

Laboratoire de Physique Nucléaire et des Hautes Énergies
Université Paris VII - Denis Diderot
UFR de Physique
École doctorale 517 - "Particules, Noyaux et Cosmos"

THÈSE DE DOCTORAT
Spécialité : **Physique des particules**

présentée par
Jonathan BROWN

Search for WH associated production in the $\ell\nu b\bar{b}$ final state using the DØ detector at the Tevatron

pour obtenir le grade de Docteur de l'Université Paris VII

soutenue le 15 septembre 2011 devant le jury composé de

Gregorio	BERNARDI	Directeur de thèse
Aurelio	JUSTE	Rapporteur
Reynald	PAIN	Président de jury
Yves	SIROIS	Examineur
Isabelle	WINGERTER-SEEZ	Rapporteur
Fabian	ZOMER	Examineur

To my grandfather Joseph

Acknowledgements

Je souhaite en premier lieu adresser mes remerciements à mon directeur de thèse Gregorio Bernardi, pour m'avoir permis de passer trois ans au sein du groupe qu'il dirige. J'ai ainsi eu la chance de pouvoir rester pendant plus d'un an à Fermilab où j'ai pu apprécier un cadre de travail dans lequel je me suis plu, me permettant de satisfaire ma curiosité. Je suis reconnaissant de son soutien et son optimisme qui m'ont permis de donner le meilleur de moi-même. Je remercie également Reynald Pain, en tant que directeur du LPNHE, ainsi que pour avoir gentiment accepté d'être le président du jury de thèse. Aurelio Juste et Isabelle Wingerter-Seez ont gracieusement accepté d'être rapporteurs de la thèse et je les remercie d'avoir pris le temps de participer au jury, ainsi que Fabian Zomer et Yves Sirois qui les ont rejoint.

Bien qu'ayant passé une grande partie de la thèse à Chicago, je souhaite remercier toutes les personnes que j'ai pu rencontrer au laboratoire. Philippe Schwemling, en tant que directeur de l'école doctorale, m'a énormément aidé et conseillé, je l'en remercie. Je remercie aussi Lydia Roos pour sa sympathie. Je garderai d'excellents souvenirs de l'école FAPPS dont elle est l'organisatrice. Eli Ben-Haim a été mon enseignant en Master, sa gentillesse et son enthousiasme ont évidemment été des éléments qui m'ont poussé dans cette voie.

J'ai effectué mes premiers pas de doctorant au LPNHE avec Stefania Bordoni et François Brun. Bien que ce fut souvent bref, j'ai apprécié les moments que nous avons partagé au labo ou en dehors. Je salue aussi les autres étudiants du laboratoire, vendeurs de café, représentant des étudiants ou buveurs de thé. Enfin, merci à Sébastien Bongard, pour sa camaraderie et avoir pu supporter un étudiant en fin de thèse dans son bureau durant quelques mois!

Je dois énormément à Sophie Trincaz-Duvoid et Didier Lacour. Leur soutien à travers les discussions que nous avons eu m'a aidé et ils ont toujours trouvé les mots justes pour m'encourager. Je remercie également l'équipe administrative du LPNHE pour leur aide, bonne humeur et patience, en particulier lors de mes retours de mission!

Durant la thèse, j'ai eu la chance de rencontrer des personnes avec qui j'ai eu le plaisir de passer du temps, au bureau ou en dehors. Tout d'abord, je ne pourrai oublier mes "prédécesseurs" Michiel Sanders, Jérémie Lellouch et Nils Huske. Merci pour votre gentillesse et votre aide, m'avoir fait découvrir l'analyse *WH* ainsi que Fermilab et d'avoir jeté un (énorme) coup d'oeil sur le manuscrit. Evidemment, mon apprentissage du travail d'analyse s'est fait grâce au soutien du groupe *WH*: Duncan Brown, Mike Cooke, Sébastien Greder, Ken Herner et en particulier Yuji Enari. Son aide, sa gentillesse et sa sympathie ont été un énorme soutien ainsi qu'un moteur. Je suis extrêmement heureux d'avoir pu faire sa connaissance et travailler à ses côtés. J'espère avoir la chance de répéter l'expérience. Arigato-gosaimasu! Je souhaite aussi bon courage et bonne continuation aux actuels

ACKNOWLEDGEMENTS

doctorants du groupe, Florian Miconi, Dikai Li et Huong Nguyen. Aussi, un grand bonjour (dans leur langue préférée) aux étudiants que j'ai rencontré au Japon et Lake Louise!

A Fermilab, l'accueil du groupe de Saclay, pour manger, se détendre et parfois même dormir, a toujours été sympathique. Je salue la bande de thésards saclaysiens: Cécile, Solène, Aurélien et Romain. Je suis heureux que l'on ait partagé de bons moments ensemble, de pouvoir discuter, s'aider et découvrir des choses avec vous (et souvent grâce à vous). Je suis certain que l'on aura à nouveau l'opportunité de passer une soirée ensemble, boire un café, ou manger du Nut***a dans une control room! Merci aussi à Christophe, pour sa sympathie, son aide à plusieurs reprises et les discussions que nous avons pu avoir. Je voudrais aussi remercier la communauté DØ France, pour leur gentillesse et les échanges que nous avons pu avoir.

Bien entendu, mon séjour a été plus agréable grâce à la présence de certaines personnes: Katy Tschann-Grimm, Gabriel Facini, Jorge Martinez et Murilo Rangel. Thanks for the good time we had together and hope to see you soon! Many thanks also to Ruti and Meir, for their help and welcoming me as part of the family, I owe you so much!

Je voudrais saluer et remercier les personnes qui m'ont accompagné avant et pendant la thèse, j'ai apprécié leur soutien et amitié. Tout d'abord Julien et Laurette, j'ai tellement de raisons pour lesquelles je dois les remercier, je leur dois énormément que je ne saurais pas comment leur rendre un dixième de ce qu'ils ont fait pour moi. Un grand merci aussi à Thomas et Marianne, et bon courage à vous pour la suite, les galères en RER et la fin de la thèse! Merci aussi à Max et Mathieu d'être venu pour la soutenance, et aussi à Chloé, en plus de m'avoir permis de passer un été Aloha! Bien que je n'ai pas pu souvent les voir durant ces trois dernières années, je salue très chaleureusement mes amis du Secteur: Adri, Marie, Dric, Sandra, Nano, Aurélie, Pierrot, Louloute, Eric, Herman, Yuv (tu reviens quand?), KRP, Julie et Vince. Ce fut toujours un plaisir de vous voir après mes absences prolongées, en particulier durant cette semaine en Espagne qui fut bonne et bien trop courte! Merci à vous d'avoir tout fait pour vous intéresser à mon travail, même si parfois ce fut pour en rire!

Je n'ai plus besoin d'anticiper en disant félicitations à Amélie et Stéphane pour leur jolie petite Alice. A très vite! Pour avoir habité avec eux, je remercie Arnaud, Joe et Jennifer pour tous les bons moments et délires que l'on a partagé et m'avoir aidé à me changer les idées en leur compagnie. Je pense que l'on partage une bonne quantité de souvenirs que je n'oublierai pas de sitôt! Je souhaite également remercier Bernard et Patricia pour m'avoir aidé, hébergé, nourri et donné leurs commentaires pendant la répétition de la soutenance.

Enfin, je ne serais pas en train d'écrire ces lignes sans l'aide et le soutien de mes parents et ma famille. Cela m'a toujours donné la force d'aller de l'avant et faire ce que j'aime. Je vous dois énormément pour tout ce que vous avez fait pour moi. Merci beaucoup à vous.

Contents

Acknowledgements	5
Introduction	11
1 The Standard Model and the Higgs boson	13
1.1 The Standard Model	13
1.1.1 The elementary constituents of matter	13
1.1.2 The fundamental forces	14
1.2 The Quantum Electrodynamics Field Theory	15
1.3 The Quantum Chromodynamics Field Theory	16
1.4 The electroweak sector	17
1.5 The Higgs mechanism	18
1.5.1 The scalar Higgs field	18
1.5.2 Mass generation for the Standard Model particles	20
1.6 Higgs searches	21
1.6.1 Theoretical constraints	21
1.6.2 Experimental constraints	23
1.7 Conclusion	27
2 Experimental apparatus : The Tevatron and the DØ detector	29
2.1 The chain of accelerators and the Tevatron	29
2.1.1 Proton beam production	30
2.1.2 The Main Injector and Recycler	32
2.1.3 Antiproton beam production	32
2.1.4 The Tevatron	34
2.2 The DØ detector	37
2.2.1 Coordinate system	38
2.2.2 Tracking and vertexing system	38
2.2.3 Preshower detectors	41
2.2.4 Calorimeter	43
2.2.5 Muon system	46
2.2.6 Luminosity monitor	48
2.2.7 Trigger and data acquisition system	50
2.2.8 Data format and detector simulation	52

3	Objects reconstruction and identification	55
3.1	Tracks	55
3.2	Primary Vertices	57
3.3	Electrons	58
3.3.1	Reconstruction and identification criteria	58
3.3.2	Electron resolution and energy scale correction	59
3.3.3	Identification efficiency between data and simulation	60
3.4	Muons	60
3.4.1	Reconstruction and identification criteria	60
3.4.2	Muon energy resolution	63
3.4.3	Reconstruction efficiency in data and simulation	63
3.5	Jets	63
3.5.1	Jet reconstruction	64
3.5.2	Jet identification and vertex confirmation	66
3.5.3	Jet Energy Scale	69
3.5.4	Jet Shifting, Smearing and Removal	75
3.6	Missing transverse energy	80
3.7	b -jet identification	80
3.7.1	b -jets properties	80
3.7.2	b jet taggability	83
3.7.3	Individual b -jet identification algorithms	83
3.7.4	The NN b -tagger	89
4	Selection and physics processes modeling in the WH analysis	95
4.1	Overview	95
4.2	Foreword on the analysis work flow	97
4.3	Data and Monte Carlo used in the WH analysis	97
4.3.1	Data samples	97
4.3.2	Monte Carlo samples and generators	97
4.3.3	Trigger selection	106
4.4	Event selection	106
4.4.1	Primary vertex selection	107
4.4.2	Lepton selection	108
4.4.3	Missing E_T selection	109
4.4.4	Jet selection	109
4.4.5	Triangular cut	110
4.4.6	Vetoos	110
4.5	Reweightings of W +jets and Z +jets samples	112
4.6	Multijet background estimation	117
4.6.1	Multijet background modeling strategy	118
4.6.2	Lepton fake rates	119
4.7	Simulation normalization scheme	123
4.7.1	Multijet sample normalization	123

4.7.2	Experimental K factors for the W +jet background	123
4.7.3	Heavy flavor scale factor	123
4.8	Pre-tag summary	129
4.9	b -tagging applied in the WH analysis	131
4.9.1	Taggability scale factors	131
4.9.2	b -tagging efficiency scale factors	132
4.9.3	Post b -tagging results	137
5	Multivariate classification, limit derivation and results for the WH analysis	145
5.1	Improving sensitivity using a Random Forest	146
5.1.1	Principle of a Random Forest	146
5.1.2	Input variables and training parameters	148
5.1.3	Rebinning of the Random Forest output	151
5.1.4	Random Forest distributions	157
5.2	The CL_s method	160
5.2.1	Principle of the method	160
5.2.2	Systematic uncertainties : sources and treatment	165
5.3	Limits obtained in the WH analysis	173
6	Future improvements and prospects	179
6.1	Improvements in the WH analysis	180
6.1.1	Jet Energy Resolution	180
6.1.2	Changes in the analysis	195
6.1.3	Summary of potential improvements	208
6.2	Prospects at the Tevatron and LHC	209
6.2.1	Prospects at the Tevatron	209
6.2.2	Prospects at the LHC	210
6.2.3	Conclusion	213
	Conclusion	215
	Bibliography	217
A	Summary of vetoes	227
B	Electron fake rate parameterizations	231
C	RunIIa kinematic variables distributions	237
D	Random Forest input variables distributions	247
E	Random Forest output distributions	259
F	Systematic uncertainties distributions	265
F.1	RunIIb electron 2 jets channel	266

CONTENTS

F.2 RunIIb muon 2 jets channel	280
--	-----

Introduction

Since the beginning of Science, physicists have been trying to establish the foundations of a theory which can describe Nature as it appears in its most fundamental manifestations. Human instinct and curiosity questioned observed phenomena, and how a theoretical framework, relying on a few fundamental principles, could describe them.

In modern Physics, such framework should be valid in the Quantum Field Theory framework and give a fundamental description of different observations in the most unified way.

One major breakthrough was in proposing a model comprised of elementary matter particles and interaction particles which are fundamental force carriers. The so-called *Standard Model* describes the electromagnetic, weak and strong interactions, where quark particles can interact through all of them, whereas leptons undergo only electromagnetic and weak forces.

As technological developments were going on in particle acceleration and detection, the confidence in the Standard Model became stronger. Its validity was proven when particles predicted by the Standard Model, such as the W and Z particles or heavier quarks from the 3rd generation, have been discovered and their properties measured with a high precision.

However, even if the Standard Model is successful in predicting the particle content of Nature so far, the theory is only correct for describing massless particles. An additional term in the theory has to be introduced in order to generate massive particles. A new scalar field is proposed to generate a mass for each particle, also known as the Higgs field. The associated particle, the Higgs boson, has also a mass term, but is a free parameter of the theory.

Searches for the Higgs boson have started as soon as the underlying model for particle mass generation has been proposed. Starting from very low energies, this work still continues to the highest energy available with today's technology since no evidence has appeared yet on its existence.

The work presented here is part of the ongoing effort on the search of the Higgs boson, in the $WH \rightarrow \ell \nu b \bar{b}$ channel which is sensitive to the low mass Higgs boson production. First, the theoretical framework will be introduced in Section 1, as well as the proposed mechanism which allows to generate mass terms for particles, called the electroweak spontaneous symmetry breaking.

The instrumental tools used to conduct this search are then presented in Section 2. Particles can be created from $p\bar{p}$ collisions as it is the case at the Tevatron accelerator, located at the Fermi National Accelerator Laboratory. As the collisions energy is high from proton and antiproton beams circulating in the accelerator close to the speed of light, massive particles can be created, before decaying in a very short time scale, typically between 10^{-10} and 10^{-20} s for relevant processes in this type of study. The decay products are generally highly energetic. In order to detect such events, the DØ detector is used and well suited for this purpose.

In order to probe a specific process, objects from its decays need to be accurately reconstructed and identified in order to be analyzed. This step requires a good understanding of the detector. For example, in the specific case of a Higgs boson decaying in a $b\bar{b}$ pair, jets produced in the final state might be identified as originating from b quarks, therefore reducing the background from jets originated from other partons. Moreover, the final state of the topology which is searched contains as well leptons, the signature in the detector of all physics objects and how they are identified will be discussed in Section 3.

After collecting data for the $p\bar{p}$ collisions using the DØ detector, selected events are analyzed and have to be compared to the simulation of physical processes since the data is composed of backgrounds with a potential signal contribution. A particular attention has to be drawn on the modeling of the simulation, in order to not bias the final result in case of a mismatch between data and simulation, thus leading to a false result (either discovery or exclusion of the searched signal in the worse case). Section 4 will focus on this crucial aspect, as well as how the background can be reduced.

The last step of the analysis is presented in Section 5. The sensitivity is additionally increased by the usage of a multivariate analysis technique, which brings discrimination between signal-like and background-like events. Since no significant excess consistent with signal is observed, upper limits are set on the cross section of the WH process times the $H \rightarrow b\bar{b}$ branching ratio, across a Higgs boson mass M_H ranging from 100 to 150 GeV. The observed limit corresponds to a cross section 4.6 times higher than the prediction from the Standard Model.

In order to increase the sensitivity of the analysis, several attempts have been conducted within the DØ collaboration and the WH analysis. Some of them will be discussed in Section 6. Finally, the Tevatron is expected to shutdown in September 2011, while the LHC collider has started to collect data. Prospects on the Higgs boson search from both experiments will be mentioned.

This result is based on 5.3 fb^{-1} of integrated luminosity, which is about half of the dataset which will be available after the Tevatron will stop collecting data. This search has been published by the DØ collaboration in Physics Letter Review B in February 2011 [1].

1

The Standard Model and the Higgs boson

The Standard Model is the theoretical framework which allows to describe all fundamental constituent of matter and their interaction. This theory relies on the quantum field theory (QFT) formalism and is based on the symmetries of the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ gauge group. The Standard Model is built from the different types of interactions existing in Nature, which are unified in the same gauge group. Since gravity cannot be described by a renormalizable quantum theory so far, it is not included in the Standard Model. After reminding the particle content in the theory, each type of interaction which are part of it will be introduced. The mechanism which generates the mass of the elementary particles of the Standard Model will be presented. Finally, a short overview of status of Higgs boson searches will be given.

1.1 The Standard Model

The elementary particles are divided in two categories. Particles having half-integer spin are called fermions. They obey the Pauli exclusion principle, which states that two fermions cannot be in the same quantum state. The other family of particles are called bosons and have integer spin value. Their nature confers on them a different behaviour. Elementary fermions and bosons are described here.

1.1.1 The elementary constituents of matter

Fermions are further subdivided in two families: leptons and quarks, which are sensitive to different types of interaction. Quarks and leptons are each existing in three generations, with the only distinction being their mass. The first generation of fermions is the most common in the Universe, u and d quarks are present in nucleons and form nuclei. Electrons are around the nuclei to form atoms and the electronic neutrino is produced in large amounts in stars. The second generation is

comprised of the c and s quarks, the muon and its associated neutrino. Finally the third generation is composed of the t and b quark, the tau lepton and its associated neutrino. The t quark is so far the heaviest known elementary particle.

The neutrino has the characteristic of interacting very rarely with matter since it only undergoes the weak interaction. Quarks are paired in each generation where one of them has a $\frac{2}{3}$ electric charge (u , c and t quarks, also denoted as *up* type quarks), whereas the other has a $-\frac{1}{3}$ electric charge (d , s and b quarks, denoted as *down* type quarks). Given their baryonic number quarks can be grouped in $q\bar{q}$ pairs to form mesons or baryons with three quarks, with respectively a baryonic number equals to 0 or 1. This quantum number has to be conserved, similarly to the lepton number. In order to conserve these laws, fermions are produced in particle-antiparticle pair.

A summary of fermions is given in Table 1.1 along with some of their fundamental properties.

	Generation	Particle	Mass (MeV)	Electric charge	Baryonic number	Leptonic number
Quarks	1 st	u	1.7-3.1	2/3	1/3	0
		d	4.1-5.7	-1/3	1/3	0
	2 nd	c	1.29×10^3	2/3	1/3	0
		s	100	-1/3	1/3	0
	3 rd	t	172.9×10^3	2/3	1/3	0
		b	4.19×10^3	-1/3	1/3	0
Leptons	1 st	e	0.511	-1	0	1
		ν_e	<2.2	0	0	1
	2 nd	μ	105.66	-1	0	1
		ν_μ	<170	0	0	1
	3 rd	τ	1.78×10^3	-1	0	1
		ν_τ	<15.5	0	0	1

Table 1.1: Summary of Standard Model elementary fermions [2]. Their antiparticles have the mass with an opposite electric charge, baryonic and leptonic number.

1.1.2 The fundamental forces

Three forces are described in the Standard Model, the electromagnetic, weak and strong forces, each of them relying on a different gauge group. This dictates their behavior with other particles. Each force is mediated by a bosonic field which is exchanged between particles when interacting.

The electromagnetic force manifests itself by the exchange of a photon and couples to any electrically charged particle. The strong force is responsible for binding nucleons inside nuclei. Gluons are the strong force carriers which is exchanged between color-charged particles. Only quarks and gluons have a color charge (where the color quantum number can be red, green or blue), therefore undergoing such interactions. This leads to the particularity that gluons can interact with themselves. There are eight gluons arising from the combination of two colors. Nine combinations are possible, however one of them is invariant under rotation in the color space. This leads to an unphysical particle which does not interact. Moreover, the strong force has the feature that its amplitude increases as the spatial separation between particles exchanging gluons is growing. Nuclear decays (for example $p \rightarrow ne^+\nu_e$) are the demonstrations of the weak force.

Particles carrying a weak isospin quantum number, which are all particles introduced here except gluons, undergo such interaction, which is of two types: either through electrically neutral Z boson or through charged W^\pm bosons. Z and W^\pm bosons are the only massive force carrier bosons, the photon and gluons being massless. A brief summary of elementary bosons is given in Table 1.2.

Force	Gauge Boson	Mass (GeV)	Interaction range	Interacts with
electromagnetic	Photon (γ)	0	infinite	charged particles
weak	W^\pm	80.399	$\sim 10^{-18}\text{m}$	quark, leptons, W^\pm , Z
	Z	91.1876		
strong	gluons (g)	0	$\sim 10^{-15}\text{m}$	quarks, gluons

Table 1.2: Summary of Standard Model elementary bosons [2].

1.2 The Quantum Electrodynamics Field Theory

Let us start from the theory described by the simplest gauge group, $U(1)$. The equation describing the motion of a free spin $\frac{1}{2}$ particle, invariant under rotations and boosts, is given by the Dirac equation. Its lagrangian is expressed as

$$\mathcal{L} = \underbrace{i\bar{\psi}\gamma_\mu\partial^\mu\psi}_{\text{kinetic term}} - \underbrace{m\bar{\psi}\psi}_{\text{mass term}}, \quad (1.1)$$

with γ^μ being 4×4 matrices defined from Pauli matrices σ^i :

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad (1.2)$$

This lagrangian is invariant under global phase transformation, or global $U(1)$ symmetry with

$$\psi(x) \rightarrow e^{i\Lambda}\psi(x), \quad \bar{\psi}(x) \rightarrow e^{-i\Lambda}\bar{\psi}(x) \quad (1.3)$$

given Λ is a free parameter. In order to get a more general formulation, the Λ parameter can varied as function of space-time coordinates x . Equation 1.3 is expressed as $\psi(x) \rightarrow e^{i\Lambda(x)}\psi(x)$. Such transformation is denoted as local symmetry or gauge symmetry.

However, the lagrangian invariance is not preserved when requiring local $U(1)$ invariance since an additional term arises when computing the field derivative. Nevertheless, the invariance can be restored by redefining the derivative ∂_μ such as the introduced derivative D_μ , called covariant derivative, would transforms under gauge symmetry as

$$D_\mu \rightarrow e^{i\Lambda(x)}D_\mu, \quad (1.4)$$

thus keeping the lagrangian invariant. Such covariant derivative is defined as

$$D_\mu = \partial_\mu - ieA_\mu \quad (1.5)$$

with e being a coupling constant and A_μ transforming as

$$A_\mu \rightarrow A_\mu + \frac{1}{e} \partial_\mu \Lambda(x)$$

Given the introduction of the covariant derivative, the lagrangian can be expressed as

$$\mathcal{L} = i\bar{\psi}\gamma_\mu D^\mu \psi - m\bar{\psi}\psi \quad (1.6)$$

$$= \underbrace{\bar{\psi}(i\gamma_\mu \partial^\mu - m)\psi}_{\text{free particle motion}} + \underbrace{e\bar{\psi}\gamma_\mu \psi A_\mu}_{\text{coupling to } A_\mu} \quad (1.7)$$

One observe that in order to preserve gauge invariance, an additional term which couples fermions to a new vector field is introduced. This shows that an interaction is naturally introduced in theory, where A_μ is the photon field. The photon is massless since adding a term as $\frac{1}{2}M^2 A_\mu A^\mu$ in the lagrangian would violate gauge invariance.

The last piece in this theory is a kinetic term to describe the A_μ field propagation. Such gauge invariant term can be written using the field tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, leading to the Quantum Electrodynamics (QED) lagrangian:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma_\mu D^\mu - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad (1.8)$$

$$= \underbrace{\bar{\psi}(i\gamma_\mu \partial^\mu - m)\psi}_{\text{free particle motion}} + \underbrace{e\bar{\psi}\gamma_\mu \psi A_\mu}_{\text{coupling to } A_\mu} - \underbrace{\frac{1}{4} F_{\mu\nu} F^{\mu\nu}}_{\text{photon kinetic term}} \quad (1.9)$$

1.3 The Quantum Chromodynamics Field Theory

Following the same steps as in QED, the Quantum Chromodynamics (QCD) lagrangian, describing the strong interaction, will be introduced. This formalism was introduced by Gell-Mann and Zweig in 1963. Based on the $SU(3)$ gauge group, particles composing the strong sector are the u , d , c , s , t and b quarks and gluons. They were proposed in order to describe the hadron spectroscopy. A new quantum number is introduced, the color, which is either red, green or blue, and is only assigned to quarks and gluons. Given the $SU(3)$ local gauge symmetry, eight gluon fields G_μ^a are present. In a similar way as it is describing QED, the QCD lagrangian is given as

$$\mathcal{L} = \sum_q \bar{\psi}_q(i\gamma_\mu D^\mu - m_q)\psi_q - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \quad (1.10)$$

and the covariant derivative is

$$D_\mu = \partial_\mu + ig_s T_a G_\mu^a \quad (1.11)$$

with the index q running over the 6 quark flavors, the index a running over the 8 colors. g_s is the coupling constant and T_a are the generators of the $SU(3)$ group. The field tensor $G_{\mu\nu}^a$ is given as

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc} G_\mu^b G_\nu^c \quad (1.12)$$

where f^{abc} are the structure constant of the $SU(3)$ group, defined from $[T^a, T^b] = if^{abc}T^c$. An important observation has to be made about the product of gluon field tensors expression. As the photon, gluons have a kinetic term. But, an additional term gives them self-interaction terms, leading to three and four legs diagrams between gluons with terms as $g_s f_{abc} G_\mu^a G_\nu^b (\partial^\mu G^{\nu c})$ and $g_s^2 f_{abc} f_{ed}^a G_\mu^b G_\nu^c G^{\mu e} G^{\nu d}$. A direct consequence of the gluon self-interaction is the confinement of quarks and gluons in hadrons, arranged such that the total color charge is zero.

Given these definitions and omitting the summation over k , the QCD lagrangian read

$$\mathcal{L}_{QCD} = \bar{\psi}(i\gamma_\mu D^\mu - m)\psi - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu} \quad (1.13)$$

$$= \underbrace{\bar{\psi}(i\gamma_\mu \partial^\mu - m)\psi}_{\text{free particle motion}} + \underbrace{g_s \bar{\psi} T^a G_\mu^a \gamma_\mu \psi}_{\text{quark-gluon coupling}} - \underbrace{\frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}}_{\text{gluon kinetic + self-interaction terms}} \quad (1.14)$$

1.4 The electroweak sector

Moving towards a theory unification, the electromagnetic and weak forces can be included in the same gauge group $SU(2)_L \otimes U(1)_Y$, which are the weak isospin and hypercharge group. Each fermion field can be decomposed in a left and right-handed component, $\psi = \psi_L + \psi_R$, corresponding to its helicity. In this representation, left-handed fermions are arranged in a doublet, whereas right-handed fermions are singlets. This assumption is based on the observation that only left-handed neutrinos interact with matter. Therefore, each lepton generation is described by

$$\psi_L = \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L, \quad \psi_R = \ell_R^- \quad (1.15)$$

and quark generations by

$$\psi_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \quad \psi_R^u = u_R, \quad \psi_R^d = d_R \quad (1.16)$$

The $U(1)_Y$ group impose an interaction of fermions with the B_μ field for both left and right-handed components, with its group generator Y . The $SU(2)_L$ group introduces three gauge bosons W_μ^a (with $a = 1, 2$ or 3) with $T^a = \frac{1}{2}\sigma^a$ as generators, σ^a being the Pauli matrices. This field couples to left-handed components only. The associated quantum number for the weak isospin is $I_3 = \frac{1}{2}$ and $I_3 = -\frac{1}{2}$, respectively for up and down components for left handed fermions, and 0 for right handed singlets. The hypercharge is defined as $Y = 2(Q - I_3)$.

The associated covariant derivative associated to the $SU(2)_L \otimes U(1)_Y$ gauge invariance reads

$$D_\mu = \partial_\mu + igT_a W_\mu^a + ig' \frac{Y}{2} B_\mu \quad (1.17)$$

for left handed fermions, with g and g' being respectively the coupling constant to the W_μ^a and B_μ field, and

$$D_\mu = \partial_\mu + ig' \frac{Y}{2} B_\mu \quad (1.18)$$

for right handed fermions. Similarly to QCD, the electroweak lagrangian is composed of three terms: the quarks and lepton kinetic term, an interaction term with the weak gauge bosons and the gauge bosons kinetic term. By introducing the $W_{\mu\nu}^a$ and $B_{\mu\nu}$ field tensors, one can notice that similarly to the gluon field, interactions between W fields is possible from

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \quad W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g\epsilon_{abc}W_\mu^b W_\nu^c \quad (1.19)$$

$$\mathcal{L}_{EW} = \underbrace{\bar{\psi}_L \gamma^\mu i(\partial_\mu + igT_a W_\mu^a + ig' \frac{Y}{2} B_\mu) \psi_L}_{\text{left handed fermion kinetic + interaction terms}} + \underbrace{\bar{\psi}_R \gamma^\mu i(\partial_\mu + ig' \frac{Y}{2} B_\mu) \psi_R}_{\text{right handed fermion kinetic + interaction terms}} \quad (1.20)$$

$$- \underbrace{\frac{1}{4} W_a^\mu W_\mu^a - \frac{1}{4} B^\mu B_\mu}_{\text{gauge boson kinetic + } W_\mu \text{ self interaction terms}} \quad (1.21)$$

1.5 The Higgs mechanism

The description given previously is only valid for massless particles and interaction bosons. As previously mentioned, mass terms for bosons would violate their associated gauge symmetry. This statement is also true for fermions, where a term in the lagrangian such as $m(\bar{\ell}_R \ell_L + \bar{\ell}_L \ell_R)$ is not invariant under $SU(2)_L \otimes U(1)_Y$.

In order to generate masses to particles and keeping symmetries preserved in the lagrangian, a formulation has been proposed by Higgs-Englert-Brout-Kibble-Guralnik-Hagen to satisfy such requirements [3, 4, 5]. Such mechanism is called the Higgs mechanism or the electroweak spontaneous symmetry breaking.

1.5.1 The scalar Higgs field

In order to introduce the Higgs mechanism, an assumption is made that the Higgs field forms an isospin doublet $\phi(x)$ of complex scalar fields:

$$\phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \quad (1.22)$$

undergoing a potential $V(\phi)$:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (1.23)$$

The lagrangian describing the dynamics of such field is given as:

$$\mathcal{L}_{Higgs} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi) \quad (1.24)$$

where the covariant derivative, with $Y_H = 1$, is the same as the one defined in Equation 1.17:

$$D^\mu = \partial_\mu - igT_a W_\mu^a - ig' \frac{Y_H}{2} B_\mu. \quad (1.25)$$

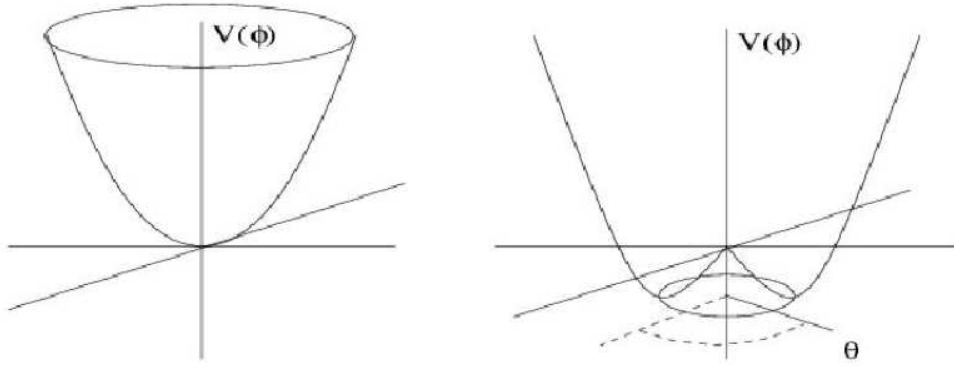


Figure 1.1: Left figure: local minimum of the Higgs potential for $\mu^2 > 0$ at $\phi = 0$. Right figure: for $\mu^2 < 0$, the vacuum acquires a non-zero value.

If $\mu^2 > 0$, the potential has a parabolic shape and the ground state, also called the vacuum, corresponds to $\phi = 0$. However, by choosing $\mu^2 < 0$, the potential acquires a “mexican hat” shape (see Figure 1.1).

Hence, the minimum of $V(\phi)$ is reached for $\phi = \pm v/\sqrt{2}$, with $v = \sqrt{\frac{-\mu^2}{\lambda}}$ being the vacuum expected value (vev).

$$|\phi|^2 = \phi^\dagger \phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda} \quad (1.26)$$

By choosing one solution, one obtains:

$$\phi_1 = \phi_2 = \phi_4 = 0, \quad \phi_3^2 = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2} \quad (1.27)$$

to represent the ground state of the vacuum. This choice breaks the $SU(2)_L \otimes U(1)_Y$ symmetry, leaving the vacuum invariant under the $U(1)$ symmetry.

Choosing one of the solutions for the non-zero ground state is equivalent to the other, thus breaking the original symmetry.

The fluctuation of the vacuum around its minimum can be expanded in terms of real field $\vec{\theta}$ and H , which are field excitation along the 3 $SU(2)_L$ directions and radial direction, such as

$$\phi_0(x) = e^{i\vec{\sigma} \cdot \vec{\theta}(x)/v} \begin{pmatrix} 0 \\ \frac{v+h(x)}{\sqrt{2}} \end{pmatrix}. \quad (1.28)$$

Since the choice of the phase is arbitrary, by placing ourselves in the unitary gauge, the field can be expressed as:

$$\phi_0(x) = \begin{pmatrix} 0 \\ \frac{v+h(x)}{\sqrt{2}} \end{pmatrix}. \quad (1.29)$$

Let us consider momentarily only this term in the lagrangian by omitting W_μ^a and B_μ contribu-

tions

$$\mathcal{L}'_{Higgs} = (\partial^\mu \phi_0)^\dagger \partial_\mu \phi_0 - \mu^2 \phi_0^\dagger \phi_0 + \lambda (\phi_0^\dagger \phi_0)^2 \quad (1.30)$$

$$= \frac{1}{2} \partial^\mu h \partial_\mu h - \mu^2 \frac{(v+h)^2}{2} - \frac{\lambda}{4} (v+h)^4 \quad (1.31)$$

$$= \frac{1}{2} \partial^\mu h \partial_\mu h - \left\{ \underbrace{h(\mu^2 v + \lambda v^3)}_{=0} + h^2 \underbrace{\left(\frac{\mu^2}{2} + \frac{3}{2} \lambda v^2 \right)}_{\lambda v^2 h^2} + \lambda v h^3 + \frac{\lambda}{4} h^4 + \text{const.} \right\}. \quad (1.32)$$

By defining $M_H^2 = 2\lambda v^2$, Equation 1.32 now reads

$$\mathcal{L}'_{Higgs} = \frac{1}{2} \partial^\mu h \partial_\mu h - \frac{1}{2} M_H^2 h^2 - \lambda v h^3 - \frac{\lambda}{4} h^4 \quad (1.33)$$

which describes the motion of a massive scalar real field h , the Higgs boson, with cubic and quartic self interaction terms. Knowing the Fermi constant G_F which is measured with high precision, v can be expressed as a function of this value and the mass of the W boson. However, λ is a free parameter, which leaves the Higgs boson mass unconstrained.

1.5.2 Mass generation for the Standard Model particles

Let us consider now only the kinetic term of Higgs field lagrangian:

$$|D_\mu \phi_0|^2 = (D^\mu \phi_0)^\dagger (D_\mu \phi_0) \quad (1.34)$$

$$= \left| \left(\partial_\mu - ig T_a W_\mu^a - ig' \frac{Y_H}{2} B_\mu \right) \phi_0 \right|^2 \quad (1.35)$$

$$= \frac{1}{2} \left| \partial_\mu - \frac{1}{2} \begin{pmatrix} gW_\mu^3 + g'B_\mu & g(W_\mu^1 - iW_\mu^2) \\ g(W_\mu^1 + iW_\mu^2) & -gW_\mu^3 + g'B_\mu \end{pmatrix} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \right|^2 \quad (1.36)$$

$$= \frac{1}{2} |\partial_\mu \phi_0|^2 \quad (1.37)$$

$$+ \frac{1}{8} v^2 g^2 [(W_\mu^1)^2 + (W_\mu^2)^2] + \frac{1}{8} v^2 [gW_\mu^3 - g'B_\mu]^2 + 0 [gW_\mu^3 + g'B_\mu]^2 \quad (1.38)$$

By combining W_μ^i and B_μ fields, one can obtain mass terms for the mass eigenstates such as:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2) \quad (1.39)$$

$$Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2 + g'^2}} \quad (1.40)$$

$$A_\mu = \frac{gW_\mu^3 + g'B_\mu}{\sqrt{g^2 + g'^2}} \quad (1.41)$$

Therefore, using these defined fields, Equation 1.38 reads

$$\left(\frac{vg}{2} \right)^2 W_\mu^+ W^{-\mu} + \frac{1}{2} \left(\frac{v}{2} \right)^2 (g^2 + g'^2) Z_\mu Z^\mu + \frac{1}{2} (0)^2 A_\mu A^\mu \quad (1.42)$$

where the mass of electroweak bosons are:

$$M_{W^\pm} = \frac{1}{2}vg \quad (1.43)$$

$$M_Z = \frac{1}{2}\sqrt{g^2 + g'^2} \quad (1.44)$$

$$M_\gamma = 0 \quad (1.45)$$

From Equations 1.40 and 1.41, one can note the relation between Z_μ and A_μ can be expressed by a simple rotation matrix

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \quad (1.46)$$

with θ_W known as the Weinberg mixing angle:

$$\cos \theta_W = \frac{g}{\sqrt{g^2 + g'^2}} \quad (1.47)$$

$$\sin \theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \quad (1.48)$$

In the fermion sector, masses are introduced by a Yukawa type interaction between the fermion and the Higgs fields. This interaction is characterized by a coupling constant f between the spinor and scalar field. An example is given here for the first generation of lepton such as, for each fermion, a new term is introduced:

$$\mathcal{L}_{FH} = f\bar{\psi}_L\psi_R\phi_0 \quad (1.49)$$

$$= f \begin{pmatrix} \bar{\nu} & \bar{e} \end{pmatrix}_L \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} e_R \quad (1.50)$$

$$= f\bar{e}_L e_R \frac{v}{\sqrt{2}} + f\bar{e}_L e_R \frac{h}{\sqrt{2}} \quad (1.51)$$

with a mass term appearing as $M_f = \frac{fv}{\sqrt{2}}$, invariant under the gauge rotation. One notices that neutrinos do not acquire mass by the Higgs mechanism.

1.6 Higgs searches

1.6.1 Theoretical constraints

Although the Standard Model does not predict the Higgs boson mass, arguments from theory provide boundaries on the allowed mass such as the theory remain stables.

As shown in Equation 1.32, the Higgs boson mass scales with λ . This free parameter dictates also the shape of the Higgs potential. The theory is valid for a stable vacuum (or metastable in the limit that its life-time is longer than the age of the Universe). Hence a constraint of the theory validity translates as a boundary on the Higgs boson mass.

Moreover, by taking into account the energy dependence of λ with higher order calculations, one can express the value of this parameter as function of an energy scale

$$\frac{\partial \lambda}{\partial t} = \frac{12}{16\pi^2} \lambda^2, \text{ with } t = \log \frac{Q^2}{Q_0^2}, \quad (1.52)$$

where Q_0^2 is an energy scale taken as reference. By requiring the coupling constant to remain in the perturbative regime of the regime, $\frac{\lambda}{4\pi} < 1$ must remain true. Given that $M_H = \sqrt{2\lambda}v$ and $v = \frac{1}{\sqrt{G_F\sqrt{2}}} = 246$ GeV, this constraint translates as $M_H \lesssim 1$ TeV. This is called the triviality argument.

Theoretical calculations of the amplitude of the $W_L^+ W_L^- \rightarrow W_L^+ W_L^-$ with W , Z and H bosons exchange in the s and t channels set limits on the Higgs boson mass [6]. By requiring the process to fulfill unitarity requirements and $s \gg M_H^2$, the constraint

$$M_H < 870 \text{ GeV} \quad (1.53)$$

is obtained. If the $W_L^+ W_L^- \rightarrow Z_L Z_L$ scattering is considered, the constraint can be lowered to

$$M_H < 710 \text{ GeV} \quad (1.54)$$

By combining these constraints, the allowed domain for a Higgs boson mass is shown in Figure 1.2. If the Standard Model is required to be valid up to an energy scale $\Lambda \sim 10^{16}$ GeV, the Higgs boson mass is constrained to $M_H \sim 160$ GeV.

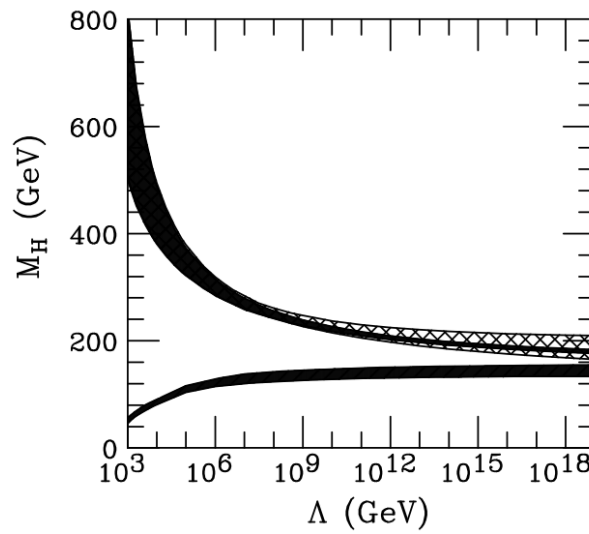


Figure 1.2: Allowed Higgs boson mass as function of Λ . The upper edge corresponds to the triviality argument constraints. The lower solid area represents the lower bounds derived from stability requirements, using $M_t = 175$ GeV and $\alpha_s = 0.118$. The hatched area shows the additional uncertainty when varying M_t from 150 to 200 GeV.

1.6.2 Experimental constraints

Indirect searches

Indirect searches aim at looking higher order effects on observables, involving Higgs boson loops. A dependence on the Higgs boson mass can be determined for such observables as Γ_Z or $A_{FB}^{0,l}$ which have been precisely measured at the LEP experiments [7], and is shown in Figure 1.3.

A global fit over all electroweak observables is performed, as a function of the Higgs boson mass. The result of this analysis is reported as $\Delta\chi^2 = \chi^2 - \chi_{min}^2$ with χ_{min}^2 yielding the preferred value of the Higgs boson mass from the fit, which is

$$M_H = 91_{-24}^{+31} \text{ GeV} \quad (1.55)$$

at 68% confidence level [2]. A light Higgs boson is favored. Figure 1.5 shows the allowed region for the Higgs boson mass as function of m_W and m_t , taking into account their dependence on M_H . Here again, the allowed region in blue is at the edge of the region where the Higgs boson is not excluded (in green), thus yielding to a light Higgs boson. One can notice that the sensitivity to the W mass measurement is dominant, thus a higher precision on this measurement will be able to indicate if the allowed region for the Higgs boson is still in agreement with the Standard Model expectation.

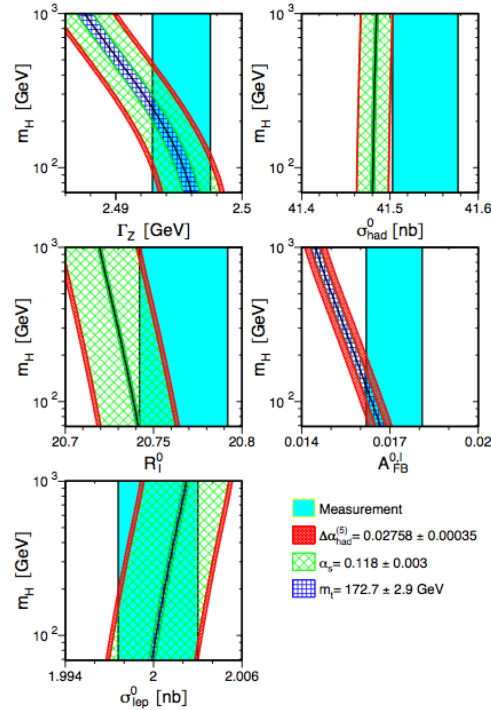


Figure 1.3: Measurement of electroweak parameters (blue band) and prediction from the Standard Model (shaded regions), shown as function of the Higgs boson mass. One can observe the weak (strong) dependence of σ_{had}^0 (Γ_Z and $A_{FB}^{0,l}$) on M_H .

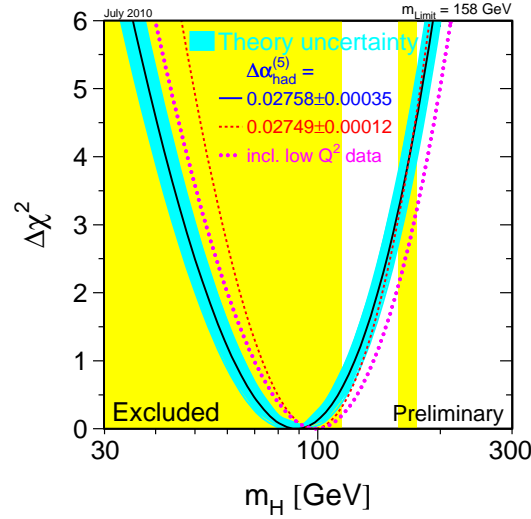


Figure 1.4: Result of the global fit of the electroweak parameters. In yellow is represented the region where the Higgs boson is excluded. The best value of M_H for the fit is obtained with $M_H = 91^{+31}_{-24}$ GeV at 68% confidence level.

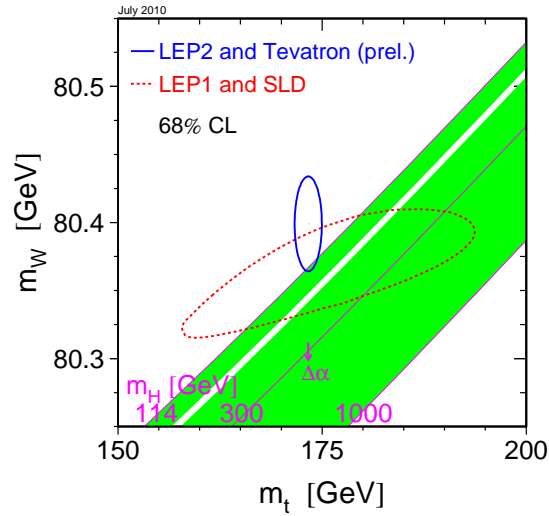


Figure 1.5: Indirect measurements of m_W and m_t (LEP-I+SLD data) (solid contour) and the direct measurements (Tevatron and LEP-II data) (dashed contour). In both cases the 68% confidence level contours are shown. The green band represent the allowed mass region by Standard Model as a function of M_t and M_W .

Direct searches

Since the mass of the Higgs boson is not predicted by the Standard Model, direct searches began for very low mass Higgs bosons [8, 9]. At higher energy, on the order of $M_H \simeq 100$ GeV, stringent limits have been set by LEP experiments. The center-of-mass energy in e^+e^- collisions can be finely controlled, thus allowing to scan the mass range for an excess consistent with a Higgs boson signal. When the experiment finished taking data, a lower limit on the Higgs boson mass has been set at

$$M_H > 114.4 \text{ GeV}. \quad (1.56)$$

Figure 1.6 shows the ratio between the limit on the coupling constant g_{HZZ} to its Standard Model expectation. The region corresponding to a ratio smaller than 1 is excluded at the 95% confidence level.

