranking and importance value occur upon repetition of RuleFitJF running. The reason for this is that, in each iteration, random numbers are being used to generate a new ensemble of unweighted events from weighted events in the initial signal and background samples. This is needed, because the current RuleFitJF implementation assumes that all events have a uniform (= 1) weight. Different samplings may lead to different rankings, making the result not completely stable. But the change in importance is not drastic for the variables that are top-ranked; this is how one is still able to use this algorithm for filtering out the less-important variables. At this point it is also worth to emphasize that the BNN performance does not depend on the relative ranking of the variables being used in the training, but rather on the set of variables itself and what is their signal-to-background discrimination power. Keeping or dropping a variable might result in a more or less optimal BNN, but it is not right or wrong in any way. The chosen variable selection criteria is thus the one that is simpler to implement. Each analysis channel is considered separately (it has its own variables), but all them are iterated together. It is considered that, if in the majority ($\sim 80\%$) of the channels the corresponding selected subsets of variables are not reduced at all in at least 3 iterations (N_{it}, $N_{it}+1$, $N_{it}+2$) while in the other channels the selected subsets are not reduced by more than 2 variables (out of ~ 25), then any of the 3 iterations, N_{it}, $N_{it}+1$ or $N_{it}+2$, could be considered final and the subsets of variables are deemed "stable enough".

Different importance thresholds are used for different jet bins. For the 2-jet and 4-jet bins the used threshold is 0.1 (corresponding to 10% of the maximum possible importance). For the 3-jet bins the threshold is increased to 0.2 to reduce the number of selected variables to less than 30 (as discussed above, it is desirable to keep the number of variables relatively low). The resulting set of selected (and well modeled) variables, ranked according to their importance, are shown in Tables E.1 to E.24 in Appendix E for each of the twenty four analysis channels.

7.6 BNN Training and Verification

7.6.1 Training

In each channel, a training set T is built consisting of an admixture of 10,000 signal and 10,000 background unweighted events, which are randomly sampled from the signal

and background weighted events in the training samples, and with each event having a probability of being picked by the sampling proportional to its weight. (The number 10,000 has nothing to do with the available statistics in the training samples. The number of total signal and background events in these samples are shown in Table 7.1.) All sources of signal and background considered in the analysis are used in the training. From the training set T, a posterior density $p(\underline{w}|T)$ over the network parameter space is constructed.

A Markov Chain Monte Carlo (MCMC) technique [105,106] is used to draw a sample of networks out of the posterior density $p(\underline{w}|T)$. A complete cycle of running through the entire training sample is called an "epoch". Each iteration of the MCMC consists of 20 epochs and the result of each iteration is the average of the outputs of the 20 training epochs. The MCMC is run with 300 iterations and a sample of 100 networks pertaining to the last 100 iterations is used to approximate the discriminant $D(\underline{x})$, (7.2), via a Bayesian neural network, $\overline{n}(\underline{x})$, defined by

$$\overline{n}(\underline{x}) \approx \frac{1}{100} \sum_{k=201}^{K=300} n(\underline{x}, \underline{w}_k). \tag{7.13}$$

Variants in the total number of MCMC iterations and in the number of networks used to build the BNN, have been tried. Increasing the number of iterations up to 500 did not show a significant improvement in the BNN performance ($\lesssim 5\%$), not even taking more than 100 networks to construct the BNN, but increased the training time by more than a factor of two[35]. On the other hand, lowering the number of iterations to 200-250, while still building the BNN out of the last 100 ones, led to a worse BNN performance (~ -5-10%). Thus, running 300 iterations is a good compromise between BNN performance and training time. A measure of the BNN performance is given by the area below the signal vs. background efficiency curve, which will be discussed in Sect. 7.6.3.

7.6.2 Treatment of Noise in the Training Data

The Bayesian neural network is sensible to the presence of "noise" in the training data, which degrades its performance. By noise it is meant MC events with large weights. When training on such samples, the prior densities may become excessively broad (see Fig. 7.2) and the MCMC sampling yields networks whose parameters have values that scatter over a large range. This leads to discriminant distributions that are excessively jagged and therefore bad approximations to the discriminant $D(\underline{x})$, Eq. (7.2). To overcome this problem, the network parameter prior widths are held fixed instead of letting them adapt to the training data. This limits the range of the network parameter values thereby favoring smoother approximations to the discriminant $D(\underline{x})$. A single

[35]The time required to train a BNN with 300 iterations in the DØ cluster is about 12-16 hours.

neural network is trained for each of the twenty four analysis channels separately, using the back-propagation algorithm as implemented in the JETNET program [107], which aims to minimize the error (7.5). The structure of the JETNET-trained networks is identical to that used for the BNNs, as are the training data. For each of the twenty four JETNET-trained networks (one per channel), the root-mean-square of the network parameters in each of the three groups (input-to-hidden weights, hidden-to-output weights, and hidden biases) is calculated. These root-mean-squares serve then as the standard deviations of the zero mean Gaussian priors in the BNN. The three standard deviations used in each of the twenty four channels are shown in Table 7.2. The width σ_b is set to a large value (100) so that the output bias is only minimally constrained. With this choice of prior densities, the BNN parameters are no longer driven to large values, resulting in smooth discriminants $D(\underline{x})$.

Values of the Gaussian Prior Widths used in the BNN

	Electron channel						Muon channel					
	1 b-tag			2 b-tags			1 b-tag			2 b-tags		
	2jet	3jet	4jet	2jet	3jet	4jet	2jet	3jet	4jet	2jet	3jet	4jet
	Run IIa											
σ_u	0.17	0.17	0.16	0.16	0.17	0.15	0.18	0.16	0.18	0.18	0.14	0.18
σ_a	0.34	0.26	0.19	0.27	0.23	0.23	0.36	0.27	0.27	0.31	0.17	0.27
σ_v	0.56	0.58	0.56	0.54	0.64	0.63	0.58	0.56	0.64	0.66	0.50	0.67
	Run IIb											
σ_u	0.19	0.15	0.15	0.16	0.13	0.15	0.18	0.19	0.17	0.17	0.16	0.17
σ_a	0.32	0.19	0.15	0.27	0.19	0.19	0.31	0.34	0.16	0.29	0.26	0.25
σ_v	0.56	0.50	0.54	0.53	0.49	0.59	0.61	0.57	0.66	0.61	0.58	0.64

Table 7.2 : Standard deviations for the gaussian priors for the input-to-hidden, hidden biases and hidden-to-output weights, obtained from a single neural network trained separately for each of the 24 channels, using JETNET.

7.6.3 Verification

As mentioned in Sect. 7.6.1, the networks used to construct the Bayesian neural network $\overline{n}(\underline{x})$ are sampled from the posterior density $p(\underline{w}|T)$ using a Markov chain. To check the convergence of the Markov chain, the diagnostic algorithm described below is followed.

Two independent ensembles of signal events, $E_{\rm sig}$ and $E'_{\rm sig}$, are buil containing each half of the statistics in the signal yield samples. A third ensemble, $E_{\rm bkg}$, is defined as containing all the statistics in the background yield samples. The three ensembles are normalized to the same number, here denoted by N. Then, the signal events in $E_{\rm sig}$ and the background events in $E_{\rm bkg}$ are weighted by their corresponding BNN output values.

Mathematically, this is equivalent to multiply the densities $f_{E_{\rm sig}}(\underline{x}|S)$ and $f_{E_{\rm bkg}}(\underline{x}|B)$ by the BNN function $\overline{n}(\underline{x})$:

$$f_n(\underline{x}|S) = f_{E_{\rm sig}}(\underline{x}|S)\,\overline{n}(\underline{x}), \tag{7.14}$$
$$f_n(\underline{x}|B) = f_{E_{\rm bkg}}(\underline{x}|B)\,\overline{n}(\underline{x}). \tag{7.15}$$

Given that the number of events in the ensembles is large, the densities $f_{E_{\rm sig}}(\underline{x}|S)$ and $f_{E_{\rm bkg}}(\underline{x}|B)$ can be replaced by $Nf(\underline{x}|S)$ and $Nf(\underline{x}|B)$, since their shape is given by $f(\underline{x}|S)$ and $f(\underline{x}|B)$ and their normalization by N. Thus, the above equations can be rewritten as

$$f_n(\underline{x}|S) = N\,f(\underline{x}|S)\,\overline{n}(\underline{x}), \tag{7.16}$$
$$f_n(\underline{x}|B) = N\,f(\underline{x}|B)\,\overline{n}(\underline{x}). \tag{7.17}$$

Let define $g(\underline{x})$ as the sum of these two weigthed densities:

$$g(\underline{x}) = f_n(\underline{x}|S) + f_n(\underline{x}|B) = N\,[f(\underline{x}|S) + f(\underline{x}|B)]\,\overline{n}(\underline{x}). \tag{7.18}$$

If, as the Markov chain proceeds, $\overline{n}(\underline{x}) \to D(\underline{x})$ given by (7.2), which is the desired outcome, then one should find that $g(\underline{x}) \to Nf(\underline{x}|S)$, that is, one should recover the signal density. It is thus possible to verify the convergence of the Markov chain by comparing $g(\underline{x})$ extracted from $E_{\rm sig} + E_{\rm bkg}$, with the distributions obtained from the independent signal-only sample $E'_{\rm sig}$. In particular, one should recover all possible marginal densities, that is, projections to 1-dimension:

$$\begin{aligned}\int_{x\notin\underline{z}} g(\underline{z})\,d\underline{z} &= \int_{x\notin\underline{z}} f_n(\underline{z}|S)\,d\underline{z} + \int_{x\notin\underline{z}} f_n(\underline{z}|B)\,d\underline{z},\\ &= m_n(x|S) + m_n(x|B) \to\ N\,m(x|S).\end{aligned} \tag{7.19}$$

In Fig. 7.3, verification plots for three of the variables used in the Run IIa e+jets 1-tag 2-jet channel are shown. Similar sets of plots are shown in Appendix F for each of the twenty four analysis channels. For plotting simplicity, the values of each of these variables are shifted and scaled in an event-by-event basis such that the resulting distributions have zero mean and unit variance:

$$x_j \to x'_j = \frac{x_j - x_{\rm mean}}{\sigma_x}, \tag{7.20}$$

where x_j is the value that the variable x takes for event j, and $x_{\rm mean}$ and σ_x are the mean and standard deviation of variable x in the corresponding ensemble. The black dots represent the signal distribution $N\,m(x|S)$ for variable x, extracted from ensemble $E'_{\rm sig}$. The blue and green histograms correspond to the distributions $m_n(x|S)$ and $m_n(x|B)$, extracted from the BNN weighted ensembles $E_{\rm sig}$ and $E_{\rm bkg}$ respectively. The red histogram (sum of the blue and green ones) is the sum $g(\underline{x})$. Very good agreement is observed between the signal distribution in the $E'_{\rm sig}$ sample (black dots)

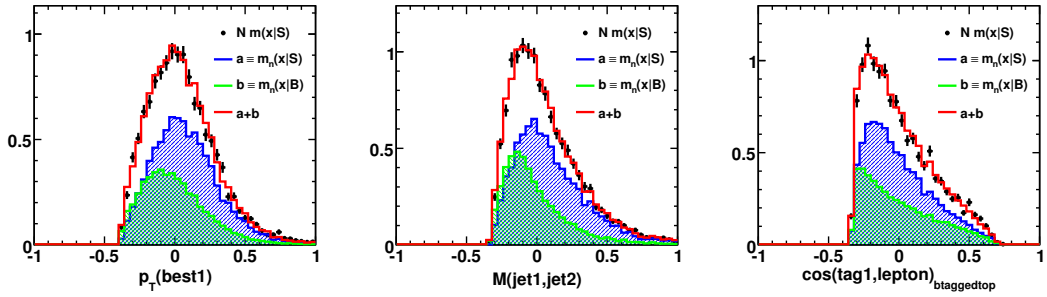

Figure 7.3 : Verification plots for three of the variables used by the BNN in the Run IIa electron channel with two jets and one b-tag. See main text for figure explanation.

and the prediction from the BNN weighted $E_{\rm sig}+E_{\rm bkg}$ samples (histogram in red), both for the variables in Fig. 7.3 and those shown in Appendix F, thus indicating that the Markov chains have converged. The convergence is checked for all the input variables used in each channel, and same level of agreement is observed.

An equivalent verification test may be performed directly on the BNN output. In fact, integrating Eq. (7.18) over all variables $\underline{z}$ with the constrain $\overline{n}(\underline{z}) = n_0 = \text{constant}$, one gets that

$$\begin{aligned}\int_{\overline{n}(\underline{z})=n_0} g(\underline{z})\, d\underline{z} &= \int_{\overline{n}(\underline{z})=n_0} f_n(\underline{z}|S)\, d\underline{z} + \int_{\overline{n}(\underline{z})=n_0} f_n(\underline{z}|B)\, d\underline{z}, \\ &= N\, S(n_0)\, n_0 + N\, B(n_0)\, n_0 \rightarrow\ N\, S(n_0), \end{aligned} \tag{7.21}$$

where $S(n_0)$ and $B(n_0)$ are the BNN-output distributions for signal and background events respectively. Equation (7.21) may be conveniently rewritten as

$$\frac{S(n_0)}{B(n_0)} = \frac{n_0}{1-n_0}, \tag{7.22}$$

which is nothing else than the statistical interpretation of the BNN output. If for an individual event n_0 represents its probability to belong to the signal class S, then, when the BNN is applied to a large sample with equal number of signal and background events, from all the events resulting in a BNN output equal to n_0, a fraction n_0 of them should belong to the signal class S, and the rest, a fraction $(1-n_0)$, to the background class B. Figure 7.4 (left) shows the BNN-output distributions, normalized to unity, for the signal (blue curve) and background (green curve) yield samples in the Run IIa e+jets 1-tag 2-jet channel. The BNN output is a number between 0 and 1, and as expected, the signal is enhanced for high values of the BNN output, while the background is enhanced at low values. It can also be observed that in each bin of the histogram, the ratio of signal to background content respects the relation (7.22) with good approximation.

Let assume a lower cut is placed on the BNN output with the aim of rejecting most of the background and keep most of the signal, and denote with S and B the total amount of signal and background left after the cut. Using these definitions, Fig. 7.4 (center) shows two measures of the BNN performance, namely $S/\sqrt{B}$ (blue curve) and $S/\sqrt{S+B}$ (red curve), as a function of the cut on the BNN output. These quantities aim to represent how significant is the amount of signal left after the cut, compared to the statistical error in the data. For $S/\sqrt{B}$ (blue curve) in Fig. 7.4, its maximum occurs for a cut BNN-ouput > 0.58 (vertical line) at which point $S/\sqrt{B} = 1.7$ (horizontal line). It is important to recall that this is only a performance test and that no cut on the BNN output is ultimately done; the cross section is measured taking into account the whole range of the BNN discriminant.

Finally, Fig. 7.4 (right) shows the signal and background efficiencies as a function of the BNN output cut. The efficiencies are defined as $\epsilon_S = S/S_0$ and $\epsilon_B = B/B_0$, where S_0 and B_0 correspond to S and B for a BNN cut at 0, which means 'no cut'. A cut at 0 corresponds to the point $\epsilon_S = \epsilon_B = 1$, while a cut at 1 to the point $\epsilon_S = \epsilon_B = 0$. In this plot, the area above the curve is shown as another way of quantifying the BNN performance. The lower this number, the better is the signal-to-background separation.

Similar sets of three plots as in Fig. 7.4 are shown in Appendix F for each of the twenty four analysis channels. In general, all of them indicate that the BNNs are performing as expected.

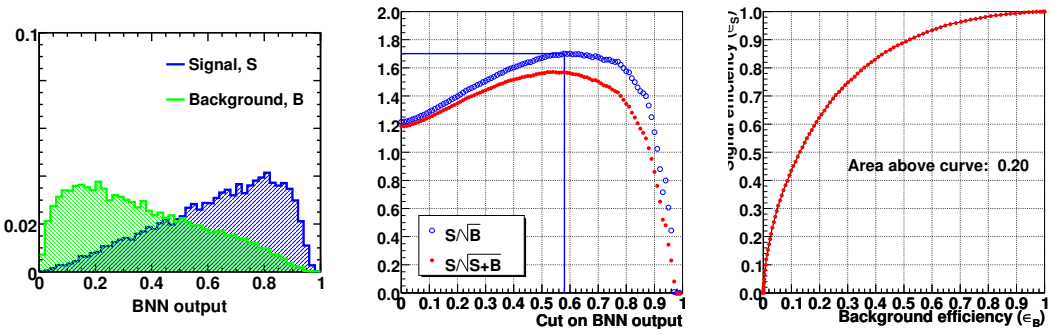

Figure 7.4 : Left: BNN-output distributions normalized to unity for the background (green) and signal (blue) yield samples. Center: two measures of the BNN performance, $S/\sqrt{B}$ and $S/\sqrt{S+B}$, as a function of a lower cut on the BNN output. Right: signal vs. background efficiency when moving the BNN output cut from 0 to 1. These plots are for the Run IIa electron channel with two jets and one b-tag. See main text for a more detailed explanation of the graphs.

7.7 BNN Output Transformation and Histogram Binning

After applying the BNN filter (7.13) to the signal and background events of the yield sample, BNN discriminant distributions with a uniform binning are obtained. The BNN output is a number between 0 and 1, with background-like (signal-like) events populating mostly the low (high) discriminant region. This introduces a problem when trying to calculate the cross section, related with the fact that in the high discriminant region there may be some bins in which there are some signal but no background events. To avoid this, a monotone binning transformation is applied to the BNN outputs before feeding them to the cross section extraction code, that ensures a minimum amount of effective background events in each bin. This transformation also gives to the background a $1/x$-like shape. All the discriminant plots shown in this thesis are made after applying this transformation (except for the BNN output plots shown in Sect. 7.6.3 and Appendix F, which are only for the purpose of checking the convergence of the Markov chains).

7.7.1 Monotone Transformation

The tranformation function is derived based only on the original BNN-output distributions for the backgrounds. The background sum is first very finely resolved in 10,000 uniform bins and normalized to unity. These 10,000 (input) bins are then rearranged (piled up together) from right (BNN output = 1) to left such that the transformed BNN output for the background sum follows a $1/x$-curve between 0.05 and 0.8 (normalized to 1 when integrated between 0.05 and 0.9) and a linear slope from the intersect of the $1/x$ graph at 0.8 down to zero at 0.95. There are no shape constraints between 0.95 and 1.0; the only condition that is also applied there, is that there has to be at least 40 background events in each of the output bins of width 0.02 in order to keep the (MC) statistical uncertainty reasonably small. This rebinning is done individually for each of the twenty four analysis channels. Figure 7.5 shows the transformation function (upper plot), and the signal (blue) and background (red) BNN-output distributions before and after the binning transformation (lower left and right plots respectively) for the Run IIa e+jets 1-tag 2-jets channel. This binning transformation increases the resolution in the signal region without introducing too much statistical fluctuations.

7.7.2 Binning Choice

In addition to the monotone transformation described in Sect. 7.7.1, a simple bin size optimization is made by varying the number of output bins in the monotone transformation. The considered number of bins are 20, 25, 50 and 100. For the 25, 50 and 100 bins, also a "variable binning" is considered, meaning that in the low statistics

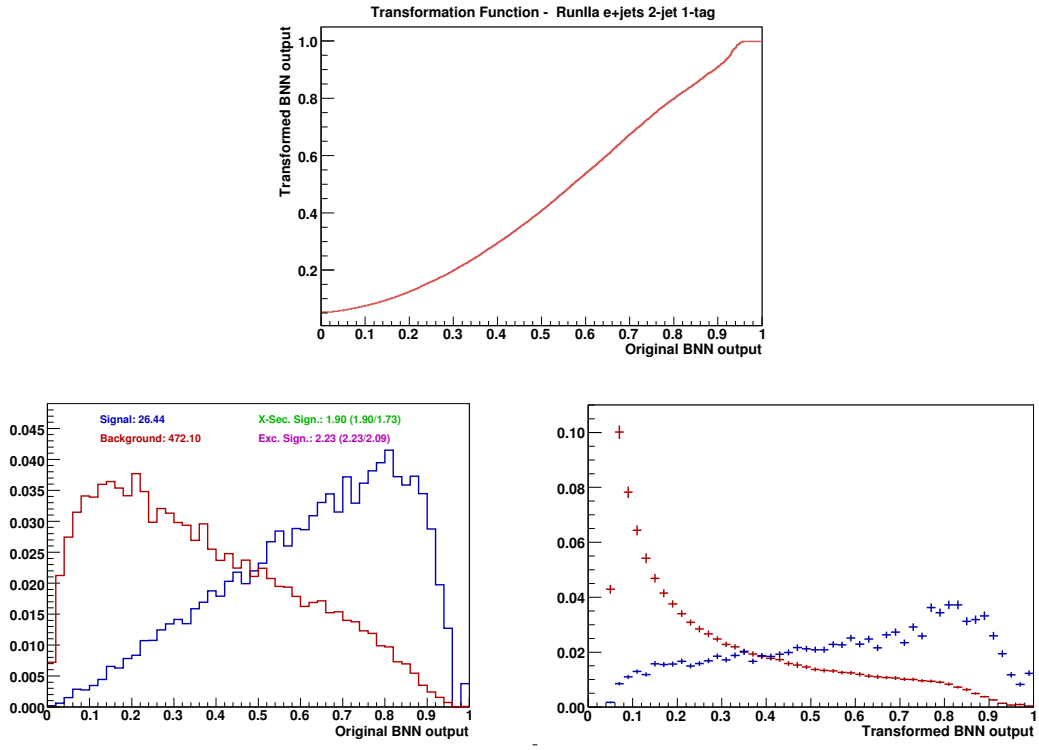

Figure 7.5 : Top: The monotone transformation function. Bottom left: The original BNN discriminant for tb+tqb signal (blue) and all backgrounds combined (red). Bottom right: Same distributions as in the left plot, but after the monotone transformation. These plots correspond to the Run IIa e+jets 2-jet 1-tag channel. Both the original and the transformed BNN-output histograms are shown using 50 bins.

channels fewer bins are used (a factor 2 less in the 4-jet 1-tag and 3-jet 2-tag channels and a factor 5 less in the 4-jet 2-tag channel). For each case, the expected peak over half-width is calculated (see Sect. 8.3 for the definition of this quantity) for the all channels combined case. It is seen that using 50 bins in all channels (i.e. not variable binning) gives the best performance (highest peak over half-width) of all binning choices considered. Therefore, this is the number of bins used in the BNN-output histograms when deriving from them all the measurements.

To further support this binning choice, the expected cross sections and peak over half-widths are calculated for numbers of bins from 5 to 200. For this test, the constraints on the trigger and $t\bar{t}$ cross section uncertainties to be asymmetric are removed. This is to avoid shifts in the expected cross section with respect to to the SM value coming from the treatment of the systematic errors. Thus, the trigger error becomes $\pm 5\%$, and the $t\bar{t}$ error $\pm 12.7\%$. The results are shown in Fig. 7.6. For a number of bins higher than 50 the expected cross section is outside the prefered range of ± 0.1 pb from the SM cross section value (3.46 pb). On the other hand, the peak over half-width increases

with the number of bins. Hence 50 bins is a reasonable choice.

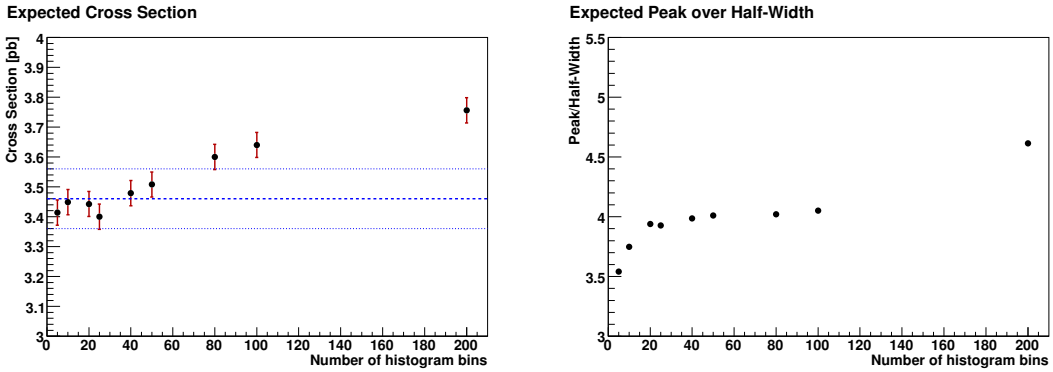

Figure 7.6 : Left: Expected Cross Section versus number of histogram bins used. For histograms with 5-50 bins, a similar expected cross sections is seen. For histograms with more than 50 bins, the cross section gets further away from the standard model value. Right: Peak over Half-Width versus number of histogram bins used. We see an improvement in the binning choices up to 50 bins.

7.8 BNN Outputs

The BNN-output distributions from the yield samples are shown in Figs. G.1 to G.8 in Appendix G for the twenty four analysis channels. These are exactly the histograms that are used in the analysis by the limit setting code, except for the overall signal normalization, which is the quantity to be measured. In the present section, the BNN-output distributions for the combination of the electron and muon channels in Run IIa+b data are presented, using a linear (Fig. 7.7) and a logarithmic scale (Fig. 7.8). Both here and in Appendix G the total single top contribution in the graphs is scaled to the measured cross section, keeping the relative *tb* to *tqb* contributions according to the ratio of their SM cross sections. Overall, a good agreement is observed, within uncertainties, between the data and the background-plus-signal model.

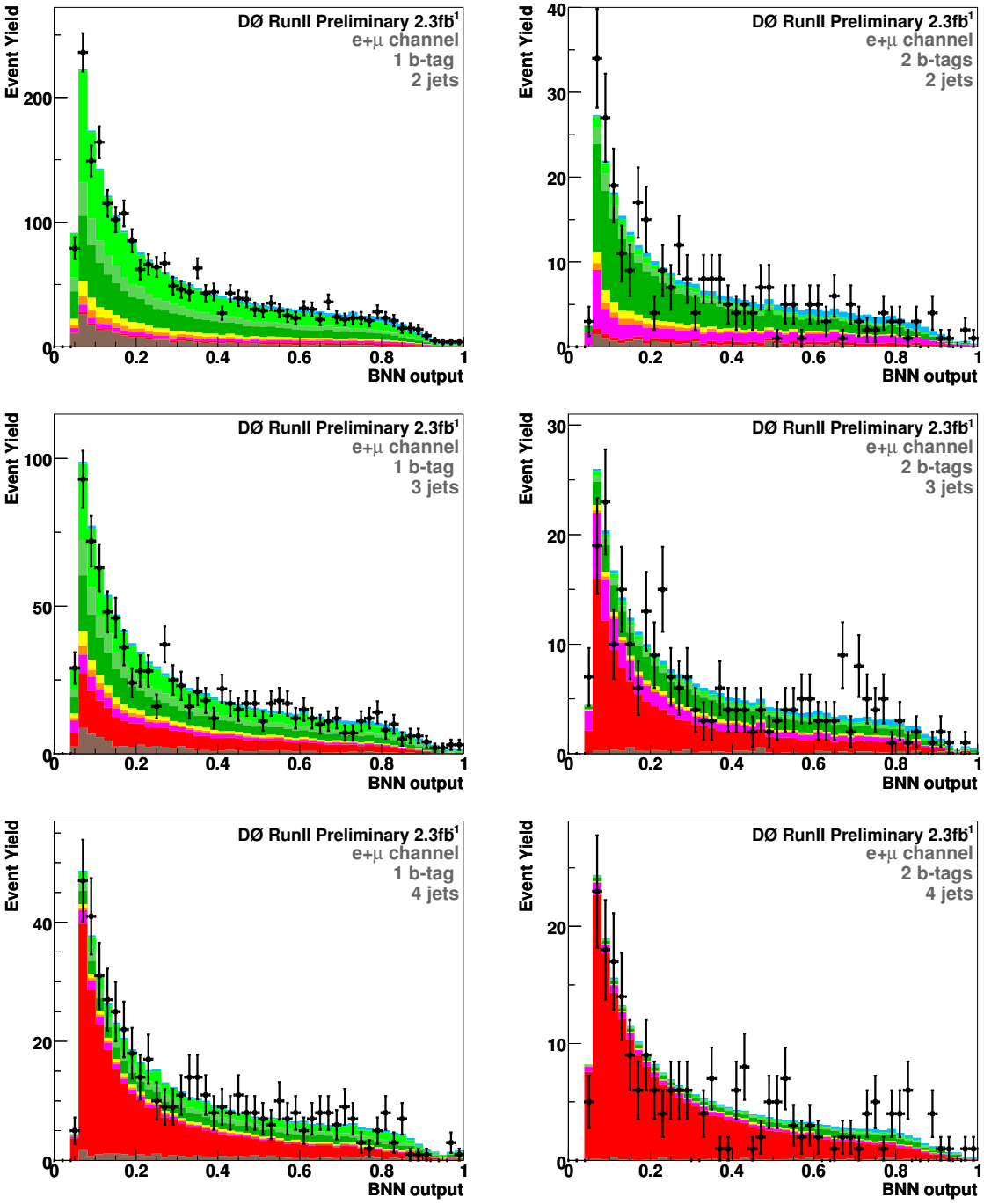

Figure 7.7 : BNN-output distributions for the e,μ+jets combination in Run IIa+b data shown in linear scale. The total signal contribution is normalized to the measured cross section, while the relative tb to tqb contributions are given by the ratio of their SM cross sections. Different rows correspond to different jet multiplicities: 2 jets (up), 3 jets (middle), 4 jets (right). The left column corresponds to 1 b-tag and the right column to 2 b-tags.

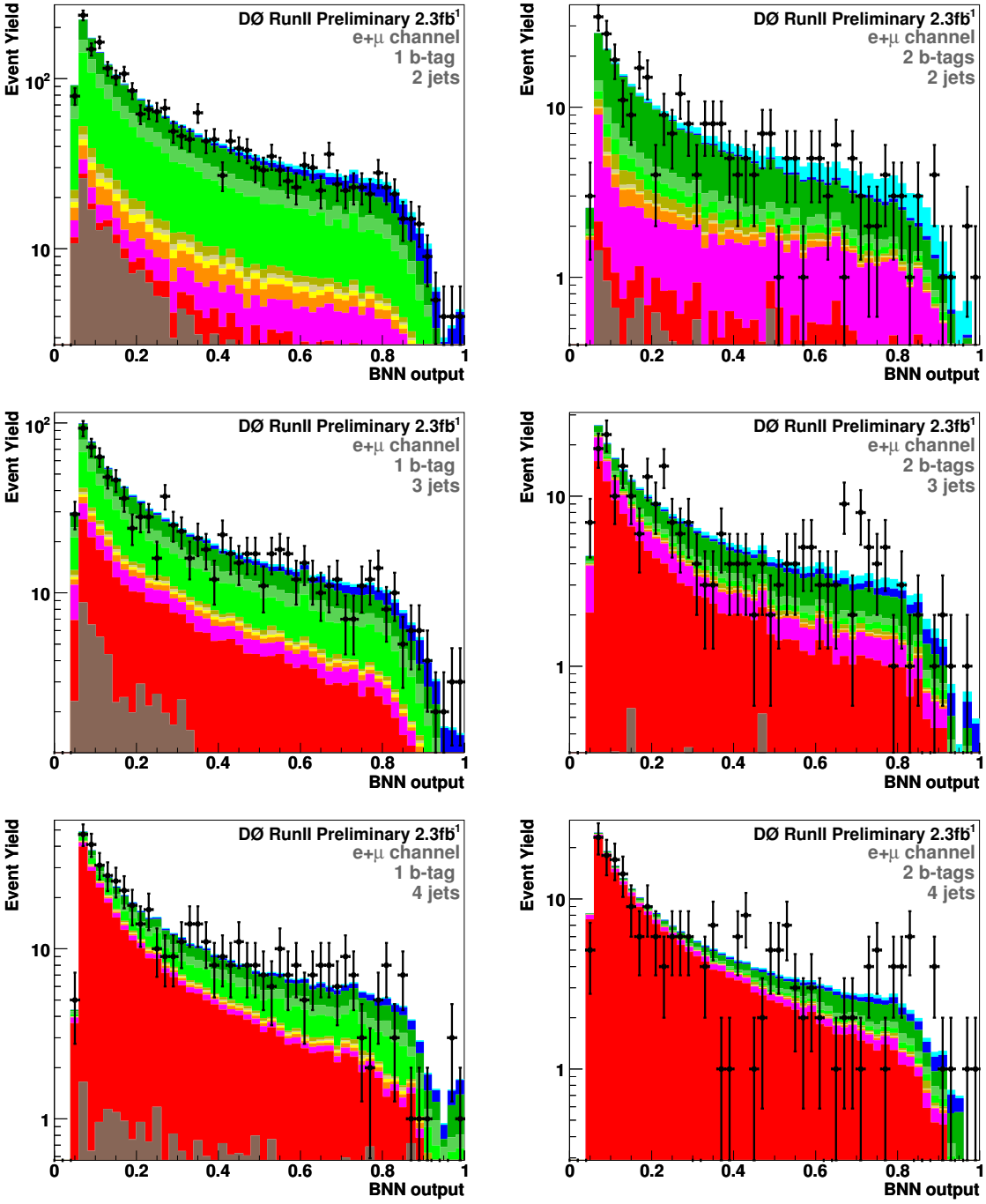

Figure 7.8 : BNN-output distributions for the e,μ+jets combination in Run IIa+b data shown in logarithmic scale. The total signal contribution is normalized to the measured cross section, while the relative tb to tqb contributions are given by the ratio of their SM cross sections. Different rows correspond to different jet multiplicities: 2 jets (up), 3 jets (middle), 4 jets (right). The left column corresponds to 1 b-tag and the right column to 2 b-tags.

8 Measurements

This chapter presents the measurements of the single top quark production cross section and the CKM matrix element $|V_{tb}|$. First, the bayesian method used for the measurements is discussed in detail in Sect. 8.1. The results and validation/calibration tests are presented in the subsequent sections of the chapter.

8.1 Methodology

This section describes the method used to extract the single top cross section. It is however straightforward to extend it to measure $|V_{tb}|$ as the cross section is proportional to $|V_{tb}|^2$. The method applies to any model with one signal process and several backgrounds. The goal is to define a Bayesian posterior probability for different values of the signal cross section. The methodology description presented here is based on Ref. [108].

8.1.1 Bayesian Approach

The probability to observe a count D, if the mean count is d, is given by the Poisson distribution:

$$p(D|d) = \frac{\exp(-d)\, d^D}{\Gamma(D+1)}, \tag{8.1}$$

where Γ is the Gamma function. The mean count d is a sum of the predicted contributions from the signal and the N_{bkg} sources of background:

$$d = s + \sum_{i=1}^{N_{\text{bkg}}} b_i = (\alpha L)\,\sigma + \sum_{i=1}^{N_{\text{bkg}}} b_i \equiv a\,\sigma + \sum_{i=1}^{N_{\text{bkg}}} b_i, \tag{8.2}$$

where s and $\underline{b} = (b_1, b_2, ..., b_{N_{\text{bkg}}})$ are the mean yields for the signal and the background sources, α is the signal acceptance, L the integrated luminosity, σ the signal cross section, and $a \equiv \alpha L$ is the effective luminosity for the signal. Both D and d can be either the total yield (observed and predicted, respectively), or the counts in bins of a kinematic distribution that provides separation between signal and backgrounds.

Using Bayes' theorem one obtains that:

$$p(\sigma, a, \underline{b}|D) = \frac{\mathcal{L}(D|d)\pi(\sigma, a, \underline{b})}{\mathcal{N}}, \tag{8.3}$$

where the likelihood function $\mathcal{L}(D|d)$ is given by the probability $p(D|d)$ and $\mathcal{N}$ is an overall normalization constant.

In case of combining two or more independent channels, the likelihood function is taken as the product of all single-channel likelihoods. If in addition one has a distribution of observed and predicted mean counts for each channel, then each single-channel likelihood is factorized as well in single-bin likelihoods, which holds since the probability to observe a count in a given bin is independent of the counts observed in the other bins:

$$\mathcal{L}(\mathbf{D}|\sigma, \mathbf{a}, \underline{\mathbf{b}}) = \prod_{ich=1}^{N_{\text{chns}}} \prod_{ibin=1}^{N_{\text{bins}}} \mathcal{L}(D_{ich,ibin}|\sigma, a_{ich,ibin}, \underline{b}_{ich,ibin}), \tag{8.4}$$

where $\mathbf{D}$, $\mathbf{a}$ and $\underline{\mathbf{b}}$, represent arrays of $N_{\text{chns}} \times N_{\text{bins}}$ components with the observed counts, effective signal luminosities and background yields, respectively. In this analysis, the distribution corresponds to the BNN discriminant.

Integrating over the parameters $\mathbf{a}$ and $\underline{\mathbf{b}}$ one obtains a general expression for the posterior probability density for the signal cross section, given the observed count $\mathbf{D}$:

$$p(\sigma|\mathbf{D}) = \frac{1}{\mathcal{N}} \iint \mathcal{L}(\mathbf{D}|\sigma, \mathbf{a}, \underline{\mathbf{b}}) \pi(\sigma, \mathbf{a}, \underline{\mathbf{b}})\, d\mathbf{a} d\underline{\mathbf{b}}. \tag{8.5}$$

The normalization constant $\mathcal{N}$ is obtained from the requirement $\int_0^{\sigma_{\max}} p(\sigma|\mathbf{D})\, d\sigma = 1$, where the integration goes up to an upper bound (in this analysis $\sigma_{\max} = 12$) where the posterior is sufficiently close to zero. $\pi(\sigma, \mathbf{a}, \underline{\mathbf{b}})$ is the prior probability that encodes all the knowledge on the parameters σ, $\mathbf{a}$ and $\underline{\mathbf{b}}$. Assuming that the effective signal luminosities $\mathbf{a}$ and the predicted background yields $\underline{\mathbf{b}}$ are independent of any prior knowledge of σ (which is true in this analysis), one can factorize the prior as:

$$\pi(\sigma, \mathbf{a}, \underline{\mathbf{b}}) = \pi(\mathbf{a}, \underline{\mathbf{b}}|\sigma)\, \pi(\sigma) = \pi(\mathbf{a}, \underline{\mathbf{b}})\, \pi(\sigma). \tag{8.6}$$

The prior $\pi(\mathbf{a}, \underline{\mathbf{b}})$ summarizes all the systematic uncertainties on the knowledge of the parameters $\mathbf{a}$ and $\underline{\mathbf{b}}$, while $\pi(\sigma)$ represents all the prior knowledge on the signal cross section. It is a convention in the field to express maximal ignorance on the signal cross section by choosing a prior $\pi(\sigma)$ that is flat in σ:

$$\begin{aligned} \pi(\sigma) &= \frac{1}{\sigma_{\max}}, \qquad && 0 < \sigma < \sigma_{\max} \\ &= 0, && \text{otherwise.} \end{aligned} \tag{8.7}$$

The posterior probability density for the signal cross section is, therefore:

$$p(\sigma|\mathbf{D}) = \frac{1}{\mathcal{N}\, \sigma_{\max}} \iint \mathcal{L}(\mathbf{D}|\sigma, \mathbf{a}, \underline{\mathbf{b}}) \pi(\mathbf{a}, \underline{\mathbf{b}})\, d\mathbf{a} d\underline{\mathbf{b}}. \tag{8.8}$$

The mode (peak) of the posterior density distribution is regarded as the cross section measurement, and the 68.27% asymmetric interval about the mode as an estimate of its uncertainty, as shown in Fig. 8.1. The 68.27% interval is constructed starting with a horizontal line at the posterior peak and moving it down until the area contained

between the points $(\sigma_{\text{peak}} - \Delta\sigma^{(-)}, \sigma_{\text{peak}} + \Delta\sigma^{(+)})$ at which the line intersects with the posterior is 68.27%. In the case the posterior is too broad or the peak is too close to zero, the lower bound may reach and stay at zero and the upper bound is further moved until 68.27% of the area is reached.

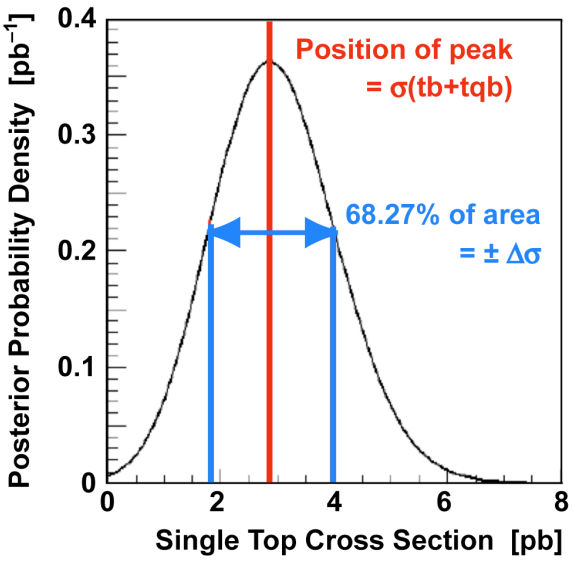

Figure 8.1 : Illustration of a posterior density distribution $p(\sigma|\mathbf{D})$. The position of the peak is used as the measurement result and the uncertainty is given by the 68.27% interval around the peak, defined in this case by the condition $p(\sigma_{\text{peak}} - \Delta\sigma^{-}|\mathbf{D}) = p(\sigma_{\text{peak}} + \Delta\sigma^{+}|\mathbf{D})$.

8.1.2 Numerical Calculation – Prior Modeling

In practice, the integral in Eq. (8.8) is done numerically using Monte Carlo importance sampling. A large number (in this analysis $N_{\text{sample}} = 20{,}000$) of points $(\mathbf{a}_n, \underline{\mathbf{b}}_n)$ is generated randomly by sampling the prior density $\pi(\mathbf{a}, \underline{\mathbf{b}})$. The posterior density is then estimated as:

$$p(\sigma|\mathbf{D}) \approx \frac{1}{\mathcal{N} N_{\text{sample}}\, \sigma_{\text{max}}} \sum_{n=1}^{N_{\text{sample}}} \mathcal{L}(\mathbf{D}|\sigma, \mathbf{a}_n, \underline{\mathbf{b}}_n). \tag{8.9}$$

Modeling the Effects of Systematics that Affect Only the Normalizations

The systematic uncertainties enter in the modeling of the prior density $\pi(\mathbf{a}, \underline{\mathbf{b}})$. Uncertainties that affect the overall normalization only, are modeled by sampling the effective signal luminosities and the background yields from a multivariate Gaussian

$$\pi(\mathbf{a}, \underline{\mathbf{b}}) = \text{Gaussian}(\underline{\mathbf{c}}; \underline{\mathbf{C}}, \Sigma_{\underline{\mathbf{C}}}), \quad \text{where} \quad \underline{\mathbf{c}} = (\mathbf{a}, \underline{\mathbf{b}}), \tag{8.10}$$

whose mean $\underline{\mathbf{C}}$ is defined by the estimates of the parameters $\underline{\mathbf{c}}$. Note that upper case is now used to denote the predicted effective luminosities and the background yields, and lower case for these parameters thought as variables. Equation (8.10) only models overall normalization effects; the effective luminositites and background yields in this equation (and in general in this subsection) are therefore summed over all bins of the discriminant. The standard deviation $\Sigma_{\underline{\mathbf{C}}}$ is derived from a covariance matrix of the uncertainties associated with the measured parameters, which takes into account all correlations of the systematic uncertainties across different sources of signal and backgrounds and analysis channels. Each element of the covariance matrix is calculated as

$$\mathrm{Cov}_{ij} = C_i C_j \sum_{k=1}^{N_{\mathrm{sys}}} f_{ik} f_{jk}, \quad i,j = 1, ..., N_{\mathrm{chns}} \times (N_{\mathrm{bkg}} + 1) \tag{8.11}$$

where $C_{i(j)}$ are the components of $\underline{\mathbf{C}}$, N_{sys} is the number of normalization-only systematic uncertainties, and $f_{i(j)k}$ is the fractional uncertainty from the k^{th} systematic component for the $i^{th}(j^{th})$ parameter. These fractional uncertainties are the ones enumerated in Sect. 6.8. The yields generated from the sampling of the normalization-only systematic effects are denoted with a ($'$), and the corresponding generated shifts with respect to the estimated values as $\underline{\mathbf{\Delta}}^{\mathrm{norm}}$. Thus,

$$(c_{ipar,ich})' = C_{ipar,ich} + \Delta^{\mathrm{norm}}_{ipar,ich}, \tag{8.12}$$

where the index $ipar$ runs over a the components of $\underline{b}$ ($ipar = 1, ..., N_{\mathrm{bkg}} + 1$).

Modeling the Effects of Shape-Changing Systematics

Shape-changing systematics are those that affect the shape of the discriminant, regardless of the effect they have on the overall normalization. In this analysis, they include three sources: jet energy scale, b-tag probabilities and ALPGEN reweightings. The shape (plus normalization) effects are modeled by changing, one at a time, the corresponding systematic sources by plus or minus one standard deviation with respect to their nominal values. As a result, three model distributions are created for each of these systematics: the nominal one, and those resulting from the plus and minus shifts. Denote the plus and minus shifts of a parameter $ipar$ for a systematic effect $isys$ in a given channel and bin $(ich, ibin)$ by $\delta^{+/-}_{isys,ipar,ich,ibin}$. Gaussian fluctuations, with standard deviation defined by the plus and minus shifts, are generated for each bin of a parameter:

$$\Delta^{\mathrm{shape}}_{isys,ipar,ich,ibin} = g(0,1)_{isys} \times \delta_{isys,ipar,ich,ibin}, \tag{8.13}$$

where $g(0,1)_{isys}$ is a random number sampled from a Gaussian with zero mean and unit standard deviation for the $isys^{th}$ systematic, and

$$\begin{aligned} \delta_{isys,ipar,ich,ibin} &= \delta^{+}_{isys,ipar,ich,ibin}, \quad \text{if } g(0,1)_{isys} \text{ is a positive number,} \\ &= \delta^{-}_{isys,ipar,ich,ibin}, \quad \text{otherwise.} \end{aligned} \tag{8.14}$$

The fluctuations (8.13) in a given bin are added linearly to the yields (8.12) generated from the sampling of the normalization-only systematic effects weighted by the nominal yield in that bin:

$$(c_{ipar,ich,ibin})'' = (c_{ipar,ich})' \times \frac{C_{ipar,ich,ibin}}{C_{ipar,ich}} + \sum_{isys} \Delta^{\text{shape}}_{isys,ipar,ich,ibin}, \tag{8.15}$$

where $C_{ipar,ich}$ is the total estimated yield summed over all bins of the discriminant as it appears in Eq. (8.12).

8.2 Definition of Ensemble Tests, and Expected and Observed Measurements

In order to validate the method, it is necessary to study its behavior on pseudo-data samples with characteristics as close as possible to those of the real data. Such ensembles can be used to determine whether an analysis is able to extract a cross section from a signal masked by large backgrounds, or if the claimed accuracy is warranted. Moreover, by running different analyses on exactly the same ensembles one can study in detail the correlations across analyses and the frequency properties of combined results and their significance. These all are called *ensemble tests*. For faster calculations and as a first glance, one can also simply set the count D in any bin to the predicted yield in that bin, getting only one set of pseudo-data events. Results from the subsequent statistical analysis are called *expected* results, while those corresponding to real data are termed as *observed* results.

8.3 Expected Results

In this section, the expected results for the single top cross section measurement and its sensitivity are presented. The input distributions are the BNN outputs shown in Appendix G for each of the twenty four individual channels, but where the data are replaced in each bin by the background model content plus an amount of single top ($tb + tqb$) at the Standard Model cross section value of 3.46 pb. From the BNN discriminants, a 1-dimensional binned likelihood posterior density for the combined $tb+tqb$ production cross section is computed as described in Sect. 8.1 and the measured cross section defined as the value at which the posterior has its maximum.

Figure 8.2 shows the Bayesian posterior probability density distribution for expected results, for the combination of all the twenty four analysis channels, and including all the systematics. The vertical solid line shows the position of the peak and the vertical dashed lines show the bounds of the 68.3%, 95.4%, 99.7%, 99.994% and 99.99994% two-sided confidence intervals.

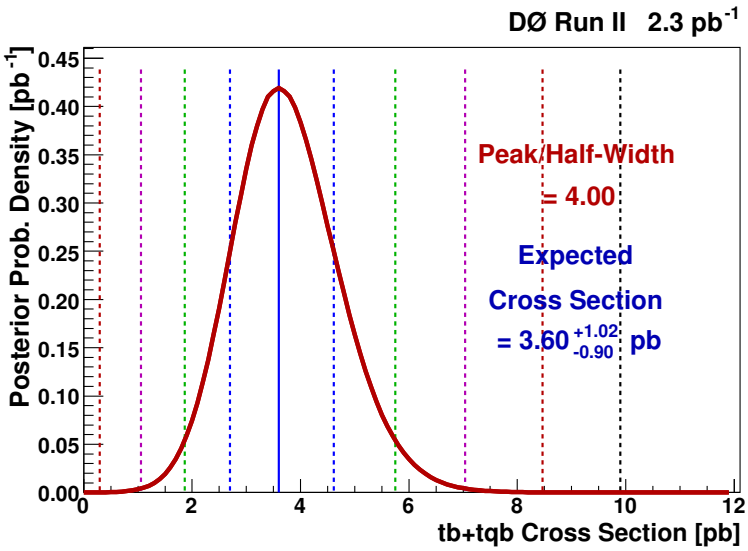

Figure 8.2 : Posterior probability density for expected results for all the twenty four channels combined as a function of the $tb+tqb$ production cross section. The theoretical cross section value is 3.46 pb. All systematic uncertainties are included.

Table 8.1 shows the expected cross sections with its uncertainties for various channel combinations. All of them are consistent within errors with the Standard Model cross section of 3.46 pb used as input. The deviations from 3.46 pb are expected given the handling of MC statistics uncertainties. The deviations are larger in channels that have very little signal.

The sensitivity of the measurements is estimated from the Bayes factor significance and the Bayes ratio significance. The Bayes factor, B_{10}, for a given set of data $\mathbf{D}$, is calculated as [108]

$$B_{10} = \frac{p(\mathbf{D}|H_1)}{p(\mathbf{D}|H_0)} = \frac{\int \mathcal{L}(\mathbf{D}|\sigma)\,\pi(\sigma)\,d\sigma}{\mathcal{L}(\mathbf{D}|\sigma=0)}, \tag{8.16}$$

where H_0 (H_1) is the background-only (signal+background) hypothesis, $\pi(\sigma)$ is the signal cross section prior and $\mathcal{L}(\mathbf{D}|\sigma)$ the likelihood given by the double integral in the r.h.s of Eq. (8.8).

where $\mathbf{a}$ is the effective luminosity for the signal and $\underline{\mathbf{b}}$ are the yields for all background sources. The Bayes ratio, B_r, is an approximation to the Bayes factor given by the posterior at the peak over the posterior at zero:

$$B_r = \frac{p(\sigma_{peak}|\mathbf{D})}{p(\sigma=0|\mathbf{D})}, \tag{8.17}$$

which to one significant digit gives the same result as the Bayes factor. Table 8.2 shows the Bayes ratio significance calculated as $\sqrt{2\ln(B_r)}$ for the measurements presented in Table 8.1. Another quantity that is used to estimate the sensitivity of the measurements

is the "peak over half-width", defined as the ratio of the peak position of the posterior over the size of the interval given by the peak position and the position of the lower bound of the 68.27% two-sided confidence interval:

$$\mathrm{P/HW} = \frac{\sigma_{\mathrm{peak}}}{\Delta\sigma^{(-)}}. \tag{8.18}$$

The peak over half-width basically says how many sigmas differs from zero the measured cross section. As with the Bayes ratio and Bayes factor significances, the higher the value the more accurate the measurement. Note that by definition $\mathrm{P/HW} \geq 1$ (it is equal to 1 when the lower bound of the 68.27% interval is at 0) or $\mathrm{P/HW} = 0$ when the peak of the posterior is at 0. Table 8.3 shows the corresponding peak over half-width values for the same channel combinations as in Table 8.1.

Finally, Table 8.4 presents the expected cross section, peak over half-width and Bayes ratio significance for each of the twenty four analysis channels.

Expected Cross Sections

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets	
	e-chan	μ-chan	1 tag	2 tags
Run IIa	$3.61^{+1.98}_{-1.71}$	$3.70^{+1.74}_{-1.60}$	$3.66^{+1.58}_{-1.38}$	$3.64^{+2.51}_{-2.29}$
Run IIb	$3.69^{+1.99}_{-1.78}$	$3.72^{+1.98}_{-1.72}$	$3.85^{+1.61}_{-1.57}$	$3.54^{+2.64}_{-2.29}$
Run IIa+b	$3.66^{+1.46}_{-1.27}$	$3.69^{+1.32}_{-1.22}$	$3.67^{+1.15}_{-1.05}$	$3.62^{+1.92}_{-1.74}$

	e,μ + 1,2tags			All
	2 jets	3 jets	4 jets	channels
Run IIa	$3.70^{+1.68}_{-1.43}$	$3.81^{+2.84}_{-2.47}$	$4.00^{+5.46}_{-4.00}$	$\mathbf{3.66^{+1.35}_{-1.24}}$
Run IIb	$3.89^{+1.86}_{-1.65}$	$3.73^{+2.83}_{-2.37}$	$3.86^{+5.29}_{-3.86}$	$\mathbf{3.73^{+1.44}_{-1.30}}$
Run IIa+b	$3.77^{+1.34}_{-1.13}$	$3.78^{+2.08}_{-1.88}$	$3.80^{+4.32}_{-3.43}$	$\mathbf{3.60^{+1.02}_{-0.90}}$

Table 8.1 : Expected cross section, with all systematic uncertainties taken into account, for many combinations of analysis channels. The values from all channels combined are shown in bold type.

Expected Bayes Ratio Significances

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets		e,μ + 1,2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
Run IIa	2.2	2.5	2.9	1.6	2.9	1.5	0.6	**3.3**
Run IIb	2.2	2.3	2.7	1.5	2.6	1.6	0.7	**3.1**
Run IIa+b	3.1	3.3	3.9	2.2	3.8	2.1	0.9	**4.4**

Table 8.2 : Expected Bayes ratio significance, with all systematic uncertainties taken into account, for many combinations of analysis channels. The values from all channels combined are shown in bold type.