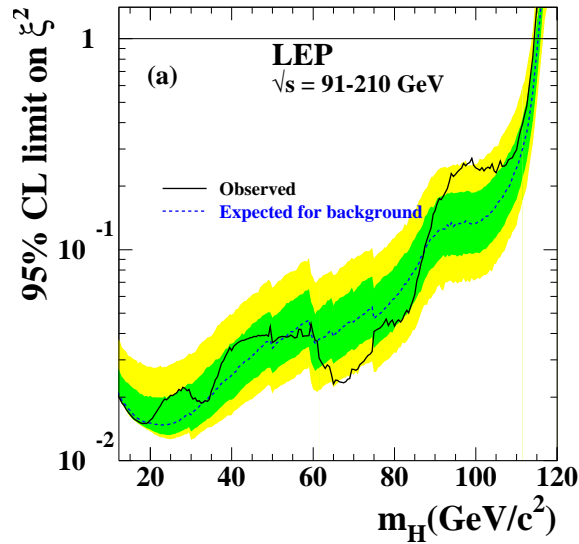


Figure 1.6: The 95% confidence level upper bound on the ratio between the limit on the HZZ coupling constant and its Standard Model prediction $\xi^2 = (g_{HZZ}/g_{HZZ}^{SM})^2$. The region below $\xi^2 = 1$, translates to the $M_H < 114.4$ GeV mass region, is excluded [10].

After the LEP shutdown in 2001, searches for the Higgs boson have been carried on by the Tevatron, later introduced in Section 2. Since the partonic center-of-mass energy is not known in $p\bar{p}$ collisions, the search is focused on two main regions, for a low or high mass Higgs boson. This distinction comes from the dominant decay mode which is $H \rightarrow b\bar{b}$ for $M_H \lesssim 135$ GeV, while for $M_H \gtrsim 135$ GeV, the $H \rightarrow W^+W^-$ decay mode is dominant, as illustrated on Figure 1.7b. A brief summary is given about the channels employed at the Tevatron for the Higgs boson search.

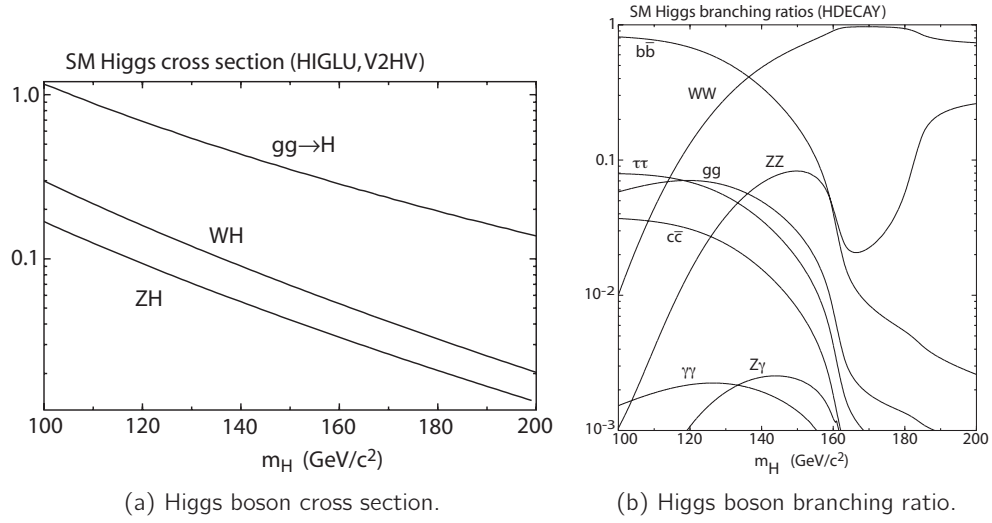


Figure 1.7: Higgs boson cross sections and branching ratios for different production and decays modes, as function of the Higgs boson mass. The cross section values (given in pb) are reported for $p\bar{p}$ collisions at 1.96 TeV. The search for a lower mass Higgs boson ($M_H \lesssim 135$ GeV) makes use of the $b\bar{b}$ decay mode, whereas the search for a higher mass Higgs boson ($M_H \gtrsim 135$ GeV) is sensitive to the W^+W^- decay mode.

Low mass Higgs search As shown on Figure 1.7b, the Higgs boson decays dominantly in a resonant $b\bar{b}$ pair. The highest cross section for the Higgs production is from gluon fusion (through a loop of top quarks or W^\pm bosons) as shown in Figure 1.7a. However, the production of a $b\bar{b}$ pair by QCD production is a background with a rate several orders of magnitude higher, therefore impossible to exploit. Hence, the lower cross section modes are exploited, such as WH and ZH , where a W or Z boson radiates a Higgs boson. The following channels are used for the search for a low mass Higgs boson:

- $ZH \rightarrow \nu\bar{\nu}b\bar{b}$, exploiting the invisible decay $Z \rightarrow \nu\bar{\nu}$ which has the highest branching ratio after $Z \rightarrow q\bar{q}$ [11],
- $WH \rightarrow \ell\nu b\bar{b}$, which will be described in this document,
- $ZH \rightarrow \ell^+\ell^-b\bar{b}$, taking advantage of the clean signature of a pair of leptons and jets. The kinematic of the objects can be constrained since no missing transverse energy from a neutrino is expected in the final state [12].

High mass Higgs search The search for a Higgs boson with $M_H \gtrsim 135$ GeV mainly relies on the $H \rightarrow W^+W^-$ decay mode, which gets maximal around $M_H = 165$ GeV. The gluon fusion production can be exploited with a ratio S/B of the order of 1/100. The channels having the highest sensitivity are $H \rightarrow \ell\nu\ell\nu$, in the $e\nu e\nu$, $\mu\nu\mu\nu$ and $e\nu\mu\nu$ [13]. The $\mu\nu\tau\nu$ channel has been recently introduced in the leptonic final state searches [14]. Another channel is used at the DØ collaboration, $H \rightarrow WW \rightarrow \ell\nu jj$ with one W boson decaying leptonically, the other hadronically [15].

Other channels are also exploited such as $H \rightarrow \gamma\gamma$ [16], $XH \rightarrow \tau\tau jj$ [17] or $VH \rightarrow VVV \rightarrow \ell^\pm \ell^\pm + X$, where $V = W$ or Z [18].

All the channels used for the Higgs boson search at the Tevatron are combined to set limits on the Higgs boson production in the mass range $100 < M_H < 200$ GeV. Analyses using up to 8.2 fb^{-1} lead to an exclusion of a Higgs boson with $158 < M_H < 173$ GeV. More details on limit derivation and potential improvements at the Tevatron are given in Section 5 and 6.

1.7 Conclusion

The Standard Model, described by a gauge theory, is successful in predicting the phenomenology of quarks, leptons and the bosons vehiculating the electromagnetic, weak and strong forces. A model is used to explain how particles acquire mass, while the theory remains invariant under the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ gauge group, and is called the Higgs mechanism. Considering experimental and theoretical constraints, the particle associated to the Higgs field is favored with a low mass, which makes a discovery or exclusion possible at the Tevatron. At the Tevatron, the $WH \rightarrow \ell\nu b\bar{b}$ channel is one of the most sensitive. This search performed at the DØ experiment will be presented in this document.

2

Experimental apparatus : The Tevatron and the DØ detector

This section is dedicated to a presentation of the Tevatron hosted at the Fermi National Accelerator Laboratory (Fermilab) and the DØ detector. Since the search for the WH signal involves, either an electron or a muon, jets originating from b quarks and missing transverse energy from the escaping neutrino, all parts of the detector are used to reconstruct these objects. An attention will be drawn on the production of proton and antiproton beams used in collisions at the Tevatron as well as the whole acceleration chain in Section 2.1. Then, the main components of the DØ detector and data acquisition will be introduced in Section 2.2.

2.1 The chain of accelerators and the Tevatron

Located near Chicago (USA), the Fermilab hosts many particle physics experiments and accelerators, in particular the Tevatron, a $p\bar{p}$ collider operating since 1988. Several discoveries have been made in this place such as the Υ meson, demonstrating the existence of the bottom quark from the E288 collaboration led by Leon Lederman in 1977 [19], the top quark by the CDF and DØ experiments in 1995 [20][21] and the tau neutrino in 2000 by the DONUT collaboration [22].

In order to provide high energy collisions, Fermilab benefits from a complex acceleration chain, the most powerful being the Tevatron accelerating beams of protons and antiprotons with an energy of 980 GeV to produce collisions at a center-of-mass energy of 1.96 TeV. The beams are crossing in two interaction points where are located the CDF and DØ detectors. An aerial view is shown in Figure 2.1 along with a diagram depicting the acceleration chain.

Collisions at the Tevatron occurred during two major periods:

- the “RunI” when the Tevatron was operating at center-of-mass energy of 1.8 TeV, from 1988

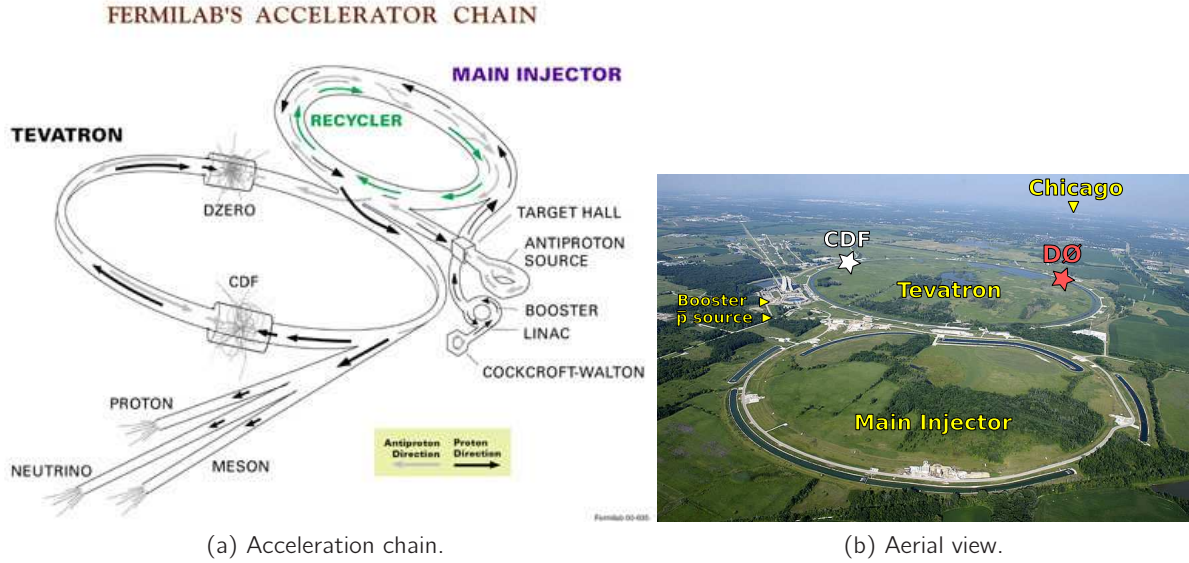


Figure 2.1: Protons are first created and accelerated at the *Cockcroft-Walton* pre-accelerator, LINAC and Booster. Part of the protons are used to create antiprotons. The antiprotons are stored in the Main Injector before being injected into the Tevatron.

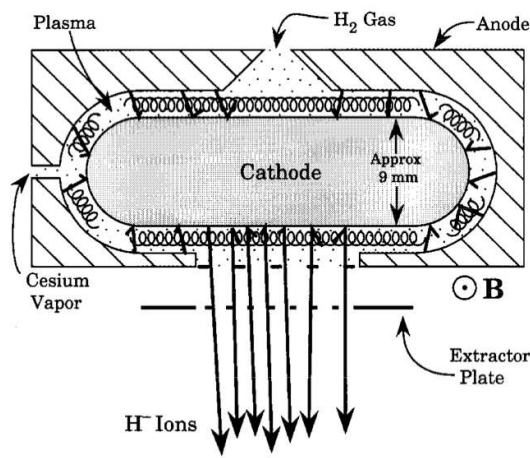
to 1996, collecting 125 pb^{-1} of data

- the “RunII”, starting in 2001 after the Tevatron had its center-of-mass energy upgraded to 1.96 TeV. This period is separated in two phases. The RunIIa period corresponds to an integrated luminosity of about 1 fb^{-1} recorded between March 2001 and March 2006. The RunIIb characterized, among others things, by the installation of an additional layer of detector to the Silicon Microstrip Tracker (SMT) at the closest to the beam pipe (more details in Section 2.2.2). The expected delivered luminosity is to be about 12 fb^{-1} by September 2011, when the Tevatron will shut down.

2.1.1 Proton beam production

Hydrogen gas (H_2) is first injected into an ionization chamber (also denoted as magnetron ion source). Through electric pulses, a plasma comprised of electrons and protons is created in the chamber. The released protons will be attracted towards the negative electrode of the chamber coated with Cesium. After being trapped by the electrode, they will be hit by the next protons. Since the binding energy of the electrons in Cesium is relatively low, protons will be released after capturing two electrons. Negatively charged H^- ions are created. A schematic view of a magnetron is given in Figure 2.2a. H^- ions are then focused towards the first step of the acceleration chain, which is the *Cockcroft-Walton*.

The electric power within the Cockcroft-Walton (see Figure 2.2b) is of 750 kV, therefore accelerating hydrogen ions to an energy of 750 keV due to its static electric field.



(a) Basic configuration of a magnetron.



(b) Cockcroft-Walton generator.

Figure 2.2: (a) Schematic view of a magnetron, producing H^- ions. (b) Photography of the Cockcroft-Walton.

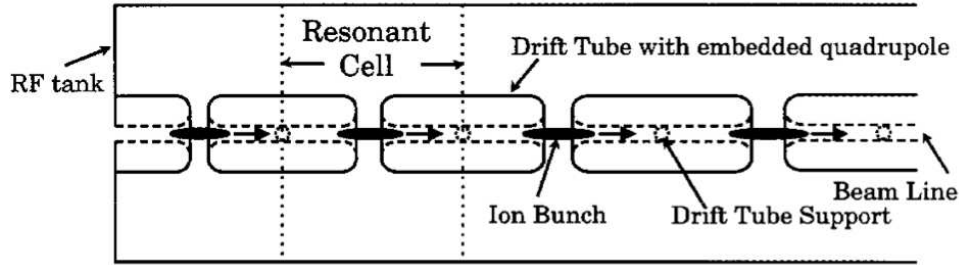


Figure 2.3: Schematic view of RF cavities in the LINAC.

The next step is a further acceleration of ions through the Linear Accelerator (LINAC) which is 130 meter long. A pulsed beam of 400 MeV is produced by the mean of radio frequency (RF) cavities. By alternating the electric field at a given frequency, the motion of particles is constrained and bunches are formed according to the frequency of the RF cavities. Quadrupole magnets present in drift tubes allows to focus the beam in traverse plane to the direction of the ions.

Before entering the Booster, ions pass through a thin foil of carbon. This will strip off the loosely bound electrons and let protons be accelerated to an energy of 8 GeV. The Booster is a 475 meters long synchrotron, formed by 96 magnets bending the trajectory of the protons and RF cavities operating at 1 GHz to accelerate them. Whereas the time of travel of ions in the LINAC is $20\mu s$, a complete revolution in the Booster takes about $2.2\mu s$. The Booster is typically filled with about 3×10^{12} protons. This step is the last one before injection to the Main Injector.



Figure 2.4: Photography of the Main Injector tunnel, where are located the Recycler (in green) and the Main Injector.

2.1.2 The Main Injector and Recycler

After having undergone an acceleration to reach an energy of 8 GeV in the Booster, protons are injected to the Main Injector [23], a synchrotron with a 3.3 km circumference, built between the RunI and RunII. The purpose of the Main Injector has different aspects. It accelerates protons to 120 GeV before sending them to a Nickel target to produce antiprotons through the reaction $p + p \rightarrow p + p + p + \bar{p}$. It is interesting to keep in mind that the antiproton production efficiency is at the order of 3×10^{-6} produced antiprotons per proton on target.

Protons and antiprotons travel in the same beam pipe, surrounded by 344 dipole magnets and 208 quadrupole magnets to focus beams. Before injecting particles in the Tevatron, particles are accelerated up to 150 GeV.

A limiting factor for the Tevatron integrated luminosity is the amount of antiprotons in the accelerator. Since their production efficiency is low, the Recycler [24] can recover antiprotons that are still present after the end of the previous *store* (period during which protons and antiprotons collide in the Tevatron). The recycler is located in the same tunnel as the Main Injector, as shown in Figure 2.4, and was installed during the Tevatron upgrade before the beginning of the RunII. Since its purpose isn't to accelerate particles, it is only comprised of permanent magnets.

2.1.3 Antiproton beam production

The main motivation for building a $p\bar{p}$ collider was to initiate reactions from opposite charge particles in order to search for a top quark pair production. This allows as well to make use of a single magnet system where both proton and antiproton beams circulate. However, the creation of an antiproton beam requires three devices: a fixed Nickel target from which antiprotons are created and two synchrotrons to accumulate and *cool* them.

As previously emphasized, antiprotons are produced from the collisions between 120 GeV protons and the Nickel target, with a rate close to one antiprotons created for one million protons on target.

The outgoing antiprotons are created with an average energy of 8 GeV and with a large angular spread. Both Debuncher and Accumulator which will be described here are operating at 8 GeV. First, produced particles are focused with a magnetic lens and then are passing through a mass spectrometer in order to separate antiprotons from other type of particles that can be created at the same time. This process is depicted on Figure 2.5.

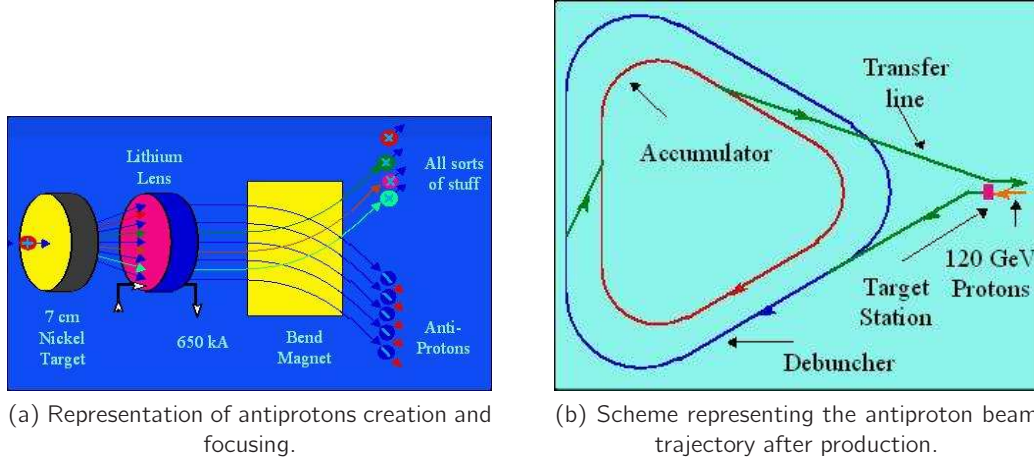


Figure 2.5: After protons hitting the Nickel target, produced antiprotons are focused and filtered from other particles (a). Antiprotons have then their energy spread reduced in the Debuncher.

As the protons hit the target by bunches, outgoing antiprotons do have the same structure. The Debuncher is a triangular 505 meters circumference synchrotron. Lower energy particles will travel closer to the inner part of the cavity whereas higher energy particles will tend to have an outward trajectory. By traveling at a different radial distance from the beam center, beams will travel in a different RF field intensity, thus bringing them to the desired trajectory and speed, in a 100 ms process. This setup will destroy the bunch structure of the beam, hence the name of this part of the acceleration chain.

The beam being stable, the process of *stochastic cooling* [25] can be initiated. This allows to reduce the transverse oscillation of the beam and consequently the energy spread. This process is achieved by detecting fluctuations in the momentum and slightly correcting the trajectory each time a particle travels at the vicinity of an electrode called *pickup*. By measuring its position relative to the nominal beam trajectory, this measurement is converted in an electric signal through the electric field created in the pickup. This signal is then sent to the *kicker* which will modify its electric field to bend the particle according to the received signal. Figure 2.6 shows a brief description of the Debuncher and how the stochastic cooling is performed.

After 2.4 seconds, antiprotons are then sent to the Accumulator, located in the same tunnel. Here, the bunch structure of the antiproton beam will be formed again and antiprotons are cooled further by interacting with a low emittance electron beam until thermal equilibrium. The process of antiproton accumulation is called *stacking* and lasts about 8 hours. Potentially the life time of the \bar{p} beam can be extended to several days without major losses as a result of the beam stability. Once the Tevatron has dumped the beam from the previous store, antiprotons stored in the Accumulator

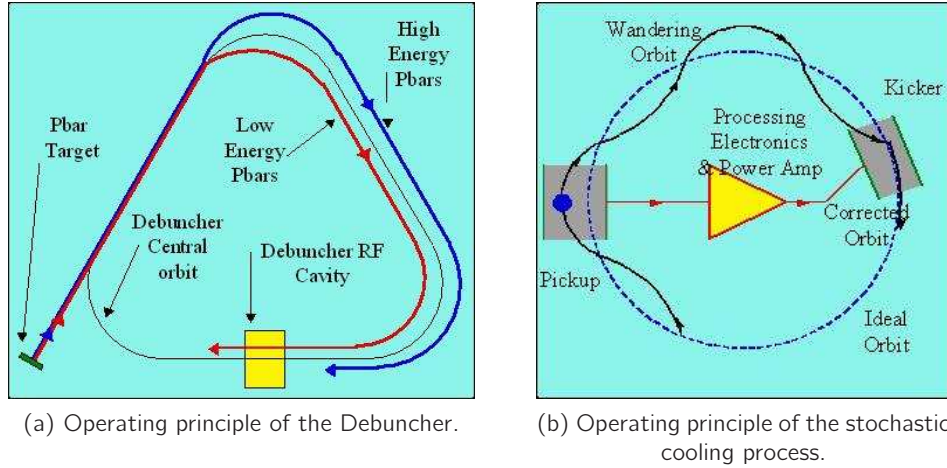


Figure 2.6: Description of the two devices allowing the reduction of the beam spread, the Debuncher (a) and the stochastic cooling process (b).

will be injected to the Tevatron.

2.1.4 The Tevatron

The last accelerator in the chain is the Tevatron, where physics collisions take place in two points, CDF and DØ, where the p and \bar{p} beams cross. The circular synchrotron has a 1 km radius. It is comprised of 8 accelerating cavities, 816 dipole superconducting magnets and 204 quadrupole magnets. Dipole magnets are made of Niobium-Titanium alloy wire. The superconducting behaviour is reached when cooled to liquid Helium temperatures (4.3 K). Beams from the Main Injector are accelerated from 150 GeV to 980 GeV by the RF cavities operating at a frequency of 53 MHz. Dipole magnets are generating then a 4.2 T magnetic field, a full revolution of particles in the Tevatron is achieved in 21 μ s.

Optimization studies from the Accelerator Division yielded the most efficient way to produce collisions at the highest rate is to fill the Tevatron with 36 bunches of approximately 3×10^{11} protons and about 10^{10} antiprotons at the beginning of a store. The bunch structure of the proton and antiproton beams, as shown in Figure 2.7, is formed by 3 "super bunches" separated by 2.64 μ s, each of them containing 12 bunches separated by 396 ns. After p and \bar{p} beams, traveling in the same beam pipe in an helical motion (see Figure 2.8), reached 980 GeV, collisions can start after beam focusing and halo removal. 1.96 TeV $p\bar{p}$ interactions take place in the interaction region, located around the center of the CDF and DØ detectors. The distribution of the luminous region along the beam axis corresponds to a gaussian distribution with a spread $\sigma_z = 18$ cm.

Major changes have been made for continually improving the Tevatron performance, see Table 2.1. The instantaneous luminosity has been increasing also during the RunII as shown in Figure 2.9.

2.1. THE CHAIN OF ACCELERATORS AND THE TEVATRON

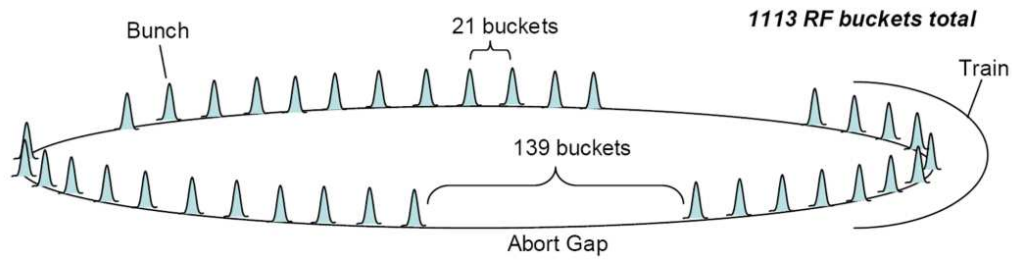


Figure 2.7: Bunch structure of the proton and antiproton beams.

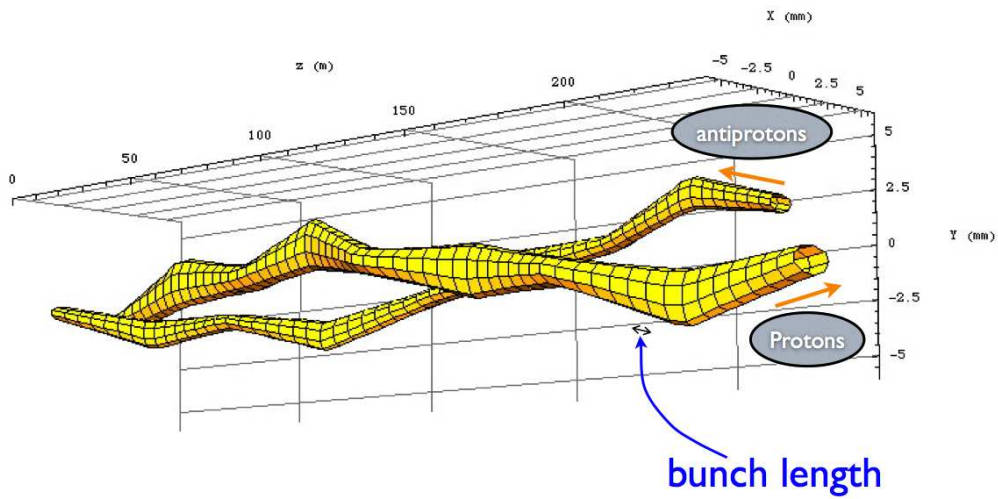


Figure 2.8: Beam envelopes along their motion axis.

	RunI	RunII	
		RunIIa	RunIIb
Period	1989-1996	2001-2006	2006-2011
Beam energy (GeV)	900	980	
Bunch spacing	$3.5 \mu\text{s}$	396 ns	
Number of bunches	6×6	36×36	
Protons per bunch	2.3×10^{11}	2.7×10^{11}	3×10^{11}
Antiprotons per bunch	5.5×10^{10}	3×10^{10}	7×10^{10}
Peak luminosity ($\text{cm}^{-2}\text{s}^{-1}$)	16×10^{30}	100×10^{30}	$200\text{--}400 \times 10^{30}$
Integrated Luminosity	160 pb^{-1}	1 fb^{-1}	$\sim 11 \text{ fb}^{-1}$

Table 2.1: Summary table of Tevatron characteristics for the RunI and RunII periods.

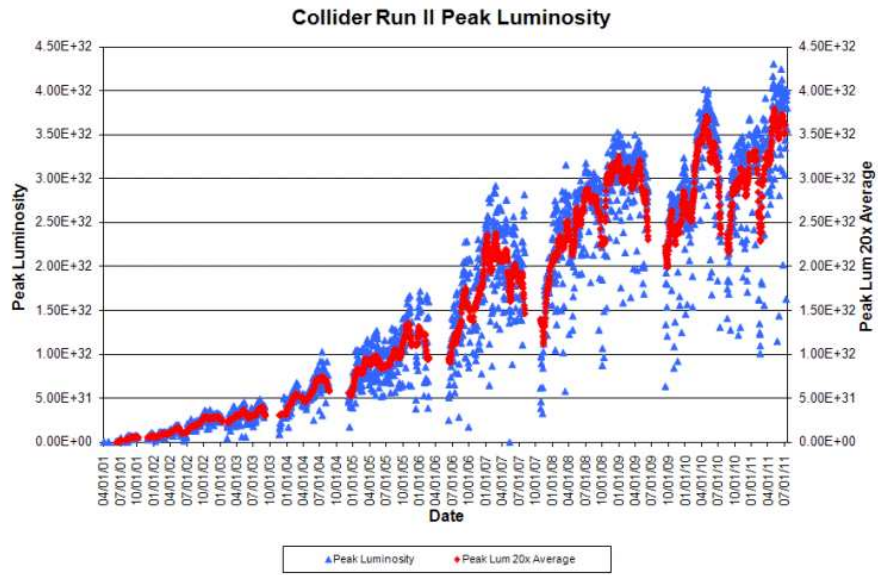


Figure 2.9: Peak instantaneous luminosity, reached in beginning of store, shown as function of time.

2.2 The DØ detector

After the production of $p\bar{p}$ collisions at 1.96 TeV being introduced, the detector used to record and study their outcome is described in this section. The DØ detector [26] [27] is a multi-purpose detector with a cylindric geometry. It is comprised of different parts, listed here from the innermost to the outermost of the beam pipe, that will be described in more details in this section:

- the Silicon Microstrip Tracker (SMT)
- the Central Fiber Tracker (CFT)
- the superconducting solenoid magnet
- preshower detectors
- the electromagnetic and hadronic calorimeter
- the muon spectrometer

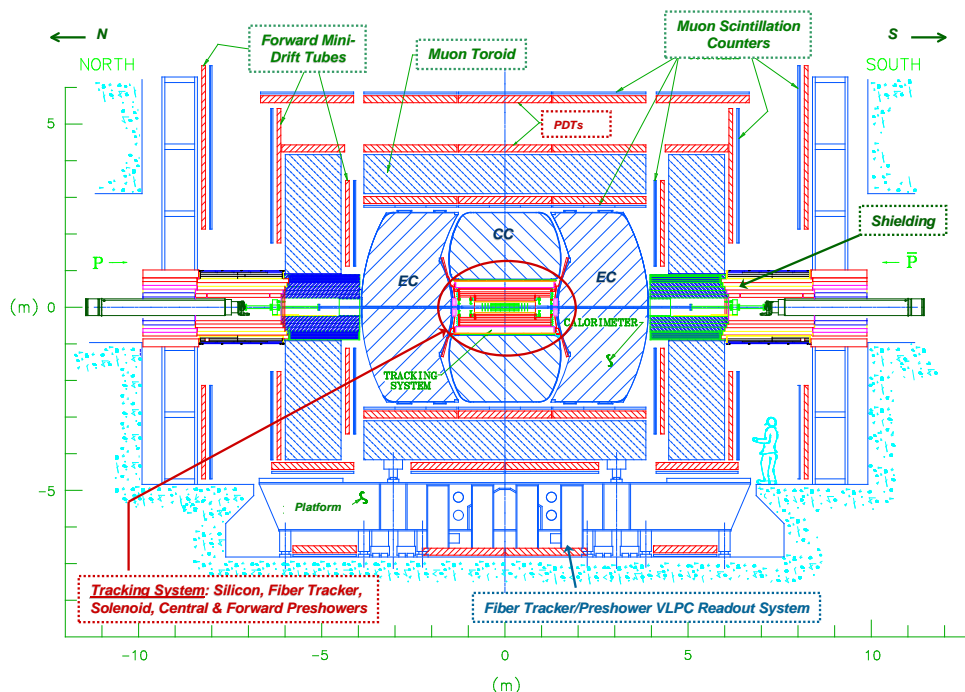


Figure 2.10: Overview of the DØ detector.

During the transition between the RunI and RunII, the detector has been upgraded. One of the major change is the replacement of the whole central tracking system, in order to introduce a solenoid magnet to measure the momentum of charged tracks. More recently, between the RunIIa and RunIIb periods, an additional layer of tracking detector as been placed at the innermost part of the detector in order to improve track resolution and identification of jets originating from b quarks.

Layer 0

During the 2006 shutdown, an additional layer of the Silicon Microstrip Tracker (detailed in the next Section) has been introduced in the DØ detector [28]. The motivation of adding a new layer at the closest point to the beam pipe are the following:

- recover tracking performances coming from radiation damages,
- better track finding efficiency which is deteriorated at higher instantaneous luminosity,
- impact parameter resolution improvement (see Figure 3.19), impacting b -tagging performances.

The annular space between the beam pipe and the first layer of the SMT has a radius of 22.90 mm, where the detector, cables and support structure have to fit in this gap.

This component is built with silicon microstrips sensors, similarly to the SMT. After the addition of the Layer 0, the impact parameter resolution is improved for tracks $p_T < 5$ GeV as shown in Figure 2.12, therefore yielding an overall improvement of 15% in b -tagging performance with respect to the beginning of RunII.

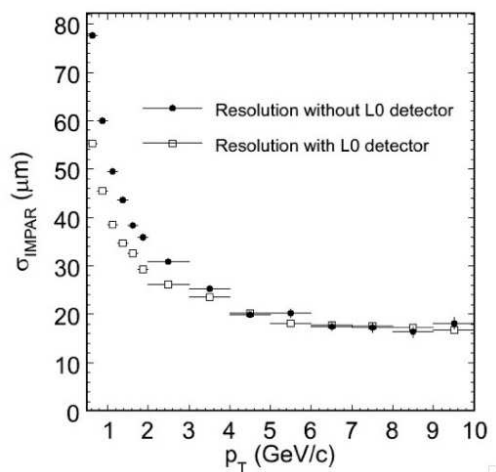


Figure 2.12: The impact parameter resolution with and without the Layer 0 detector.

Silicon Microstrip Tracker

Located at the innermost part of the detector, the Silicon Microstrip Tracker (SMT) is the main ingredient to reconstruct tracks from charged particles created from the $p\bar{p}$ collisions, bended by a magnetic field. Such objects are crucial in order to reconstruct primary vertices, measure lepton transverse momentum, identify b quark generated jets and others, therefore impacting the whole range of the physics program at DØ : search for New Physics, b physics and of course Higgs boson searches. The motivation to use in the DØ collaboration a silicon-based detector was to have a

high resolution and fast response device.

Since the length of the interaction region is relatively big, the challenge is to cover the volume where tracks can be reconstructed. The extremities of the SMT cover up ± 64 cm along the z axis. The central part covers the $0 < |z| < 60$ cm region, corresponding to $|\eta_{det}| < 1.5$, whereas the forward part covers the $60 < |z| < 100$ cm or $\eta_{det} < 3$ region. An isometric view of the SMT is given in Figure 2.13.

Since tracks are usually perpendicular to the beam, the SMT has been designed with modules parallel to the beam pipe to measure the $r - \phi$ coordinates. Disk detectors measure $r - z$ as well as $r - \phi$ components.

The central part of the SMT is comprised of two subparts:

- 6 barrels, each of them comprised of 72 modules ordered in 4 layers. They are arranged along the z axis. The distance with respect to the beam pipe is 2.7, 4.5, 6.6 and 9.4 cm respectively for each layer. Silicon strips have three different orientations: longitudinal, 90° stereo and 2° stereo with respect to beam axis.
- 12 “F-disks”, where 6 are located between the barrels except for the two most central ($|z| = 12.5, 25.3$ and 38.2 cm) and the remaining are placed at higher $|\eta_{det}|$ region ($|z| = 43.1, 48.1$ and 53.1 cm). Each disks is built from 12 wedges covering 30° in the transverse plane.

The forward part of the SMT is comprised of two “H-disks” arranged on each side of the detector at $|z| = 100.4$ and 121.0 cm. Their diameter is larger than the F-disks in order to cover the $|\eta_{det}| > 2$ region. Each section of the disks covers 15° in the transverse plane.

Disks are comprised of semi-conductive doped silicon sensors. By passing through the sensors, charged particles create electron-hole pairs. The signal is measured by collecting drifting charges by applying a voltage difference in the sensors. Each barrel has four silicon readout layers. The silicon modules installed in the barrels are called “ladders”. Layers 1 and 2 have twelve ladders each; layers 3 and 4 have twenty-four ladders each, for a total of 432 ladders. The SMT is cooled down to -5°C in order to reduce silicon radiation damages.

The SMT achieves a spatial resolution of around $20\ \mu\text{m}$ for hits.

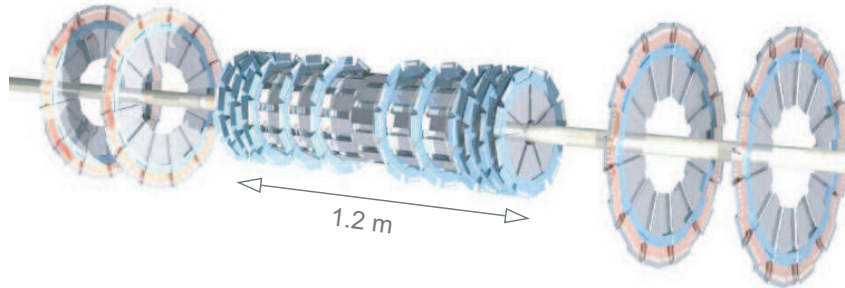


Figure 2.13: Overview of the SMT.

Central Fiber Tracker

The Central Fiber Tracker (CFT), located between 20 to 52 cm from the center of the beam pipe, is used to determine the momentum of charged particles and reconstruct their tracks. Located around the SMT, the coverage is $|\eta_{det}| \lesssim 1.7$. Its central part is 1.66 m long due to limitations from the presence of the SMT H-disks which have a larger radius than components in the central part of the SMT, while the outer part is 2.52 m long. Concentric layers of scintillating fibers are arranged in 8 cylinders of 4 fibers, 2 of them aligned along the z direction, whereas the other layers of fibers have a stereo angle of $+3^\circ$ or -3° , respectively denoted as u and v layers.

The fiber diameter is $835 \mu\text{m}$ and made of polystyrene doped with paraterphenyl to provide fluorescence. Scintillator fibers are connected to waveguide fibers (having $\sim 90\%$ reflexion power, therefore reducing losses of signal) to transfer the light produced when particles passing through. The signal, with a wavelength distributed around 530 nm, is collected by avalanche photodetectors (VLPC, standing for Visible Light Photon counter Cassettes) operating nominally at 9K. In total, 1000 km of fibers are used : 200 km of scintillating fibers and 800 km of waveguides.

Let us consider a simple example where a particle go through a unit squared detector cell, representing a single layer. The spatial hit resolution is given by the size of the unit cell divided by $\sqrt{12}$. Given that this measure is repeated 8 times (for the 8 layers of the CFT), decreasing as $1/\sqrt{8}$, thus $835/\sqrt{12 \times 8} \sim 100 \mu\text{m}$. This estimate corresponds to the track spatial resolution in the $r - \phi$ plane.

Solenoid

In order to measure the momentum of tracks, a superconducting solenoid magnet, cooled down by liquid helium, has been added after RunI. Its overall size was determined by the available space between the CFT and the calorimeter: 2.73 m in length and the outer diameter is 1.42 m, for approximately 1 radiation length at $\eta_{det} = 0$. Since the tracking volume is relatively compact, a 2 T magnet is used in order to have good track momentum resolution despite the limited tracking radius. The generated field is uniform at the 0.5% level. In order to remove potential systematic effects, the polarity of the DØ solenoid magnet is regularly reversed (every two weeks).

2.2.3 Preshower detectors

Before describing the DØ calorimeter, the Preshower detectors are introduced in this section. A thin layer of scintillators is placed in front of the calorimeter, helping in electron identification and hadrons/electromagnetic particles discrimination during both triggering and offline reconstruction. They operate as calorimeters as well as tracking detectors. There are 3 preshower detectors. The Central Preshower detector (CPS) is located in the central part ($|\eta_{det}| < 1.3$), between the solenoid magnet and the central calorimeter. Two Forward Preshower detectors (FPS) are located in the forward region ($1.5 < |\eta_{det}| < 2.5$) in front of the end calorimeters, around the luminosity monitors (see Section 2.2.6).

Both preshower detectors are made from triangular strips of scintillators, as shown in Figure 2.15, in order to minimize dead space. Scintillator are made of polystyrene plastic doped with

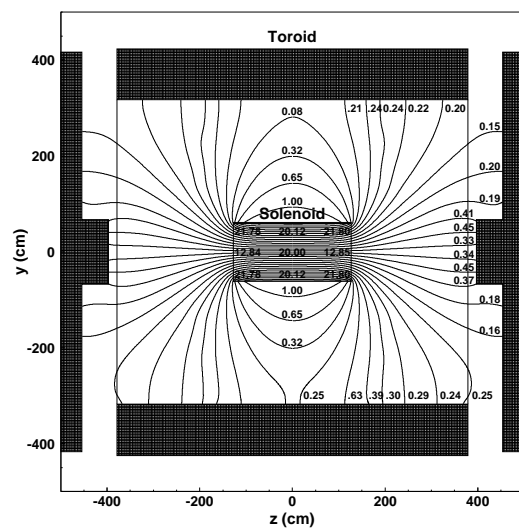


Figure 2.14: Side view of the magnetic field measured in kG. Lines are representing the field lines in y - z plane.

1% p-terphenyl and 150 ppm diphenyl stilbene. At the center of each triangular strip, wavelength-shifting fiber are connected to fibers in order to collect and carry the light to the end of the detector. The light is then sent to VLPC cassettes for readout.

In the CPS, 3 layers of scintillator strips are arranged in axial- u - v geometry, with u and v layers having respectively a -23.774° and a $+24.017^\circ$ stereo angle with respect to the beam axis. Each layer contains 1280 fibers for readout.

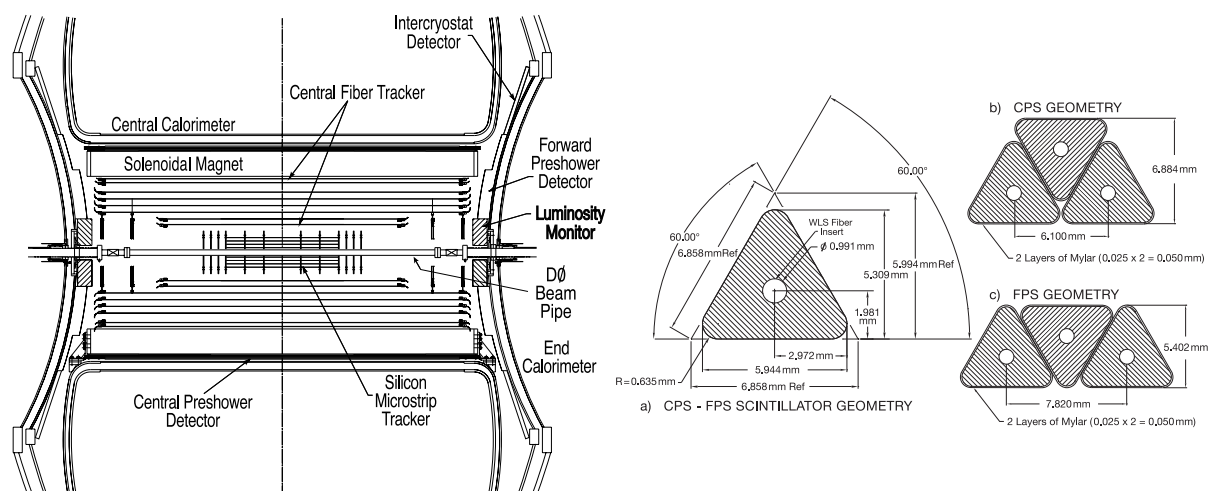


Figure 2.15: Left figure: The CPS detector is located in the gap between the solenoid magnet and the calorimeter. FPS detectors are placed between the CFT and the End Calorimeters. Right figure: Preshower detectors geometry.

2.2.4 Calorimeter

One important observable in the WH analysis, described in Section 4, is the reconstructed mass of a pair of jets to evaluate the mass of a candidate Higgs boson. Therefore, efficient jet reconstruction, identification and a precise energy measurement are crucial aspects in the analysis presented here. These aspects, later described, rely on the performance of the calorimeter which is now introduced.

Using this device, the energy of photons, electrons and jets can be directly measured, and the missing transverse energy can be inferred, indirectly giving an indication of the presence of a neutrino. The calorimeter measures the energy while a particle goes through the material and interacts through atomic or nuclear interactions and creates a “shower”. The DØ calorimeter is essentially composed of depleted Uranium and liquid Argon, respectively acting as absorbing and active materials, which is the operating mode of a sampling calorimeter. The energy of incoming particles is measured in the active medium of the detector. The sampling fraction of the calorimeter, defined as the ratio of energy deposited in the active medium to the total energy deposit in the calorimeter and determined by simulation, is approximately 5%, depending on the Uranium/Argon ratio in the different parts of the detector.

Since particles can interact by different types of interactions, the showers of particles created in the calorimeter can be differentiated between hadronic and electromagnetic showers:

- photons and electrons initiate electromagnetic showers. In this range of energy, electrons undergo bremsstrahlung: after traveling one interaction length X_0 in a material, a photon is emitted, leaving the electron with a lower energy. This process is favoured for light particles as energy loss scales as $1/m^2$. The radiation length, normalized to the material density, is 19.6 g/cm^2 for liquid Argon [29] and 5 g/cm^2 for Uranium [30]. This process appears in the vicinity of the Coulomb field of a nucleus, the electrons interacting with a virtual photon emitted by the nucleus. A photon produces a e^+e^- pair after traveling the same characteristic length X_0 . Similarly to the bremsstrahlung process, pair production occurs under the influence of the atomic nucleus Coulomb field. This process is repeated until the energy of particle is not sufficient to initiate another photon radiation or e^+e^- pair production;
- hadrons mainly undergo strong interactions in the detector. By hitting nuclei, a spray of pions, protons and fragments of the nuclei will be created. Similarly to X_0 , the nuclear interaction length is defined as the mean free path before an inelastic collision with a nucleus. Neutral pions will have a similar signature as an electromagnetic particle since it mainly decays to two photons. Other charged hadrons will interact successively until forming a shower.

In order to accurately measure such objects, the DØ calorimeter is segmented in two main parts, the hadronic and electromagnetic calorimeter. The hadronic part of the calorimeter is further divided in two parts: the fine and coarse hadronic calorimeter, where the granularity is lower for the latter. The calorimeter is also segmented longitudinally in two main parts: the central calorimeter (CC) and two end calorimeters (EC) located in the forward regions¹.

A unit cell of the DØ calorimeter is shown in figure 2.17.

The liquid Argon active region, located in the 2.3 mm gap between two plates, is ionized by secondary particles produced in the showers and induced charges drift (typically in 450 ns) and

¹EC calorimeters are commonly referred as North and South calorimeters.

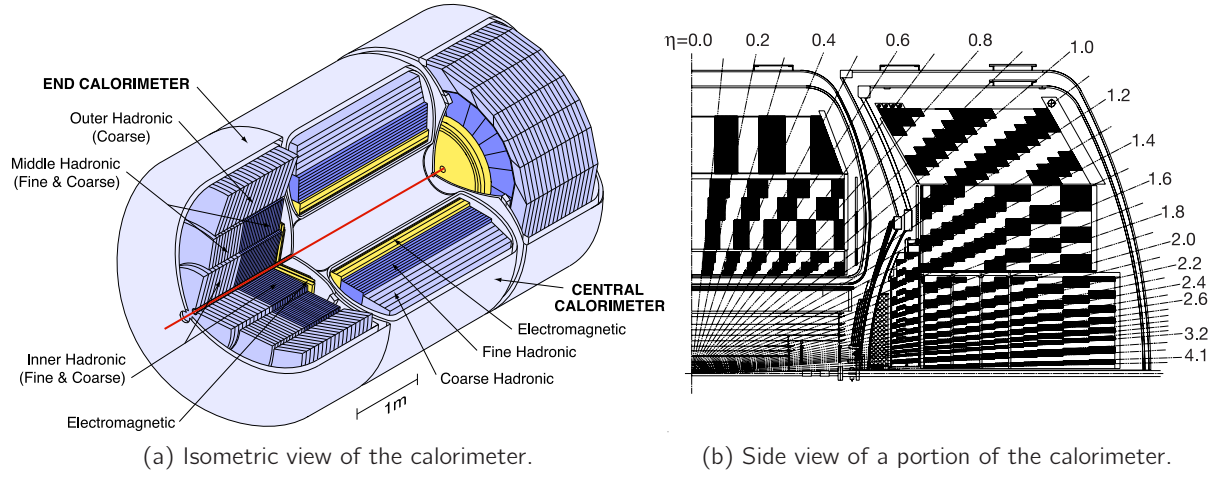


Figure 2.16: Overview of the DØ calorimeter.

are collected by electrodes connected to 2.0 kV voltage. Collected charges are then converted to electric signal. Cells which have the same η_{det} and ϕ coordinates with their center aligned with the center of the detector form *calorimetric towers* as shown in Figure 2.16b (the projection of a “point” in the (η_{det}, ϕ) plane form a tower, shown alternatively as a shaded region in the $r - z$ plane).

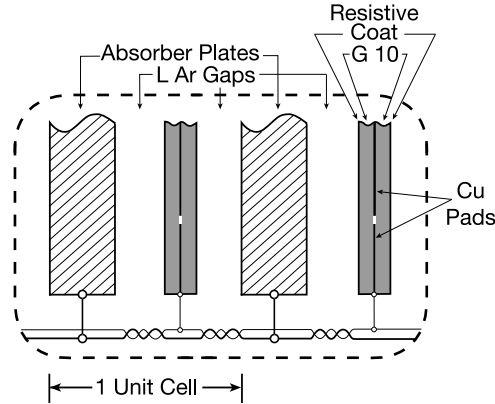


Figure 2.17: Schematic view of a unit cell of the DØ calorimeter.

Each calorimeter part, CC and both EC, is located in a separate cryostat to keep the Argon in a liquid state, down to a temperature of 90 K. A small gap exists between the CC and EC, therefore the *intercryostat detector* is placed in this region to ensure detection in this η_{det} range. These three components will now be described.

Central calorimeter

This part of the calorimeter covers the $|\eta_{det}| \lesssim 1$ region. In the electromagnetic region, absorbers are 3 mm thick depleted Uranium plates. The fine hadronic calorimeter uses 6 mm thick plates made from Uranium-Niobium alloy, whereas the coarse hadronic region, 46.5 mm thick Copper absorber plates are used. The choice of Uranium, which has a short radiation length ($X_0 \simeq 0.32$ cm [30]), allows the calorimeter to be relatively compact.

Layers of each subdetector parts have different thickness. In the electromagnetic calorimeter, layers are 1.4, 2.0, 6.8 and 9.8 X_0 thick, starting from the closest to the center of the detector. For the fine hadronic detector, layers are respectively 1.3, 1.0 and 0.9 λ_f thick, with λ defined as the absorption length which a hadron can travel before suffering another inelastic scattering event. Finally, the coarse hadronic calorimeter, layers are 3.2 λ_f thick.

The segmentation in the $\Delta\eta \times \Delta\phi$ plane (or granularity of calorimeter cells) is 0.1×0.1^1 , except for the third electromagnetic layer where it is twice finer since in RunI, electromagnetic showers were maximally developing in this part.

End calorimeters

Two identical end calorimeters denoted “North” and “South” are located on both sides of the CC. These detectors cover the acceptance $0.7 \lesssim |\eta_{det}| \lesssim 4$. Their structures is similar to the central part of the calorimeter, except a few differences. Unlike the CC, stainless steel absorber plates are used. Layers in the electromagnetic calorimeter are 1.6, 2.6, 7.9 and 9.3 X_0 thick.

The hadronic calorimeter is subdivided in three parts (opposed to the CC were there is only the fine and coarse hadronic calorimeter):

- the inner hadronic calorimeter which consists of four 1.1 λ_f thick layers,
- the middle hadronic calorimeter with four internal layers having a thickness of 0.9 λ_f and an external layers with 4.4 λ_f ,
- the single-layered outer hadronic calorimeter which is 4.1 λ_f thick,

Intercryostat detector

As previously written, both CC and EC parts are cooled down to 90 K by surrounding cryostats. In the narrow $0.8 < |\eta_{det}| < 1.4$ region, is placed a different type of detectors,

- cells which doesn't contain absorber plates (*massless gaps*) covering the $0.8 < |\eta_{det}| < 1.2$ and $1 < |\eta_{det}| < 1.3$ are located inside the CC and EC cryostats respectively. Cryostats are here acting as absorbers.
- the *Intercryostat detector* (ICD) consists of scintillating tiles that are attached to the exterior surfaces of the end cryostats (see Figure 2.15). Each $\Delta\eta \times \Delta\phi = 0.3 \times 0.4$ tile is subdivided. Hence, the granularity in this region is 0.1×0.1 .

¹The azimuthal segmentation made in 64 parts, thus $\Delta\phi = 2\pi/64 \simeq 0.1$.

Energy resolution of the calorimeter

The energy resolution σ_E in sampling calorimeter can be expressed in the following way:

$$\left(\frac{\sigma_E}{E}\right)^2 = C^2 + \frac{S^2}{E} + \frac{N^2}{E^2} \quad (2.1)$$

The detector can be characterized by the following parameters:

- the constant term C , taking into account non-uniformity of material thickness, non-uniformity in charge collection, mechanical imperfections, fluctuations in the amount of the upstream energy deposit and shower leakage. This term is independent of the energy and dominant at the high energy regime,
- the sampling term S , determined by the choice of the absorber and active material, the thickness of sampling layer, represents the statistical fluctuations in the amount of measured energy,
- the noise term N is associated with electronic noise in the detector and is dominant at low energy.

A detailed explanation of each term is given in [31] and [32].

The resolution for the DØ calorimeter are given in Table 2.2. They are determined separately for electrons and jets and in different calorimeter regions [32, 33]. Figure 2.18 shows the jet energy resolution for central jets, along with simple examples when one of the jet energy parameters are multiplied by a factor 2, in order to emphasize the energy resolution dependence on the parameters. Since the sensitivity of the WH analysis mostly relies on the resolution of the invariant mass of two jets, one can understand that the energy calorimeter is a limiting factor for the search of dijet resonance.

	e (CC)	jet ($ \eta < 0.4$)	e (EC)	jet ($2.4 < \eta < 2.8$)
C	0.030 ± 0.001	$\simeq 0.049$	0.028 ± 0.001	$\simeq 0.068$
S ($\text{GeV}^{1/2}$)	0.24 ± 0.01	$\simeq 0.757$	0.27 ± 0.01	$\simeq 0.523$
N (GeV^2)	0.47 ± 0.09	$\simeq 0.860$	0.73 ± 0.07	$\simeq 0.860$

Table 2.2: Energy resolution parameters for electrons and jets in RunII.

2.2.5 Muon system

At the energy range at which they are produced, muons are minimum ionizing particles. They cannot have their energy measured or be identified with the calorimeter only where they deposit 2 to 3 GeV on average. These particles have a dedicated part of the detector, the muon spectrometer, to detect them. It is located at the furthest of the interaction point. Its coverage is up to $|\eta| \simeq 2$. As represented in Figure 2.19, it consists of several elements:

- drift chambers,

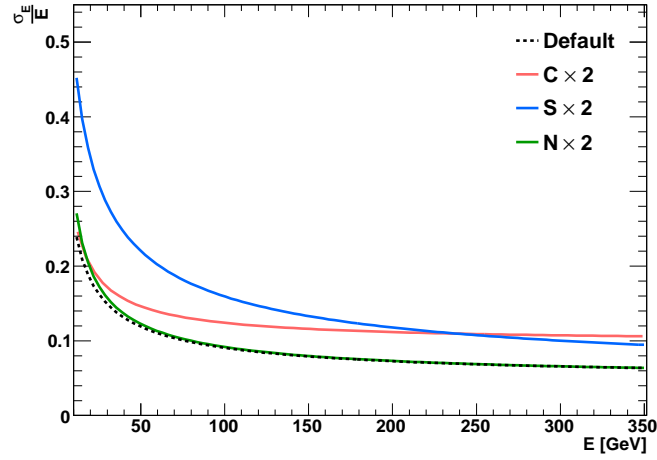


Figure 2.18: Energy resolution for jets with $|\eta| < 0.4$ as function of the jet energy. The dotted black line is obtained from the parameters quoted in Table 2.2. Colored lines are obtained by multiplying C (in red), S (in blue) and N (in green) by a factor 2 independently of the other parameters.

- scintillators chambers,
- toroidal magnets generating a 1.8 T magnetic field. Similarly to the solenoid magnet, the polarity is reversed about every two weeks.

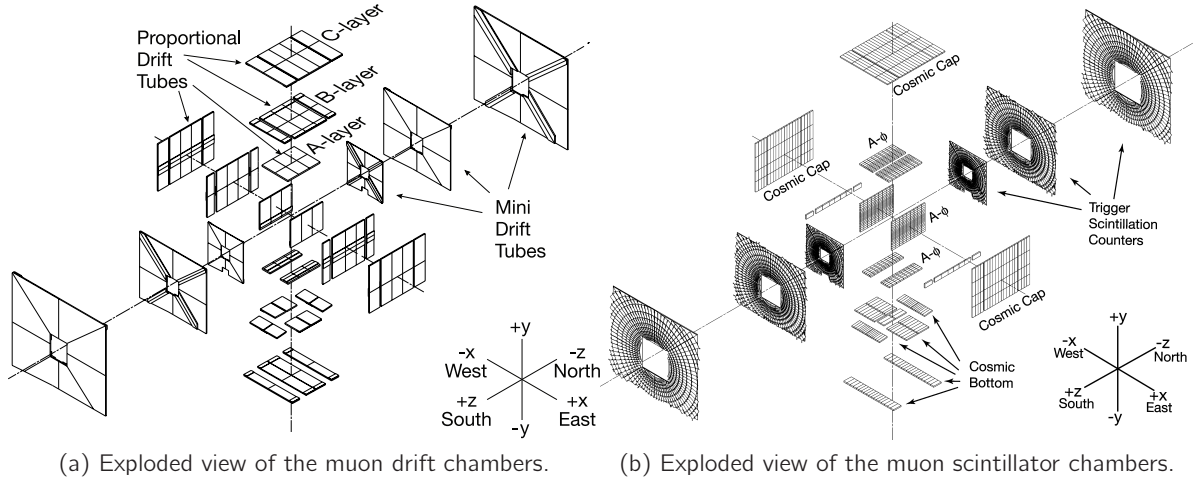


Figure 2.19: Exploded view of the muon drift chambers (a) and muon scintillator chambers (b).

The muon system is built in 3 planes called layer A, B and C. The layer A is placed between the calorimeter and the toroidal magnets, where layers B and C are in the outermost part of the detector. Since The detector coverage is limited over the ϕ angle, notably there is a hole in the A layer coverage at the bottom as visible in Figure 2.19, to allow for the calorimeter support, also

called “phi-hole” or “phi-crack”.

Muon drift chambers

The central and forward of the muon drift chambers system is equipped with two different kinds of technologies: Proportional Drift Tubes (PDT) and Mini Drift Tubes (MDT).

PDTs, covering the $|\eta_{det}| \lesssim 1$ region, are divided in cells, and their number is different according to the layer: 96 in layer A and 72 for layer B and C. Each cell is composed of an anode wire connect to a 4.7 kV high voltage and 2 pads operating with a high voltage of 2.3 kV. When muons go through cells, an mixture of gas (84% Argon, 8% methane, and 8% CF₄) is ionized and secondary electrons are created in the medium. These charges are collected by the wires. The drift velocity is approximately 10 cm/ μ s, for a maximum drift time of about 500 ns.

Similarly to PDTs, MDTs are also arranged in ABC layers in the $1 < |\eta_{det}| < 2.15$ region. They consist of tubes divided in 8 cells, each one containing a W-Au anode wire embedded in an aluminum compartment. The operating principle is similar to PDTs, where a 90% CF₄ – 10%CH₄ gas mixture is ionized and charges are collected by the wires. The charge drift times is between 40 and 50 ns.

Muon scintillator chambers

Scintillator chambers have their signal sent to photomultipliers (see Figure 2.20). The most central part is covered by the “A ϕ ” layer in the $|\eta| < 1$ region, upstream to the A layer and are used for triggering, their timing response being of the order of 2 ns. They can also help in identification of muons stopped in the toroidal magnets since they are located before this magnet. The ϕ segmentation is identical to the CFT. Fast timing detector located in layer C are used to associate a muon in a PDT with the appropriate bunch crossing and discriminate against the cosmic ray background. Finally, scintillators in the forward region ($1 < |\eta_{det}| < 2$) are used for triggering.

2.2.6 Luminosity monitor

In order to measure the integrated and delivered luminosity by the machine, luminosity monitors are used since the number of particules in each bunch and the size cannot be measured when protons and antiprotons are present in the Tevatron. The purpose of such a detector is not to measure the energy of particles or identify them, but to estimate the instantaneous luminosity by estimating the number of inelastic collisions.

The instantaneous luminosity can be expressed as function of the parameters of the accelerator:

$$L = \frac{fBN_pN_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)} \cdot F \quad (2.2)$$

with $f \simeq 43.7$ kHz is rotation frequency of the bunches, $B = 36$ being the number of bunches circulating, N_p ($N_{\bar{p}}$) the number of protons (antiprotons) per bunch, σ_p ($\sigma_{\bar{p}}$) the transverse proton (antiproton) beam size and F being a form factor depending on the bunch length. Since parameters

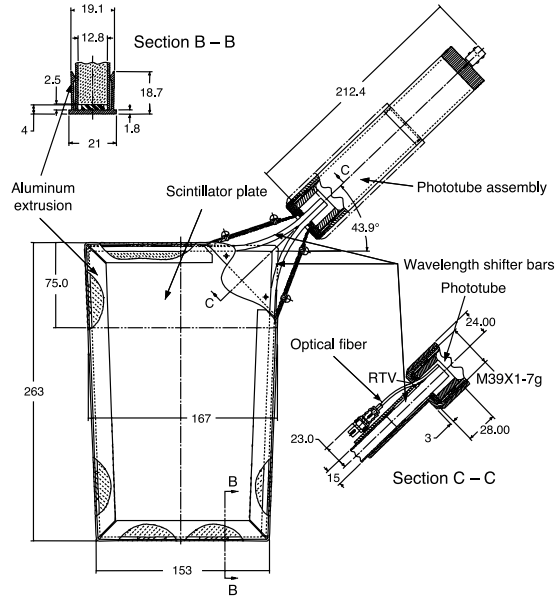


Figure 2.20: Schematic view of a scintillating plate connected to a photomultiplier for the forward muon system.

describing the instantaneous luminosity may change during a store, or from store to store, this quantity needs to be measured.

Located at each extremity of the detector ($z = \pm 140$ cm, which corresponds to the forward region $2.7 < |\eta_{det}| < 4.4$), luminosity monitors are placed around the beam pipe. They consist of 24 Bicron BC-408 scintillators wedges, a photomultiplier directly glued to the outer surface of each wedge to collect scintillation light. Wedges are arranged as shown in Figure 2.21. Since scintillators are exposed to a high radiation flux, they were replaced four times during the RunII period.

Knowing when bunches are crossing inside the DØ detector, the luminosity monitors will only count the number of crossings for which no collisions occurred. This technique is called the “counting empties” method and can avoid the difficulty for the case that multiple interactions occurs during the same crossing.

The luminosity \mathcal{L} can be expressed as $\mathcal{L} = \frac{f \bar{N}_{LM}}{\sigma_{LM}}$, where f is the beam crossing frequency, \bar{N}_{LM} is the average number of interaction per bunch crossing and σ_{LM} is the effective $p\bar{p}$ cross section estimated from previous experiments or in RunI [34].

Given that the probability of having N inelastic collisions per crossing follows Poisson statistics, $\mathcal{P}(N) = \frac{\mu^N}{N!} e^{-\mu}$, the probability to measure zero inelastic collision per crossing is $\mathcal{P}(0) = e^{-\mu}$, with $\mu = \mathcal{L} \frac{\sigma_{LM}}{f}$.

Luminosity monitors count the amount of “no collision” events occurring in coincidence in both sides of the detector, the rate of inelastic collisions can be expressed as $R_c = f(1 - \mathcal{P}(0))$. The measured instantaneous luminosity is then deduced from this expression.

An uncertainty on the measured instantaneous luminosity is found to be 6.1%, coming from the determination of σ_{LM} and detector efficiency and acceptance.

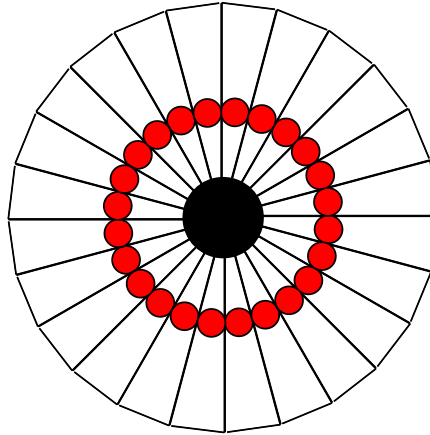


Figure 2.21: Schematic view of the luminosity monitor geometry and location of photomultipliers (red circles).

2.2.7 Trigger and data acquisition system

The time between two bunch crossing being of 396 ns, the frequency at which events would be recorded by the DØ detector would be around 2.5 MHz. Recording such amount of data is not achievable from a technical point of view, taking into account that some of these events could be detector noise or have a high cross section, which may necessitate to be filtered.

A preliminary selection on objects reconstructed in the detector is imposed in order to reduce the acceptance and the recording rate. This allows to reject for example low p_T multijet events which have a very high cross section, and not used in analyses. The trigger system is used for this purpose and is designed in multiple layers (see Figure 2.22), each of them relying on detector information as shown in Figure 2.23.

Only events in which an inelastic collision occurred are triggered. Therefore both luminosity monitors are required to have hits in coincidence with the bunch crossing. Such events are called “Minimum Bias” events.

Sets of conditions are defined in order to “label” events. The data can be then described by logical “and” and “or” of several trigger conditions.

Since a high rate of events may still satisfy all trigger requirements, events are removed randomly in order to accommodate with the recording bandwidth. These “prescales” are different according to the event topology.

Level 1 trigger

First, the decision at Level 0 to retain an event is taken in 3.5 μ s, lowering the event rate from 2.5 MHz to 2 kHz. the Level 1 relies on the fast electronics readout of parts of the detector (CFT, preshower detectors, calorimeter, muon spectrometer).

Information from the CFT is used in order to reject events with high p_T fake tracks. This is done by comparing recorded data to predefined models of tracks in 4.5° angular sections of the CFT.

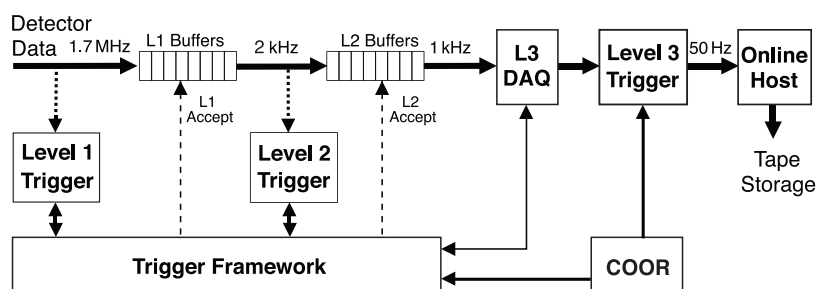


Figure 2.22: Data flow in the DØ trigger system. The event rate is reduced from 1.7 MHz to 100 Hz.

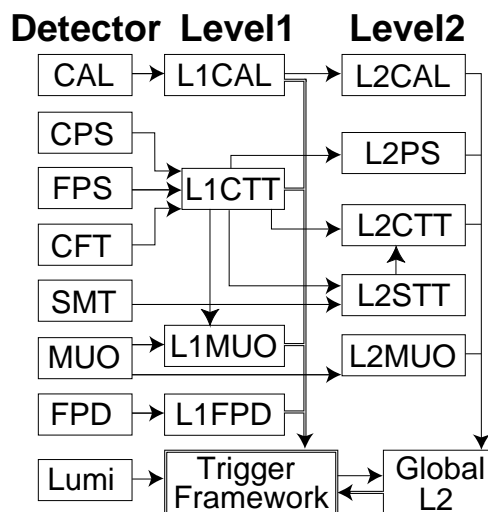


Figure 2.23: Level 1 and 2 of the DØ trigger system. Arrows represent the data stream.

Calorimetric triggers rely on reconstructed “trigger towers” in $\Delta\phi \times \Delta\eta = 0.2 \times 0.2$, passing above predefined energy thresholds. These thresholds are different depending if only electromagnetic or calorimetric cells are considered to reconstruct towers. Only towers in the $|\eta_{det}| < 3.2$ region are used.

Muon triggers are based on the matching of tracks produced in the CFT, given that they fulfill the above mentioned requirement, and hits obtained at least in the A or B layer of the muon system.

Level 2 trigger

The Level 2 trigger decision reduce the rate from 2 kHz to 1 kHz. The trigger must be taken in less than 100 μs . All parts of the detector are used at this level. Since the allowed is longer than in Level 1, more complex algorithms can be used and a fast reconstruction of electrons, photons and jets can be performed.

Jets (electromagnetic objects) are reconstructed in groups of 5×5 (3×3) trigger towers, with the centermost tower having an energy higher than 2 GeV (1 GeV). Electrons and photons can be identified if a Level 1 track is match to clusters in the preshower detectors. An estimate of the missing transverse can be obtained by these “raw” calorimetric objects.

At this level, trigger requirements on tracks are tighter by using SMT and CFT informations. At least 3 SMT hits around tracks found at the Level 1 are required. This allows to have a good impact parameter resolution.

The muon Level 2 trigger uses information from Level 1, combined with additional detector informations from the wire chambers and scintillator chambers to give a better precision of the object coordinates (η , ϕ) and transverse momentum.

Correlations between objects are checked before events are passed to the Level 3.

Level 3 trigger

This last step reduces the event rate down to around 50 Hz, depending on the Tevatron instantaneous luminosity. The trigger decision is taken in less than 150 ms. After all events successively passed Level 1 and 2, a partial event reconstruction is performed on a farm of 400 computers. The output of the Level 3 trigger is the electronic signal from all detector parts, to be sent to the processor which will reconstruct objects as it is detailed in Section 3.

2.2.8 Data format and detector simulation

The format of the data after passing the 3 levels of the trigger system is the electronic signal coming from all parts of the detector, called also “raw data”. In order to analyze the data, objects have to be reconstructed. This process is carried by the `d0reco`¹ software.

In simulated events obtained from Monte Carlo event generators, objects have to be reconstructed as well, but after detector simulation. This part is taken care by the `d0gstar`² software, which is based on the description of the structure and electronics of the DØ detector by Geant3.

¹standing for DØ offline RECOstruction program

²standing for DØ Geant Simulation of the Total Apparatus Response.

Hence, simulated events are in the same format as raw data, and reconstructed in a similar manner. In order to reproduce multiple bunch crossing in simulation and detector noise, randomly selected events from data (Zero Bias data) are overlayed to the simulation.

Data and simulation events are stored in ROOT files, with sufficient informations from detector parts. A common format, *CAFe* [35], is used across the DØ collaboration in order to unify software. In order to facilitate data analysis without dealing with unnecessary big files of data containing irrelevant events for a given analysis, both data and simulation are subdivided in smaller samples. This splitting is performed by a logical OR of several basic object selection and trigger requirements. In the *WH* analysis, the *EMinclusive* and *MUinclusive* skims are used respectively for the electron and muon channels.

3

Objects reconstruction and identification

The DØ detector and main subparts have been described in Section 2.2. Informations about particles produced from $p\bar{p}$ collisions are gathered by nearly individual channels from all parts of the detector. This raw information is at this level only digital values, output of every channel. This data needs to go through the DØ reconstruction software in order provide useful information that can be latter on analyzed, such as energy, momentum, spatial position, charge. . .

The WH analysis is aimed to look for a final state with isolated leptons, jets and imbalance in the transverse plane. For this type of event, almost all parts of the DØ detector come in use. An accurate reconstruction of objects and identification is crucial. Starting from basic objects, such as tracks and vertices, an overview is given on how electromagnetic objects like photons and electrons, jets, muons and the missing transverse energy are reconstructed and identified. Treatments for data and simulation will be discussed, as well as tagging of jets originating from b quarks, crucial for searches for a low mass Higgs boson.

3.1 Tracks

Tracks are the basic ingredient to reconstruct objects created from $p\bar{p}$ collisions. Their presence is used for the following:

- reconstruction of vertices from the hard scatter process or displaced vertices coming from the decay of long lived particles,
- electrons and photons have a similar signature except the presence of tracks associated to energy deposit in the calorimeter for electrons,
- muons are reconstructed from hits in the muon spectrometer and their associated track in the SMT and CFT system.

They also provide a measurement of the momentum of charged particles.

Track relies on the SMT and CFT systems and is achieved by two different algorithms: the Histogram Track Finder (HTF) [36] and the Alternative Algorithm (AA) [37], both using the Kalman Fitter technique [38, 39].

The amount of tracks reconstructed in the tracking systems is high: when protons and antiprotons collide, residual fragments of the hard scatter process (spectator particles) have also a signature in the detector. In this busy environment in the vicinity of the interaction point, efficient track reconstruction is very challenging.

The motion of a particle in a magnetic field can be characterized by the curvature of its track $\rho = \frac{qB}{p_T}$, with q being the charge of the track and B the amplitude of the magnetic field, given the track transverse momentum p_T .

Let us consider a simple example to illustrate the principle of the HTF algorithm, assuming that the primary interaction vertex occurred in $(x, y) = (0, 0)$ and a hit in the tracking system in (x_h, y_h) in the same coordinates. In the presence of a single hit, an infinity of “track solutions” can be found. Using the Hough transformation, each solution can be defined in the (ϕ_i, ρ_i) coordinated for the i^{th} track solution. Given the fine spatial resolution, trajectories can be extrapolated to a line in this same plane.

Repeated for several hits, each line should intersects at a same point, thus defining the track parameter for a charged particle. However, this technique relies on 2D histograms, representing the finite segmentation of the tracking system. Histograms are filled for each hit possibly originating from the same charged particle, the track parameters can be extracted with associated measurement errors. This procedure, well-suited for tracks with a small distance of closest approach to the primary interaction vertex, is summarized and depicted in Figure 3.1.

The second algorithm used at DØ is the AA algorithm. The principle of this technique is to built “track hypotheses” and then filter them according to defined criteria.

Starting from innermost SMT hits, a second point from a downstream SMT layer can be added to the track hypothesis if the angular separation between tracks $\Delta\Phi < 0.08$ is fulfilled. Then a third point located further down the trajectory of the track hypothesis is added if the radius of circle formed by hits is greater than 30 cm, corresponding to a track p_T of 180 MeV in a 2 T magnetic field. Finally, in order to be retained, further requirements need to be fulfilled such as: χ^2 from the track fit is computed and must be less than 16 for the hypothesis to be retained, or a maximum number of layers (axial and stereo) not containing hits along the track hypothesis. Since duplicate tracks can be found, a removal procedure is carried.

Since the two presented algorithms are used in parallel, only candidate tracks reconstructed by both methods are kept. The efficiency for finding a track is $0.966 \pm 0.001\%$ at an instantaneous luminosity of $2 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$, for the case that no additional events from elastic interactions occurred in the same bunch crossing [40].

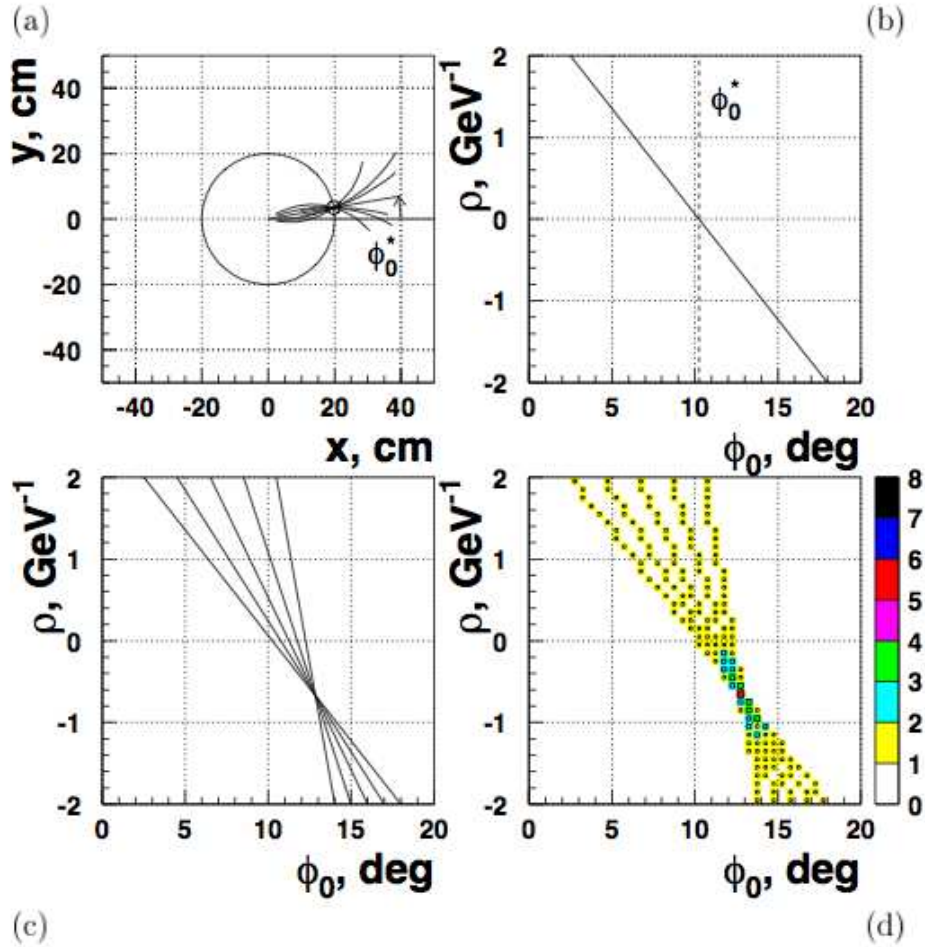


Figure 3.1: The HTF method applied to a single 1.5 GeV track with 5 hits: (a) a set of track solutions is defined in the (x, y) plane for a given hit. (b) Trajectories of tracks in the (ϕ, ρ) plane for a given hit. (c) Trajectories for several hits intersect at the same point in the (ϕ, ρ) plane. (d) The intersection can be seen as a peak in a 2D distribution.

3.2 Primary Vertices

In order to characterize an event topology, one needs to reconstruct the initial proton antiproton collision. An accurate measurement can be critical especially for the missing transverse energy measurement, determination of the η component of jets, as well as their identification (See Section 3.5.2), which are sensitive to the primary vertex (PV) selection.

Since the Tevatron is running at high instantaneous luminosity, multiple interactions can occur during a single bunch crossing, but usually there is only one vertex produced from the $p\bar{p}$ hard scatter process. The other vertices, coming from elastic $p\bar{p}$ collisions are referred to as *minimum bias* (MB) events. The physics of these additional collisions are not relevant in the context of

the search presented here. It is possible to suppress this background by selecting the PV. This is achieved in two steps: first all vertices are reconstructed, then the least consistent vertex with a MB hypothesis is selected.

Candidate tracks attached to the vertex from the hard scatter are first selected with $p_T > 0.5$ GeV and at least two hits in the SMT. In order to suppress tracks originating from secondary vertices (probing the presence of long lived particles, such as K_S^0 , Λ or B mesons), $\frac{DCA}{\sigma(DCA)}$ is required to be less than 5, where DCA is the distance of closest approach of the track to the beam line in the $r-\phi$ plane, and $\sigma(DCA)$ is its associated uncertainty. Selected tracks are then clustered along the z direction. Using the Kalman Fit method, each cluster has a χ^2 value assigned, denoting how likely are the tracks to originate from the origin. If the χ^2 is larger than 10, the track which contributes the most to the χ^2 is removed. This process is repeated until $\chi^2 < 10$ or only one track is left. Following this procedure, each cluster is a PV candidate.

Since tracks from MB interactions have a lower p_T than tracks from the hard scatter, a MB track probability is computed for each track as following:

$$P(p_T) = \frac{\int_{\log_{10}(p_T)}^{\infty} \log_{10}(p_T^{track}) dp_T^{track}}{\int_{\log_{10}(0.5)}^{\infty} \log_{10}(p_T^{track}) dp_T^{track}}. \quad (3.1)$$

It is then possible to assign to each PV candidate a MB vertex probability:

$$PMB = \Pi \sum_{k=0}^{N-1} \frac{-\ln \Pi}{k!} \quad (3.2)$$

with Π being the product of MB track probabilities for N tracks. The vertex having the lowest MB vertex probability is chosen to be the correct PV, with an efficiency close to 100%.

In order to take into account the difference in instantaneous luminosity between data-taking periods and imperfect detector simulation of the tracking system (which may affect the track selection), the z position of the primary vertex is corrected in order to match observations in data.

3.3 Electrons

With a view to probe the leptonic decay from a W boson as in the WH analysis, electrons must be accurately identified in order to not introduce backgrounds with an electromagnetic object (photon or pion) which is misidentified as an electron. This section will focus on the reconstruction and identification of electrons, as well as the treatment used to match simulated resolution to the one obtained in data.

3.3.1 Reconstruction and identification criteria

Electrons are calorimetric objects which interact mostly in the electromagnetic calorimeter by producing a narrow shower compared to jets. Since they are charged particles, an associated track may be found. By exploiting these characteristics, candidate electrons are reconstructed the following way:

1. calorimetric towers are clustered in a cone if their measured p_T is greater than 500 MeV and are called “seeds”,
2. calorimetric towers in $\Delta R < 0.3^1$ are added to the seeds if their p_T is greater than 50 MeV,
3. if the clustered energy of a group of towers is higher than 1 GeV, all electromagnetic towers in $\Delta R < 0.4$ are added.

The next step is to enhance the electron purity since such objects could be photons or jets matching the previous criteria. A set of variables, calculated for each electron candidate, defines the electron identification criteria [41]:

- electromagnetic fraction EMf , calculated as the ratio of energy in the electromagnetic calorimeter in a $\Delta R = 0.2$ cone to the energy in the calorimeter in a $\Delta R = 0.4$ cone, $EMf = \frac{E_{EM(0.2)}}{E_{tot(0.4)}}$. This value is expected to be close to 1 for electrons for which most of their energy is deposit in the electromagnetic layers of the calorimeter,
- calorimeter isolation Iso , the fraction of total calorimeter energy in a $\Delta R = 0.4$ cone after subtracting electromagnetic calorimeter in a $\Delta R = 0.2$ cone to the total calorimeter energy, $Iso = \frac{E_{tot(0.4)} - E_{EM(0.2)}}{E_{EM(0.2)}}$. This observable can discriminate electrons from jets, where this value would be respectively close to 0 and 1, for electrons and jets, since electrons will produce showers mainly contained in the EM layers,
- track isolation of the EM cluster $IsoHC4$, calculated as the total track p_T (for tracks with $p_T > 0.5$ GeV) in the hollow cone $0.05 < \Delta R < 0.4$ around the EM cluster,
- “track match χ^2 probability” with an electromagnetic cluster, efficient in photon background removal,
- $HMatrix$ variable characterizing the lateral shower shape of EM cluster,
- electron likelihood variable², combined and trained to discriminate electrons from jets [42].

Using the above defined variables, several working points for electron identification are defined. Those used in the WH analysis are given in Table 3.1. Two types of criteria are used and events containing electrons in the CC and EC are analyzed separately, thus 4 types of electrons are defined.

3.3.2 Electron resolution and energy scale correction

Electrons being calorimetric objects, their resolutions follows Equation 2.1. A procedure is set up in order to match the electron resolution to data, since it is better in simulation [33]. The electron energy is transformed using the relation:

$$E' = E \times [\alpha + \xi = Gaus(0, \sigma = \alpha c)] \quad (3.3)$$

¹ ΔR denotes a distance in the (η, ϕ) plane and is defined as $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\phi$ and $\Delta\eta$ are the azimuthal angle and pseudorapidity differences between two objects in the detector.

²The 7 likelihood variables are: spatial track match χ^2 probability, E_T/p_T , H-matrix7, EMF, distance of closest approach to primary vertex, number of tracks in a $\Delta R = 0.05$ cone, total p_T of tracks in 0.4 cone around candidate track.

Variables	CC		EC	
	LOOSE	MEDIUM	<i>top_loose</i>	<i>top_tight</i>
<i>Iso</i> <	0.10	0.07	0.15	0.15
<i>EMf</i> >	0.95	0.97	0.9	0.9
<i>HMatrix</i> <	35	25	50	50
<i>IsoHC4</i> <	3.0	2.5	–	–
Track Match χ^2 >	0.0	0.0	0.0	0.0
Likelihood >	–	0.2	0.0	0.85

Table 3.1: Electron identification point used in the WH analysis, defined by the listed cuts on identification variables.

where α is an energy scale multiplicative factor and ξ is an oversmearing factor following a gaussian distribution centered in 0 and a width of αc . The α and c parameters are determined by adjusting these values through a fit of the di-electron mass distribution around the Z mass peak in simulation to data.

3.3.3 Identification efficiency between data and simulation

The electron efficiencies are determined using a “tag-and-probe” method. $Z \rightarrow e^+e^-$ candidate events are selected by requiring two electron candidates with a di-electron invariant mass consistent with that of a Z boson. One of the electron candidates, denoted as the tag object, is required to pass stringent cuts to improve the purity of the sample, while the other candidate, the probe, is required to pass the cuts relevant to the working point efficiency to be determined. The efficiency ϵ is measured by counting the number of probe electrons passing and failing the identification cuts: $\epsilon = \frac{N_{pass}}{N_{pass} + N_{fail}}$. A detailed overview of the method used to determine the identification efficiency is given in [43].

An electron satisfying the MEDIUM identification criteria with $p_T = 30$ GeV is found to have an efficiency of ~ 0.85 in MC and ~ 0.77 in data.

Due to imperfect detector simulation, the measured identification efficiency in MC (ϵ_{MC}) is better than in data (ϵ_{Data}). A correction is applied to the simulation where the event weight is multiplied by $\epsilon_{MC}/\epsilon_{Data}$. This correction is parametrized as function of the η_{det} and ϕ components of the electron, as shown in Figure 3.2 for MEDIUM electrons.

3.4 Muons

The reconstruction and identification of muons relies on both tracking sub-systems (SMT and CFT) and muon spectrometer since only a small fraction of their energy is deposit in the calorimeter.

3.4.1 Reconstruction and identification criteria

Two independent parts of the detector are used for muon identification. Muons reconstructed only in the three layers of the muon detector are called *local* muons. In the tracking system, tracks

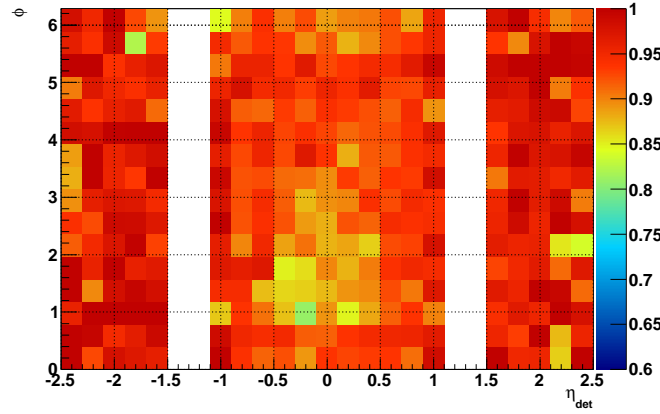


Figure 3.2: $\epsilon_{MC}/\epsilon_{Data}$ for MEDIUM electrons.

can be matched to local muons which are extrapolated back to the innermost part of the detector. They are called *central muons*.

Starting from these definitions, muons are classified by different criteria [44, 45]:

- the muon type which characterizes the matching between local and central muon,
- the muon quality, estimated from the number of hits in the muon spectrometer
- the track quality, analogous to the muon quality
- the muon isolation.

The different muon types are defined by the *nseg* parameter, an integer value between -3 and 3. The absolute value indicates the number of layers (or segments) that are reconstructed. A positive *nseg* value denotes if a matching with a track occurred. The definition for the different types are given in Table 3.2.

The muon quality used for muon identification in the *WH* analysis, called Medium $|nseg| = 3$, is defined by the following requirements:

- A layer: at least two hits in the drift chambers and at least one hit in the scintillator chambers,
- BC layers: at least two hits in the drift chambers and at least one hit in the scintillator chambers (except for central muons with less than four hits in the BC drift chambers).

The muon track quality used in the analysis fulfills the medium criteria, defined as:

- $|DCA| < 0.2$ cm. If the track has SMT hits, the cut is tighten to $|DCA| < 0.02$ cm.
- $\chi^2/d.o.f. < 4$.

$nseg$	Muon Type	Central track matching algorithm
3	Central track + local muon track (A, B and C segments)	Muon to central if local muon track fit converged. Central to muon otherwise.
2	Central track + BC only	central to muon
1	Central track + A only	central to muon
0	Central track + muon hit in layer A, B or C	central to muon
-1	A segment only	no match
-2	BC segment only	no match
-3	local muon track (A, B and C segments)	no match

Table 3.2: Muon types at DØ . Criteria requirement on track, segment hits and matching of local and central muons.

Muon isolation variables are defined in order to separate non-isolated muons background coming from semi-leptonic decays in jet (such as $B \rightarrow \mu + X$) and muons originating from W or Z bosons. Three of the five variables, used later on, are defined:

- $TrackHalo = |\sum^{tracks} p_T|$ in $\Delta R(track, muon track) < 0.5$ cone.
- $CalorimeterHalo = |\sum^{cells} E_T|$ in $0.1 < \Delta R(cal-cells, muon cal-track) < 0.4$.
- $\Delta R(\mu, jet) =$ Distance to closest jet in $\eta - \phi$ space.

Using these variables, a set of isolation working points are defined. The ones which are used in the WH analysis are defined as following:

- $\Delta R = \Delta R(\mu, jet) > 0.5$.
- $NPTight = TrackHalo < 2.5 \text{ GeV}$ and $CalorimeterHalo < 2.5 \text{ GeV}$.

Finally a cosmic muon veto is applied, using information from the scintillator hit times and coincidence with the bunch crossing. The hits timing requirement with respect to bunch crossing in A and BC layers are $|t_A| < 10 \text{ ns}$ and $|t_{BC}| < 10 \text{ ns}$.

The efficiency in data and simulation is obtained separately from the tag-and-probe method. In a similar way to electron identification efficiencies determination, $Z \rightarrow \mu^+ \mu^-$ events are selected: one “tag” muon is selected with rather tight requirements, the efficiency for each identification criteria being measured with the “probe” muon.

The efficiency for each identification criteria are given in Table 3.3 for 30 GeV muons in data.

Systematic uncertainties are estimated from selection cuts in the tag-and-probe method for each component of identification criteria. All muon identification criteria defined in DØ can be found in [44, 45].

	Muon quality: mediumnseg3	Track quality: mediumtrack	Muon isolation: NPTight
Efficiency	$\sim 72\%$	$\sim 87\%$ (instantaneous luminosity average)	$\sim 85\%$

Table 3.3: Identification criteria efficiencies for 30 GeV muons in data.

3.4.2 Muon energy resolution

By comparing the width of the Z mass peak in data and simulation, it is found that the resolution is better in MC (6.85 ± 0.12 GeV in data compared to 4.84 ± 0.03 GeV in MC). The simulation is smeared to better agree with observation in data [46] [47], using the following formula:

$$\frac{q}{p_T} \rightarrow \frac{q}{p_T} + AG_1 + \frac{B\sqrt{\cosh\eta_{det}}}{p_T}G_2,$$

where A takes into account the tracking resolution and B is the associated term for multiple interactions which can modify the track trajectory. The parameters of the formula are estimated using di-muon events in the J/ψ and Z mass peak in data and simulation.

3.4.3 Reconstruction efficiency in data and simulation

Typically,

In order to match the identification efficiency to what is measured in data, a scale factor is obtained for each component of the muon identification:

- the muon quality is parametrized in ϕ_μ and η_{det}^μ bins,
- the track quality has a dependence in z and η_{CFT} (pseudorapidity measured in the CFT), hence parametrized along these variables,
- the DeltaR and NPTight muon isolations have respectively their efficiencies parametrized in $\Delta R(\mu, jet)$ and in the $\Delta R(\mu, jet) \times p_T^\mu$ plane.

3.5 Jets

At the Tevatron, jet production events have a very high cross section and are produced at a high rate. These complex objects are the product of parton hadronization and fragmentation, leading to a spray of colorless particles according to the QCD theory. Jets are the experimental signature of partons. For the WH analysis, these objects are of utmost importance, since the mass of the Higgs boson candidate which decays in a $b\bar{b}$ resonance is reconstructed from the corresponding jets.

This section will focus on the reconstruction and identification of jets at $D\bar{D}$. The energy calibration will also be introduced, as well as resolution treatment for simulation.

3.5.1 Jet reconstruction

The reconstruction of jet objects relies only on the calorimeter. Since partons cannot be detected, their experimental signature is reconstructed using a jet algorithm. Such procedure needs some prerequisites:

- In QCD, emission of soft and collinear gluons can lead to divergences. The jet algorithm has to be insensitive to such radiations, if a soft gluon is radiated between two partons, the outcome of the reconstruction has to remain identical to the case there is no radiation (also referred as *infrared safety*, see Figure 3.3a). In case a parton is replaced by two partons (with same total momentum), the reconstruction has to be stable under this test (*collinear safety*, see Figure 3.3b).
- Jets cannot have a reconstructed energy above the kinematic threshold ($\sqrt{s}/2$ in the transverse plane).
- The reconstructed jet variables must be boost-invariant along the beam axis

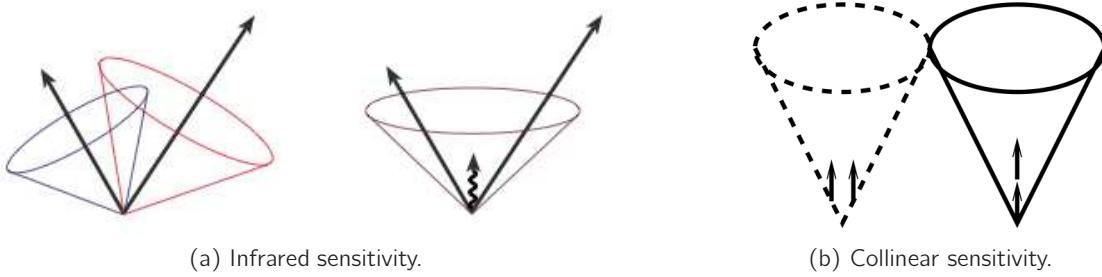


Figure 3.3: Examples depicting infrared and collinear sensitivity in jet reconstruction. Arrows represents jets, with their length proportional to the energy. (a) Infrared safety: by adding a soft emission between two jets, the result of the reconstruction scheme can be a merged jet with the three components inside. (b) Collinear safety: on the left, two candidate jets (or towers) which do not have enough energy to be reconstructed as a single jet can pass the reconstruction threshold if merged.

$D\phi$ uses the *RunII cone algorithm* [48]: all calorimeter towers or calorimeter cells which have energy deposited are clustered in a cone with its origin taken as the interaction point. The cone radius is fixed in the (Y, ϕ) plane. If the center of the cone corresponds to the direction of particles entering the cone, a jet is formed. Several steps are carried to reconstruct a jet and will be now detailed.

Combination of calorimetric cells and towers

Cells are considered as massless objects with a 4-vector $\mathbf{P}^{cell} = (E^{cell}, \vec{p}^{cell})$, where defined E^{cell} is the energy deposited in the cell and \vec{p}^{cell} is defined from the interaction point to the center of the cell. Noisy cells are removed with two algorithms. Let us consider σ^{cell} the energy corresponding to the measured width of the signal due to electronics noise. Isolated hot cells are removed if their

energy is above $2.5 \sigma^{cell}$ [49, 50]. Isolated low energy cells are removed if their energy is less than $4 \sigma^{cell}$ and they do not have neighbours above $4 \sigma^{cell}$ [51, 52, 53].