Expected Posterior Peak over Half-Widths

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets		e,μ + 1,2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
Run IIa	2.1	2.3	2.7	1.6	2.6	1.5	1.0	**3.0**
Run IIb	2.1	2.2	2.5	1.5	2.4	1.6	1.0	**2.9**
Run IIa+b	2.9	3.0	3.5	2.1	3.3	2.0	1.1	**4.0**

Table 8.3 : Expected posterior peak over half-width, with all systematic uncertainties taken into account, for many combinations of analysis channels. The values from all channels combined are shown in bold type.

Expected Results in Individual Channels

	Electron Channel			Muon Channel		
	$\sigma \pm \Delta\sigma^{(\pm)}$	P/HW	BRS	$\sigma \pm \Delta\sigma^{(\pm)}$	P/HW	BRS
Run IIa / 1tag / 2jets	$3.6^{+2.6}_{-2.1}$	1.7	1.8	$3.7^{+2.3}_{-1.9}$	1.9	2.0
Run IIa / 1tag / 3jets	$4.0^{+4.3}_{-3.6}$	1.1	0.9	$4.2^{+3.9}_{-3.4}$	1.2	1.1
Run IIa / 1tag / 4jets	$4.3^{+10.9}_{-4.3}$	1.0	0.5	$4.7^{+11.5}_{-4.7}$	1.0	0.4
Run IIa / 2tags / 2jets	$3.8^{+4.1}_{-3.4}$	1.1	1.0	$3.8^{+3.8}_{-3.1}$	1.2	1.1
Run IIa / 2tags / 3jets	$3.9^{+6.6}_{-3.9}$	1.0	0.6	$3.9^{+5.8}_{-3.9}$	1.0	0.7
Run IIa / 2tags / 4jets	$4.0^{+13.3}_{-4.0}$	1.0	0.3	$3.7^{+13.5}_{-3.7}$	1.0	0.3
Run IIb / 1tag / 2jets	$3.8^{+2.7}_{-2.3}$	1.7	1.7	$3.9^{+2.7}_{-2.3}$	1.7	1.8
Run IIb / 1tag / 3jets	$3.7^{+4.0}_{-3.3}$	1.1	1.0	$4.4^{+4.2}_{-3.6}$	1.2	1.1
Run IIb / 1tag / 4jets	$4.9^{+10.5}_{-4.9}$	1.0	0.5	$4.8^{+10.1}_{-4.8}$	1.0	0.6
Run IIb / 2tags / 2jets	$3.6^{+4.2}_{-3.4}$	1.1	0.9	$4.0^{+4.4}_{-3.5}$	1.2	1.1
Run IIb / 2tags / 3jets	$3.7^{+6.5}_{-3.7}$	1.0	0.7	$4.3^{+6.3}_{-4.3}$	1.0	0.8
Run IIb / 2tags / 4jets	$3.7^{+13.1}_{-3.7}$	1.0	0.3	$4.3^{+12.4}_{-4.3}$	1.0	0.4

Table 8.4 : Expected cross section, posterior peak over half-width and Bayes ratio significance, with all systematic uncertainties taken into account, for each of the twenty four analysis channels.

8.3.1 Stability Test

A stability test is performed in which a BNN is trained, in all the twenty four analysis channels, with a reduced set of variables. At the end of Appendix E, Tables E.25 to E.30 show the overlap of the selected variables used as inputs for the BNNs, between channels that have the same b-tag and jet multiplicities. So, for example, Table E.25 shows the variables that are used by the Bayesian neural networks in the four channels with 1 b-tag and 2 jets. Variables that are common in at least two of the four channels are presented in the two left columns, while variables that are only used in one of the four channels, are presented in the two right columns. The test consists in retraining the BNNs and rederiving the expected results, after removing the "channel-specific" variables (those not shared by at least two of the four channels). So, for example, in the

1 b-tag and 2 jets case, four variables are removed in the p17 e+jets channel (p_T(jet1), E(jet2), p_T(best1) and cos(lepton,Q(lepton)$\times z$)$_{\text{besttop}}$), one variable is removed in the p20 e+jets channel (cos(best1,lepton)$_{\text{besttop}}$), five variables in the p17 μ+jets channel (p_T(lepton), and so on.

Tables 8.5 and 8.6 show the expected cross section and peak over half-width respectively, measured for different combinations of channels after applying to the background and signal models the new BNNs. Table 8.7 presents the same two quantities for each of the twenty four analysis channels. Comparing these numbers with those presented in Sect. 8.3, one sees that the agreement is very high and thus concludes that the expected results are stable under the removal of the channel-specific variables.

Expected Cross Sections
(reduced set of variables)

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets	
	e-chan	μ-chan	1 tag	2 tags
Run IIa	$3.7^{+2.0}_{-1.8}$	$3.7^{+1.7}_{-1.6}$	$3.7^{+1.5}_{-1.4}$	$3.5^{+2.6}_{-2.2}$
Run IIb	$3.8^{+2.0}_{-1.8}$	$3.7^{+1.9}_{-1.7}$	$3.8^{+1.6}_{-1.5}$	$3.6^{+2.7}_{-2.3}$
Run IIa+b	$3.7^{+1.4}_{-1.3}$	$3.6^{+1.3}_{-1.2}$	$3.7^{+1.2}_{-1.0}$	$3.6^{+2.0}_{-1.7}$

	e,μ + 1,2tags			All
	2 jets	3 jets	4 jets	channels
Run IIa	$3.7^{+1.7}_{-1.5}$	$3.7^{+2.8}_{-2.5}$	$3.8^{+5.5}_{-3.8}$	$\mathbf{3.7^{+1.3}_{-1.3}}$
Run IIb	$3.9^{+1.9}_{-1.6}$	$3.8^{+2.8}_{-2.4}$	$4.0^{+5.4}_{-4.0}$	$\mathbf{3.8^{+1.4}_{-1.4}}$
Run IIa+b	$3.8^{+1.3}_{-1.2}$	$3.8^{+2.1}_{-1.9}$	$4.0^{+4.3}_{-3.5}$	$\mathbf{3.7^{+1.0}_{-0.9}}$

Table 8.5 : Expected cross section, with all systematic uncertainties taken into account, for many combinations of analysis channels, corresponding to a BNN trained with only the variables that are shared by at least two of the four channels with same b-tag and jet multiplicities. The values from all channels combined are shown in bold type.

Expected Posterior Peak over Half-Widths (reduced set of variables)

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets		e,μ + 1,2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
Run IIa	2.0	2.3	2.6	1.6	2.5	1.5	1.0	**2.9**
Run IIb	2.1	2.2	2.5	1.6	2.3	1.6	1.0	**2.8**
Run IIa+b	2.8	3.1	3.5	2.1	3.2	2.0	1.1	**4.1**

Table 8.6 : Expected posterior peak over half-width, with all systematic uncertainties taken into account, for many combinations of analysis channels, corresponding to a BNN trained with only the variables that are shared by at least two of the four channels with same b-tag and jet multiplicities. The values from all channels combined are shown in bold type.

Expected Results in Individual Channels (reduced set of variables)

	Electron Channel		Muon Channel	
	$\sigma \pm \Delta\sigma^{(\pm)}$	P/HW	$\sigma \pm \Delta\sigma^{(\pm)}$	P/HW
Run IIa / 1tag / 2jets	$3.7^{+2.5}_{-2.2}$	1.7	$3.7^{+2.3}_{-1.9}$	1.9
Run IIa / 1tag / 3jets	$4.1^{+4.3}_{-3.8}$	1.1	$4.0^{+3.9}_{-3.3}$	1.2
Run IIa / 1tag / 4jets	$4.2^{+10.8}_{-4.2}$	1.0	$4.6^{+10.7}_{-4.6}$	1.0
Run IIa / 2tags / 2jets	$3.8^{+4.1}_{-3.4}$	1.1	$3.8^{+3.9}_{-3.2}$	1.2
Run IIa / 2tags / 3jets	$3.7^{+6.7}_{-3.7}$	1.0	$3.8^{+5.8}_{-3.8}$	1.0
Run IIa / 2tags / 4jets	$4.4^{+13.0}_{-4.4}$	1.0	$3.5^{+13.1}_{-3.5}$	1.0
Run IIb / 1tag / 2jets	$3.9^{+2.6}_{-2.3}$	1.7	$3.8^{+2.7}_{-2.2}$	1.7
Run IIb / 1tag / 3jets	$3.8^{+4.1}_{-3.4}$	1.1	$4.3^{+4.2}_{-3.5}$	1.2
Run IIb / 1tag / 4jets	$4.9^{+10.0}_{-4.9}$	1.0	$5.3^{+10.0}_{-5.3}$	1.0
Run IIb / 2tags / 2jets	$3.6^{+4.2}_{-3.5}$	1.0	$4.1^{+4.4}_{-3.5}$	1.2
Run IIb / 2tags / 3jets	$3.9^{+6.6}_{-3.9}$	1.0	$4.3^{+6.3}_{-4.3}$	1.0
Run IIb / 2tags / 4jets	$3.8^{+13.4}_{-3.8}$	1.0	$4.2^{+12.7}_{-4.2}$	1.0

Table 8.7 : Expected cross section and posterior peak over half-width, with all systematic uncertainties taken into account, for each of the twenty four analysis channels, corresponding to a BNN trained with only the variables that are shared by at least two of the four channels with same b-tag and jet multiplicities.

8.4 Ensemble Tests

To cross check the ability of the method to measure the single top cross section, several ensembles of pseudo-datasets are generated from the total pool of signal and background model events. All systematic uncertainties are taken into account during the ensemble generation for each pseudo-dataset individually. For details about the generation of the ensembles the reader should refer to Ref. [108].

Eight ensembles are generated, each with at least 7,600 entries (pseudo-experiments), and each with a different amount of single top, but keeping the relative tb to tqb contribution as given by the ratio of their SM cross sections. In each such ensemble, the cross section is measured treating each pseudo-dataset as real data. The measured cross section distribution in each ensemble can be seen in Fig. 8.3. A Gaussian fit around the mean of the distribution is performed and the resulting Gaussian mean (and its error) used as the measured cross section in the ensemble. Figure 8.4 shows the measured single top cross section (as described above) as a function of the input value of the single top cross section used in the ensemble generation. A linear fit to the points is performed, which gives $\sigma^{tb+tqb}_{\text{meas}} = 0.032 + 0.993 \times \sigma^{tb+tqb}_{\text{input}}$, consistent, within two standard deviations, with no bias.

An additional very large ensemble with no single top is generated which is used to measure the sensitivity of the analysis (see Sect. 8.6).

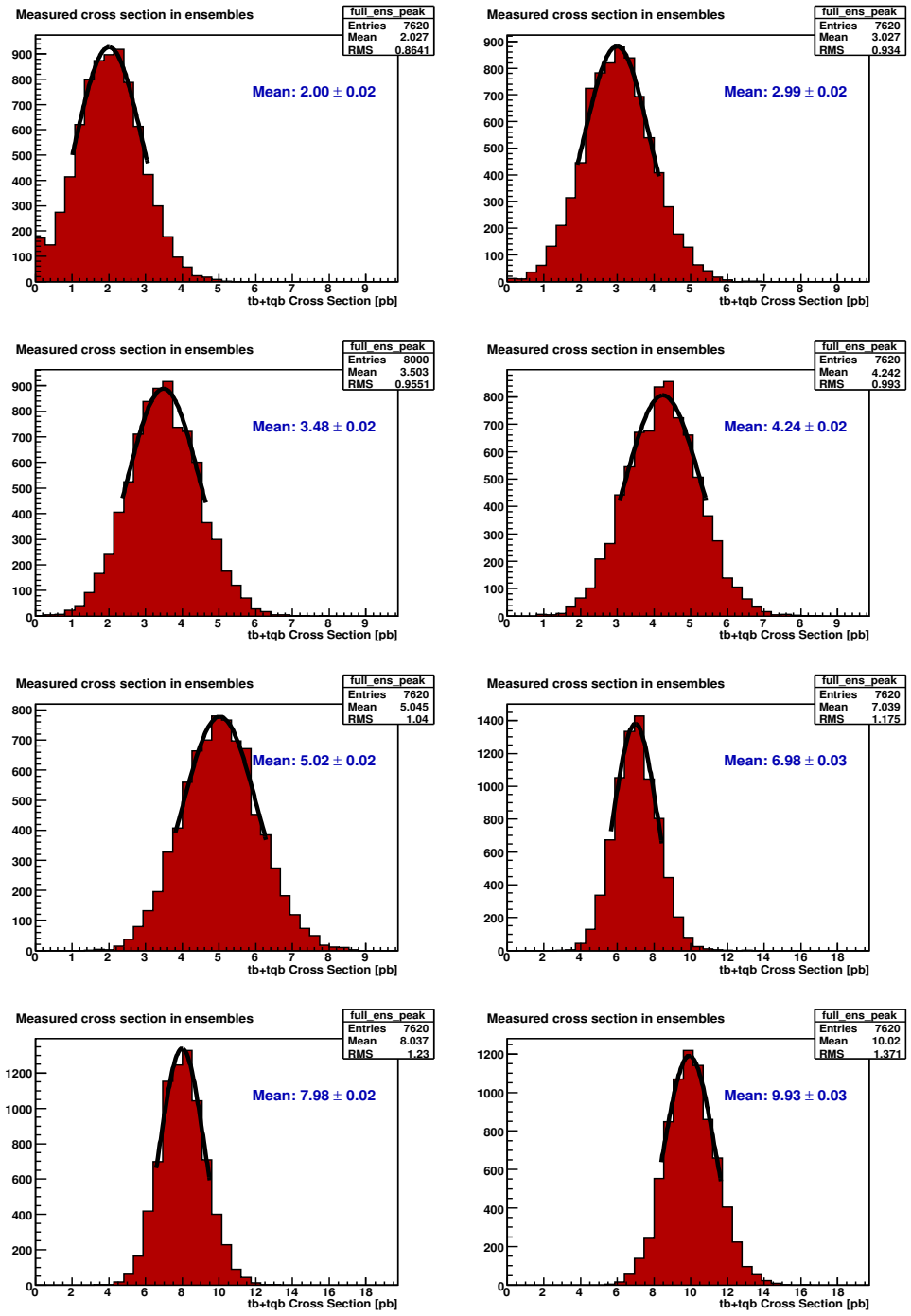

Figure 8.3 : Measured single top cross sections in ensembles with various amounts of single top. The first row shows the results from the 2 pb and the 3 pb ensembles, the second row shows the results from the SM (3.46 pb) and the 4.2 pb ensembles, the third row shows the results from the 5 pb and the 7 pb ensembles, and the fourth row shows the results from the 8 pb and the 10 pb ensembles.

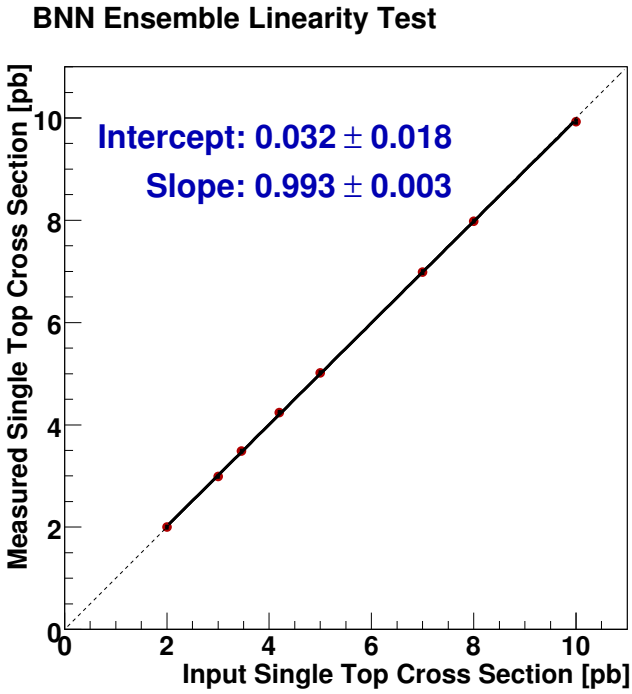

Figure 8.4 : Linear fit through the means from the gaussian fits (Fig. 8.3) of the measured cross sections in ensembles generated with different amounts of single top. The fit is constrained to the range $[2, 10]$. The correct cross section is measured on average.

8.5 Observed Results

This section presents the BNN observed cross section results using the 2.3 fb^{-1} dataset. Figure 8.5 shows the BNN output for all 24 channels combined, when the largest sensitivity is obtained. The BNN-output distributions for the Run IIa+b e+μ combination, but separate b-tag and jet multiplicities are shown in Appendix G.

Figure 8.6 shows the posterior probability density when combining all twenty four analysis channels. The measured cross section is given by the peak position of the observed posterior probability density and corresponds to $4.70^{+1.18}_{-0.93}$ pb.

Table 8.8 shows the $tb + tqb$ measured cross sections for various channel combinations. The corresponding Bayes ratio significance and peak over half-width values are shown in Tables 8.9 and 8.10 respectively. The corresponding results for the twenty four analysis channels are shown in Table 8.11. The individual channels are not sensitive by themselves, but the combination of all channels is. In particular, the "0.0" measured cross section values that appear in Table 8.11 correspond to individual channels with low statistics.

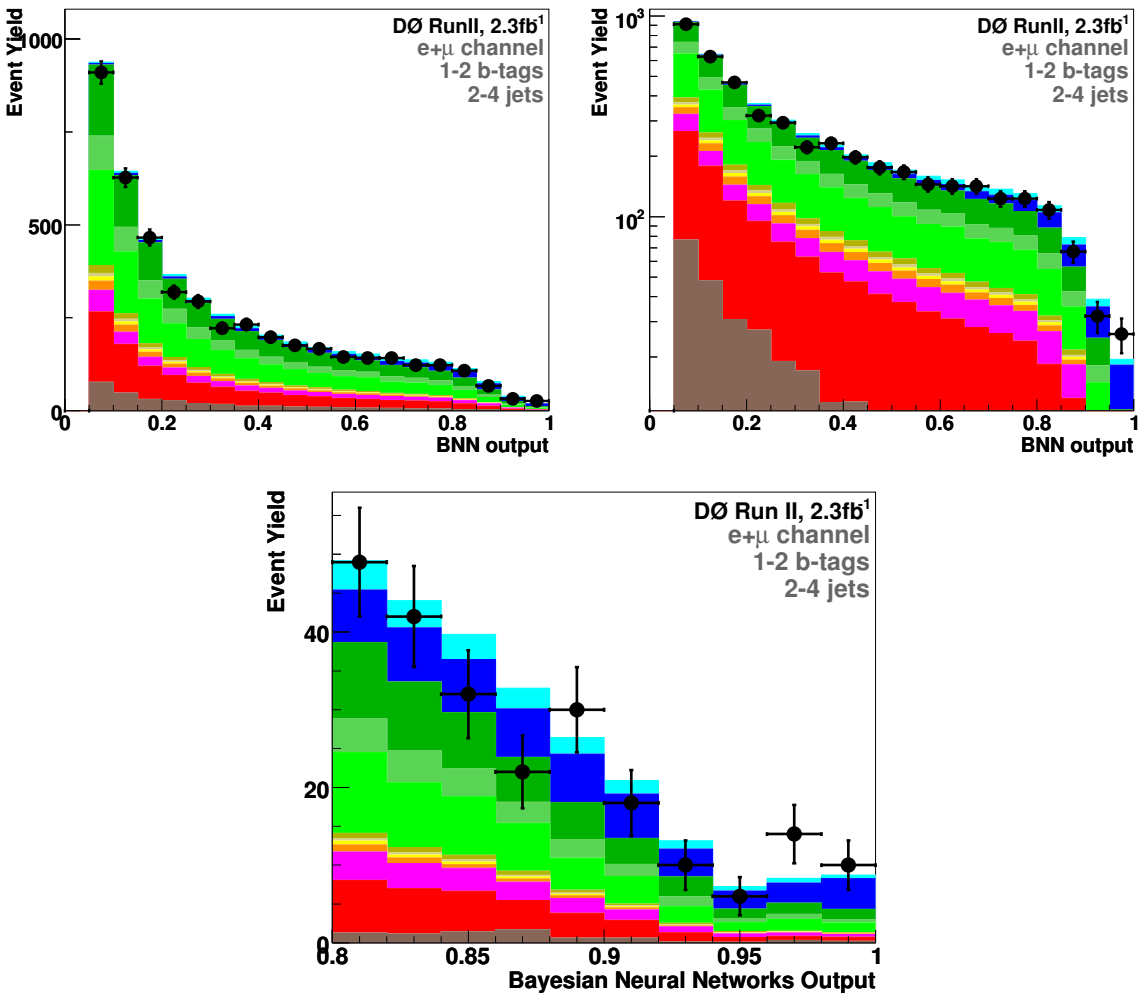

Figure 8.5 : Bayesian neural network discriminant output for all twenty four channels combined in linear scale (upper left plot) and logarithmic scale (upper right plot). The total single top contribution in these plots is normalized to the measured cross section (the relative tb to tqb contributions are given by the ratio of their SM cross sections). The lower plot is a zoom of the upper left plot in the high discriminant region.

Finaly, Kolmogorov-Smirnov (KS) test p-values for the BNN output distribution for the all channels combined case are calculated in different intervals of the discriminant. The upper limit of the intervals is fixed to 1 for all cases, while the lower limit varies between 0 and 0.9. Figure 8.7 illustrates the results. Data are compared to the background model only (green curve) and to background+signal (blue curve). The points at 0 correspond to the KS-test p-values calculated in the interval [0,1], the points at 0.1 correspond to the interval [0.1,1], and so on. One can conclude from this plot that the data are far more consistent with the presence of single top signal than with its absence, in addition to the predicted background.

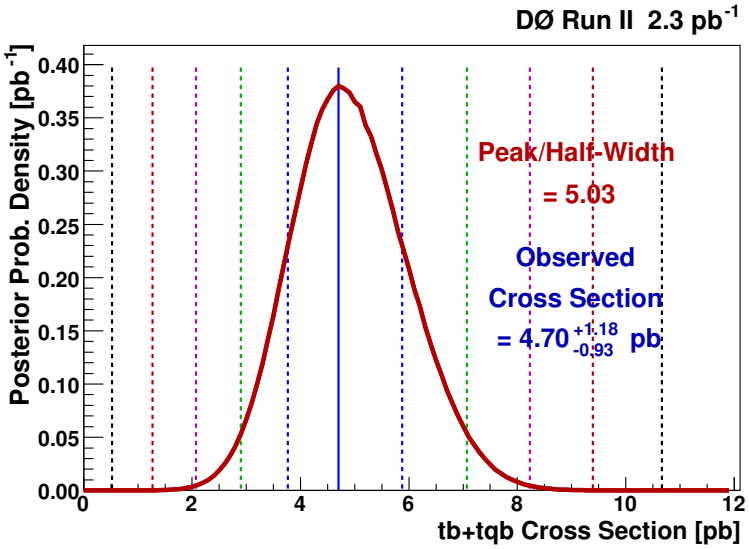

Figure 8.6 : Observed posterior probability density for all twenty four channels combined as a function of the $tb+tqb$ production cross section. All systematic uncertainties were included.

Observed Cross Section Measurements

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets	
	e-chan	μ-chan	1 tag	2 tags
Run IIa	$3.79^{+2.08}_{-1.86}$	$2.98^{+1.74}_{-1.60}$	$2.81^{+1.54}_{-1.35}$	$3.14^{+2.61}_{-2.25}$
Run IIb	$6.38^{+2.23}_{-2.01}$	$4.96^{+1.98}_{-1.69}$	$6.35^{+1.89}_{-1.67}$	$5.10^{+2.80}_{-2.30}$
Run IIa+b	$5.29^{+1.50}_{-1.37}$	$3.96^{+1.26}_{-1.16}$	$4.73^{+1.36}_{-1.14}$	$4.43^{+1.90}_{-1.80}$

	e,μ + 1,2tags			All
	2 jets	3 jets	4 jets	channels
Run IIa	$1.95^{+1.41}_{-1.24}$	$4.97^{+3.23}_{-2.78}$	$8.48^{+8.73}_{-6.38}$	$\mathbf{3.40^{+1.42}_{-1.32}}$
Run IIb	$4.49^{+2.08}_{-1.78}$	$7.69^{+3.37}_{-2.80}$	$11.83^{+11.06}_{-6.90}$	$\mathbf{5.74^{+1.47}_{-1.28}}$
Run IIa+b	$3.08^{+1.41}_{-1.16}$	$6.60^{+2.36}_{-2.05}$	$10.93^{+8.10}_{-5.67}$	$\mathbf{4.70^{+1.18}_{-0.93}}$

Table 8.8 : Observed cross section, with all systematic uncertainties taken into account, for many combinations of analysis channels. The best values from all channels combined, with systematics, are shown in bold type.

Observed Bayes Ratio Significance

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets		e,μ + 1,2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
Run IIa	2.2	1.9	2.2	1.3	1.5	1.8	1.4	**2.7**
Run IIb	3.7	3.3	4.5	2.5	3.0	3.4	2.2	**5.1**
Run IIa+b	4.3	3.8	4.7	2.7	3.0	3.8	2.5	**5.7**

Table 8.9 : Observed Bayes ratio significance, with all systematic uncertainties included, for many combinations of analysis channels. The best values from all channels combined, with systematics, are shown in bold type.

Observed Posterior Peak over Half-Width

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets		e,μ + 1,2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
Run IIa	2.0	1.9	2.1	1.4	1.6	1.8	1.3	**2.6**
Run IIb	3.2	2.9	3.8	2.2	2.5	2.7	1.7	**4.5**
Run IIa+b	3.9	3.4	4.2	2.5	2.7	3.2	1.9	**5.0**

Table 8.10 : Observed posterior peak over half-width, with all systematic uncertainties taken into account, for many combinations of analysis channels. The best values from all channels combined, with systematics, are shown in bold type.

Observed Results in Individual Channels

	Electron Channel			Muon Channel		
	$\sigma \pm \Delta\sigma^{(\pm)}$	P/HW	BRS	$\sigma \pm \Delta\sigma^{(\pm)}$	P/HW	BRS
Run IIa / 1tag / 2jets	$2.3^{+2.0}_{-1.8}$	1.3	1.2	$2.6^{+2.0}_{-1.8}$	1.4	1.3
Run IIa / 1tag / 3jets	$10.0^{+6.7}_{-5.4}$	1.8	2.0	$2.0^{+3.3}_{-2.0}$	1.0	0.6
Run IIa / 1tag / 4jets	$13.1^{+13.4}_{-10.1}$	1.3	1.4	$0.0^{+10.3}_{-0.0}$	0.0	0.0
Run IIa / 2tags / 2jets	$0.0^{+3.2}_{-0.0}$	0.0	0.0	$3.6^{+3.8}_{-3.1}$	1.1	1.0
Run IIa / 2tags / 3jets	$6.0^{+7.0}_{-5.7}$	1.1	0.9	$0.0^{+6.5}_{-0.0}$	0.0	0.0
Run IIa / 2tags / 4jets	$6.0^{+15.4}_{-6.0}$	1.0	0.6	$10.1^{+13.8}_{-10.1}$	1.0	0.9
Run IIb / 1tag / 2jets	$4.0^{+2.8}_{-2.3}$	1.7	1.8	$5.9^{+3.0}_{-2.5}$	2.3	2.7
Run IIb / 1tag / 3jets	$7.0^{+4.9}_{-3.9}$	1.8	2.1	$6.7^{+5.2}_{-4.3}$	1.6	1.6
Run IIb / 1tag / 4jets	$14.7^{+15.6}_{-11.5}$	1.3	1.5	$6.9^{+11.6}_{-6.9}$	1.0	0.8
Run IIb / 2tags / 2jets	$2.6^{+4.1}_{-2.6}$	1.0	0.6	$5.4^{+4.6}_{-3.6}$	1.5	1.6
Run IIb / 2tags / 3jets	$9.4^{+8.9}_{-6.4}$	1.5	1.7	$5.4^{+6.3}_{-4.8}$	1.1	1.0
Run IIb / 2tags / 4jets	$23.2^{+21.9}_{-14.5}$	1.6	2.0	$8.0^{+12.3}_{-8.0}$	1.0	0.8

Table 8.11 : Observed cross section, posterior peak over half-width and Bayes factor significance, with all systematic uncertainties taken into account, for the twenty four analysis channels.

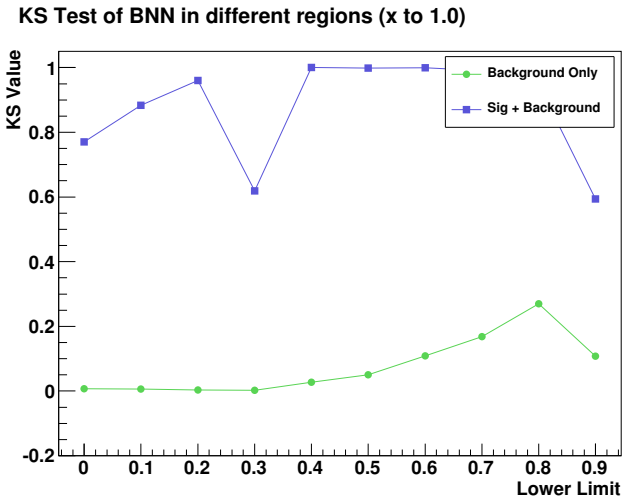

Figure 8.7 : KS-test p-values for the BNN output distribution (with all channels combined) in the intervals $[x,1]$, for ten different values of x. The green curve shows the KS test values when comparing data against the background model only, while the blue curve shows the KS test values when comparing data against background model plus $tb + tqb$ scaled to the measured cross section.

8.6 Significance

To measure the significance of the measured signal cross section, a large ensemble of zero-signal pseudo-datasets is used to represent experiments under the background-only hypothesis. Each such dataset corresponds to 2.3 fb^{-1} of data without any single top. The aim is to get an estimate of how probable it is that the measured value of the cross section is just due to a statistical fluctuation of the background instead to the presence of real signal. The single top cross section is measured in each of the pseudo-datasets in exactly the same way as it is measured in the real dataset. The distribution of the measured cross sections is shown in Figs. 8.8 and 8.9.

From the measured cross sections in the ensemble, the probability that data containing no single top quark events could fluctuate up to or above the measured cross section in the real data set, is calculated dividing the number of pseudo-datasets above the measured cross section by the total number of pseudo-datasets. This probability is known as the "p-value" (α). The significance of the measurement is then estimated as the quantile function (the inverse of the normal cumulative distribution function) at probability $1 - \alpha$; so, it is measured in units of "sigmas", the standard deviation of a Gaussian distribution:

$$N_\sigma = \Phi^{-1}(1-\alpha) = \sqrt{2} \cdot \mathrm{erf}^{-1}(1-2\alpha), \quad \text{or} \tag{8.19}$$

$$\Phi(N_\sigma) = \int_{-\infty}^{N_\sigma} \mathrm{Gauss}(x;0,1)\, dx = 1-\alpha. \tag{8.20}$$

The p-value and the corresponding number of standard deviations for the expected and the observed excess of signal over background are given in Figs. 8.8 and 8.9 respectively. Out of the 62.8 million zero-signal pseudo-datasets, only 1457 of them are expected to give a cross section at least as high as the SM prediction of 3.46 pb, and only 2 are expected to fluctuate as high as the measured value of 4.70 pb, giving a signal significance of 5.4σ.

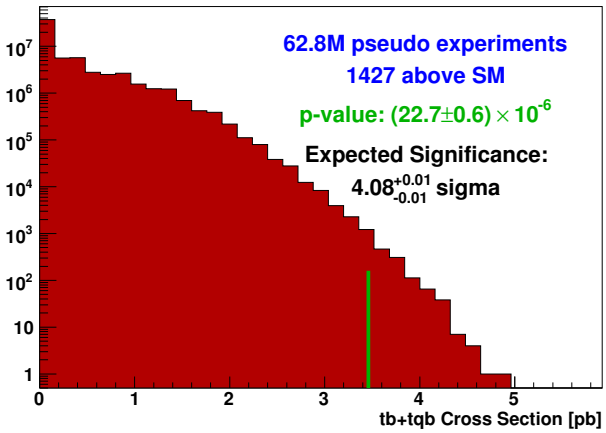

Figure 8.8 : Measured cross section from a large ensemble of pseudo-datasets containing no single top. The expected significance is calculated from the number of pseudo-datasets measuring a cross section higher than the standard model cross section (green line).

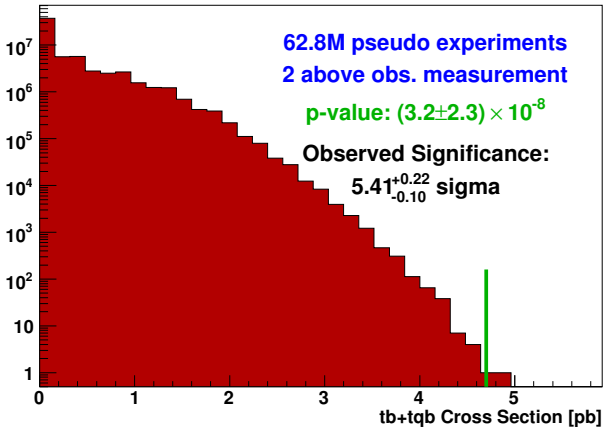

Figure 8.9 : Measured cross section from a large ensemble of pseudo-datasets containing no single top. The observed significance is calculated from the number of pseudo-datasets measuring a cross section higher than the measured cross section on the real dataset (green line).

8.7 Measurement of $|V_{tb}|$

By means of analytical expressions for the single top quark production cross section, the same approach that is used to measure the cross section (see Sect. 8.1.1) can also be used to measure $|V_{tb}|^2$. This section presents the measurement of $|V_{tb}|$ for the Bayesian neural network analysis.

The measurement does not assume any restriction on the number of generations nor the unitarity of the CKM matrix. However, what it does assume is that the only existing production mechanism of single top quarks involves the interaction with a W boson, implying that extensions of the SM where single top quark events can be produced, e.g. via flavor-changing neutral current (FCNC) interactions or heavy scalar or vector boson exchange, are not considered. A second assumption is that $|V_{td}|^2+|V_{ts}|^2 \ll |V_{tb}|^2$, something that in Sect. 1.3.3 was shown to be consistent with measurements from DØ and CDF. The requirement that $|V_{td}|^2+|V_{ts}|^2 \ll |V_{tb}|^2$ implies that $B(t \to Wb) \simeq 100\%$ and that single top quark production is completely dominated by the Wtb interaction. This assumption has been made explicitly when measuring the combined $tb+tqb$ cross section assuming the SM ratio of $\sigma(tb)/\sigma(tqb)$ [109], as well as in the generation of single top and $t\bar{t}$ MC samples. Finally, it is assumed that the Wtb interaction is CP-conserving and of the type $V-A$, but it is allowed to have an anomalous strength. To understand why, consider the most general Lorentz invariant Wtb vertex, which is given by [110]

$$-\frac{g}{\sqrt{2}}V_{tb}^{*}W_{\mu}^{+}\bar{b}[\gamma^{\mu}(f_1^L\mathbb{P}_L+f_1^R\mathbb{P}_R)-\frac{i\sigma^{\mu\nu}}{m_W}(f_2^L\mathbb{P}_L+f_2^R\mathbb{P}_R)]t+h.c., \tag{8.21}$$

where the f_1 and f_2 anomalous couplings can a-priori be CP-violating. In the case of the SM, CP is conserved in the Wtb vertex and $f_1^L=1$ and $f_1^R=f_2^L=f_2^R=0$. Because the single top and $t\bar{t}$ MC used to perform this analysis assumes the SM Wtb vertex, the conditions $f_1^R=f_2^L=f_2^R=0$ are explicitly implied. However, this MC can still be used under the assumption of an anomalous f_1^L coupling, since *i)* the $t\bar{t}$ cross section and kinematics are completely unaffected, and *ii)* it would only rescale the single top tb and tqb cross sections allowing them to be larger or smaller than the SM prediction, but maintaining the SM ratio. Therefore, strictly speaking, what is measured here is the strength of the $V-A$ coupling, i.e. $|V_{tb}f_1^L|$, which is allowed to be >1. Limiting the measurement to the range [0,1] implies the additional assumption that $f_1^L=1$.

The $|V_{tb}|$ measurement forms a Bayesian posterior based on the BNN discriminants shown in Appendix G. The prior is chosen to be flat in $|V_{tb}f_1^L|^2$, following standard convention for parameters that multiply the cross section. Additional theoretical uncertainties need to be considered related to the terms entering in the analytical expression of the cross section. These uncertainties are applied separately to the tb and tqb samples in order to take the correlations into account properly. The additional uncertainties are shown in Table 8.12.

	tb	tqb
Top quark mass	5.56	3.48
Factorization scale	3.7	1.74
PDF	3.0	3.0
α_s	1.4	0.01

Table 8.12 : Systematic uncertainties in percent on the cross section factor required to extract $|V_{tb}|$. The uncertainty on the cross section factor due to the top quark mass corresponds to a mass uncertainty of 2.1 GeV/c^2.

The Bayesian posterior density for $|V_{tb}f_1^L|^2$ is shown in Fig. 8.10. The most probable value for $|V_{tb}f_1^L|^2$ is given by the peak of the posterior and is $|V_{tb}f_1^L|^2 = 1.31^{+0.37}_{-0.28}$, which results in $|V_{tb}f_1^L| = 1.14^{+0.15}_{-0.13}$. Thus, the measured strength of the $V-A$ coupling, or in other words, the CKM matrix element times coupling strength, is above the SM expectation by about 1 standard deviation. Restricting the prior to be non-zero only in the range [0,1], which corresponds to setting $f_1^L = 1$, one gets the $|V_{tb}|^2$ posterior, shown in Fig. 8.11. The posterior for $|V_{tb}|^2$ gives $|V_{tb}|^2 = 1.0^{+0.0}_{-0.14}$, which results in $|V_{tb}| = 1.0^{+0.0}_{-0.07}$. This posterior is also used to obtain a 95% confidence level (CL) lower limit on $|V_{tb}|^2$:

$$\int_{(|V_{tb}|^2)_{95\%\ \mathrm{CL\ limit}}}^{1} p(|V_{tb}|^2|D)\, d|V_{tb}|^2 = 0.95, \tag{8.22}$$

which gives $|V_{tb}|^2 > 0.68$, corresponding to $|V_{tb}| > 0.82$.

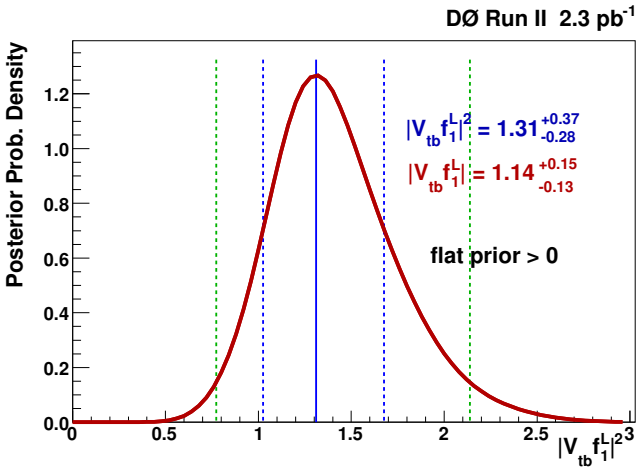

Figure 8.10 : Posterior probability density for $|V_{tb}f_1^L|^2$. The blue and green dashed lines indicate the limits of the 68% and 95% coverage areas respectively.

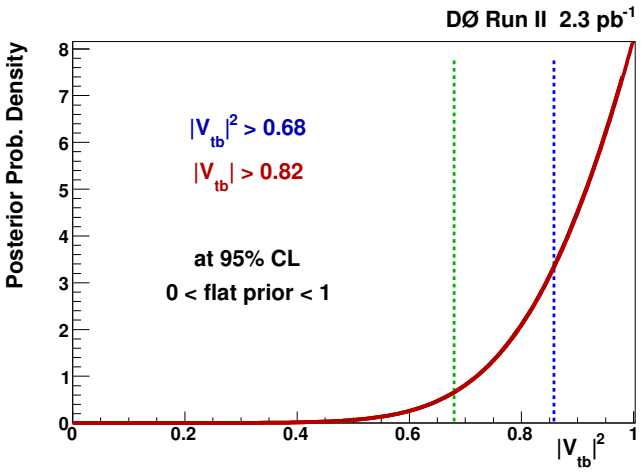

Figure 8.11 : Posterior probability density for $|V_{tb}|^2$ using a prior restricted to the region [0,1]. The blue and green dashed lines indicate the lower limit of the 68% and 95% coverage areas respectively.

8.8 Combination with the Boosted Decision Trees and Matrix Elements Analyses

In addition to the Bayesian neural network analysis, two other analyses are performed in DØ in parallel, which differ in the multivariate technique used to separate the single top signal from the overwhelming background. The two other techniques are boosted decision trees (BDT) and matrix element calculations (ME). The dataset, background modeling and event selection, are common to the three analyses. To measure the correlation between analyses, an ensemble of pseudo-datasets containing SM signal (3.46 pb) and background is generated using the final discriminant outputs to fill histograms for BNN, BDT and ME simultaneously. The produced pseudo-datasets are then processed by each of the three methods. The correlation between the methods are shown in Fig. 8.12. The correlation is higher between BNN and BDT (74%) since they are both machine learning techniques based on the same idea −exploit the differences in event kinematics between signal and background−, while the matrix element technique is essentially different −it consists on direct calculation (by numerical integration) of the probability for an event to be signal−, and therefore its correlation to the other methods is smaller (57% with BNN, 60% with BDT).

Because the three methods are not 100% correlated, it is worth to combine the three analyses to make use of as much information as possible, expecting to get a better expected (and maybe also observed) significance. The combination is done using another Bayesian neural network, which takes as inputs the outputs of the three individual methods, and provides its own discriminant output [98]. To the output of the combination, the same statistical analysis used by the other methods (described

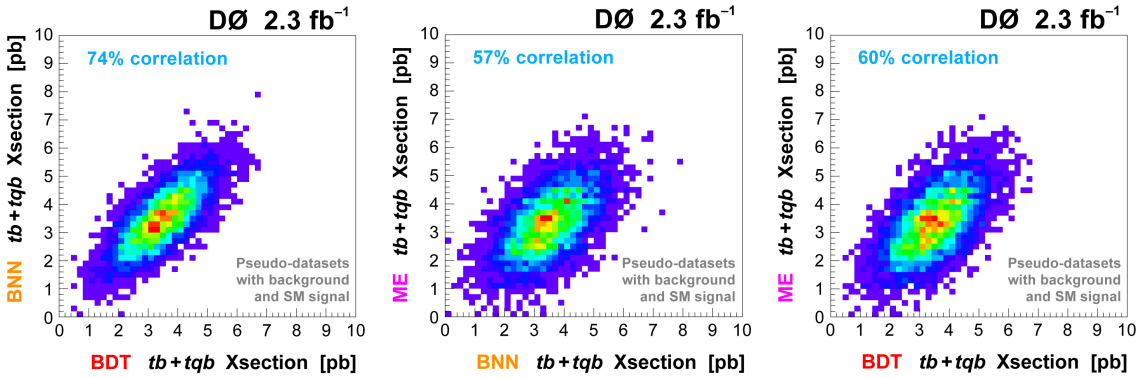

Figure 8.12 : Two-dimentional plots showing the measured cross sections in an ensemble of pseudo-datasets with SM signal and background for pairs of discriminants (left: BNN vs. BDT, middle: ME vs. BNN, right: ME vs. BDT).

in Sect. 8.1) is applied to produce the combined measurements, and a corresponding ensemble of pseudo-datasets with only background (no signal content) is generated to measure the significance. Results are shown in Fig. 8.13. These are the final results published by the collaboration. Table 8.13 shows the measured single top cross section, expected and observed significances, Wtb coupling strength ($|V_{tb}f_1^L|$) and 95% confidence level lower limit on $|V_{tb}|$ as measured by the three methods and their combination. One can see, indeed, that the expected significance improves when analyses are combined.

DØ 2.3 fb^{-1} Single Top Results

Analysis method	Single Top cross section	Significance expected	Significance observed	$\lvert V_{tb}f_1^L\rvert$	95% CL lower limit on $\lvert V_{tb}\rvert$
BNN	$4.70^{+1.18}_{-0.93}$ pb	4.1σ	5.4σ	$1.14^{+0.15}_{-0.13}$	> 0.82
BDT	$3.74^{+0.95}_{-0.79}$ pb	4.3σ	4.6σ	$1.05^{+0.13}_{-0.12}$	> 0.77
ME	$4.30^{+0.99}_{-1.20}$ pb	4.1σ	4.9σ	–	–
Combination	3.94 ± 0.88 pb	4.5σ	5.0σ	1.07 ± 0.12	> 0.78

Table 8.13 : Measured single top cross section, expected and observed significances, Wtb coupling strength ($|V_{tb}f_1^L|$) and 95% confidence level lower limit on $|V_{tb}|$ as measured by the three analysis techniques used by DØ and their combination.

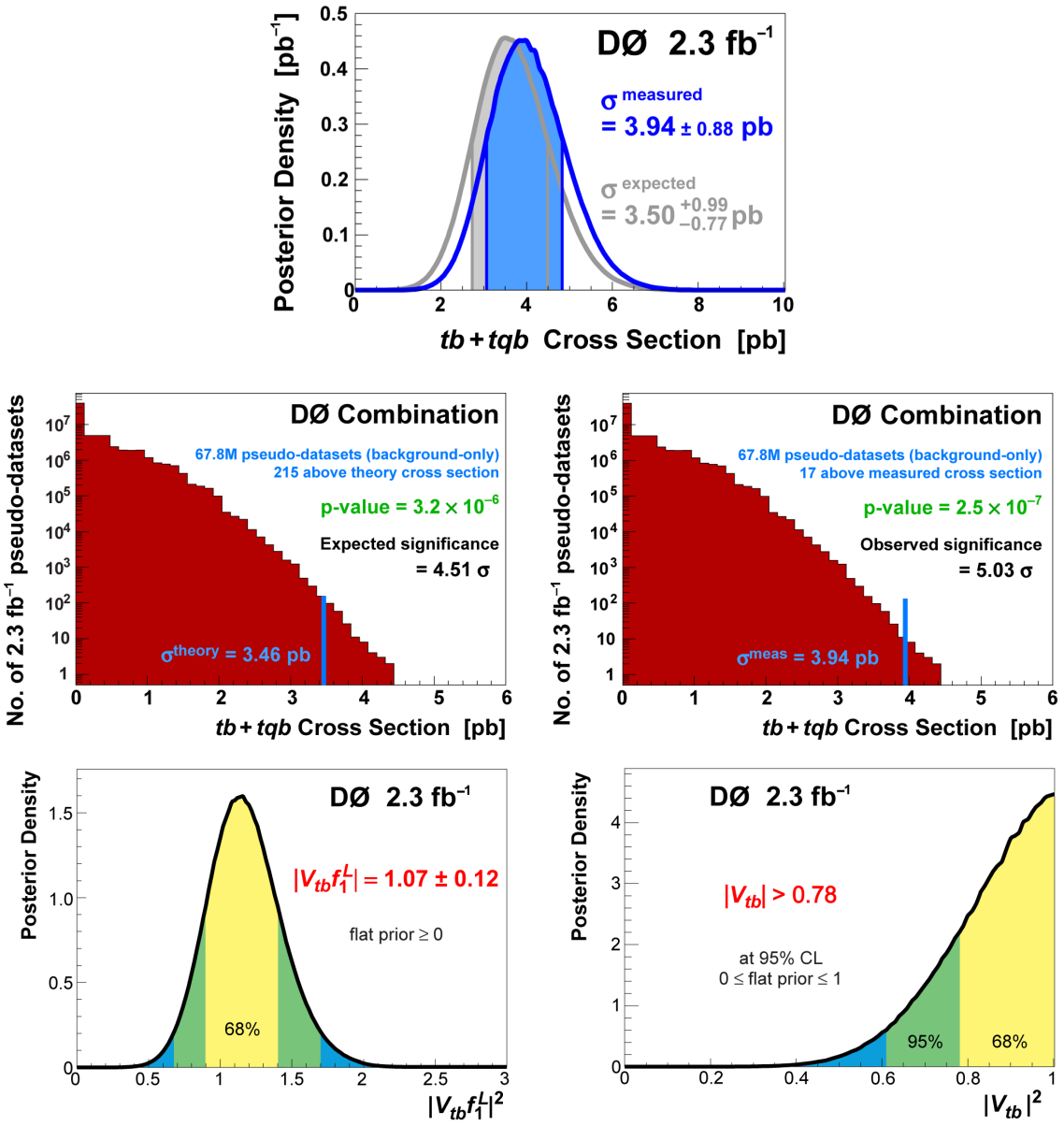

Figure 8.13 : Results obtained by the combination of the BNN, BDT and ME analysis: expected and observed posterior probability densities (upper plot), expected significance (middle left plot), observed significance (middle right plot), $|V_{tb}f_1^L|^2$ posterior (lower left plot) and $|V_{tb}|^2$ posterior (lower right).

9 Summary

Bayesian neural networks are used to separate expected single top quark signals from background in a sample of electron+jets and muon+jets events selected from 2.3 fb^{-1} of Tevatron Run II data. The Bayesian neural network output distributions across the twenty four independent channels used in the analysis are combined using a binned likelihood and the tb+tqb single top cross section is measured using a Bayesian method. The measured $tb + tqb$ single top quark production cross section is:

$$\sigma\left(p\bar{p} \rightarrow tb + X, tqb + X\right) = 4.70^{+1.18}_{-0.93} \text{ pb}.$$

The probability to measure a cross section of at least this value in absence of single top events is $(3.2 \pm 2.3) \times 10^{-8}$, corresponding to a 5.4σ fluctuation from zero. Following the convention in the field that a significance greater than 5σ is regarded as discovery/observation, the Bayesian neural network analysis could claim observation of single top production. (The observation result published by the DØ collaboration [14] corresponds to the combination of this Bayesian neural network analysis with other two analyses methods done in parallel −boosted desicion tress and matrix element calculations−, which gives a 5.03σ significance.)

The same Bayesian technique is used to measure, on the same dataset and based also on the Bayesian neural network discriminant distributions, the strength of the Wtb coupling. This measurement makes no assumption on the number of generations nor the unitarity of the CKM matrix. The result is:

$$|V_{tb} f_1^L| = 1.14^{+0.15}_{-0.13}, \tag{9.1}$$

By further imposing the conditions $f_1^L = 1$ and $|V_{tb}| < 1$, as predicted by the Standard Model, a 95% confidence level lower limit is set on $|V_{tb}|$:

$$|V_{tb}| > 0.82.$$

A Studies on the Instrumental Background in Electron+Jets+$\not\!\!E_T$ Final State Selections

In this appendix, studies performed to improve the understanding and modeling of the instrumental background in lepton+jets+$\not\!\!E_T$ analyses at DØ are presented. The studies were done based on p17 Monte Carlo samples, and although focused on the case of the lepton being an electron, some of the studies (specifically the ones presented in Sect. A.5) are also applicable to muons.

QCD events are a background to any analysis that involves a lepton+jets+$\not\!\!E_T$ final state. In the case of electrons, this happens when one of the jets is misreconstructed as an electron and at the same time the energy of a jet (or jets) is mismeasured resulting in fake missing transverse energy. For electrons, photon+jets (QED) events are also a background, because of a misassignment of a track to the electromagnetic cluster produced by the photon. The studies presented in here involve an electron+jets+$\not\!\!E_T$ final state selection and therefore both kinds of events are considered, QCD and γ+jets. QCD events will also be sometimes referred to as "multijets". In the everyday nomenclature in Tevatron physics analyses however, it is very common to ignore the γ+jets contribution and therefore refer to the composite background just as the multijets or the QCD background. To avoid confusion, we will adopt a less common, but probably more precise terminology, and will call them the "instrumental background" emphasizing the fact that the real object that was reconstructed as an electron is not really an electron, but was reconstructed as such, because of instrumental limitations (either at the detector or at the algorithm levels).

W+jets production is typically a major physics contribution to a lepton+jets+$\not\!\!E_T$ event selection. The leptonic decay of the W boson gives origin not only to the reconstructed lepton, but also to substantial missing transverse energy accounting for the undetected neutrino. In QCD and γ+jets background events the high value of missing transverse energy comes from the error in the measurement of the energy of the jets (jet energy resolution).

Since it is very difficult to simulate the instrumental background, its model is typically derived from data. The shape of the instrumental background is usually modeled using a data sample which is orthogonal to the data sample under analysis. Then, the normalization of this background is typically derived from a set of two equations known as the "matrix method" equations [82], from where the normalization of the W+jets Monte Carlo sample is also obtained. All the studies shown in this appendix are directly connected to the matrix method and the parameters involved. So, the next section is dedicated to describe the matrix method and to motivate the studies presented here.

A.1 The Matrix Method

Section A.1.1 is devoted to explain the traditional use of the matrix method as a tool to determine the relative contribution to the analysis data sample of the instrumental background and one of the physics backgrounds (for us W+jets). In Sect. A.1.2, a formula to correct the bias that a "loose-but-not-tight" data sample introduces when modeling the instrumental background is derived from the matrix method. This correction allows a better estimation, both in shape and normalization, of the real contribution of this background in the data.

A.1.1 The Matrix Method as a Tool to get the Normalization of Backgrounds

The matrix method consists of defining two samples of events, a *loose* and a *tight* set, the latter being a subset of the first, where the same event selection has been applied, but using a tighter lepton identification (ID) criteria. In general, the *tight* sample is the final selected sample. Moreover, the matrix method assumes that there is only one lepton in both samples that satisfies the loose lepton ID. Using the fact that in each of these two samples there are events with a real lepton and events with a fake lepton, one can write:

$$N_{\text{loose}} = N_{\text{loose}}^{\text{fake}} + N_{\text{loose}}^{\text{real}}, \tag{A.1}$$

$$N_{\text{tight}} = N_{\text{tight}}^{\text{fake}} + N_{\text{tight}}^{\text{real}}, \tag{A.2}$$

where N means number of events (or more generally event yield), the subscripts "loose" and "tight" refer to the *loose* and *tight* data samples respectively, and the labels "fake" and "real" refer to the leptonic identity of the reconstructed lepton. Thus, N_{loose} and N_{tight} are the number of events in the *loose* and *tight* data samples respectively. $N_{\text{loose}}^{\text{fake}}$ and $N_{\text{tight}}^{\text{fake}}$ represent the number of events in the *loose* and *tight* data samples in which the reconstructed lepton is not really a lepton (lets call them the "*loose-with-fake-lepton*" and "*tight-with-fake-lepton*" samples). These are the instrumental background events. Similarly, $N_{\text{loose}}^{\text{real}}$ and $N_{\text{tight}}^{\text{real}}$ represent the number of events in the *loose* and *tight* data samples in which the reconstructed lepton corresponds to a real lepton (lets call them the "*loose-with-real-lepton*" and "*tight-with-real-lepton*" samples), and they include physics background events plus signal events. These last four quantities are all unknown, except for the signal and all but one of the $n_{\text{bkg,phys}}$ components of the physics background which typically have a small cross section (e.g., Z+jets, $t\bar{t}$, diboson, single top) and can be estimated from MC if the cross sections are known with enough precision.

The set of Eqs. (A.1) and (A.2) can be rewritten as

$$N_{\text{loose}} = N_{\text{loose}}^{\text{fake}} + N_{\text{loose}}^{\text{real}}, \tag{A.3}$$

$$N_{\rm tight} = \epsilon_{\rm fake}\, N^{\rm fake}_{\rm loose} + \epsilon_{\rm real}\, N^{\rm real}_{\rm loose}, \tag{A.4}$$

where two new quantities have been introduced, $\epsilon_{\rm fake}$ and $\epsilon_{\rm real}$, which are defined as the ratios of the number of instrumental background and physics background plus signal events in the *tight* sample to the corresponding number of events in the *loose* sample:

$$\epsilon_{\rm fake} = \frac{N^{\rm fake}_{\rm tight}}{N^{\rm fake}_{\rm loose}}, \tag{A.5}$$

$$\epsilon_{\rm real} = \frac{N^{\rm real}_{\rm tight}}{N^{\rm real}_{\rm loose}}. \tag{A.6}$$

They represent therefore the probability for each of these two types of events to pass the tighter lepton ID given that they passed the looser lepton ID (and, of course, all the other cuts involved in the analysis). From the physical point of view, they represent the efficiency for a fake and a real lepton that passed the loose lepton ID to also pass the tight lepton ID. Since the instrumental background and/or the physics background may contain more than one component (as already explained it happens for the instrumental background in electron+jets final state analyses, where there are multijets and γ+jets components), $\epsilon_{\rm fake}$ and $\epsilon_{\rm real}$ are, in the general case, averages of sample specific efficiencies. The reader should be also warn about that Eqs. (A.5) and (A.6) are the mathematical definitions of the efficiencies needed to connect (A.4) to (A.2), but they do not refer to how these efficiencies are measured, something that is explained later. This is understandable, since one can not isolate the fake lepton events in the analysis data sample (otherwise there would be no instrumental background).

Knowing $N_{\rm loose}$ and $N_{\rm tight}$, and having measured $\epsilon_{\rm fake}$ and $\epsilon_{\rm real}$, one can solve the system of Eqs. (A.3) and (A.4) to get $N^{\rm fake}_{\rm loose}$ and $N^{\rm real}_{\rm loose}$. Multiplying then these quantities by $\epsilon_{\rm fake}$ and $\epsilon_{\rm real}$ respectively, one obtains $N^{\rm fake}_{\rm tight}$ and $N^{\rm real}_{\rm tight}$. The solutions are:

$$N^{\rm fake}_{\rm tight} = \epsilon_{\rm fake}\, \frac{(N_{\rm loose}\epsilon_{\rm real} - N_{\rm tight})}{(\epsilon_{\rm real} - \epsilon_{\rm fake})}, \tag{A.7}$$

$$N^{\rm real}_{\rm tight} = \epsilon_{\rm real}\, \frac{(N_{\rm loose}\epsilon_{\rm fake} - N_{\rm tight})}{(\epsilon_{\rm fake} - \epsilon_{\rm real})}. \tag{A.8}$$

Equation (A.7) gives the number of instrumental background events in the *tight* sample (in the data), while Eq. (A.8) gives the number of physics background plus signal events. Subtracting from $N^{\rm real}_{\rm tight}$ the signal yield and the yield of the $n_{\rm bkg,phys} - 1$ known physics backgrounds, one finally gets the normalization for the unknown physics background, e.g. W+jets:

$$N^{W+\rm jets}_{\rm tight} = N^{\rm real}_{\rm tight} - N^{\rm sig}_{\rm tight} - \sum_{i=1}^{n_{\rm bkg,phys}-1} N^{{\rm bkg,phys}_i}_{\rm tight}. \tag{A.9}$$

Another alternative is to first subtract from $N_{\rm loose}$ and $N_{\rm tight}$ the signal yield and the yield of the $n_{\rm bkg,phys}-1$ known physics backgrounds and then solve for the instrumental background and W+jets yields:

$$N_{\rm tight}^{\rm fake} = \epsilon_{\rm fake}\,\frac{(N'_{\rm loose}\epsilon_{\rm real}-N'_{\rm tight})}{(\epsilon_{\rm real}-\epsilon_{\rm fake})}, \tag{A.10}$$

$$N_{\rm tight}^{W+{\rm jets}} = \epsilon_{\rm real}\,\frac{(N'_{\rm loose}\epsilon_{\rm fake}-N'_{\rm tight})}{(\epsilon_{\rm fake}-\epsilon_{\rm real})}, \tag{A.11}$$

where

$$N'_{\rm loose} = N_{\rm loose}-N_{\rm loose}^{\rm sig}-\sum_{i=1}^{n_{\rm bkg,phys}-1} N_{\rm loose}^{{\rm bkg,phys}_i}, \tag{A.12}$$

$$N'_{\rm tight} = N_{\rm tight}-N_{\rm tight}^{\rm sig}-\sum_{i=1}^{n_{\rm bkg,phys}-1} N_{\rm tight}^{{\rm bkg,phys}_i}. \tag{A.13}$$

In this case, the real lepton efficiency is mathematically defined as

$$\epsilon_{\rm real} = \frac{N_{\rm tight}^{W+{\rm jets}}}{N_{\rm loose}^{W+{\rm jets}}}. \tag{A.14}$$

So, this second alternative would be more useful if the real lepton efficiency has some sample dependence.

If the efficiencies $\epsilon_{\rm fake}$ and $\epsilon_{\rm real}$ are just constant numbers, then the set of Eqs. (A.7) and (A.8) just needs to be solved once for the entire sample. However, the efficiencies may depend on kinematic variables, like for example the transverse momentum of the reconstructed lepton ($p_T^{\rm lepton}$). In that case, it is better to smartly choose a binning in $p_T^{\rm lepton}$ and solve one set of Eqs. (A.7) and (A.8) per bin of $p_T^{\rm lepton}$. Another possibility would be to use Eqs. (A.7) and (A.8) in an event by event basis, weighting the events from the *loose* and the *tight* samples by an eventweight. This would give:

$$N_{\rm tight}^{\rm fake} = \sum_{i=1}^{N_{\rm loose}} w_{{\rm l},i}^{\rm fake}+\sum_{i=1}^{N_{\rm tight}} w_{{\rm t},i}^{\rm fake}, \tag{A.15}$$

$$N_{\rm tight}^{\rm real} = \sum_{i=1}^{N_{\rm loose}} w_{{\rm l},i}^{\rm real}+\sum_{i=1}^{N_{\rm tight}} w_{{\rm t},i}^{\rm real}, \tag{A.16}$$

with $\sum_{i=1}^{N_{\rm loose}}$ and $\sum_{i=1}^{N_{\rm tight}}$ being sums over the events of the *loose* and *tight* samples

respectively, and

$$w_{\rm l}^{\rm fake} = \frac{\epsilon_{\rm fake}\ \epsilon_{\rm real}}{(\epsilon_{\rm real} - \epsilon_{\rm fake})}, \quad \text{(A.17)}$$

$$w_{\rm t}^{\rm fake} = \frac{\epsilon_{\rm fake}}{(\epsilon_{\rm fake} - \epsilon_{\rm real})}, \quad \text{(A.18)}$$

$$w_{\rm l}^{\rm real} = \frac{\epsilon_{\rm real}\ \epsilon_{\rm fake}}{(\epsilon_{\rm fake} - \epsilon_{\rm real})}, \quad \text{(A.19)}$$

$$w_{\rm t}^{\rm real} = \frac{\epsilon_{\rm real}}{(\epsilon_{\rm real} - \epsilon_{\rm fake})}, \quad \text{(A.20)}$$

where $\epsilon_{\rm fake}$ and $\epsilon_{\rm real}$, and thus the eventweights $w_{\rm l}^{\rm fake}$, $w_{\rm t}^{\rm fake}$, $w_{\rm l}^{\rm real}$ and $w_{\rm t}^{\rm real}$, are functions of, let's say again, $p_T^{\rm lepton}$, and should be therefore evaluated at the $p_T^{\rm lepton}$ of the event i. This option of using eventweights seems to be, from the logical point of view, the most accurate to estimate $N_{\rm tight}^{\rm fake}$ and $N_{\rm tight}^{\rm real}$. The problems arise when measuring the fake rate, because it can be different in the sample where it is measured than in the analysis data sample. In this case there is no guarantee that Eqs. (A.15) and (A.16) work better than Eqs. (A.7) and (A.8).

As any quantity related to the instrumental background, the fake rate can not be measured in Monte Carlo due to the difficulty of having a good simulation of this background. Thus, it is usually measured in data in a sample largely dominated by instrumental background events, obtained by applying the same selection cuts as in the analysis, but reverting the missing transverse energy cut, e.g. $\not\!\!E_T < 10$ GeV. Lets call this sample the low-$\not\!\!E_T$ data sample. With this last cut the W+jets events are extremely reduced. Z+jets events are not, but their cross section is at least one order of magnitude smaller than the W+jets cross section. In spite of that, these events are still present in a small fraction and constitute a contamination to the low-$\not\!\!E_T$ data sample that should be subtracted before measuring the fake rate. It is usually argued that the fake rate should be independent of $\not\!\!E_T$ and therefore the same in the low-$\not\!\!E_T$ and in the analysis data samples[36]. But, as will be shown in Sect. A.3, the fake rate is sample dependent (it is different for multijets than for γ+jets). So, even if the individual γ+jets and QCD fake rates are both independent of $\not\!\!E_T$, if these two background contributions are not similar in the low-$\not\!\!E_T$ and in the analysis data samples, a bias may be present. Section A.4 is dedicated to show how the composition of the instrumental background may be kept the same in the two $\not\!\!E_T$ regions to assure same fake rate in these two samples.

The real lepton efficiency is usually measured in W+jets MC events applying the same event selection as in the analysis, and matching the reconstructed lepton to the particle level lepton coming from the W boson decay. Then, it is corrected with a data to MC scale factor measured in Z+jets.

[36]Indeed, it has been shown in many opportunities that the fake rate has no dependence with $\not\!\!E_T$ at least in the range 0-10 GeV.

A.1.2 Using the Matrix Method to improve the Instrumental Background Model

Section A.1.1 explained the simplest and most common use of the matrix method, i.e. how to use it to normalize backgrounds. The present section shows how to use the matrix method equations to extract a correction to the shape of the instrumental background model. The key point is that the shape of the instrumental background is obtained from data from a sample that is orthogonal to the analysis data sample. One usually defines a "*loose-minus-tight*" sample as those events that pass the same set of selection cuts as in the analysis, except that the reconstructed lepton is required to pass a given (loose) lepton ID of looser quality than the one used in the analysis, but fail the (tight) lepton ID from the analysis. This sample is expected to be dominated by events with fake leptons, as a real lepton should pass the tight criteria with a high probability. This sample is also in principle the most suitable one to model the shape of the instrumental background in the analysis data sample, as it has the same event selection and only differs in the lepton ID criteria.

The number of events in the *loose-minus-tight* sample ($N_{\text{loose}-\text{tight}}$) is given by $N_{\text{loose}} - N_{\text{tight}}$, which, using the matrix method Eqs. (A.3) and (A.4), can be written as

$$N_{\text{loose}-\text{tight}} = \frac{(1-\epsilon_{\text{fake}})}{\epsilon_{\text{fake}}} N_{\text{tight}}^{\text{fake}} + \frac{(1-\epsilon_{\text{real}})}{\epsilon_{\text{real}}} N_{\text{tight}}^{\text{real}}. \tag{A.21}$$

Solving (A.21) for $N_{\text{tight}}^{\text{fake}}$ one gets

$$N_{\text{tight}}^{\text{fake}} = \frac{\epsilon_{\text{fake}}}{(1-\epsilon_{\text{fake}})} N_{\text{loose}-\text{tight}} - \frac{\epsilon_{\text{fake}}}{(1-\epsilon_{\text{fake}})} \frac{(1-\epsilon_{\text{real}})}{\epsilon_{\text{real}}} N_{\text{tight}}^{\text{real}}. \tag{A.22}$$

Equation (A.22) shows the relationship between the *loose-minus-tight* sample and the sample to be modeled (the instrumental background contribution to the *tight* sample). If ϵ_{fake} and ϵ_{real} can not be taken as constant numbers, then one can think of dividing the phase space in bins where ϵ_{fake} and ϵ_{real} are fairly constant and write one equation like (A.22) for each of these bins. But then this means that the shape of the *loose-minus-tight* model does not represent the shape of the instrumental background contribution in the *tight* sample, because it has a kinematic bias given by $(1-\epsilon_{\text{fake}})/\epsilon_{\text{fake}}$ and it contains also a contamination of real lepton events.

One can go to the continuous limit (bins size going to zero) and rewrite Eq. (A.22) in an event by event basis form as done with Eqs. (A.7) and (A.8) when they were rewriten as (A.15) and (A.16):

$$N_{\text{tight}}^{\text{fake}} = \sum_{i=1}^{N_{\text{loose}-\text{tight}}} w_{\text{l}-\text{t},i}^{\text{kin.bias}} + \sum_{i=1}^{N_{\text{tight}}^{\text{real}}} w_{\text{t}_{\text{real}},i}^{\text{sig.cont.}}, \tag{A.23}$$

where $\sum_{i=1}^{N_{\rm loose-tight}}$ and $\sum_{i=1}^{N_{\rm tight}^{\rm real}}$ are sums over the events of the *loose-minus-tight* and the *tight-with-real-lepton* (physics background plus signal) samples respectively, and

$$w_{\rm l-t}^{\rm kin.bias} = \frac{\epsilon_{\rm fake}}{(1-\epsilon_{\rm fake})}, \tag{A.24}$$

$$w_{\rm t_{real}}^{\rm sig.cont.} = \frac{-\epsilon_{\rm fake}}{(1-\epsilon_{\rm fake})}\,\frac{(1-\epsilon_{\rm real})}{\epsilon_{\rm real}}, \tag{A.25}$$

are the eventweights to be applied to the *loose-minus-tight* and the *tight-with-real-lepton* samples respectively.

Equation (A.23) is not only saying how to get the number of events of the instrumental background in the *tight* sample. Because it is formulated in an event by event basis, it is also saying that the instrumental background phase space distribution is given by the *loose-minus-tight* distribution reshaped by $w_{\rm l-t}^{\rm kin.bias}$ minus the *tight-with-real-lepton* distribution reshaped by $|w_{\rm t_{real}}^{\rm sig.cont.}|$ (notice that $w_{\rm t_{real}}^{\rm sig.cont.}$ is negative). This second term in Eq. (A.23) can also be put together with the *tight-with-real-lepton* contribution to the *tight* sample, and in this way one keeps separate the two main samples with which one works in the analysis. The following equation summarizes this:

$$\begin{aligned} N_{\rm tight} &= N_{\rm tight}^{\rm fake} + N_{\rm tight}^{\rm real}, \\ &= \sum_{i=1}^{N_{\rm loose-tight}} w_{\rm l-t,i}^{\rm kin.bias} + \sum_{i=1}^{N_{\rm tight}^{\rm real}} (1 - |w_{{\rm t_{real}},i}^{\rm sig.cont.}|). \end{aligned} \tag{A.26}$$

A quick estimation of the magnitude of the weights $w_{\rm l-t}^{\rm kin.bias}$ and $w_{\rm t_{real}}^{\rm sig.cont.}$ can be done. Typical values of the efficiencies are ~ 0.1 for the fake rate and ~ 0.9 for the real lepton efficiency. Then $w_{\rm l-t}^{\rm kin.bias} \approx \frac{1}{9}$ and $|w_{\rm t_{real}}^{\rm sig.cont.}| \approx (w_{\rm l-t}^{\rm kin.bias})^2 \approx \frac{1}{81}$. So, the signal contamination to the *loose-minus-tight* sample would be a second order effect and it could be neglected, simplifying things a lot. Equation (A.26) would then become

$$N_{\rm tight} = \sum_{i=1}^{N_{\rm loose-tight}} w_{{\rm l-t},i}^{\rm kin.bias} + N_{\rm tight}^{\rm real}, \tag{A.27}$$

and only the *loose-minus-tight* sample would need to be reshaped. Notice that in this case the first term in (A.27) gives both the shape and the normalization of the instrumental background contribution in the data.

A.2 Monte Carlo Samples

All MC samples used for these studies correspond to the p17 production. To simulate the γ+jets and the multijets instrumental backgrounds, the photon+jet and the dijet (γ-like) MC samples introduced in Sect. 4.3.2 are respectively used. The dijet (γ-like)

sample was derived applying a set of preselection cuts at the particle level to enrich its content in jets faking electromagnetic objects.

Also, three p17 ALPGEN MC samples are used to simulate events with a real lepton: W+jets, Z+jets and top-pair, where the final lepton is always an electron. The p17 W+jets MC samples correspond to the SAM definitions TOP_w$n$lp_qMWPtW_v1 (for w+$n$lp), TOP_w2c$n$lp_HFskim_v2 (for w+2c+$n$lp) and TOP_w2b$n$lp_HFskim_v2 (for w+2b+$n$lp), while the p17 Z+jets and $t\bar{t}$ MC samples are given in list files in the `caf_mc_util (p21-br-45)` package.

All the samples are normalized to an integrated luminosity of ~ 1 fb^{-1}.

A.3 Fake Rate

The fake rate represents the efficiency for a fake lepton reconstructed with a given loose lepton ID to also satisfy the tight lepton ID used in the analysis. It is therefore a relative fake rate, because it depends on how much tighter is the tight lepton ID with respect to the loose one. It is also important to remember that for the matrix method to make sense and for ϵ_{fake} (and of course also ϵ_{real}) to be meaningful, the loose lepton ID must define a set of events that is a superset of the events selected with the tight lepton ID. Thus, the fake rate can be defined and measured as a ratio of number of selected events, in the numerator using the tight lepton ID and in the denominator using the loose lepton ID, in a sample dominated by fake leptons.

Given that it is very difficult to achieve a good simulation of the instrumental background, the fake rate is measured in data. As already said in Sect. A.1.1, the data sample used to measure the fake rate in lepton+jets+$\not\!\!E_T$ analyses is a sample constructed by applying the same event selection as in the analysis, but with $\not\!\!E_T < 10$ GeV. The aim of this reversed $\not\!\!E_T$ cut is to reduce as much as possible the W+jets events contribution, the main real lepton background, where the lepton is produced in association with a neutrino. This data sample is dominated by instrumental background events with some little amount of events with a real lepton (signal-like events). These small amount of events is called a "signal contamination".

In the case of the lepton being an electron, there are two kind of instrumental backgrounds, multijets and γ+jets. In γ+jets events, what fakes an electron is of course the photon, while in multijets it is a jet that fluctuates to a leading π^0 (with $\pi^0 \to \gamma\gamma$). Therefore, these two different backgrounds do not need in principle to have the same fake rate for the same set of selection cuts. The multijets and γ+jets Monte Carlo samples are used to look at their fake rates as a function of different kinematic variables with very basic event selection given by:

- at least one electron (or positron) with $p_T > 15$ GeV/c and $|\eta_{\text{det}}| < 1.1$,

- at least one jet with $p_T > 15$ GeV/c and $|\eta_{\rm det}| < 2.5$,
- $\not\!\!E_T < 10$ GeV.

The lepton IDs used here are top_loose(v3) and top_tight(v3), which are the same as the loose and tight isolated electron definitions introduced in Sect. 3.4.2.

Figure A.1 shows the dependence of the fake rate with the transverse momentum of the reconstructed electron (p_T^e). The first thing one can see is that the fake rate for γ+jets events is much higher than for multijets. This is in principle expected from the fact that it should be much easier for a photon coming from a direct γ+jet production to fake an electron than for a photon coming from a π^0 decay, since in the latter case the photon will be usually surrounded by hadronic activity resulting from the fragmentation of the original parton. The plot also shows that the fake rate in γ+jets events increases with p_T^e, while in QCD events it is approximately constant. One may be tempted to say that the p_T^e dependence of the combined fake rate is just a consequence of the p_T^e dependence of the γ+jets fake rate. But this may not be completely true if the relative contribution of both processes changes with p_T^e, in which case the combined fake rate will still have a dependence with p_T^e even if the γ+jets fake rate is flat (but different than for multijets).

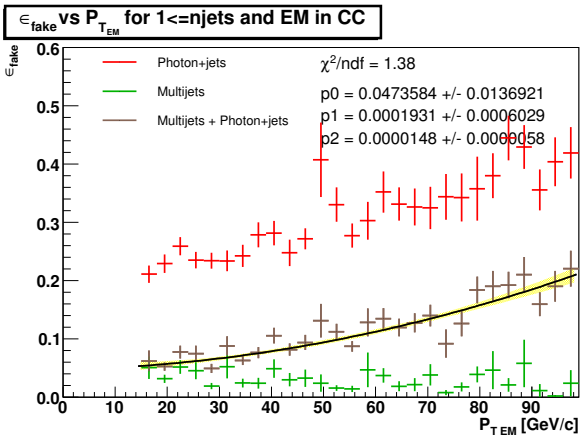

Figure A.1 : Fake rate as a function of the p_T of the reconstructed electron for the multijets (green) and γ+jets (red) Monte Carlo samples and for both samples together (brown) when applying a very basic event selection described in the text above. A quadratic fit (p0 + p1·p_T^e + p2·$(p_T^e)^2$) was done on the combined fake rate just for illustration.

The dependence seen for other fundamental kinematic variables is consistent with constant, so only the dependence on p_T^e is considered in the studies presented here.

A.4 The Triangle Cut

To reduce the instrumental background, a cut in the azimuthal angle (ϕ) between the reconstructed lepton (electron) and the missing transverse energy of the event ($\not\!\!E_T$) versus $\not\!\!E_T$ is often applied. This is a cut in a 2D plane. The excluded region has the form of a triangle, thus this cut is often called the "triangle cut" or "delta phi cut".

Since Run I, the triangle cut was always tuned (sometimes by eye) to exclude the region where the signal-over-background acceptance is minimal. But there is another important role the triangle cut seems to play in electron+jets+$\not\!\!E_T$ analysis. As said in Sect. A.1.1, the fake rate is measured in a low-$\not\!\!E_T$ region ($\not\!\!E_T < 10$ GeV). But then it is used in the matrix method equations where all the other quantities belong to a high-$\not\!\!E_T$ region. It was always assumed that the fake rate is independent of $\not\!\!E_T$, validating this statement by the fact that it is flat in the $0-10$ GeV range of $\not\!\!E_T$. On the other hand, the fake rate uncertainty does not contribute significantly to the uncertainties in $N_{\text{tight}}^{\text{fake}}$ and $N_{\text{tight}}^{\text{real}}$, so that the error in the fake rate could even be inflated to cover a possible discrepancy between both $\not\!\!E_T$ regions.

As shown in Fig. A.1, the fake rate for γ+jets events is different (higher) than for multijets events. Therefore, when measured in data, the fake rate is a weighted average between these two contributions (plus some small contamination from real lepton events, which here are neglected). If the fake rate for each of these two backgrounds is really independent of $\not\!\!E_T$, then keeping the same composition of the instrumental background in the low and high-$\not\!\!E_T$ regions ensures that ϵ_{fake} is the same in these two regions. The studies presented here suggest that, a reasonably well chosen triangle cut, does exactly this job. Table A.1 shows the percent fraction of γ+jets events in a sum of γ+jets and multijets events when using the following selection cuts, which correspond to a $t\bar{t}$ analysis, and including or not the $\Delta\phi(e,\not\!\!E_T)$ cut:

- at least one electron (or positron) with $p_T > 15$ GeV/c and $|\eta_{\text{det}}| < 1.1$,
- second electron veto with electron ID = top_tight(v3), $p_T > 15$ GeV/c and $|\eta_{\text{det}}| < 1.1$,
- at least one jet with $p_T > 20$ GeV/c and $|\eta_{\text{det}}| < 2.5$,
- leading jet must have $p_T > 40$ GeV/c,
- *optional* : $\Delta\phi(e,\not\!\!E_T) > 0.7\pi - 0.045\not\!\!E_T/\text{GeV}$ (triangle cut).

The fractions shown are for $\not\!\!E_T < 10$ GeV, region where the fake rate is usually measured, and $\not\!\!E_T > 20$ GeV, the analysis region. From Table A.1, one can see that when the triangle cut is applied the γ+jets fractions become the same in both $\not\!\!E_T$ regions, while when the triangle cut is not applied the γ+jets fractions are different. The fractions presented in Table A.1 are for the loose lepton ID (top_loose(v3)). In

Sect. A.5 it is shown that when the triangle cut is applied the fake becomes the same in both $\not\!\!E_T$ regions.

Fractional Composition of the Instrumental Background

Set of selection cuts	Photon+jets fraction [%] at $\not\!\!E_T < 10$ GeV	Photon+jets fraction [%] at $\not\!\!E_T > 20$ GeV
without the triangle cut	16	10
with the triangle cut	17	17

Table A.1 : Fraction of γ+jets events in a sum of γ+jets plus multijets events when applying the $t\bar{t}$ event selection described in the text and applying or not the triangle cut. The fractions shown correspond to the regions $\not\!\!E_T < 10$ GeV (where the fake rate is measured) and $\not\!\!E_T > 20$ GeV (where the analysis is performed).

A.5 Effect of the Triangle Cut on the Fake Rate

In Sect. A.4 it has been shown that, at least for the event selection from the $t\bar{t}$ analysis [111] and with electron ID equal to top_loose(v3), the inclusion of the triangle cut as part of the selection cuts made the fraction of γ+jets events in the total instrumental background to become the same (17%) in the low- and high-$\not\!\!E_T$ regions ($\not\!\!E_T <$ 10 GeV and $\not\!\!E_T > 20$ GeV). When the triangle cut is not applied, the γ+jets fraction in the low-$\not\!\!E_T$ region is 16% and in the high-$\not\!\!E_T$ region 10%. Here, the fake rates are compared for this same event selection. Fig. A.2 shows the fake rate as a function of the transverse momentum of the reconstructed electron in the low- and high-$\not\!\!E_T$ regions. The left plot corresponds to the case in which the triangle cut is not applied, where it is seen that the fake rate measured for $\not\!\!E_T < 10$ GeV is approximately 30% higher than the fake rate for $\not\!\!E_T > 20$ GeV. This is consistent with a higher γ+jets contribution in $\not\!\!E_T < 10$ GeV, since γ+jets has a higher fake rate than multijets (see Fig. A.1). On the other hand, when the triangle cut is applied (right plot in Fig. A.2), both fake rates become the same within statistics, consistent with similar γ+jets fractions in both $\not\!\!E_T$ regions.

A.6 Reshaping of the loose-minus-tight Model

This section is dedicated to the second part of the studies, i.e. to the correction of the *loose-minus-tight* model of the instrumental background. As we said in Sect. A.1.2, a *loose-minus-tight* sample is typically used to model the shape of the instrumental background that one has in the selected *tight* sample. But this model has a kinematic bias for which one would like to correct, and a contamination of real lepton events

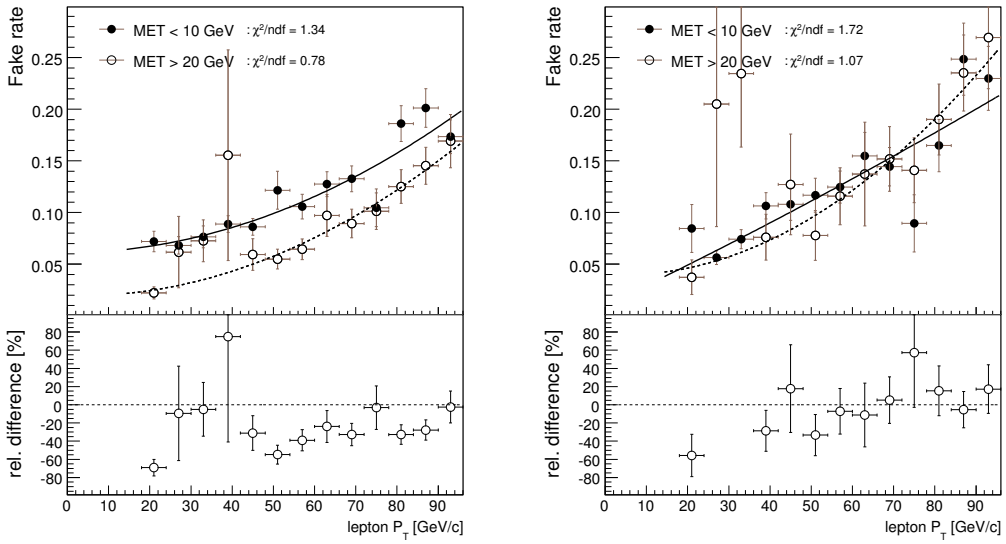

Figure A.2 : Left: Fake rate as a function of the p_T of the reconstructed electron in the low- and the high-$\not\!\!E_T$ regions, using the multijets and γ+jets Monte Carlo samples and applying the set of selection cuts listed in Sect. A.4 without the optional triangle cut. Right: Same plot as on the left, but adding the triangle cut on the 2D plane of $\Delta\phi(e,\not\!\!E_T)$ vs. $\not\!\!E_T$.

which one would like to subtract (see Eq. (A.22)). Since the signal contamination is a second order effect, we will first concentrate on the kinematic bias. To correct for the kinematic bias one needs to apply the weight given in Eq. (A.24) to each event in the *loose-minus-tight* sample. After this reweighting, the *loose-minus-tight* sample should better model the instrumental background one really has in the selected *tight* sample.

As a start point, the following electron+jets+$\not\!\!E_T$ event selection is applied to all the Monte Carlo samples used in these studies:

- only one top_tight(v3) electron (or positron) with $p_T > 15$ GeV/c and $|\eta_{\rm det}| < 1.1$,
- at least one jet with $p_T > 15$ GeV/c and $|\eta_{\rm det}| < 3.4$,
- leading jet must have $p_T > 25$ GeV/c and $|\eta_{\rm det}| < 2.5$,
- $\not\!\!E_T > 15$ GeV,
- veto muon,
- $\Delta\phi(e,\not\!\!E_T) > 0.595 - 0.01\not\!\!E_T/\text{GeV}$ (triangle cut).

The selected events represent the main contributions to the *tight* sample of this emulated electron+jets+$\not\!\!E_T$ analysis. In particular, the selected events from the γ+jet plus multijets MC samples represent the real contribution of the instrumental background to the *tight* sample. Its distributions will be compared against the ones of the *loose-minus-tight* model. This model is built by applying the same event selection as listed above, except for the electron ID (which is choosen to be top_loose(v3) plus the additional requirement $\mathcal{L}_{EM} < 0.85$ to exclude the electrons that satisfy the top_tight(v3) electron ID criteria), to all the MC samples (because in a real analysis the *loose-minus-tight* model is derived from data).

To get the correction weight from Eq. (A.24), the fake rate is measured using only the γ+jets plus multijets MC samples and in the region of $\not\!\!E_T < 10$ GeV. W+jets and Z+jets samples are not included in the fake rate measurement, since it is assumed that these contributions can be subtracted when measuring the fake rate in a real analysis in data, or can be neglected. The left plot in Fig. A.3 shows the measured fake rate as a function of the transverse momentum of the reconstructed electron, including a quadratic fit to the points $(p0 + p1p_T^{\text{lepton}} + p2(p_T^{\text{lepton}})^2)$.

Figure A.3 also shows in the right plot the real lepton efficiency as a function of the transverse momentum of the reconstructed electron. This efficiency is measured using the W+jets plus Z+jets MC samples. In W+jets, the reconstructed electron is matched with the particle level electron from the W boson decay using a cone of radius 0.3. A fit to the points is done with an exponential function $(p0 + p1\exp(p2p_T^{\text{lepton}}))$. The real lepton efficiency is needed in the weight given by Eq. (A.25) to subtract the real lepton contamination existing in the *loose-minus-tight* sample. Actually, as said in Sect. A.1.2, to avoid to define the instrumental background model by a mixture of the *loose-minus-tight* and the *tight-with-real-lepton* samples (as suggested by Eq. (A.23)), the idea is not to subtract the signal contamination from the *loose-minus-tight* sample itself, but from the *tight-with-real-lepton* sample by weighting the events in that sample by

$$1 - |w_{\text{t}_{\text{real}}}^{\text{sig.cont.}}| = \frac{\epsilon_{\text{real}} - \epsilon_{\text{fake}}}{(1 - \epsilon_{\text{fake}})\epsilon_{\text{real}}}, \tag{A.28}$$

as expressed in (A.26).

Figure A.4 shows the reweighting factors given in Eqs. (A.24) and (A.28) (left and right plots respectively), when using for the efficiencies the results of the fits shown in Fig. A.3. One can see that the signal contamination is a $1-3\%$ effect, while the kinematic bias is on average a 10% effect.

Figures A.5 to A.8 show a comparison, in different distributions, between the instrumental background contribution in the *tight* sample (shown as points) and the *loose-minus-tight* model of the instrumental background (shown as histograms in colors). On the left side are the comparisons before applying any reweighting, while on the right side are the comparisons after reweighting the *loose-minus-tight* sample by the

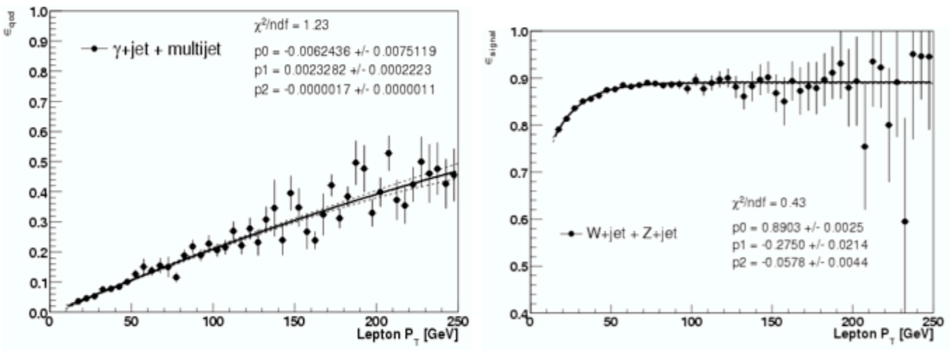

Figure A.3 : Left: Fake rate (labeled as ϵ_{qcd} in the figure) as a function of the electron transverse momentum as measured using the γ+jet and multijets MC samples in the region $\not\!\!E_T < 10$ GeV. Right: Signal efficiency (labeled as ϵ_{signal} in the figure) as a function of the electron transverse momentum as measured using the W+jet and Z+jet Monte Carlo samples. In W+jets, the reconstructed electron was matched with the particle level electron from the W boson decay in a cone of radius 0.3.

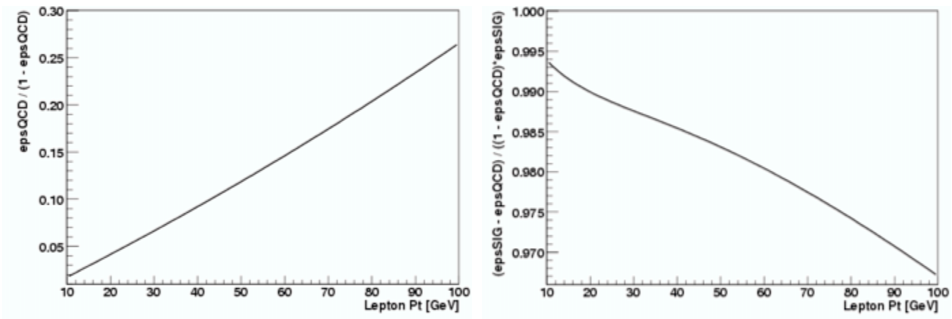

Figure A.4 : Left: Reweighting factor given by Eq. (A.24) calculated using for ϵ_{fake} (labeled as epsQCD in the figure) the result of the quadratic fit in the left plot of Fig. A.3. Right: Reweighting factor given by Eq. (A.28) calculated using for ϵ_{fake} and ϵ_{real} (labeled as epsSIG in the figure) the results of the quadratic and exponential fits in the plots of Fig. A.3.

left curve in Fig. A.4. The upper plots are in linear scale, while the lower plots are in logarithmic scale to allow a better view of the tail of the distributions. The distribution in Fig. A.5 corresponds to the electron p_T. This is the variable for which we see the biggest improvement, which is expected, since this is the variable against which the fake rate was measured. We also see some improvement in the W transverse mass and the leading jet p_T distributions (Figs. A.6 and A.7 respectively). But other distributions like $\not\!\!E_T$ (Fig. A.8) did not improve.

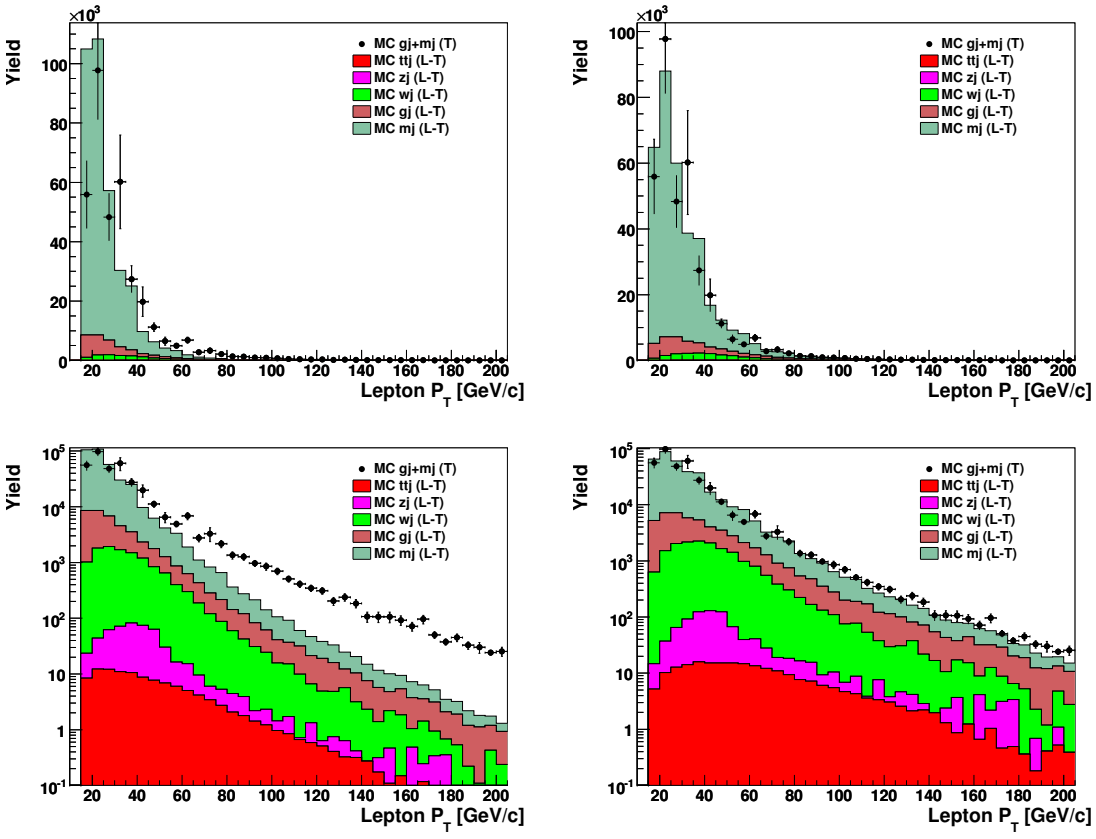

Figure A.5 : Comparison between the instrumental background contribution in the selected *tight* (T) sample (black points) and the *loose-minus-tight* (L-T) model of that background (colored histograms), using the Monte Carlo samples enumerated in Sect. A.2. The upper plots are in linear scale and the lower plots in logarithmic scale. Left: Comparison of the lepton transverse momentum distribution before reshaping the *loose-minus-tight* sample. Right: Same plots as on the left, but after reshaping the *loose-minus-tight* sample using the reweighting factor given by Eq. (A.24).

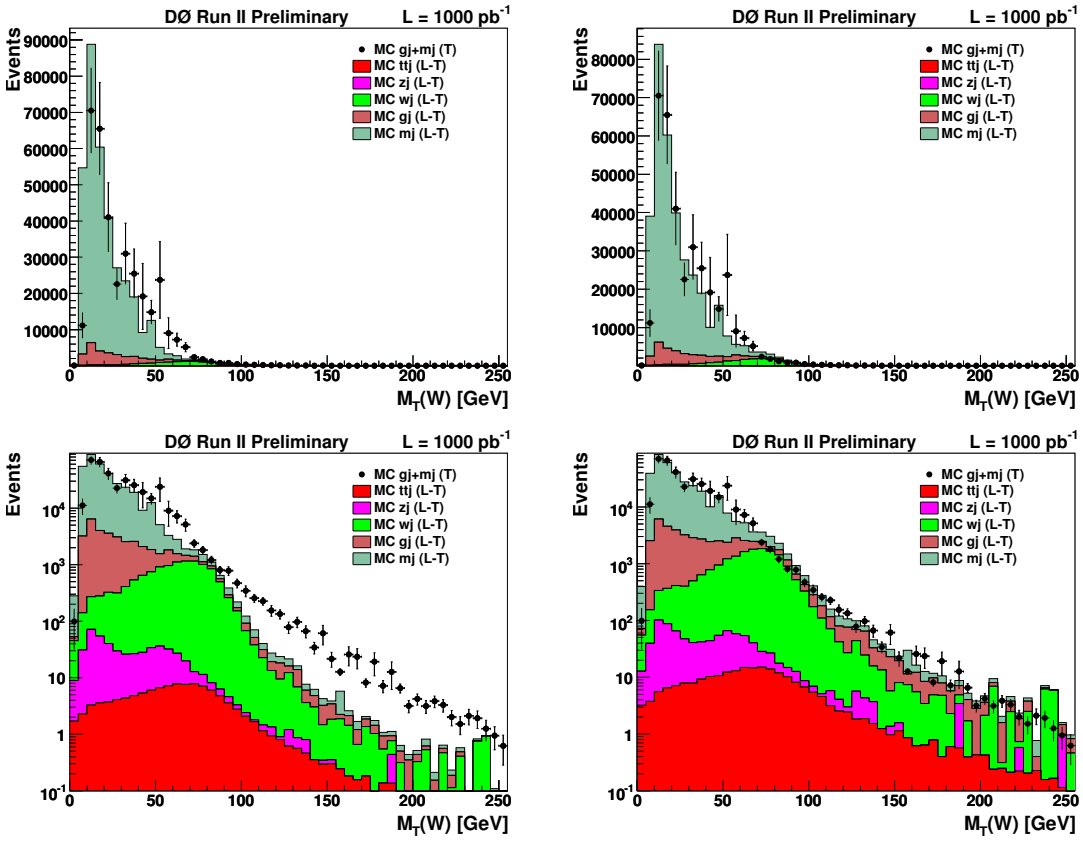

Figure A.6 : Comparison between the instrumental background contribution in the selected *tight* (T) sample (black points) and the *loose-minus-tight* (L-T) model of that background (colored histograms), using the Monte Carlo samples enumerated in Sect. A.2. The upper plots are in linear scale and the lower plots in logarithmic scale. Left: Comparison of the W boson transverse mass distribution before reshaping the *loose-minus-tight* sample. Right: Same plots as on the left, but after reshaping the *loose-minus-tight* sample using the reweighting factor given by Eq. (A.24).

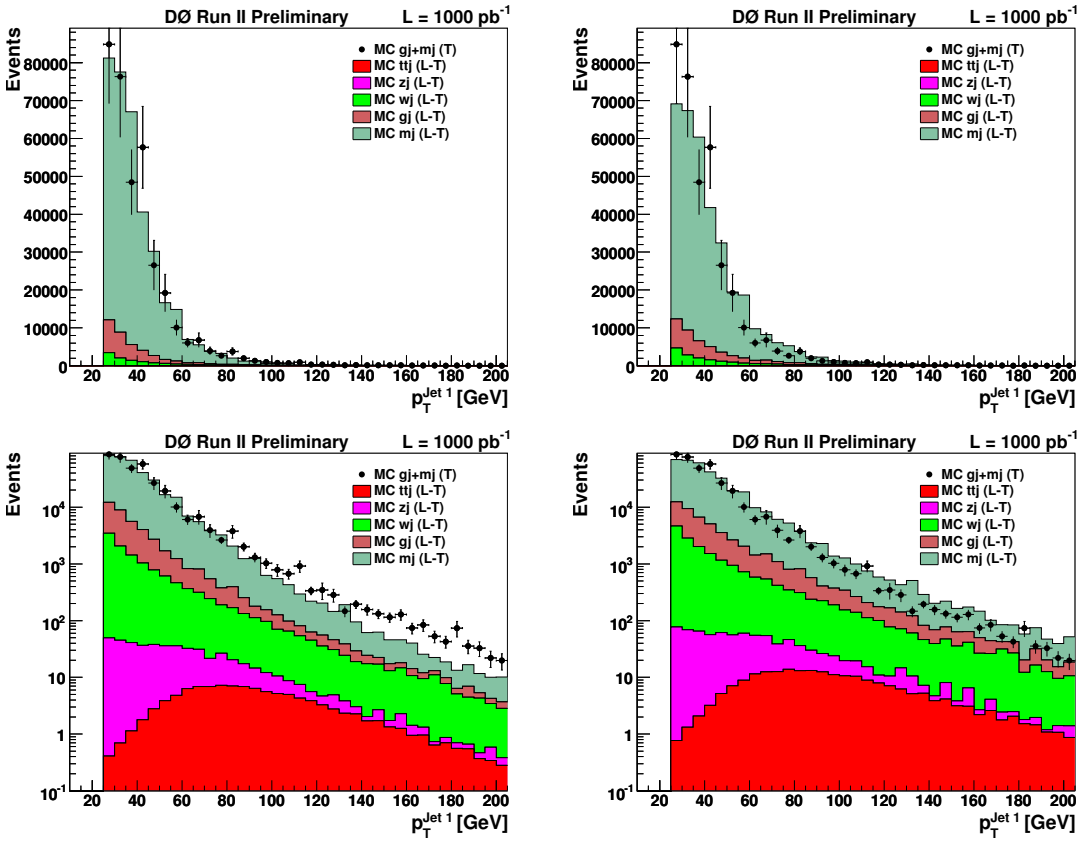

Figure A.7 : Comparison between the instrumental background contribution in the selected *tight* (T) sample (black points) and the *loose-minus-tight* (L-T) model of that background (colored histograms), using the Monte Carlo samples enumerated in Sect. A.2. The upper plots are in linear scale and the lower plots in logarithmic scale. Left: Comparison of the leading jet transverse momentum distribution before reshaping the *loose-minus-tight* sample. Right: Same plots as on the left, but after reshaping the *loose-minus-tight* sample using the reweighting factor given by Eq. (A.24).

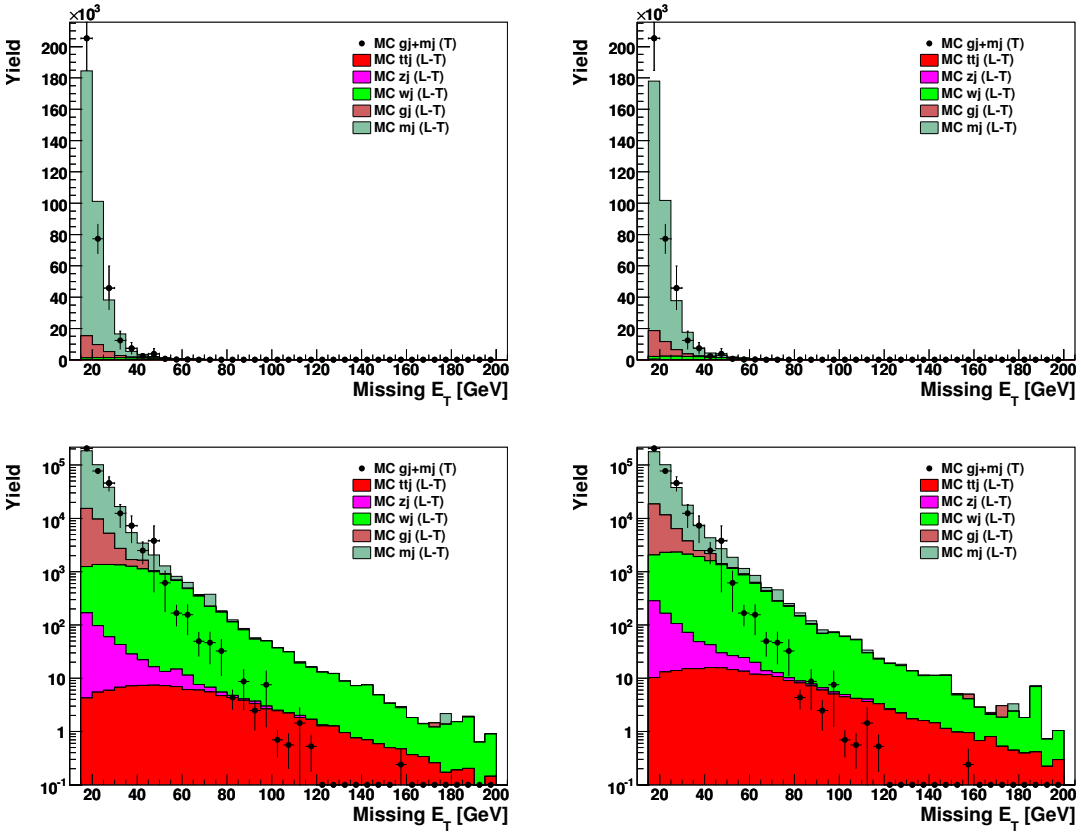

Figure A.8 : Comparison between the instrumental background contribution in the selected *tight* (T) sample (black points) and the *loose-minus-tight* (L-T) model of that background (colored histograms), using the Monte Carlo samples enumerated in Sect. A.2. The upper plots are in linear scale and the lower plots in logarithmic scale. Left: Comparison of the missing transverse energy distribution before reshaping the *loose-minus-tight* sample. Right: Same plots as on the left, but after reshaping the *loose-minus-tight* sample using the reweighting factor given by Eq. (A.24).

The other comparison one can do is between the selected *tight* sample and the sum of the background (plus signal) models. This comparisons are shown in Figs. A.9 to A.12, where the *tight* sample is shown in black points and consists of all the Monte Carlo samples (γ+jets, multijets, W+jets, Z+jets and $t\bar{t}$) since it represents the "data", and the models are shown in colored histograms. The W+jets, Z+jets and $t\bar{t}$ models are exactly the same as their contributions in the *tight* sample, at least on the left plots where no reshaping has been applied to them. On the right plots we applied the reweighting factor given by Eq. (A.28) to the W+jets, Z+jets and $t\bar{t}$ samples. For the instrumental background we again used the *loose-minus-tight* model from the previous set of figures (Figs. A.5 to A.8), on the left plots without the reshaping and on the right plots with the reshaping given by Eq. (A.24). Again, the distributions where one can see the biggest improvement in the "data"-model agreement are the electron p_T and the W transverse mass. This comes from the improvement in the *loose-minus-tight* sample as a model of the γ+jets plus multijets contribution in the "data", as showed in Figs. A.5 and A.6. But now, also the $\not\!\!E_T$ distribution had improved in the range $35 < \not\!\!E_T < 70$ GeV, and this must be because of the reweighting of the *tight-with-real-lepton* sample, since Fig. A.8 shows that in that range the *loose-minus-tight* model of the instrumental background does not get better after its reweighting. However, we are not able to conclude that this is a general feature, since the fake rate that we used to calculate the reweighting factors may be biased, because for the studies of this section the triangle cut was not optimized as suggested in Sect. A.4 to ensure the same instrumental background composition in the low- and high-$\not\!\!E_T$ regions. Therefore, the fake rate we measured in the region of $\not\!\!E_T < 10$ GeV may not exactly represent the fake rate in the analysis region of $\not\!\!E_T > 15$ GeV.

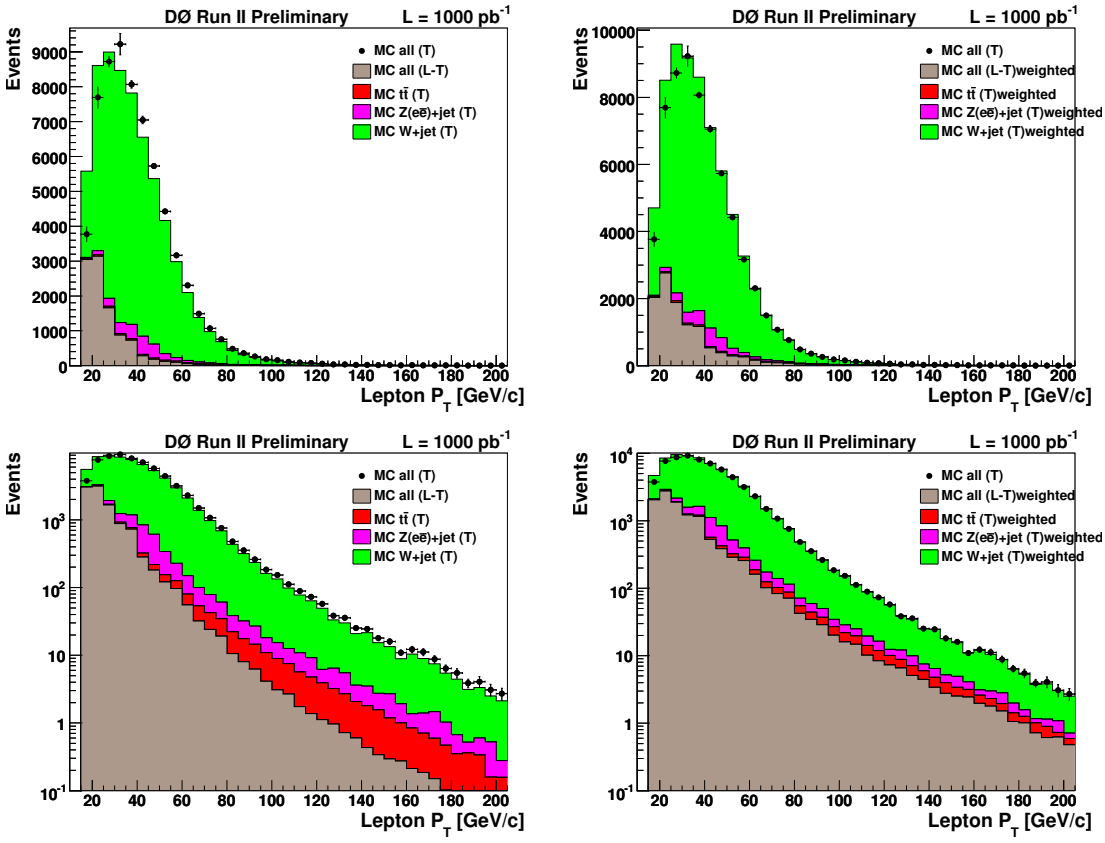

Figure A.9 : Comparison between the *tight* (T) sample (black points) and the sum of the *loose-minus-tight* (L-T) model of the instrumental background (brown histogram) plus the real lepton contributions of W+jets, Z+jets and $t\bar{t}$ (green, pink and red histograms respectively), using the Monte Carlo samples enumerated in Sect. A.2. The upper plots are in linear scale and the lower plots in logarithmic scale. Left: Comparison of the lepton p_T distribution before any corrections. Right: Same plots as on the left, but after reshaping the *loose-minus-tight* sample using the reweighting factor given by Eq. (A.24) and the W+jets, Z+jets and $t\bar{t}$ samples by the reweighting factor given by Eq. (A.28).