A reconstructed tower is obtained from calorimetric towers containing at least one cell that passed the above mentioned procedure. Hence, its 4-vector as:

$$\mathbf{P}^{tower} = (E^{tower}, \vec{p}^{tower}) = \sum_{i=\text{cells in tower}} (E_i, \vec{p}_i). \quad (3.4)$$

Kinematic properties of such object can be computed and are Lorentz invariant. The prescription for calorimetric elements combination is called the *E-scheme*.

Constructing preclusters: building blocks of jets

Starting from the list of towers obtained from the *E-scheme* prescription ordered in decreasing p_T , towers will be iteratively added to clusters. First a tower I with $p_T^I > 500$ MeV is taken to initiate a precluster¹. If the next tower J in the list has a separation $\Delta R(\text{precluster}, J) < 0.3$ and $p_T^J > 1$ MeV, it is combined to the formed precluster. This step is repeated until no towers are left. If the precluster has $p_T > 1$ GeV and composed of more than one tower, the object is retained in the list of preclusters for a particular event. Preclusters will be now used to form jets using the RunII cone algorithm.

RunII cone algorithm

The RunII cone algorithm forms proto-jets with cone radius $\Delta R = 0.5^2$, using preclusters as “seeds”. Similarly to the previous step, preclusters are ordered in decreasing p_T . A loop over preclusters is performed : if a precluster P fulfills the condition $\Delta R(P, \text{proto-jet}) < \frac{R_{cone}}{2}$, it is added to the proto-jet. In case a precluster isn't matched to a proto-jet, it will be used to form a new proto-jet.

A stable cone is obtained from proto-jets as towers are added to the cone, and its center is evaluated using the *E-scheme* procedure. The iterative process ends if one of the following conditions are fulfilled:

- p_T of the candidate is less than 5.5 GeV,
- the cone is stable *i.e.* $\Delta R(P_i, P_{i+1}) < 0.001$, where P_i is the proto-jet formed at the i^{th} iteration,
- the maximal number of iterations is reached $N_{it} = 50$.

In order to ensure infrared safety during reconstruction, a procedure is introduced by adding *mid-points* [54], *i.e.* points formed by two proto-jets A and B with $R_{cone} < \Delta R(A, B) < 2 \times R_{cone}$. The clustering method mentioned above is repeated by using midpoints instead of preclusters, with the exception that there is no requirement on the distance between the preclusters and midpoints.

A list of proto-jets formed by midpoints and proto-jets is then obtained.

¹If a tower contains its highest p_T cell C in the coarse hadronic layer of the calorimeter or in the massless gaps, then it is required that $p_T^C - p_T^{\bar{C}} > 500$ MeV

²Some analysis, especially in QCD measurements use jets with a cone radius $\Delta R = 0.7$. The procedure to reconstruct these jets is similar.

Overlap treatment

Since they may have towers in common, energy can be double-counted. Hence, a treatment of overlapping towers is needed. The *Merging/Splitting* algorithm loops over all proto-jets and if a proto-jet X has at least one tower in common with proto-jet Y , the sum of p_T of the towers shared with B is calculated. If it exceeds 50% of p_T^B , proto-jets are merged. If this value is lower than 50%, proto-jets are splitted. This process is repeated until no proto-jet is left.

The final reconstructed objects are those fulfilling this procedure with $p_T > 8$ GeV.

3.5.2 Jet identification and vertex confirmation

The identification procedure is carried in order to reduce the amount of jets not originating from outgoing partons of the hard scatter process, such as reconstructed fake jets from hardware noise.

A serie of cuts on variables are introduced:

- the electromagnetic fraction EMf , defined in Section 3.3.1 is required to be less than 0.95 in order to rejected electron or photon-like objects. Moreover, a minimal requirement $EMf > 0.05$ has to be fulfilled to reject objects with most of their energy in the hadronic part of the calorimeter, which is more subject to hardware noise,
- the coarse hadronic fraction CHf , fraction of energy in the hadronic layers of the calorimeter and analogous to EMf , is required to be less than 0.15,
- the *Hot Fraction* is calculated as the energy ratio of the most energetic tower to the second most enegetic one. A minimal value of 0.40 aims at rejecting fake jets having unusually high energetic towers.
- at most two towers must contain 90% of the jet energy.
- a Level 1 trigger confirmation is obtained by requiring that the ratio of the jet p_T to the p_T of towers firing the Level 1 trigger to be greater than 0.5.

Jets passing the above mentioned selection are denoted as “Good jets”. A description of all cuts applied can be found in [55] and [56] for the Level 1 confirmation.

Since the beginning of the RunII, the instantaneous luminosity has been constantly increasing. Therefore, multiple interactions can occur during the same bunch crossing. In order to get rid of additional jets not belonging to the hard scatter, additional criteria are required on jets, on top of the “Good jet” requirements and must be “Vertex Confirmed”. This means that jets must have at least 2 tracks attached to the primary vertex. These tracks must have:

- $p_T > 0.5$ GeV,
- at least one hit in the SMT,
- distance of closest approach to the beam line in the transverse plane $DCA_{xy} < 0.5$ cm,
- distance of closest approach to the beam line in the z direction $DCA_z < 1$ cm,

- $\Delta z(\text{vertex, track}) < 2.0 \text{ cm}$.

This requirement, after Jet Energy Scale being applied (see Section 3.5.3), allows to reduce significantly jets from minimum bias, as shown on Figure 3.4. For simulated events, jets are randomly removed according to their associated scale factors, shown as function of η_{det} and z_{PV} in Figure 3.5.

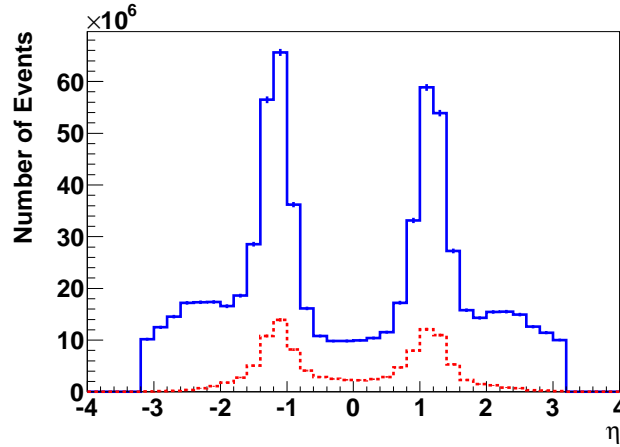


Figure 3.4: η distribution for jets in data from zero-bias (ZB) events (events with the only requirement of being produced in coincidence with the bunch crossing. There is no explicit requirement on luminosity monitors hits as opposed to minimum bias events.). Reconstructed jets with and without Vertex Confirmation requirement are shown respectively in dashed red line and blue line. One can see an important reduction of such jets by requiring at least 2 tracks attached to jets pointing to the primary vertex [57].

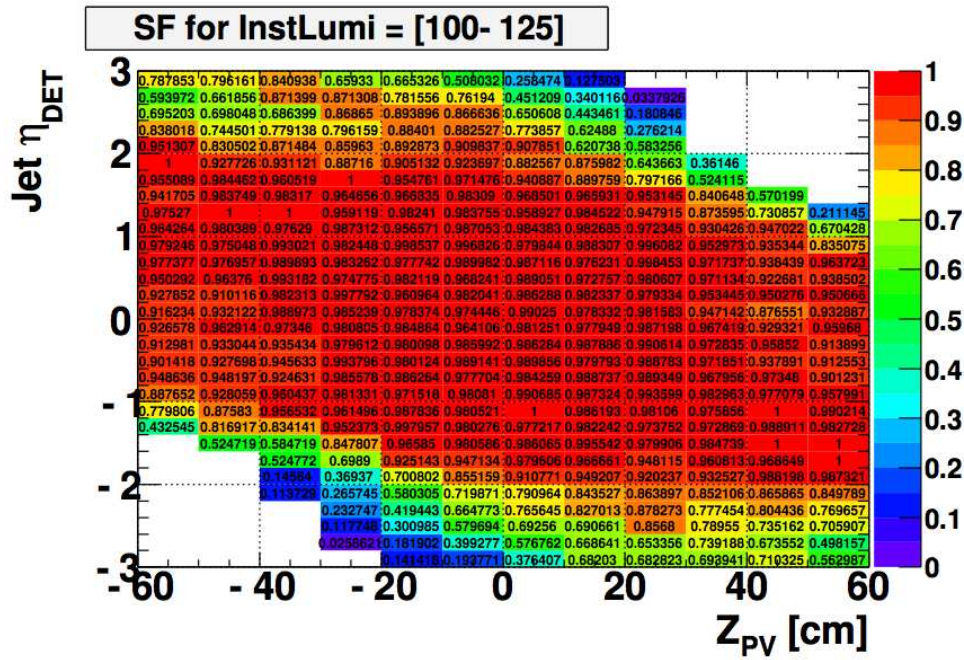


Figure 3.5: Parametrization of the Vertex Confirmation scale factors in the (η_{det}, z_{PV}) plane, shown for $100 \cdot 10^{30} < \mathcal{L}_{inst} < 125 \cdot 10^{30} \text{cm}^{-2}\text{s}^{-1}$.

3.5.3 Jet Energy Scale

Now that the jet reconstruction and identification processes have been detailed, one need to be able to make measurements with such objects, originating from the quark hadronization and interaction in the calorimeter. In the scope of the WH analysis, an important variable is the invariant mass of the two jets, since the Higgs boson decays to $b\bar{b}$ pair. A precise measurement of the jet energy is therefore capital. A calibration is performed and described here.

The jet evolution in time can be separated at three different levels, as the jet develops in the detector:

- partons are produced from the hard scatter process from a $p\bar{p}$ process. They can eventually radiate additional partons, also denoted as final state radiations (FSR), and form a “parton jet”, which is only accesible at the simulation level,
- undergoing a process of hadronization by color reconnection, hadrons are produced, forming a “particle jet” which can be reconstructed in the simulation by clustering and adding energy of all stable particles,
- the spray of produced hadrons is interacting in the detector, from the beam pipe and downstream material, until reaching the calorimeter, where a “calorimetric jet” is defined and its energy measured.

The evolution of a jet object is shown in Figure 3.6, with the types of jet listed above are inside a jet cone.

The measured calorimetric energy E_{jet}^{raw} is different from energy of the associated parton jet and particle jet. In order to be able to compare theoretical predictions at the particle jet level, The jet energy scale is designed to infer the particle jet energy from E_{jet}^{meas} . JES corrections take all calorimeter effects into account to be able to give the best possible prediction of the particle jet energy E_{jet}^{ptcl} .

The method consists of several corrections to deal with each different effect. First the energy not associated to the hard scatter (offset correction O) is subtracted from E_{jet}^{raw} . Then, from γ -jet samples, the jet response in the calorimeter is corrected mainly to account for effects such as energy loss in uninstrumented detector regions (absolute response correction R). Since jets have a different response in the calorimeter in different pseudorapidity regions, a relative response correction (η -intercalibration F_η) is determined, in order to deal with the different depths of calorimeter to go through. A final correction is determined to take into account for energy flowing inside the jet cone from other contributions or outside due to the effect of the magnetic field or shower development (showering correction S). Potential biases introduced from the previous corrections are removed using a overall factor k_{bias} .

The correspondence between E_{jet}^{raw} and E_{jet}^{ptcl} is given by the following expression:

$$E_{jet}^{raw} = \frac{E_{jet}^{ptcl} - O}{F_\eta \cdot R \cdot S} \cdot k_{bias} \quad (3.5)$$

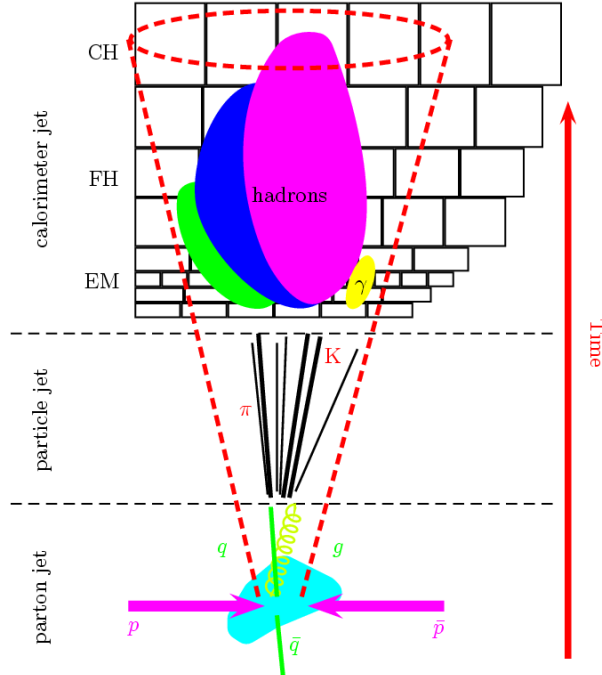


Figure 3.6: Representation of the evolution of a jet inside the detector, from the jet of partons produced from the hard scatter interaction to the interaction of hadrons inside the calorimeter.

where each term corresponds to the corrections listed above (except for k_{bias} , which has a small effect on the total JES, more details can be found in [58] and [59]). They will now be briefly detailed before evaluating the final result of the JES method.

Offset correction

Energy measured by the calorimeter may originate from different contributions and not associated to the primary interaction. Different effects such as electronic calorimeter noise, electronic signal from previous collisions not yet read-out (pile-up) need to be taken into account. Moreover, multiple interactions $p\bar{p}$ can occur during the same bunch-crossing inside the calorimeter.

To extract the offset correction, minimum-bias (MB) and zero-bias (ZB) events are used¹ to determine the energy density of calorimeter towers with the requirement of having no reconstructed primary vertices in ZB events. The MB tower energy density is determined in bins of primary vertex multiplicity (N_{PV}) and instantaneous luminosity (\mathcal{L}_{inst}).

The correction is estimated as following:

$$O(N_{PV}, \eta, \mathcal{L}_{inst}) = MI(N_{PV}, \eta, \mathcal{L}_{inst}) - MI(N_{PV} = 1, \eta, \mathcal{L}_{inst}) + NP(\eta, \mathcal{L}_{inst}) \quad (3.6)$$

¹As a reminder zero-bias events are triggered at every bunch crossing in DØ and minimum-bias are triggered with the only requirement of hits in the luminosity monitors.

where MI is the contribution from multiple interaction in MB events, and NP from ZB events with no primary interaction detected. The offset correction is show in Figure 3.7.

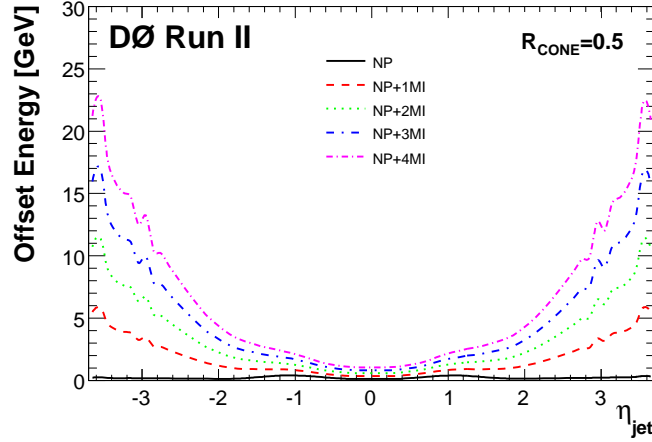


Figure 3.7: Offset energy for different primary vertex multiplicities, as function of η_{det}^{jet} .

Absolute response correction

The dominant correction for the JES is the absolute response correction. This correction is derived by using γ +jet events, with a photon with tight identification criteria and $|\eta^\gamma| < 1$ and a jet in the $|\eta^\gamma| < 0.4$ range. Using a two body process, photon and jet candidates are required to have $\Delta\phi > 3.1$ to select back-to-back events. The correction is derived using the “Missing E_T Projection Fraction method” (MPF). At the particle level, such selected events satisfies the relation:

$$\vec{p}_T^\gamma + \vec{p}_T^{had} = \vec{0}, \quad (3.7)$$

with \vec{p}_T^{had} being the hadronic recoil to the photon. In the ideal case, this relation should be preserved at the detector level. However, at the detector level, missing energy in the transverse may be measured coming from the fact that the response of the hadronic part isn't calibrated, leading to the following relation:

$$\vec{p}_T^\gamma + R_{had} \vec{p}_T^{had} = -\vec{\cancel{E}}_T, \quad (3.8)$$

with R_{had} being the hadronic response of the calorimeter. This quantity can be extracted via the following formula:

$$R_{had} = \frac{\vec{\cancel{E}}_T \cdot \vec{p}_T^\gamma}{\vec{p}_T^{\gamma 2}}, \quad (3.9)$$

estimated in bins of $E' = E_T^\gamma \cosh \eta_{jet}$ since the photon energy resolution and jet η direction is much better than the raw jet energy resolution. The correction is shown in Figure 3.8. A correspondence between E' and E_{jet}^{raw} is then performed to report the absolute response factor as function of the

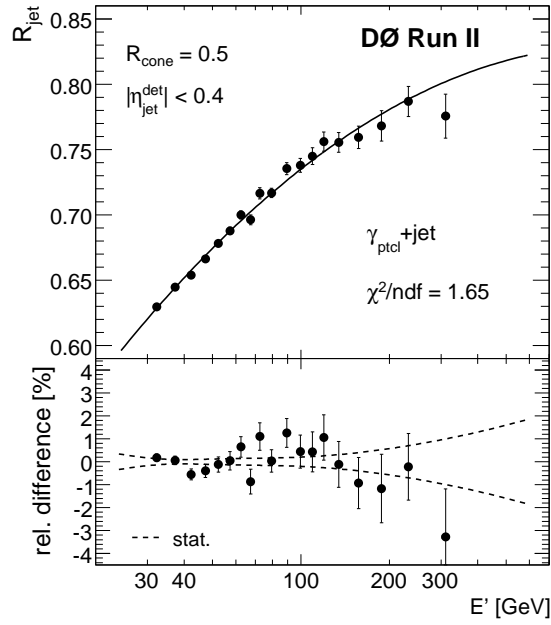
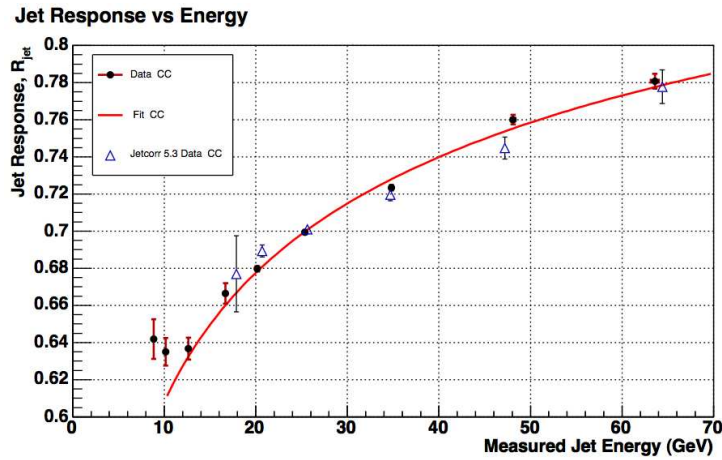

 Figure 3.8: Absolute response correction as function of E' .


Figure 3.9: Measured absolute response correction as function of the measured jet energy. Data points are in black and the fit (red line) is performed with the two first points excluded since they are biased by the jet reconstruction threshold.

measured (raw) jet energy, shown in Figure 3.9. The response is fitted in order to be extrapolated to higher jet energy regimes.

The resultant systematic error for this correction mainly comes from the event selection used to derive the correction and the photon energy uncertainty. More details on the correction derivation as well as consideration of sources of systematic uncertainties can be found in [60].

Relative response correction

The aim of this correction is calibrate forward jets with respect to central jets. Two different samples are used for the correction derivation which relies on the MPF method previously mentioned: γ +jet events are used for the low transverse momentum region since it is statistically limited in the higher energy regime and dijet samples are then used for high transverse momentum.

Corrections obtained by the MPF method are different due to different composition of jets initiated by quarks and gluons in both samples. Gluon initiated jets are typically wider and contain more particles than quark initiated jets. In average, particles are less energetic than in case of jets originating from quark, for equal energy jets. The response is therefore smaller for gluon jets.

The result for the F_η correction is shown in Figure 3.10.

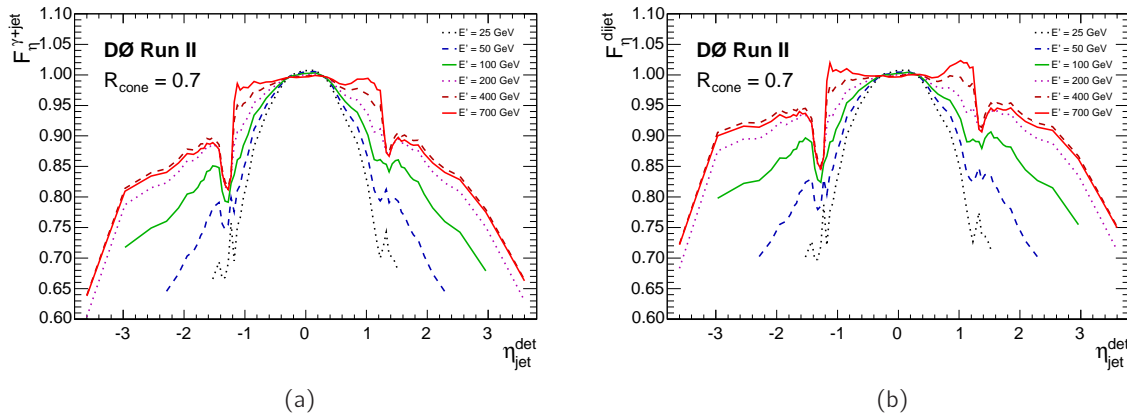


Figure 3.10: Relative response correction obtained from both γ +jet (a) and jet+jet (b) samples for different E' values (colored lines).

Showering correction

The showering correction is determined to take into account energy leaking outside (inside) the jet cone coming from particles inside (outside) the jet cone. This energy leakage is intended to scale the calorimeter jet energy to the particle jet energy. However, detector effects mentioned earlier could modify the energy flow inside the cone.

Using particle jet from γ +jet MC events with exactly one primary vertex reconstructed, it is possible to determine the energy density profile as function of the distance between towers and the jet axis. Contributions from particle jet particles energy and “out-of-cone” energy (*i.e.* energy not clustered in the particle jet) can be determined in MC and data from a template fit method. The correction, shown in Figure 3.11, is then determined as a ratio of the visible energy in a cone of $\Delta R < 0.5$ with respect to the visible energy from all particle jet particles.

Systematic uncertainty associated to the showering correction is driven by the uncertainty on the purity of γ +jet sample, the quality of the template fit and the physics modeling of jet fragmentation (which impact the evolution of the showering effect).

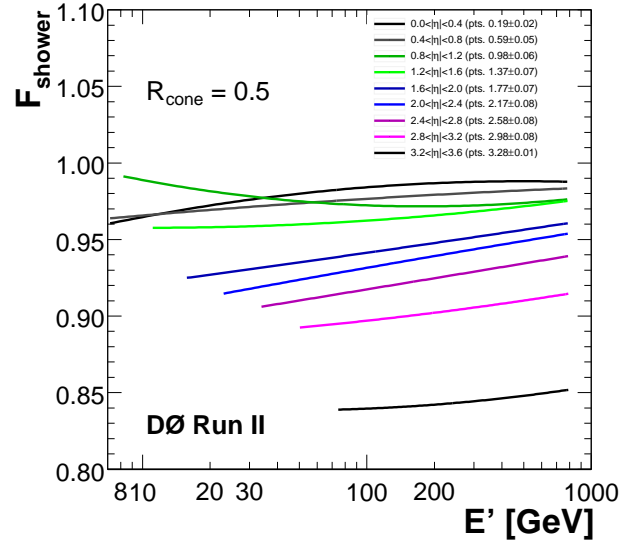


Figure 3.11: Showering correction in data as function of E' for different η_{det} regions (colored lines).

Result of the Jet Energy Scale determination

The derivation of the JES is estimated by relying as much as possible on data, with the attempt of taking into account every potential effect which would bias the jet energy calibration. Separate corrections are estimated and applied for data and simulation. The obtained total uncertainty, after evaluating the correlations between each correction, is estimated to be less than 3% for raw $p_T^{jet} > 10$ GeV, as shown in Figure 3.12.

Semi-muonic jet correction

During JES corrections determination, the assumption is made that in a γ +jet event, objects are balanced in the transverse plane. However, some of the jet energy may escape detection in the case a neutrino is produced during the development of the jet shower. This occurs in about 20% of B mesons decays, through direct decays $b \rightarrow \mu\nu X$ or cascade decays $b \rightarrow cX \rightarrow \mu\nu X'$. Hence, the JES corrections are not sufficient to calibrate the jet energy back to the particle jet energy, since the energy taken by the neutrino is not measured in the calorimeter. An extension of the jet energy scale, called JESMU, is derived to correct the energy of jets containing semi-muonic decays. Since products from the decay of low mass Higgs boson are jets initiated by b quarks, this correction is important for the WH channel.

The correction is derived as following:

1. events with muon inside jets are first selected. Muons are minimum ionizing particle and deposit on average $E_\mu^{loss} = 2$ GeV in the calorimeter. This energy is subtracted from the jet

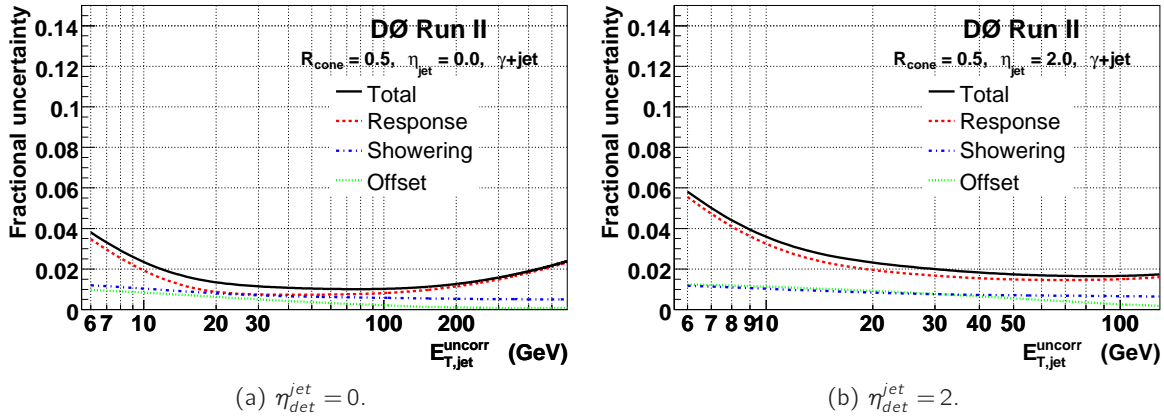


Figure 3.12: Jet energy scale uncertainty in data for jets with (a) $\eta_{det} = 0$ and (b) $\eta_{det} = 2$, as a function of the raw jet transverse energy.

to later avoid double-counting of energy

2. the semi-muonic correction $C = \frac{E'_\mu + E'_\nu}{E_\mu}$ is determined from MC to access the muon and neutrino energy E'_μ and E'_ν from the generator level information, E_μ being the muon energy measured by the DØ detector. C is a function of E_μ and parametrized in different bins of p_T^{rel} , defined as the projection of transverse momentum of the muon on the muon+jet axis (see Figure 3.13). The correction is fitted for each p_T^{rel} bin.
3. the jet energy is corrected as following: $E_{JESMU}^{jet} = E_{JES} - E_\mu^{loss} + CE_\mu$, with E_{JES} being the JES-corrected energy.

A 3.1% improvement is achieved on the invariant mass from $Z \rightarrow b\bar{b}$ MC events with the semi-muonic correction. A detailed overview of the correction is given in [61]. The algorithm presented here is used in the WH analysis. Current developments are ongoing and will be briefly presented in Section 6.1.1.

3.5.4 Jet Shifting, Smearing and Removal

After JES being applied in data and MC, it is observed that the modeling for jet kinematics distributions in simulation isn't perfect. A difference in jet resolution is observed and is higher in simulation than in data. This disagreement is coming from the fact that the jet energy resolution is not corrected by the JES procedure. A standard jet correction is in place in the DØ collaboration in order to improve the resolution modeling of jets, the so-called *JSSR* method for *Jet Smearing, Shifting and Removal*. This method relies on the measurement of the p_T imbalance in γ +jet and $Z \rightarrow e^+e^-$ +jet events. The ΔS variable is perfectly suited for this purpose and is defined as:

$$\Delta S = \frac{p_T^{jet} - p_T^{\gamma/Z}}{p_T^{\gamma/Z}} \quad (3.10)$$

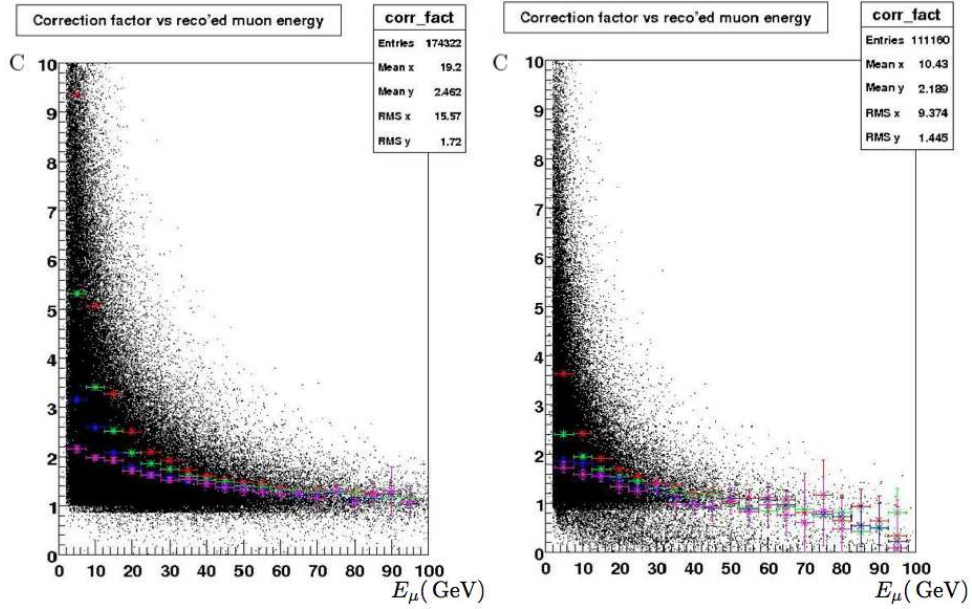


Figure 3.13: Correction factor for jets with semi-muonic (a) direct and (b) cascade decays in $0.8 \leq |\eta_{det}^{jet}| \leq 1.5$. The dependence on p_T^{rel} is visible, where the median value of the correction for each bin along the horizontal axis is represented for $0 < p_T^{rel} \leq 0.5$ (red points), $0.5 < p_T^{rel} \leq 1$ (green points), $1.0 < p_T^{rel} \leq 2.0$ (blue points) and $2.0 < p_T^{rel} \leq 5.0$ (magenta points).

This quantity is computed in several $p_T^{\gamma/Z}$ bins and in 4 $|\eta_{det}^{jet}|$ region:

- $|\eta_{det}^{jet}| < 0.8$ (CC)
- $0.8 < |\eta_{det}^{jet}| < 1.6$ (ICR)
- $1.6 < |\eta_{det}^{jet}| < 2.4$ (EC)
- $2.4 < |\eta_{det}^{jet}| < 3.2$ (VEC)

Figure 3.14 shows an example of the ΔS distribution. One can separate two different regimes according to the $p_T^{\gamma/Z}$ value. For higher $p_T^{\gamma/Z}$ values (above 45 GeV in data, 30 GeV in simulation), the ΔS distribution can be fitted by a gaussian function. If $p_T^{\gamma/Z}$ values are lower, the effect of the jet reconstruction threshold ($P_T > 6$ GeV) is visible and the ΔS distribution needs to be fitted a gaussian function multiplied by an "Error" function to reproduce the jet reconstruction turn-on.

Several parameters will be extracted and are intended to correct the following features in MC and bring them as close as possible as in data:

- **Jet energy resolution** (*Smearing* procedure) : The gaussian width $(\sigma_{\Delta S})_{Gauss} = \frac{\sigma_{p_T}}{p_T}$ is extracted for both data and simulation.
- **Residual difference in the jet energy scale** (*Shifting* procedure) : The gaussian mean $\langle \Delta S \rangle$ is extracted for both data and simulation.

- **Jet reconstruction and identification efficiency** (*Removal* procedure) : "Error" functions are evaluated from the fit of the ΔS distributions in MC.

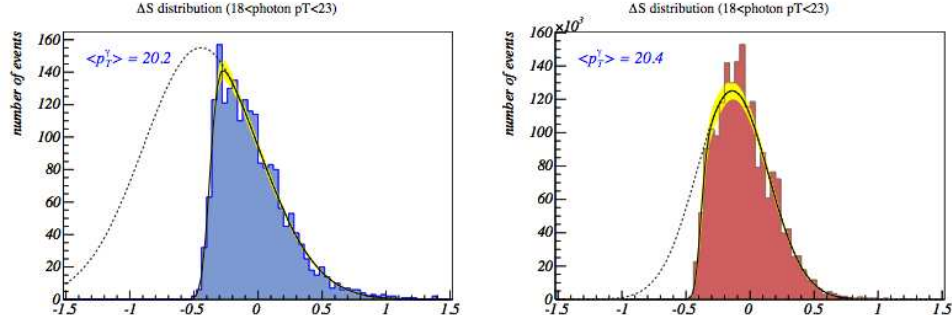


Figure 3.14: ΔS distributions for $18 < p_T^\gamma < 23$ GeV for data (left) and Monte Carlo (right). The fitting function is the product of an error function and a gaussian function (full curves), with statistical error (yellow bands). The dashed curves correspond to the extrapolation of the gaussians in the regions affected by the turn-on.

Smearing

The gaussian width of the ΔS distribution is measured in data and Monte Carlo. A new gaussian function is built with a width computed as :

$$\sigma_{smearing} = \sqrt{\sigma_{\Delta S}^2(Data) - \sigma_{\Delta S}^2(MC)} \quad (3.11)$$

where $\sigma_{\Delta S}^2(Data)$ ($\sigma_{\Delta S}^2(MC)$) is the gaussian width of the ΔS distribution is measured in data (Monte Carlo). The oversmearing factor will be randomly evaluated from this gaussian function. $\sigma_{smearing}$ is shown in Figure 3.15.

Shifting

A similar procedure is used to evaluate the relative energy scale factor by measuring $\langle \Delta S \rangle$:

$$\mathcal{D} = \langle \Delta S_{Data} \rangle - \langle \Delta S_{MC} \rangle, \quad (3.12)$$

which is shown in Figure 3.16.

Removal

The jet removal procedure relies on the determination of the jet "turn-on curves" due to reconstruction threshold effects at low jet transverse momentum. It is parametrized as

$$f(p_T) = 1 + \text{erf} \left(\frac{p_T - \alpha}{\sqrt{2}\beta} \right) \quad (3.13)$$

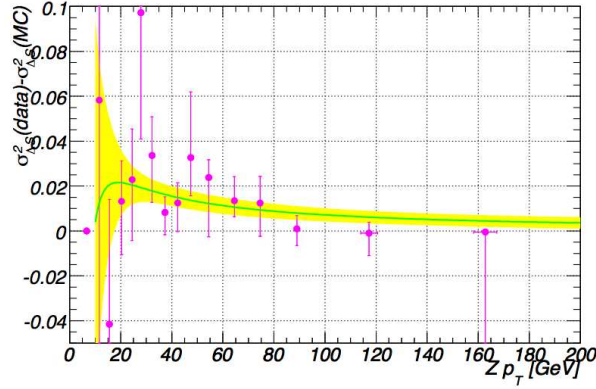


Figure 3.15: $\sigma_{smearing}$ as function p_T^Z for CC jets. The parametrization and the statistical error are represented by the green curve and the yellow band.

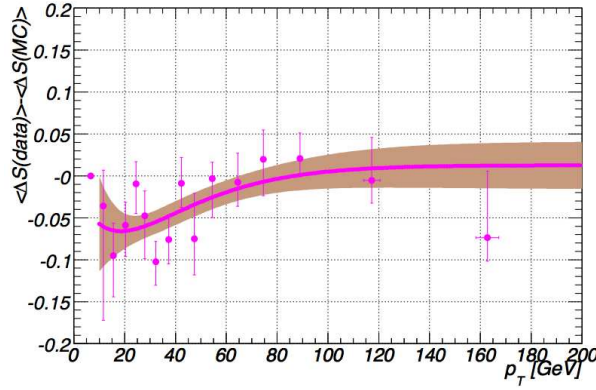


Figure 3.16: The shifting factor \mathcal{D} as function of p_T^Z for CC jets. The parametrization and the associated error from the fit are represented by the green curve and the yellow band.

A scale factor $SF_{removal}$ is computed as function of the jet p_T . A random number r generated uniformly is drawn. If, for a given jet, the condition

$$r < SF_{removal} \quad (3.14)$$

is fulfilled, the jet is removed from the event. The fitted "turn-on" curves are shown in Figure 3.17.

The corrected jet p_T can be expressed as:

$$p_T^{corr} = p_T + p_T^\gamma (shift(p_T^\gamma) + Gauss(0, smear(p_T^\gamma))) \quad (3.15)$$

where p_T^γ is the transverse momentum of the photon, which is assumed to be equal to the jet p_T , $shift(p_T^\gamma)$ is the relative jet energy scale as shown in Figure 3.16 and $Gauss(0, smear(p_T^\gamma))$ being a gaussian function centered at 0 and with a width function of p_T^γ as shown in Figure 3.15.

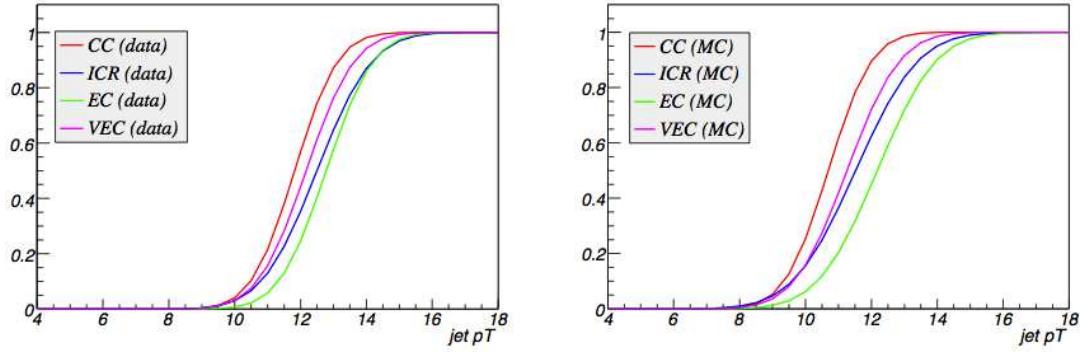


Figure 3.17: Average efficiency turn-on functions represented for different calorimeter regions for data (left) and simulation (right).

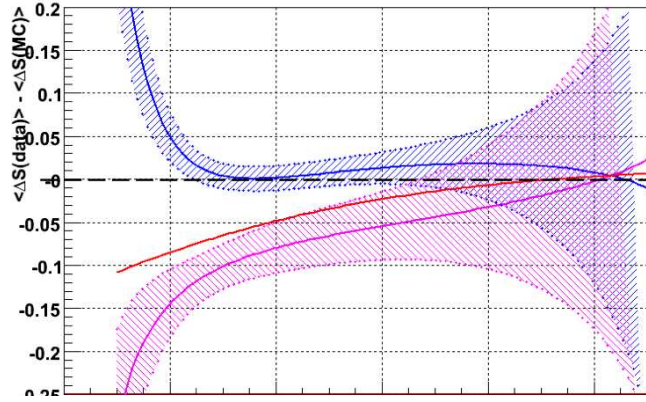


Figure 3.18: Relative Jet Energy Scale correction for quark (blue) and gluon (magenta) jets. The red line shows the nominal shifting, as shown on Figure 3.16.

Since this method is derived on γ +jets and Z +jets samples, the content of quark of gluon and quark jets are different. Since the Jet Energy Scale is different for both types of jets, this also leads to different shifting factor. This difference in the relative jet energy scale is shown Figure 3.18. The shifting factor for quark jets is consistent with 0, whereas gluon jets tend to be shifted downwards, consistently as the nominal shifting factor. Hence, in the WH analysis, jets are treated differently. The following prescription is therefore used:

- Light jets (*i.e.* initiated by a gluon or u , d , or s quarks) do not have the shifting factor applied,
- b -jets are known to be broader than quark jets which makes them gluon-like : the shifting factor is then applied,
- c -jets are treated as light quark jets : no shifting is applied.

More details on the JSSR method can be found in [62], [63], [64] and [65].

3.6 Missing transverse energy

From the principle of momentum conservation, one expect the vectorial sum of all interacting objects in the detector to be equal to zero. However, partons not participate in the hard scatter process have a very high rapidity and the incoming energy of partons is not known on an event by event basis due to parton density functions. Hence, the momentum conservation is only accessible in the transverse plane at the detector level. In practice, the vectorial sum in the transverse plane is rarely zero, due to finite resolution of measured objects in the final state or the presence of weakly interacting particle such as neutrinos. The missing transverse energy is measured as the opposite of the calorimeter cells transverse energy [66]:

$$\vec{E}_T^{raw} = - \sum^{cells} \vec{E}_T \quad (3.16)$$

The energy of cells from the coarse hadronic layers of the calorimeter are not included in the sum since they suffer from relatively high noise.

This relation needs to incorporate changes in the energy scale of calorimetric objects: JES and electron energy scale are propagated back to the measured \vec{E}_T . Moreover, muons can deposit a small amount of energy inside the calorimeter. It is subtracted to avoid energy double-counting since the muon transverse momentum measured by the muon spectrometer is also added to the vectorial sum, which translates the previous equation to:

$$\vec{E}_T = - \sum^{energy\ corrected\ cells} \vec{E}_T - \sum^{muons} \vec{p}_T. \quad (3.17)$$

3.7 b -jet identification

The particularity of a Higgs Boson with $M_H < 135$ GeV is that it decays dominantly in a pair of b quarks. These partons hadronize and form jets. Since B hadrons have a lifetime of the order of a picosecond before decaying, one can exploit their specificities and their experimental signature within the detector to identify them. By rejecting events with jets that are not identified as coming from b quarks, the intrinsic sensitivity of the analysis can be increased. As it will be see in Section 4.9, backgrounds which do not contain such jets (e.g. W +light jets, Multijet, ...) can be significantly reduced while retaining the signal which is composed of pure b jets. The DØ collaboration has developed a powerful tool in order to identify jet originating from b quarks. This algorithm, using the neural network technique, will be detailed as well as its usage in the WH analysis in this section.

3.7.1 b -jets properties

Jets originating from b quarks fragmentation have a slight different signature in the detector. First, the mass of the b quark is 5 GeV, which is the heaviest in this fermion sector (except from the t quark which is heavier than the W boson and decays before hadronization). Also, b hadrons have a longer lifetime (1.518 ± 0.007 ps for B^0 mesons [2]) than hadrons comprised only by light quarks.

Therefore hadrons can fly a relatively long distance (up to a few mm) before decaying. These characteristics are exploited in order to differentiate light jet (originating from u, d, c, s quarks or from gluons) from b -jets by making use of detector informations :

- Displaced secondary vertices can be reconstructed from charged tracks,
- Identification of particles having a high impact parameter (closest distance between the charged track and the primary vertex)
- 11% of b hadrons decays are semi-leptonic, where a muon can be found in a jet.

Several tools have developed in the DØ collaboration in order to identify b jets, by investigating the different features of these objects described previously:

- CSIP (*Counting Signed Impact Parameters*): this algorithm make use of tracks within the $R=0.5$ jet cone to calculate their impact parameter,
- SVT (*Secondary Vertex Tagger*): after reconstructing a secondary vertex from charged tracks, the ΔR between the secondary vertex (SV) and the closest jet is calculated. If $\Delta R(SV, jet) < 0.5$, the jet is tagged by this algorithm,
- JLIP (*Jet Lifetime Probability Tagger*): This algorithm combines tracks-related information and computes a final variable (JLIP-proba) which can be interpreted as the confidence level that all tracks in a jet originate from the (selected) primary interaction point. Jets having a JLIP-proba value close to 1 are more likely to come from b quarks, whereas jets having a value close to 0 are more likely to be light jets,
- SLT (*Soft Lepton Tagger*) If muon is found within a jet, it is then tagged.

Since the algorithms are performing well separately and are exploiting different informations from jets, they can be combined in a single variable by using a neural network (NN). This technique increases the tagging efficiency with respect to a single algorithm and provide the best discrimination between b and light jets. The usage of the NN b -tagger will be described in 3.7.4, after the inputs to this tagger will be described.

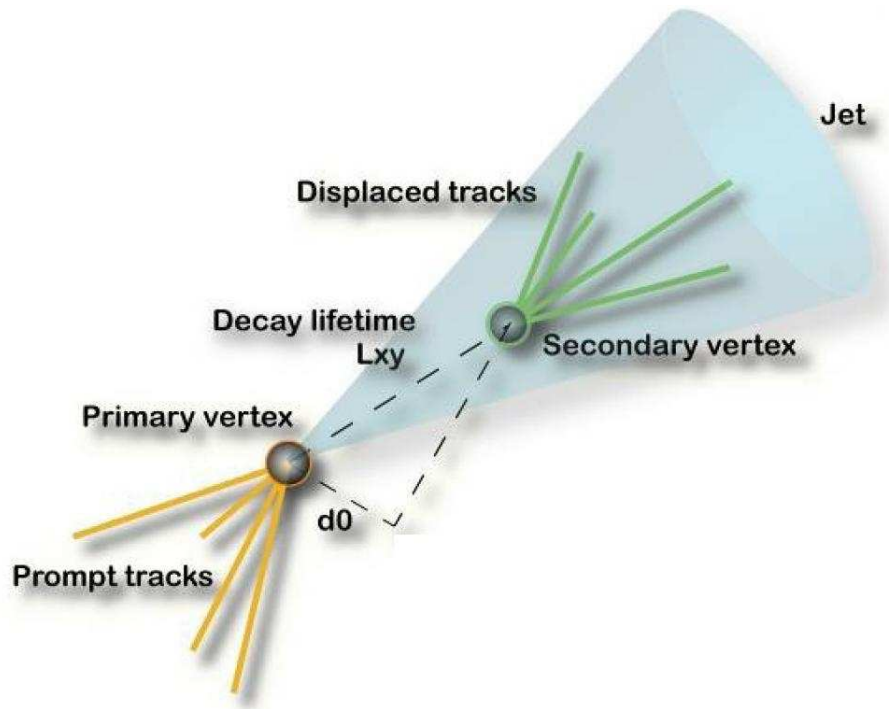


Figure 3.19: As b -hadrons have a long life-time, they can fly for a visible distance, denoted L_{xy} on this figure, before decaying. Where the decay occurs, a secondary vertex can be reconstructed from tracks produced by the decay products that are not attached to the primary vertex. A jet can be b -tagged if a secondary vertex is reconstructed or if the impact parameter d_0 can be measured (distance of closest approach of tracks to the primary vertex (DCA), defined as a straight line parallel to the z axis of the detector with (x,y) position as the one of the reconstructed primary vertex).

3.7.2 *b* jet taggability

Before applying the identification tools that will be discussed next, candidate *b* jets have to satisfy the taggability requirement. This property ensures that a jet can be considered by the several algorithms and potentially be identified as a *b*-jet. The requirement for a jet to be taggable is that it should be within $\Delta R < 0.5$ from a so-called track jet. Track jets are reconstructed starting from tracks having at least one hit in the SMT, a distance to the selected primary vertex less than 2 mm in the transverse plane and less than 4 mm in the *z* direction, and $p_T > 0.5$ GeV. Starting with “seed” tracks having $p_T > 1$ GeV, tracks are clustered within cones of radius $R = 0.5$ using a jet algorithm.