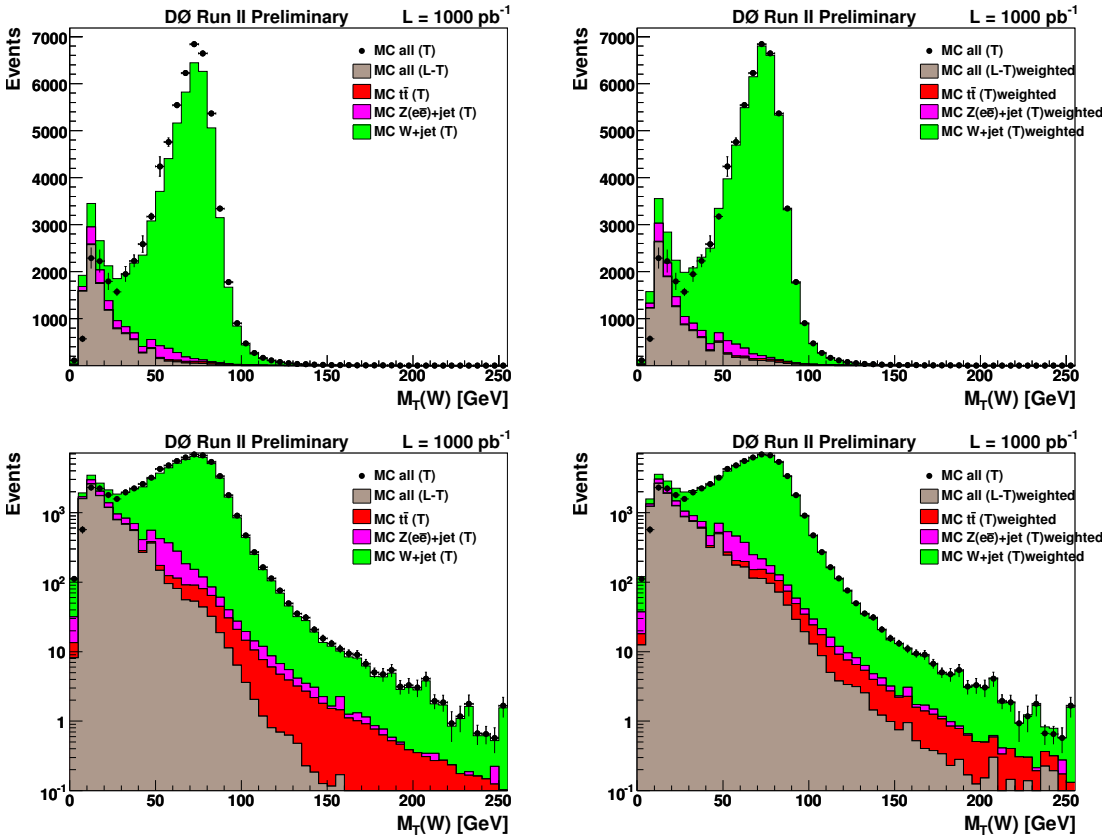

Figure A.10 : Comparison between the *tight* (T) sample (black points) and the sum of the *loose-minus-tight* (L-T) model of the instrumental background (brown histogram) plus the real lepton contributions of W+jets, Z+jets and $t\bar{t}$ (green, pink and red histograms respectively), using the Monte Carlo samples enumerated in Sect. A.2. The upper plots are in linear scale and the lower plots in logarithmic scale. Left: Comparison of the W transverse mass distribution before any corrections. Right: Same plots as on the left, but after reshaping the *loose-minus-tight* sample using the reweighting factor given by Eq. (A.24) and the W+jets, Z+jets and $t\bar{t}$ samples by the reweighting factor given by Eq. (A.28).

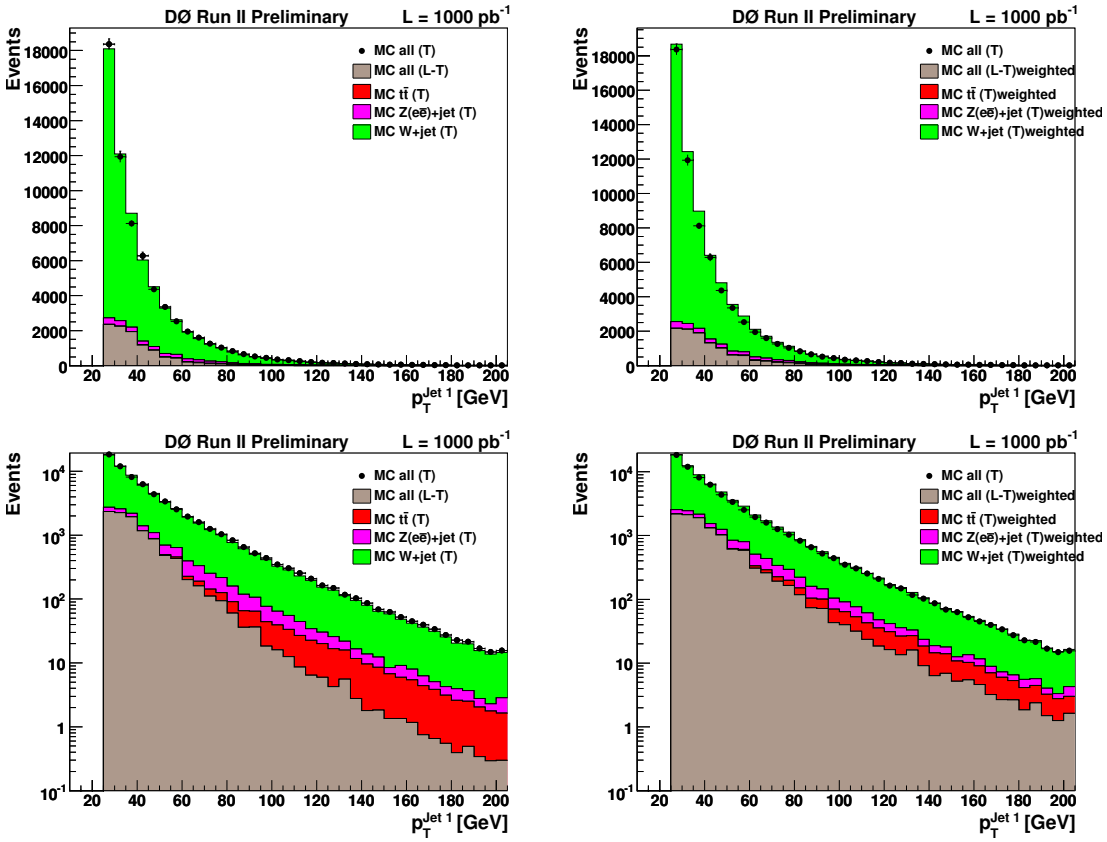

Figure A.11 : Comparison between the *tight* (T) sample (black points) and the sum of the *loose-minus-tight* (L-T) model of the instrumental background (brown histogram) plus the real lepton contributions of W+jets, Z+jets and $t\bar{t}$ (green, pink and red histograms respectively), using the Monte Carlo samples enumerated in Sect. A.2. The upper plots are in linear scale and the lower plots in logarithmic scale. Left: Comparison of the leading jet transverse momentum distribution before any corrections. Right: Same plots as on the left, but after reshaping the *loose-minus-tight* sample using the reweighting factor given by Eq. (A.24) and the W+jets, Z+jets and $t\bar{t}$ samples by the reweighting factor given by Eq. (A.28).

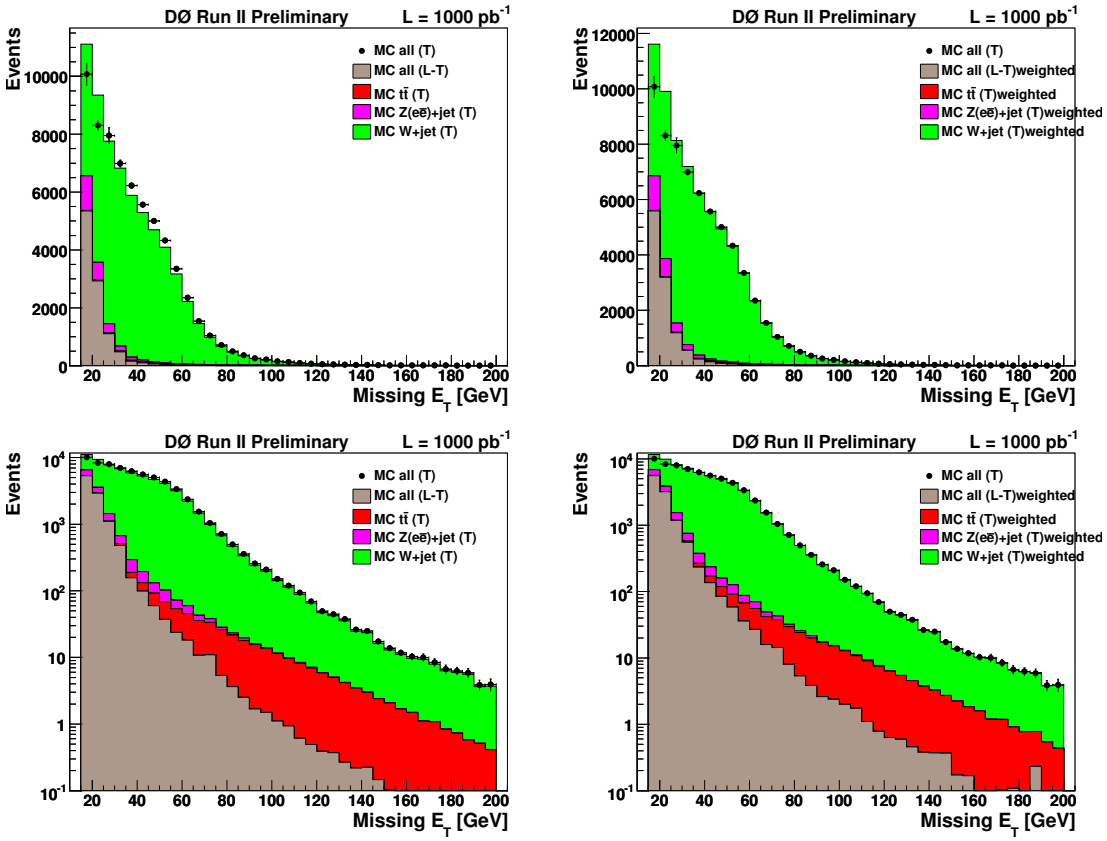

Figure A.12 : Comparison between the *tight* (T) sample (black points) and the sum of the *loose-minus-tight* (L-T) model of the instrumental background (brown histogram) plus the real lepton contributions of W+jets, Z+jets and $t\bar{t}$ (green, pink and red histograms respectively), using the Monte Carlo samples enumerated in Sect. A.2. The upper plots are in linear scale and the lower plots in logarithmic scale. Left: Comparison of the missing transverse energy distribution before any corrections. Right: Same plots as on the left, but after reshaping the *loose-minus-tight* sample using the reweighting factor given by Eq. (A.24) and the W+jets, Z+jets and $t\bar{t}$ samples by the reweighting factor given by Eq. (A.28).

B Trigger Efficiency

This appendix presents the measurement of the efficiency of the trigger Mega-OR used in the analysis and the uncertainty assigned to it. As explained in Sect. 5.1, the Mega-OR includes all "reasonable" triggers for leptons and leptons+jets in the DØ trigger list. Therefore, an efficiency close to 100% is expected. The efficiency is measured by comparing the ratio of Single-Lepton-OR (where only a lepton is required) to Mega-OR selected data, with the ratio of MC simulation that has the known Single-Lepton-OR turn-on curves applied to 100% efficient MC simulation. The multijets contribution is estimated, and subtracted, separately for the Single-Lepton-OR and the Mega-OR data samples, to look only for events with real leptons. The MC samples contain the sum of MC backgrounds. These ratios are taken on selected pretagged data requiring two or more good jets. A ratio of these two ratios is then taken to represent the efficiency of the Mega-OR selection. Assuming the Single-Lepton-OR is modeled correctly, then the ratio of ratios (MC over data) gives the Mega-OR efficiency in data divided by the Mega-OR efficiency in MC (100%), which should be less than one. Any value larger than one indicates imperfect modeling of the multijet background or the Single-Lepton-OR trigger.

These ratios are shown in Figs. B.1 to B.4 with respect to the topological variables that are likely to be most sensitive to the trigger selection (p_T(jet1), p_T(jet2), p_T(lepton), $\not\!\!E_T$, and $M_T(W)$) to observe any possible inefficiency in specific kinematic regions that the Mega-OR may introduce. In these plots, the vertical lines delineate the relevant region containing 90% of the data (with 5% of the data on either side). A horizontal line is drawn at 100% efficiency (red), and at ±5% (blue).

The figures show that there is no indication of efficiency loss for the Mega-OR trigger set. The uncertainty on the trigger efficiency is estimated from the 90% bands in the ratio-of-ratios plots. Since almost all the points are in between the 0.95 and 1.05 lines, an uncertainty of 5% is assigned to the trigger efficiency estimate of 100%. In the Run IIb muon channels the uncertainty is raised to 10%. In the statistical analysis, this uncertainty is modeled as a Gaussian with mean 1 and width 0.05, and it is allowed to vary only between 0.9 and 1.

Similar plots in Ref. [85] but as a function of the multivariate output in the boosted decision tree analysis, show that the trigger efficiency uncertainty has no shape dependence in the background region (BDToutput ∈ [0, 0.8].)

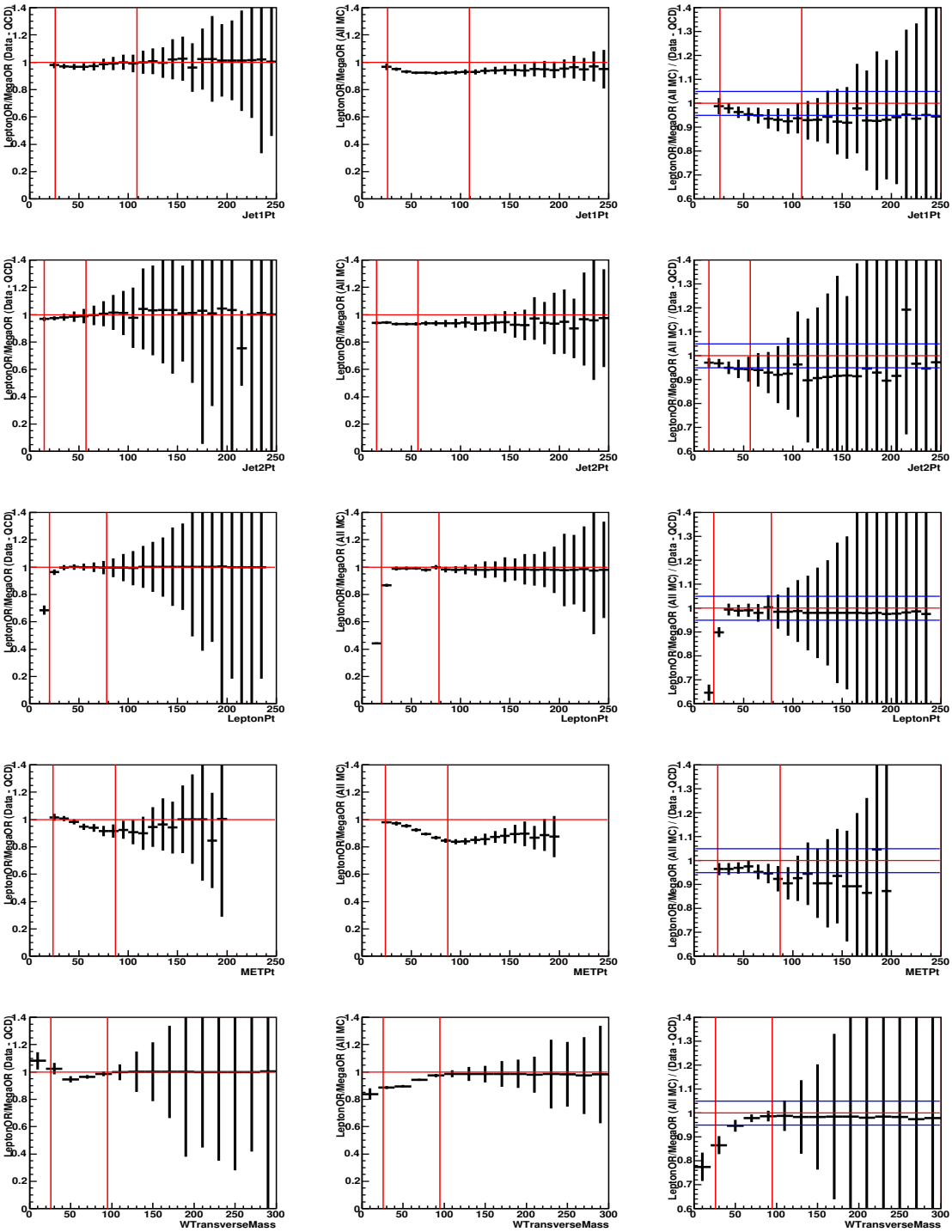

Figure B.1 : Run IIa electron channel: Ratio of Single-Lepton-OR to Mega-OR in data-minus-multijets (left column); ratio of Single-Lepton-OR to 100% in the sum of MC backgrounds (center column); ratio of these "efficiencies" in MC to data (right column). Rows show these ratios for the five chosen variables of interest.

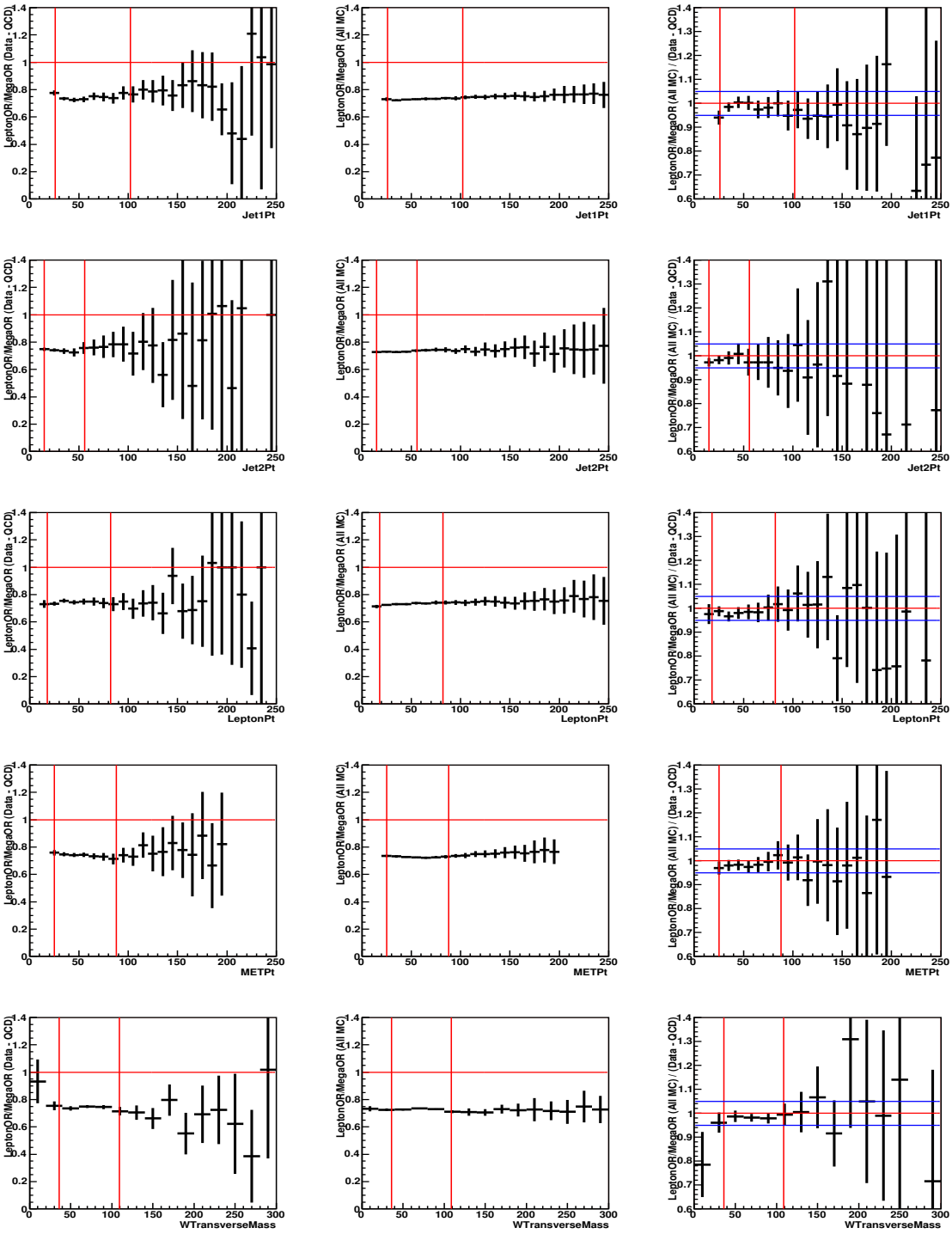

Figure B.2 : Run IIa muon channel: Ratio of Single-Lepton-OR to Mega-OR in data-minus-multijets (left column); ratio of Single-Lepton-OR to 100% in the sum of MC backgrounds (center column); ratio of these "efficiencies" in MC to data (right column). Rows show these ratios for the five chosen variables of interest.

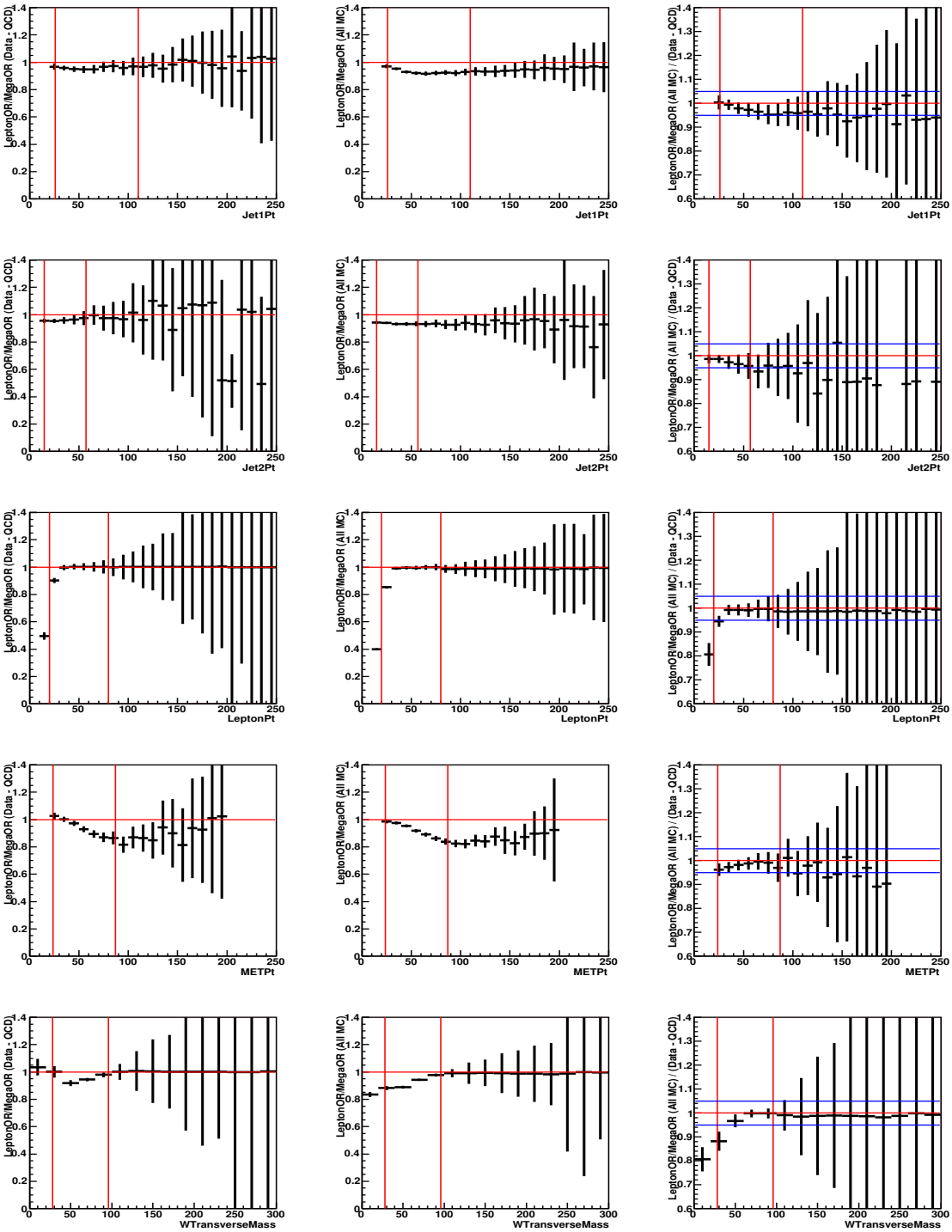

Figure B.3 : Run IIb electron channel: Ratio of Single-Lepton-OR to Mega-OR in data-minus-multijets (left column); ratio of Single-Lepton-OR to 100% in the sum of MC backgrounds (center column); ratio of these "efficiencies" in MC to data (right column). Rows show these ratios for the five chosen variables of interest.

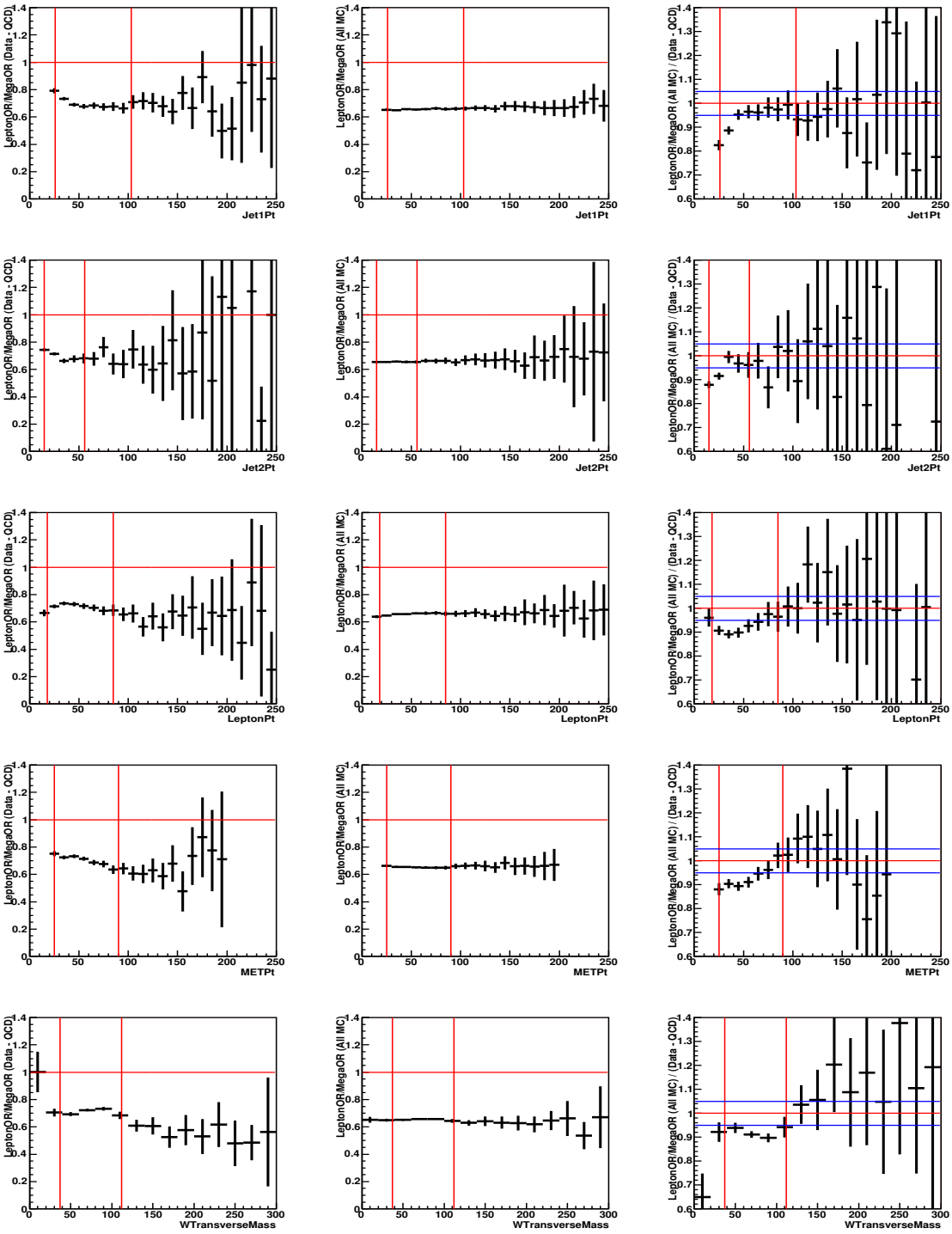

Figure B.4 : Run IIb muon channel: Ratio of Single-Lepton-OR to Mega-OR in data-minus-multijets (left column); ratio of Single-Lepton-OR to 100% in the sum of MC backgrounds (center column); ratio of these "efficiencies" in MC to data (right column). Rows show these ratios for the five chosen variables of interest.

C Plots from the Cross-Check Samples

This section presents first, many distributions showing the degree of agreement between the data and the background model for the cross-check samples defined in Sect. 6.7. Figures C.1 and C.2 correspond to the "W+jets" cross-check sample, while Figs. C.3 and C.4 correspond to the "$t\bar{t}$" cross-check sample. Each row shows a given variable in the four channels: Run IIa electron, Run IIb electron, Run IIa muon, Run IIb muon.

In order to validate every step of the Bayesian neural network analysis, these cross-check samples are further used to decide whether the background model and the data are in agreement after applying a BNN filter. Figures C.5 and C.6 show the Bayesian neural network output distributions for the cross-check samples.

In all cases, good agreement is observed between the predicted background and the observed data in both the samples, for each of the electron and muon channels and both Run IIa and Run IIb data. Thus, one concludes that the background model (W+jets and $t\bar{t}$ principally) describes the data well within uncertainties.

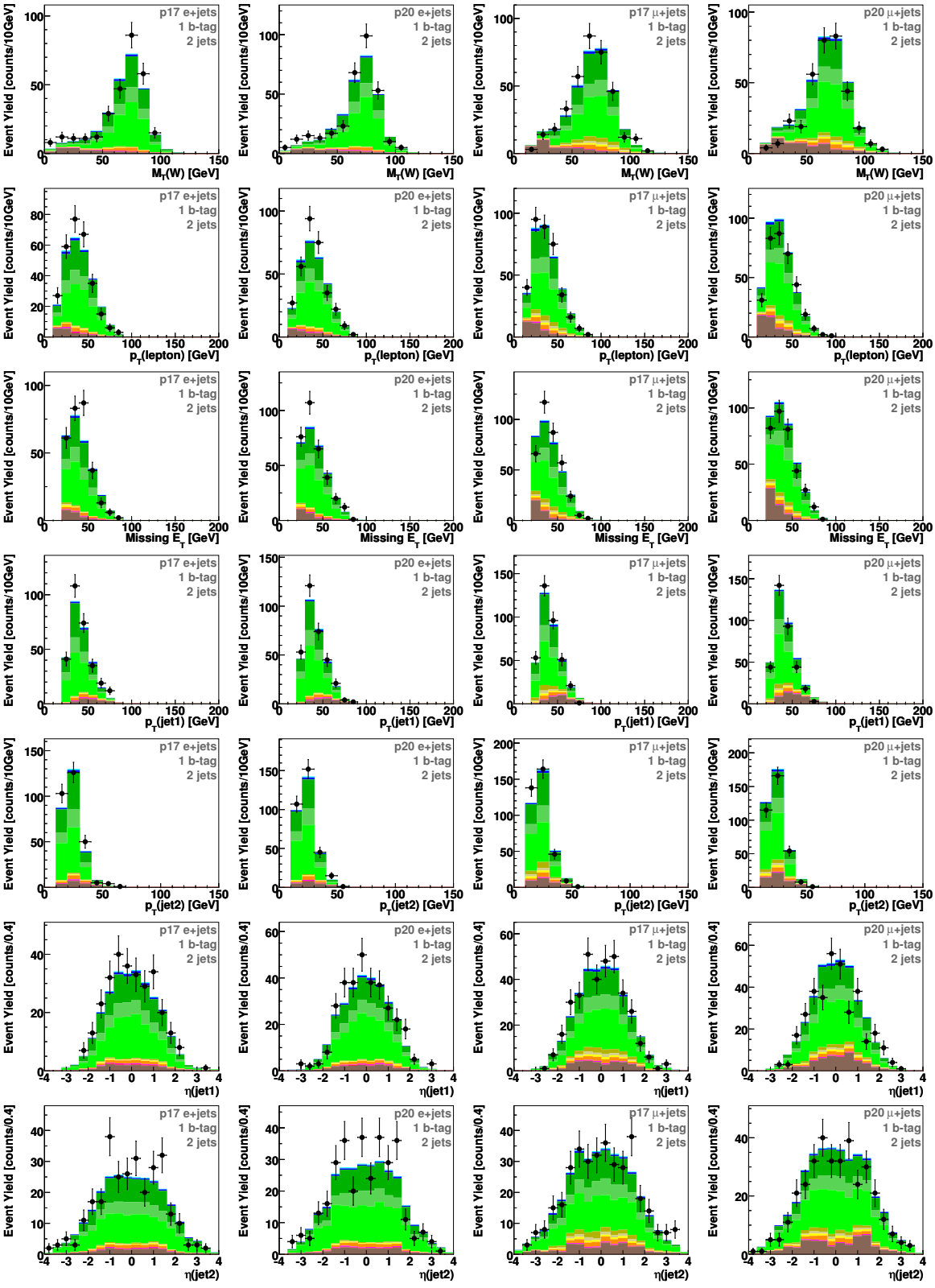

Figure C.1 : "W+jets" cross-check sample: comparison between the data (black points) and the background model (histograms) for several variables distributions.

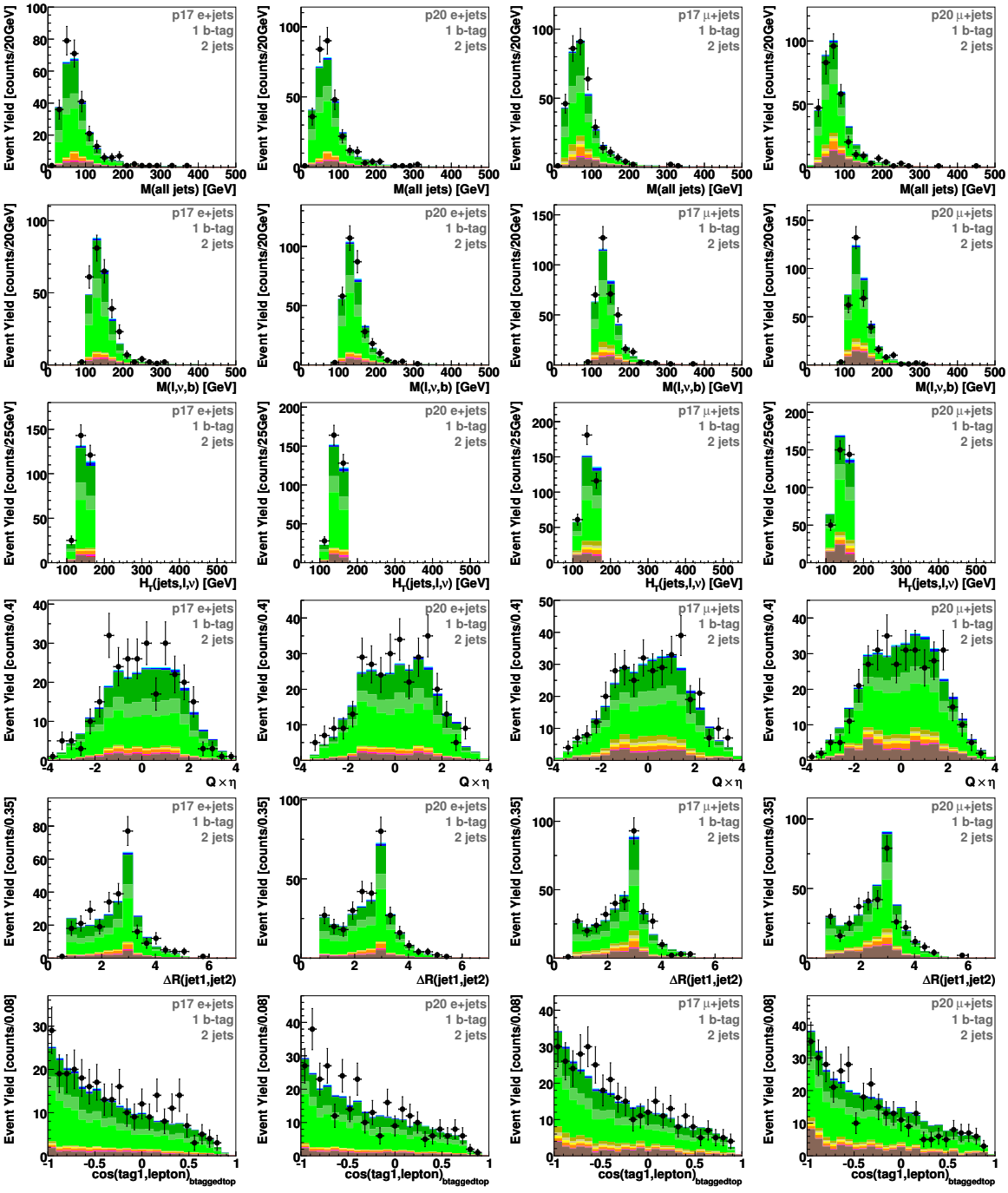

Figure C.2 : "W+jets" cross-check sample: comparison between the data (black points) and the background model (histograms) for several variables distributions.

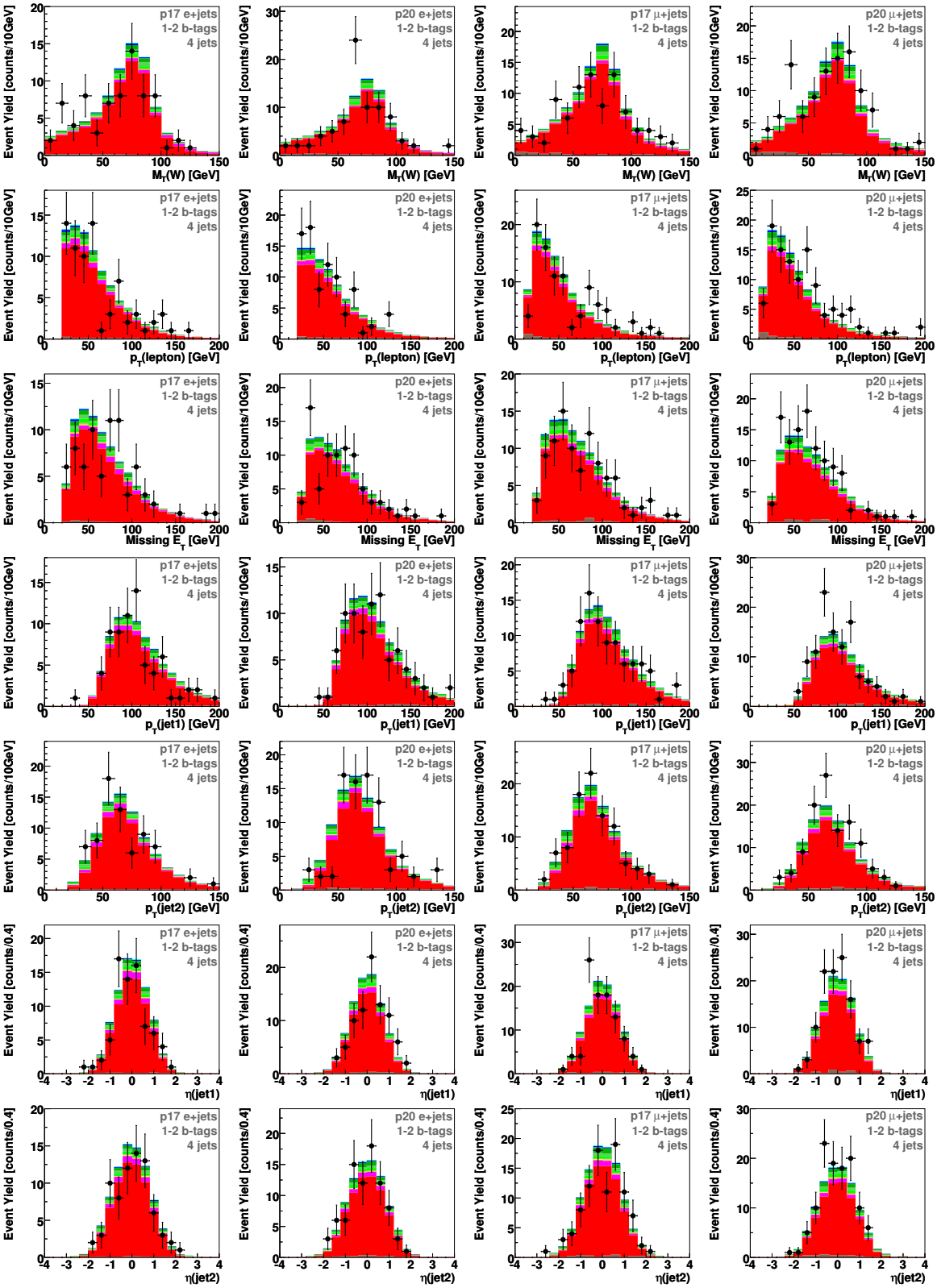

Figure C.3 : "$t\bar{t}$" cross-check sample: comparison between the data (black points) and the background model (histograms) for several variables distributions.

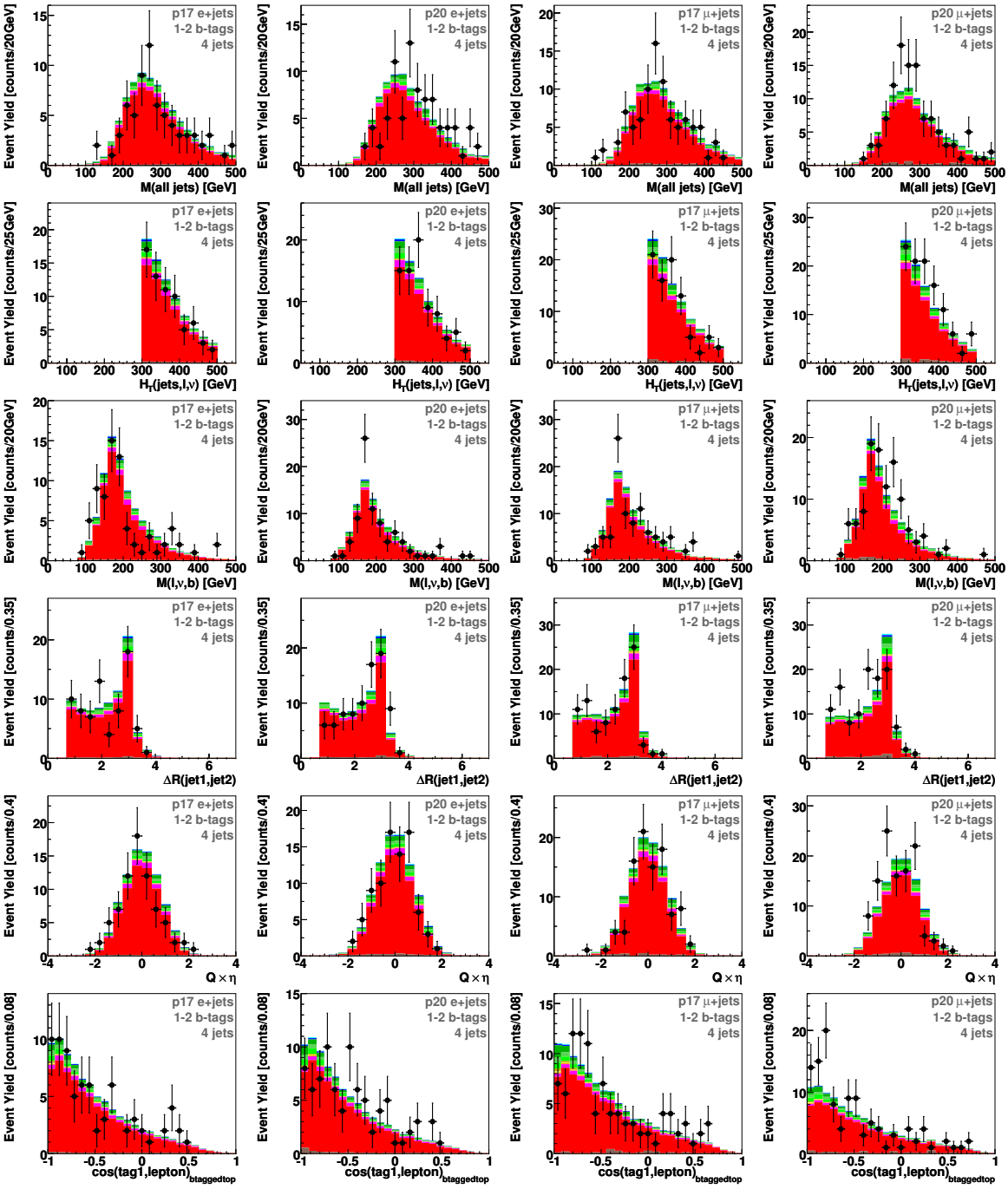

Figure C.4 : "$t\bar{t}$" cross-check sample: comparison between the data (black points) and the background model (histograms) for several variables distributions.

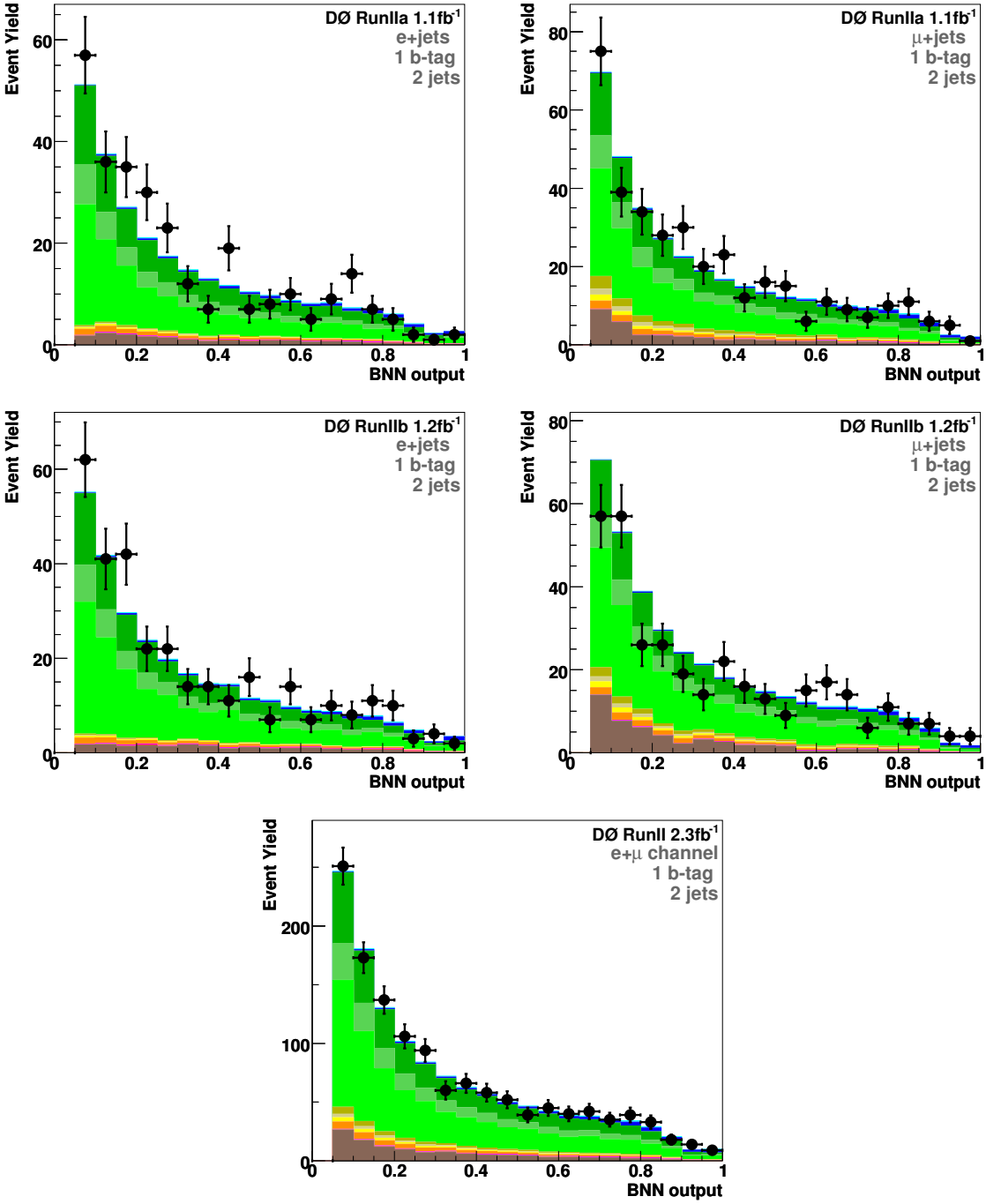

Figure C.5 : BNN outputs from the "W+jets" cross-check samples for the electron (left column) and muon (right column) channels in Run IIa (upper row) and Run IIb (central row) data. The plot at the bottom is the combination of the upper four plots.

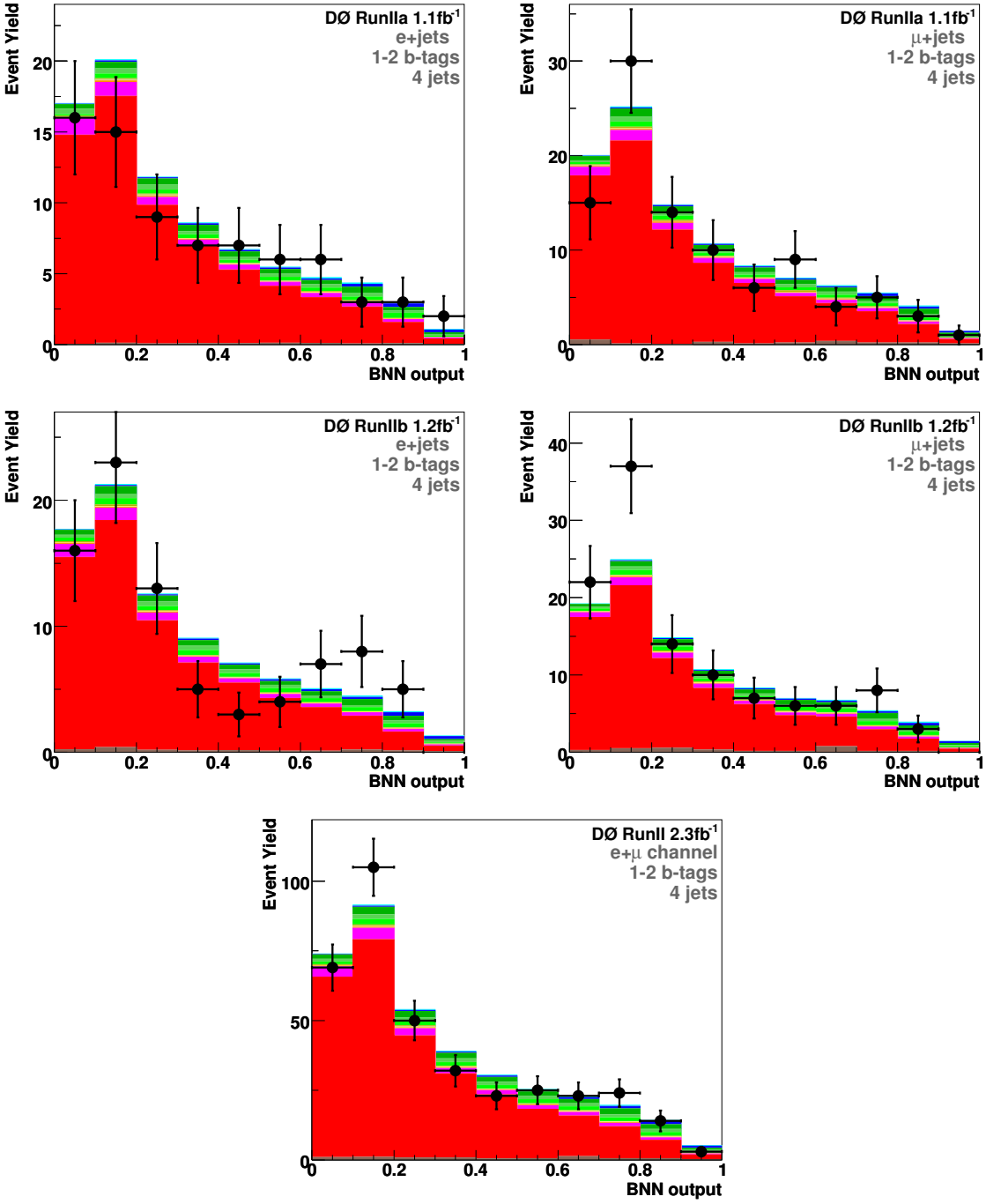

Figure C.6 : BNN outputs from "$t\bar{t}$" cross-check samples for the electron (left column) and muon (right column) channels in Run IIa (upper row) and Run IIb (central row) data. The plot at the bottom is the combination of the upper four plots.

D Shape-Changing Systematic Uncertainties

To evaluate the effect of the jet energy scale, b-tag probability and ALPGEN reweighting systematics, two additional samples are created for each of these systematics and each MC background and/or signal source affected by the systematic, with the corresponding correction/efficiency/reweighting factors shifted by plus or minus one standard deviation respectively. The Bayesian neural network filter derived by training on the nominal (not shifted) samples is then applied to these shifted samples. The resulting shifts in each BNN-output histograming bin, $\delta^{+/-}_{isys,ipar,ich,ibin}$, are then used when calculating the uncertainty in the cross section and $|V_{tb}|$ measurements (see Sect. 8.1.2). In this appendix, the relative shifts in the BNN-ouputs are shown. For each channel and each shape-changing systematic, two plots are presented, one for the sum of the affected MC backgrounds (in the ALPGEN reweighting case this includes only W+jets) and another one for the signal (not affected by the ALPGEN reweighting). Figures D.1 to D.4 correspond to the jet energy scale (JES) systematic, Figs. D.5 to D.8 to the b-tagging efficiency (TRF) systematic, and Figs. D.9 and D.10 to the ALPGEN reweighting (RWT) systematic. The red (green) curves show the plus (minus) shifts.

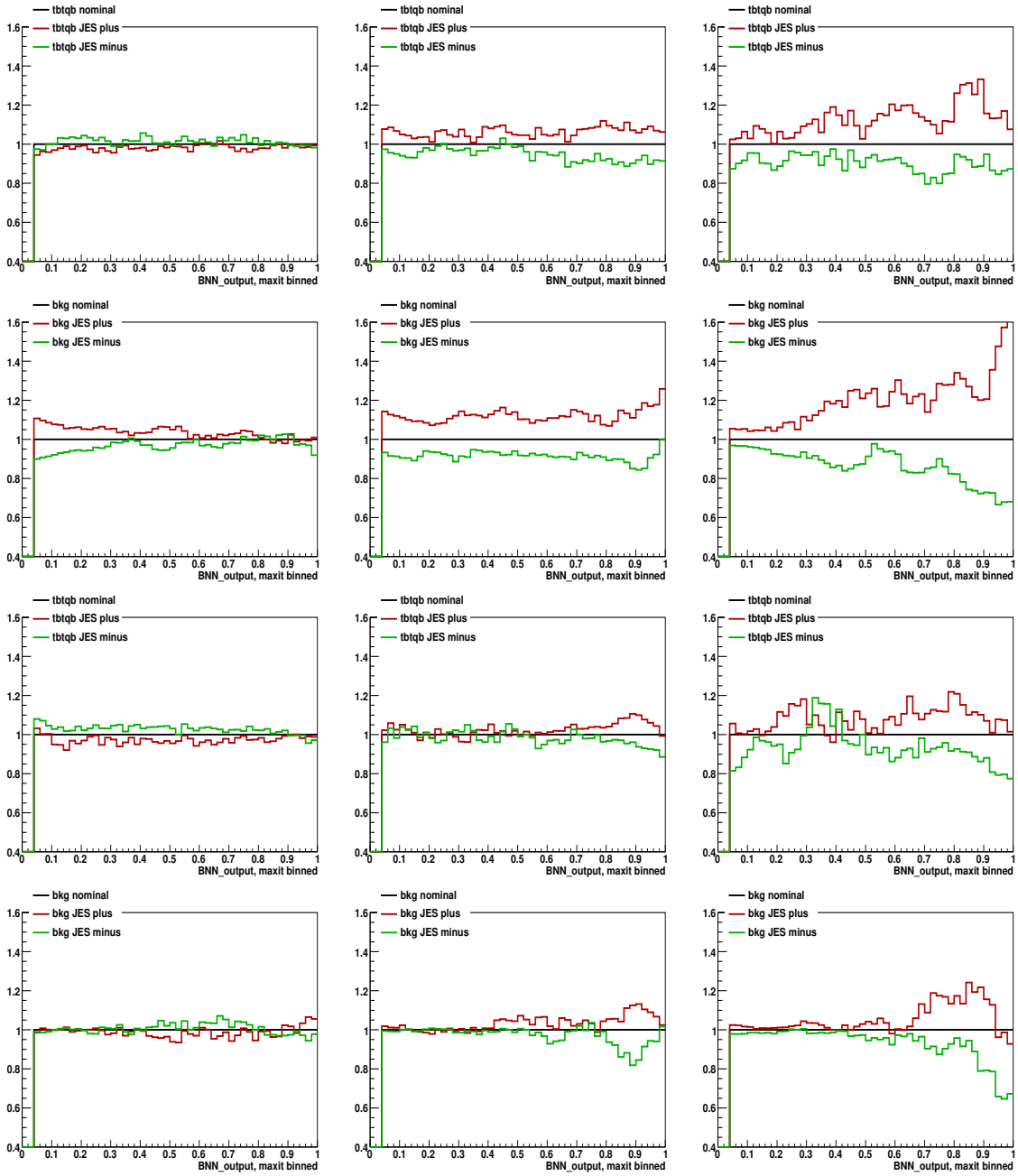

Figure D.1 : Systematic shifts when varying the JES by $\pm 1\sigma$ versus the BNN output, for signal (odd rows) and total background (even rows) in the Run IIa electron channel. Left column: events with 2 jets. Middle column: events with 3 jets. Right column: events with 4 jets. First and second rows: events with 1 b-tagged jet. Third and fourth rows: events with 2 b-tagged jets.

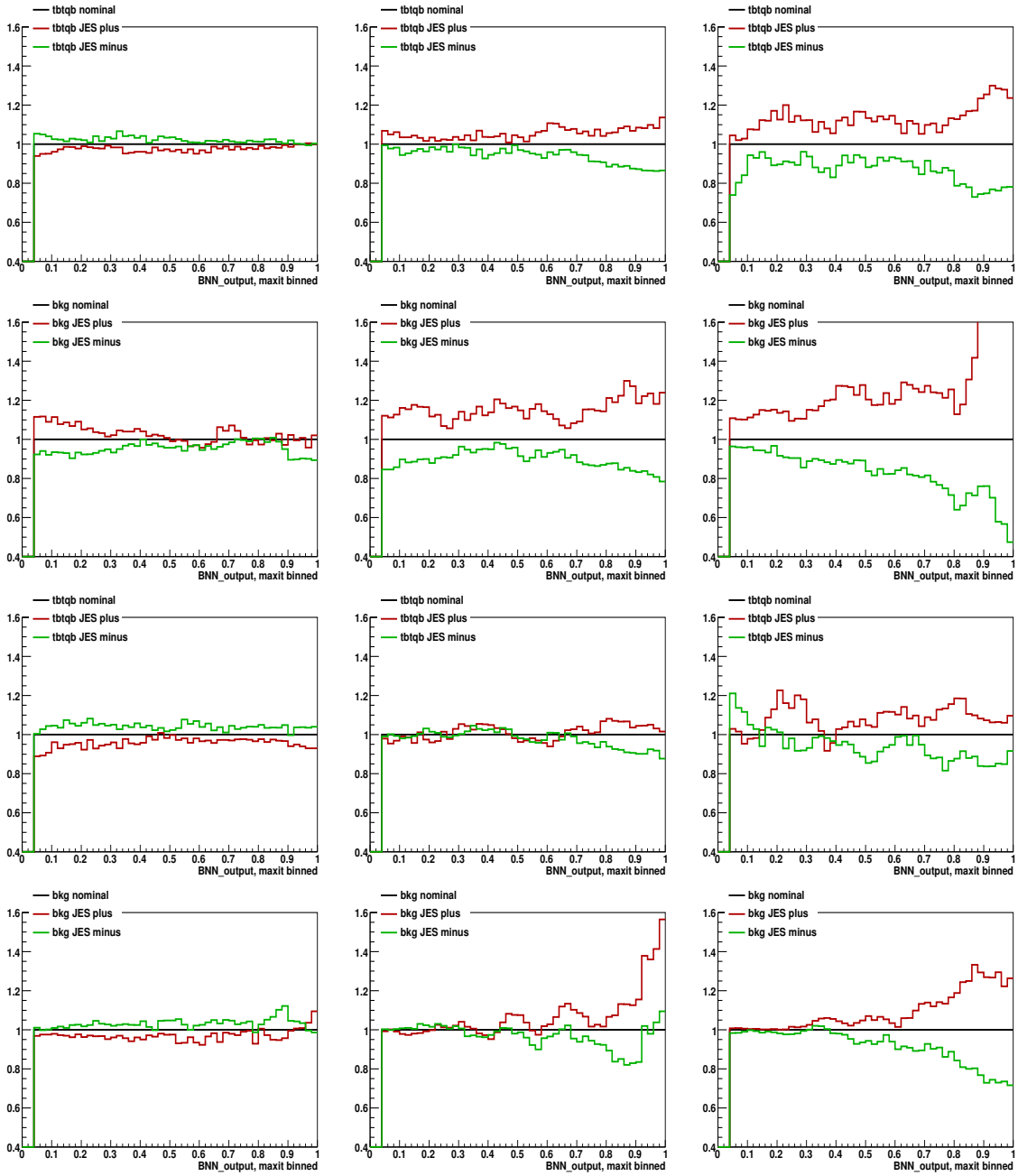

Figure D.2 : Systematic shifts when varying the JES by $\pm 1\sigma$ versus the BNN output, for signal (odd rows) and total background (even rows) in the Run IIb electron channel. Left column: events with 2 jets. Middle column: events with 3 jets. Right column: events with 4 jets. First and second rows: events with 1 b-tagged jet. Third and fourth rows: events with 2 b-tagged jets.

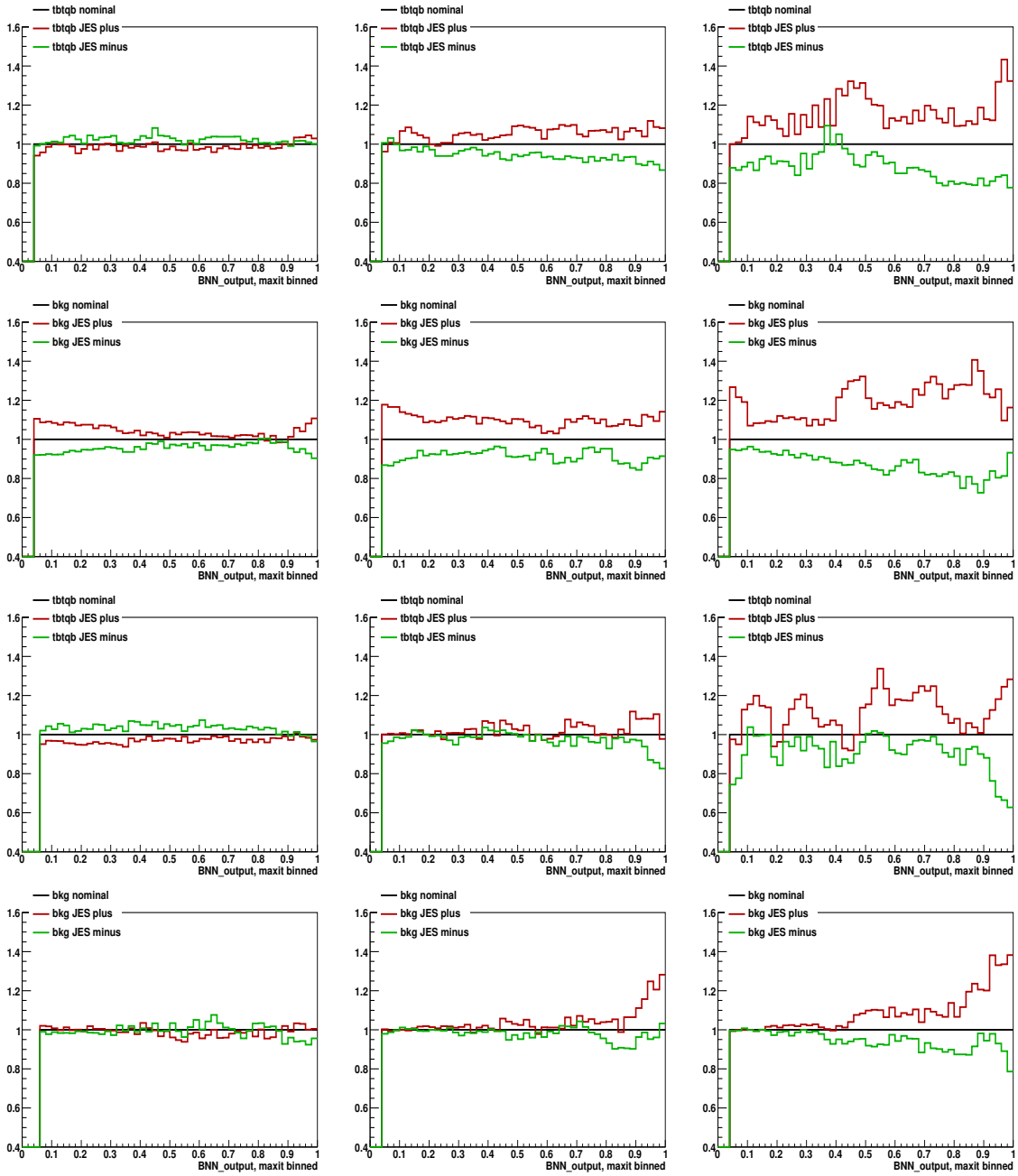

Figure D.3 : Systematic shifts when varying the JES by $\pm 1\sigma$ versus the BNN output, for signal (odd rows) and total background (even rows) in the Run IIa muon channel. Left column: events with 2 jets. Middle column: events with 3 jets. Right column: events with 4 jets. First and second rows: events with 1 b-tagged jet. Third and fourth rows: events with 2 b-tagged jets.

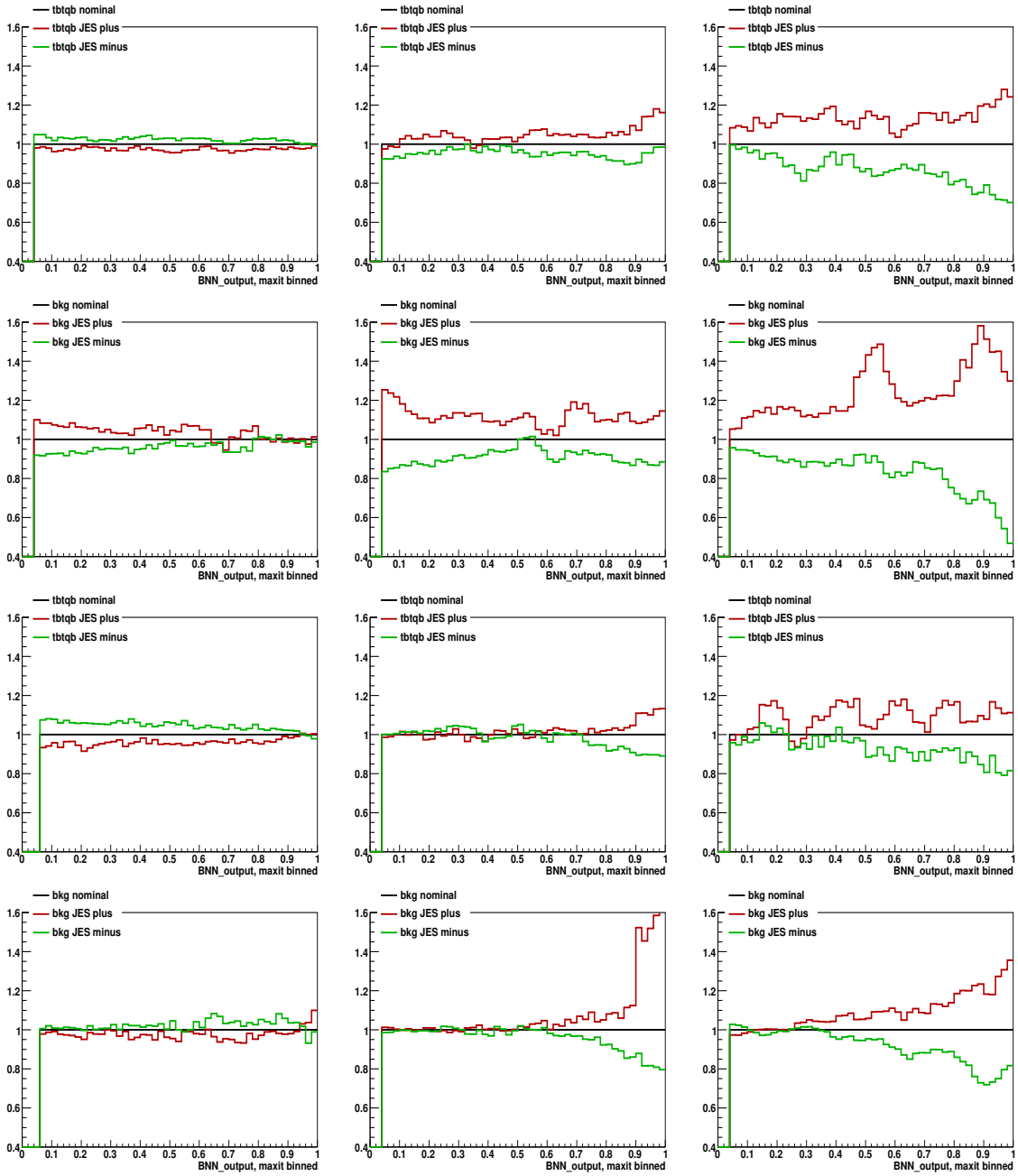

Figure D.4 : Systematic shifts when varying the JES by $\pm 1\sigma$ versus the BNN output, for signal (odd rows) and total background (even rows) in the Run IIb muon channel. Left column: events with 2 jets. Middle column: events with 3 jets. Right column: events with 4 jets. First and second rows: events with 1 b-tagged jet. Third and fourth rows: events with 2 b-tagged jets.

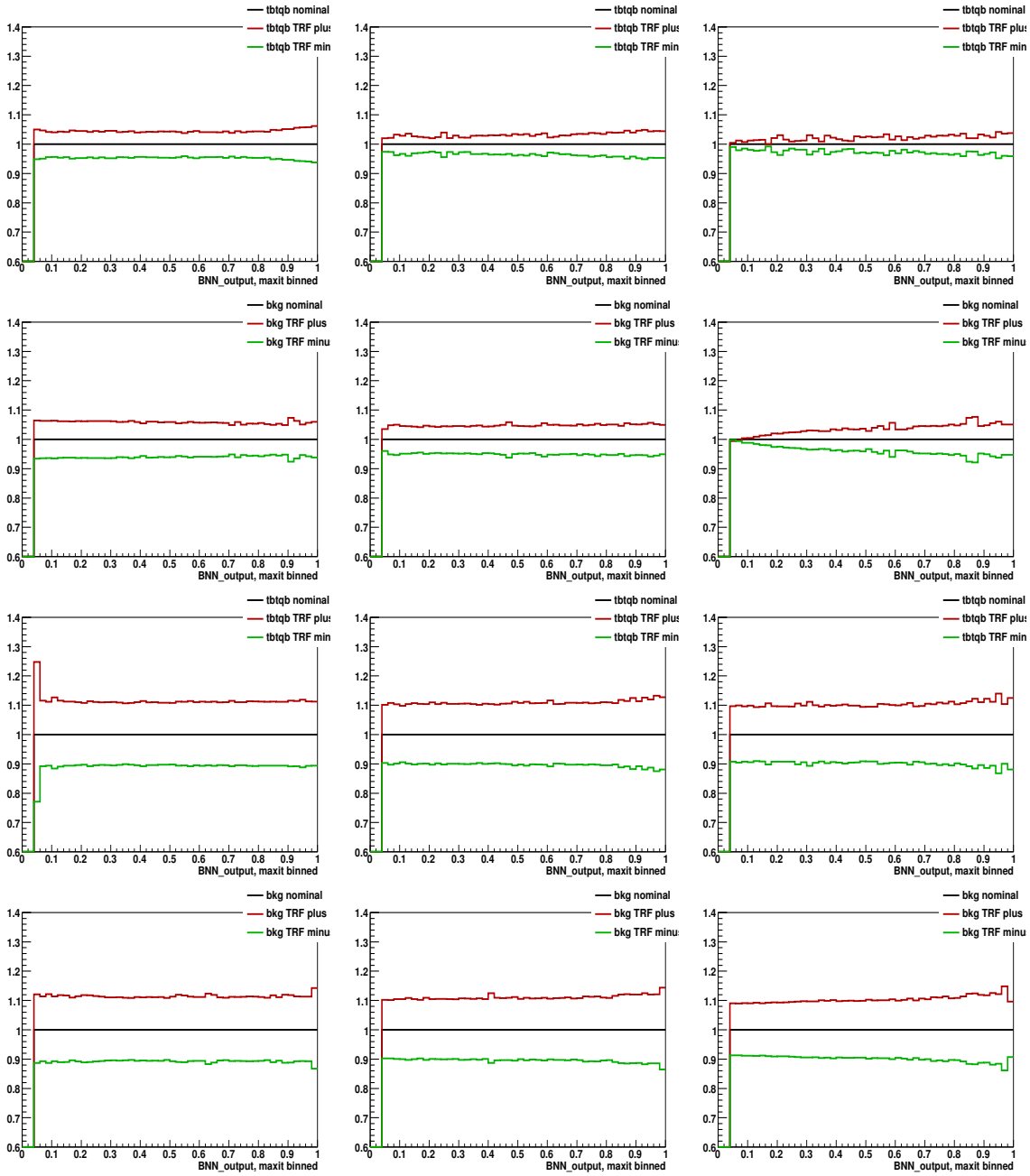

Figure D.5 : Systematic shifts when varying the TRF by $\pm 1\sigma$ versus the BNN output, for signal (odd rows) and total background (even rows) in the Run IIa electron channel. Left column: events with 2 jets. Middle column: events with 3 jets. Right column: events with 4 jets. First and second rows: events with 1 b-tagged jet. Third and fourth rows: events with 2 b-tagged jets.

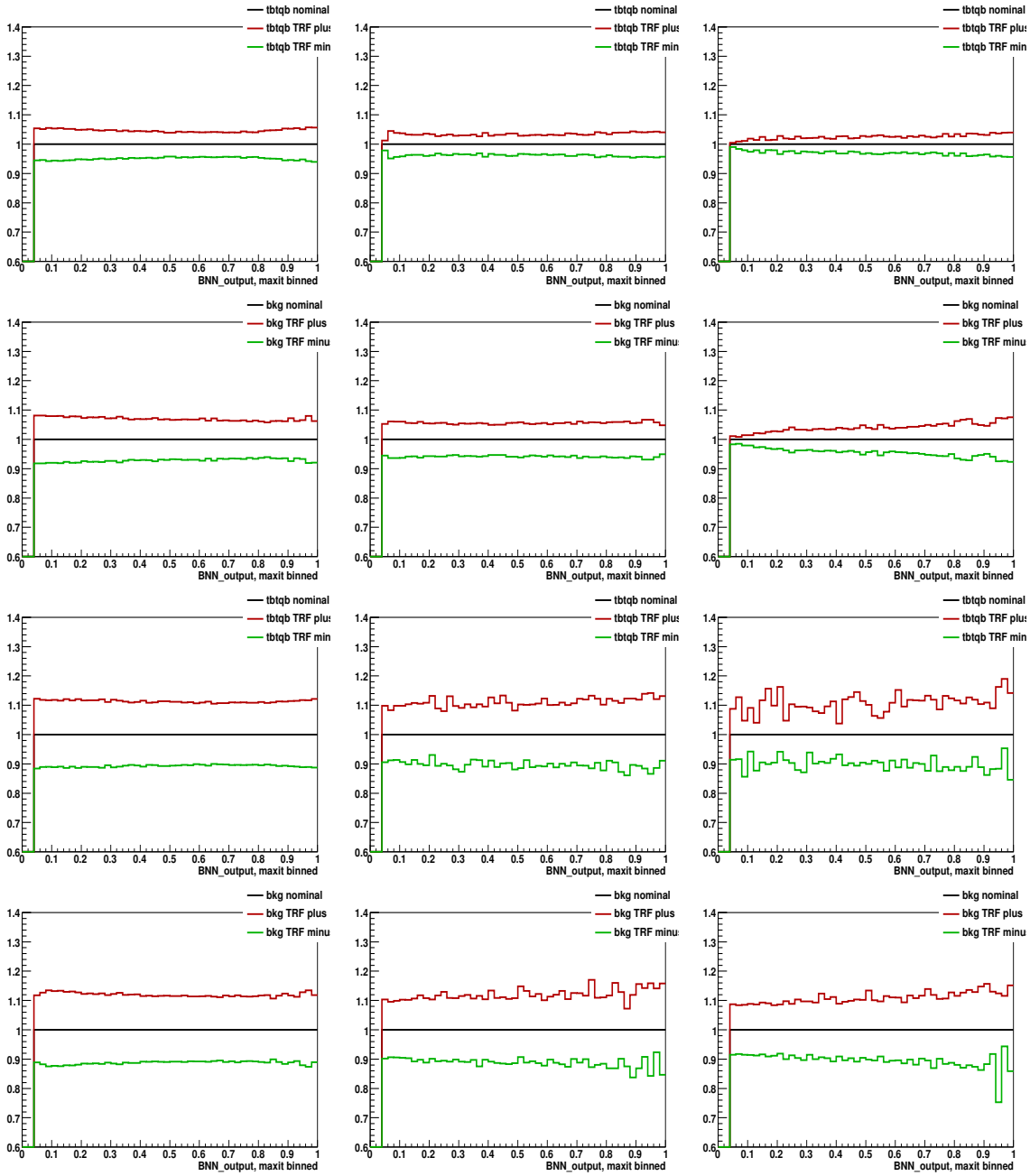

Figure D.6 : Systematic shifts when varying the TRF by $\pm 1\sigma$ versus the BNN output, for signal (odd rows) and total background (even rows) in the Run IIb electron channel. Left column: events with 2 jets. Middle column: events with 3 jets. Right column: events with 4 jets. First and second rows: events with 1 b-tagged jet. Third and fourth rows: events with 2 b-tagged jets.

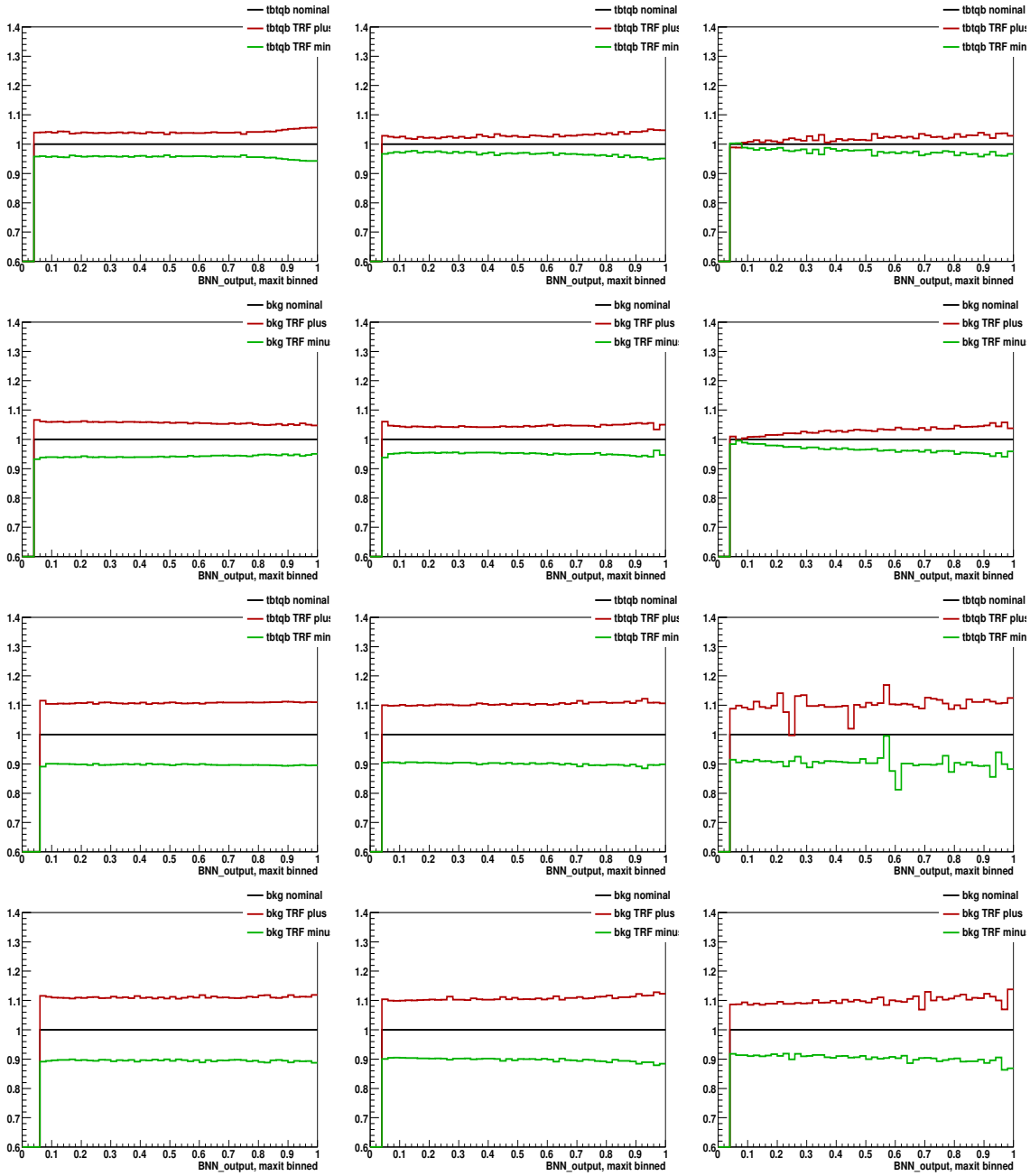

Figure D.7 : Systematic shifts when varying the TRF by $\pm 1\sigma$ versus the BNN output, for signal (odd rows) and total background (even rows) in the Run IIa muon channel. Left column: events with 2 jets. Middle column: events with 3 jets. Right column: events with 4 jets. First and second rows: events with 1 b-tagged jet. Third and fourth rows: events with 2 b-tagged jets.

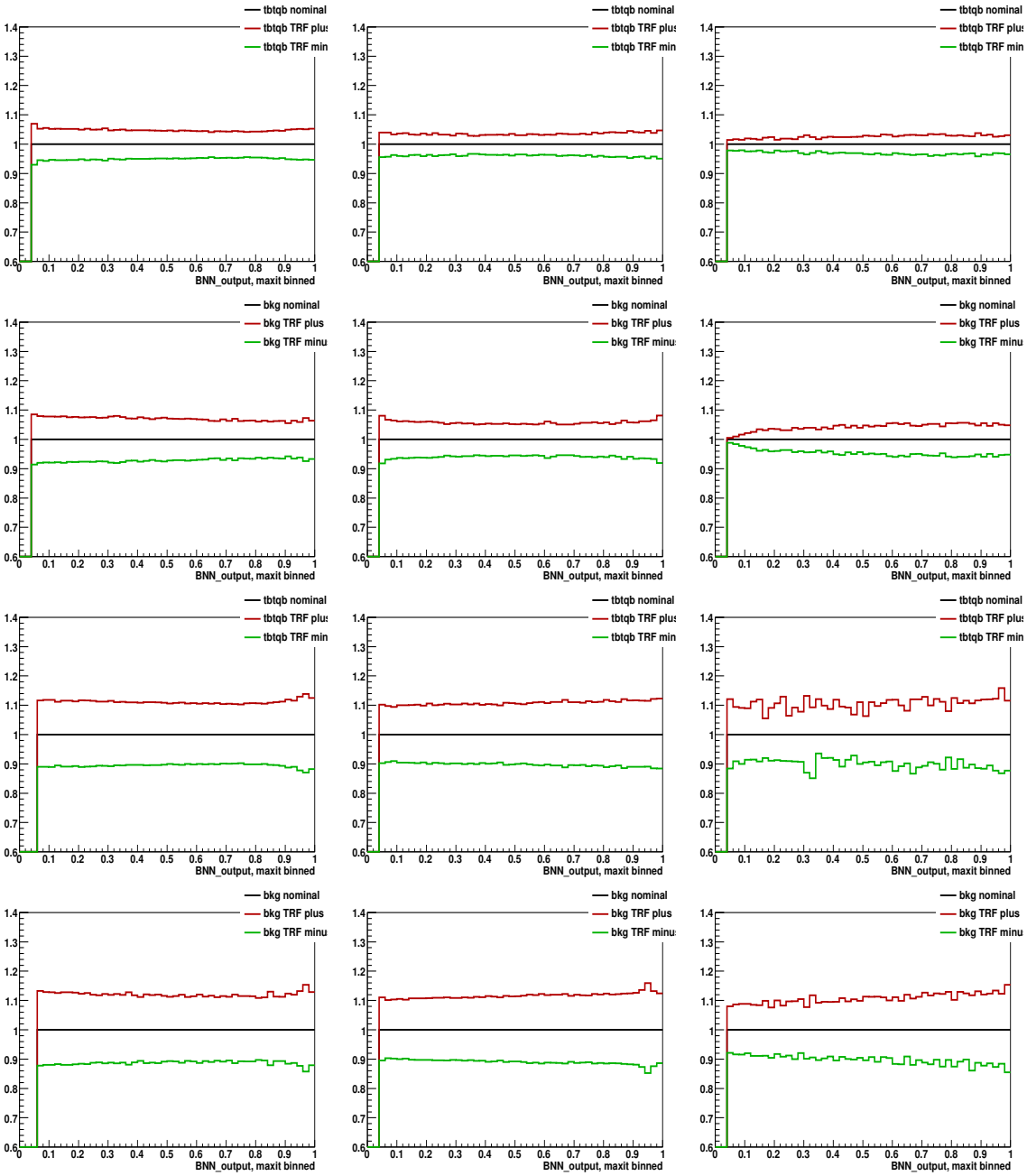

Figure D.8 : Systematic shifts when varying the TRF by $\pm 1\sigma$ versus the BNN output, for signal (odd rows) and total background (even rows) in the Run IIb muon channel. Left column: events with 2 jets. Middle column: events with 3 jets. Right column: events with 4 jets. First and second rows: events with 1 b-tagged jet. Third and fourth rows: events with 2 b-tagged jets.

RWT uncertainty – Run IIa (up) and Run IIb (down) e+jets

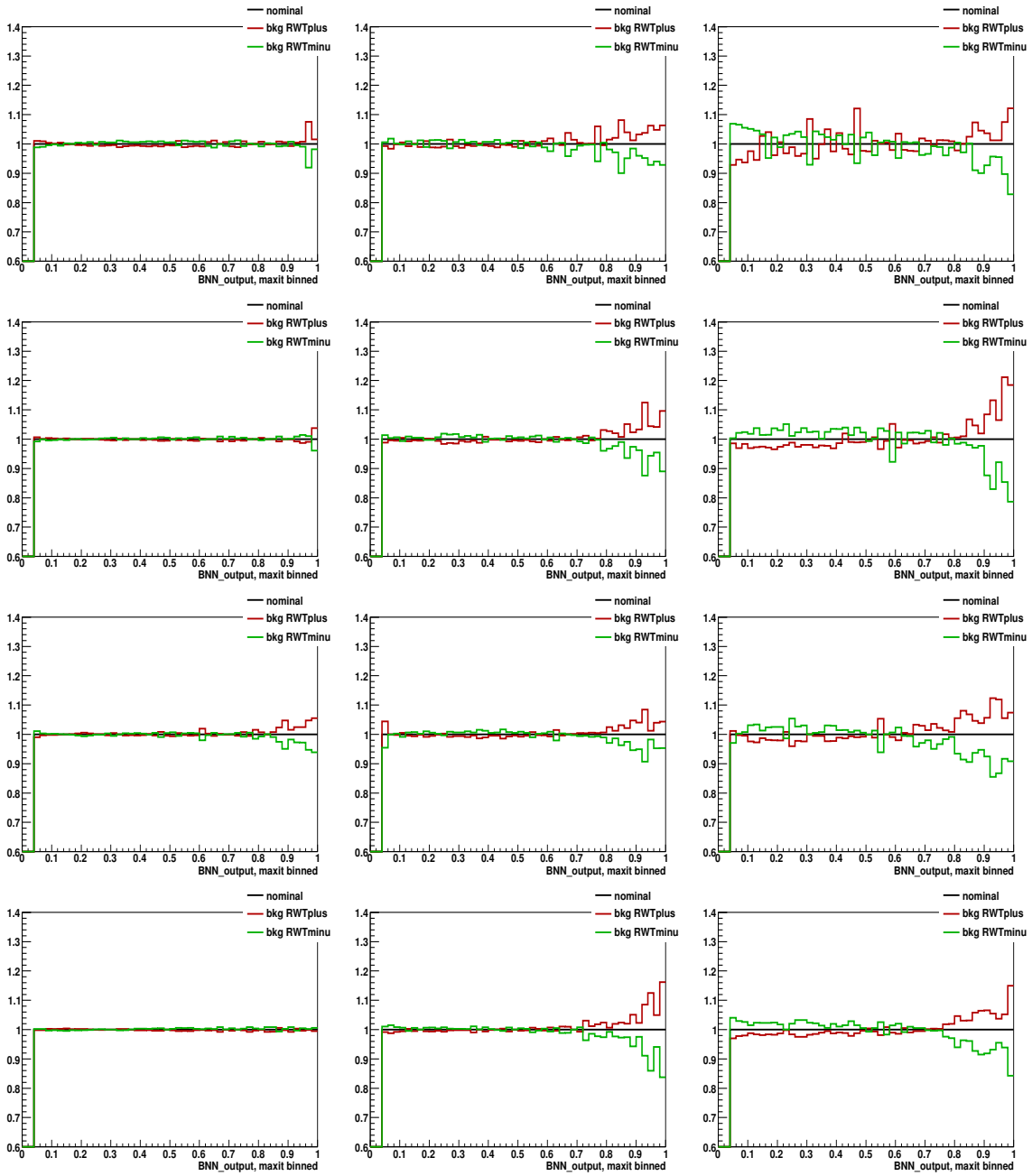

Figure D.9 : Systematic shifts when varying the ALPGEN reweighting by $\pm 1\sigma$ versus the BNN output, for total background in the Run IIa (first two rows) and Run IIb (last two rows) electron channels. Left column: events with 2 jets. Middle column: events with 3 jets. Right column: events with 4 jets. First and third rows: events with 1 b-tagged jet. Second and fourth rows: events with 2 b-tagged jets.

RWT uncertainty – Run IIa (up) and Run IIb (down) μ+jets

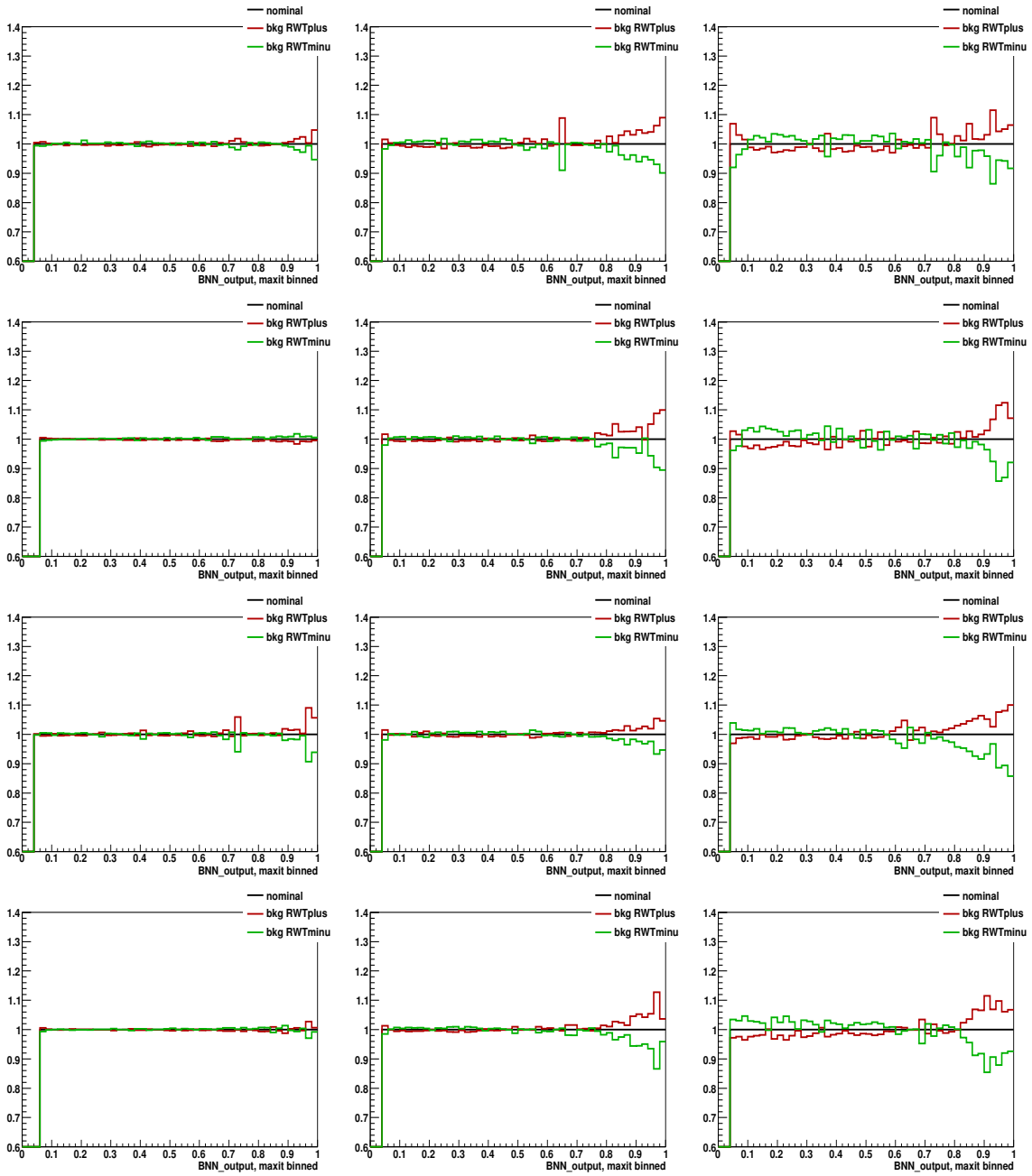

Figure D.10 : Systematic shifts when varying the ALPGEN reweighting by $\pm 1\sigma$ versus the BNN output, for total background in the Run IIa (first two rows) and Run IIb (last two rows) muon channels. Left column: events with 2 jets. Middle column: events with 3 jets. Right column: events with 4 jets. First and third rows: events with 1 b-tagged jet. Second and fourth rows: events with 2 b-tagged jets.

E Selected BNN Input Variables

This appendix shows first, for each of the 24 channels, the list of variables that are used as inputs in the training of the Bayesian neural networks, after making sure they are well modeled and optimum as explained in Sect. 7.5. Tables E.1 to E.12 show the variables used in p17, and Tables E.13 to E.24 the ones used in p20. In the last column of the tables, the K-S test value for each variable is shown together with the number of histograming bins (the number between parentheses) used to do the K-S test.

The definition of some of the variables is given in Sect. 6.6. Some variables, like the neutrino transverse momentum $p_T(\nu)$, cannot be measured directly and are instead calculated from other kinematic variables. This leads sometimes to the fact that there can be two possible values for the derived variable, coming from two possible solutions to the equations. To distinguish between the two solutions, a subscript "S2" is added to refer to the second solution.

After these twenty four tables, six more are presented (Tables E.25 to E.30), including in each of them the variables used in the four channels that have same number of b-tagged jets and same jet multiplicity. So, for example, Table E.25 shows the variables used in the four channels with 1 b-tag and 2 jets. In these tables the variables are separated in the following five categories: "Object kinematics", "Event kinematics", "Top quark reconstruction", "Angular correlations" and "Jet reconstruction". The aim of these tables is to show the degree of overlap of the variables in channels with same number of jets and b-tagged jets. Thus, to the right of the variable name, a short list is added of the channels (reco version and lepton type) in which the variable is used, together with its importance in that channel. The following abbreviations are used: $e1 \equiv$ p17 $e+$jets, $e2 \equiv$ p20 $e+$jets, $\mu1 \equiv$ p17 $\mu+$jets and $\mu2 \equiv$ p20 $\mu+$jets. For each category, variables that are shared by at least two of the four channels are shown in the two left columns, while variables that have no channel overlap appear in the two right columns. A short conclusion about the degree of overlap seen in each of the tables is given next.

- In the 1-tag 2-jets channels, 34 variables are used in total. 14 (41%) of them are common to all reco versions and lepton types, 2 (6%) are common in only three channels, 5 (15%) are common in only two channels, and 13 (38%) do not have overlap with other channels. But from the point of view that each channel has ~ 21 variables, one can conclude that there is a high degree of total overlap, since $\sim 65\%$ of those variables are shared with all the other channels. One can also say that, in average, each channel has 3.25 variables that are not present in the other channels. The category with more overlap is "Event kinematics".

- In the 1-tag 3-jets channels, 39 variables are used in total. 9 (23%) of them are common to all reco versions and lepton types, 7 (18%) are common in only three

channels, 8 (21%) are common in only two channels, and 15 (38%) do not have overlap with other channels. In average, each channel has ~ 22 variables, meaning that the degree of total overlap is $\sim 41\%$. Also, in average, each channel has 3.75 variables not present in the other channels. The category with more overlap is again "Event kinematics".

- In the 1-tag 4-jets channels, 51 variables are used in total. 11 (21.5%) of them are common to all reco versions and lepton types, 7 (14%) are common in only three channels, 11 (21.5%) are common in only two channels, and 22 (43%) do not have overlap with other channels. Each channel has in average ~ 27 variables. Therefore, the degree of total overlap is $\sim 41\%$. However, each channel has, in average, 5.5 variables that are not shared with other channels. The category with more overlap is "Event kinematics". The "Angular correlation" variables have a much less degree of overlap than in channels with fewer number of jets. Similarly with the "Top quark reconstruction" variables, which is understandable since the more jets are present in the event the more possibilities are to reconstruct the top quark.

- In the 2-tag 2-jets channels, 38 variables are used in total. 12 (31.5%) of them are common to all reco versions and lepton types, 4 (11%) are common in only three channels, 9 (23.5%) are common in only two channels, and 13 (34%) do not have overlap with other channels. In average, each channel has ~ 23 variables. Since 12 variables are shared with all other channels, the total overlap is $\sim 52\%$. In average, each channel has 3.25 non-shared variables. All categories except "Jet reconstruction" present a high degree of variables overlap.

- In the 2-tag 3-jets channels, 39 variables are used in total. 13 (33%) of them are common to all reco versions and lepton types, 9 (23%) are common in only three channels, 7 (18%) are common in only two channels, and 9 (26%) do not have overlap with other channels. Each channel has in average ~ 27 variables. The degree of total overlap is then $\sim 48\%$. This is the tag and jet combination where each channel has the less number of non-shared variables, only 2.5, and the majority of them correspond to the "Object kinematics" and "Angular correlations" categories. All other categories have a high degree of overlap.

- In the 2-tag 4-jets channels, 44 variables are used in total. 14 (32%) of them are common to all reco versions and lepton types, 6 (14%) are common in only three channels, 8 (18%) are common in only two channels, and 16 (36%) do not have overlap with other channels, giving 4 non-shared variables in average per channel. Each channel has in average ~ 27 variables, giving a degree of total overlap of $\sim 52\%$. The same as it happend in the 1-tag 4-jet case, the category with the lesser overlap is "Top quark reconstruction". The category with the highest overlap is "Event kinematics".

From the six tables in general, one can conclude that the overlap of variables is high enough to make one confident of the variable selection method used in the BNN analysis.