The taggability is defined as the ratio of the number of taggable jets over the number of calorimeter jets. One observe a dependency of the taggability on the jet kinematics (p_T and η). Since taggability also relies on tracking performance, it can be parametrized as function of the *z* component of the primary vertex (see Figure 3.20). The z' variable is used in order to account for correlation between $|z|$ and η and because tracking performance on the amount of material being crossed by charged particles:

$$z' \equiv |z| \cdot \text{sign}(\eta \cdot z) \quad (3.18)$$

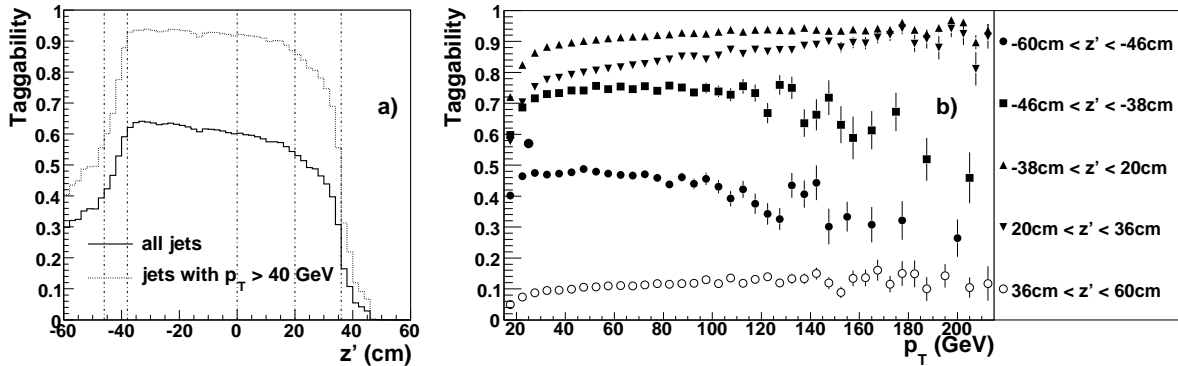


Figure 3.20: (a): Taggability as a function of $z' \equiv |z| \cdot \text{sign}(\eta \cdot z)$. The vertical lines denote the boundaries chosen for the parametrization in p_T and $|\eta|$. (b): taggability as a function of jet p_T , in different bins of z' . The curves for the two central bins are very similar and have been combined.

3.7.3 Individual *b*-jet identification algorithms

Since the NN *b*-tagger makes use of three of these algorithms, CSIP, SVT and JLIP, a particular attention will be drawn on these methods. SLT has a lower efficiency from the fact that the branching ratio of semi-muonic decays in jets is low. Therefore, the addition of this variable as input to the NN does not provide more discrimination. Moreover, this tagger is used independently to measure tagging efficiency, more details can be found in [67].

Counting Signed Impact Parameter (CSIP)

This method [68] does not make use of a reconstructed secondary vertex but relies on the measurement of the impact parameter of tracks with respect to the primary vertex. Before introducing the method, some quantities need to be introduced. The sign of the projection of the impact parameter d_0 onto the jet axis (obtained from the calorimeter measurement) can either be positive or negative. One can define the signed impact parameter IP_s as

$$IP_s = \text{sign} \times d_0 \quad (3.19)$$

From this variable, one can define the impact parameter significance SIP as:

$$SIP = \frac{IP_s}{\sigma(IP_s)} \quad (3.20)$$

where $\sigma(IP_s)$ is obtained from the error on the measurement of IP_s . The distribution of tracks in jets for light and b jets is shown in 3.21. An asymmetric distribution shape is visible for b jets from the fact that d_0 is larger for these jets, and also because $\sigma(IP_s)$ is better measured as d_0 increases.

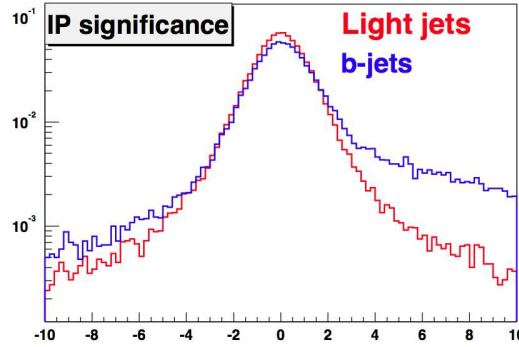


Figure 3.21: Impact parameter significance for light and b jets

The main steps of the CSIP method applied to all “good” tracks (criteria defined in [68] are the following:

1. Jets are selected with $p_T > 15$ GeV
2. Selected tracks must be within a cone of $\Delta R < R_B$ around the jet axis (where $R_B = 0.5$)
3. Tracks are required to have at least 2 hits in the SMT and $p_T^{min} > 1$ GeV
4. The impact parameter d_0 , its error $\sigma(d_0)$ and the sign of the projection of the impact parameter are evaluated for all “good” tracks). Tracks with $d_0 > 2$ mm are rejected. The renormalized significance RS is computed,

$$RS = \frac{SIP}{a} \quad (3.21)$$

where a is a scale factor chosen to be 1.2.

5. Jets are tagged as a *b* jets if it contains at least 2 tracks with $RS > 3$ or at least 3 tracks with $RS > 2$.

Even though this method is designed to retain *b* jets efficiently, some contamination from light jets can occur from various reasons:

- finite SMT resolution
- SMT hits associated to another track than the one they belong to
- multiple scattering in the detector can produce additional tracks
- secondary interactions occurring with the detector material

Another effect can influence the *b*-jet selection efficiency which is from light hadrons having a long life-time like mainly from $K_S \rightarrow \pi^+ \pi^-$, Λ or converted $\gamma \rightarrow e^+ e^-$ that can fake heavy flavor hadrons decays. Therefore, the V^0 rejection procedure is used to reduce light flavor background contributions [67]. After selecting candidate background decays, these tracks are removed.

For a fixed value of $a = 1.2$ which gives the best compromise between *b*-tagging efficiency and mis-tagging efficiency, one found $\epsilon_{btag} = 57\%$ and a mis-tagging rate of $\epsilon_{mistag} = 2.6\%$. The variation of the a value and track p_T threshold provides several operating points. The *b*-tagging efficiency and mis-tagging rate are shown in Figure 3.22 for different operating points.

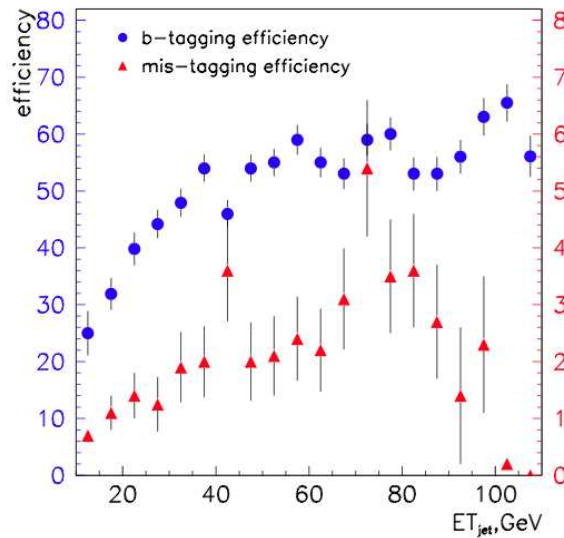


Figure 3.22: *b*-tagging efficiency and mis-tagging rate as function of the jet transverse momentum.

Using this algorithm, one of the input variable to the NN is built. It relies on the tracks' "closeness" to a jet. The closeness is empirically chosen as the difference in the azimuthal angle

between the track and jet directions, $\Delta\varphi$, being less than $\Delta\varphi_0 = 20$ mrad. Four categories of tracks are counted separately:

- tracks with $\mathcal{S}_d/a > 3$, $|\Delta\varphi| > \Delta\varphi_0$ (“3 σ -strong” tracks, their total number to be denoted as N_{3s}),
- tracks with $2 < \mathcal{S}_d/a < 3$, $|\Delta\varphi| > \Delta\varphi_0$ (“2 σ -strong” tracks, N_{2s}),
- tracks with $|\mathcal{S}_d/a| > 3$, $|\Delta\varphi| < \Delta\varphi_0$ (“3 σ -weak” tracks, N_{3w}),
- tracks with $2 < |\mathcal{S}_d/a| < 3$, $|\Delta\varphi| < \Delta\varphi_0$ (“2 σ -weak” tracks, N_{2w}).

From these definitions, one can form the N_{CSIP} variable:

$$N_{CSIP} = 6 \times N_{3s} + 4 \times N_{2s} + 3 \times N_{3w} + 2 \times N_{2w}. \quad (3.22)$$

Secondary Vertex Tagger (SVT)

Since many b hadrons have a significant life-time, this algorithm exploits this characteristic and relies on secondary vertices reconstruction to identify b -jets. It allows, from a number of tracks, to reconstruct a displaced intersection of these tracks, which hopefully is consistent with a heavy flavor hadron which had a decay in flight.

First the identification and selection the primary vertex, a *track-jet* is formed by clustering a jet from tracks (instead of calorimeter cells) within $\Delta R = 0.5$. It has to be matched to a calorimeter jet. The secondary vertex will be then built from these objects. Additional requirements are made for tracks within the track-jet: they should have at least two SMT hits, $p_T > 0.5$ GeV, transverse impact parameter with respect to the primary vertex $|d| > 1.5$ mm, and the z component of the DCA < 4 mm. The V^0 rejection procedure previously introduced is applied.

The SVT algorithm is an iterative process that stops when the extrapolation of at least 2 tracks to a vertex converges, by the minimization of the χ^2 value. The following method is used, more details can be found in [69] and [70]:

1. Tracks within track jets with large transverse impact parameter significance, $|SIP| > 3$, are selected.
2. Vertices are reconstructed from all pairs of tracks using a Kalman vertex fitting technique [39], and are retained if the vertex fit yields a goodness-of-fit $\chi^2 < \chi^2_{\max} = 100$. Tracks are added if their contribution in the fit is not too large ($\Delta\chi^2 < 15$),
3. The resulting vertex is selected if in addition, the angle ζ between the reconstructed momentum of the displaced vertex (computed as the sum of the constituent tracks' momenta) and the direction from the primary to the displaced vertex (in the transverse plane) satisfies $\cos \zeta > 0.9$, and the vertex decay length in the transverse direction $L_{xy} < 2.6$ cm.
4. Many displaced vertex candidates may result, with individual tracks possibly contributing to multiple candidates. Duplicate displaced vertex candidates are removed until no candidates are associated with the identical sets of tracks.

5. Secondary vertices are associated with the nearest calorimeter jets if $\Delta R(vertex, jet) < 0.5$. Here, the vertex direction is computed as the difference of the secondary and primary vertex positions.

Figure 3.23 shows distributions that characterize the properties of *b* quark-jet and light-flavor secondary vertices reconstructed in $t\bar{t}$ events simulated using the ALPGEN [71] event generator: the multiplicity of vertices found in a track jet (N_{vtx}), the number of tracks associated with the vertex (N_{trk}), the mass of the vertex (m_{vtx}) calculated as the invariant mass of all track four-momentum vectors assuming that all particles are pions, and the largest decay length significance, $S_{xy} \equiv L_{xy}/\sigma(L_{xy})$, where $\sigma(L_{xy})$ represents the uncertainty on L_{xy} .

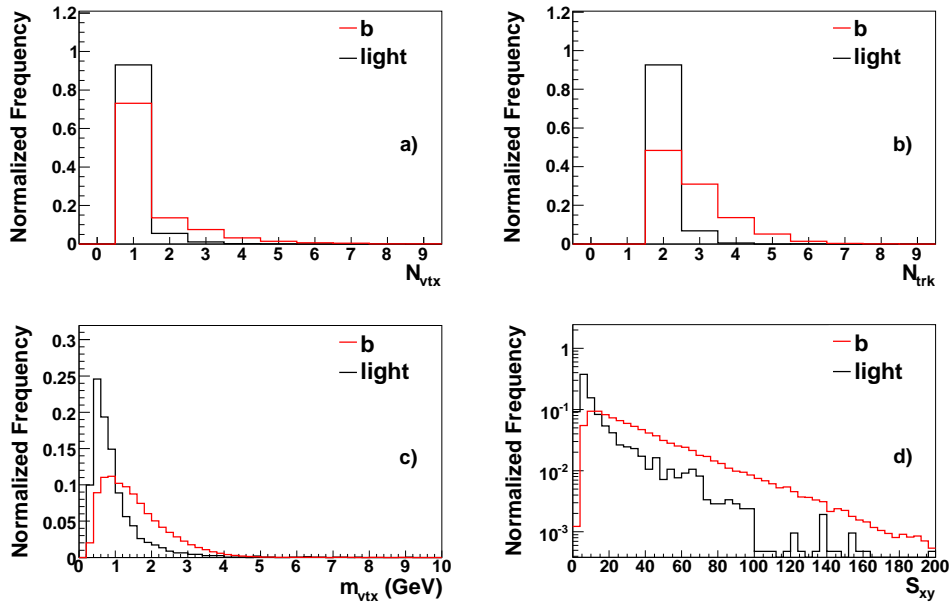


Figure 3.23: Properties of the secondary vertices for tagged *b* quark and light jets in $t\bar{t}$ MC: multiplicity of vertices found in a track jet (a), the number of tracks associated with the vertex (b), the mass of the vertex (c), and the decay length significance (d).

Various operating points have been designed for the SVT algorithm. The *b* tagging efficiency ϵ_{btag} and light jet mis-tag rate ϵ_{mistag} (also called *light jet efficiency* or *fake rate*) vary from $\epsilon_{btag} = 53.6\%$ and $\epsilon_{mistag} = 1.86\%$ for the loosest operating point to $\epsilon_{btag} = 31.8\%$ and $\epsilon_{mistag} = 0.27\%$ for the loosest.

Jet Lifetime Probability (JLIP)

By combining the impact parameter of tracks associated to a calorimeter jet, one can compute a single variable, P_{JLIP} , which can be interpreted as the probability of a jet to originate from the primary interaction point. The smaller P_{JLIP} is, the more likely it is to originate from a *b* quark.

The negative part of the *SIP* (see Figure 3.21), denoted as the impact parameter resolution function $R(SIP)$, can be fitted with the sum of 4 gaussian functions.

$$P_{trk}(SIP) = \frac{\int_{-\infty}^{-|SIP|} R(s) ds}{\int_{-\infty}^0 R(s) ds} \quad (3.23)$$

Then the track probabilities are combined for the N_{trk} tracks with $SIP > 0$ (and separately for $SIP < 0$) to form P_{JLIP} , the probability that a jet originates from the primary vertex, shown on Figure 3.24 is expressed as:

$$P_{JLIP} \equiv P_{JLIP}^+ = \Pi \times \sum_{j=0}^{N_{trk}-1} \frac{(-\ln \Pi)^j}{j!} \quad (3.24)$$

$$\text{with } \Pi = \prod_{i=1}^{N_{trk}} P_{trk}(S_{d,i}^{corr}) .$$

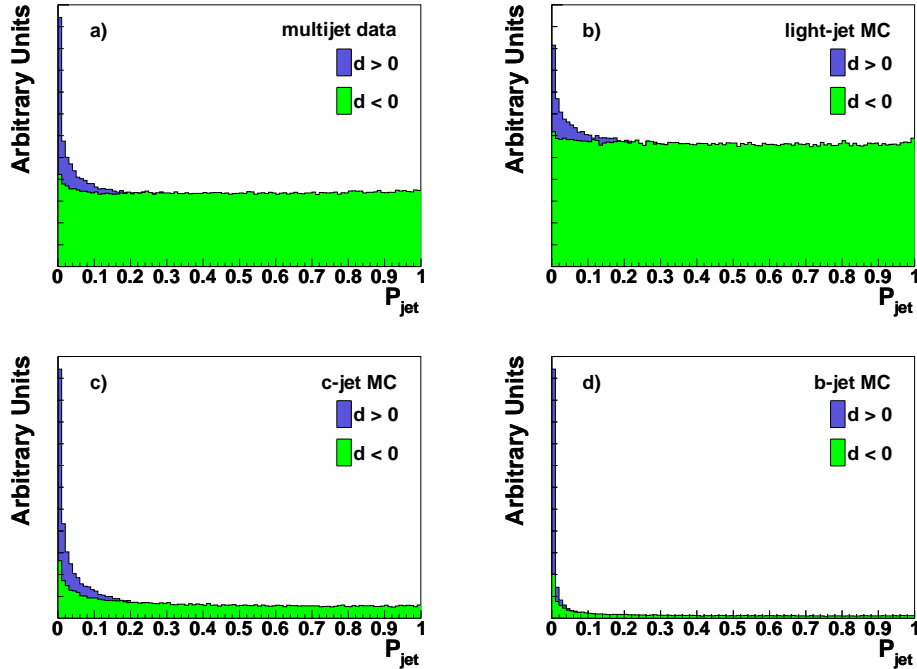


Figure 3.24: Jet probability (P_{JLIP}) distribution in multijet data (a) and QCD MC simulation of light-flavor (b), c quark (c), and b quark (d) jets, for positive (blue histograms) and negative (green histograms) impact parameter values.

The JLIP algorithm has 6 operating points, obtained by varying a cut on the P_{JLIP} value. ε_{btag} and ε_{mistag} are shown in 3.25 for $35 < p_T^{jet} < 55$ GeV and $|\eta| < 1.2$, with the overall statistical and systematic uncertainty.

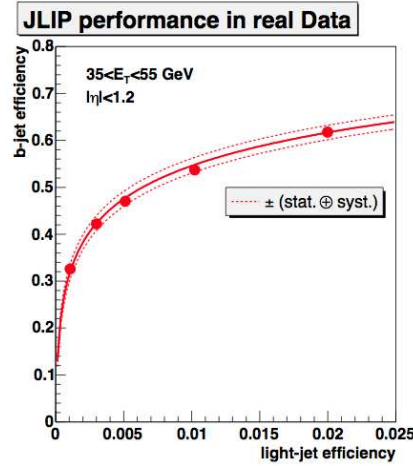


Figure 3.25: ϵ_{btag} and ϵ_{mistag} in data, across the JLIP algorithm operating points (the last point out of bounds), for $35 < p_T^{jet} < 55$ GeV and $|\eta| < 1.2$. The dashed curves represent the overall statistical and systematic uncertainty.

3.7.4 The NN *b*-tagger

The individual *b*-tagging tool have been introduced and have already been used in several analyses to identify *b*-jets [72, 73, 74]. The efficiencies of the algorithms have been discussed briefly in the previous section. A more powerful method can be used, by combining several quantities obtained from the CSIP, SVT and JLIP algorithms into a multivariate analysis technique, the neural network, in order to discriminate light jets from *b* jets.

The principle of neural network is now briefly described, more details on this technique can be found in [75]. This technique, introduced in computing science, is based on a simple model of the human brain, composed of neurons and synapses. Two classes of population are defined, knowing *a priori* their properties (signal or background type of event in population, described by a set of variables). A multi-layered structure is built with nodes (or neurons) and synapses. Nodes are arranged in “layers” as shown in Figure 3.26a, where each node is connected by a synapse to all nodes of the next layer. A neurons performs a mathematical operation by taking a linear combination of input values (either a single value or the output values of neurons from the precedent layer) and getting the output of this value through an “activation function” (either a sigmoid or linear function). This process is schematized in Figure 3.26b. The output node of the neural network is a continuous distribution bounded between 0 and 1, remapping the class of population which are given to be discriminated. The network can be defined by its structure and N_p parameters which are the weights used in each linear combination performed by the nodes. In order to get a discrimination between two types of events, a “learning” phase of the technique is performed (or training), where the difference between the neural network output and the desired value for each event has to be minimized. This iterative learning phase is performed over all events of each population by adjusting the parameters in the N_p -dimensional space, hence the total error is minimized.

The NN *b*-tagger will categorize classes of events, “signal” (*b*-jets) and “background” (light

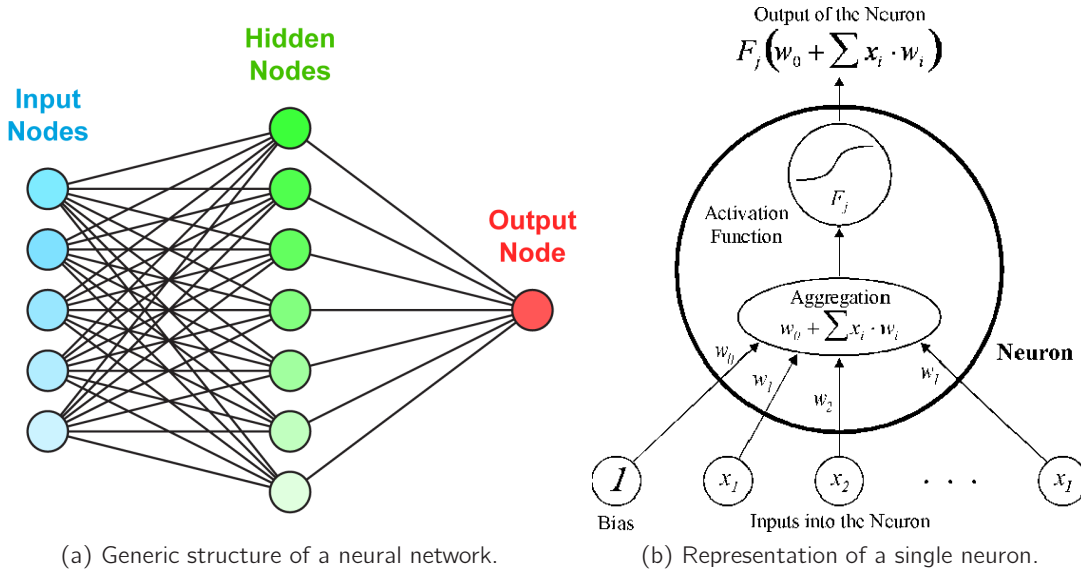


Figure 3.26: Scheme of a neural network structure shown in (a), where each of the circles is a neuron, represented in (b).

jets). The choice of input variables is optimized in a way that the discrimination between the two populations has to be good and the correlations between them has to be as small as possible. The variables used as inputs to the NN b -tagger are:

SVT S_{xy} : the decay length significance (the decay length in the transverse plane divided by its uncertainty) of the secondary vertex with respect to the primary vertex.

SVT χ^2_{dof} : the χ^2 per degree of freedom of the secondary vertex fit.

SVT N_{trk} : the number of tracks used to reconstruct the secondary vertex.

SVT m_{vtx} : the mass of the secondary vertex, calculated as the invariant mass of all track four-momentum vectors, assuming all particles were pions.

SVT N_{vtx} : the number of secondary vertices reconstructed in the jet.

SVT ΔR : the distance in (η, ϕ) space between the jet axis and the difference between the secondary and primary vertex positions.

JLIP P_{JLIP} : the “jet lifetime probability”.

JLIP P_{RedJLIP} : P_{JLIP} re-calculated with the track with the highest significance removed from the calculation.

CSIP N_{CSIP} : a combined variable based on the number of tracks with an impact parameter significance greater than an optimized value (see Equation 3.22).

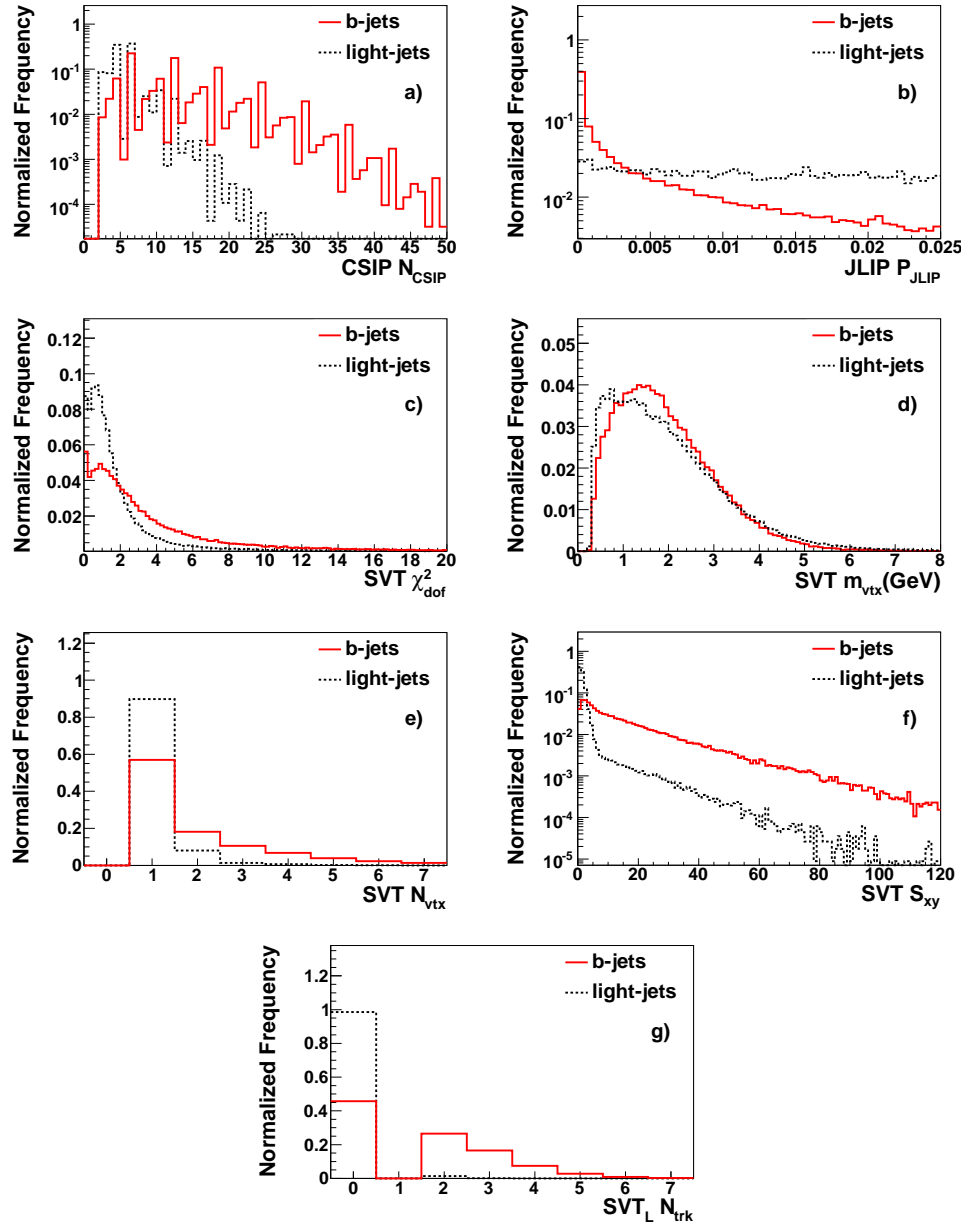


Figure 3.27: The NN variables $CSIP N_{CSIP}$ (a), $JLIP P_{JLIP}$ (b), $SVT \chi^2_{dof}$ (c), m_{vtx} (d), N_{vtx} (e), S_{xy} (f), and $SVT N_{trk}$ (g) for QCD $b\bar{b}$ MC (continuous lines), and light jet QCD MC (dashed lines). All histograms are normalized to unit area.

The distributions of some of the input variables can be found in Figure 3.27.

The training of the NN is performed with the ROOT package, using the TMultiLayerPerceptron class. The structure of the NN is designed in the following way:

- the NN is built with 7 input nodes, 24 hidden nodes in the first layer and 1 hidden node in the second layer and one output node,
- the number of training epochs is set to 400,
- the BFGS training algorithm is used [76, 77, 78, 79].

This choice allows to obtain the optimal rejection between b -jets and light jets. By construction, the output of the NN will assign values close to 0 to light jets and values close to 1 to b jets. This discriminating variable is shown in Figure 3.28.

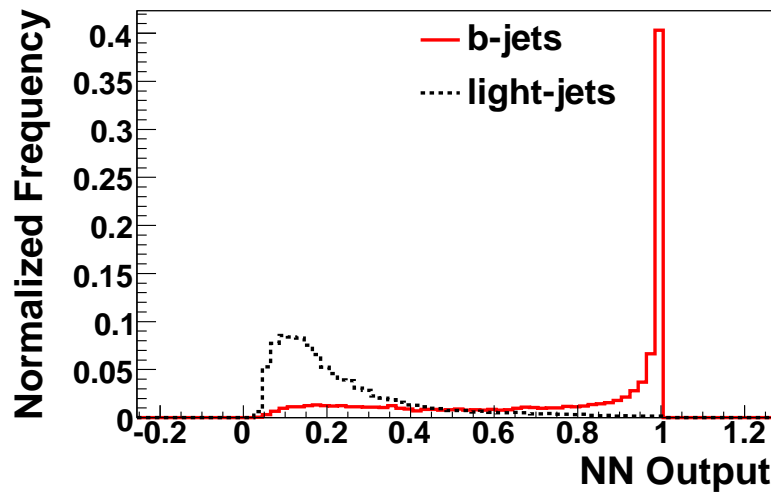


Figure 3.28: The NN output for light-flavor (dashed line) and b quark (continuous line) jets (with $p_T > 15$ GeV and $|\eta| < 2.5$) in simulated QCD events. Both distributions are normalized to unity.

More details on the method to choose the optimal training parameters, variables and NN structure can be found in [75].

The average b -tagging efficiency of JLIP is 33% for a light jet efficiency of 0.15% for its tighter operating point. By combining several observables from the different algorithms, one can observe that the NN-tagger performs much better than an individual b -tagging tool, as shown in Figure 3.29, where the fake rate is reduced by a factor of between two or three for fixed signal efficiencies.

As for the CSIP, SVT and JLIP taggers, several operating points are designed with their associated efficiency/fake rate by varying a cut value on the NN output. Twelve operating points are provided, where the tightest one will give the purest sample of b jets. A summary of all operating points is given in Table 3.4, with the associated b jet efficiencies, light jet mis-identification rate and the systematic errors assigned to a given operating point. The quoted numbers are for jets in the central region of the calorimeter ($|\eta| < 1.2$).

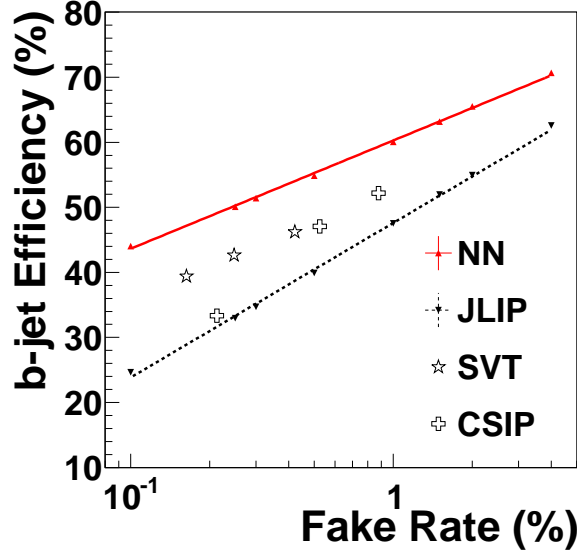


Figure 3.29: Performance of the NN (up triangles, continuous line), JLIP (down triangles, dotted line), CSIP (hollow crosses) and SVT (hollow stars) taggers computed on simulated $Z \rightarrow b\bar{b}$ and $Z \rightarrow b\bar{b}$ jet samples. The performance for each of the operating points of the CSIP and SVT taggers are shown instead of a continuous line due to the discrete nature of the algorithms.

Operating point	L6	L5	L4	L3	L2	Loose
NN output	> 0.1	> 0.15	> 0.2	> 0.25	> 0.325	> 0.45
<i>b</i> jet efficiency	77%	74.9%	72.2%	69.6%	65.9%	60.8%
Fake rate	11.1%	8.16%	6.06%	4.66%	3.28%	2.02%
Systematic error	1.67%	1.58%	1.47%	1.31%	1.29%	1.37%

Operating point	oldLoose	Medium	Tight	VeryTight	UltraTight	MegaTight
NN output	> 0.5	> 0.65	> 0.775	> 0.85	> 0.9	> 0.925
<i>b</i> jet efficiency	59.3%	53.7%	47.6%	43.3%	39.5%	37.1%
Fake rate	1.68%	0.958%	0.546%	0.343%	0.226%	0.169%
Systematic error	1.45%	1.34%	1.52%	1.51%	1.33%	1.43%

Table 3.4: The different NN-tagger operating points, with their associated NN output cut value, *b* jet efficiency, light jet mis-identification, and systematic error for jets in the central region of the calorimeter ($|\eta| < 1.2$).

4

Selection and physics processes modeling in the WH analysis

4.1 Overview

As seen in 1.5, the WH production is one the most sensitive channels for the search of a low-mass Higgs boson by looking for its signature in the $WH \rightarrow \ell\nu b\bar{b}$ mode. The leptonic branching ratio $W \rightarrow \ell\nu$ is 3 times lower than $W \rightarrow q\bar{q}'$ which is dominant, but backgrounds from QCD processes are significantly lowered. This allows to deal with events that have a much cleaner signature in the DØ detector. Therefore, the intrinsic power of this channel comes from a reasonable rate of background and a relatively high cross-section for the signal compared to other production modes. The high branching ratio from the $b\bar{b}$ decay mode in the mass range that is analyzed also plays a major role.

The event topology corresponding to the WH channel is one isolated high transverse momentum lepton (either an electron or a muon) and missing transverse energy, which are the experimental signature of the leptonic decay of a W boson. Since the Higgs boson has a resonant decay in a $b\bar{b}$ pair, events must also contain two or three jets, for which b -tagging will be applied. The reconstruction of the invariant mass of the two jets is computed in order to reconstruct the candidate Higgs boson mass. The channel with three jets in the final state allows to select events having an additional gluon radiation, not selected in the dominant channel with exactly two jets in the final state.

This analysis has first been realized in 2004 and has been published, using 174 pb^{-1} of Tevatron RunII data [80]. Another result was published in 2008 using the full RunIIa dataset which represents 1 fb^{-1} of data [81]. The result which is presented in this document has been published in 2011 with 5.4 fb^{-1} [1] and a longer version of the article is in preparation. Updated results have been presented at least once a year at major conferences.

This chapter will describe the strategy that is used for this analysis. In Section 4.4 will be detailed the criteria used to select events corresponding to the WH associated production, in the $\ell\nu b\bar{b}$ final state. Two important points in this analysis are to have the largest acceptance for the WH signal, and to reject background as much as possible. Since the detector simulation described in 2.2.8 is not perfect as well as generators used have limitations in describing underlying processes, some corrections need to be applied to the simulation (Section 4.5). The multijet background which has an instrumental origin, cannot be properly modeled by Monte Carlo generators that are currently available. Therefore, a data-driven technique is used to model this background component. This will be described in Section 4.6.1. Finally, in order to identify jets initiated by b -quarks, the NN b -tagger described in 3.7 is used in the analysis to select jets initiated by b quarks. Its usage for data and Monte Carlo samples will be described in 4.9.

An efficient data analysis framework for this search has been developed and maintained. The data analysis shown in this document mainly relies on the usage of this software which was used at many steps of the analysis:

- data to simulation comparisons to verify the data is accurately described by simulation,
- derivation of corrections to account for residual simulation mismodeling (Section 4.5),
- training of multivariate analysis techniques and production of final histograms for results extraction (which are later detailed in Section 5).

All figures shown in this document related to the WH analysis, published in 2011 [1], have been produced using this software. The color legend for the different components are show on Figure 4.1, except when the legend is displayed on the figures.



Figure 4.1: Color legend for data to simulation comparison. Backgrounds denoted as W_{jj} , W_{cc} and W_{bb} respectively includes Z_{jj} , Z_{cc} and Z_{bb} processes. More details on the simulated background samples used in the analysis are given in Section 4.3.2. The multijet background is denoted as QCD in this legend.

4.2 Foreword on the analysis work flow

The ROOT software is used in the WH analysis to analyze data, build histograms to validate background modeling and extract limits. Each data and Monte Carlo sample is grouped in *requestIDs*, after reconstruction and detector simulation for Monte Carlo has been performed. These samples are processed within the $CAFe^1$ framework which is used to run the selection step of the analysis (loose selection of reconstructed objects, application of corrections. . .). The average run time to produce these skimmed files is about 24 hours. The output format is a light-weight ROOT tree for each sample with all relevant information to run the final selection code (final objects selection, b -tagging application, application of multivariate techniques and creation of input histograms for the final result extraction). This step is run in approximately an hour.

4.3 Data and Monte Carlo used in the WH analysis

The data and Monte Carlo samples used in the WH analysis are obtained from skimmed simulated and recorded data samples produced in a common format for the DØ collaboration. The EMinclusive and MUinclusive skims are respectively used for the electron and muon channels. They are obtained by respectively requiring at least one reconstructed electron or muon, by a logical OR between different subskims having different selection criteria.

Two main data-taking period are analyzed: the RunIIa dataset which was taken until the 2006 shutdown and the RunIIb dataset (June 2006 - June 2009). The RunIIb data used here are divided into two main periods: the RunIIb1 dataset, taken from June 2006 to August 2007, and the RunIIb2 dataset, taken from October 2007 to June 2009. These two periods are characterized by a change in instantaneous luminosity. Important detector and trigger updates have been made between the RunIIa and RunIIb periods [28, 82]. They are treated separately to properly account for the changed detector response by using different sets of Monte Carlo samples and associated correction factors.

4.3.1 Data samples

The RunIIa dataset has an integrated luminosity of $1.04 (1.08) \text{ fb}^{-1}$ for the electron (muon) channel. The RunIIb1 dataset has $1.20 (1.20) \text{ fb}^{-1}$, and the RunIIb2 dataset has $3.08 (3.08) \text{ fb}^{-1}$ after requiring data quality and removing bad luminosity blocks from the luminosity system. The total integrated luminosity is 5.32 fb^{-1} and 5.36 fb^{-1} for the electron and muon channel, respectively. Most of the distributions shown here are from the RunIIb dataset (2 jets channels) which provides the highest sensitivity in the analysis, distributions from the RunIIa dataset are given in Appendix C.

4.3.2 Monte Carlo samples and generators

This section details the Monte Carlo samples used to simulate signal and background contributions, used to compare with data. All Standard Model processes sharing a similar topology as the WH

¹which stands for Common Analysis Format Environment.

signal and susceptible to enter the analysis after all selection cuts applied are considered and generated by Monte Carlo, except for the multijet background which is determined from data as later explained in 4.6. Processes containing two leptons in the final state, such as $Zjj \rightarrow \ell^+\ell^-jj$ or purely leptonic decay modes of $t\bar{t}$, are included as well in case one of the leptons is not reconstructed or is outside of the detector acceptance. Hence these event would pass the selection cuts. The Higgs mass being a free parameter of the analysis, the WH signal is simulated for different mass hypotheses.

The following processes are used:

- W/Z +light jets. The W or Z boson decays leptonically to one or two charged leptons in association with jets originated from light partons, such as u , d , s quarks or gluons. This is the dominant background before applying b -tagging.
- W/Z +heavy jets, from c or b quarks. After applying b -tagging, it is the dominant background.
- $t\bar{t}$ production. A pair of top quarks are produced from strong interaction. The t quark decays almost 100% of the time to a b quark and a W boson. The W boson can successively decays either leptonically ($W \rightarrow \ell\nu$) or hadronically ($W \rightarrow q\bar{q}'$). Only two configurations are used in the analysis: the semileptonic and dilepton decays, respectively when one W decaying leptonically and the other hadronically or when both W bosons decay leptonically. The process with both W bosons decaying hadronically, leading to events with at least 6 jets in the final state is not included in the analysis due to its small contribution in the 2 or 3 jets channels. The mass of the generated t quark is 172.5 GeV.
- Single top production, when the t quark is produced via electroweak interaction, either from decays of W bosons to a t and b quark (s-channel), or in the t-channel: a gluon splits into a $b\bar{b}$ pair, a quark from the other incoming (anti-)proton radiates a W boson which will interact with one of the b quarks to produce a t quark. The mass of the generated t quark is 172.5 GeV.
- Diboson production. A pair of weak bosons, WW , WZ or ZZ are produced. These processes lead to one or two leptons (from one of the bosons) and two jets (coming from the decay of the other boson). The $WZ \rightarrow \ell\nu b\bar{b}$ background has the most similar event characteristics to the signal: a similar topology and a resonant production of a pair of b jets with a dijet invariant mass which may be close to the Higgs mass. It is denoted as irreducible background.
- WH signal. Several mass points are generated, from 100 to 150 GeV, by steps of 5 GeV. Another signal contribution is from $ZH \rightarrow \ell^+\ell^-b\bar{b}$ when one of the lepton is out of the detector acceptance. The Higgs boson decay branching ratio predictions are calculated with HDECAY [83].

Feynman diagrams for samples used are displayed in Figure 4.3.2.

Different Monte Carlo generators are used to produce the samples described above. A brief introduction to these programs is given now.

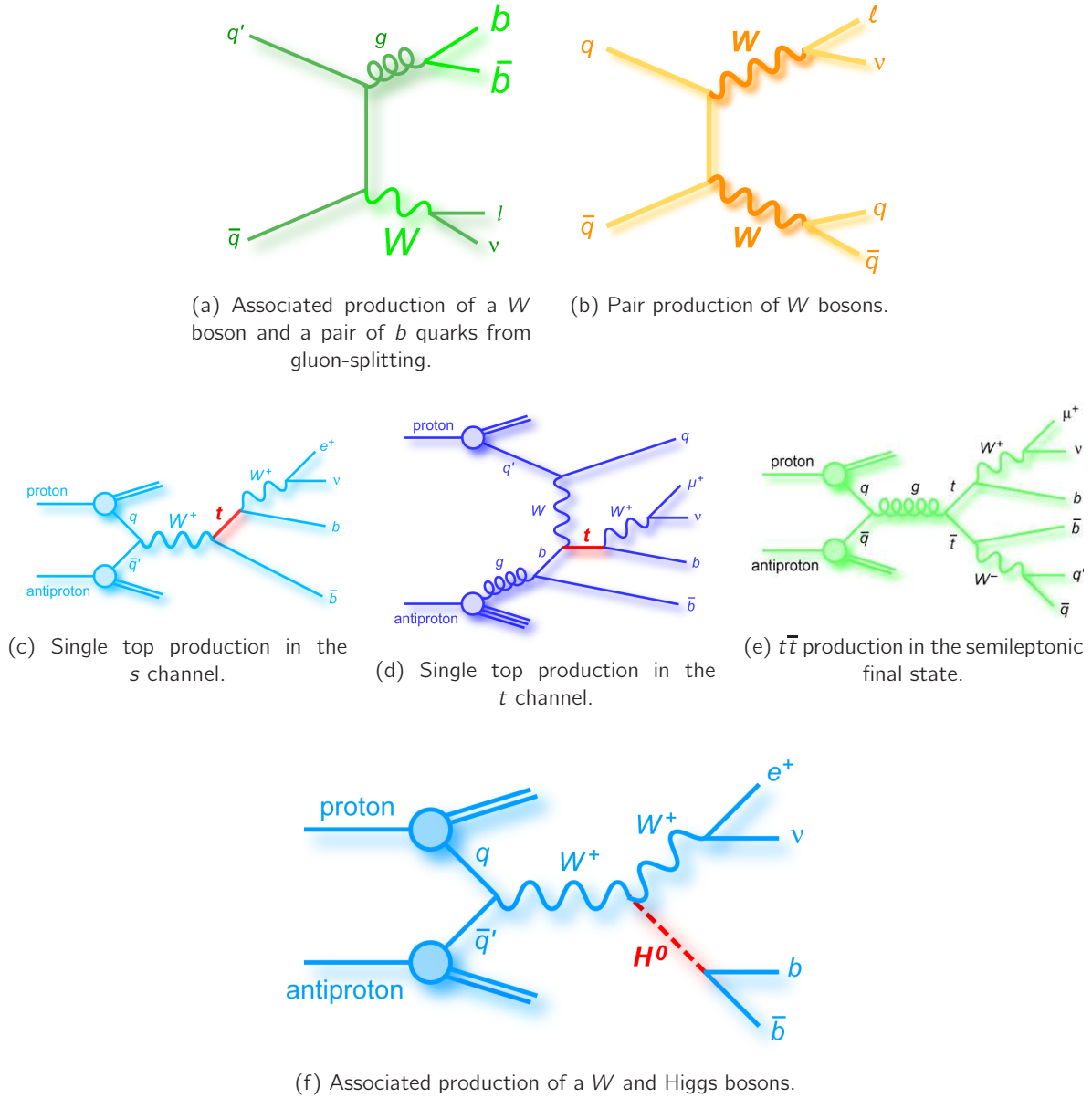


Figure 4.2: Example Feynman diagrams of Monte Carlo processes used in the analysis.

PYTHIA

PYTHIA is Monte Carlo event generator calculating Feynman diagrams at the tree-level (or Leading Order perturbation theory (LO)), higher order loop calculations not being performed [84]. Process cross sections are calculated using the exact matrix element calculation. PYTHIA uses the parton shower model to simulate higher multiplicity final states, with a parton emission seen as a probabilistic process which occurs with a given probability. The hadronization process, when colour-

less particles are produced from partonic final state, is modeled with a phenomenological “String Model” [85]. The colored field (analogous to the electric field in electromagnetism) is described by a string attached at both extremities to $q\bar{q}$ pair. If the string carries enough potential energy, the string breaks to another $q\bar{q}$ pair created from the vacuum. The process ends until quarks hadronize to mesons when the energy of the string is not sufficient enough to create new pairs of quarks. PYTHIA is interfaced with the leading order parton distribution functions CTEQ6L [86, 87].

ALPGEN

ALPGEN is a tree-level matrix element (ME) calculator for a fixed number of partons in the final-state ($2 \rightarrow N$ processes). The evaluation of matrix elements gives a better description for processes with high jet multiplicities. However, the showering and hadronization aren't handled by ALPGEN. Therefore, after generating partons in the final state, an additional step needs to be performed.

MLM matching prescription Since ALPGEN is a matrix element generator and fails to properly model the showering and hadronization, it is interfaced with PYTHIA in order to deal with this remaining aspect of the event generation. When this combination of generators is applied, higher jet multiplicities samples can be artificially produced if a soft radiation from pythia is created. Such migration can lead to a double counting of events, therefore matching criteria are introduced, such as the ΔR between partons and the relative momentum between partons. A more detailed description is given in [88].

COMPHEP

CompHEP is a Matrix Element generator approximating NLO calculations [89]. It can compute cross sections and distributions with many particles (up to 4-6) in the final state. It takes into account the spin correlation between W bosons and t quarks in single top production make its use appreciated.

Cross sections and event tables

The number of generated events (passing the data quality criteria) and their corresponding cross sections are given in Tables 4.2–4.7.

Some cross sections for processes used in the analysis are calculated at the Leading Order (LO) in perturbation theory. In order to correct the cross section values to match the higher orders predictions¹ (NLO) calculated from the matrix element MCFM generator, a factor

$$K = \frac{\sigma^{NLO}}{\sigma^{LO}}$$

¹higher order predictions refers to the next-to-leading matrix element calculation, as well as logarithm factors arising from the summation of soft radiations not taken into account by generators since they are below the generation threshold. This change in cross section is introduced as a series expansion where only the leading logarithm term (LL) or next-to-leading logarithm term (NLL) are taken into account.

4.3. DATA AND MONTE CARLO USED IN THE WH ANALYSIS

is applied as a multiplicative term to the cross section given from ALPGEN. The cross sections of WH and ZH processes come from [90].

The corrective K factors for W/Z +jets processes are given in Table 4.1. The K factor for heavy flavor jets is slightly different because of non-negligible mass terms for b and c quarks.

The factors for other processes ($t\bar{t}$, single top, WW , WZ and ZZ) are given along with their theoretical cross sections in Tables 4.2–4.7.

Sample	K factor
W +light jets	1.3
Z +light jets	1.3
$W + c\bar{c}$	1.3×1.47
$W + b\bar{b}$	1.3×1.47
$Z + c\bar{c}$	1.3×1.67
$Z + b\bar{b}$	1.3×1.52

Table 4.1: K factors for W/Z +jets processes.

The ALPGEN samples have been produced in five exclusive bins of “light” (i.e. gluons or u, d, s quarks) parton multiplicity except for the “highest” multiplicity bin obtained in an inclusive way, i.e. it includes higher multiplicities as well. ALPGEN uses the matching prescription as described in [91].