Analysis Channel: p17 e+jets 1-tag 2-jets

Rank	Variable	Importance	K-S (N of bins)
1	M(jet1,jet2)	1.0000	0.235 (500)
2	$M_T(W)$	0.8103	0.435 (300)
3	$M(W$,tag1)	0.8014	0.550 (500)
4	$\Delta M_{\rm top}^{\rm min}$	0.7819	0.155 (5000)
5	H_T(lepton,$\not\!\!E_T$,jet1,jet2)	0.7263	0.154 (500)
6	$M(W$,tag1,$S2)$	0.5957	0.386 (500)
7	$\not\!\!E_T$	0.5734	0.465 (250)
8	Q(lepton)$\times\eta$(light1)	0.4845	0.353 (80)
9	cos(lepton$_{\rm btaggedtop}$,btaggedtop$_{\rm CMframe}$)	0.4329	0.928 (100)
10	cos(lepton,tag1)$_{\rm btaggedtop}$	0.3980	0.775 (100)
11	p_T(jet1)	0.3639	0.552 (250)
12	Width$_\eta$(jet2)	0.3433	0.507 (100)
13	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.2956	0.826 (80)
14	Width$_\phi$(jet2)	0.2550	0.101 (100)
15	p_T(jet2)	0.2487	0.981 (250)
16	Q(lepton)$\times\eta$(best1)	0.2374	0.657 (80)
17	E(jet2)	0.1894	0.149 (250)
18	p_T(best1)	0.1732	0.756 (250)
19	$p_T^{\rm rel}$(jet1,μ)	0.1598	0.968 (100)
20	cos(lepton,light1)$_{\rm btaggedtop}$	0.1248	0.931 (100)
21	cos(lepton,Q(lepton)$\times z$)$_{\rm besttop}$	0.1140	0.497 (100)

Table E.1 : Set of selected variables, ordered by their importance, for the e+jets 1-tag 2-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 e+jets 1-tag 3-jets

Rank	Variable	Importance	K-S (N of bins)
1	$M_T(W)$	1.0000	0.862 (300)
2	$\Delta M_{\text{top}}^{\text{min}}$	0.8748	0.490 (5000)
3	$M(W,\text{tag1},S2)$	0.8363	0.240 (500)
4	H_T(alljets)	0.7968	0.183 (800)
5	M(jet1,jet2)	0.7488	0.470 (500)
6	H_T(lepton,$\not\!\!E_T$,alljets)	0.7062	0.091 (500)
7	Q(lepton)$\times\eta$(light1)	0.6782	0.937 (80)
8	$\cos(\text{lepton}_{\text{besttop}},\text{besttop}_{\text{CMframe}})$	0.6690	0.068 (100)
9	$M(W,\text{tag1})$	0.6470	0.180 (500)
10	Width_η(tag1)	0.6115	0.845 (100)
11	$\cos(\text{lepton,best1})_{\text{besttop}}$	0.5730	0.580 (100)
12	H_T(lepton,$\not\!\!E_T$)	0.5402	0.338 (500)
13	Centrality(alljets)	0.5238	0.720 (25)
14	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.4691	0.837 (80)
15	p_T(light2)	0.4671	0.690 (250)
16	Width_ϕ(notbest1)	0.4107	0.914 (100)
17	M(light1,light2)	0.3969	0.609 (500)
18	M(alljets)	0.3919	0.965 (1000)
19	Width_η(light2)	0.3408	0.080 (100)
20	$\cos(\text{lepton}_{\text{btaggedtop}},\text{btaggedtop}_{\text{CMframe}})$	0.3338	0.613 (100)
21	$M(W,\text{jet2},S2)$	0.3220	0.104 (500)
22	p_T^{rel}(jet1,μ)	0.2637	1.000 (100)
23	Width_ϕ(light2)	0.2553	0.804 (100)

Table E.2 : Set of selected variables, ordered by their importance, for the e+jets 1-tag 3-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 e+jets 1-tag 4-jets

Rank	Variable	Importance	K-S (N of bins)
1	Centrality(alljets)	1.0000	0.435 (25)
2	M(alljets−tag1)	0.7294	0.320 (500)
3	p_T(notbest2)	0.7222	0.827 (250)
4	$M_T(W)$	0.7048	0.602 (300)
5	p_T(jet4)	0.6652	0.236 (250)
6	M(alljets)	0.6465	0.304 (1000)
7	Q(lepton)×η(light1)	0.6259	0.775 (80)
8	M(W,tag1)	0.6051	0.474 (500)
9	H_T(lepton,$\not\!\!E_T$,alljets)	0.5516	0.961 (500)
10	M(alljets−best1)	0.4768	0.065 (500)
11	cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	0.4347	0.410 (100)
12	Width$_\eta$(light2)	0.4332	0.777 (100)
13	Q(lepton)×η(best1)	0.4258	0.559 (80)
14	Width$_\eta$(tag1)	0.3372	0.981 (100)
15	Width$_\phi$(light2)	0.3271	0.800 (100)
16	ΔR(jet1,jet2)	0.3072	0.415 (50)
17	cos(lepton,tag1)$_{\text{btaggedtop}}$	0.2885	0.997 (100)
18	Width$_\eta$(jet4)	0.2838	0.565 (100)
19	H_T(alljets−best1)	0.2732	0.975 (500)
20	cos(lepton$_{\text{besttop}}$,besttop$_{\text{CMframe}}$)	0.2600	0.839 (100)
21	Width$_\phi$(jet4)	0.2499	0.549 (100)
22	M(W,jet1,$S2$)	0.2397	0.988 (500)
23	H_T(alljets)	0.2365	0.886 (800)
24	$\not\!\!E_T$	0.2072	0.217 (250)
25	p_T(jet3)	0.1491	0.842 (250)

Table E.3 : Set of selected variables, ordered by their importance, for the e+jets 1-tag 4-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 e+jets 2-tags 2-jets

Rank	Variable	Importance	K-S (N of bins)
1	$\Delta M_{\rm top}^{\rm min}$	1.0000	0.193 (5000)
2	M(jet1,jet2)	0.8015	0.200 (1000)
3	$M_T(W)$	0.8008	0.990 (300)
4	Width$_\eta$(tag1)	0.6549	0.855 (100)
5	H_T(jet1,jet2)	0.6074	0.482 (500)
6	Q(lepton)$\times\eta$(notbest1)	0.4875	0.557 (80)
7	$M(W$,best1)	0.4695	0.861 (500)
8	p_T(best1)	0.4618	0.004 (250)
9	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.4604	0.618 (80)
10	$\Delta R^{\rm min}$(alljets)	0.4311	0.064 (50)
11	Q(lepton)$\times\eta$(best1)	0.4277	0.903 (80)
12	cos(lepton,jet1)$_{\rm btaggedtop}$	0.4030	0.958 (100)
13	$\not\!\!E_T$	0.3816	0.487 (250)
14	$\sqrt{\hat{s}}$	0.3790	0.254 (1000)
15	H_T(lepton,$\not\!\!E_T$,jet1,jet2)	0.3649	0.157 (500)
16	Centrality(alljets)	0.3352	0.710 (25)
17	Significance$_{\rm min}(M_{\rm top})$	0.3091	0.300 (1000)
18	$M(W$,best1,$S2)$	0.3039	0.975 (500)
19	$M_{\rm top}^{\Delta M^{\rm min}}$	0.2405	0.384 (500)
20	$M(W$,notbest1,$S2)$	0.2362	0.410 (500)
21	M(tag1)	0.1832	0.616 (100)

Table E.4 : Set of selected variables, ordered by their importance, for the e+jets 2-tags 2-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 e+jets 2-tags 3-jets

Rank	Variable	Importance	K-S (N of bins)
1	p_T(notbest2)	1.0000	0.310 (250)
2	H_T(alljets)	0.9219	0.047 (800)
3	Q(lepton)$\times\eta$(light1)	0.9051	0.580 (80)
4	M(alljets)	0.8222	0.785 (1000)
5	$M_T(W)$	0.8203	0.587 (300)
6	$M(W$,tag1,$S2)$	0.8076	0.642 (500)
7	H_T(lepton,$\not\!\!E_T$)	0.8012	0.103 (500)
8	Width$_\eta$(light2)	0.7963	0.294 (100)
9	M(alljets$-$best1)	0.7652	0.443 (500)
10	$M(W$,tag1)	0.6400	0.074 (500)
11	M(jet1)	0.5976	0.160 (100)
12	M(alljets$-$tag1)	0.5955	0.304 (500)
13	Width$_\eta$(tag1)	0.5923	0.882 (100)
14	$M(W$,light1,$S2)$	0.5656	0.287 (500)
15	Width$_\eta$(tag2)	0.5299	0.460 (100)
16	$\Delta R^{\min}$(alljets)	0.5015	0.985 (50)
17	Centrality(alljets)	0.4931	0.975 (25)
18	M(jet1,jet2)	0.4856	0.560 (500)
19	Width$_\phi$(jet1)	0.4641	0.296 (100)
20	$\Delta M^{\min}_{\text{top}}$	0.4076	0.463 (5000)
21	Width$_\eta$(notbest2)	0.4048	0.426 (100)
22	Width$_\phi$(light1)	0.3610	0.861 (100)
23	p_T(lepton)	0.3339	0.141 (250)
24	Width$_\eta$(jet2)	0.3278	0.335 (100)
25	Q(lepton)$\times\eta$(best1)	0.3271	0.824 (80)
26	Width$_\phi$(light2)	0.3119	0.899 (100)
27	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.2149	0.060 (80)

Table E.5 : Set of selected variables, ordered by their importance, for the e+jets 2-tags 3-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 e+jets 2-tags 4-jets

Rank	Variable	Importance	K-S (N of bins)
1	M(alljets−tag1)	1.0000	0.137 (500)
2	M(alljets−best1)	0.9449	0.117 (500)
3	Centrality(alljets)	0.8646	0.105 (25)
4	Width$_\eta$(jet4)	0.6309	0.990 (100)
5	ΔR(jet1,jet2)	0.6174	0.850 (50)
6	M(alljets)	0.6139	0.190 (1000)
7	Width$_\phi$(jet4)	0.5345	0.280 (100)
8	Width$_\eta$(light2)	0.5192	0.995 (100)
9	H_T(alljets−tag1)	0.4748	0.627 (500)
10	H_T(lepton,$\not\!\!E_T$,alljets)	0.4475	0.355 (500)
11	Width$_\phi$(light2)	0.4247	0.921 (100)
12	p_T(jet4)	0.4214	1.000 (250)
13	Width$_\eta$(tag2)	0.4156	0.998 (100)
14	$M_T(W)$	0.4130	0.981 (300)
15	p_T(light2)	0.3949	0.900 (250)
16	p_T(notbest2)	0.3607	0.862 (250)
17	cos(lepton$_{\rm btaggedtop}$,btaggedtop$_{\rm CMframe}$)	0.3322	0.696 (100)
18	H_T(lepton,$\not\!\!E_T$)	0.3229	0.725 (500)
19	Q(lepton)×η(light1)	0.3098	0.017 (80)
20	Width$_\eta$(tag1)	0.2894	0.319 (100)
21	cos(lepton$_{\rm besttop}$,besttop$_{\rm CMframe}$)	0.2759	0.293 (100)
22	M(jet3,jet4)	0.2612	0.275 (500)
23	Q(lepton)×η(jet1)	0.2484	0.339 (80)
24	Width$_\eta$(notbest2)	0.2203	0.990 (100)
25	Aplanarity(W,alljets)	0.1966	0.748 (50)
26	H_T(alljets−best1)	0.1919	0.820 (500)
27	M(W,tag2)	0.1673	0.770 (500)

Table E.6 : Set of selected variables, ordered by their importance, for the e+jets 2-tags 4-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 μ+jets 1-tag 2-jets

Rank	Variable	Importance	K-S (N of bins)
1	$\cos(\text{lepton,light1})_{\text{btaggedtop}}$	1.0000	0.868 (100)
2	$M(W,\text{tag1})$	0.8574	0.497 (500)
3	$M_T(W)$	0.8335	0.897 (300)
4	$M(\text{jet1,jet2})$	0.7870	0.196 (500)
5	$\Delta M_{\text{top}}^{\text{min}}$	0.7196	0.856 (5000)
6	$H_T(\text{lepton},\not\!\!E_T,\text{jet1,jet2})$	0.6198	0.046 (500)
7	$H_T(\text{lepton},\not\!\!E_T)$	0.5908	0.596 (500)
8	$\cos(\text{lepton,tag1})_{\text{btaggedtop}}$	0.5860	0.598 (100)
9	$Q(\text{lepton})\times\eta(\text{light1})$	0.5327	0.960 (80)
10	$H_T(\text{jet1,jet2})$	0.4892	0.039 (500)
11	$\Delta\phi(\text{lepton},\not\!\!E_T)$	0.4711	0.391 (80)
12	$M(W,\text{tag1},S2)$	0.4497	0.618 (500)
13	$p_T(\text{jet2})$	0.4367	0.022 (250)
14	$E(\text{light1})$	0.3823	0.497 (250)
15	$\text{Width}_\phi(\text{jet2})$	0.3754	0.948 (100)
16	$\not\!\!E_T$	0.3641	0.712 (250)
17	$\Delta R(\text{lepton,jet1})$	0.3188	0.876 (50)
18	$p_T^{\text{rel}}(\text{jet1},\mu)$	0.2468	1.000 (100)
19	$M(W,\text{light1},S2)$	0.1914	0.288 (500)
20	$\text{Width}_\eta(\text{tag1})$	0.1791	0.738 (100)
21	$p_T(\text{jet1,jet2})$	0.1688	0.676 (500)
22	$\text{Width}_\eta(\text{jet2})$	0.1114	0.958 (100)
23	$p_T(\text{lepton})$	0.1103	0.800 (250)

Table E.7 : Set of selected variables, ordered by their importance, for the μ+jets 1-tag 2-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 μ+jets 1-tag 3-jets

Rank	Variable	Importance	K-S (N of bins)
1	$M_T(W)$	1.0000	0.948 (300)
2	$M(W$,tag1)	0.8662	0.166 (500)
3	H_T(alljets)	0.7336	0.062 (800)
4	$\cos(\text{lepton}_{\text{btaggedtop}},\text{btaggedtop}_{\text{CMframe}})$	0.6023	0.321 (100)
5	$M(W$,jet2)	0.5839	0.830 (500)
6	$\Delta\phi(\text{lepton},\not\!\!E_T)$	0.5701	0.596 (80)
7	$Q(\text{lepton})\times\eta(\text{light1})$	0.5615	0.200 (80)
8	$H_T(\text{lepton},\not\!\!E_T,\text{alljets})$	0.5120	0.099 (500)
9	$M(W$,tag1,$S2)$	0.5087	0.653 (500)
10	Width_η(tag1)	0.5015	0.878 (100)
11	M(alljets)	0.4946	0.894 (1000)
12	ΔR(lepton,tag1)	0.4870	0.260 (50)
13	Width_ϕ(best1)	0.4839	0.990 (100)
14	M(jet1,jet2)	0.4830	0.725 (500)
15	Width_η(light2)	0.4506	0.332 (100)
16	Width_η(jet2)	0.4343	0.661 (100)
17	$\Delta M_{\text{top}}^{\text{min}}$	0.4257	0.436 (5000)
18	$H_T(\text{lepton},\not\!\!E_T)$	0.4136	0.055 (500)
19	Centrality(alljets)	0.4028	0.999 (25)
20	$\cos(\text{lepton}_{\text{besttop}},\text{besttop}_{\text{CMframe}})$	0.4018	0.166 (100)
21	p_T(light2)	0.3908	0.365 (250)
22	$M(W$,jet1)	0.3793	0.217 (500)
23	$\cos(\text{lepton},\text{best1})_{\text{besttop}}$	0.3317	0.021 (100)
24	p_T^{rel}(jet1,μ)	0.2218	1.000 (100)

Table E.8 : Set of selected variables, ordered by their importance, for the μ+jets 1-tag 3-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 μ+jets 1-tag 4-jets

Rank	Variable	Importance	K-S (N of bins)
1	M(alljets$-$tag1)	1.0000	0.871 (500)
2	Centrality(alljets)	0.9316	0.237 (25)
3	p_T(light2)	0.7046	0.886 (250)
4	Width$_\eta$(jet4)	0.5519	1.000 (100)
5	$M_T(W)$	0.4845	0.710 (300)
6	cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	0.4789	0.441 (100)
7	Q(lepton)$\times\eta$(light1)	0.4707	0.845 (80)
8	M(W,tag1)	0.4375	0.285 (500)
9	M(W,tag1,$S2$)	0.4040	0.183 (500)
10	p_T(jet4)	0.3932	0.059 (250)
11	H_T(alljets)	0.3748	0.326 (800)
12	M(alljets)	0.3487	0.594 (1000)
13	M(W,light1,$S2$)	0.3392	0.010 (500)
14	M(jet1,jet2)	0.3275	0.075 (500)
15	p_T(notbest2)	0.3066	0.987 (250)
16	H_T(alljets$-$best1)	0.2580	0.611 (500)
17	Width$_\phi$(jet2)	0.2569	0.385 (100)
18	cos(lepton,tag1)$_{\text{btaggedtop}}$	0.2532	0.576 (100)
19	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.2492	0.463 (80)
20	Width$_\eta$(light2)	0.2400	0.479 (100)
21	H_T(lepton,$\not\!\!E_T$,jet1,jet2)	0.2364	0.137 (500)
22	M(W,jet4)	0.2007	0.145 (500)
23	M(alljets$-$best1)	0.1998	0.748 (500)
24	Width$_\phi$(light2)	0.1978	0.876 (100)
25	Sphericity(W,alljets)	0.1954	0.631 (25)
26	$\Delta\phi$(lepton,tag1)	0.1593	0.611 (80)
27	M(W,jet3,$S2$)	0.1346	0.427 (500)
28	Q(lepton)$\times\eta$(light2)	0.1311	0.832 (80)

Table E.9 : Set of selected variables, ordered by their importance, for the μ+jets 1-tag 4-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 μ+jets 2-tags 2-jets

Rank	Variable	Importance	K-S (N of bins)
1	$M_T(W)$	1.0000	0.278 (300)
2	$\Delta M_{\text{top}}^{\text{min}}$	0.9719	0.740 (5000)
3	M(jet1,jet2)	0.6920	0.954 (1000)
4	H_T(jet1,jet2)	0.5179	0.771 (500)
5	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.4430	0.209 (80)
6	Q(lepton)$\times\eta$(best1)	0.4339	0.991 (80)
7	p_T(lepton)	0.4178	0.110 (250)
8	Width$_\eta$(jet1)	0.3554	0.976 (100)
9	cos(lepton$_{\text{besttop}}$,besttop$_{\text{CMframe}}$)	0.3553	0.437 (100)
10	H_T(lepton,$\not\!\!E_T$,jet1,jet2)	0.3342	0.880 (500)
11	$\sqrt{\hat{s}}$	0.3116	0.976 (1000)
12	ΔR(lepton,tag1)	0.3015	0.042 (50)
13	$\not\!\!E_T$	0.2878	0.826 (250)
14	Width$_\phi$(notbest1)	0.2869	0.237 (100)
15	$M(W$,tag1,$S2)$	0.2712	0.419 (500)
16	p_T(best1)	0.2710	0.340 (250)
17	ΔR^{min}(lepton,alljets)	0.2602	0.321 (50)
18	Q(lepton)$\times\eta$(notbest1)	0.2445	0.925 (80)
19	cos(lepton,best1)$_{\text{besttop}}$	0.2414	0.035 (100)
20	$M(W$,best1)	0.2246	0.571 (500)
21	p_T(tag2)	0.2197	0.284 (250)
22	ΔR^{min}(alljets)	0.1872	0.929 (50)
23	$M_{\text{top}}^{\Delta M^{\text{min}}}$	0.1614	0.813 (500)
24	Width$_\phi$(best1)	0.1493	0.200 (100)
25	cos(lepton,Q(lepton)$\times z)_{\text{besttop}}$	0.1212	0.840 (100)
26	Significance$_{\text{min}}(M_{\text{top}})$	0.1147	0.883 (1000)

Table E.10 : Set of selected variables, ordered by their importance, for the μ+jets 2-tags 2-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 μ+jets 2-tags 3-jets

Rank	Variable	Importance	K-S (N of bins)
1	p_T(notbest2)	1.0000	0.968 (250)
2	Centrality(alljets)	0.9929	0.761 (25)
3	Width$_\eta$(light2)	0.9007	0.442 (100)
4	$M_T(W)$	0.8736	0.889 (300)
5	M(jet1)	0.7754	0.747 (100)
6	$M(W$,tag1)	0.7009	0.281 (500)
7	H_T(lepton,$\not\!\!E_T$,alljets)	0.6942	0.708 (500)
8	ΔR(lepton,tag1)	0.6755	0.991 (50)
9	M(alljets)	0.6541	0.957 (1000)
10	cos(lepton$_{\rm btaggedtop}$,btaggedtop$_{\rm CMframe}$)	0.6166	0.325 (100)
11	Width$_\phi$(light1)	0.5919	0.926 (100)
12	$M(W$,tag2)	0.5423	0.534 (500)
13	H_T(alljets$-$tag1)	0.5334	0.899 (500)
14	Width$_\eta$(tag2)	0.5164	0.313 (100)
15	M(alljets$-$tag1)	0.5050	0.929 (500)
16	M(jet1,jet2)	0.5011	0.727 (500)
17	Width$_\eta$(tag1)	0.4995	0.443 (100)
18	M(alljets$-$best1)	0.4979	0.972 (500)
19	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.4347	0.958 (80)
20	$\Delta M_{\rm top}^{\rm min}$	0.4047	0.786 (5000)
21	Width$_\eta$(notbest2)	0.2811	0.998 (100)
22	p_T(lepton)	0.2778	0.868 (250)
23	Q(lepton)$\times\eta$(light1)	0.2714	0.018 (80)
24	Width$_\phi$(light2)	0.2503	0.465 (100)

Table E.11 : Set of selected variables, ordered by their importance, for the μ+jets 2-tags 3-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p17 μ+jets 2-tags 4-jets

Rank	Variable	Importance	K-S (N of bins)
1	Centrality(alljets)	1.0000	0.854 (25)
2	M(alljets−tag1)	0.9044	0.858 (500)
3	H_T(alljets−tag1)	0.7015	0.992 (500)
4	Width$_\eta$(jet4)	0.6837	0.666 (100)
5	Width$_\eta$(notbest2)	0.6572	0.748 (100)
6	p_T(notbest2)	0.6300	0.829 (250)
7	M(alljets−best1)	0.5827	0.282 (500)
8	p_T(jet4)	0.5490	0.088 (250)
9	cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	0.5245	0.225 (100)
10	Q(lepton)×η(light1)	0.4822	0.639 (80)
11	H_T(lepton,$\not\!\!E_T$,alljets)	0.4407	0.885 (500)
12	H_T(lepton,$\not\!\!E_T$)	0.3954	0.464 (500)
13	M(light1,light2)	0.3543	0.943 (500)
14	Width$_\phi$(light2)	0.3411	0.864 (100)
15	M(jet3,jet4)	0.3388	0.484 (500)
16	M(W,jet4)	0.3383	0.448 (500)
17	M(alljets)	0.3299	0.535 (1000)
18	M(jet1,jet2)	0.2992	0.845 (500)
19	Width$_\phi$(tag2)	0.2877	0.247 (100)
20	Width$_\phi$(jet4)	0.2840	0.999 (100)
21	cos(lepton,tag1)$_{\text{btaggedtop}}$	0.2739	0.346 (100)
22	Q(lepton)×η(best1)	0.2574	0.953 (80)
23	p_T(lepton)	0.2214	0.241 (250)
24	$\Delta M_{\text{top}}^{\text{min}}$	0.2101	0.885 (5000)
25	p_T(jet3)	0.1742	0.764 (250)
26	M(W,light1)	0.1304	0.938 (500)
27	M_T(W)	0.1268	0.374 (300)

Table E.12 : Set of selected variables, ordered by their importance, for the μ+jets 2-tags 4-jets channel in p17 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 e+jets 1-tag 2-jets

Rank	Variable	Importance	K-S (N of bins)
1	M(jet1,jet2)	1.0000	0.148 (1000)
2	$\Delta M_{\rm top}^{\rm min}$	0.7551	0.551 (5000)
3	$M(W$,tag1)	0.6603	0.007 (500)
4	$M(W$,tag1,$S2)$	0.5785	0.014 (500)
5	H_T(lepton,$\not\!\!E_T$,jet1,jet2)	0.5638	0.017 (500)
6	$M_T(W)$	0.4784	0.729 (300)
7	H_T(jet1,jet2)	0.4742	0.016 (500)
8	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.4056	0.412 (80)
9	Q(lepton)$\times\eta$(light1)	0.4000	0.860 (80)
10	$\not\!\!E_T$	0.3566	0.245 (250)
11	cos(light1,lepton)$_{\rm btaggedtop}$	0.3397	0.118 (100)
12	p_T(jet2)	0.3308	0.428 (250)
13	E(light1)	0.2944	0.676 (250)
14	$p_T^{\rm rel}$(jet1,μ)	0.2923	1.000 (100)
15	cos(lepton,jet1)$_{\rm btaggedtop}$	0.2436	0.740 (100)
16	Width$_\phi$(jet2)	0.2057	0.650 (100)
17	cos(lepton,best1)$_{\rm besttop}$	0.1888	0.823 (100)
18	Width$_\eta$(jet2)	0.1657	0.345 (100)

Table E.13 : Set of selected variables, ordered by their importance, for the e+jets 1-tag 2-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 e+jets 1-tag 3-jets

Rank	Variable	Importance	K-S (N of bins)
1	M(alljets)	1.0000	0.920 (1000)
2	H_T(lepton,$\not\!\!E_T$,alljets)	0.9116	0.180 (500)
3	M(W,tag1)	0.8451	0.682 (500)
4	Q(lepton)$\times\eta$(light1)	0.8102	0.555 (80)
5	cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	0.7678	0.995 (100)
6	H_T(jet1,jet2)	0.6736	0.015 (500)
7	$M_T(W)$	0.6705	0.164 (300)
8	M(W,tag1,$S2$)	0.6612	0.640 (500)
9	Width$_\phi$(light1)	0.6077	0.796 (100)
10	cos(lepton,tag1)$_{\text{btaggedtop}}$	0.5917	0.796 (100)
11	M(light1,light2)	0.5418	0.995 (500)
12	$\Delta M_{\text{top}}^{\text{min}}$	0.5162	0.988 (5000)
13	Centrality(alljets)	0.5138	0.037 (25)
14	Width$_\eta$(jet2)	0.4887	0.532 (100)
15	p_T(notbest2)	0.4751	0.157 (250)
16	H_T(alljets$-$tag1)	0.4717	0.005 (500)
17	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.4245	0.673 (80)
18	Width$_\phi$(light2)	0.4114	0.943 (100)
19	M(W,jet1)	0.2155	0.779 (500)
20	Width$_\eta$(tag1)	0.2099	0.687 (100)

Table E.14 : Set of selected variables, ordered by their importance, for the e+jets 1-tag 3-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 e+jets 1-tag 4-jets

Rank	Variable	Importance	K-S (N of bins)
1	Centrality(alljets)	1.0000	0.859 (25)
2	M(alljets−tag1)	0.9930	0.850 (500)
3	$M_T(W)$	0.7847	0.669 (300)
4	H_T(lepton,$\not\!\!E_T$,alljets)	0.7716	0.533 (500)
5	p_T(light2)	0.6802	0.934 (250)
6	$M(W$,tag1)	0.6144	0.597 (500)
7	Width$_\phi$(light2)	0.6022	0.586 (100)
8	$M(W$,tag1,$S2)$	0.5943	0.447 (500)
9	Q(lepton)×η(light1)	0.5899	0.899 (80)
10	H_T(jet1,jet2)	0.5489	0.840 (500)
11	$\not\!\!E_T$	0.5281	0.251 (250)
12	M(alljets)	0.5169	0.660 (1000)
13	ΔR(jet1,lepton)	0.5097	0.949 (50)
14	p_T(jet4)	0.4652	0.919 (250)
15	cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	0.4480	0.318 (100)
16	M(alljets−best1)	0.4355	0.902 (500)
17	Width$_\eta$(jet4)	0.4155	0.288 (100)
18	cos(lepton,best1)$_{\text{besttop}}$	0.4042	0.334 (100)
19	p_T(notbest2)	0.4036	0.403 (250)
20	M(light1,light2)	0.3557	0.891 (500)
21	Width$_\eta$(tag1)	0.3470	0.885 (100)
22	Width$_\phi$(jet4)	0.3384	0.240 (100)
23	$M(W$,jet2,$S2)$	0.3095	0.266 (500)
24	Width$_\eta$(jet2)	0.3024	0.200 (100)
25	Width$_\eta$(best1)	0.2903	0.397 (100)
26	cos(best1,notbest1)$_{\text{besttop}}$	0.2899	0.075 (100)
27	Q(lepton)×η(light2)	0.1911	0.858 (80)
28	M(jet1,jet2)	0.1361	0.230 (500)

Table E.15 : Set of selected variables, ordered by their importance, for the e+jets 1-tag 4-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 e+jets 2-tags 2-jets

Rank	Variable	Importance	K-S (N of bins)
1	M(jet1,jet2)	1.0000	0.352 (500)
2	$\Delta M_{\text{top}}^{\text{min}}$	0.5523	0.963 (5000)
3	$M_T(W)$	0.4549	0.142 (300)
4	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.3177	0.062 (80)
5	M(W,best1)	0.3054	0.916 (500)
6	H_T(lepton,$\not\!\!E_T$,jet1,jet2)	0.3000	0.026 (500)
7	$\not\!\!E_T$	0.2673	0.532 (250)
8	cos(lepton,Q(lepton)$\times z$)$_{\text{besttop}}$	0.2666	0.834 (100)
9	Width$_\eta$(tag1)	0.2587	0.997 (100)
10	$M_{\text{top}}^{\Delta M^{\text{min}}}$	0.2510	0.998 (500)
11	Q(lepton)$\times\eta$(best1)	0.2509	0.930 (80)
12	$\sqrt{\hat{s}}$	0.2213	0.025 (1000)
13	H_T(jet1,jet2)	0.2194	0.106 (500)
14	p_T(best1)	0.1873	0.419 (250)
15	M(W,notbest1,$S2$)	0.1803	0.071 (500)
16	cos(lepton,best1)$_{\text{besttop}}$	0.1717	0.823 (100)
17	cos(best1,notbest1)$_{\text{besttop}}$	0.1663	0.833 (100)
18	M(W,tag1,$S2$)	0.1622	0.776 (500)
19	Width$_\eta$(tag2)	0.1597	0.548 (100)
20	p_T(lepton)	0.1586	0.398 (250)
21	Significance$_{\text{min}}$(M_{top})	0.1563	0.511 (1000)

Table E.16 : Set of selected variables, ordered by their importance, for the e+jets 2-tags 2-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 e+jets 2-tags 3-jets

Rank	Variable	Importance	K-S (N of bins)
1	Centrality(alljets)	1.0000	0.096 (25)
2	M(alljets)	0.8534	0.165 (1000)
3	$M_T(W)$	0.8329	0.822 (300)
4	M(W,tag1)	0.7501	0.872 (500)
5	Width$_\eta$(light2)	0.6590	0.974 (100)
6	M(alljets$-$best1)	0.6340	0.971 (500)
7	p_T(notbest2)	0.6316	0.557 (250)
8	M(light1,light2)	0.5964	0.541 (500)
9	M(jet1)	0.5333	0.861 (100)
10	M(jet1,jet2)	0.5166	0.498 (500)
11	$\Delta M_{\rm top}^{\rm min}$	0.4823	0.404 (5000)
12	Width$_\phi$(light2)	0.4741	0.999 (100)
13	H_T(lepton,$\not\!\!E_T$)	0.4680	0.164 (500)
14	Q(lepton)$\times\eta$(light1)	0.4576	0.269 (80)
15	Width$_\phi$(light1)	0.4278	0.975 (100)
16	M(W,tag1,$S2$)	0.4192	0.731 (500)
17	H_T(alljets$-$tag1)	0.4028	0.130 (500)
18	M(alljets$-$tag1)	0.3979	0.883 (500)
19	Width$_\phi$(jet1)	0.3359	1.000 (100)
20	cos(lepton$_{\rm besttop}$,besttop$_{\rm CMframe}$)	0.3198	0.556 (100)
21	H_T(lepton,$\not\!\!E_T$,alljets)	0.3182	0.398 (500)
22	Width$_\eta$(tag1)	0.2682	0.992 (100)
23	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.2559	0.942 (80)
24	Width$_\eta$(tag2)	0.2479	0.815 (100)
25	M(W,light1,$S2$)	0.2428	0.970 (500)
26	p_T(light2)	0.2348	0.589 (250)

Table E.17 : Set of selected variables, ordered by their importance, for the e+jets 2-tags 3-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 e+jets 2-tags 4-jets

Rank	Variable	Importance	K-S (N of bins)
1	M(alljets−tag1)	1.0000	0.569 (500)
2	Centrality(alljets)	0.8503	0.041 (25)
3	p_T(light2)	0.6641	0.453 (250)
4	H_T(lepton,$\not\!\!E_T$,alljets)	0.6341	0.750 (500)
5	M(alljets)	0.5389	0.142 (1000)
6	Width$_\phi$(light2)	0.4892	0.748 (100)
7	Width$_\eta$(jet4)	0.4636	0.201 (100)
8	M(alljets−best1)	0.4403	0.284 (500)
9	p_T(notbest2)	0.4401	0.962 (250)
10	Width$_\phi$(tag2)	0.4255	0.250 (100)
11	ΔR(jet1,jet2)	0.4088	0.971 (50)
12	M(jet3,jet4)	0.3911	0.432 (500)
13	M(light1,light2)	0.3801	0.928 (500)
14	H_T(alljets−best1)	0.3794	0.918 (500)
15	p_T(jet4)	0.3671	0.399 (250)
16	Width$_\phi$(jet4)	0.3430	0.037 (100)
17	H_T(alljets−tag1)	0.3201	0.361 (500)
18	M(W,light1,$S2$)	0.3162	0.986 (500)
19	Q(lepton)×η(light1)	0.2979	0.510 (80)
20	M(W,tag1,$S2$)	0.2864	0.805 (500)
21	$M_T(W)$	0.2850	0.730 (300)
22	Width$_\phi$(light1)	0.2591	0.906 (100)
23	$\Delta R^{\min}$(alljets)	0.2131	0.916 (50)
24	cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	0.2121	0.212 (100)
25	M(jet1,jet2)	0.2065	0.829 (500)

Table E.18 : Set of selected variables, ordered by their importance, for the e+jets 2-tags 4-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 μ+jets 1-tag 2-jets

Rank	Variable	Importance	K-S (N of bins)
1	$M(W,\text{tag1})$	1.0000	0.157 (500)
2	$M(\text{jet1,jet2})$	0.8793	0.456 (500)
3	$\Delta M_{\text{top}}^{\text{min}}$	0.8576	0.925 (5000)
4	$\Delta\phi(\text{lepton},\not\!\!E_T)$	0.8155	0.837 (80)
5	$H_T(\text{lepton},\not\!\!E_T)$	0.6414	0.518 (500)
6	$p_T(\text{jet2})$	0.5586	0.980 (250)
7	$M(W,\text{tag1},S2)$	0.5497	0.755 (500)
8	$Q(\text{lepton})\times\eta(\text{light1})$	0.5413	0.976 (80)
9	$H_T(\text{jet1,jet2})$	0.5019	0.106 (500)
10	$M_T(W)$	0.4992	0.304 (300)
11	$\cos(\text{lepton}_{\text{btaggedtop}},\text{btaggedtop}_{\text{CMframe}})$	0.4745	0.348 (100)
12	$\Delta R(\text{lepton,tag1})$	0.4626	0.870 (50)
13	$\Delta\phi(\text{jet1},\not\!\!E_T)$	0.3837	0.797 (80)
14	$\text{Width}_\eta(\text{jet2})$	0.3675	0.365 (100)
15	$p_T^{\text{rel}}(\text{jet1},\mu)$	0.3388	1.000 (100)
16	$H_T(\text{lepton},\not\!\!E_T,\text{jet1,jet2})$	0.3171	0.735 (500)
17	$Q(\text{lepton})\times\eta(\text{best1})$	0.3044	0.581 (80)
18	$E(\text{light1})$	0.2918	0.677 (250)
19	$\cos(\text{lepton,light1})_{\text{btaggedtop}}$	0.2743	0.341 (100)
20	$\not\!\!E_T$	0.2730	0.319 (250)
21	$\text{Width}_\phi(\text{jet2})$	0.2711	0.236 (100)
22	$M(W,\text{best1})$	0.2669	0.910 (500)
23	$\cos(\text{lepton,jet1})_{\text{btaggedtop}}$	0.2014	0.546 (100)

Table E.19 : Set of selected variables, ordered by their importance, for the μ+jets 1-tag 2-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 μ+jets 1-tag 3-jets

Rank	Variable	Importance	K-S (N of bins)
1	H_T(lepton,$\not\!\!E_T$,alljets)	1.0000	0.123 (500)
2	$M_T(W)$	0.9515	0.995 (300)
3	$M(W$,tag1)	0.9095	0.052 (500)
4	M(alljets)	0.7392	0.298 (1000)
5	H_T(lepton,$\not\!\!E_T$)	0.7224	0.795 (500)
6	Q(lepton)$\times\eta$(light1)	0.6922	0.552 (80)
7	M(jet1,jet2)	0.6092	0.100 (500)
8	cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	0.5010	0.865 (100)
9	$M(W$,tag1,$S2)$	0.4926	0.363 (500)
10	cos(lepton,tag1)$_{\text{btaggedtop}}$	0.4667	0.036 (100)
11	p_T(light2)	0.4606	0.679 (250)
12	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.4378	0.330 (80)
13	Centrality(alljets)	0.4156	0.971 (25)
14	H_T(alljets)	0.4006	0.793 (800)
15	Width$_\phi$(jet2)	0.3982	0.712 (100)
16	$M(W$,jet1,$S2)$	0.3677	0.025 (500)
17	$M(W$,best1)	0.2917	0.054 (500)
18	Width$_\eta$(light2)	0.2309	0.143 (100)
19	cos(lepton,light1)$_{\text{btaggedtop}}$	0.2284	0.571 (100)
20	ΔR(lepton,best1)	0.2263	0.688 (50)
21	p_T(best1)	0.2209	0.511 (250)

Table E.20 : Set of selected variables, ordered by their importance, for the μ+jets 1-tag 3-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 μ+jets 1-tag 4-jets

Rank	Variable	Importance	K-S (N of bins)
1	Centrality(alljets)	1.0000	0.675 (25)
2	M(W,tag1)	0.7807	0.003 (500)
3	M(alljets−tag1)	0.7525	0.137 (500)
4	H_T(alljets)	0.7155	0.393 (800)
5	$M_T(W)$	0.6217	0.756 (300)
6	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.5413	0.965 (80)
7	Width$_\eta$(jet4)	0.5120	0.791 (100)
8	Width$_\eta$(light2)	0.4992	0.693 (100)
9	Q(lepton)×η(light1)	0.4420	0.301 (80)
10	ΔR(lepton,tag1)	0.4343	0.603 (50)
11	M(alljets−best1)	0.4305	0.096 (500)
12	M(W,jet1,$S2$)	0.4227	0.636 (500)
13	$p_z(\nu,S2)$	0.4127	0.212 (1000)
14	H_T(alljets−tag1)	0.4078	0.374 (500)
15	M(jet1,jet2)	0.3946	0.429 (500)
16	p_T(notbest2)	0.3932	0.980 (250)
17	H_T(lepton,$\not\!\!E_T$,alljets)	0.3489	0.070 (500)
18	cos(lepton$_{\rm btaggedtop}$,btaggedtop$_{\rm CMframe}$)	0.3308	0.381 (100)
19	Width$_\eta$(jet2)	0.3274	0.866 (100)
20	cos(lepton,light1)$_{\rm btaggedtop}$	0.3140	0.205 (100)
21	Q(lepton)×η(light2)	0.3123	0.663 (80)
22	$\Delta M_{\rm top}^{\rm min}$	0.2951	0.319 (5000)
23	p_T(light2)	0.2830	0.953 (250)
24	p_T(jet4)	0.2313	0.974 (250)
25	M(alljets)	0.1919	0.237 (1000)
26	H_T(lepton,$\not\!\!E_T$,jet1,jet2)	0.1471	0.174 (500)
27	M(light1,light2)	0.1359	0.985 (500)
28	Width$_\phi$(notbest1)	0.1190	0.490 (100)

Table E.21 : Set of selected variables, ordered by their importance, for the μ+jets 1-tag 4-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 μ+jets 2-tags 2-jets

Rank	Variable	Importance	K-S (N of bins)
1	M(jet1,jet2)	1.0000	0.912 (1000)
2	$\Delta M_{\rm top}^{\rm min}$	0.7676	0.276 (5000)
3	$M_T(W)$	0.7126	0.489 (300)
4	p_T(lepton)	0.6406	0.805 (250)
5	H_T(jet1,jet2)	0.6011	0.288 (500)
6	$\Delta\phi$(lepton,$\not\!\!E_T$)	0.5367	0.505 (80)
7	Q(lepton)$\times\eta$(notbest1)	0.5289	0.969 (80)
8	$M_{\rm top}^{\Delta M^{\rm min}}$	0.5272	0.303 (500)
9	Q(lepton)$\times\eta$(best1)	0.4050	0.301 (80)
10	$\Delta R^{\rm min}$(lepton,alljets)	0.3956	0.801 (50)
11	M(W,best1,$S2$)	0.3870	0.153 (500)
12	H_T(lepton$\not\!\!E_T$,jet1,jet2)	0.3490	0.263 (500)
13	ΔR(lepton,tag1)	0.3467	0.591 (50)
14	cos(lepton,best1)$_{\rm besttop}$	0.3262	0.456 (100)
15	M(W,jet1,$S2$)	0.2987	0.492 (500)
16	$\not\!\!E_T$	0.2965	0.169 (250)
17	M(W,tag2)	0.2796	0.254 (500)
18	p_T(best1)	0.2682	0.126 (250)
19	$\sqrt{\hat{s}}$	0.2508	0.760 (1000)
20	M(best1)	0.2444	0.780 (100)
21	M(W,best1)	0.2144	0.204 (500)
22	M_T(jet1,jet2)	0.1602	0.526 (500)
23	cos(lepton,jet1)$_{\rm btaggedtop}$	0.1466	0.632 (100)

Table E.22 : Set of selected variables, ordered by their importance, for the μ+jets 2-tags 2-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 μ+jets 2-tags 3-jets

Rank	Variable	Importance	K-S (N of bins)
1	p_T(notbest2)	1.0000	0.441 (250)
2	M(alljets)	0.8958	0.345 (1000)
3	$M(W$,tag1)	0.7705	0.925 (500)
4	Centrality(alljets)	0.7474	0.790 (25)
5	$M_T(W)$	0.7138	0.247 (300)
6	M(alljets$-$tag1)	0.7019	0.782 (500)
7	Q(lepton)$\times\eta$(tag2)	0.6080	0.098 (80)
8	$M(W$,tag1,$S2)$	0.5546	0.422 (500)
9	$\Delta M^{\rm min}_{\rm top}$	0.5249	0.929 (5000)
10	Q(lepton)$\times\eta$(light1)	0.5229	0.979 (80)
11	Width$_\eta$(light2)	0.5023	1.000 (100)
12	Width$_\phi$(light2)	0.4999	0.947 (100)
13	H_T(lepton,$\not\!\!E_T$)	0.4826	0.452 (500)
14	cos(lepton,tag1)$_{\rm btaggedtop}$	0.4692	0.437 (100)
15	$M(W$,tag2)	0.4596	0.494 (500)
16	H_T(alljets$-$tag1)	0.4590	0.394 (500)
17	ΔR(jet1,jet2)	0.4463	0.993 (50)
18	H_T(alljets)	0.4366	0.498 (800)
19	Q(lepton)$\times\eta$(light2)	0.3851	0.119 (80)
20	ΔR(lepton,tag1)	0.3826	0.390 (50)
21	M(alljets$-$best1)	0.3798	0.677 (500)
22	M(jet1,jet2)	0.3241	0.232 (500)
23	Width$_\phi$(jet1)	0.3086	0.520 (100)
24	M(jet1)	0.2523	0.728 (100)
25	H_T(lepton,$\not\!\!E_T$,alljets)	0.2492	0.929 (500)
26	cos(lepton$_{\rm btaggedtop}$,btaggedtop$_{\rm CMframe}$)	0.2324	0.224 (100)

Table E.23 : Set of selected variables, ordered by their importance, for the μ+jets 2-tags 3-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channel: p20 μ+jets 2-tags 4-jets

Rank	Variable	Importance	K-S (N of bins)
1	Centrality(alljets)	1.0000	0.876 (25)
2	p_T(notbest2)	0.9001	0.004 (250)
3	M(alljets−tag1)	0.8225	0.516 (500)
4	M(alljets−best1)	0.5262	0.115 (500)
5	Width$_\eta$(jet4)	0.4602	0.595 (100)
6	Width$_\phi$(light2)	0.4374	1.000 (100)
7	H_T(lepton,$\not\!\!E_T$,alljets)	0.4250	0.031 (500)
8	ΔR(jet1,jet2)	0.3983	0.383 (50)
9	M(jet3,jet4)	0.3625	0.879 (500)
10	Q(lepton)×η(light1)	0.3562	0.526 (80)
11	cos(lepton$_{\rm btaggedtop}$,btaggedtop$_{\rm CMframe}$)	0.3501	0.155 (100)
12	Width$_\phi$(tag2)	0.3239	0.811 (100)
13	Q(lepton)×η(light2)	0.3218	0.810 (80)
14	p_T(lepton)	0.3192	0.112 (250)
15	M(W,jet1,$S2$)	0.2581	0.962 (500)
16	p_T(jet4)	0.2495	0.752 (250)
17	Width$_\phi$(jet4)	0.2460	0.716 (100)
18	Width$_\eta$(light2)	0.2438	1.000 (100)
19	Width$_\phi$(light1)	0.2180	0.521 (100)
20	$\Delta M_{\rm top}^{\rm min}$	0.2147	0.411 (5000)
21	M(jet1,jet2)	0.2111	0.599 (500)
22	$\Delta R^{\rm min}$(alljets)	0.2005	0.998 (50)
23	p_T(light2)	0.1967	0.822 (250)
24	M(alljets)	0.1727	0.793 (1000)
25	H_T(alljets−tag1)	0.1700	0.592 (500)
26	M(light1,light2)	0.1635	0.212 (500)
27	$\Delta R^{\rm min}$(lepton,alljets)	0.1279	0.832 (50)

Table E.24 : Set of selected variables, ordered by their importance, for the μ+jets 2-tags 4-jets channel in p20 data. For the angular variables, the subscript indicates the reference frame. A K-S test value for each variable is also shown in the right most column, and the number of histograming bins used to get this value is given in parentheses.

Analysis Channels: 1-tag 2-jets

Object kinematics			
$\not\!\!E_T$	$e1/0.57$ $\mu1/0.36$ $e2/0.36$ $\mu2/0.27$	p_T(jet1)	$e1/0.36$
p_T(jet2)	$\mu2/0.56$ $\mu1/0.44$ $e2/0.33$ $e1/0.25$	E(jet2)	$e1/0.19$
Q(lepton)$\times\eta$(light1)	$\mu2/0.54$ $\mu1/0.53$ $e1/0.48$ $e2/0.40$	p_T(best1)	$e1/0.17$
E(light1)	$\mu1/0.38$ $e2/0.29$ $\mu2/0.29$	p_T(lepton)	$\mu1/0.11$
Q(lepton)$\times\eta$(best1)	$\mu2/0.30$ $e1/0.24$		
Event kinematics			
M(jet1,jet2)	$e2/1.00$ $e1/1.00$ $\mu2/0.88$ $\mu1/0.79$	p_T(jet1,jet2)	$\mu1/0.17$
$M_T(W)$	$\mu1/0.83$ $e1/0.81$ $\mu2/0.50$ $e2/0.48$		
H_T(lepton$\not\!\!E_T$,jet1,jet2)	$e1/0.73$ $\mu1/0.62$ $e2/0.56$ $\mu2/0.32$		
H_T(jet1,jet2)	$\mu2/0.50$ $\mu1/0.49$ $e2/0.47$		
H_T(lepton$\not\!\!E_T$)	$\mu2/0.64$ $\mu1/0.59$		
Top quark reconstruction			
$M(W$,tag1)	$\mu2/1.00$ $\mu1/0.86$ $e1/0.80$ $e2/0.66$	$M(W$,best1)	$\mu2/0.27$
$\Delta M_{\text{top}}^{\text{min}}$	$\mu2/0.86$ $e1/0.78$ $e2/0.76$ $\mu1/0.72$	$M(W$,light1,$S2)$	$\mu1/0.19$
$M(W$,tag1,$S2)$	$e1/0.60$ $e2/0.58$ $\mu2/0.55$ $\mu1/0.45$		
Angular correlations			
cos(lepton,light1)$_{\text{btaggedtop}}$	$\mu1/1.00$ $e2/0.34$ $\mu2/0.27$ $e1/0.12$	ΔR(lepton,tag1)	$\mu2/0.46$
$\Delta\phi$(lepton,$\not\!\!E_T$)	$\mu2/0.82$ $\mu1/0.47$ $e2/0.41$ $e1/0.30$	$\Delta\phi$(jet1,$\not\!\!E_T$)	$\mu2/0.38$
cos(lepton,tag1)$_{\text{btaggedtop}}$	$\mu1/0.59$ $e1/0.40$	ΔR(lepton,jet1)	$\mu1/0.32$
cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	$\mu2/0.47$ $e1/0.43$	cos(lepton,best1)$_{\text{besttop}}$	$e2/0.19$
cos(lepton,jet1)$_{\text{btaggedtop}}$	$e2/0.24$ $\mu2/0.20$	cos(lepton,Q(lepton)$\times z)_{\text{besttop}}$	$e1/0.11$
Jet reconstruction			
Width$_\phi$(jet2)	$\mu1/0.38$ $\mu2/0.27$ $e1/0.25$ $e2/0.21$	Width$_\eta$(tag1)	$\mu1/0.18$
Width$_\eta$(jet2)	$\mu2/0.37$ $e1/0.34$ $e2/0.17$ $\mu1/0.11$		
p_T^{rel}(jet1,μ)	$\mu2/0.34$ $e2/0.29$ $\mu1/0.25$ $e1/0.16$		

Table E.25 : Set of selected variables for the 1-tag 2-jets channels. For the angular variables, the subscript indicates the reference frame. For each variable, it is shown to its right the list of channels (reco versions and lepton types) in which the variable is used for the BNN training, together with its importance in each of the channels. (See main text for abbreviations convention.)

Analysis Channels: 1-tag 3-jets

Object kinematics			
Q(lepton)$\times\eta$(light1)	$e2/0.81$ $\mu2/0.69$ $e1/0.68$ $\mu1/0.56$	p_T(notbest2)	$e2/0.48$
p_T(light2)	$e1/0.47$ $\mu2/0.46$ $\mu1/0.39$	p_T(best1)	$\mu2/0.22$
Event kinematics			
$M_T(W)$	$\mu1/1.00$ $e1/1.00$ $\mu2/0.95$ $e2/0.67$	H_T(jet1,jet2)	$e2/0.67$
H_T(lepton,$\not\!\!E_T$,alljets)	$\mu2/1.00$ $e2/0.91$ $e1/0.71$ $\mu1/0.51$	H_T(alljets$-$tag1)	$e2/0.47$
M(alljets)	$e2/1.00$ $\mu2/0.74$ $\mu1/0.49$ $e1/0.39$		
Centrality(alljets)	$e1/0.52$ $e2/0.51$ $\mu2/0.42$ $\mu1/0.40$		
H_T(alljets)	$e1/0.80$ $\mu1/0.73$ $\mu2/0.40$		
M(jet1,jet2)	$e1/0.75$ $\mu2/0.61$ $\mu1/0.48$		
H_T(lepton,$\not\!\!E_T$)	$\mu2/0.72$ $e1/0.54$ $\mu1/0.41$		
M(light1,light2)	$e2/0.54$ $e1/0.40$		
Top quark reconstruction			
$M(W$,tag1)	$\mu2/0.91$ $\mu1/0.87$ $e2/0.85$ $e1/0.65$	$M(W$,jet2)	$\mu1/0.58$
$M(W$,tag1,$S2)$	$e1/0.84$ $e2/0.66$ $\mu1/0.51$ $\mu2/0.49$	$M(W$,jet1,$S2)$	$\mu2/0.37$
$\Delta M_{\text{top}}^{\text{min}}$	$e1/0.87$ $e2/0.52$ $\mu1/0.43$	$M(W$,jet2,$S2)$	$e1/0.32$
$M(W$,jet1)	$\mu1/0.38$ $e2/0.22$	$M(W$,best1)	$\mu2/0.29$
Angular correlations			
cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	$e2/0.77$ $\mu1/0.60$ $\mu2/0.50$ $e1/0.33$	ΔR(lepton,tag1)	$\mu1/0.49$
$\Delta\phi$(lepton,$\not\!\!E_T$)	$\mu1/0.57$ $e1/0.47$ $\mu2/0.44$ $e2/0.42$	ΔR(lepton,best1)	$\mu2/0.23$
cos(lepton$_{\text{besttop}}$,besttop$_{\text{CMframe}}$)	$e1/0.67$ $\mu1/0.40$	cos(lepton,light1)$_{\text{btaggedtop}}$	$\mu2/0.23$
cos(lepton,tag1)$_{\text{btaggedtop}}$	$e2/0.59$ $\mu2/0.47$		
cos(lepton,best1)$_{\text{besttop}}$	$e1/0.57$ $\mu1/0.33$		
Jet reconstruction			
Width$_\eta$(tag1)	$e1/0.61$ $\mu1/0.50$ $e2/0.21$	Width$_\phi$(light1)	$e2/0.61$
Width$_\eta$(light2)	$\mu1/0.45$ $e1/0.34$ $\mu2/0.23$	Width$_\phi$(best1)	$\mu1/0.48$
Width$_\eta$(jet2)	$e2/0.49$ $\mu1/0.43$	Width$_\phi$(notbest1)	$e1/0.41$
Width$_\phi$(light2)	$e2/0.41$ $e1/0.26$	Width$_\phi$(jet2)	$\mu2/0.40$
p_T^{rel}(jet1,μ)	$e1/0.26$ $\mu1/0.22$		

Table E.26 : Set of selected variables for the 1-tag 3-jets channels. For the angular variables, the subscript indicates the reference frame. For each variable, it is shown to its right the list of channels (reco versions and lepton types) in which the variable is used for the BNN training, together with its importance in each of the channels. (See main text for abbreviations convention.)

Object kinematics			
p_T(notbest2)	$e1/0.72$ $e2/0.40$ $\mu2/0.39$ $\mu1/0.31$	Q(lepton)$\times\eta$(best1)	$e1/0.43$
p_T(jet4)	$e1/0.67$ $e2/0.47$ $\mu1/0.39$ $\mu2/0.23$	p_T(jet3)	$e1/0.15$
Q(lepton)$\times\eta$(light1)	$e1/0.63$ $e2/0.59$ $\mu1/0.47$ $\mu2/0.44$		
p_T(light2)	$\mu1/0.70$ $e2/0.68$ $\mu2/0.28$		
Q(lepton)$\times\eta$(light2)	$\mu2/0.31$ $e2/0.19$ $\mu1/0.13$		
$\not\!\!E_T$	$e2/0.53$ $e1/0.21$		
Event kinematics			
Centrality(alljets)	$\mu2/1.00$ $e2/1.00$ $e1/1.00$ $\mu1/0.93$	H_T(jet1,jet2)	$e2/0.55$
M(alljets$-$tag1)	$\mu1/1.00$ $e2/0.99$ $\mu2/0.75$ $e1/0.73$	H_T(alljets$-$tag1)	$\mu2/0.41$
$M_T(W)$	$e2/0.78$ $e1/0.70$ $\mu2/0.62$ $\mu1/0.48$	Sphericity(W,alljets)	$\mu1/0.20$
M(alljets)	$e1/0.65$ $e2/0.52$ $\mu1/0.35$ $\mu2/0.19$		
M(alljets$-$best1)	$e1/0.48$ $e2/0.44$ $\mu2/0.43$ $\mu1/0.20$		
H_T(lepton$\not\!\!E_T$,alljets)	$e2/0.77$ $e1/0.55$ $\mu2/0.35$		
H_T(alljets)	$\mu2/0.72$ $\mu1/0.37$ $e1/0.24$		
M(jet1,jet2)	$\mu2/0.39$ $\mu1/0.33$ $e2/0.14$		
M(light1,light2)	$e2/0.36$ $\mu2/0.14$		
H_T(alljets$-$best1)	$e1/0.27$ $\mu1/0.26$		
H_T(lepton$\not\!\!E_T$,jet1,jet2)	$\mu1/0.24$ $\mu2/0.15$		
Top quark reconstruction			
$M(W$,tag1)	$\mu2/0.78$ $e2/0.61$ $e1/0.61$ $\mu1/0.44$	$p_z(\nu,S2)$	$\mu2/0.41$
$M(W$,tag1,$S2)$	$e2/0.59$ $\mu1/0.40$	$M(W$,light1,$S2)$	$\mu1/0.34$
$M(W$,jet1,$S2)$	$\mu2/0.42$ $e1/0.24$	$M(W$,jet2,$S2)$	$e2/0.31$
		$\Delta M_{\rm top}^{\rm min}$	$\mu2/0.30$
		$M(W$,jet4)	$\mu1/0.20$
		$M(W$,jet3,$S2)$	$\mu1/0.13$
Angular correlations			
cos(lepton$_{\rm btaggedtop}$,btaggedtop$_{\rm CMframe}$)	$\mu1/0.48$ $e2/0.45$ $e1/0.43$ $\mu2/0.33$	ΔR(lepton,jet1)	$e2/0.51$
$\Delta\phi$(lepton$\not\!\!E_T$)	$\mu2/0.54$ $\mu1/0.25$	ΔR(lepton,tag1)	$\mu2/0.43$
cos(lepton,tag1)$_{\rm btaggedtop}$	$e1/0.29$ $\mu1/0.25$	cos(lepton,best1)$_{\rm besttop}$	$e2/0.40$
		cos(lepton,light1)$_{\rm btaggedtop}$	$\mu2/0.31$
		ΔR(jet1,jet2)	$e1/0.31$
		cos(best1,notbest1)$_{\rm besttop}$	$e2/0.29$
		cos(lepton$_{\rm besttop}$,besttop$_{\rm CMframe}$)	$e1/0.26$
		$\Delta\phi$(lepton,tag1)	$\mu1/0.16$
Jet reconstruction			
Width$_\eta$(jet4)	$\mu1/0.55$ $\mu2/0.51$ $e2/0.42$ $e1/0.28$	Width$_\eta$(best1)	$e2/0.29$
Width$_\phi$(light2)	$e2/0.60$ $e1/0.33$ $\mu1/0.20$	Width$_\phi$(jet2)	$\mu1/0.26$
Width$_\eta$(light2)	$\mu2/0.50$ $e1/0.43$ $\mu1/0.24$	Width$_\phi$(notbest1)	$\mu2/0.12$
Width$_\eta$(tag1)	$e2/0.35$ $e1/0.34$		
Width$_\phi$(jet4)	$e2/0.34$ $e1/0.25$		
Width$_\eta$(jet2)	$\mu2/0.33$ $e2/0.30$		

Table E.27 : Set of selected variables for the 1-tag 4-jets channels. For the angular variables, the subscript indicates the reference frame. For each variable, it is shown to its right the list of channels (reco versions and lepton types) in which the variable is used for the BNN training, together with its importance in each of the channels. (See main text for abbreviations convention.)