All ALPGEN W/Z +jets, $W+b\bar{b}$ and $W+c\bar{c}$ have undergone a process of heavy-flavor (HF) skimming: events with additional heavy-flavor partons generated by PYTHIA are removed so as not to double-count them with those produced by ALPGEN. Specifically, W +light jets samples are skimmed to remove $W+c\bar{c}$ and $W+b\bar{b}$ events. $W+c\bar{c}$ samples are not skimmed, and $W+b\bar{b}$ samples are skimmed to remove $W+c\bar{c}$ events. W +jets and $t\bar{t}$ samples include all three lepton flavors. Other samples, such as Z +jets and single-top, have been produced in bins of lepton flavor; they are given in Table 4.2–4.7.

Since the number of generated MC events is different than the expectation from integrated luminosity and cross section, events are given individual weights $w = \frac{\sigma \times \mathcal{L}_{int}}{N_{tot}}$, where σ is the cross-section of the process corrected by its K factor if needed, \mathcal{L}_{int} is the integrated luminosity used in the analysis and N_{tot} is the total number of generated events in the MC sample.

All simulated events have been processed through the $D\emptyset$ detector simulation $d\emptyset gstar$ [92] based on the detailed detector material simulation package GEANT [93], the electronics simulation $d\emptyset sim$ and the reconstruction software $d\emptyset reco$. Because of major upgrades, i.e installing Layer0, between RunIIa and RunIIb, different sets of Monte Carlo samples are used in this analysis.

Process		# events (RunIIa)	# events (RunIIb)	$\sigma(\times\text{BR})[\text{pb}]$
$WH \rightarrow \ell\nu + b\bar{b}$	$m_H = 100 \text{ GeV}$	194715	320322	0.0251
	$m_H = 105 \text{ GeV}$	193580	293813	0.0209
	$m_H = 110 \text{ GeV}$	199080	316140	0.0173
	$m_H = 115 \text{ GeV}$	196937	279087	0.0141
	$m_H = 120 \text{ GeV}$	194767	321634	0.0112
	$m_H = 125 \text{ GeV}$	193882	278550	0.0087
	$m_H = 130 \text{ GeV}$	193045	553407	0.0065
	$m_H = 135 \text{ GeV}$	193795	460095	0.0047
	$m_H = 140 \text{ GeV}$	197115	495718	0.0032
	$m_H = 145 \text{ GeV}$	195850	446339	0.0021
	$m_H = 150 \text{ GeV}$	194676	320201	0.0013
$ZH \rightarrow \ell\ell + b\bar{b}$	$m_H = 100 \text{ GeV}$	394432	319375	0.0046
	$m_H = 105 \text{ GeV}$	403338	279205	0.00384
	$m_H = 110 \text{ GeV}$	396185	320038	0.00320
	$m_H = 115 \text{ GeV}$	400115	279468	0.00263
	$m_H = 120 \text{ GeV}$	404973	322296	0.00212
	$m_H = 125 \text{ GeV}$	388415	279466	0.00166
	$m_H = 130 \text{ GeV}$	385764	321965	0.00125
	$m_H = 135 \text{ GeV}$	387729	279647	0.00091
	$m_H = 140 \text{ GeV}$	376238	320005	0.000637
	$m_H = 145 \text{ GeV}$	387119	279310	0.000410
	$m_H = 150 \text{ GeV}$	387219	316756	0.000249

Table 4.2: List of simulated signal processes generated with PYTHIA, along with the generator that was used for production, number of events and cross section times branching ratio, for one lepton flavor.

Process	# events (RunIIa)	# events (RunIIb)	$\sigma \times \text{BR}[\text{pb}]$
WW inclusive	1905k	709.9k	1.03×11.6
WZ inclusive	1059k	632.3k	1.06×3.25
ZZ inclusive	590.6k	540.3k	1.33

Table 4.3: List of simulated diboson processes generated with PYTHIA, along with the generator that was used for production, number of events and cross section times branching ratio.

4.3. DATA AND MONTE CARLO USED IN THE WH ANALYSIS

Process	# events (RunIIa)	# events (RunIIb)	$K \times (\sigma \times \text{BR})[\text{pb}]$
$t\bar{t} \rightarrow b\bar{b} + \ell^+ \nu \ell'^- \bar{\nu}_{\ell'}$	+ 0 light parton	1516107	749642
	+ 1 light parton	492647	452177
	+ 2 light partons	288992	281453
$t\bar{t} \rightarrow b\bar{b} + 2j + \ell\nu$	+ 0 light parton	771271	777068
	+ 1 light parton	492647	457782
	+ 2 light partons	288992	321166
Single-top s -channel ($t\bar{b} \rightarrow e\nu b\bar{b}$)	290262	247517	0.99×0.112
Single-top s -channel ($t\bar{b} \rightarrow \mu\nu b\bar{b}$)	287994	225286	0.99×0.11
Single-top s -channel ($t\bar{b} \rightarrow \tau\nu b\bar{b}$)	287991	248722	0.99×0.117
Single-top t -channel ($tq\bar{b} \rightarrow e\nu bq\bar{b}$)	290262	272573	0.99×0.243
Single-top t -channel ($tq\bar{b} \rightarrow \mu\nu bq\bar{b}$)	287994	273354	0.99×0.239
Single-top t -channel ($tq\bar{b} \rightarrow \tau\nu bq\bar{b}$)	289106	246552	0.99×0.254

Table 4.4: List of simulated $t\bar{t}$ and single-top processes, along with the generator that was used for production, number of events and K factor times cross section times branching ratio, where $\ell = e, \mu$ or τ . K factors scale $\sigma(p\bar{p} \rightarrow t\bar{t} \rightarrow X)$ to 7.3 pb, $\sigma(p\bar{p} \rightarrow t\bar{b} \rightarrow \ell\nu b\bar{b})$ to 0.34 pb, $\sigma(p\bar{p} \rightarrow tq\bar{b} \rightarrow \ell\nu bq\bar{b})$ to 0.73 pb. $t\bar{t}$ samples are generated using ALPGEN + PYTHIA generators and single-top samples are generated using CompHEP + PYTHIA.

Process	# events (RunIIa)	# events (RunIIb)	$\sigma(\times \text{BR})[\text{pb}]$
$W \rightarrow \ell\nu$	+ 0 light parton	12.5M	46.4M
	+ 1 light parton	18.9M	19.9M
	+ 2 light partons	13.3M	18.1M
	+ 3 light partons	3.5M	3.75M
	+ 4 light partons	2.5M	2.6M
	+ 5 light partons	781k	2.0M
$Wb\bar{b} \rightarrow \ell\nu b\bar{b}$	+ 0 light parton	1.4M	1.1M
	+ 1 light parton	667k	782k
	+ 2 light partons	249k	524k
	+ 3 light partons	377k	413k
$Wc\bar{c} \rightarrow \ell\nu c\bar{c}$	+ 0 light parton	1.2M	934k
	+ 1 light parton	740k	739k
	+ 2 light partons	342k	554k
	+ 3 light partons	446k	470k

Table 4.5: List of simulated W +jets processes using ALPGEN + PYTHIA generators, along with the generator that was used for production, number of events and cross section times branching ratio (the light partons are requested to have $p_T > 8$ GeV and $|\eta| < 5$). ALPGEN samples have been produced in bins of light parton multiplicity. All bins are exclusive except for the last one, which is inclusive. ℓ must be understood as lepton i.e. either e, μ or τ .

Process		# events (RunIIa)	# events (RunIIb)	$\sigma(\times\text{BR})[\text{pb}]$
$Z \rightarrow ee$ 15-75 GeV	+ 0 light parton	577k	1.9M	337
	+ 1 light parton	479k	956k	40.3
	+ 2 light partons	191k	549k	10
	+ 3 light partons	96k	536k	2.76
$Z \rightarrow ee$ 75-130 GeV	+ 0 light parton	3.0M	1.2M	132
	+ 1 light parton	1.9M	567k	40.8
	+ 2 light partons	982k	268k	10
	+ 3 light partons	988k	127k	3.15
$Z \rightarrow ee$ 130-250 GeV	+ 0 light parton	295k	352k	0.89
	+ 1 light parton	192k	179k	0.37
	+ 2 light partons	98k	160k	0.09
	+ 3 light partons	98k	300k	0.03
$Z \rightarrow \mu\mu$ 15-75 GeV	+ 0 light parton	577k	1.7M	336
	+ 1 light parton	483k	570k	39.7
	+ 2 light partons	192k	275k	9.9
	+ 3 light partons	96k	268k	2.8
$Z \rightarrow \mu\mu$ 75-130 GeV	+ 0 light parton	3.0M	1.5M	132
	+ 1 light parton	2.0M	604k	40.6
	+ 2 light partons	1.1M	401k	9.8
	+ 3 light partons	1.1M	146k	3.1
$Z \rightarrow \mu\mu$ 130-250 GeV	+ 0 light parton	484k	351k	0.88
	+ 1 light parton	391k	170k	0.35
	+ 2 light partons	298k	160k	0.1
	+ 3 light partons	299k	142k	0.03
$Z \rightarrow \tau\tau$ 75-130 GeV	+ 0 light parton	2.9M	1.5M	133
	+ 1 light parton	2.0M	528k	40.6
	+ 2 light partons	963k	274k	10
	+ 3 light partons	978k	174k	3.2
$Z \rightarrow \tau\tau$ 130-250 GeV	+ 0 light parton	288k	359k	0.88
	+ 1 light parton	194k	171k	0.34
	+ 2 light partons	97k	162k	0.09
	+ 3 light partons	100k	158k	0.03

Table 4.6: List of simulated Z +light jets processes generated with ALPGEN + PYTHIA generators, along with the generator that was used for production, number of events and cross section times branching ratio (the light partons are requested to have $p_T > 8$ GeV and $|\eta| < 5$). ALPGEN samples have been produced in bins of light parton multiplicity. Bins are exclusive except for the last bin, which is inclusive.

Process	# events (RunIIa)	# events (RunIIb)	$\sigma(\times\text{BR})[\text{pb}]$	
$Zb\bar{b} \rightarrow ee + b\bar{b}$	+ 0 light parton	201k	196k	0.40
	+ 1 light parton	101k	93k	0.19
	+ 2 light partons	50k	44k	0.10
$Zc\bar{c} \rightarrow ee + c\bar{c}$	+ 0 light parton	202k	182k	0.90
	+ 1 light parton	105k	89k	0.48
	+ 2 light partons	49k	47k	0.29
$Zb\bar{b} \rightarrow \mu\mu + b\bar{b}$	+ 0 light parton	194k	206k	0.40
	+ 1 light parton	99k	96k	0.19
	+ 2 light partons	50k	45k	0.11
$Zc\bar{c} \rightarrow \mu\mu + c\bar{c}$	+ 0 light parton	194k	194k	0.93
	+ 1 light parton	102k	93k	0.50
	+ 2 light partons	51k	51k	0.29
$Zb\bar{b} \rightarrow \tau\tau + b\bar{b}$	+ 0 light parton	202k	193k	0.41
	+ 1 light parton	101k	98.2k	0.19
	+ 2 light partons	50k	44k	0.09
$Zc\bar{c} \rightarrow \tau\tau + c\bar{c}$	+ 0 light parton	196k	260k	0.91
	+ 1 light parton	97k	101k	0.51
	+ 2 light partons	48k	51k	0.28

Table 4.7: List of simulated Z +heavy flavor jets processes generated with ALPGEN + PYTHIA generators, along with the generator that was used for production, number of events and cross section times branching ratio (the light partons are requested to have $p_T > 8$ GeV and $|\eta| < 5$). ALPGEN samples have been produced in bins of light parton multiplicity. Bins are exclusive except for the last bin, which is inclusive.

4.3.3 Trigger selection

Triggers used in the Electron Channel

In the electron channel analysis, a logical OR of the Single EM and EM+JET trigger suites is used. These trigger terms are fired by events having at least a good electromagnetic object. Therefore the trigger efficiency of these terms can be modeled by the leading EM object. They are measured using a tag-and-probe method on the $Z \rightarrow ee$ sample. A detailed description of the trigger efficiency measurement for EM+JET triggers can be found in [94, 95, 96].

The event weight applied to simulated samples is calculated through measuring the trigger efficiencies as a function of p_T and η_{det} , and the prescale of each trigger term while normalizing the integrated luminosity of each trigger list by using the `caf_trigger` package [97]. When applying trigger efficiency corrections, trigger terms are matched with the leading electron candidate in the analysis.

Triggers used in the Muon Channel

In order to increase acceptance in the muon channel, for which the trigger efficiency of the OR of single-muon triggers is approximately 70%, no explicit trigger in the muon analysis are required, also called inclusive trigger approach. The validity of this procedure has been established on RunIIa data [98] and verified it on RunIIb data.

First, the analysis is performed through the full selection stage using single muon triggers only. Using the difference between the inclusive-trigger data and single muon only data, a correction factor is derived for the Monte Carlo intended to take into account the non-single muon trigger contribution to the data:

$$P_{corr} = \frac{(Data - QCD)_{incl} - (Data - QCD)_{single\mu}}{MC_{incl}} \quad (4.1)$$

where QCD is the number of multijet events and MC_{incl} refers to the Monte Carlo with the trigger probability set to 1. The correction is parametrized in H_T (defined as the scalar sum over all p_T^{jet} in the event) and in muon η_{det} for $|\eta_{det}| < 1.6$. As seen in Section 2.2.5, the coverage of the muon detector is up to $|\eta| \lesssim 2$. Very forward muons are not triggered by the single-muon limiting the muon acceptance up to $|\eta| < 1.6$. For each Monte Carlo event, the trigger probability is set to be

$$P_{trig} = P_{corr}(H_T, \eta_{det}) + P_{single\mu} \quad (4.2)$$

where $P_{single\mu}$ is the probability for a given event to fire a single muon trigger, obtained from the `caf_trigger` package. P_{trig} is required to not exceed unity for each event. If $P_{trig} > 1$ for an event it is forced to be $1 - P_{single\mu}$ so that $P_{corr} + P_{single\mu}$ never exceeds 100%. The inclusive trigger approach is used for both RunIIa and RunIIb for the remainder of the analysis.

Figure 4.3 shows the trigger correction as a function of H_T for RunIIb.

4.4 Event selection

The following selection criteria are designed to accept events with a maximal acceptance for the WH signal and keep the volume of data to analyze as low as possible. The same selection criteria

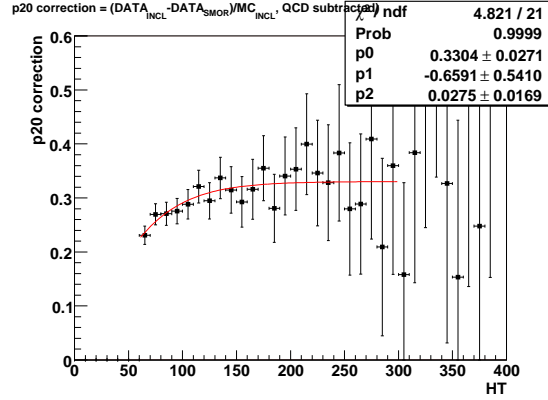


Figure 4.3: Trigger correction, P_{corr} in Eq. 4.1, for RunIIb, parameterized with a tanh function.

is applied to both RunIIa and RunIIb datasets; this simplifies the dataset combination and allows us to use the same multivariate technique. The selection stage consists of the following criteria:

- Exactly one electron (muon) with transverse momentum $p_T > 15$ GeV within a pseudorapidity range of $|\eta| < 2.5$ where the ICR region ($1.1 < |\eta_{det}| < 1.5$) is excluded for electrons and $|\eta| < 1.6$ for muons. Additional lepton flavor-specific requirements are detailed in Section 4.4.2;
- $\cancel{E}_T > 20$ GeV;
- A primary vertex with at least 3 associated tracks and $|z| < 40$ cm;
- Exactly two or three jets, analyzed in separate channels, with $p_T > 20$ GeV after jet energy scale correction with ICD hot cell removal (only in RunIIa data), and $|\eta| < 2.5$;
- The sum of the p_T of the jets (H_T) is required to exceed 60 GeV (80 GeV in the 3 jets case) in order to ensure that events lies above the trigger “turn-on” and reduce the contribution of low p_T QCD events. This cut also increases the intrinsic sensitivity of analysis based on $S/\sqrt{S+B}$;
- A 2-dimensional “triangular” cut $M_W^T > -0.5 \cdot \cancel{E}_T + 40$ GeV is applied to reduce the multijet background.

Since the detector simulation for MC is not perfectly modeled, some corrections need to be applied to simulation. These corrections such as lepton reconstruction efficiency, jet reconstruction efficiency are described in the sections related to each object selection.

4.4.1 Primary vertex selection

Events are selected from collisions occurring in the central part of the detector, where the activity from the primary interaction can be covered by the SMT. This ensures that analyzed $p\bar{p}$ collisions

contain final state objects within the detector acceptance, hence removing events which potentially would not contain all reconstructed objects. Therefore, the z position of the reconstructed primary vertex is required to be within $|PV_z| < 40$ cm.

4.4.2 Lepton selection

The final state that is studied contains an isolated lepton from the decay of the W boson. Although the hadronic decay of the W boson has a much higher branching ratio than the leptonic one, requiring an isolated lepton allows to deal with a significantly reduced multijet background. Both muon and electron channels are analyzed separately. In this section, lepton-specific selection criteria will be detailed. Since electrons are calorimetric objects, this channel has a bigger multijet contribution coming from misidentified jets and photons. In the muon channel, semi-leptonic decays in jets produce muon inside the jets, but only isolated leptons are selected. In both channels, events are required to have exactly one lepton with $p_T > 15$ GeV.

Electron selection

The lepton selection criteria in the electron channel are optimized separately for the central calorimeter region ($|\eta| < 1.1$) and end-cap region ($1.5 < |\eta| < 2.5$). In order to maximize the signal acceptance, the selection for electrons in the central part of the detector is relatively loose, as opposed to the forward region where the multijet contribution is more important, therefore using a tighter operation point.

As mentioned in Section 3.3.1, the WH analysis is using the LOOSE and MEDIUM electron identification criteria for electrons in the CC. In the EC, the `top_loose` and `top_tight` criteria are used. The electron selection is based on cuts on variables listed in Table 3.1. Electrons passing the MEDIUM and `top_tight` requirements are considered “tight” for the purpose of determining the multijet background and used for the signal sample from which the final result will be extracted.

Electrons in the intercryostat region are not considered.

Muon selection

Muons are reconstructed using information of two independent subdetector systems which are the muon detector and the central tracker. Muon identification is based on a track in the muon system, referred to as local track. Muons are required to have hits in all layers of the muon system inside and outside the toroid. The superior spatial resolution of the central tracker is used to improve the accuracy of the muon’s kinematic properties and to confirm that the muon originated from the primary vertex.

The muon candidates are required to fulfill the following criteria, defined in Section 3.4.1:

- *medium* muon type,
- muon quality $nseg = +3$, signifying matching muon segments from A and BC layers, as well as a central track match. This requirement rejects about 10% of high transverse momentum

muons, mainly in the bottom region of the detector where no full muon coverage in the A and BC layers is available,

- *medium* track quality,
- Cosmic muon veto.

Muons coming from the leptonic decay of W bosons tend to be isolated from jets, while muons originating from semi-leptonic decays of heavy flavored hadrons are typically non-isolated due to jet fragmentation of the partial hadronic decay. A “loose” isolation criterion, DeltaR which is defined in Section 3.4.1, is used to estimate the multijet background. The “tight” muon sample, used as the signal sample in this analysis, requires the additional NPTight isolation criteria.

Scale factors are applied to the simulation based on these identification, track matching and isolation criteria.

The distributions of the energy, p_T , η and ϕ of the electron and muon channels for the $W + 2$ jet events are shown in Figure 4.12. These distributions are described by the sum of the simulation and the multijet background derived from the data, as described below.

4.4.3 Missing E_T selection

The W boson decaying leptonically, one of the outgoing particle is a neutrino. This particle doesn't interact with the detector and thus its presence is characterized by missing energy in the transverse plane to the beam direction. As seen in Section 3.6, the missing transverse energy is recalculated after all jets are corrected with JES, and energy deposit from muons in the calorimeter is taken into account. The applied cut is set to $\cancel{E}_T > 20$ GeV and is efficient for the removal of mismodeled events and reduces the multijet dominated region with low \cancel{E}_T .

\cancel{E}_T distributions for the the electron and muon channels are shown in Figure 4.12 for the RunIIb period.

4.4.4 Jet selection

For $M_H < 135$ GeV, the Standard Model predicts that the Higgs boson decays dominantly in a $b\bar{b}$ pair. By analyzing the 2 jets channel, most of the sensitivity is obtained. But one can recover events which radiated an additional jet from an initial or state. Therefore, the analysis is split in two orthogonal channels and are analyzed separately and then combined at the limit calculation step. After satisfying the identification criteria described in 3.5 and applying the corresponding corrections, jets have to fulfill the following requirements:

- $p_T > 20$ GeV
- $|\eta_{det}| < 2.5$
- jets must be *vertex confirmed*.

Figure 4.13 shows the p_T , η and ϕ distributions for the leading p_T and next-to-leading p_T jets.

4.4.5 Triangular cut

The multijet background consists of jet+jet events where the jet is reconstructed as an isolated lepton and photon+jet events where the photon would be an electron candidate. In the muon channel, a muon produced inside a jet is reconstructed outside of it, thus leading to low M_T^W value. These mismeasurements lead to the same signature as other physics background, with a isolated lepton, missing transverse energy and jets. Since a straight cut on a single distribution is not the most efficient way to get rid of this background which is significant especially in the electron channel, it has been chosen to cut on a two dimensional distribution in the (M_T^W, \cancel{E}_T) plane. Events passing the following cut are kept:

$$M_T^W > -0.5 \cancel{E}_T + 40 \text{ GeV}. \quad (4.3)$$

This cut also ensures that the agreement between data and simulation is good for low \cancel{E}_T and low M_T^W regions, even though the multijet background modeling is parametrized in multiple dimensions (Section 4.6). The ratio of data events to the WH expectation in the 2D plane is shown on Figure 4.4(b). A large reduction of the multijet background is clearly visible by applying the triangular cut.

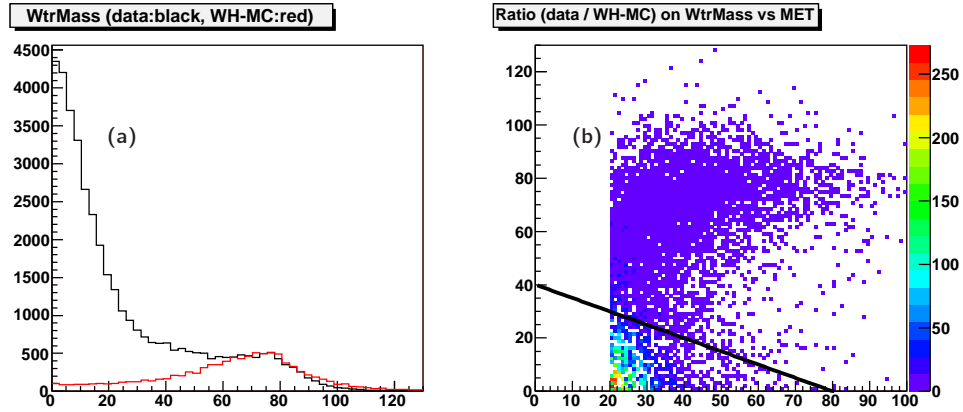


Figure 4.4: (a) M_T^W using loose electron criteria in data (black) as multijet sample, and WH MC (red). (b) 2D ratio plot of data/ WH MC in the (M_T^W, \cancel{E}_T) plane. The multijet events are lying at low \cancel{E}_T and low M_T^W and can therefore be eliminated by a triangle cut of $M_T^W > -0.5 \cancel{E}_T + 40$ as shown by the black line. The WH MC is normalized on the W peak.

4.4.6 Vetoes

Since this result is a part of the $D\bar{D}$ and Tevatron Standard Model Higgs combination, a particular attention has to be drawn to avoid biases in the process of combining several channels which can contain the same topology. It is thus important to avoid analyzing twice the same events in different analyses. Therefore a set of cuts are applied to avoid this potential bias when selecting

events passing selection in another analysis. In the case of the WH analysis, orthogonality cuts are defined to remove events selected by the $ZH \rightarrow \ell\ell b\bar{b}$ and $X + H \rightarrow \tau\tau jj$ analyses. The same object selection applied in these analyses is used to reject events that will fulfill such requirements.

In order to minimize backgrounds that contain two isolated leptons in the final state, such as those coming from Z and $t\bar{t}$ production, events are required to not contain any additional isolated lepton.

In the $WH \rightarrow e\nu b\bar{b}$ channel, events which additionally contains isolated muons¹ with $p_T > 15$ GeV, are rejected. Similarly, in the muon channel, events with an additional LOOSE electron are not considered.

The effect on yield for the vetoes described now is summarized in Appendix A.

Tau veto

Hadronic tau candidates that passes the $X + H \rightarrow \tau\tau jj$ selection are vetoed. Candidates can be any of the three DØ standard tau types [99], with the following requirements for each type:

Type1 $E_T > 10$ GeV, $p_T > 7$ GeV, $E_T/p_T > 0.5$, $NN_\tau > 0.9$

Type2 $E_T > 10$ GeV, $p_T > 5$ GeV, $NN_\tau > 0.9$

Type3 $E_T > 15$ GeV, $p_T > 7$ GeV, $\sum_{trk} p_T^{trk} > 15$ GeV, $NN_\tau > 0.95$

where p_T is the momentum of the associated track, NN_τ is the hadronic tau identification tool (analogous to the NN b -tagger) and $\sum_{trk} p_T^{trk}$ is the sum of tracks p_T associated to the τ candidate. This condition applied in both electron channel muon channels. There is a negligible loss in data and WH signal for both the electron and muon channels, while the largest difference can be seen in ZH and ZZ due to type 2 taus in the electron channel. The $t\bar{t}$ background is also reduced by 8% , mainly due to di-lepton final states where there can be a tau in the final state.

$ZH \rightarrow \ell\ell b\bar{b}$ veto

The $ZH \rightarrow \ell\ell b\bar{b}$ is split in different sub-channels. This search is performed on dilepton events with 2 or more jets. The dilepton pair can either be two muons or two electrons. A second lepton veto allows the WH analysis to be orthogonal to these two channels.

In order to ensure orthogonality with the $ZH \rightarrow \mu\mu b\bar{b}$ analysis, a veto is applied on events with two loose track-matched muons with $p_T > 10$ GeV, in the $60 < M_{\mu\mu} < 150$ GeV range. This allows as well to reject $Z \rightarrow \mu\mu$ and $t\bar{t}$ background events.

In the case of the $ZH \rightarrow ee b\bar{b}$ analysis, events are required to not contain additional electrons with $p_T > 15$ GeV that satisfy the loose electron criteria.

In order to increase their sensitivity, two alternate channels are considered in the analysis of final state with two leptons and two jets in order to increase the acceptance. The dilepton pair can also be a muon and an isolated track (to recover di-muon events with one of the lepton not

¹The definition of muons to be discarded in the analysis is: MediumNseg3 muon quality, Medium track quality, DeltaR isolation requirements.

having hits in the muon spectrometer) or an electron and a “ICR electron” (e_{ICR}).

Veto on $e + e_{ICR}$ events

To reproduce the $e + e_{ICR}jj$ analysis, the selection for the $e + e_{ICR}$ pair is done as following:

electron $p_T > 20$ GeV

ICR electron $p_T > 20$ GeV

Events with a reconstructed (electron, e_{ICR}) invariant mass greater than 5 GeV are vetoed. There is almost no data loss due to this veto. The largest differences can be seen in ZH and ZZ , which loose about 7%.

Veto on $\mu + \mu_{trk}$ events

To reproduce the $\mu + \mu_{trk}jj$ analysis, the selection for the μ +track pair is done as the following:

muon $p_T > 10$ GeV

track $p_T > 20$ GeV.

Events with a reconstructed (μ, trk) invariant mass between 60 GeV and 150 GeV are rejected. The largest differences can be seen in ZH and ZZ , where a loss of about 50% is observed, which indicates an effective $\mu\mu_{trk}$ selection.

4.5 Reweighting of W +jets and Z +jets samples

As described in Section 4.3.2, W +jets and Z +jets samples are generated by ALPGEN. Studies on MC generators have shown that some distributions are not well modeled and distributions shapes are different in data. Also, by comparing to other MC generators, one found that shapes are different [88]. Comparison plots are shown on Figure 4.5 for η^{jet1} , η^{jet2} , $\Delta R(jet_1, jet_2)$ and p_T^W .

Standard corrections for p_T^W and p_T^Z are applied by default in $D\bar{O}$ analyses¹ in order to correct the physics modeling of soft radiation emission back-to-back to the vector boson, for which the transverse momentum of the boson is very sensitive [100, 101, 102]. Nevertheless, these corrections don't seem to bring simulation distributions shapes close enough to the data ones. In order to correct these differences, reweighting functions are derived to reproduce the difference between ALPGEN W +jets and Z +jets MC samples and data where all backgrounds are subtracted except W and Z +jets MC:

$$\frac{N_{Data} - N_{QCD} - N_{SM-Vjets}}{N_{Vjets}} \quad (4.4)$$

where N_{Data} is the data yield, N_{QCD} is the multijet yield, $N_{SM-Vjets}$ is the sum of $t\bar{t}$ diboson and single-top MC yields, and N_{Vjets} is the expected yield for MC W/Z +jets. A fit $f_d(x_d)$ (where x_d

¹corrections are derived from the boson p_T distribution at the generator level and the observed spectrum in unfolded data to remove detector effects.

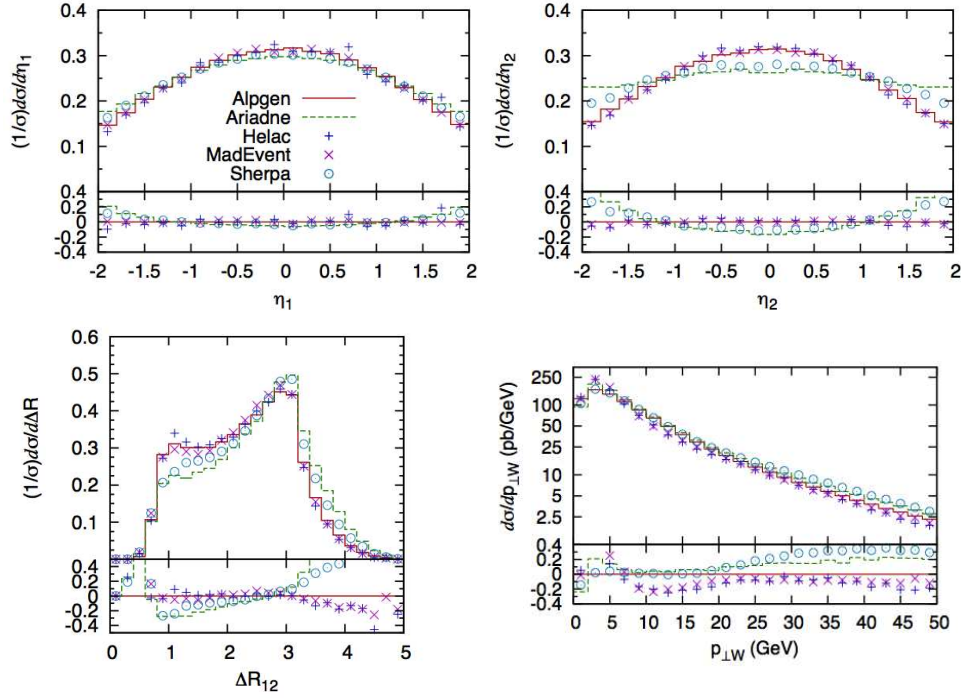


Figure 4.5: η distributions of the 2 leading jets (upper row), $\Delta R(jet_1, jet_2)$ (bottom row, left) and p_T^W (bottom row, right) for different MC generators. All curves are normalized to unit area. Figures taken from [88]

is the value of the variable of interest d for a given event) to the ratio defined in Equation 4.4 is performed. The event weight is then scaled

$$w_i \rightarrow w_i \prod_V f_V(x_V),$$

where w_i is the event weight for the i^{th} event and V runs over all distributions from which the reweighting functions have been derived. These corrections are assumed to correct generator level mismodeling.

The W +jets and Z +jets Monte Carlo samples are reweighted for both RunIIa and RunIIb in the electron and muon channel. All light and heavy flavor W/Z +jets samples are reweighted using the same functions.

In the ICD region of the detector ($0.8 < |\eta_{det}| < 1.4$), a "horn" effect is visible in the η_{det} distributions of the leading and second leading jet of the event. This feature, present in RunIIb only, comes from the detector gain loss in this region, and therefore a correction is applied only to this dataset. This effect is compensated by applying constant factors $f_{\eta_{jet_1}^{ICDnorth}}$, $f_{\eta_{jet_1}^{ICDsouth}}$, $f_{\eta_{jet_2}^{ICDnorth}}$, and $f_{\eta_{jet_2}^{ICDsouth}}$, listed below, to the corresponding ICD regions of the jet.

To reproduce the shapes observed in data, a reweighting function for the lepton η distribution is derived with a second order polynomial, $f_{\eta_{lepton}}$, and reweighting functions for the η distributions of the leading and second leading jet with fourth order polynomials, $f_{\eta_{jet_1}}$ and $f_{\eta_{jet_2}}$, listed below. In three jet events, only the leading two jets are reweighted.

Discrepancies in the shape of the $\Delta R(jet_1, jet_2)$ and p_T^W distributions are also seen. These two variables show a strong correlation, therefore a 2D reweighting function $f_{p_T^W} \times f_{\Delta R}$ is derived from the ratio of W +jets data (meaning all other backgrounds have been subtracted) to W +jets MC in the $\Delta R - p_T^W$ plane (see Fig. 4.7 (a)–(d)). Both functions are then applied to the W +jets MC sample. The ΔR function is also applied to the Z +jets MC sample conserving the event yield for this sample.

Figures 4.6 and 4.7 (e,f) show all reweighting functions applied in this analysis. In each figure, the black curve represents the nominal reweighting function, while the red and blue dashed curves represent $\pm 1\sigma$ variations on the fit function. These variations are based on selecting the parameter that affects the shape of the fit function the most (the highest order polynomial term for most fits; the slope of the sigmoid turn-on for p_T^W), varying it by 1σ based on the uncertainty of the fit, and varying the other fit parameters by the appropriate amount based on the covariance matrix of the fit.

The functions are defined as:

$$\begin{aligned}
 f_{\eta_{jet_1}^{CDsouth}} &= 0.928 \\
 f_{\eta_{jet_1}^{CDnorth}} &= 0.992 \\
 f_{\eta_{jet_2}^{CDsouth}} &= 0.862 \\
 f_{\eta_{jet_2}^{CDnorth}} &= 0.899 \\
 f_{\eta_{lepton}} &= 0.974 + 0.066 \eta_{lepton}^2 \\
 f_{\eta_{jet_1}} &= 0.977 - 0.026 \eta_{jet_1}^2 + 0.019 \eta_{jet_1}^4 \\
 f_{\eta_{jet_2}} &= 0.943 - 0.005 \eta_{jet_2}^2 + 0.018 \eta_{jet_2}^4 \\
 f_{\Delta R} &= 4.382 - 0.255 \Delta R + 0.085 \Delta R^2 - 0.006 \Delta R^3 \\
 f_{p_T^W} &= 0.216 + 0.054 (1 + \text{erf}((W_{p_T} - 7.963)/(\sqrt{2} \times (-8.974))))
 \end{aligned}$$

with the error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

The functions for η_{lepton} , η_{jet_1} and η_{jet_2} contain only quadratic and quartic terms since the detector (and the observed disagreement between data and simulation) is assumed to be symmetric around $\eta = 0$.

Systematic uncertainties are estimated from the fitted function for each reweighting function. This is described in 5.2.2. They are taken into account while setting limits.

The η_{jet_1} , η_{jet_2} , η_{lepton} , $\Delta R(jet_1, jet_2)$ and p_T^W distributions are shown in Figure 4.8, after the reweightings being applied. With this procedure, a good agreement is observed between data and simulation. The same reweighting functions are applied for the 3 jets channels since it is found

that they are efficient in correcting simulation to agree with data.

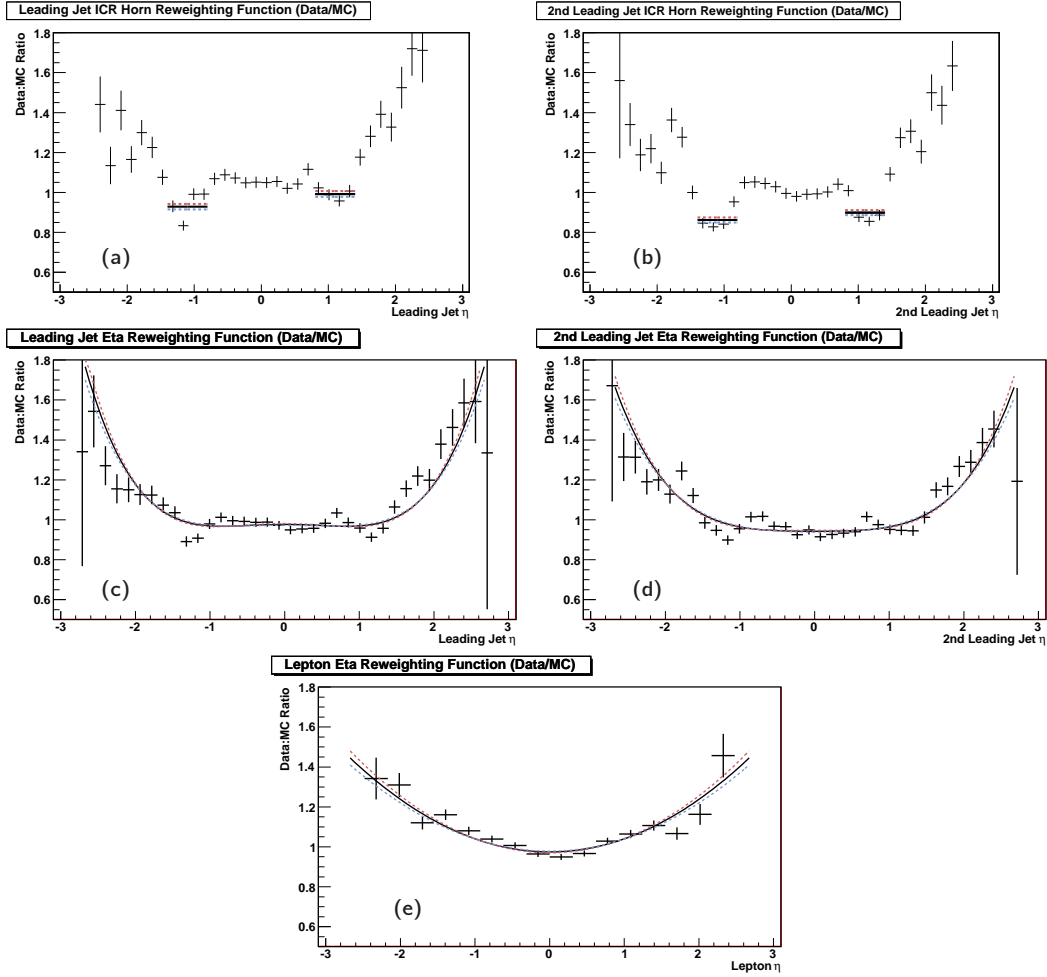


Figure 4.6: Reweighting functions applied to correct for the horns in the ICR region in η_{det} of the leading jet (a) and the second leading jet (b), constant factors are used for these correction; the mismodeling of η in the leading jet (c) and the second leading jet (d), fourth order polynomials are used in these cases; and the mismodeling of η of the lepton (e), a second order polynomial is used for the correction function. The black curve is the nominal reweighting function, while the red and blue dashed curves are systematic variations.

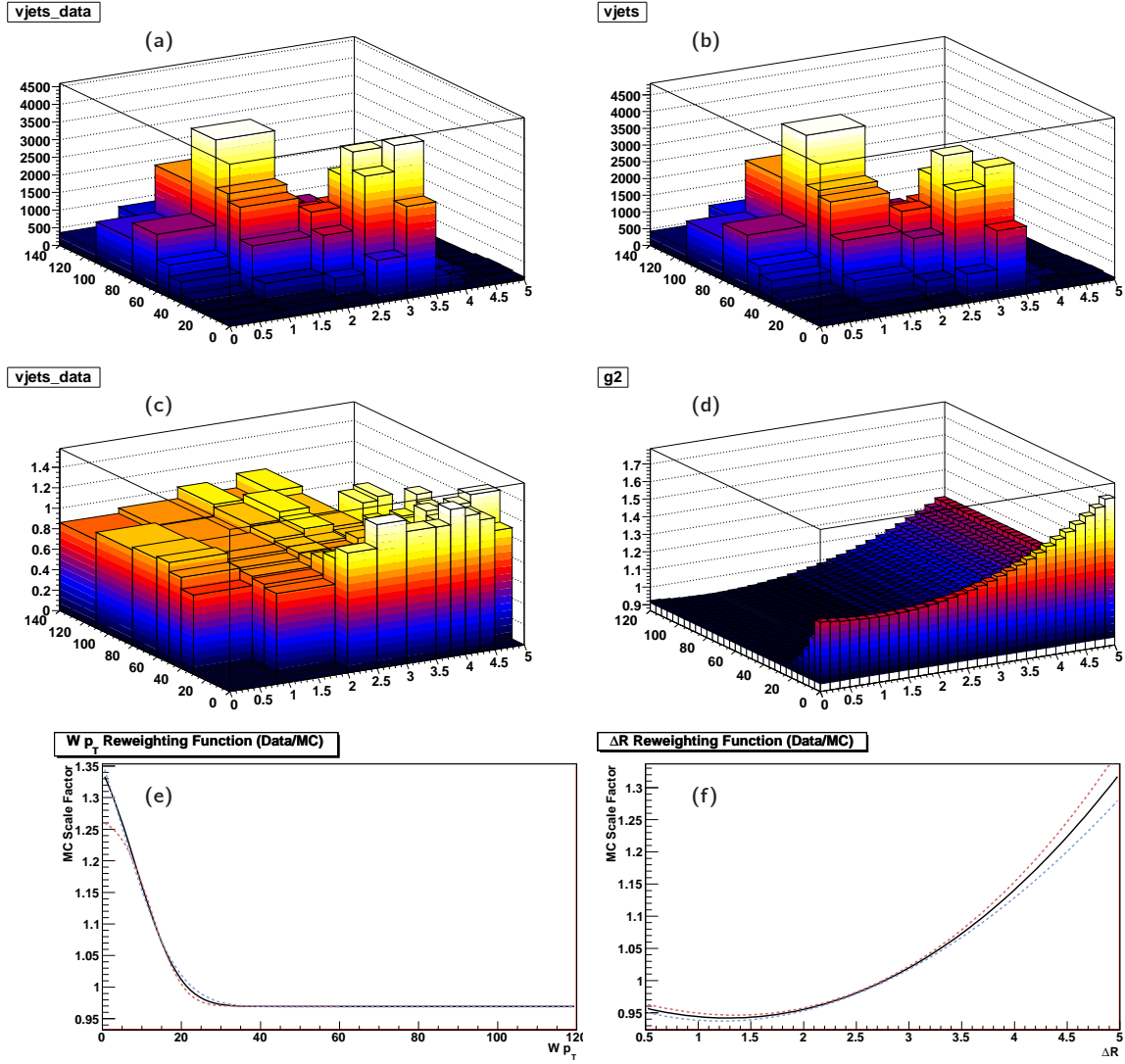


Figure 4.7: 2D (simultaneous) reweighting of ΔR and p_T^W . (a) shows the distribution of W +jets data (meaning all other MC backgrounds have been subtracted) in the $\Delta R - p_T^W$ plane, (b) shows the distribution for W +jets MC, (c) shows their ratio, (d) shows the 2D fit function applied to reweight the ratio plot. (e) shows the p_T^W reweighting function (obtained from the 2D fit) that is applied to the W +jets MC sample only, (f) shows the ΔR reweighting function (obtained from the 2D fit) that is applied to both W +jets and Z +jets MC samples (conserving the event yield for the Z +jets MC sample).

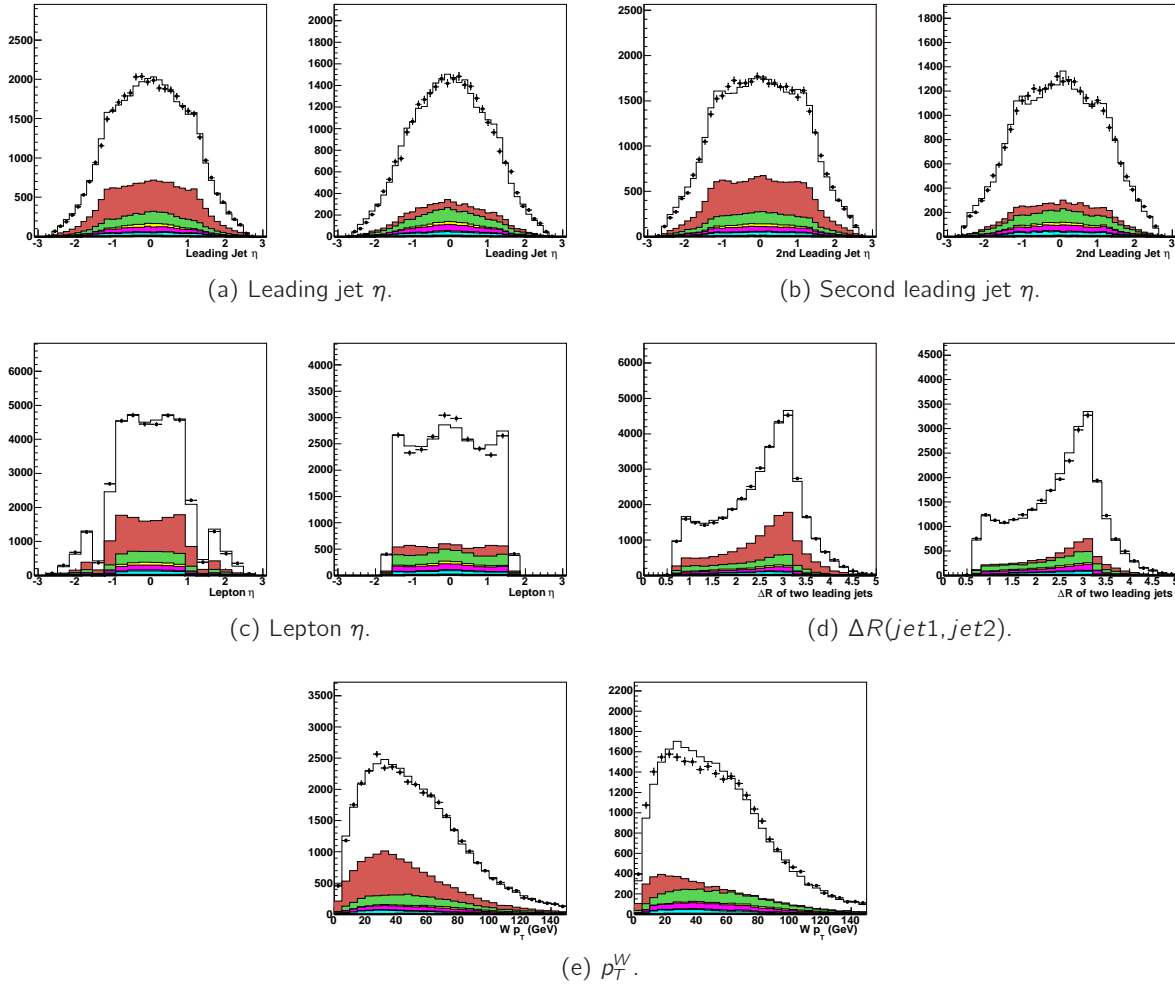


Figure 4.8: η_{jet1} , η_{jet2} , η_{lep} , $\Delta R(jet1, jet2)$ and p_T^W distributions after all reweightings applied, for the RunIIb electron and muon channels (grouped in pairs, electron channel shown in the left figure and the muon channel in the right figure).