Analysis Channels: 2-tags 2-jets

Object kinematics			
p_T(best1)	$e1/0.46$ $\mu1/0.27$ $\mu2/0.27$ $e2/0.19$	p_T(tag2)	$\mu1/0.22$
Q(lepton)$\times\eta$(best1)	$\mu1/0.43$ $e1/0.43$ $\mu2/0.41$ $e2/0.25$		
$\not\!\!E_T$	$e1/0.38$ $\mu2/0.30$ $\mu1/0.29$ $e2/0.27$		
p_T(lepton)	$\mu2/0.64$ $\mu1/0.42$ $e2/0.16$		
Q(lepton)$\times\eta$(notbest1)	$\mu2/0.53$ $e1/0.49$ $\mu1/0.24$		
Event kinematics			
M(jet1,jet2)	$\mu2/1.00$ $e2/1.00$ $e1/0.80$ $\mu1/0.69$	Centrality(alljets)	$e1/0.34$
$M_T(W)$	$\mu1/1.00$ $e1/0.80$ $\mu2/0.71$ $e2/0.45$	M_T(jet1,jet2)	$\mu2/0.16$
H_T(jet1,jet2)	$e1/0.61$ $\mu2/0.60$ $\mu1/0.52$ $e2/0.22$		
$\sqrt{\hat{s}}$	$e1/0.38$ $\mu1/0.31$ $\mu2/0.25$ $e2/0.22$		
H_T(lepton,$\not\!\!E_T$,jet1,jet2)	$e1/0.36$ $\mu2/0.35$ $\mu1/0.33$ $e2/0.30$		
Top quark reconstruction			
$\Delta M_{\rm top}^{\rm min}$	$e1/1.00$ $\mu1/0.97$ $\mu2/0.77$ $e2/0.55$	$M(W$,jet1,$S2)$	$\mu2/0.30$
$M_{\rm top}^{\Delta M^{\rm min}}$	$\mu2/0.53$ $e2/0.25$ $e1/0.24$ $\mu1/0.16$	$M(W$,tag2)	$\mu2/0.28$
$M(W$,best1)	$e1/0.47$ $e2/0.31$ $\mu1/0.22$ $\mu2/0.21$		
Significance$_{\rm min}(M_{\rm top})$	$e1/0.31$ $e2/0.16$ $\mu1/0.11$		
$M(W$,best1,$S2)$	$\mu2/0.39$ $e1/0.30$		
$M(W$,tag1,$S2)$	$\mu1/0.27$ $e2/0.16$		
$M(W$,notbest1,$S2)$	$e1/0.24$ $e2/0.18$		
Angular correlations			
$\Delta\phi$(lepton,$\not\!\!E_T$)	$\mu2/0.54$ $e1/0.46$ $\mu1/0.44$ $e2/0.32$	cos(lepton$_{\rm besttop}$,besttop$_{\rm CMframe}$)	$\mu1/0.36$
cos(lepton,best1)$_{\rm besttop}$	$\mu2/0.33$ $\mu1/0.24$ $e2/0.17$	cos(best1,notbest1)$_{\rm besttop}$	$e2/0.17$
$\Delta R^{\rm min}$(alljets)	$e1/0.43$ $\mu1/0.19$		
$\Delta R^{\rm min}$(lepton,alljets)	$\mu2/0.40$ $\mu1/0.26$		
cos(lepton,jet1)$_{\rm btaggedtop}$	$e1/0.40$ $\mu2/0.15$		
ΔR(lepton,tag1)	$\mu2/0.35$ $\mu1/0.30$		
cos(lepton,Q(lepton)$\times z)_{\rm besttop}$	$e2/0.27$ $\mu1/0.12$		
Jet reconstruction			
Width$_\eta$(tag1)	$e1/0.65$ $e2/0.26$	Width$_\eta$(jet1)	$\mu1/0.36$
		Width$_\phi$(notbest1)	$\mu1/0.29$
		M(best1)	$\mu2/0.24$
		M(tag1)	$e1/0.18$
		Width$_\eta$(tag2)	$e2/0.16$
		Width$_\phi$(best1)	$\mu1/0.15$

Table E.28 : Set of selected variables for the 2-tags 2-jets channels. For the angular variables, the subscript indicates the reference frame. For each variable, it is shown to its right the list of channels (reco versions and lepton types) in which the variable is used for the BNN training, together with its importance in each of the channels. (See main text for abbreviations convention.)

Object kinematics			
p_T(notbest2)	$\mu 2/1.00$ $\mu 1/1.00$ $e1/1.00$ $e2/0.63$	Q(lepton)$\times\eta$(tag2)	$\mu 2/0.61$
Q(lepton)$\times\eta$(light1)	$e1/0.91$ $\mu 2/0.52$ $e2/0.46$ $\mu 1/0.27$	Q(lepton)$\times\eta$(light2)	$\mu 2/0.39$
p_T(lepton)	$e1/0.33$ $\mu 1/0.28$	Q(lepton)$\times\eta$(best1)	$e1/0.33$
		p_T(light2)	$e2/0.23$
Event kinematics			
Centrality(alljets)	$e2/1.00$ $\mu 1/0.99$ $\mu 2/0.75$ $e1/0.49$	M(light1,light2)	$e2/0.60$
M(alljets)	$\mu 2/0.90$ $e2/0.85$ $e1/0.82$ $\mu 1/0.65$		
$M_T(W)$	$\mu 1/0.87$ $e2/0.83$ $e1/0.82$ $\mu 2/0.71$		
M(alljets$-$best1)	$e1/0.77$ $e2/0.63$ $\mu 1/0.50$ $\mu 2/0.38$		
M(alljets$-$tag1)	$\mu 2/0.70$ $e1/0.60$ $\mu 1/0.50$ $e2/0.40$		
M(jet1,jet2)	$e2/0.52$ $\mu 1/0.50$ $e1/0.49$ $\mu 2/0.32$		
H_T(lepton$\not\!\!E_T$)	$e1/0.80$ $\mu 2/0.48$ $e2/0.47$		
H_T(lepton$\not\!\!E_T$,alljets)	$\mu 1/0.69$ $e2/0.32$ $\mu 2/0.25$		
H_T(alljets$-$tag1)	$\mu 1/0.53$ $\mu 2/0.46$ $e2/0.40$		
H_T(alljets)	$e1/0.92$ $\mu 2/0.44$		
Top quark reconstruction			
$M(W$,tag1)	$\mu 2/0.77$ $e2/0.75$ $\mu 1/0.70$ $e1/0.64$		
$\Delta M_{\text{top}}^{\text{min}}$	$\mu 2/0.52$ $e2/0.48$ $e1/0.41$ $\mu 1/0.40$		
$M(W$,tag1,$S2)$	$e1/0.81$ $\mu 2/0.55$ $e2/0.42$		
$M(W$,light1,$S2)$	$e1/0.57$ $e2/0.24$		
$M(W$,tag2)	$\mu 1/0.54$ $\mu 2/0.46$		
Angular correlations			
$\Delta\phi$(lepton$\not\!\!E_T$)	$\mu 1/0.43$ $e2/0.26$ $e1/0.21$	ΔR^{min}(alljets)	$e1/0.50$
ΔR(lepton,tag1)	$\mu 1/0.68$ $\mu 2/0.38$	cos(lepton,tag1)$_{\text{btaggedtop}}$	$\mu 2/0.47$
cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	$\mu 1/0.62$ $\mu 2/0.23$	ΔR(jet1,jet2)	$\mu 2/0.45$
		cos(lepton$_{\text{besttop}}$,besttop$_{\text{CMframe}}$)	$e2/0.32$
Jet reconstruction			
Width$_\eta$(light2)	$\mu 1/0.90$ $e1/0.80$ $e2/0.66$ $\mu 2/0.50$	Width$_\eta$(jet2)	$e1/0.33$
M(jet1)	$\mu 1/0.78$ $e1/0.60$ $e2/0.53$ $\mu 2/0.25$		
Width$_\phi$(light2)	$\mu 2/0.50$ $e2/0.47$ $e1/0.31$ $\mu 1/0.25$		
Width$_\eta$(tag1)	$e1/0.59$ $\mu 1/0.50$ $e2/0.27$		
Width$_\phi$(light1)	$\mu 1/0.59$ $e2/0.43$ $e1/0.36$		
Width$_\eta$(tag2)	$e1/0.53$ $\mu 1/0.52$ $e2/0.25$		
Width$_\phi$(jet1)	$e1/0.46$ $e2/0.34$ $\mu 2/0.31$		
Width$_\eta$(notbest2)	$e1/0.40$ $\mu 1/0.28$		

Table E.29 : Set of selected variables for the 2-tags 3-jets channels. For the angular variables, the subscript indicates the reference frame. For each variable, it is shown to its right the list of channels (reco versions and lepton types) in which the variable is used for the BNN training, together with its importance in each of the channels. (See main text for abbreviations convention.)

Object kinematics			
p_T(notbest2)	$\mu 2/0.90$ $\mu 1/0.63$ $e2/0.44$ $e1/0.36$	Q(lepton)$\times\eta$(light2)	$\mu 2/0.32$
p_T(jet4)	$\mu 1/0.55$ $e1/0.42$ $e2/0.37$ $\mu 2/0.25$	Q(lepton)$\times\eta$(best1)	$\mu 1/0.26$
Q(lepton)$\times\eta$(light1)	$\mu 1/0.48$ $\mu 2/0.36$ $e1/0.31$ $e2/0.30$	Q(lepton)$\times\eta$(jet1)	$e1/0.25$
p_T(light2)	$e2/0.66$ $e1/0.39$ $\mu 2/0.20$	p_T(jet3)	$\mu 1/0.17$
p_T(lepton)	$\mu 2/0.32$ $\mu 1/0.22$		
Event kinematics			
M(alljets$-$tag1)	$e2/1.00$ $e1/1.00$ $\mu 1/0.90$ $\mu 2/0.82$	Aplanarity(W,alljets)	$e1/0.20$
Centrality(alljets)	$\mu 2/1.00$ $\mu 1/1.00$ $e1/0.86$ $e2/0.85$		
M(alljets$-$best1)	$e1/0.94$ $\mu 1/0.58$ $\mu 2/0.53$ $e2/0.44$		
H_T(alljets$-$tag1)	$\mu 1/0.70$ $e1/0.47$ $e2/0.32$ $\mu 2/0.17$		
H_T(lepton$\not\!\!E_T$,alljets)	$e2/0.63$ $e1/0.45$ $\mu 1/0.44$ $\mu 2/0.43$		
M(alljets)	$e1/0.61$ $e2/0.54$ $\mu 1/0.33$ $\mu 2/0.17$		
M(jet3,jet4)	$e2/0.39$ $\mu 2/0.36$ $\mu 1/0.34$ $e1/0.26$		
$M_T(W)$	$e1/0.41$ $e2/0.28$ $\mu 1/0.13$		
M(light1,light2)	$e2/0.38$ $\mu 1/0.35$ $\mu 2/0.16$		
M(jet1,jet2)	$\mu 1/0.30$ $\mu 2/0.21$ $e2/0.21$		
H_T(lepton$\not\!\!E_T$)	$\mu 1/0.40$ $e1/0.32$		
H_T(alljets$-$best1)	$e2/0.38$ $e1/0.19$		
Top quark reconstruction			
$\Delta M_{\text{top}}^{\text{min}}$	$\mu 2/0.21$ $\mu 1/0.21$	M(W,jet4)	$\mu 1/0.34$
		M(W,light1,$S2$)	$e2/0.32$
		M(W,tag1,$S2$)	$e2/0.29$
		M(W,jet1,$S2$)	$\mu 2/0.26$
		M(W,tag2)	$e1/0.17$
		M(W,light1)	$\mu 1/0.13$
Angular correlations			
cos(lepton$_{\text{btaggedtop}}$,btaggedtop$_{\text{CMframe}}$)	$\mu 1/0.52$ $\mu 2/0.35$ $e1/0.33$ $e2/0.21$	cos(lepton$_{\text{besttop}}$,besttop$_{\text{CMframe}}$)	$e1/0.28$
ΔR(jet1,jet2)	$e1/0.62$ $e2/0.41$ $\mu 2/0.40$	cos(lepton,tag1)$_{\text{btaggedtop}}$	$\mu 1/0.27$
ΔR^{min}(alljets)	$e2/0.21$ $\mu 2/0.20$	ΔR^{min}(lepton,alljets)	$\mu 2/0.13$
Jet reconstruction			
Width$_\eta$(jet4)	$\mu 1/0.68$ $e1/0.63$ $e2/0.46$ $\mu 2/0.46$	Width$_\eta$(tag2)	$e1/0.42$
Width$_\phi$(jet4)	$e1/0.53$ $e2/0.34$ $\mu 1/0.28$ $\mu 2/0.25$	Width$_\eta$(tag1)	$e1/0.29$
Width$_\phi$(light2)	$e2/0.49$ $\mu 2/0.44$ $e1/0.42$ $\mu 1/0.34$		
Width$_\phi$(tag2)	$e2/0.43$ $\mu 2/0.32$ $\mu 1/0.29$		
Width$_\eta$(notbest2)	$\mu 1/0.66$ $e1/0.22$		
Width$_\eta$(light2)	$e1/0.52$ $\mu 2/0.24$		
Width$_\phi$(light1)	$e2/0.26$ $\mu 2/0.22$		

Table E.30 : Set of selected variables for the 2-tags 4-jets channels. For the angular variables, the subscript indicates the reference frame. For each variable, it is shown to its right the list of channels (reco versions and lepton types) in which the variable is used for the BNN training, together with its importance in each of the channels. (See main text for abbreviations convention.)

F Verification Plots

This appendix presents, for each of the twenty four analysis channels, plots that verify the convergence of the Markov chains and plots of the BNN performance. They are similar to those presented in Sect. 7.6.3 where only the ones for the Run IIa electron 2-jet 1-tag channel were shown as an example. Events from the yield sample (see Sect. 7.1) are used to make these graphs. The quantities S and B appearing in the central plot of the second and fourth rows correspond to the signal and background amount left in the sample after rejecting events with a BNN output lower than a certain limit. For a more detailed explanation of the graphs, please refer back to Sect. 7.6.3.

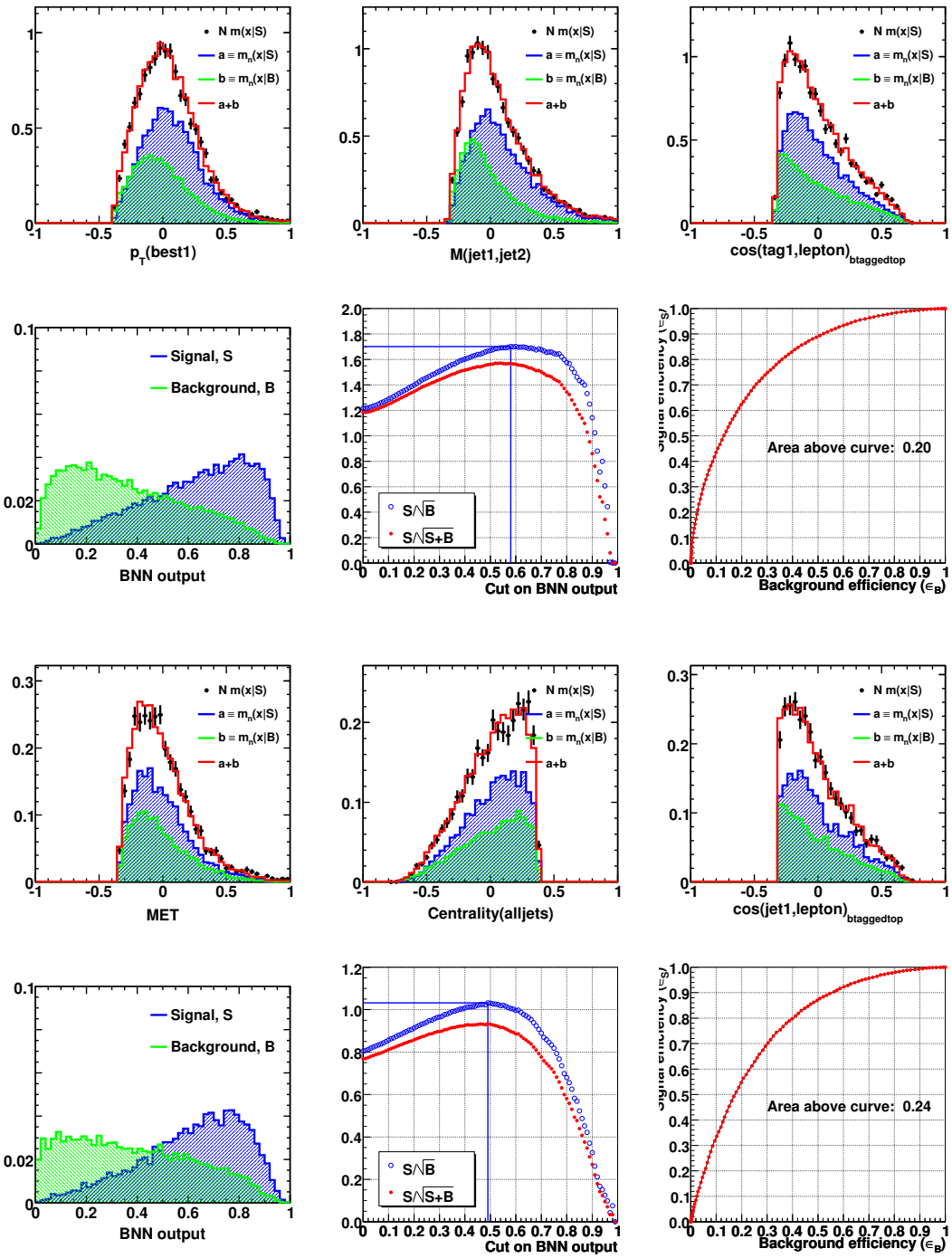

Figure F.1 : Verification plots for three of the variables used by the BNN in the Run IIa e+jets 2-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

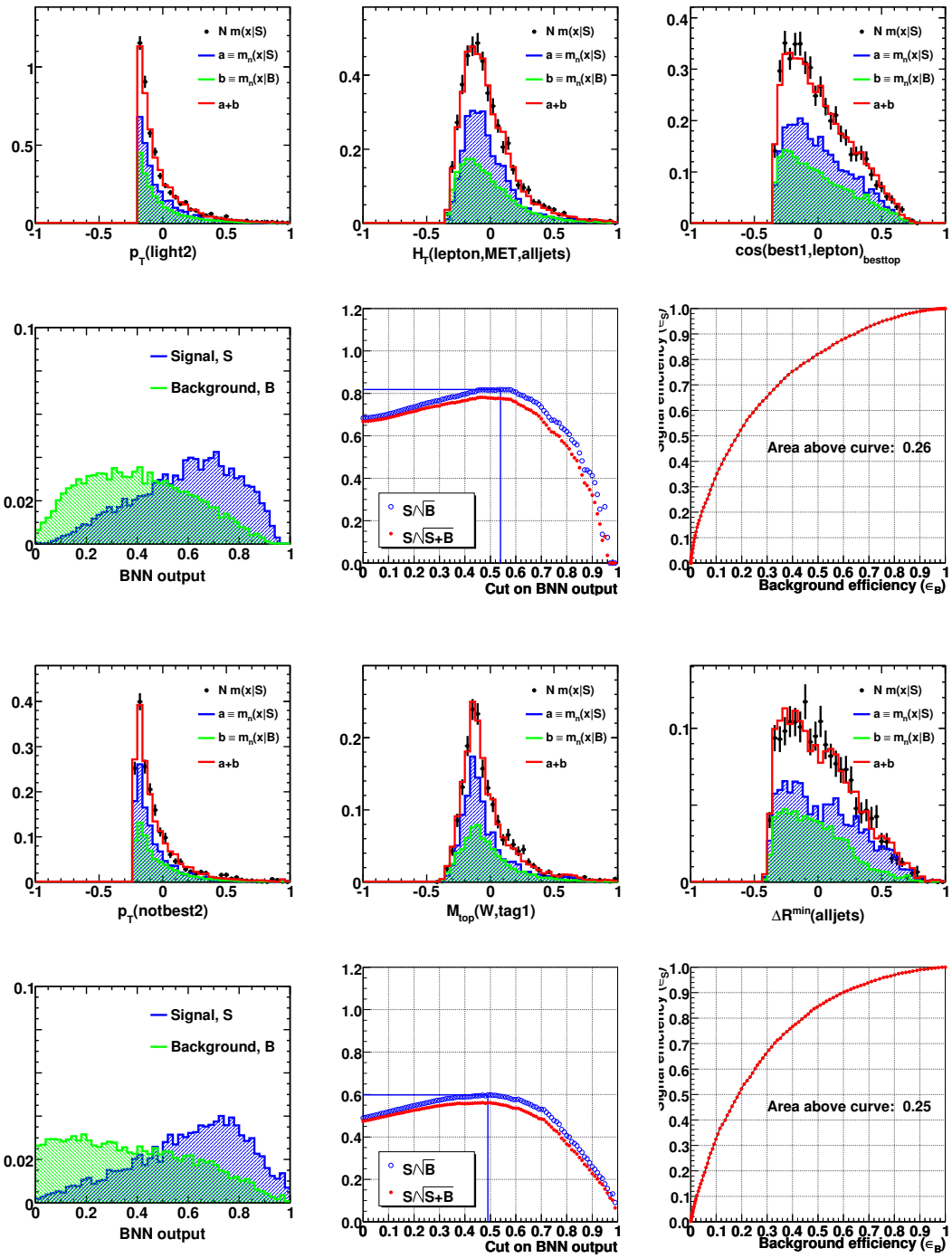

Figure F.2 : Verification plots for three of the variables used by the BNN in the Run IIa e+jets 3-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

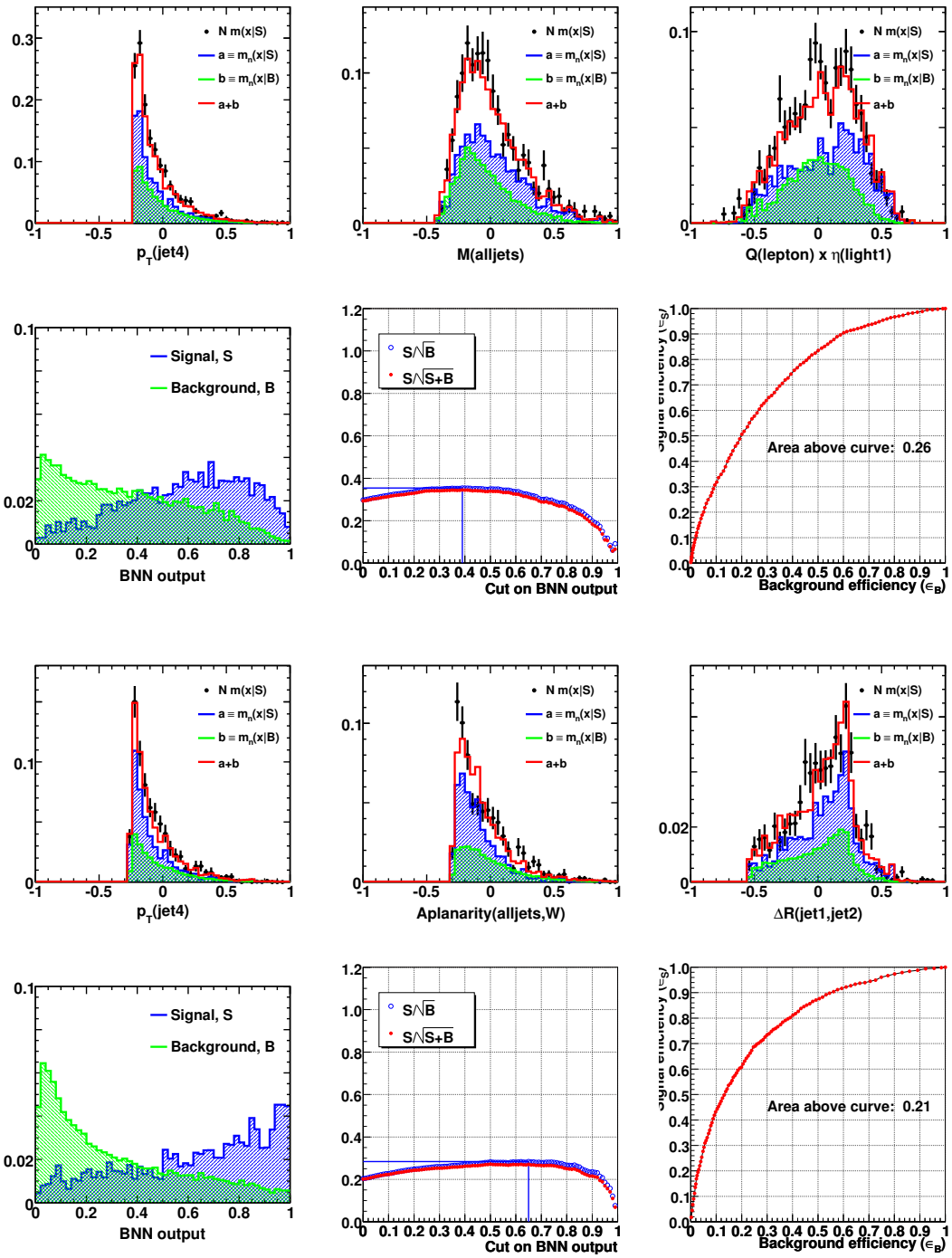

Figure F.3 : Verification plots for three of the variables used by the BNN in the Run IIa e+jets 4-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

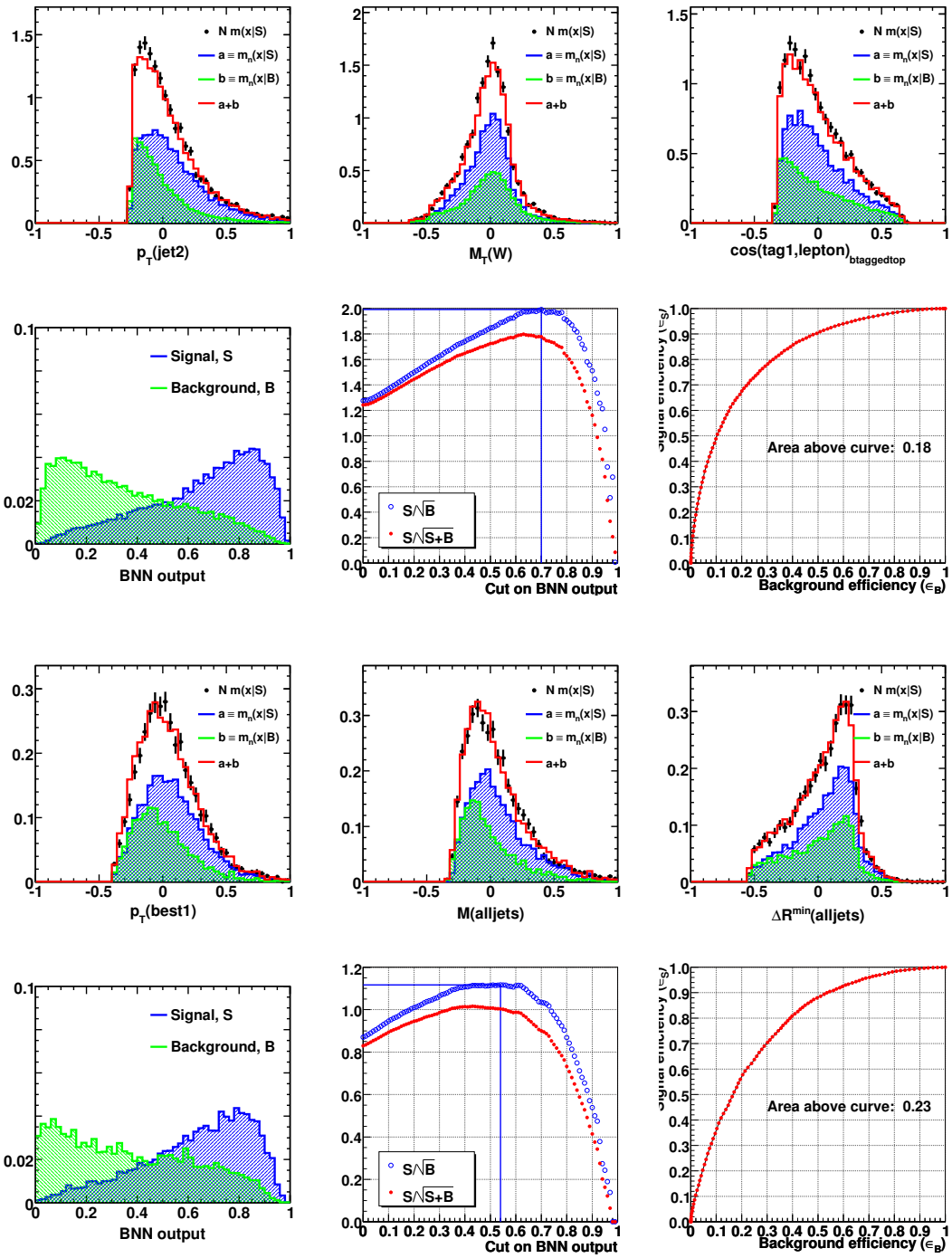

Figure F.4 : Verification plots for three of the variables used by the BNN in the Run IIa μ+jets 2-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

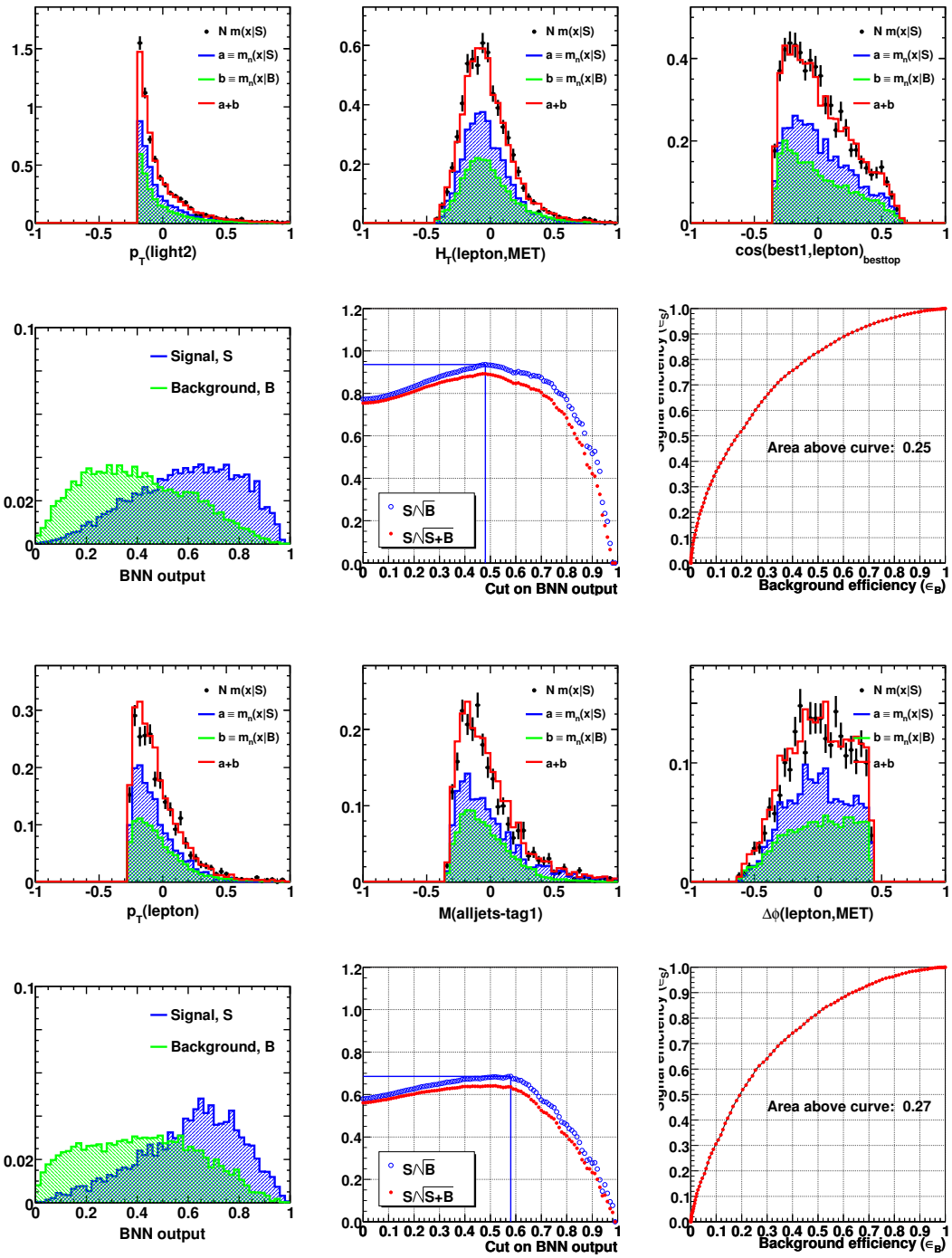

Figure F.5 : Verification plots for three of the variables used by the BNN in the Run IIa μ+jets 3-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

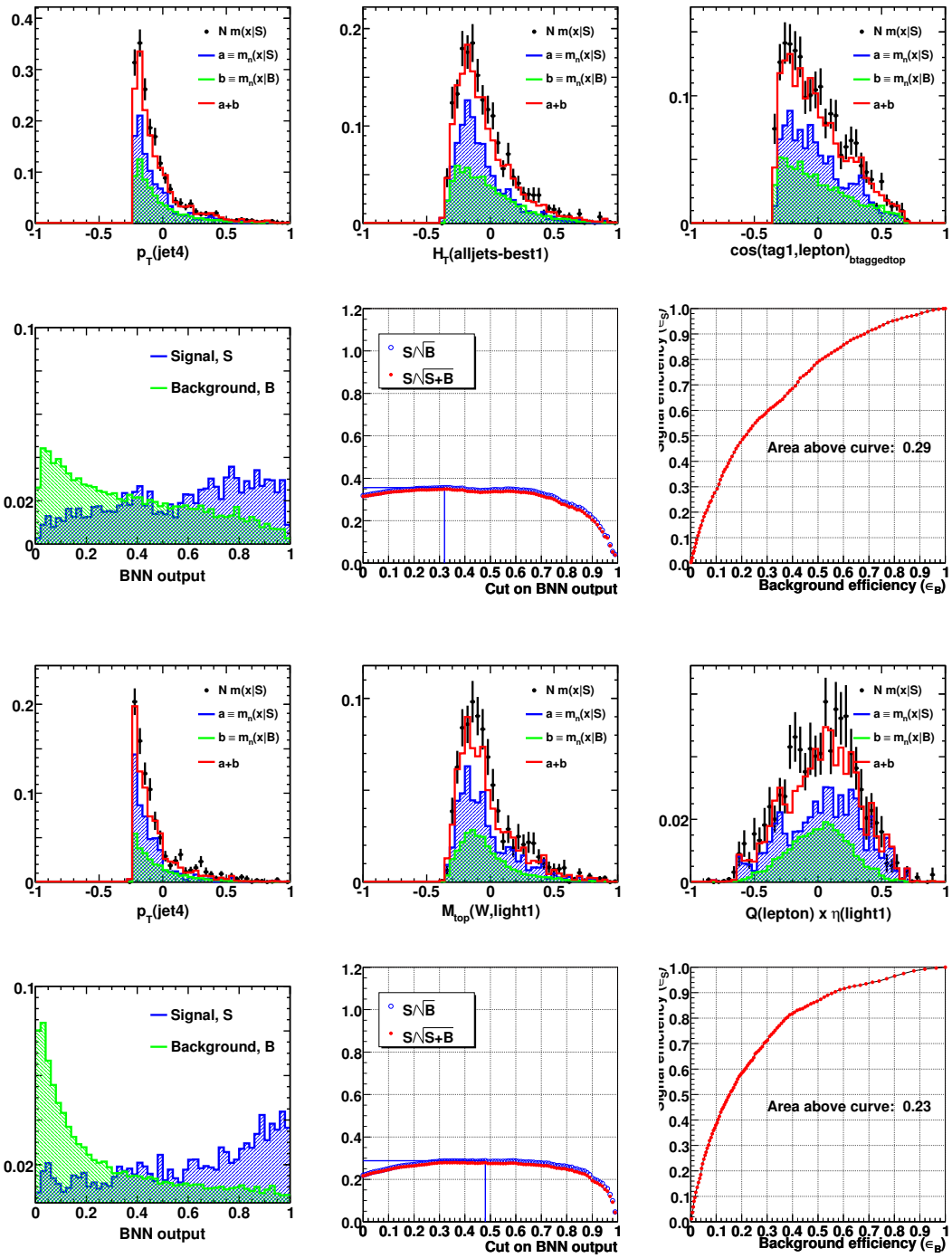

Figure F.6 : Verification plots for three of the variables used by the BNN in the Run IIa μ+jets 4-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

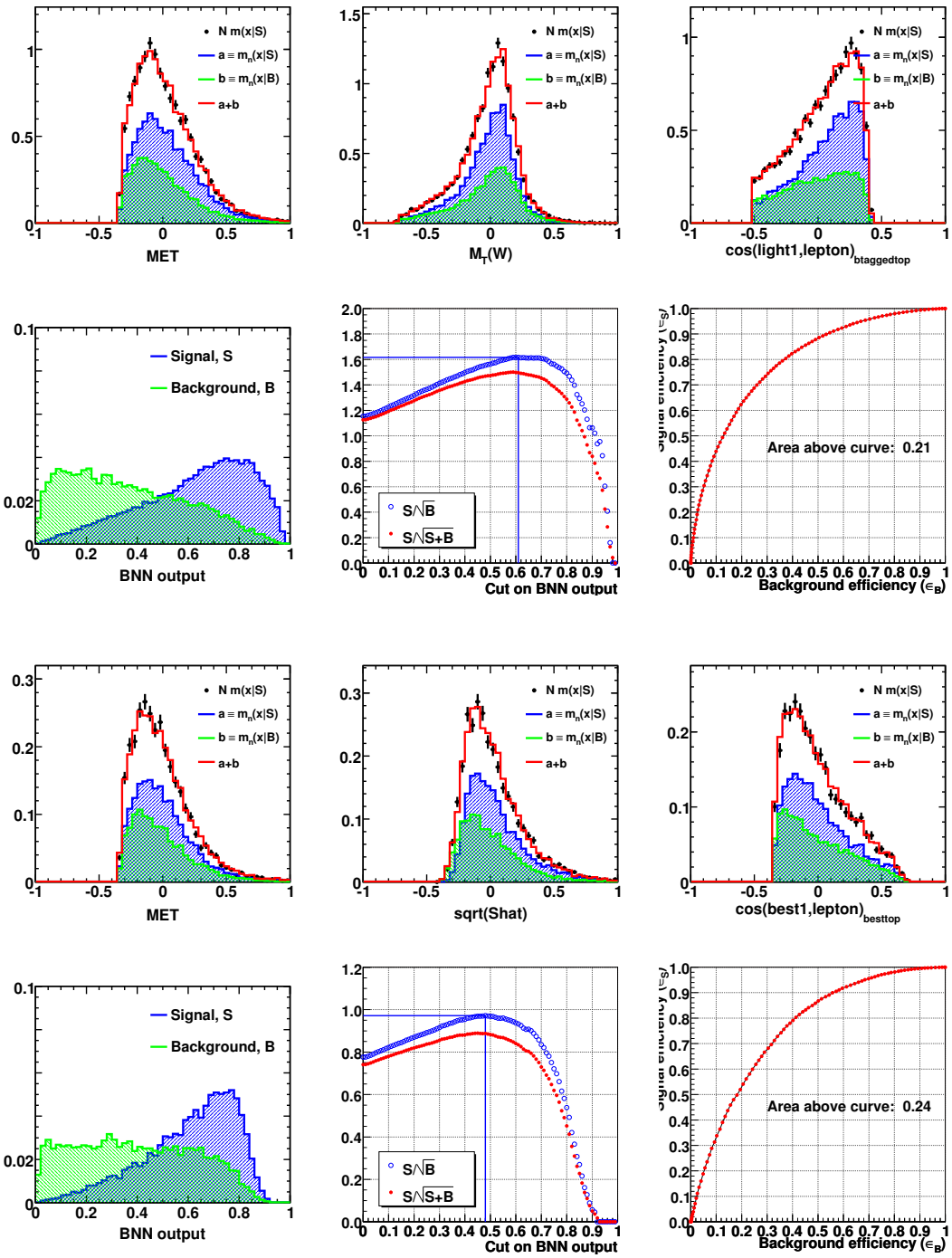

Figure F.7 : Verification plots for three of the variables used by the BNN in the Run IIb e+jets 2-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

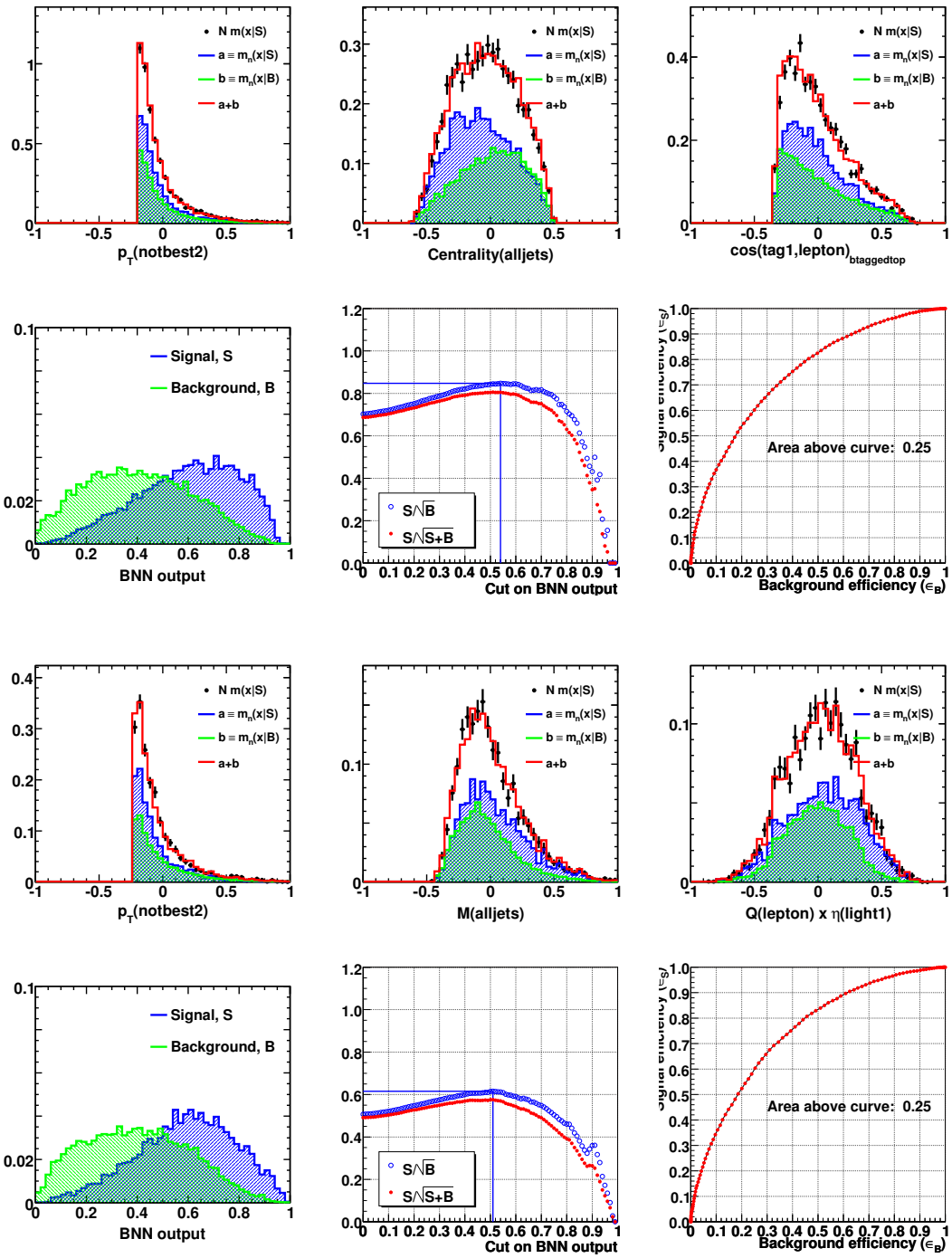

Figure F.8 : Verification plots for three of the variables used by the BNN in the Run IIb e+jets 3-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

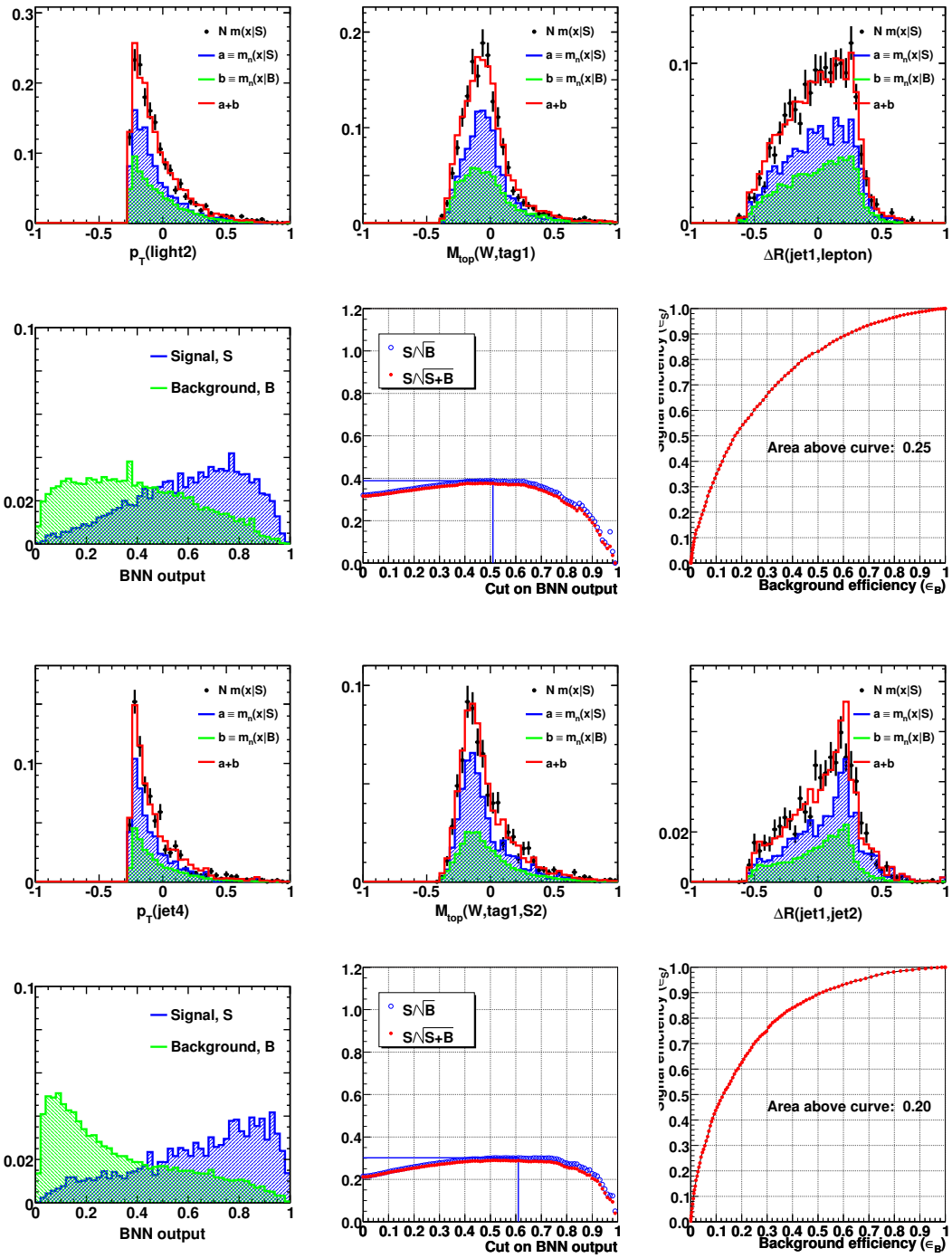

Figure F.9 : Verification plots for three of the variables used by the BNN in the Run IIb e+jets 4-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

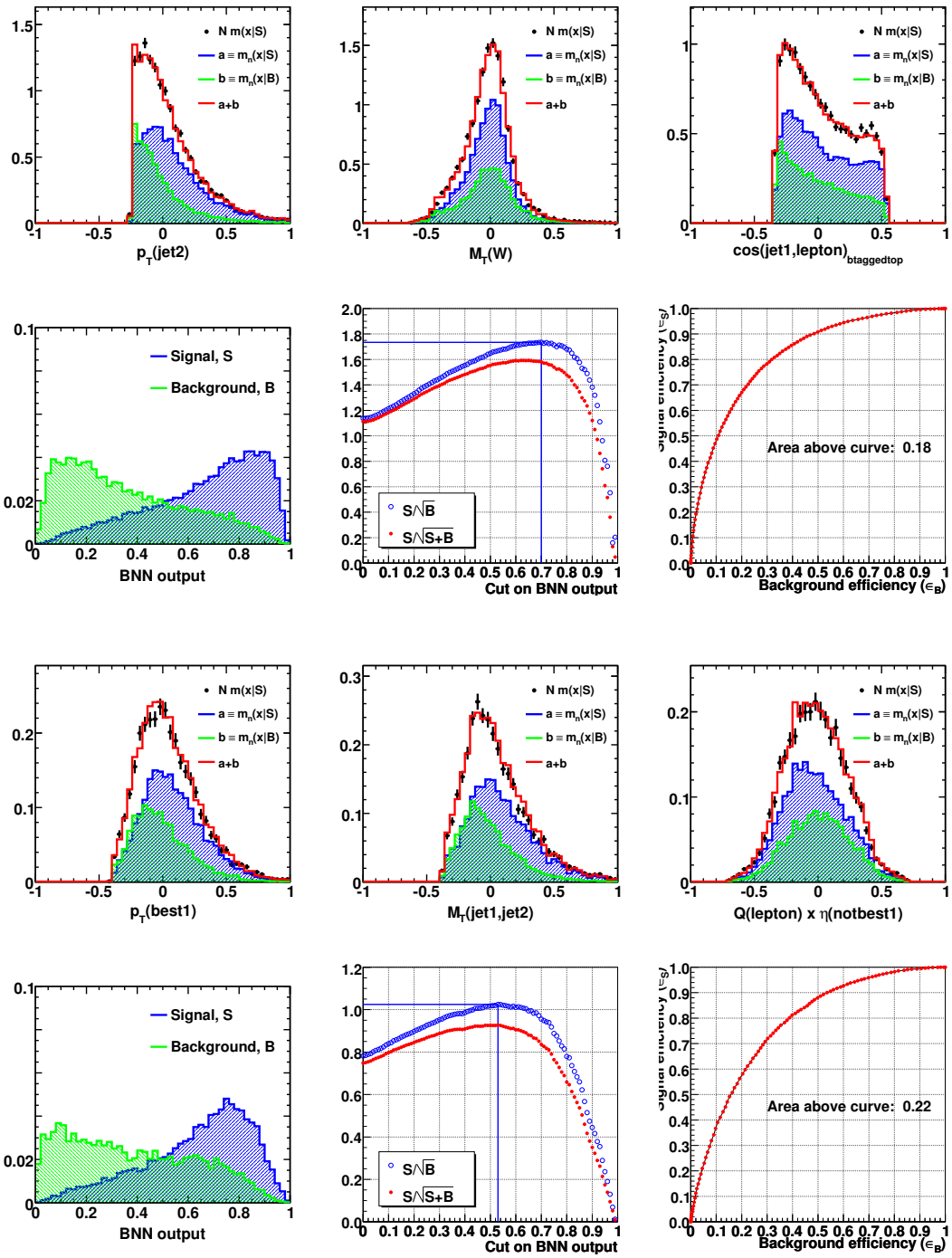

Figure F.10 : Verification plots for three of the variables used by the BNN in the Run IIb μ+jets 2-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

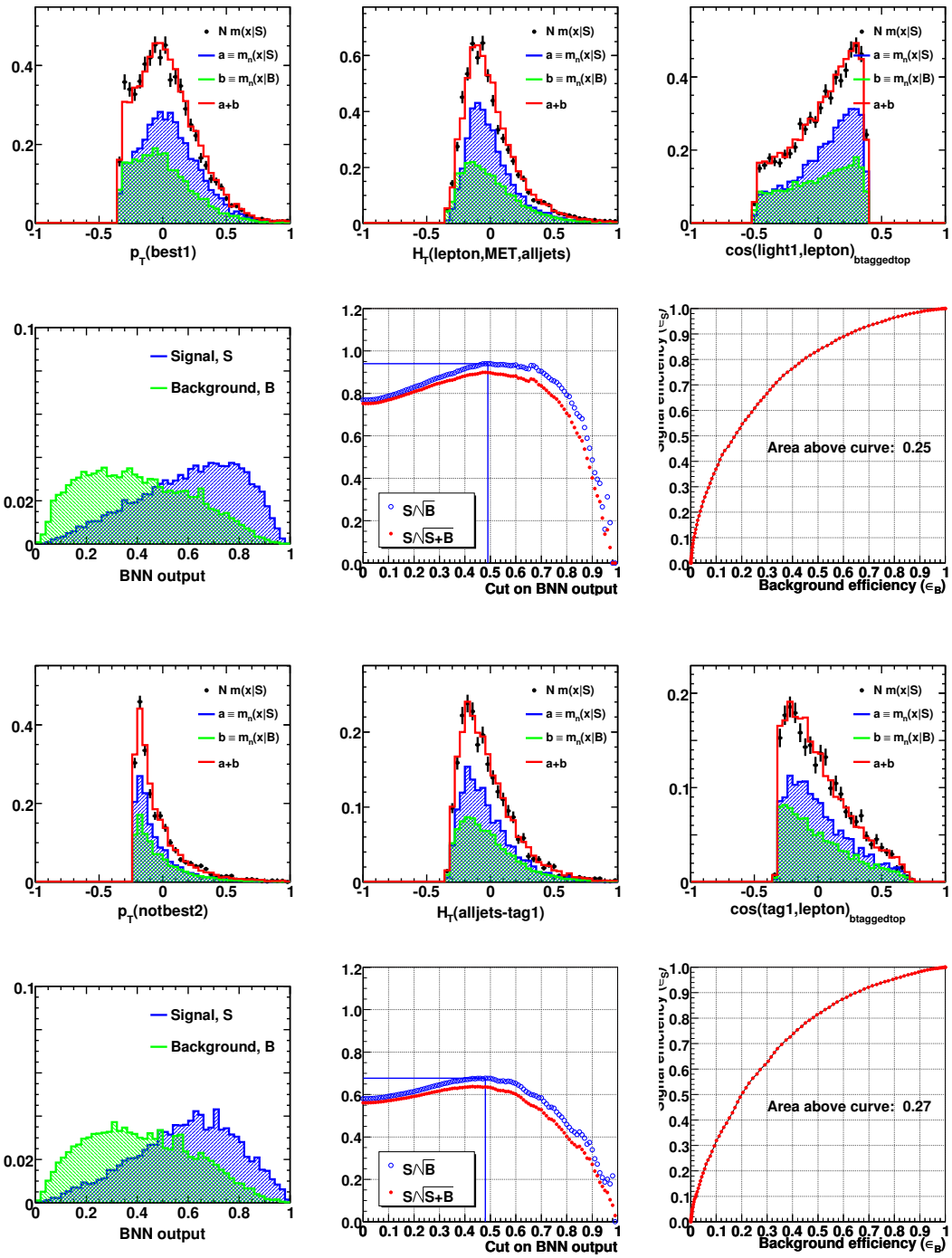

Figure F.11 : Verification plots for three of the variables used by the BNN in the Run IIb μ+jets 3-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

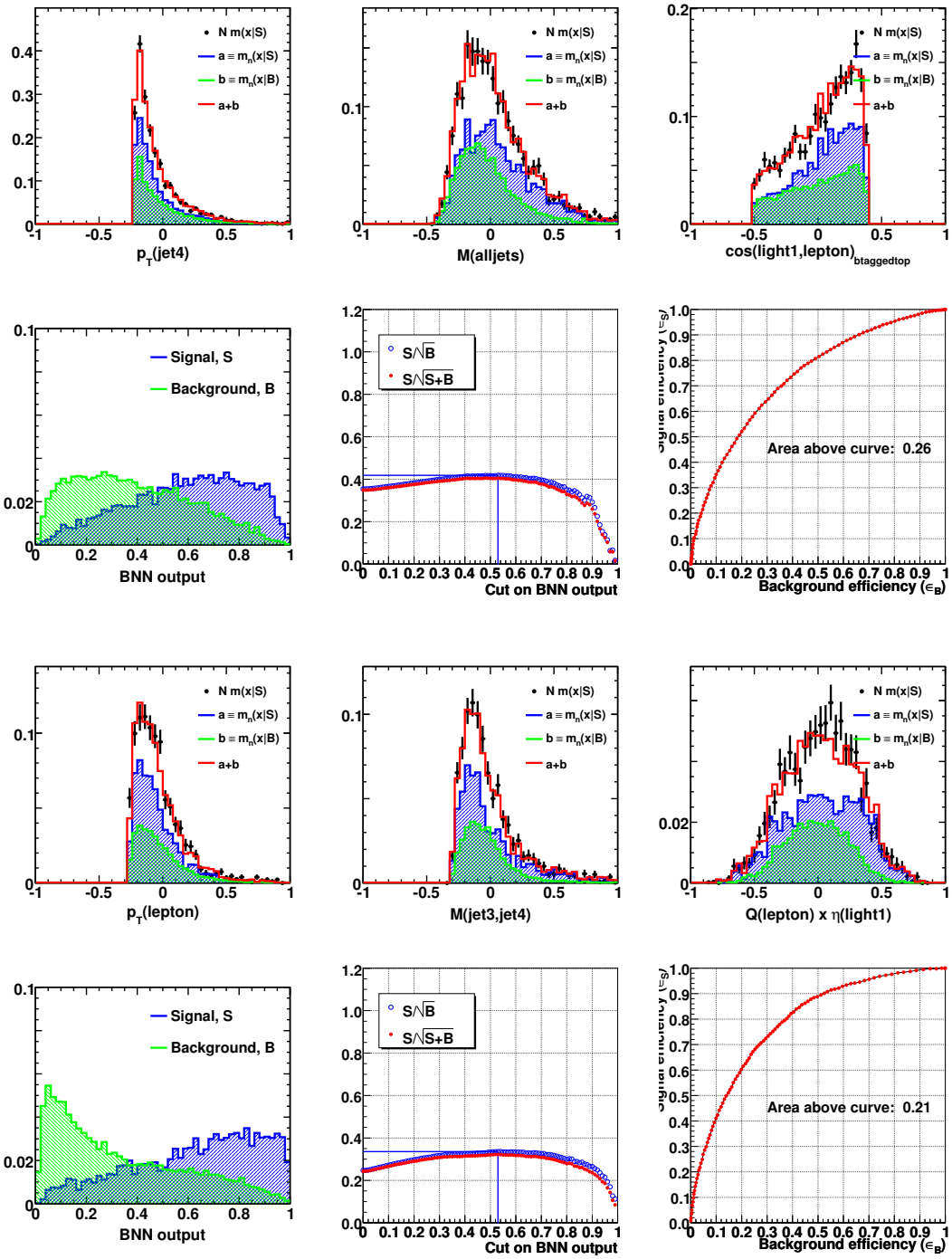

Figure F.12 : Verification plots for three of the variables used by the BNN in the Run IIb μ+jets 4-jet channel with 1 b-tag (first row) and 2 b-tags (third row). Below the verification plots, left: BNN-output distributions for background (green) and signal (blue) normalized to unity; middle: $S/\sqrt{B}$ and $S/\sqrt{S+B}$ versus a lower cut on the BNN output; right: signal and background efficiencies (ϵ_S and ϵ_B) as a function of the cut on the BNN output.

G BNN Discriminant Plots

This appendix shows the BNN discriminants separately for each of the twenty four analysis channels, using 50 histograming bins, in both linear and logarithmic scales. The single top cross section is extracted from exactly these histograms. These plots complement the ones shown in Sect. 7.8 for the Run IIa+b electron+muon channels combination.

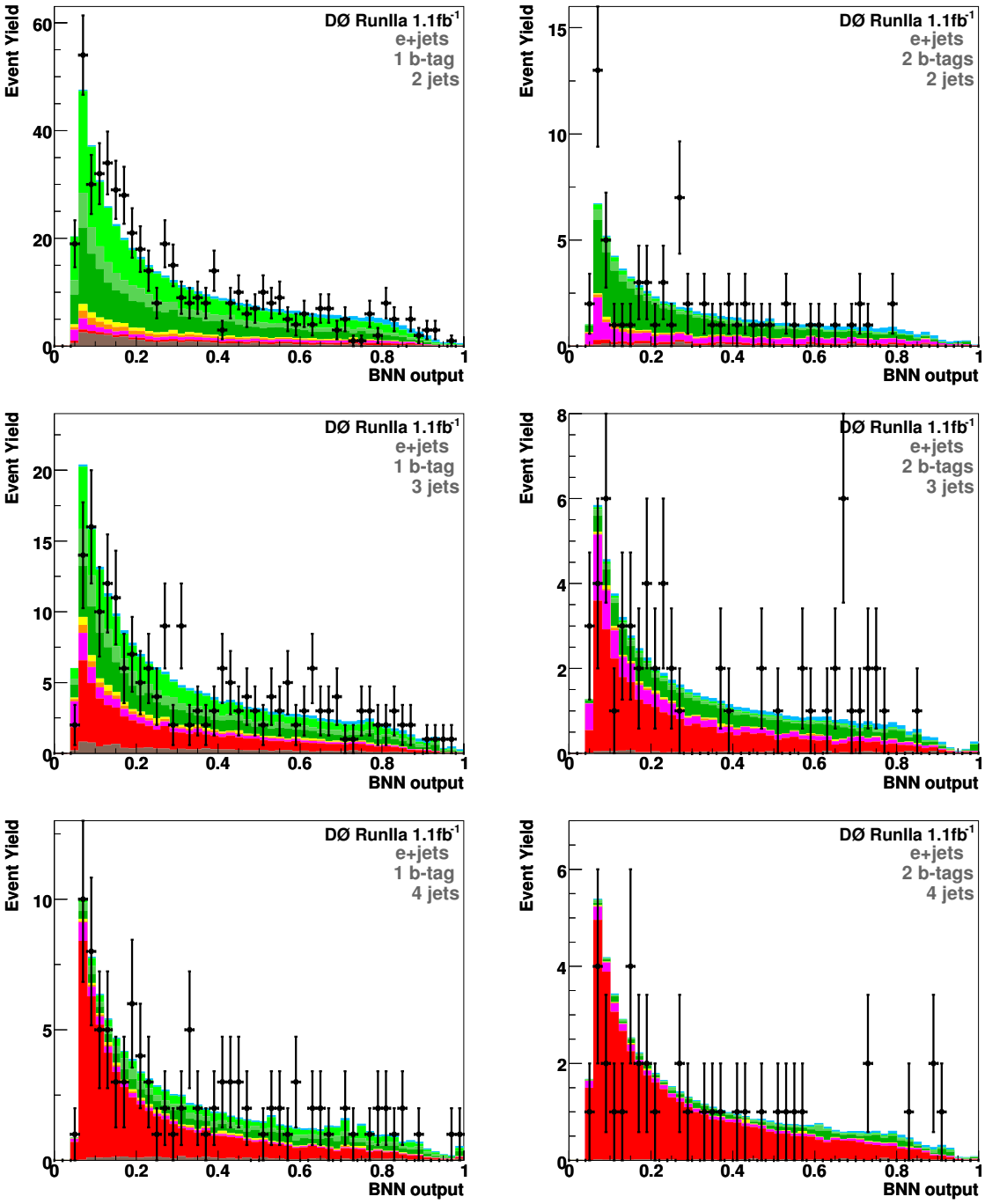

Figure G.1 : Bayesian neural network output distributions for the e+jets channel in Run IIa data shown in linear scale. The total signal contribution is normalized to the measured cross section, while the relative tb to tqb contributions are given by the ratio of their SM cross sections. Different rows correspond to different jet multiplicities: 2 jets (up), 3 jets (middle), 4 jets (down). The left column corresponds to 1 b-tag and the right column to 2 b-tags.

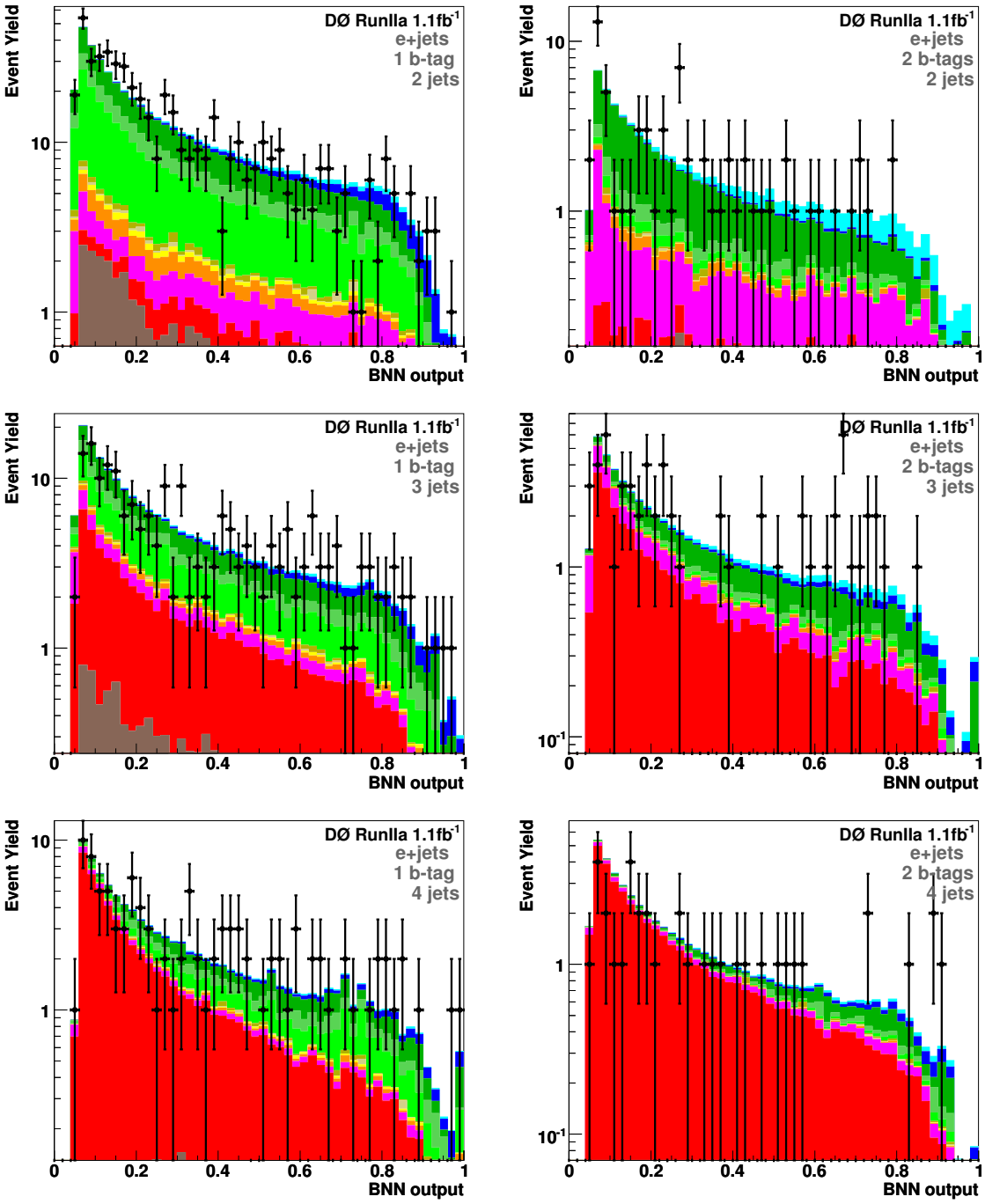

Figure G.2 : Bayesian neural network output distributions for the e+jets channel in Run IIa data shown in logarithmic scale. The total signal contribution is normalized to the measured cross section, while the relative tb to tqb contributions are given by the ratio of their SM cross sections. Different rows correspond to different jet multiplicities: 2 jets (up), 3 jets (middle), 4 jets (down). The left column corresponds to 1 b-tag and the right column to 2 b-tags.

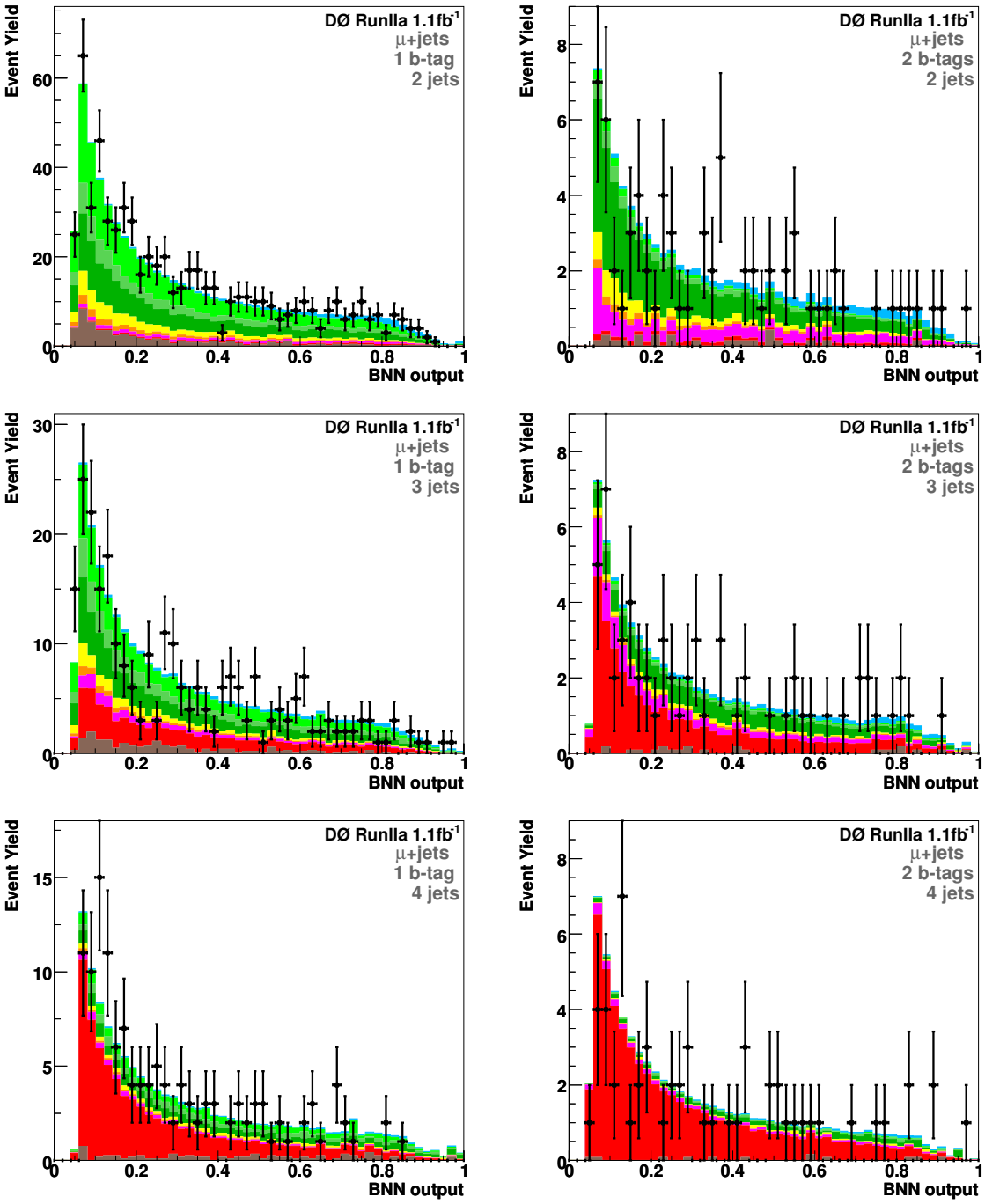

Figure G.3 : Bayesian neural network output distributions for the μ+jets channel in Run IIa data shown in linear scale. The total signal contribution is normalized to the measured cross section, while the relative tb to tqb contributions are given by the ratio of their SM cross sections. Different rows correspond to different jet multiplicities: 2 jets (up), 3 jets (middle), 4 jets (down). The left column corresponds to 1 b-tag and the right column to 2 b-tags.

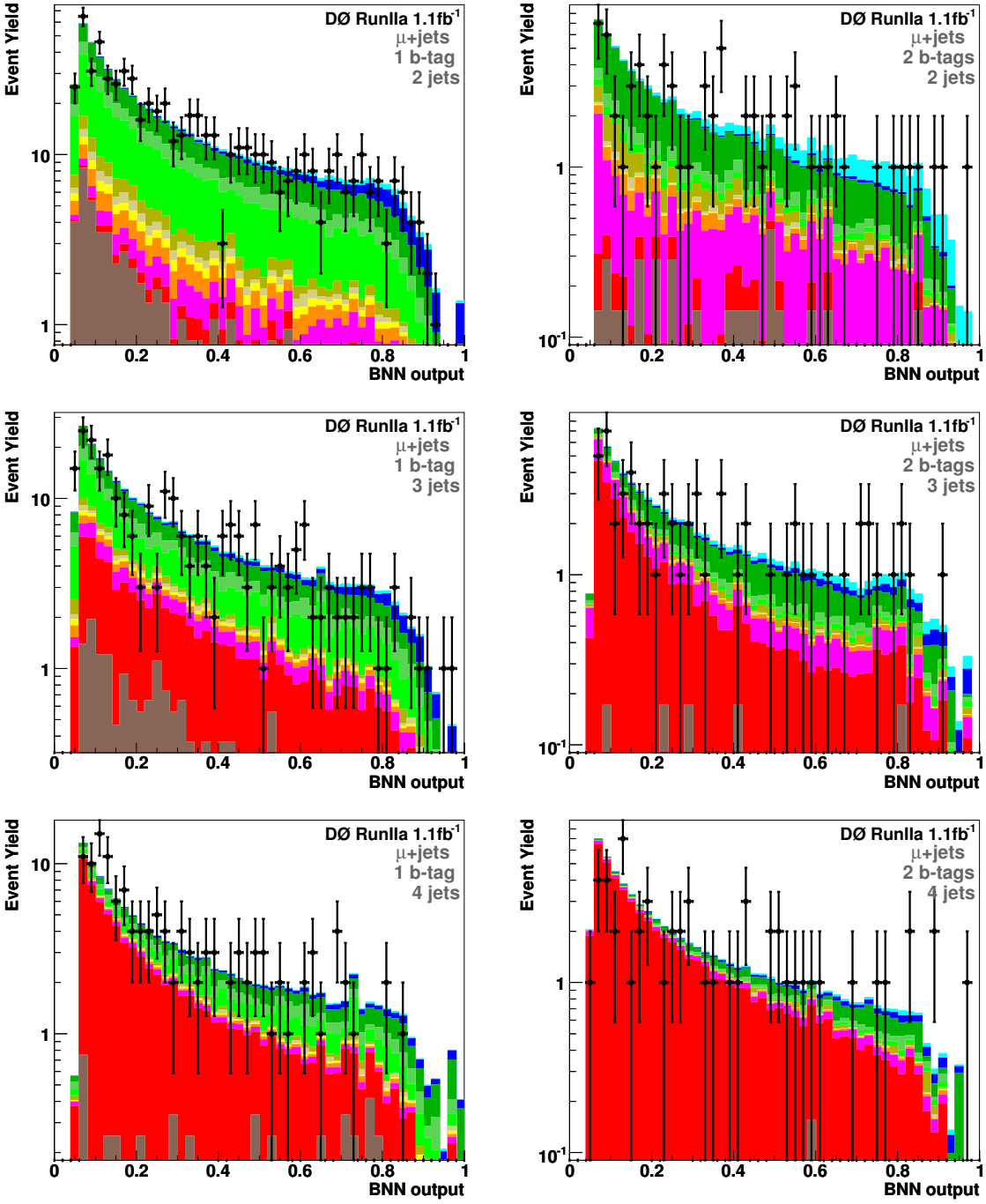

Figure G.4 : Bayesian neural network output distributions for the μ+jets channel in Run IIa data shown in logarithmic scale. The total signal contribution is normalized to the measured cross section, while the relative tb to tqb contributions are given by the ratio of their SM cross sections. Different rows correspond to different jet multiplicities: 2 jets (up), 3 jets (middle), 4 jets (down). The left column corresponds to 1 b-tag and the right column to 2 b-tags.

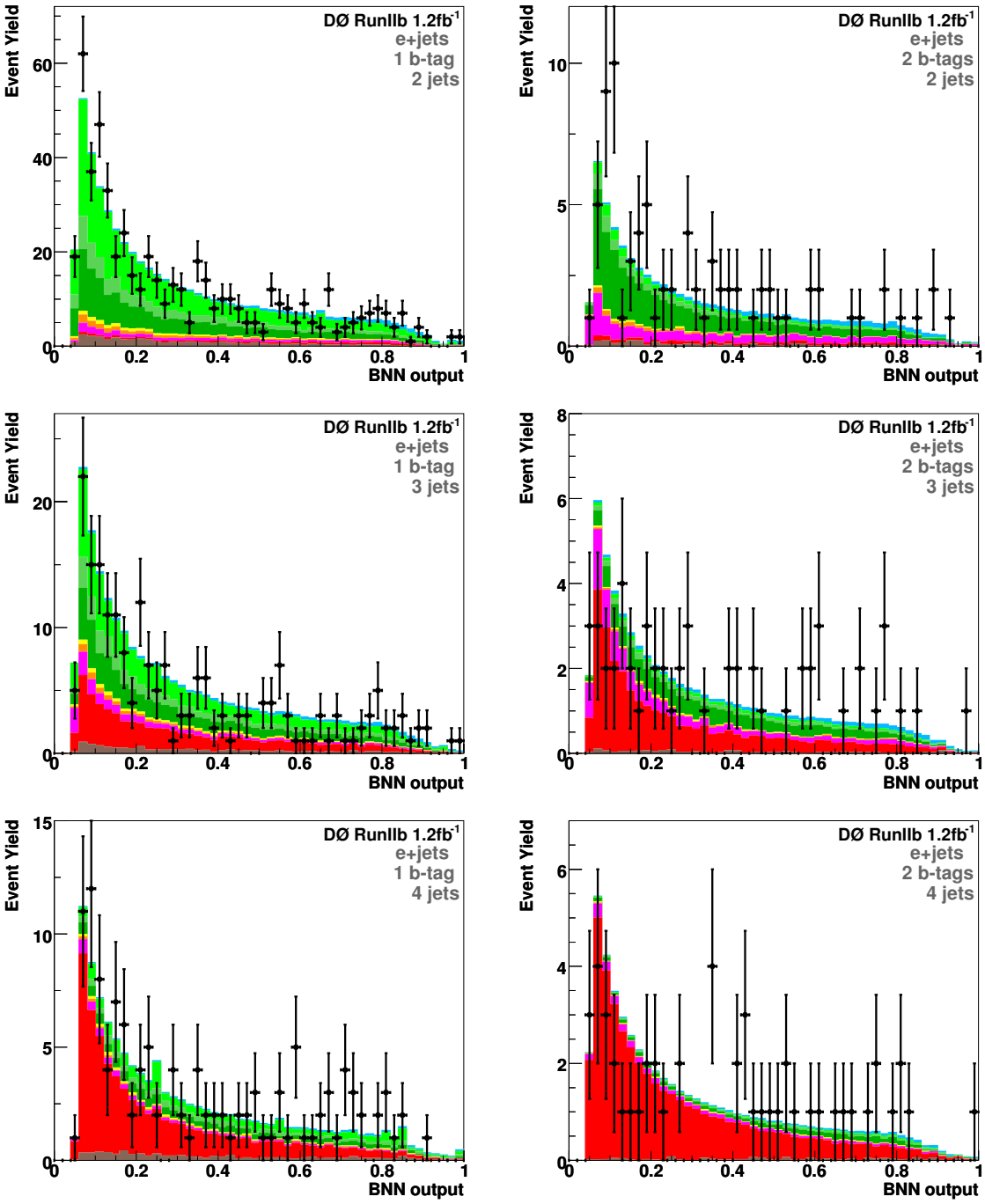

Figure G.5 : Bayesian neural network output distributions for the e+jets channel in Run IIb data shown in linear scale. The total signal contribution is normalized to the measured cross section, while the relative tb to tqb contributions are given by the ratio of their SM cross sections. Different rows correspond to different jet multiplicities: 2 jets (up), 3 jets (middle), 4 jets (down). The left column corresponds to 1 b-tag and the right column to 2 b-tags.

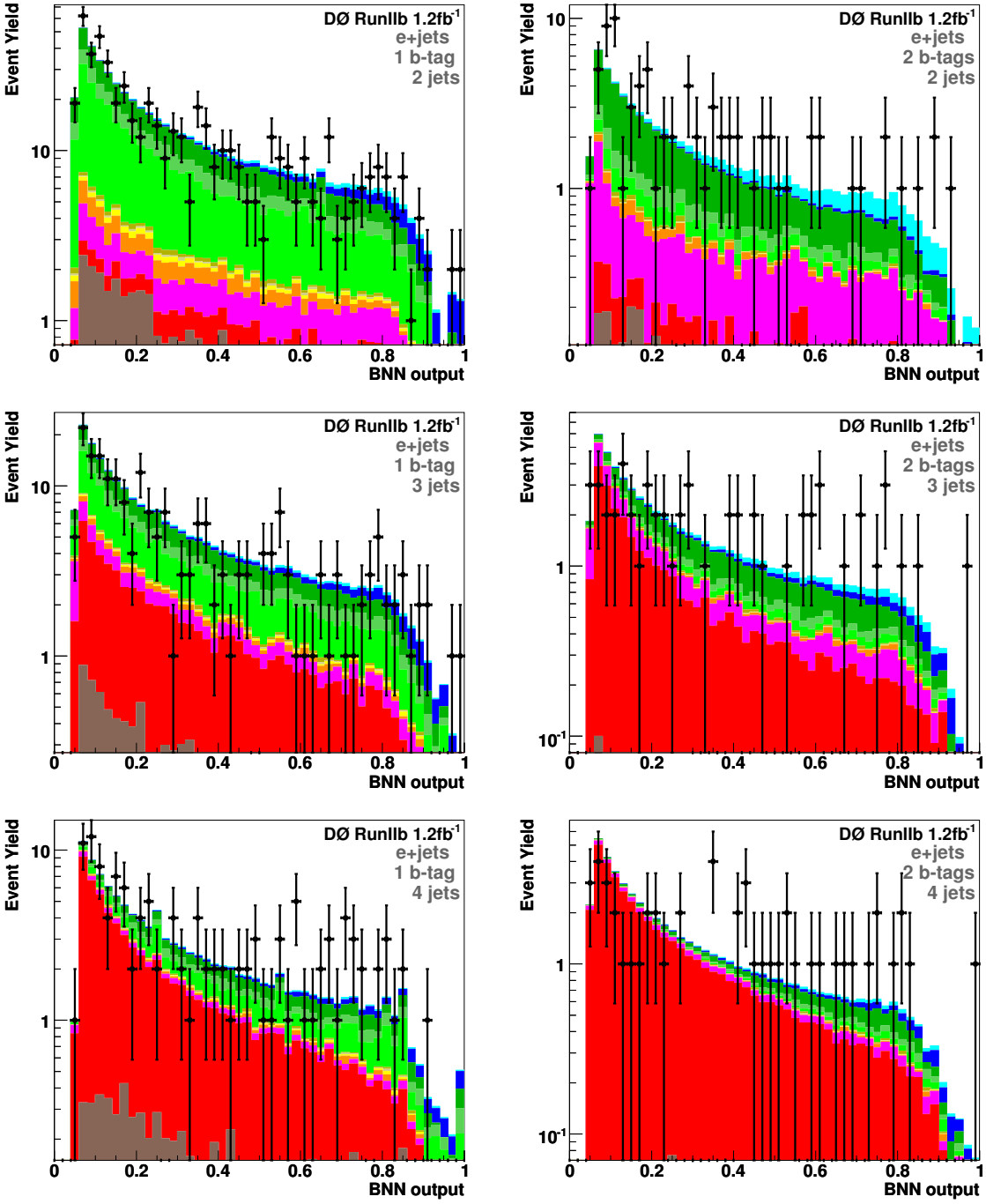

Figure G.6 : Bayesian neural network output distributions for the e+jets channel in Run IIb data shown in logarithmic scale. The total signal contribution is normalized to the measured cross section, while the relative tb to tqb contributions are given by the ratio of their SM cross sections. Different rows correspond to different jet multiplicities: 2 jets (up), 3 jets (middle), 4 jets (down). The left column corresponds to 1 b-tag and the right column to 2 b-tags.

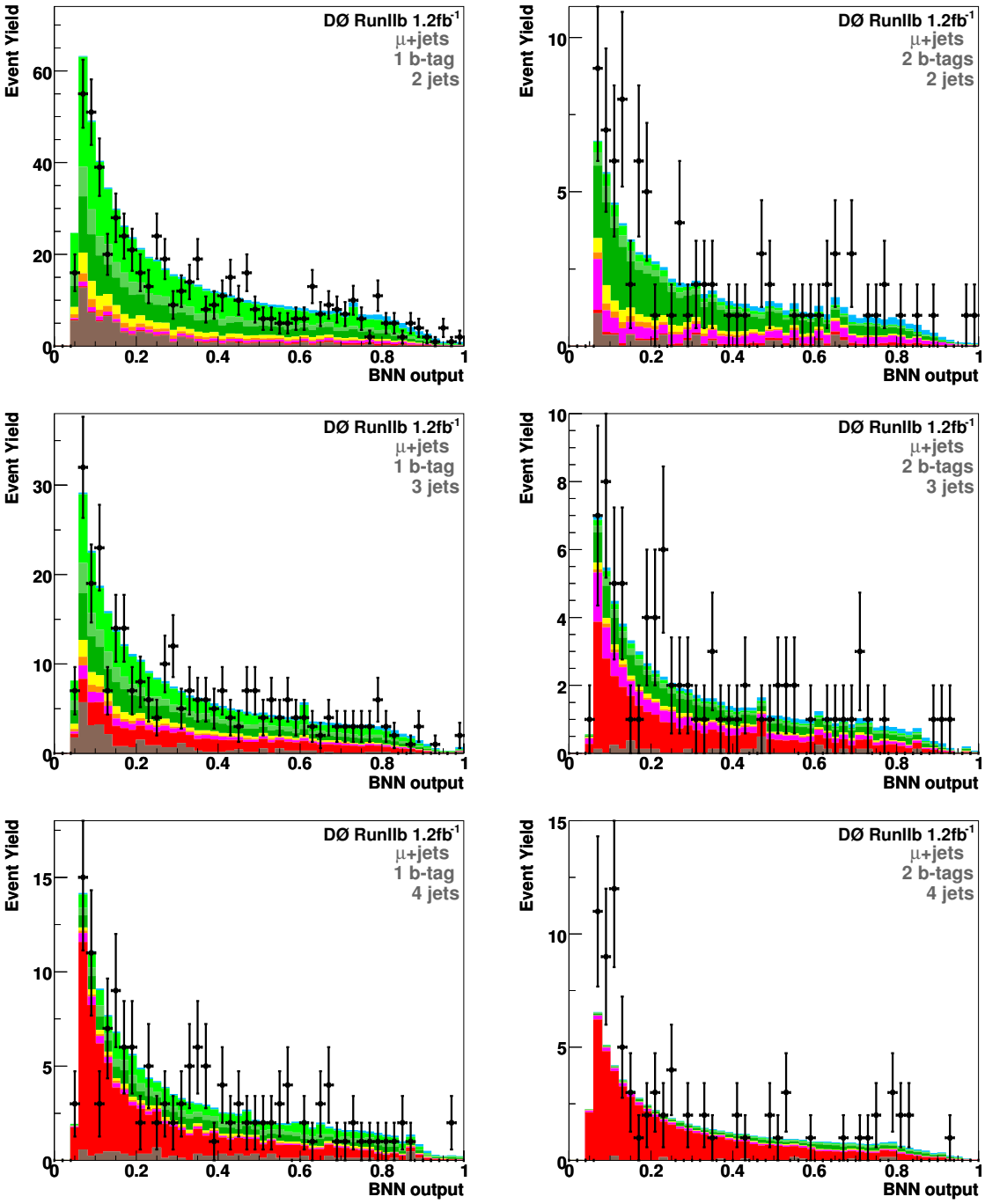

Figure G.7 : Bayesian neural network output distributions for the μ+jets channel in Run IIb data shown in linear scale. The total signal contribution is normalized to the measured cross section, while the relative tb to tqb contributions are given by the ratio of their SM cross sections. Different rows correspond to different jet multiplicities: 2 jets (up), 3 jets (middle), 4 jets (down). The left column corresponds to 1 b-tag and the right column to 2 b-tags.

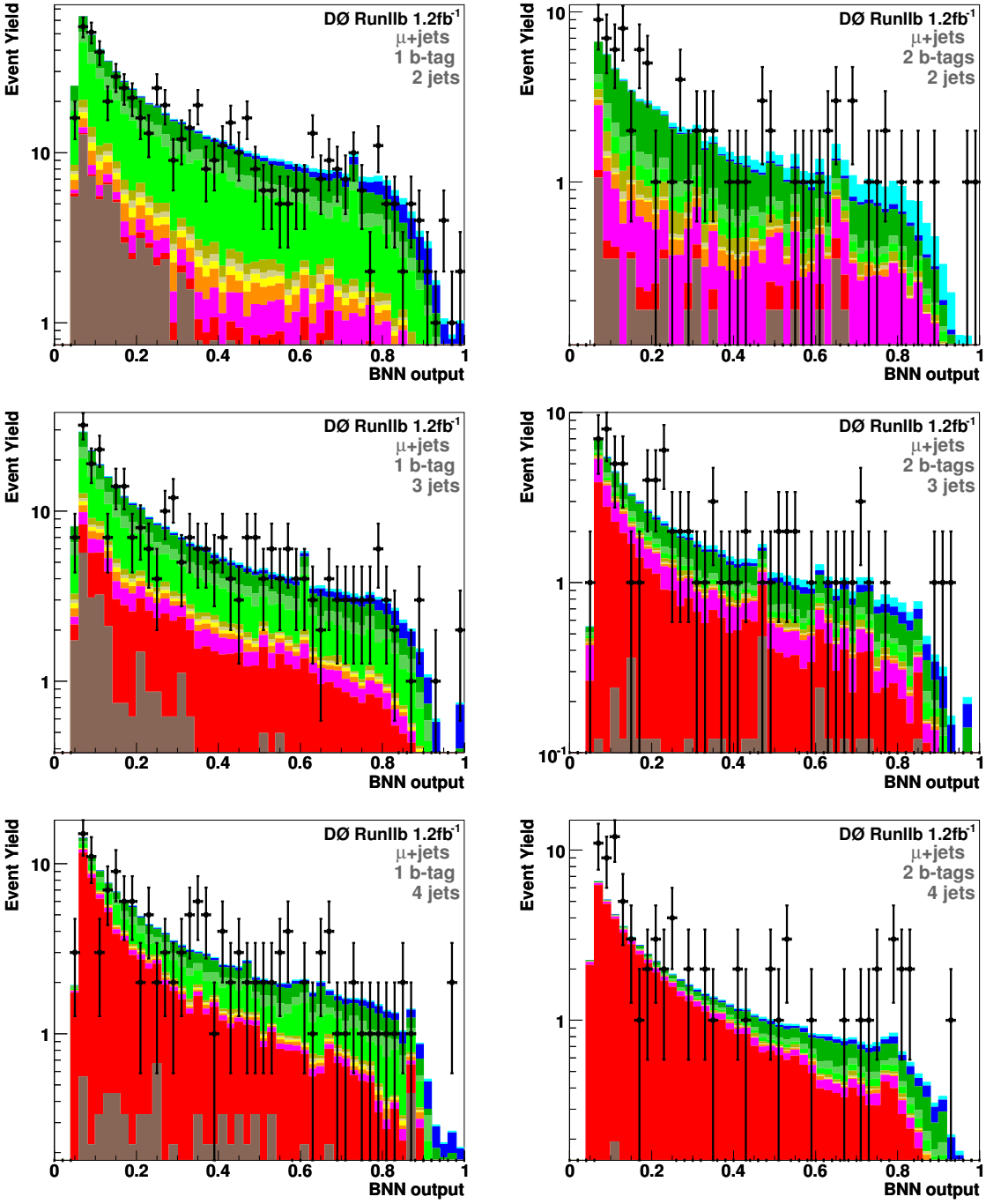

Figure G.8 : Bayesian neural network output distributions for the μ+jets channel in Run IIb data shown in logarithmic scale. The total signal contribution is normalized to the measured cross section, while the relative tb to tqb contributions are given by the ratio of their SM cross sections. Different rows correspond to different jet multiplicities: 2 jets (up), 3 jets (middle), 4 jets (down). The left column corresponds to 1 b-tag and the right column to 2 b-tags.

References

[1] C. Amsler *et al.* (Particle Data Group), Phys. Lett. B **667**, 1 (2008) and 2009 partial update for the 2010 edition.

[2] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1961);
M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).

[3] S. Weinberg, *"The Quantum Theory of Fields, Vol. I"*, (Cambridge University Press, New York, 1996).

[4] N. Kidonakis, *"Single Top Quark Production at the Fermilab Tevatron: Threshold Resummation and Finite-Order Soft Gluon Corrections"*, Phys. Rev. D **74**, 114012 (2006) [arXiv:hep-ph/0609287].

[5] The CTEQ Collaboration,
`http://www.phys.psu.edu/~cteq/`

[6] MRS/MRST/MSTW Parton Distributions,
`http://durpdg.dur.ac.uk/hepdata/mrs.html`

[7] A.D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, Phys. Lett. B **604**, 61 (2004) [arXiv:hep-ph/0410230].

[8] A.P. Heinson, A.S. Belyaev and E.E. Boos, Phys. Rev. D **56**, 3114 (1997) [arXiv:hep-ph/9612424].

[9] T.M.P. Tait and C.-P. Yuan, Phys. Rev. D **63**, 014018 (2001) [arXiv:hep-ph/0007298].

[10] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).

[11] S.W. Herb *et al.*, Phys. Rev. Lett. **39**, 252 (1977).

[12] F. Abe *et al.* [CDF Collaboration], Phys. Rev. Lett. **74**, 2626 (1995).

[13] S. Abachi *et al.* [DØ Collaboration], Phys. Rev. Lett. **74**, 2632 (1995).

[14] V.M. Abazov *et al.* [DØ Collaboration], Phys. Rev. Lett. **103**, 092001 (2009) [arXiv:hep-ex/0903.0850].

[15] T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. Lett. **103**, 092002 (2009) [arXiv:hep-ex/0903.0885].

[16] D. Acosta *et al.* [CDF Collaboration], Phys. Rev. Lett. **95**, 102002 (2005) [arXiv:hep-ex/0505091].

[17] V.M. Abazov *et al.* [DØ Collaboration], Phys. Lett. B **639**, 616 (2006) [arXiv:hep-ex/0603002].

[18] C. Amsler *et al.*, Phys. Lett. B **667**, 1 (2008).

[19] J. Alwall *et al.*, Eur. Phys. J. C **49**, 791 (2007).

[20] S. Abachi *et al.* [DØ Collaboration], Nucl. Instrum. Meth. A **338**, 185 (1994).

[21] V.M. Abazov *et al.* [DØ Collaboration], Nucl. Instrum. Meth. A **565**, 463 (2006) [arXiv:physics/0507191].

[22] R. Wigmans, *"Calorimetry: Energy Measurement In Particle Physics"*, (Oxford University Press, New York, 2000).

[23] A. Khanov, *"HTF: histogramming method for finding tracks. The algorithm description."*, DØ Note 3778 (2000).

[24] D. Adams, *"Finding Tracks"*, DØ Note 2958 (1998).

[25] G. Borissov, *"Ordering a Chaos or... Technical Details of AA Tracking"*, Presentation in the 'All DØ Meeting' (February 28, 2003).

[26] H. Greenlee, *"The DØ Kalman Track Fit"*, DØ Note 4303 (2004).

[27] H. Greenlee, *"The DØ Interacting Propagator"*, DØ Note 4293 (2003).

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

[29] Y. Peters, A. Schawartzman and M. Strauss, *"Certification of the Adaptive Primary Vertex in p17"*, DØ Note 5192 (2006).

[30] A. Schwartzman and M. Narain, *"Vertex Fitting by Means of the Kalman Filter Technique"*, DØ Note 3907 (2001).

[31] A. Schwartzman and M. Narain, *"Probabilistic Primary Vertex Selection"*, DØ Note 4042 (2002).

[32] A. Schwartzman and M. Narain, *"Secondary Vertex Reconstruction using the Kalman Filter"*, DØ Note 3908 (2001).

[33] B. Olivier, U. Bassler and G. Bernardi, *"NADA: A New Event by Event Hot Cell Killer"*, DØ Note 3687 (2000).

[34] G. Bernardi and S. Trincaz-Duvoit, *"Improvement of the NADA Algorithm: Hot Cell Killing in the DØ RunII Data"*, DØ Note 4057 (2002).

[35] U. Bassler and G. Bernardi, *"Towards a Coherent Treatment of Calorimetric Energies: Missing Transverse Energy, Jets, E.M. Objects and the T42 Algorithm"*, DØ Note 4124 (2003).

[36] J.-R. Vlimant, U. Bassler, G. Bernardi and S. Trincaz-Duvoid, *"Technical Description of the T42 Algorithm for the Calorimeter Noise Suppression"*, DØ Note 4146 (2003).

[37] G. Bernardi, E. Busato and J.-R. Vlimant, *"Improvements from the T42 Algorithm on Calorimeter Objects Reconstruction"*, DØ Note 4335 (2004).

[38] J. Hays, J. Mitrevski, C. Schwanenberger and T. Toole, *"Single Electron Efficiencies in p17 Data and Monte Carlo"*, DØ Note 5025 (2006).

[39] J. Kozminski *et al.*, *"Electron Likelihood in p14"*, DØ Note 4449 (2004).

[40] M. Aoki, *"Electron Likelihood p20"*, DØ Note 5675 (2009).

[41] A. Kumar *et al.*, *"Electron Likelihood Study"*, DØ Note 4769 (2005).

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

[43] G. Hesketh for the DØ Muon Algorithm and Identification groups, *"Content of the p17 Muon Thumbnail"*, DØ Note 4735 (2005).

[44] S. Cho, F. Deliot, M. Eads, D. Hedin and H. Lee, *"Muon Identification Certification for p20 data"*, DØ Note 5824 (2008).

[45] We use the iterative, seed-based cone algorithm including midpoints, as described in p. 47 in G.C. Blazey *et al.*, in Proceedings of the Workshop: *"QCD and Weak Boson Physics in Run II"*, edited by U. Baur, R.K. Ellis and D. Zeppenfeld, FERMILAB-PUB-00-297 (2000).

[46] JES Group, *"Jet Energy Scale Determination at DØ Run II (final p17 version)"*, DØ Note 5382 (2007).

[47] T. Scanlon, *"A Neural Network b-tagging Tool"*, DØ Note 4889 (2005).

[48] M. Anastasoaie, S. Robinson and T. Scanlon, *"Performance of the NN b-tagging Toll on p17 Data"*, DØ Note 5213 (2007).

[49] T. Gadfort, A. Haas, D. Johnston, D. Lincoln, T. Scanlon and S. Schlobohm, *"Performance of the DØ NN b-tagging Tool on p20 Data"*, DØ Note 5554 (2008).

[50] K. DeVaughan *et al.*, *"Jet Energy Scale Determination for DØ Run IIb"*, DØ Note 5801 (2008).

[51] T. Sjöstrand, L. Lönnblad, S. Mrenna and P. Skands, *"PYTHIA 6.3 Physics and Manual"*, arXiv:hep-ph/0308153 (2003).

[52] J. Pumplin *et al.*, *"New Generation of Parton Distributions with Uncertainties from Global QCD Analysis"*, JHEP 0207, 012 (2002) [arXiv:hep-ph/0201195].

[53] D. Bandurin, *"The Isolated Photon Cross Section in the Central Rapidity Region at 1.96 TeV"*, DØ Note 4672 (2004).

[54] R. Field and R. Craig Group, *"PYTHIA Tune A, HERWIG, and JIMMY in Run 2 at CDF"*, arXiv:hep-ph/0510198 (2005).

[55] V.M. Abazov *et al.* [DØ Collaboration], Phys. Rev. Lett. **94**, 221801 (2005) [arXiv:hep-ex/0409040].

[56] D. Acosta *et al.* [CMS Collaboration], Note 2006-067 (2006).

[57] A. Bhatti *et al.*, *"Determination of the Jet Energy Scale at the Collider Detector at Fermilab"*, Nucl. Instrum. Meth. A 566, 375 (2006) [arXiv:hep-ex/0510047].

[58] M. Voutilainen, Ph.D. thesis, University of Helsinki, (2008).

[59] A. Kupco, C. Royon, M. Voutilainen and M. Wobisch, *"Measurement of the Inclusive Jet Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV"*, DØ Note 4751 (2005).

[60] R. Brun and F. Carminati, *"GEANT: Detector Description and Simulation Tool"*, CERN Program Library Long Writeup, Report No. W5013 (1993).

[61] C. Royon and M. Voutilainen, *"Jet p_T resolution using JetCorr v7.1"*, DØ Note 5381 (2007).

[62] B. Casey *et al.*, *"Determination of the Run IIb Luminosity Constants"*, DØ Note 5559 (2007).

[63] Common Samples Group web-page for p17 (Run IIa) data:
`http://www-d0.fnal.gov/Run2Physics/cs/skimmming/fixPass2p170903.html`
Common Samples Group web-page for p20 (Run IIb) data:
`http://www-d0.fnal.gov/Run2Physics/cs/skimmming/p20_pass2.html`

[64] E. Aguiló *et al.*, *"Search for Single Top Quark Production in 1* fb^{-1} *of Data"*, DØ Note 5285 (2007).

[65] E.E. Boss *et al.*, *"Method for Simulating Electroweak Top-Quark Production Events in the NLO Approximation: SingleTop Generator"*, Phys. Atom. Nucl. **69**, 1317 (2006).

[66] J. Pumplin *et al.*, *"New Generation of Parton Distributions with Uncertainties from Global QCD Analysis"*, J. High Energy Phys. **0207**, 012 (2002).

[67] T. Sjöstrand, S. Mrenna and P. Skands, *"PYTHIA 6.4 Physics and Manual"*, J. High Energy Phys. **0608**, 026 (2006).

[68] S. Jadach *et al.*, *"The Tau Decay Library TAUOLA: version 2.4"*, Comput. Phys. Commun. **76**, 361 (1993).

[69] D.J. Lange, *"The EvtGen Particle Decay Simulation Package"*, Nucl. Instrum. Meth. A **462**, 152 (2001).

[70] M.L. Magnano *et al.*, *"ALPGEN, a Generator for Hard Multiparton Processes in Hadronic Collisions"*, J. High Energy Phys. **0307**, 001 (2003).

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

[72] D. Gillberg, *"Heavy Flavour Removal and Determination of Weighting Factors for* ALPGEN *W+jets Monte Carlo"*, DØ Note 5129 (2006).

[73] N. Kidonakis and R. Vogt, *"Next-to-Next-to-Leading Order Soft Gluon Corrections in Top Quark Hadroproduction"*, Phys. Rev. D **68**, 114014 (2003).

[74] The MCFM (N)NLO calculations of the diboson cross sections and their uncertainties are documented at: `http://www-clued0.fnal.gov/~nunne/cross-sections/mcfm_cross-sections.html`. This is MCFM version 3.4.5.

[75] H. Schellman, *"The Longitudinal Shape of the Luminous region at DØ"*, DØ Note 5142 (2006).
H. Schellman, *"Run IIb Longitudinal Beam Shape"*, DØ Note 5540 (2007).

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

[77] D. Boline, *"Jet Multiplicity Dependent Reweighting of the Z boson p_T in p20"*, DØ Note 5569 (2008).

[78] L. Wang, S. Eno, M. Sanders and J. Zhu, *"Measurement of the Shape of the Boson Transverse Momentum Distribution using $p\bar{p} \to Z/\gamma^* \to ee$ Events at the Tevatron Collider"*, DØ Note 5187 (2006).

[79] J. Hays *et al.*, *"Single Electron Efficiencies in p17 Data and Monte Carlo using p18.05.00 d0correct"*, DØ Note 5105 (2006).

[80] O. Atramentov *et al.*, *"Electron and Photon Identification with p20 Data"*, DØ Note 5761 (2008).

[81] N. Makovec and J.-F. Grivaz, *"Shifting, Smearing and Removing Simulated Jets"*, DØ Note 4914 (2005).

[82] E. Barberis, T. Golling, I. Iashvili, A. Juste, A. Quadt and P. Schieferdecker, *"The Matrix Method and its Error Calculation"*, DØ Note 4564 (2004).

[83] Y. Peters *et al.*, *"Study of the W+Jets Heavy Flavor Scale Factor in p17"*, DØ Note 5406 (2007).

[84] C. Gerber, E. Shabalina and G. Otero Y Garzón, *"Taggability in p14 Pass2 Data"*, DØ Note 4995 (2006).

[85] E. Aguiló *et al.*, *"Single Top Quark Production in 2.3* fb^{-1} *of Data — Signal and Background Modeling and Event Selection"*, DØ Note 5810 (2008).

[86] Z. Sullivan, *"Angular correlations in single-top quark and Wjj production at next-to-leading order"*, Phys. Rev. D **72**, 094034 (2005).

[87] N. Kidonakis and R. Vogt, *"Theoretical Top Quark Cross Section at the Fermilab Tevatron and the CERN LHC"*, Phys. Rev. D **78**, 074005 (2008).

[88] The Tevatron Electroweak Working Group for the CDF and DØ Collaborations, *"Combination of the CDF and DØ Results on the Mass of the Top Quark"*, arXiv:hep-ex/0808.1089 (2008).

[89] Y. Peters, M. Begel, K. Hamacher and D. Wicke, *"Reweighting of the Fragmentation Function for the DØ Monte Carlo"*, DØ Note 5325 (2007).

[90] R.M. Neal, *"Bayesian Learning of Neural Networks"*, (Springer-Verlag, New York, 1996).

[91] S.B. Beri, R. Kaur and H.B. Prosper, *"Bayesian Neural Networks"*, DØ Note 4846 (2006).

[92] P.C. Bhat and H.B. Prosper, *"Bayesian Neural Networks in Statistical Problems in Particle Physics, Astrophysics and Cosmology"*, ed. L. Lyons and M.K. Ünel, (Imperial College Press, London, 2006).

[93] E. Aguiló *et al.*, *"Observation of Single Top Quark Production in 2.3* fb^{-1} *of Data using Boosted Decision Trees"*, DØ Note 5811 (2008).

[94] D. Gillberg, Ph.D. thesis, Simon Fraser University, (2009).

[95] J.A. Benitez, Ph.D. thesis, Michigan State University, (2009).

[96] E. Aguiló *et al.*, *"Observation of Single Top Quark Production Using The Matrix Element Analysis Technique in 2.3* fb^{-1} *of Data"*, DØ Note 5813 (2008).

[97] M. Pangilinan, Ph.D. thesis, Brown University, (in preparation).

[98] E. Aguiló *et al.*, *"Combination of BDT, BNN and ME Single Top Analyses in 2.3 fb^{-1} of Data using a Multivariate Classifier"*, DØ Note 5814 (2009).

[99] R. Barlow, *"Event Classification Using Weighting Methods"*, J. Comp. Phys. **72**, 202 (1987).

[100] E.K. Blum and L.K. Li, *"Approximation theory and feedforward networks"*, Neural Networks, **4**, 511-515 (1991).

[101] C.M. Bishop, *"Neural Networks for Pattern Recognition"*, (Clarendon Press, Oxford, 1998).

[102] H.B. Prosper, *"Some Mathematical Comments on Feed-Forward Neural Networks"*, DØ Note 1606 (1993);
H.B. Prosper, *"More Mathematical Comments on Feed-Forward Neural Networks"*, DØ Note 4493 (2004).

[103] J.H. Friedman,*"Separating Signal From Background Using Ensembles Of Rules, in Statistical Problems in Particle Physics, Astrophysics and Cosmology"*, ed. L. Lyons and M.K. Ünel, (Imperial College Press, London, 2006);
`http://www-stat.stanford.edu/~jhf/R-RuleFit.html`

[104] TMVA (Toolkit for Multivariate Data Analysis with ROOT) Home web page: `http://tmva.sourceforge.net/`

[105] B.A. Berg, *"Markov Chain Monte Carlo Simulations and Their Statistical Analysis"*, (World Scientific Singapore, 2004).

[106] S. Duane, A.D. Kennedy, B.J. Pendleton and D. Roweth, *"Hybrid Markov Chain Monte Carlo"*, Phys. Lett. **B** 195, 216-222 (1987).

[107] C. Peterson, T. Rögnvaldsson and L. Lönnblad, *"JETNET 3.0-A Versatile Artificial Neural Network Package"*, Computer Physics Communications **81**, 185-220 (1994).

[108] S. Jain, H. Prosper and R. Schwienhorst, *"Statistical Methods Implemented in the Package* `top_statistics`*"*, DØ Note 5817, (2008).

[109] J. Alwall *et al.*, Eur. Phys. J. **C** 49, 791 (2007).

[110] G.L. Kane, F.A. Ladinsky and C.P. Yuan, Phys. Rev. D **45**, 124 (1992);
K. Whisnant *et al.*, Phys. Rev. D **56**, 467 (1997).

[111] M-A. Pleier, *"Measuremente of the Electron and Muon Fake Rates in Lepton+Jets Datasets"*, DØ Note 5469 (2007).