4.6 Multijet background estimation

The collected data passing the analysis selection cuts is compared to simulated physics backgrounds, where the detector response is simulated. Another background needs to be taken into account, which is the multijet background. Even though events produced from QCD interaction do not contain an isolated lepton, they can pass the selection in the following cases:

- a jet can be misidentified as an electron,
- a photon can be misidentified as an electron,

- jets containing semi-leptonic decays can have the lepton reconstructed outside of the jet, hence isolated, mainly contributing in the muon channel.

Furthermore, up to now, the description of QCD processes is not modeled well enough by current MC generators. Therefore, the so-called *Matrix Method* is used, a data-driven method which will be described in this section. Two inputs are used in this method, the probability that a jet (or a photon in the electron channel) is misidentified as a lepton (fake rate) and the lepton identification efficiency. A precise measurement and parametrization will allow to give an accurate modeling of this instrumental background.

4.6.1 Multijet background modeling strategy

The background from multijet events, also referred to as QCD background, is determined by forming a template that models the kinematics of multijet events and then scaling that template at the selection stage of the analysis to estimate the number of multijet events that pass the analysis selection. The template is created by reweighting individual events in data based on the Matrix Method technique.

The multijet background can be suppressed by tightening lepton isolation criteria. By choosing a “loose” and “tight” isolation working points, one can define two samples, where the tightest is a subsample of the loosest one.

The following pair of equations can be formed and are solved simultaneously in order to estimate the number of multijet events:

$$N_L = N_\ell + N_{QCD}, \quad N_T = \varepsilon_\ell N_\ell + f_j N_{QCD}, \quad (4.5)$$

where N_L (N_T) is the number of events in data with a lepton that passes a loose (tight) isolation requirement, N_ℓ is the number of events with a well identified lepton and N_{QCD} is the number of events misidentified as having an isolated electron in that data. ε_ℓ is the efficiency for a real lepton that passes the loose identification requirement to subsequently pass the tight identification requirement, and f_j is the rate at which a jet that has been misidentified as a lepton that passes the loose identification requirement subsequently passes the tight identification requirement. Solving this system of equations for the number of multijet events in the tight sample yields:

$$N_{QCD}^T = \frac{f_j}{\varepsilon_\ell - f_j} (\varepsilon_\ell N_L - N_T) \quad (4.6)$$

This method is designed to work on a binned data sample, where N_L and N_T can be counted, and constant ε_ℓ and f_j are used. However, ε_ℓ and f_j may depend on event kinematics described by \vec{k}_i . Therefore, each event with a loosely identified electron sample contributes to the multijet background estimation in the tight identification sample by assigning each event a weight, w_i , that depend on parameterized functions of ε_ℓ and f_j . Starting from Equation 4.6, the event weight can be expressed as:

$$w_i = \frac{f_j(\vec{k}_i)}{\varepsilon_\ell(\vec{k}_i) - f_j(\vec{k}_i)} (\varepsilon_\ell(\vec{k}_i) - \Theta_i^T), \quad (4.7)$$

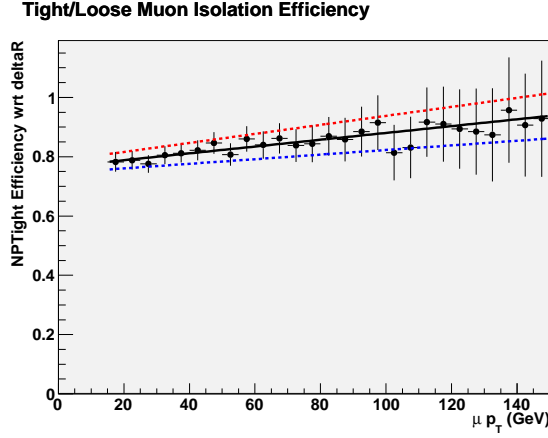


Figure 4.9: Muon efficiency parameterization, as a function of muon p_T , for RunIIb. The black line represents the fit to the data points, the red (blue) dashed lines are the $+1\sigma$ (-1σ) of the fit.

where $\Theta^T = 1$ if the event satisfies tight lepton identification criteria and 0 otherwise. The total multijet prediction in the tight sample is a sum over weights in the inclusive loose sample:

$$N_{QCD}^T = \sum_{i=1}^{N^L} w_i \quad (4.8)$$

For an electron, the efficiency ε_e , is parameterized two-dimensionally in $|\eta_{det}|$ and ϕ , while the efficiency for muons, ε_μ , is parameterized as a function of muon p_T (see Figure 4.9). These variables are chosen for the parametrization since a dependence is found. Indeed, the electron identification efficiency is found to vary as function of η_{det} given that the calorimeter performance degrades at higher pseudorapidity and the muon isolation efficiency is dependent on p_T .

As later described in Section 4.4.2, the difference from loose to tight muon selection is the isolation criteria (DeltaR for loose, NPTight for tight). In the electron channel, the same strategy is used (Loose (CC) and top_loose (EC) identification criteria for the loose electron sample, Medium (CC) and top_tight (EC) for the tight electron sample). Therefore the efficiency is defined as the ratio of tight events to loose events, in the parametrization earlier described.

4.6.2 Lepton fake rates

The fake rate f_j for jets (or photons in the electron channel) to fake isolated leptons is determined as a function of the lepton p_T in events lying at low \cancel{E}_T to select a multijet enriched region to parametrize the rate at which objects are misidentified as isolated leptons. To measure the fake rates, events are selected with all analysis cuts applied (defined in Section ??), with the difference that $5 < \cancel{E}_T < 15$ GeV is applied and the triangle cut is removed.

Separate parameterizations in p_T are determined for several two-dimensional bins of $|\eta_{det}|$ and $\Delta\phi(\cancel{E}_T, e)$ in the electron channel (all figures can be found in Appendix B). Muon fake rates are determined in bins of η and $\Delta\phi(\cancel{E}_T, \mu)$.

For electrons and muons, the fake rate for a given kinematic range is determined by the ratio of MC-subtracted data events, to isolate the multijet contribution, with loose and tight lepton identification criteria (the criteria used in the analysis are defined in Section 4.4.2):

$$f_{QCD} = \frac{N_T - MC_T}{N_L - MC_L}, \quad (4.9)$$

where N_L (N_T) is the number of data events in the multijet estimation sample with a reconstructed loose (tight) lepton and MC_L (MC_T) is the total MC estimation of the number of events with real loose (tight) leptons in them.

The electron fake rate is parameterized as:

$$\begin{aligned} f_{QCD}(p_T) &= e^{-a_0 p_T + a_1} + a_2 + a_3 p_T & (|\eta| < 1.1) \\ f_{QCD}(p_T) &= a_0 + a_1 p_T & (1.5 < |\eta| < 2.5) \end{aligned} \quad (4.10)$$

where a_N are the fit parameters. These parameters are determined separately for electrons in the EC and CC, using separate CC bins above and below $|\eta_{det}| = 0.7$. Each $|\eta_{det}|$ region is further divided into $\Delta\phi(\not{E}_T, e)$ bins between 0, 0.5, 1, 2 and π . In 2 jets events, each of these individual $|\eta_{det}|$ and $\Delta\phi(\not{E}_T, e)$ is fit separately. In 3 jets events, parameterizations for each $|\eta_{det}|$ bin are determined, and these curves are scaled based on variations in $\Delta\phi(\not{E}_T, e)$ bins. Furthermore, fake rates are determined separately for RunIIa and RunIIb data since lepton identification performance are dependent on the instantaneous luminosity, which increased between the two periods.

Figure 4.10 shows the resultant fit for fake rate in 2 jets events for RunIIb data as a function of e p_T in the $|\eta_{det}| < 0.7$ $2 < \Delta\phi(\not{E}_T, e) < \pi$ bin, with the parameterized function overlaid on the corresponding data, along with associated uncertainty bands. All parametrizations for the electron fake rate are given in Appendix B.

The muon fake rate is determined in two separate regions, $\Delta\phi(\not{E}_T, \mu) < \pi/2$ and $\Delta\phi(\not{E}_T, \mu) > \pi/2$. Parameterization functions appear in Table 4.8 and are plotted in Figure 4.10. These fake rates are applied to both 2- and 3 jets events.

	$\Delta\phi(\not{E}_T, \mu) < \pi/2$	$\Delta\phi(\not{E}_T, \mu) > \pi/2$
RunIIa	$a_0 + a_1 p_T$	Binned values used
RunIIb	$a_0 + a_1 p_T + a_2 p_T^2$	$a_0 + a_1 p_T$

Table 4.8: Muon fake rate (f_{QCD}) parameterization functions.

In order to validate the multijet background description, one can see on Figure 4.11 the distributions of $\Delta\phi(\ell, \not{E}_T)$ for all 2 jets channels. One can observe a good data to Monte Carlo agreement, especially at high $\Delta\phi(\ell, \not{E}_T)$ values, a sensitive region to the fake rate parametrization.

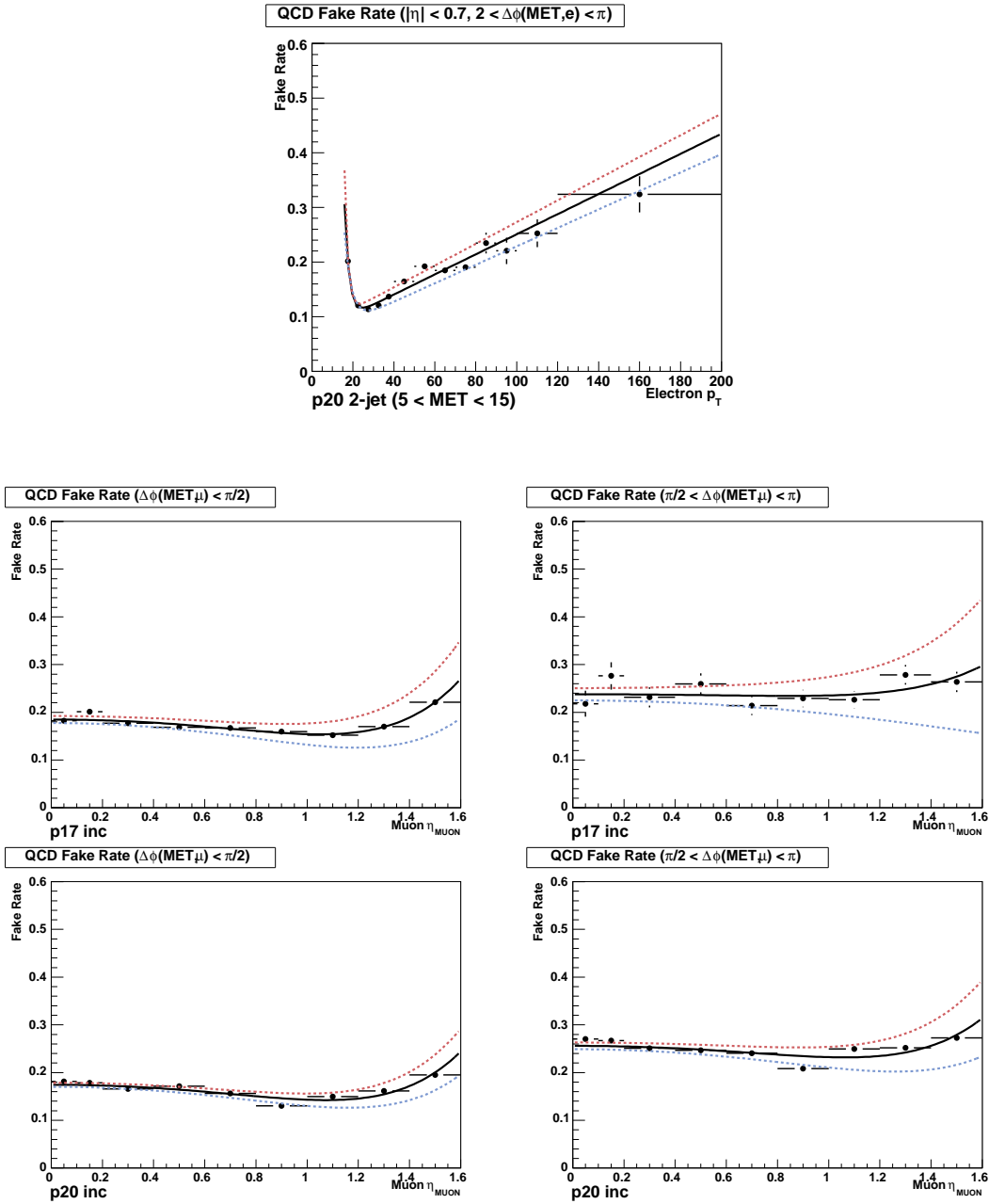


Figure 4.10: Lepton fake rate ratios in data. The electron fake rate (upper figure) in 2 jets events for RunIIb data is shown as a function of p_T^{electron} in the $|\eta_{\text{det}}| < 0.7, 2 < \Delta\phi(\not{E}_T, e) < \pi$ bin. All parametrizations are given in Appendix B. Middle (bottom) figures show the muon fake rate ratios in RunIIa (RunIIb) data as a function of muon $|\eta_{\text{det}}|$. Figures on the left show data with $\Delta\phi(\not{E}_T, \mu) < \pi/2$, while figures on the right are for $\Delta\phi(\not{E}_T, \mu) > \pi/2$.

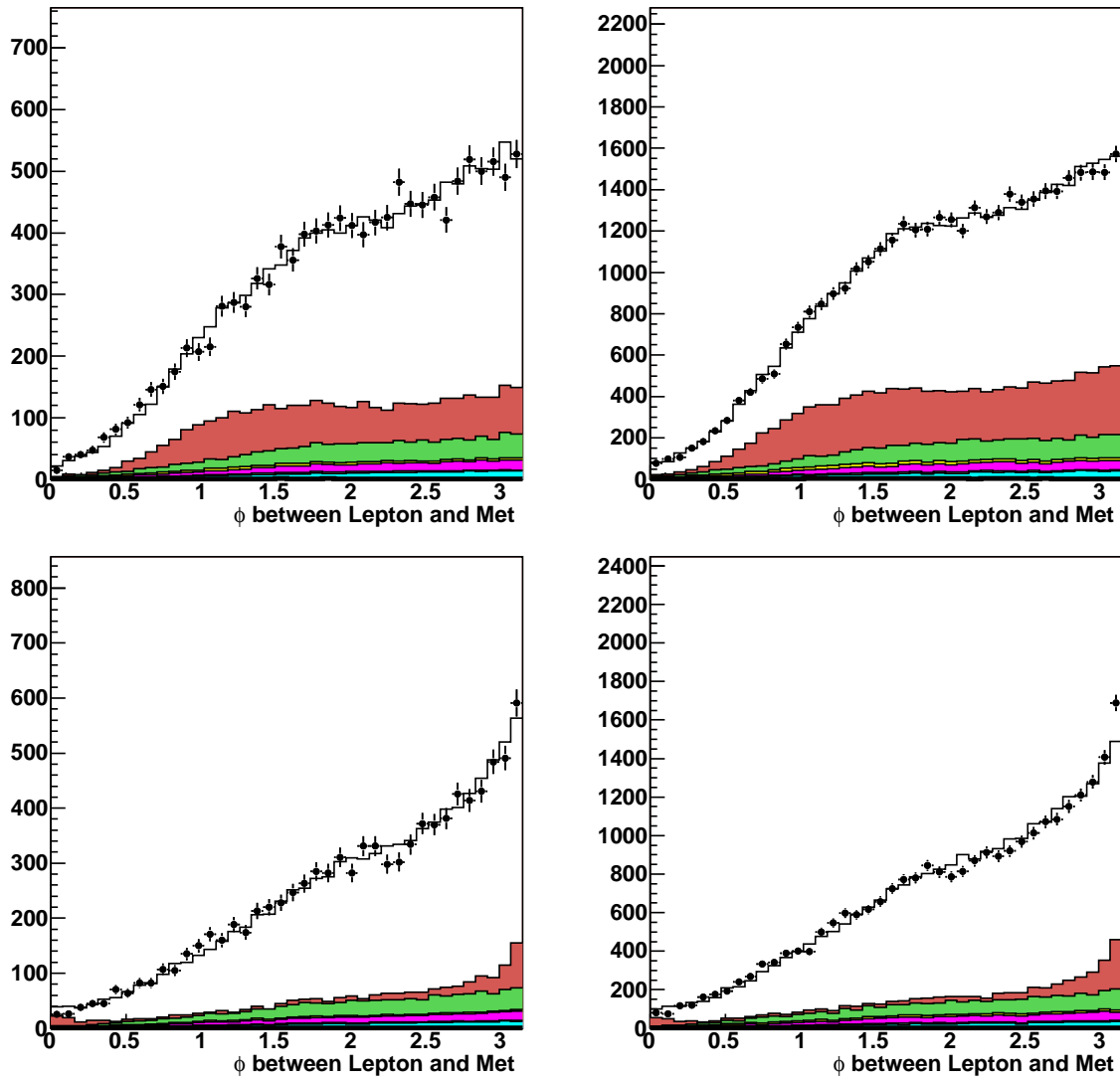


Figure 4.11: $\Delta\phi(\ell, \cancel{E}_T)$ distributions for the electron (top row) and muon (bottom row) channel, RunIIa (left) and RunIIb (right).

4.7 Simulation normalization scheme

All samples are normalized to their cross section at the highest order available. This is valid for all Monte Carlo samples produced except for W +jet processes and the multijet background estimated from data. The procedure which is followed in the WH analysis is detailed in this section.

4.7.1 Multijet sample normalization

The unbinned Matrix Method is applied in order to form a multijet template. The normalization of the template is obtained by performing a χ^2 minimization fit to the W boson transverse mass (m_T) distribution at selection stage of the analysis. While determining the multijet template normalization factor, the W +jet normalization is varied so that the total number of estimated events matches the number of data events before applying b -tagging. Separate multijet normalization factors are determined for both the loose and tight lepton ID operating points. These normalization factors remain fixed for the rest of the analysis.

4.7.2 Experimental K factors for the W +jet background

The simulated background processes are absolutely normalized to the SM prediction of their cross section except for the ALPGEN W +jets samples. W +jets samples are normalized to data before applying b -tagging. The final normalization is set in conjunction with the scaling of the multijet background template, as described in Section 4.7.1. The experimental K factor K_{LF}^{exp} for the W +jets processes is defined as

$$K_{LF}^{exp} = \frac{N_{data} - N_{SM} - N_{QCD}}{N_{W+jet}}. \quad (4.11)$$

where N_{SM} is the number of Standard Model expected events with proper isolated lepton identification but without the W +jets contribution, N_{QCD} is the number of Standard Model events with a misidentified lepton, estimated from data and N_{W+jet} is the expectation from ALPGEN. The K_{LF}^{exp} for e (μ) from each dataset is summarized in Table 4.9, and is ~ 1.1 (1.1) for RunIIa and ~ 1.0 (1.0) for RunIIb. All these numbers are consistent, given the uncertainty which is of the order of 10%, and which originates from lepton characteristics (lepton trigger, lepton efficiency, multijet background, other indirect effects linked to the lepton, like the \cancel{E}_T requirement which is corrected with the lepton energy).

4.7.3 Heavy flavor scale factor

As mentioned in 4.7.2, the W +jet background is normalized to data. However, under the assumption that the normalization scale factor may be different for W +light flavor and W +heavy flavor, this additional S_{HF} scale factor is derived. The S_{HF} factor is computed using the "0 tag" sample, where no b jets have been identified in the final state and orthogonal to the samples used to set limits in the WH analysis [103].

The cross section for the heavy flavor processes are obtained by multiplying the $Wb\bar{b}$, $Wc\bar{c}$, $Zb\bar{b}$ and $Zc\bar{c}$ ALPGEN cross sections by the factor K (described later in Section 4.7.2) and by the S_{HF}

	channel	2 jets	3 jets
RunIIa	Electron	1.10 ± 0.01	1.22 ± 0.03
	Muon	1.17 ± 0.01	1.36 ± 0.03
RunIIb	Electron	1.05 ± 0.01	1.14 ± 0.01
	Muon	1.11 ± 0.01	1.22 ± 0.01

Table 4.9: The experimental K_{LF}^{exp} factors for each sample, taking into account the theoretical K factor of 1.3. Errors are statistical only. The total uncorrelated systematic uncertainty between the e and μ determinations is approximately 7-8%, based on trigger (3-4% e , 5% μ) and lepton ID (3% e , 4% RunIIa μ , 2% RunIIb μ) uncertainties.

Heavy Flavor factor which is derived from data. With these conventions, the experimental K factor for heavy flavor is given by

$$K_{HF}^{exp} = S_{HF} \times K_{LF}^{exp}. \quad (4.12)$$

The S_{HF} factor is determined from two orthogonal samples, the tagged sample (denoted by ') and the 0 tag sample (denoted by ''), using the following equation:

$$S_{HF} = \frac{(Data' - X') \times W'' - (Data'' - X'') \times W'}{(Data'' - X'') \times B' - (Data' - X') \times B''} \quad (4.13)$$

where $W(B)$ is the number of events in the W +light jets ($Wb\bar{b}$ and $Wc\bar{c}$) inclusive sample, and X is the number of events in the MC background sample except W +jets (i.e. $t\bar{t}$, single top and diboson processes).

All channels are used to derive the S_{HF} factor and obtain $S_{HF} = 1.00$ for the luminosity weighted average, which is applied to the heavy flavor samples ($Wb\bar{b}$, $Wc\bar{c}$, $Zb\bar{b}$ and $Zc\bar{c}$). An uncertainty of 20% is assigned and S_{HF} will be allowed to vary at the limit settings level.

	Electron	muon
RunIIa	0.78 ± 0.09	0.99 ± 0.11
RunIIb	1.14 ± 0.06	1.02 ± 0.06

Table 4.10: The S_{HF} heavy flavor factor in the 0-tag sample, applied on top of the theoretical K factors listed in Section 4.3.2. The error is only data statistical errors is considered. The luminosity weighted average is 1.00.

After all corrections and selection cuts are applied, a good agreement between data and Monte Carlo is observed. These distributions are shown in Figure 4.12 for lepton related variables¹ along with the missing transverse energy distributions. Using these informations, one can construct W candidates distributions, such as the transverse mass calculated as:

$$M_W^T = \sqrt{(\cancel{E}_T + p_T^\ell)^2 - (\cancel{E}_x + p_x^\ell)^2 - (\cancel{E}_y + p_y^\ell)^2}. \quad (4.14)$$

¹The “phi-hole” is clearly visible in the muon ϕ distributions.

Figure 4.13 shows distributions related to the leading and second leading jet. Using both jets informations, the dijet system can be described by variables such as:

- the dijet invariant mass $M_{jj} = \sqrt{(E^{j_1} + E^{j_2})^2 - (p_x^{j_1} + p_x^{j_2})^2 - (p_y^{j_1} + p_y^{j_2})^2 - (p_z^{j_1} + p_z^{j_2})^2}$, constructed from the two leading jets in the event,
- the scalar sum of the transverse momentum of jets H_T ,
- the separation in the (η, ϕ) plane between the two leading jet $\Delta R(j_1, j_2)$,
- the difference in pseudorapidity of the two leading jets $\Delta\eta(j_1, j_2)$,
- the difference in azimuthal angle of the two leading jets $\Delta\phi(j_1, j_2)$,
- the transverse momentum of the dijet system $p_T(j_1, j_2)$,

shown for the 2 jets channels in Figure 4.14.

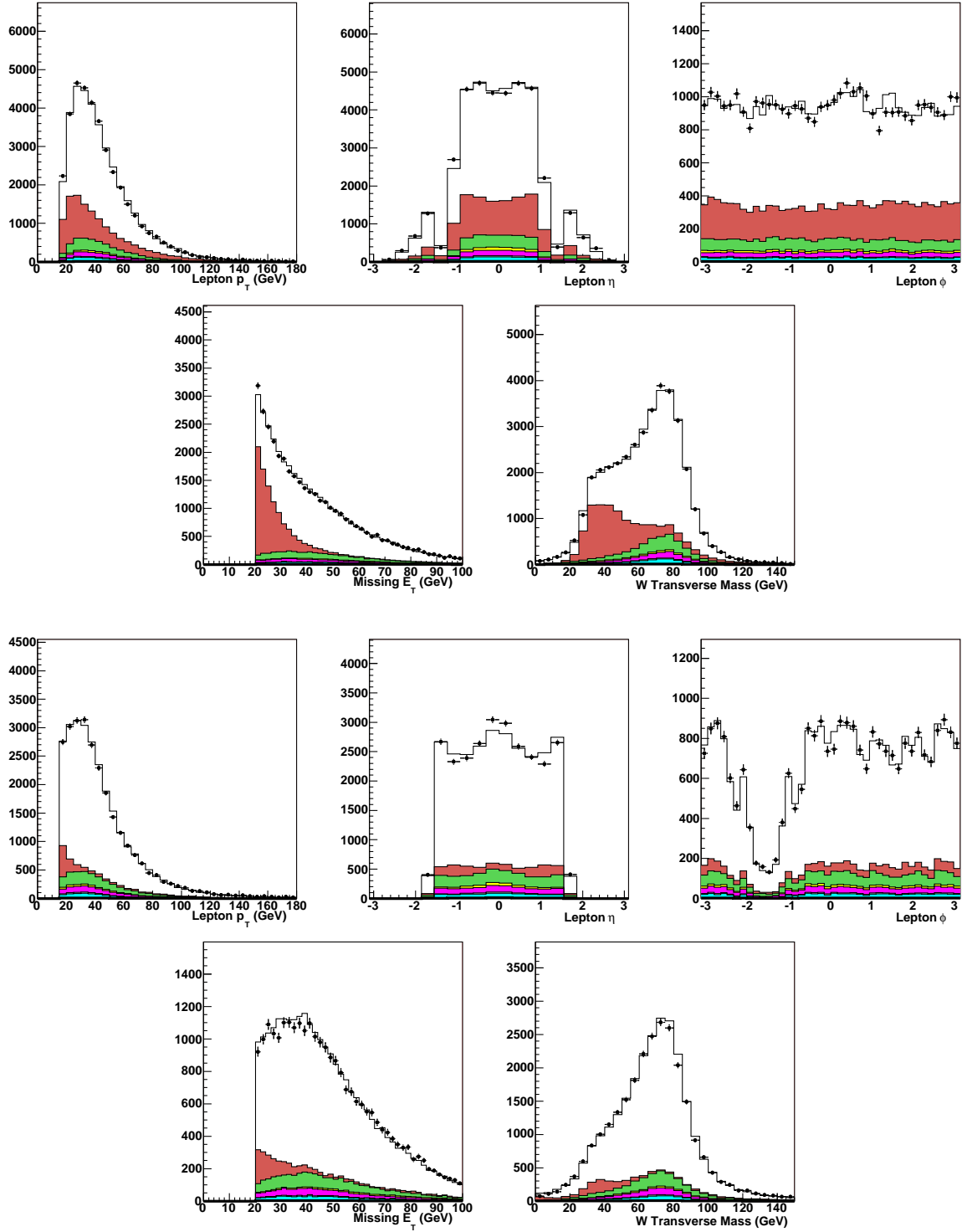


Figure 4.12: The upper five figures show the lepton p_T , lepton η , lepton ϕ , E_T and M_T^W distributions, for the 2 jets RunIIb electron channel. The bottom five figures show the same variables in the RunIIb muon channel.

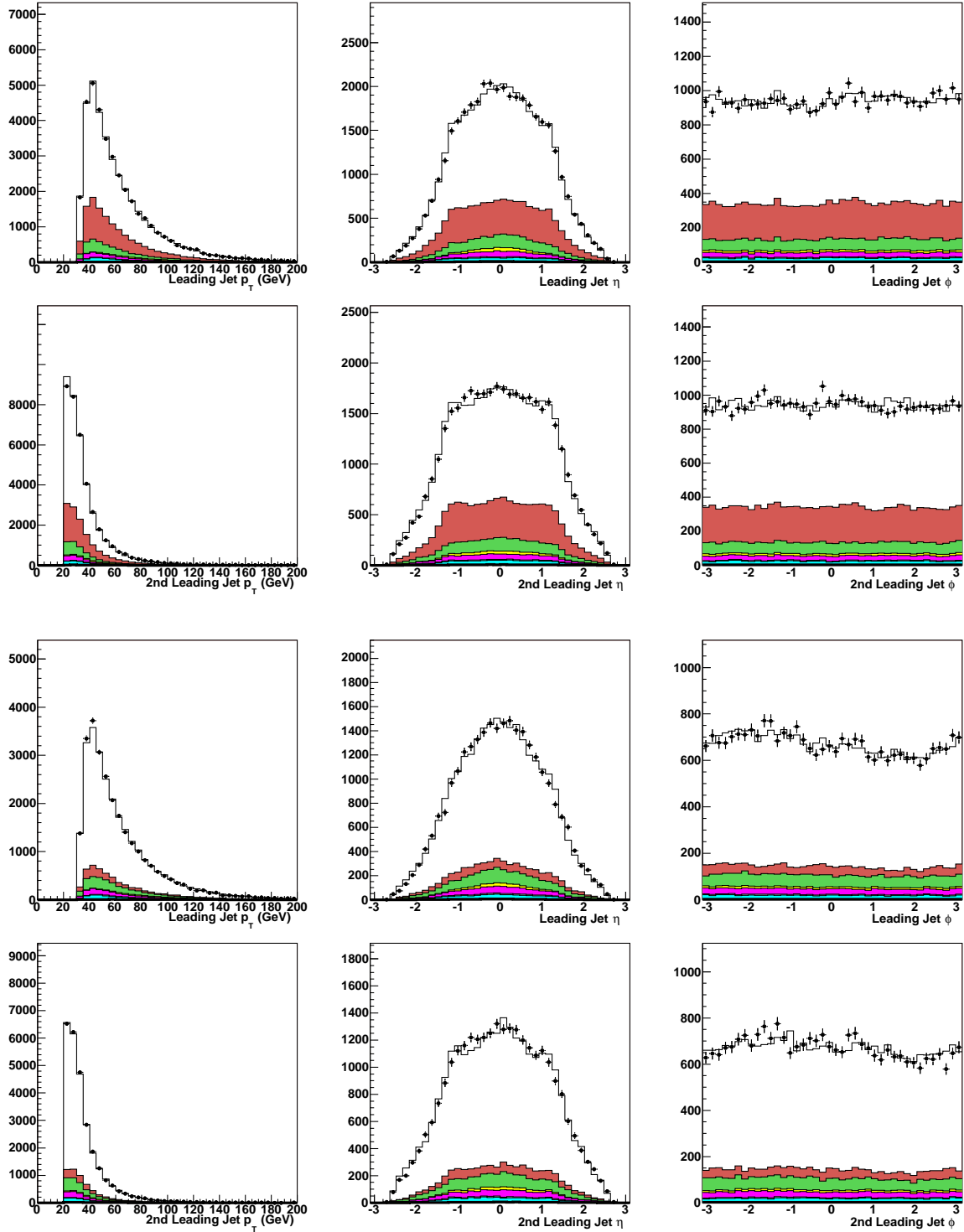


Figure 4.13: The upper six figures show the p_T^{jet} , η^{jet} and ϕ^{jet} distributions for the leading and second leading jet, in the 2 jets RunIIb electron channel. The bottom six figures show the same variables in the RunIIb muon channel.

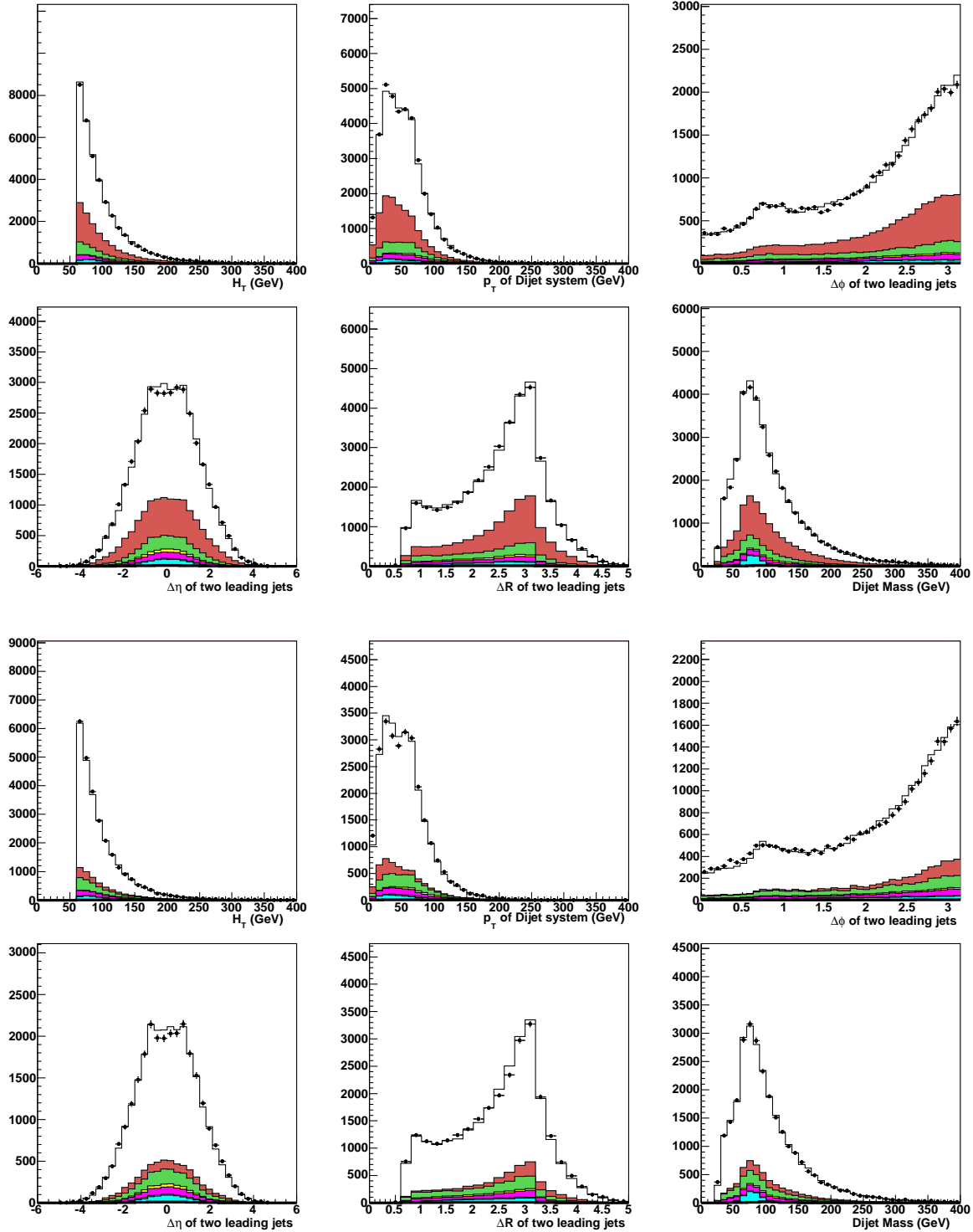


Figure 4.14: The upper six figures show the H_T , $p_T(j,j)$, $\Delta\phi(j,j)$, $\Delta\eta(j,j)$, $\Delta R(j,j)$ and M_{jj} distributions for the leading and second leading jet, in the 2 jets RunIIb electron channel. The bottom six figures show the same variables in the RunIIb muon channel.

4.8 Pre-tag summary

Our selection being done and validated by a good description of the data by the simulation, the event yields are given in Tables 4.11–4.12.

The number of events in the electron is systematically higher than in the muon channel due to the loss of acceptance from the “phi-hole” and a smaller pseudorapidity coverage than in the electron channel. Another noticeable difference between the two lepton flavors is the rate of multijet events which is higher in the electron since photons add a contribution of objects potentially as electrons. Moreover, the multijet background in the electron channel is higher in RunIIb due to the instantaneous luminosity increase, which impacts isolation performances. Since more jets are produced at high luminosity, semi-leptonic decays of jets will bring more fake isolated muons or jets and photons misidentified as electrons.

At this stage of the analysis, the S/B ratio is about $1/4000$, the dominant background Wjj representing about $2/3$ of the sample in the 2 jets channels. The next step which will be described will focus on reducing the contribution from light jets and multijet backgrounds and enhancing the signal content in the sample.

CHAPTER 4. SELECTION AND PHYSICS PROCESSES MODELING IN THE WH ANALYSIS

	$W(e) + 2 \text{ jets}$	$W(\mu) + 2 \text{ jets}$	$W(e) + 3 \text{ jets}$	$W(\mu) + 3 \text{ jets}$
WH	3.01 ± 0.33	2.45 ± 0.27	0.68 ± 0.08	0.58 ± 0.07
WW	266.45 ± 29.31	222.96 ± 24.53	60.13 ± 6.61	52.53 ± 5.78
WZ	43.96 ± 4.84	37.14 ± 4.09	10.95 ± 1.20	9.13 ± 1.00
ZZ	2.43 ± 0.27	2.14 ± 0.24	1.55 ± 0.17	0.60 ± 0.08
$Wb\bar{b}$	358.68 ± 71.74	316.62 ± 63.32	94.5 ± 18.90	89.63 ± 17.93
$Zb\bar{b}$	16.84 ± 3.37	8.93 ± 1.79	6.53 ± 1.31	3.01 ± 0.60
$Wc\bar{c}$	856.02 ± 171.20	771.01 ± 154.20	243.43 ± 48.69	237.63 ± 47.53
$Zc\bar{c}$	30.84 ± 6.17	21.37 ± 4.27	13.62 ± 2.72	6.84 ± 1.37
$t\bar{t}$	126.57 ± 17.72	80.29 ± 11.24	201.56 ± 28.22	151.45 ± 21.20
Single top	62.13 ± 8.08	52.35 ± 6.81	18.01 ± 2.34	15.48 ± 2.01
Multijet	2048.58 ± 184.37	457.19 ± 41.15	662.18 ± 59.60	132.01 ± 11.88
$W + \text{jets (light)}$	8225.6 ± 740.30	7406.36 ± 666.57	1360.77 ± 122.47	1365.48 ± 122.89
$Z + \text{jets (light)}$	373.89 ± 33.65	178.64 ± 16.08	83.77 ± 7.54	36.20 ± 3.26
Total expectation	12412.00 (n.t.d.)	9555.00 (n.t.d.)	2757.00 (n.t.d.)	2100.00 (n.t.d.)
Observed Events	12412	9555	2757	2100

Table 4.11: Summary table for the 2 jets and 3 jets final states for RunIIa. Observed events in data are compared to the expected numbers before tagging. The signal expectation is for $M_H = 115$ GeV. Quoted errors are from statistical source only. “n.t.d.” stands for “normalized to data”.

	$W(e) + 2 \text{ jets}$	$W(\mu) + 2 \text{ jets}$	$W(e) + 3 \text{ jets}$	$W(\mu) + 3 \text{ jets}$
WH	10.66 ± 1.17	8.18 ± 0.90	2.14 ± 0.24	1.68 ± 0.19
WW	819.06 ± 90.10	628.50 ± 69.14	176.06 ± 19.37	118.03 ± 12.98
WZ	134.36 ± 14.78	107.17 ± 11.79	32.84 ± 3.61	20.33 ± 2.24
ZZ	5.03 ± 0.55	7.95 ± 0.88	1.86 ± 0.22	1.90 ± 0.22
$Wb\bar{b}$	1099.18 ± 219.84	963.55 ± 192.71	256.11 ± 51.22	224.08 ± 44.82
$Zb\bar{b}$	31.25 ± 6.25	33.75 ± 6.75	9.51 ± 1.90	9.64 ± 1.93
$Wc\bar{c}$	2637.17 ± 527.43	2035.37 ± 407.07	710.24 ± 142.05	489.05 ± 97.81
$Zc\bar{c}$	60.86 ± 12.17	73.94 ± 14.79	22.09 ± 4.42	19.38 ± 3.88
$t\bar{t}$	496.90 ± 69.57	336.94 ± 47.17	681.66 ± 95.43	518.42 ± 72.58
Single top	196.40 ± 25.53	165.66 ± 21.54	53.03 ± 6.89	46.48 ± 6.04
Multijet	8319.60 ± 748.76	1477.89 ± 133.01	2084.97 ± 187.65	353.38 ± 31.80
$W + \text{jets (light)}$	23416.9 ± 2107.52	20357.1 ± 1832.14	3458.06 ± 311.23	3112.87 ± 280.16
$Z + \text{jets (light)}$	496.15 ± 44.65	615.15 ± 55.36	93.57 ± 8.42	96.44 ± 8.68
Total expectation	37713.00 (n.t.d.)	26803.00 (n.t.d.)	7580.00 (n.t.d.)	5010.00 (n.t.d.)
Observed Events	37713	26803	7580	5010

Table 4.12: Summary table for the 2 jets and 3 jets final states for RunIIb. See caption of Table 4.11.

4.9 b -tagging applied in the WH analysis

The main characteristic of a low mass Higgs boson is its resonant decay in $b\bar{b}$. By identifying both jets as originating from b quark, one can significantly increase the sensitivity of the search in the $WH \rightarrow \ell\nu b\bar{b}$ channel. By requiring 2 b -tagged jets (also referred as *double-tagged events*), the dominant V + light jets ($V = W$ or Z) can be reduced, as well as the $V + c\bar{c}$ and multijet backgrounds. The remaining dominant backgrounds are $V + b\bar{b}$ and $t\bar{t}$.

By requiring 2 b -tagged jets, the light jets background can be significantly reduced. In order to keep as much signal events passing the NN b -tagger requirements as possible, the relatively loose operating point *OldLoose* is chosen. Since the single b -tagged sample has a higher contamination from light jets backgrounds, the *Tight* NN b -tagger operating point is optimal in order to improve the signal sensitivity by reducing multijet and light jets backgrounds.

Although the *OldLoose* operating point has an overall efficiency close to 60% and fake rate of about 1.7% for jets in the central calorimeter, one can still increase the signal acceptance by requiring only one b -tagged jet, analyzed in a separate channel, if not fulfilling the double-tag requirements. This orthogonal sample has a reduced sensitivity compared to the double-tag sample, but can provide in addition about 10% sensitivity. However the background rate is still high and the dominant backgrounds are events containing heavy flavor jet such as $V + b\bar{b}$, $V + c\bar{c}$ and $t\bar{t}$, but the V + light jets remains the most important one.

The same b -tagging requirements are applied to both data and simulation. Only the 2 leading jets are considered as taggable. If both jets satisfy the *OldLoose* requirements, it is tagged, an event enters the double b -tagged sample. If not, one require one jet with a *Tight* NN b -tagger cut to populate the orthogonal single b -tagged sample. If such a jet is not found, the event is dismissed, and will not be considered downstream the analysis.

The following procedure is applied to MC in order to take in consideration the difference with data due to detector simulation and correct the tagging efficiency in simulation:

- Since taggability is higher for MC than data due to tracking differences in simulation, the taggability in simulation is corrected.
- The b -tagging efficiency scale factors are applied to MC.

Each step of the procedure will be described in this section.

4.9.1 Taggability scale factors

Since the detector isn't perfectly modeled in the simulation, taggability is affected by this difference and is higher for MC than in data. If no correction is applied at the analysis level, one would observe higher yields in MC than in data. Therefore, one need to take into account this difference and accordingly correct the MC to match the data. A taggability scale factor is measured with W +2 jets events and parametrized as function of jet η (and jet p_T for RunIIa only, since no dependency is found for RunIIb) in 4 regions of vertex position along the z direction ($-40 < z_{vtx} < -30$ cm, $-30 < z_{vtx} < 0$ cm, $0 < z_{vtx} < 30$ cm and $30 < z_{vtx} < 40$ cm). It is computed by taking the

ratio of taggability in data to taggability in simulation. The parametrizations are shown for RunIIb (Figures 4.15).

The procedure employed is checked to be correct by performing a closure test. The ratio of the data taggability divided by the simulated taggability corrected by the scale factor is plotted, as a function of p_T^{jet} and η^{jet} . If the simulation has been corrected properly in all kinematic and spatial regions, the result must be flat, at unity. The result is shown in Figures 4.16, for both the electron and muon channels. The closure test fit results are summarized in Table 4.13, which shows that all closure test fits are consistent with unity within 1σ .

		$-40 < z_{vtx} < -30$ cm	$-30 < z_{vtx} < 0$ cm	$0 < z_{vtx} < 30$ cm	$30 < z_{vtx} < 40$ cm
RunIIa	p_T	1.002 ± 0.035	1.012 ± 0.013	1.003 ± 0.013	1.031 ± 0.039
	η	1.000 ± 0.036	1.010 ± 0.010	1.001 ± 0.013	1.023 ± 0.039
RunIIb	p_T	1.000 ± 0.025	1.001 ± 0.008	1.003 ± 0.008	1.020 ± 0.030
	η	1.000 ± 0.024	1.001 ± 0.008	1.003 ± 0.008	1.015 ± 0.026

Table 4.13: Taggability closure test fit results. Column headings show the range of primary vertex z -values used for the fit. Numbers represent the result of fitting the ratio of data to taggability-modified MC to a constant, with uncertainty.

4.9.2 b -tagging efficiency scale factors

Since the b -tagging efficiency is different in data and simulation, this has to be corrected in the simulation. b -tagging efficiency scale factors are provided by the b-ID group. These *Tagging Scale Factors* (TSF) are derived as:

$$TSF = \frac{\varepsilon_{data}}{\varepsilon_{MC}}, \quad (4.15)$$

where ε_{data} (ε_{MC}) is the efficiency in data (MC) which depends on the chosen operating point. The DØ collaboration has developed a method to make an unbiased measure of the efficiency on relying mainly on data. The so-called *System8* method [104] makes use of samples having different b -jets composition. By using an independent tagger, the SLT tagger, one can form and resolve a system of 8 equations with 8 unknowns. In the case of heavy flavor jets (jets originating from b or c quark), $\varepsilon_{data,MC}$ is their b -tagging efficiencies in data or MC. For light jets, this variable is the mis-identification rate in data or MC.

The TSFs are parametrized as function of p_T^{jet} in 3 calorimeter regions (CC, ICR and EC), for each b-ID operating point. There are also different according to the jet flavour. An example of the TSF for b -jets is given in Figures 4.17a and 4.17b.

These scale factors are applied on tagged events as follows:
The scale factor weight for double tagged events, DT_{weight} is

$$DT_{weight} = TSF_{j1} \times TSF_{j2} \quad (4.16)$$

where TSF_{j1} and TSF_{j2} are the scale factors for the leading and second leading jet, respectively, parameterized in p_T , η and jet flavor of the jet.

The scale factor weight (ST_{weight}) for the single tagged events is determined by:

$$ST_{weight} = ST_{weight}^{STevent} + ST_{weight}^{DTevent} \quad (4.17)$$

where $ST_{weight}^{STevent}$ is the event weight calculated for the single exclusively tagged event, and $ST_{weight}^{DTevent}$ is the contribution from double tagged events that can migrate (due to the scale factor) into the single tagged sample. The single tagged event weights are calculated as

$$ST_{weight}^{STevent} = T SF_j \quad (4.18)$$

where SF_j is the scale factor for the tagged jet in the single tagged event. The scale factor for the other jet to migrate below the operating point of double tagged sample is given by:

$$ST_{weight}^{DTevent} = (1 - SF_j^{DT}) \times SF_j. \quad (4.19)$$

where SF_j is the scale factor of the single tag operating point and SF_j^{DT} is the scale factor for the jet with the double tagged operating point. The migration from a double tagged event to a single tagged event will happen if one jet fails to be tagged, in which case a factor of $(1 - SF_j^{DT})$ from the other jet will be applied. This parametrization is valid only for the case where the double tag operating point is looser than the single tag operating point (which is the case in this analysis, OldLoose for double tag and Tight for single tag).

The total *b*-tagging scale factors corresponds to the product of the *b*-ID scale factor and the taggability scale factors. A simple example is presented in [105].

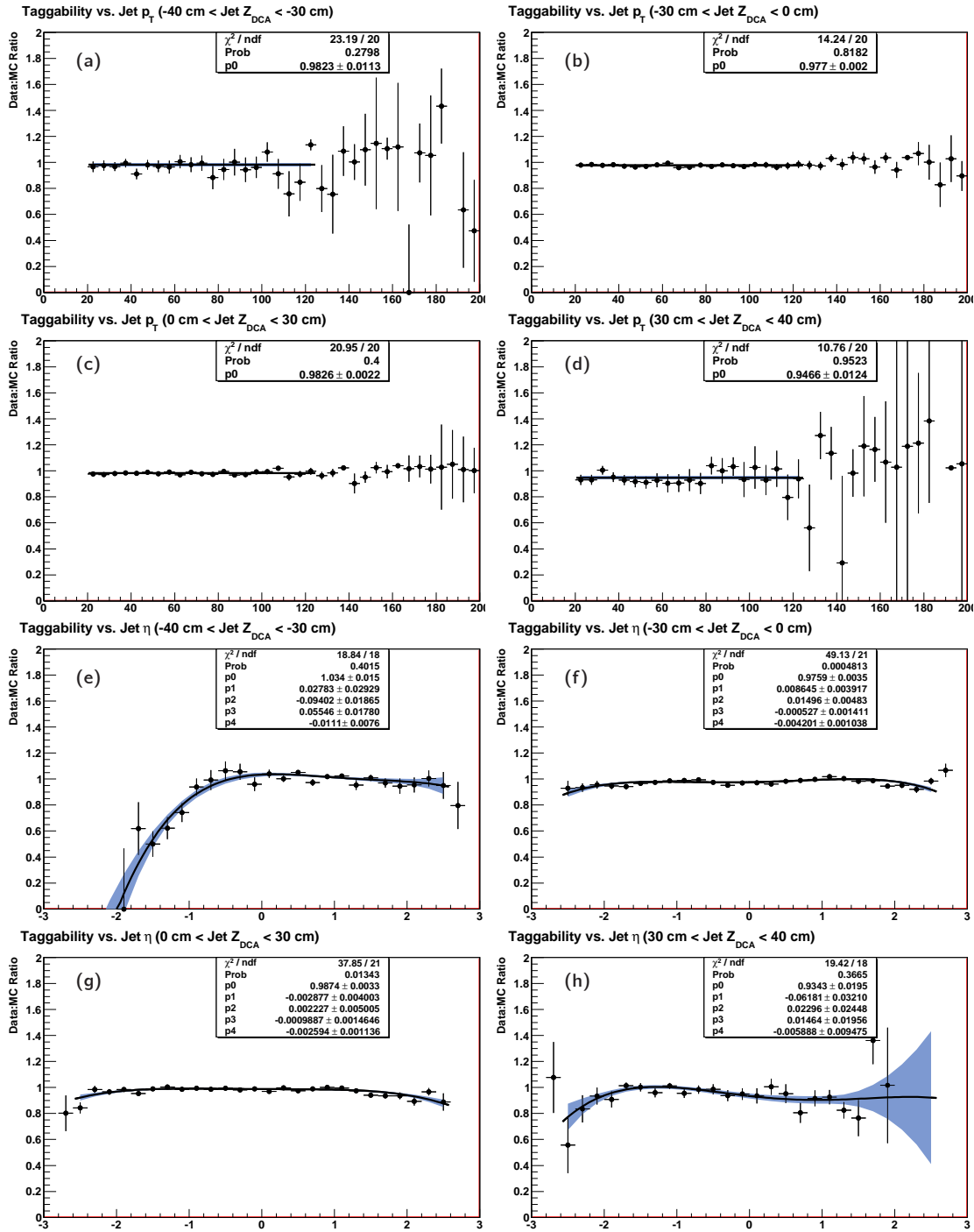


Figure 4.15: Jet taggability scale factors for muon channel as a function of p_T^{jet} ((a), (b), (c) and (d)) and η^{jet} ((e), (f), (g) and (h)) from the data and simulated samples of the RunIIb analysis. The parameterizations are shown for (a,e) $-40 < z_{vtx} < -30$ cm, (b,f) $-30 < z_{vtx} < 0$ cm, (c,g) $0 < z_{vtx} < 30$ cm and (d,h) $30 < z_{vtx} < 40$ cm. The black line represents the nominal fit and the blue band represents the $\pm 1\sigma$ uncertainty band on the fit.

4.9. B -TAGGING APPLIED IN THE WH ANALYSIS

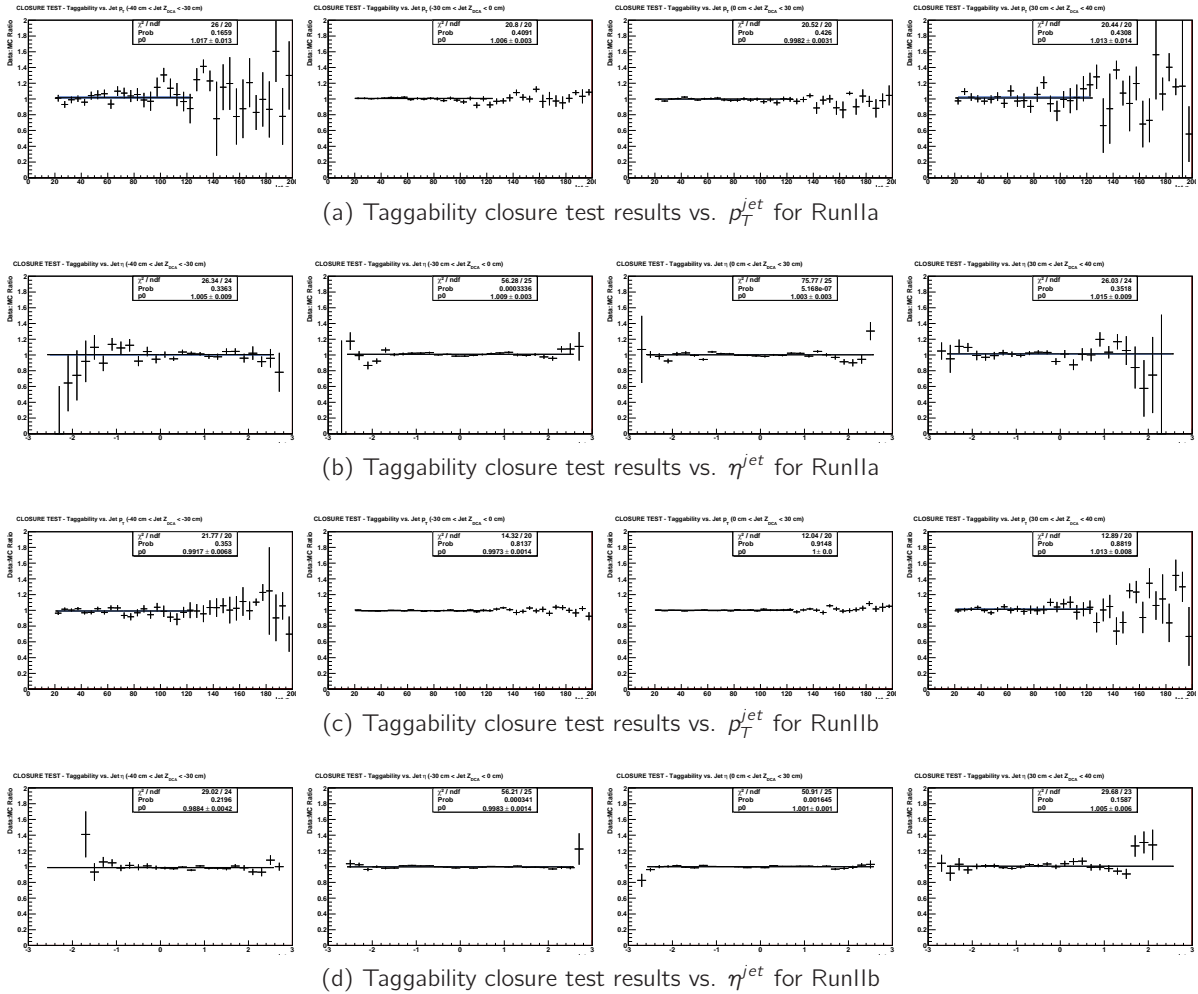
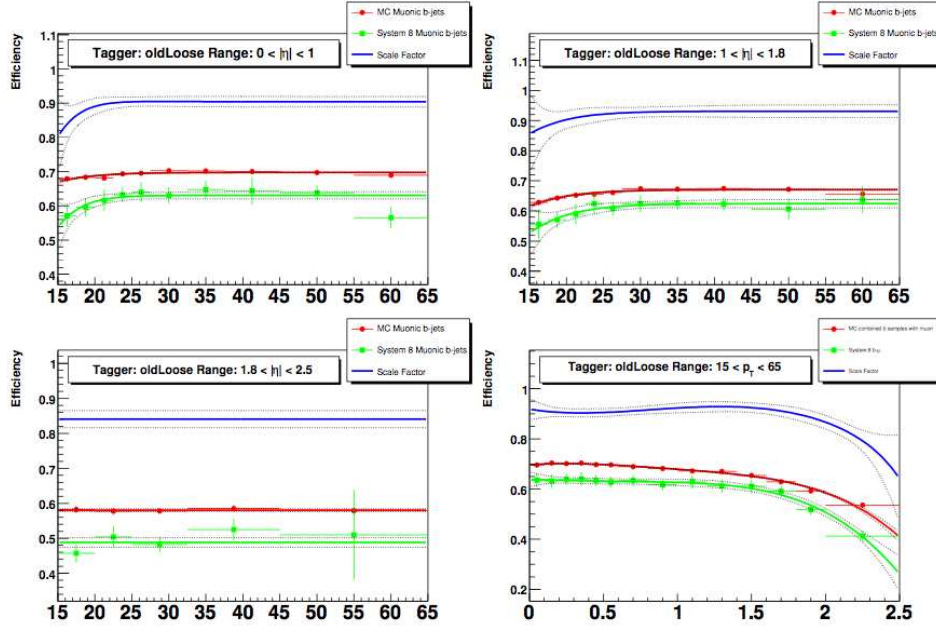
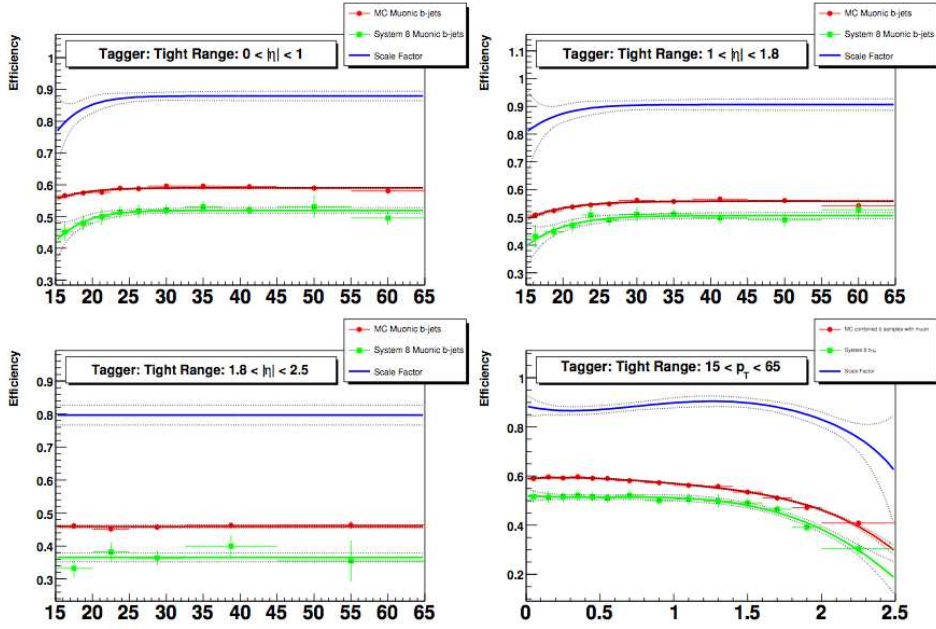


Figure 4.16: Taggability closure test results, for RunIIa (a,b) and RunIIa (c,d), parametrized in p_T^{jet} and η^{jet} for the $-40 < z_{vtx} < -30$ cm, $-30 < z_{vtx} < 0$ cm, $0 < z_{vtx} < 30$ cm, and $30 < z_{vtx} < 40$ cm bins (from left to right). The black line represents the fit to a constant.



(a) OldLoose operating point.



(b) Tight operating point.

Figure 4.17: b -jet efficiency measured in data (green) and MC (red) along with the data/MC scale factor (blue) as a function of jet p_T in the CC (top left), ICR (top right) and EC (bottom left) and as a function of η (bottom right), for the OldLoose operating point 4.17a and Tight operating point 4.17b.

4.9.3 Post *b*-tagging results

After correcting the taggability and *b*-tagging efficiency in the simulation, data and simulation are compared with all corrections previously discussed: efficiency correction for objects reconstruction, reweighting of kinematic distributions, experimental and theoretical scale factors. A summary of event yields is given in Tables 4.14–4.17.

Distributions for jet related variables are shown in Figures ??–4.21, with the signal contribution magnified by a factor 10.

One can observe that after *b*-tagging being applied, the expectation from MC agrees well with data, either in terms of number of events, within the statistical uncertainty, or distribution shapes. Considering first the 2 jets channels, the dominant backgrounds are *Wjj* and multijet events in the single *b*-tagged electron channel (the ratio between the expected number of *Wjj* and Multijet events is close to 1 compared to the pre-tag level which was about 3), as opposed to the muon channel where the multijet background has a lower rate. Even if *Wjj* is still the main background, requiring only one jet to be *b*-tagged lowers its rate by a factor 50.

After requiring 2 *b*-tagged jets, the background composition changes accordingly to the *b* quark content and cross section of the processes. Therefore, the *Wb \bar{b}* and *t \bar{t}* backgrounds become dominant.

To illustrate the power of the *b*-tagger tool employed, one can notice that the expectation from single top processes is of the same amount as the *Wjj* process which has a cross section about 6000 times higher: starting from the pre-tag sample, the associated production of *W* and light jets is reduced by a factor 1000.

For the 3 jets channel case, higher multiplicities backgrounds dominates, clearly visible as the *t \bar{t}* represents about 2/3 of the total background expectation.

In total, about 19 signal events are expected after combining all channels for $M_H = 115$ GeV. The following step in the analysis will be to separate the signal from background contributions in order to increase the discriminating power of the analysis.

	$W(e) + 2 \text{ jets}$		$W(\mu) + 2 \text{ jets}$	
	(1 b jet)	(2 b jets)	(1 b jet)	(2 b jets)
WH	1.25 ± 0.14	0.73 ± 0.08	1.02 ± 0.11	0.62 ± 0.07
WW	13.35 ± 1.47	0.36 ± 0.12	10.50 ± 1.16	0.21 ± 0.22
WZ	4.79 ± 0.53	2.17 ± 0.24	3.86 ± 0.43	1.89 ± 0.21
ZZ	0.29 ± 0.07	0.07 ± 0.14	0.23 ± 0.06	0.11 ± 0.10
$Wb\bar{b}$	124.26 ± 24.85	33.59 ± 6.72	107.56 ± 21.51	29.93 ± 5.99
$Zb\bar{b}$	5.38 ± 1.08	0.63 ± 0.14	3.26 ± 0.65	0.58 ± 0.13
$Wc\bar{c}$	68.07 ± 13.61	5.31 ± 1.06	54.81 ± 10.96	4.76 ± 0.95
$Zc\bar{c}$	2.07 ± 0.42	0.04 ± 0.25	1.38 ± 0.28	0.06 ± 0.32
$t\bar{t}$	51.97 ± 7.28	20.53 ± 2.87	32.97 ± 4.62	13.37 ± 1.87
Single top	26.39 ± 3.43	6.65 ± 0.86	22.54 ± 2.93	5.50 ± 0.72
Multijet	76.76 ± 6.91	5.07 ± 0.47	38.55 ± 3.47	1.74 ± 0.20
Wjj	141.73 ± 12.76	2.11 ± 0.19	122.56 ± 11.03	4.04 ± 0.38
Zjj	4.58 ± 0.42	0.19 ± 0.28	2.83 ± 0.26	0.00 ± 0.00
Total expectation	519.62 ± 32.90	76.73 ± 7.46	401.05 ± 27.37	62.19 ± 6.41
Observed Events	479	74	400	62

Table 4.14: Summary table for the 2 jets final state for RunIIa. Observed events in data are compared to the expected number of $W + 2$ jet, with one and two b -tagged jets. The signal expectation is for $M_H = 115$ GeV. Quoted errors are from statistical source only.

	$W(e) + 3 \text{ jets}$		$W(\mu) + 3 \text{ jets}$	
	(1 b jet)	(2 b jets)	(1 b jet)	(2 b jets)
WH	0.29 ± 0.03	0.09 ± 0.03	0.24 ± 0.03	0.09 ± 0.04
WW	2.91 ± 0.32	0.11 ± 0.21	3.16 ± 0.35	0.07 ± 0.41
WZ	1.18 ± 0.14	0.37 ± 0.11	0.98 ± 0.12	0.26 ± 0.10
ZZ	0.13 ± 0.10	0.02 ± 0.32	0.06 ± 0.11	0.02 ± 0.23
$Wb\bar{b}$	28.47 ± 5.69	4.69 ± 0.94	26.55 ± 5.31	3.58 ± 0.72
$Zb\bar{b}$	1.90 ± 0.38	0.26 ± 0.10	1.00 ± 0.20	0.21 ± 0.10
$Wc\bar{c}$	17.65 ± 3.53	1.42 ± 0.28	15.62 ± 3.12	1.07 ± 0.23
$Zc\bar{c}$	0.84 ± 0.18	0.01 ± 0.44	0.68 ± 0.17	0.00 ± 0.00
$t\bar{t}$	82.52 ± 11.55	22.07 ± 3.09	63.09 ± 8.83	16.97 ± 2.38
Single top	7.27 ± 0.95	1.60 ± 0.21	6.34 ± 0.82	1.40 ± 0.18
Multijet	30.13 ± 2.71	1.70 ± 0.22	11.28 ± 1.02	0.63 ± 0.23
Wjj	22.29 ± 2.01	0.33 ± 0.03	21.32 ± 1.92	0.95 ± 0.09
Zjj	1.14 ± 0.13	0.03 ± 0.44	0.78 ± 0.12	0.01 ± 0.63
Total expectation	196.43 ± 13.81	32.60 ± 3.26	150.85 ± 11.02	25.17 ± 2.52
Observed Events	178	32	137	25

Table 4.15: Summary table for the 3 jets final state for RunIIa dataset. Observed events in data are compared to the expected number of $W + 3$ jet, with one and two b -tagged jets. The signal expectation is for $M_H = 115$ GeV. Quoted errors are from statistical source only.

4.9. *B*-TAGGING APPLIED IN THE *WH* ANALYSIS

	<i>W</i> (<i>e</i>) + 2 jets		<i>W</i> (μ) + 2 jets	
	(1 <i>b</i> jet)	(2 <i>b</i> jets)	(1 <i>b</i> jet)	(2 <i>b</i> jets)
<i>WH</i>	4.21 ± 0.46	2.83 ± 0.31	3.23 ± 0.36	2.32 ± 0.26
<i>WW</i>	48.56 ± 5.34	1.77 ± 0.26	41.76 ± 4.59	1.73 ± 0.29
<i>WZ</i>	15.30 ± 1.68	7.47 ± 0.82	12.67 ± 1.39	6.24 ± 0.69
<i>ZZ</i>	0.56 ± 0.11	0.25 ± 0.17	0.72 ± 0.11	0.22 ± 0.14
<i>Wb\bar{b}</i>	385.74 ± 77.15	117.8 ± 23.56	348.44 ± 69.69	101.54 ± 20.31
<i>Zb\bar{b}</i>	10.09 ± 2.02	2.13 ± 0.43	11.89 ± 2.38	3.05 ± 0.61
<i>Wc\bar{c}</i>	258.79 ± 51.76	22.59 ± 4.52	207 ± 41.40	22.68 ± 4.54
<i>Zc\bar{c}</i>	5.33 ± 1.07	0.83 ± 0.23	7.11 ± 1.42	0.65 ± 0.22
<i>t\bar{t}</i>	196.71 ± 27.54	84.47 ± 11.83	135.36 ± 18.95	58.69 ± 8.22
Single top	82.81 ± 10.77	25.35 ± 3.30	71.32 ± 9.27	20.76 ± 2.70
Multijet	397.26 ± 35.75	39.65 ± 3.57	149.89 ± 13.49	10.00 ± 0.90
<i>Wjj</i>	535.17 ± 48.17	25.36 ± 2.28	457.24 ± 41.15	24.67 ± 2.22
<i>Zjj</i>	15.50 ± 1.40	0.65 ± 0.44	10.42 ± 0.94	0.45 ± 0.62
Total expectation	1951.82 ± 114.69	328.31 ± 27.30	1453.81 ± 94.49	250.68 ± 22.68
Observed Events	2002	325	1435	248

Table 4.16: Summary table for the 2 jets final state for RunIIb. Observed events in data are compared to the expected number of *W* + 2 jet, with one and two *b*-tagged jets. The signal expectation is for $M_H = 115$ GeV. Quoted errors are from statistical source only.

	<i>W</i> (<i>e</i>) + 3 jets		<i>W</i> (μ) + 3 jets	
	(1 <i>b</i> jet)	(2 <i>b</i> jets)	(1 <i>b</i> jet)	(2 <i>b</i> jets)
<i>WH</i>	0.86 ± 0.10	0.37 ± 0.05	0.69 ± 0.08	0.30 ± 0.05
<i>WW</i>	10.82 ± 1.19	0.00 ± 0.00	7.19 ± 0.80	0.05 ± 0.51
<i>WZ</i>	4.25 ± 0.47	0.95 ± 0.16	2.84 ± 0.32	0.67 ± 0.16
<i>ZZ</i>	0.19 ± 0.17	0.05 ± 0.35	0.18 ± 0.17	0.05 ± 0.37
<i>Wb\bar{b}</i>	78.68 ± 15.74	14.32 ± 2.86	72.93 ± 14.59	11.72 ± 2.34
<i>Zb\bar{b}</i>	2.81 ± 0.56	0.37 ± 0.15	2.74 ± 0.55	0.47 ± 0.15
<i>Wc\bar{c}</i>	64.89 ± 12.98	4.38 ± 0.88	39.68 ± 7.94	4.89 ± 0.98
<i>Zc\bar{c}</i>	1.80 ± 0.37	0.01 ± 0.50	1.79 ± 0.37	0.35 ± 0.31
<i>t\bar{t}</i>	274.84 ± 38.48	77.33 ± 10.83	212.20 ± 29.71	59.69 ± 8.36
Single top	20.93 ± 2.72	5.45 ± 0.71	19.04 ± 2.48	4.57 ± 0.59
Multijet	109.62 ± 9.87	7.36 ± 0.67	34.52 ± 3.11	2.99 ± 0.28
<i>Wjj</i>	88.56 ± 7.97	7.18 ± 0.66	74.49 ± 6.70	3.55 ± 0.35
<i>Zjj</i>	0.29 ± 0.36	0.00 ± 0.00	1.46 ± 0.29	0.00 ± 0.00
Total expectation	657.70 ± 45.48	117.41 ± 11.30	469.05 ± 34.94	89.01 ± 8.77
Observed Events	671	125	477	119

Table 4.17: Summary table for the 3 jets final state for RunIIb. Observed events in data are compared to the expected number of *W* + 3 jet, with one and two *b*-tagged jets. The signal expectation is for $M_H = 115$ GeV. Quoted errors are from statistical source only.

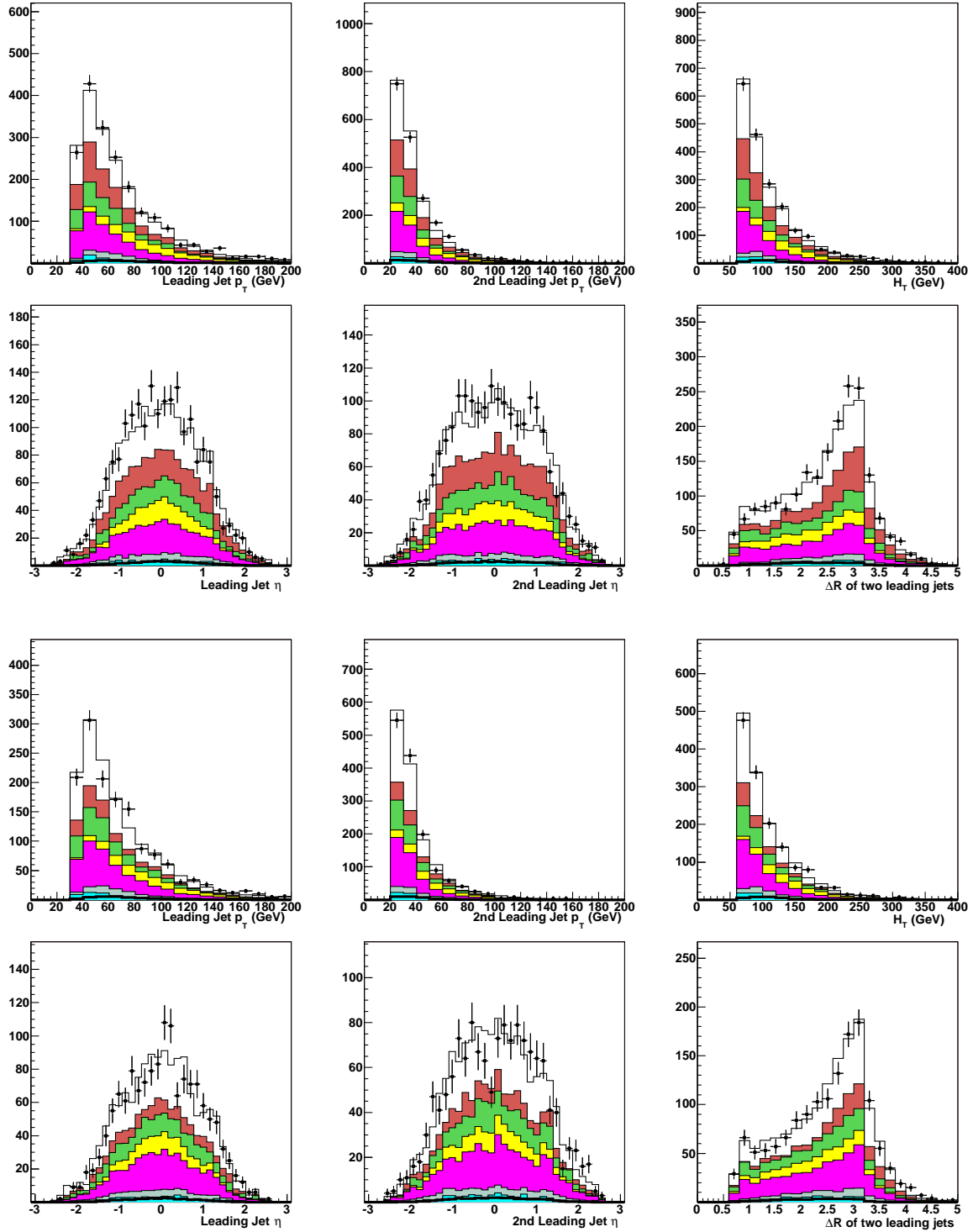


Figure 4.18: The upper six figures show the $p_T^{jet_1}$, $p_T^{jet_2}$, H_T , η^{jet_1} , η^{jet_2} and $\Delta R(j,j)$ distributions in the 2 jets RunIIb electron channel with 1 b -tagged jet. The bottom six figures show the same variables in the muon channel.

4.9. B -TAGGING APPLIED IN THE WH ANALYSIS

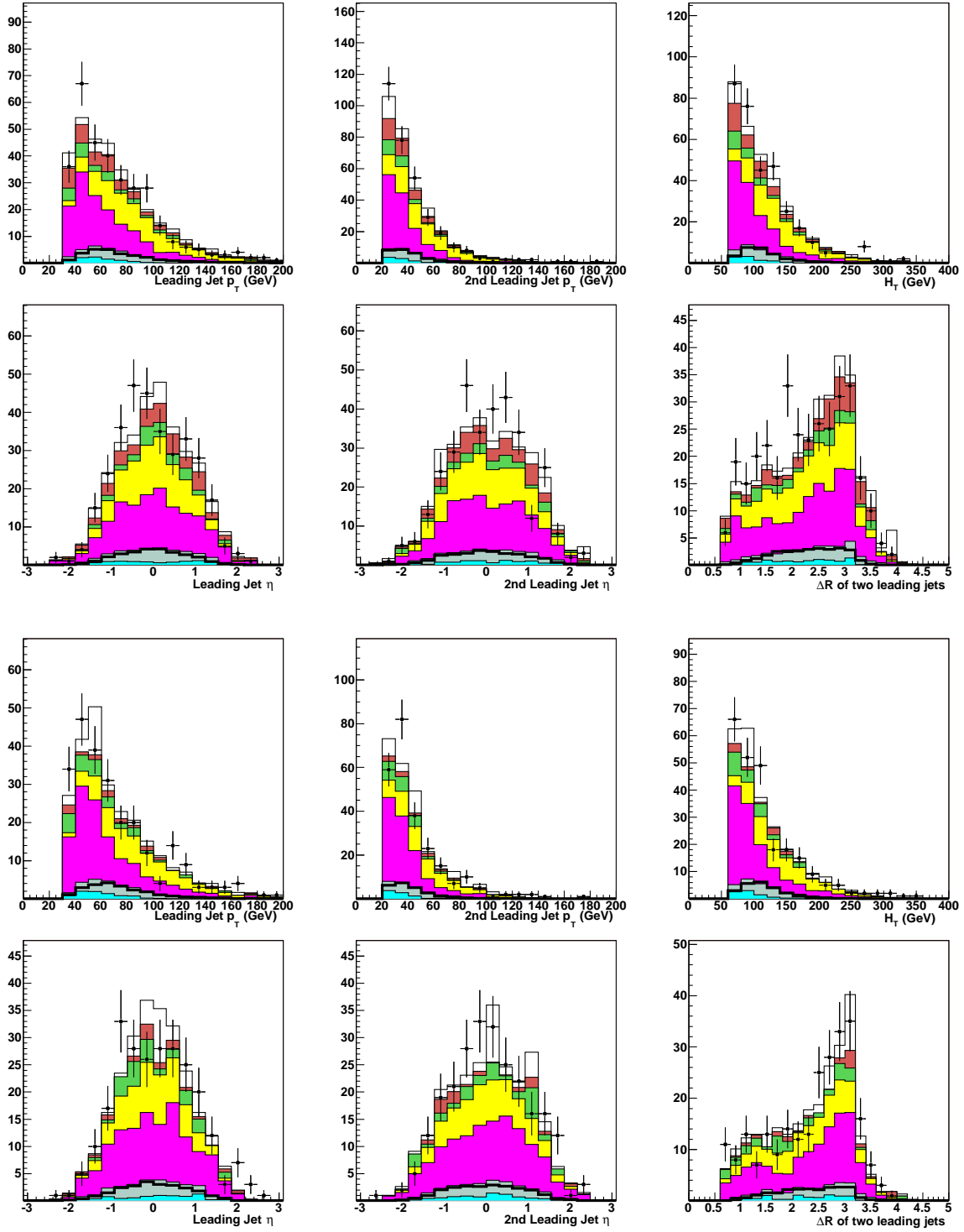


Figure 4.19: The upper six figures show the $p_T^{\text{jet}_1}$, $p_T^{\text{jet}_2}$, H_T , η^{jet_1} , η^{jet_2} and $\Delta R(j, j)$ distributions in the 2 jets RunIIb electron channel with 2 b -tagged jets. The bottom six figures show the same variables in the muon channel.

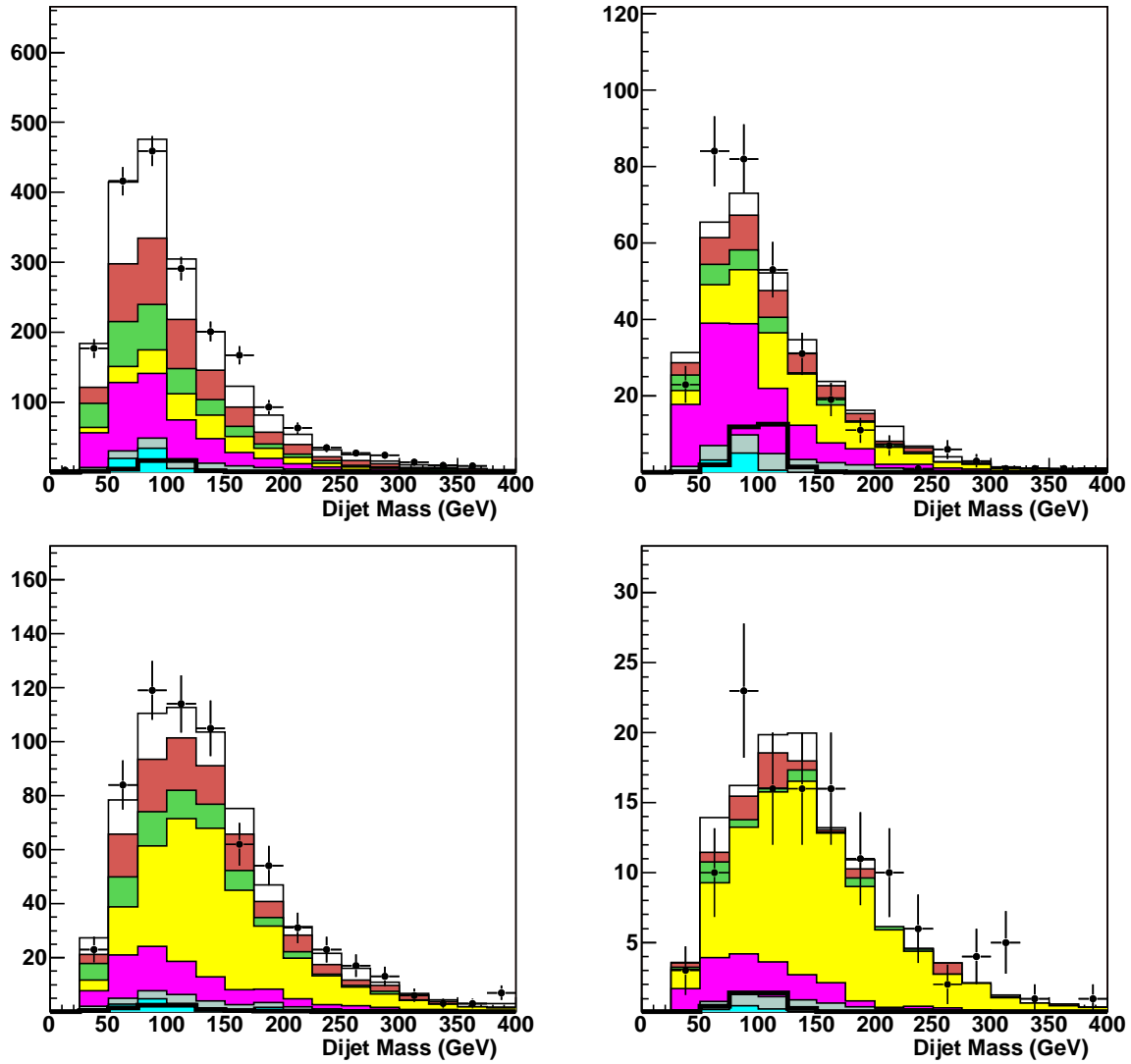


Figure 4.20: Dijet invariant mass distributions in the RunIIb electron channel, for the 2 (top figures) and 3 jets (bottom plots) channels with one (left) or two (right) b -tagged jets.

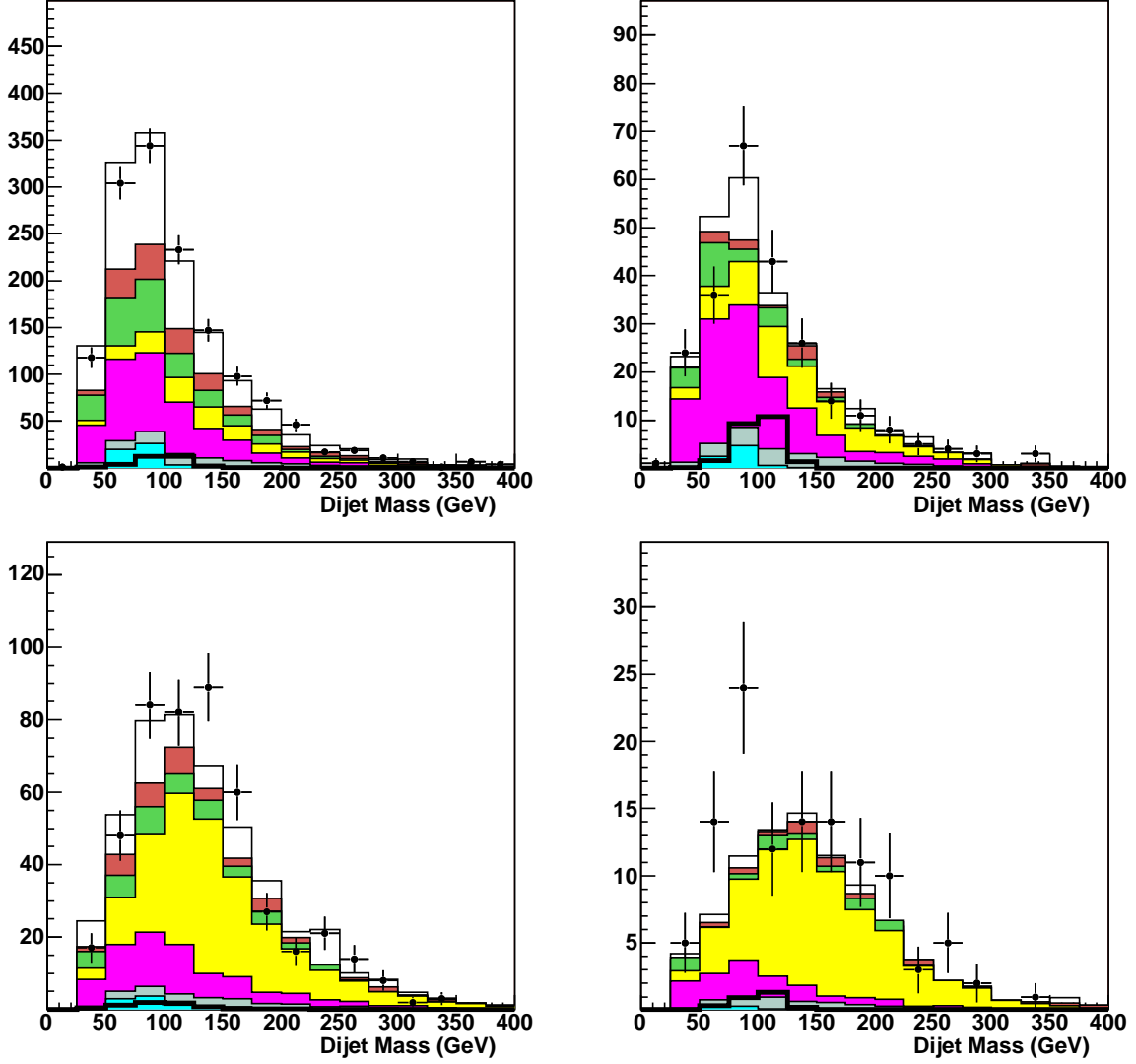


Figure 4.21: Dijet invariant mass distributions in the RunIIb muon channel, for the 2 (top figures) and 3 jets (bottom plots) channels with one (left) or two (right) *b*-tagged jets.

5

Multivariate classification, limit derivation and results for the WH analysis

The search for the $WH \rightarrow \ell \nu b \bar{b}$ signal is characterized by a S/B ratio of about 1/3500 before applying b -tagging. The sensitivity is improved after both leading jets being identified as b -jets, S/B increases to 1/100. One method to further increase the sensitivity is to provide maximal discrimination between signal and background events. These discriminating distributions are then used to extract the limits. A common method is to make use of multivariate analysis techniques, which combine information from several observables into a single discriminant distribution. This method is intended, in case of no signal excess being observed, to provide the most stringent limit on the $p\bar{p} \rightarrow WH \rightarrow \ell \nu b \bar{b}$ production.

Such techniques are widely and successfully used, not only in the DØ experiment. A previous iteration of the WH analysis [103] made use of the neural network technique and a sensitivity improvement of about 10% has been reached, with respect to the limit extraction from the dijet invariant mass which the most discriminating single variable in this search. Moreover, a more recent technique, the Random Forest method, has been introduced [106] and additional 15% improvement in the results is observed with respect to the previously mentioned result.

From Section 4, it has been observed that an overall good data to Monte Carlo agreement is achieved. However, a coherent choice of input variables to the Random Forest discriminator has to be made since some of the distributions may not be perfectly described by simulation, or variables which don not provide discriminating power may be discarded. A study has been carried to determine which variables to be used in the Random Forest has been carried by evaluating the Kolmogorov-Smirnov statistical test [107, 108, 109]. This allows to check the compatibility in shape between data and simulation distributions for a given variable. This optimization leads to the list of variables given in Table 5.1. Scaled to the same luminosity, an improvement greater than 10% is achieved with respect to the previous result of the WH analysis, which was based on

the neural network technique.

Due to low statistics samples, especially after b -tagging is applied, systematic uncertainties determination may suffer from large unphysical fluctuations. A coherent treatment has been developed and systematic uncertainties obtained by this method are used for the evaluation of the result.

In this section, the principle and usage in the WH analysis of the Random Forest technique will be described. Then, the limit setting procedure and the systematic uncertainties treatment will be detailed and finally, the results obtained for WH analysis will be presented.

5.1 Improving sensitivity using a Random Forest

A multivariate analysis technique (MVA), namely a neural network, has already been used in a previous publication of the $WH \rightarrow \ell\nu b\bar{b}$ analysis, instead of using the dijet mass distribution to set limit on the $WH \rightarrow \ell\nu b\bar{b}$ production cross section. Ultimately, multivariate analysis techniques are an approximation of

$$\Lambda = \frac{\mathcal{S}(\vec{x})}{\mathcal{S}(\vec{x}) + \mathcal{B}(\vec{x})}$$

for events characterized by N observables $\vec{x} = \{x_1, \dots, x_N\}$, $\mathcal{S}(\vec{x})$ ($\mathcal{B}(\vec{x})$) being the signal (background) probability function in the $\{x_1, \dots, x_N\}$ space.

The differences between different MVAs are how many variables they can be fed with, how well they can handle correlations between them and flexibility of the training algorithms.

Sensitivity improvement studies have been conducted in the WH analysis and a new method based on Decision Trees has been introduced, the Random Forest. Different software provide the usage of this technique, the one used here is from the *StatPatternRecognition* package [110]. More details on Random forests can be found in [111]. This technique has successfully been used for the first evidence of Diboson (WW , WZ , ZZ) production in the $\ell\nu jj$ decay mode at DØ [112, 113].

The principle of Random Forests will be explained, the choice of input variables and a rebinning procedure introduced to reduce possible statistical fluctuations will then be detailed.

5.1.1 Principle of a Random Forest

The purpose of multivariate techniques is to provide the best discrimination between two populations of events, based on characteristics or variables describing those events. In the case of the WH analysis, the populations one wants to separate are $WH + ZH$ signal events sample from the background events sample, using a set of variables x .

The Random Forest technique (RF) relies on a set of decision trees. A decision tree consists of successive cuts on chosen variables, as shown on Figure 5.1. Each cut on a variable is denoted as a “node”. A cut value is optimized, if a given event satisfies or not the condition at a node, this process will be repeated until a criteria is fulfilled, the final node of a decision tree is also called a “leaf”. Two different types of leaves are defined, either signal-like or background-like. The Random Forest is a collection a decision trees, each one built independently.

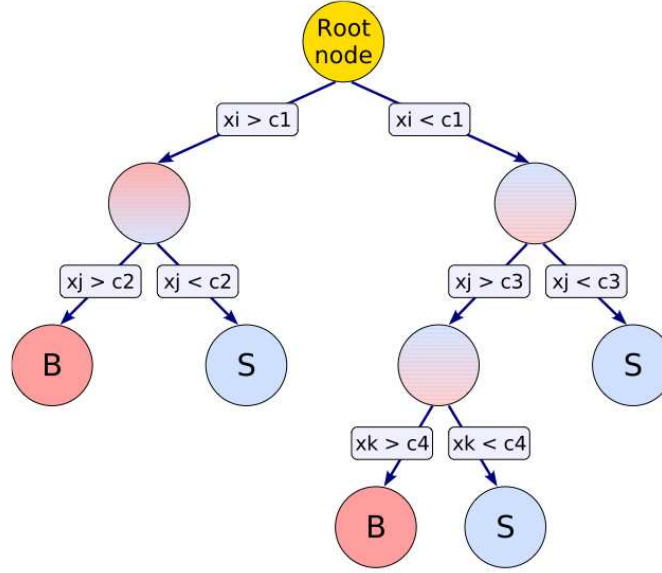


Figure 5.1: Schematic view of a decision tree.

The discrimination power of the method is obtained after a training phase being performed. The training phase consist of building decision trees, choose the discriminating variables on which cuts will be performed at each node and adjust the the cut values.

The total event sample passing all selection criteria is split in two subsamples. One subsample is used for training, the other one is only used for analysis after removing the training events in order to avoid any bias in the case the RF is "overtrained" on the training sample.

The drawback of using a single decision tree is that terminal nodes could suffer from very low statistics, causing instability when the method would be applied on an independent sample of events. The training procedure relies on three main parameters:

- N , the number of input variables
- S , the maximum number of input variables per tree
- n , the number of trees in the forest
- ℓ , the minimum leaf size.

The RF training algorithm used is designed to randomly select S input variables that will be used for the training of a decision tree, among the set of N training variables. Then, it will define cut value for each chosen variable, by maximizing the cross-entropy criteria used as figure of merit, defined as $Q(p) = -p \ln(p) - (1 - p) \ln(1 - p)$, where the purity $p = \frac{\text{signal events in the leaf}}{\text{all events in the leaf}}$. A parent node with W events and correctly classified $p \times W$ events is split into two daughters nodes if $WQ(p) < W_1Q(p_1) + W_2Q(p_2)$. This optimization is performed by considering all possible splits until this relation can not fulfilled anymore. The growth of the decision tree then stops. Another stopping criteria is determined by the minimal number of events per tree node ℓ . The overtraining

of the method is sensitive to this parameter, if it chosen to be small compared to the size of the training sample. The smaller this value is, the deeper decision trees can be grown. In the training phase, one can then end up with very high purity leaves, which can have only few events and introduce a bias since statistical fluctuations can arise when evaluating the output of the RF.

The tree continues making new nodes until it is composed of leaves only (nodes that cannot be split without a decrease in the figure of merit and nodes that cannot be split because they have too few events).

n decision trees are built using the same scheme, each time using a random subset of the training sample. Such treatment allows to take into account as much as possible correlations between variables. Therefore, the training instabilities and sensitivity to weakly discriminating variables is reduced. For a given event, the response of a decision tree is the purity p of the leaf where the event will end up after satisfying the “yes/no” decisions at each node. The response of the whole forest of decision trees is computed as the average response of every decision tree that constitutes the RF.

Advantages of using Random Forests is that the training phase can be fast. Also, the robustness introduced by the randomization of variable choice allows to a large number of input variables and the averaging of the response of each decision tree makes this method benefit of robustness against overtraining.

5.1.2 Input variables and training parameters

The dijet invariant mass is the most discriminating single variable of the analysis. Previous searches in the WH channel made use of neural networks, by discriminating the signal only from the dominant background after b -tagging, $Wb\bar{b}$. By relying on the robustness of the RF technique, one can choose an extended set of input variables and train the discriminant against all physical backgrounds at the same time, hence making use of higher statistics in the training phase. The training on the on mixture of background samples is possible since this technique the correlations between input variable are handled efficiently, therefore this technique is able to isolate regions in the input variables phase space, where, potentially, different backgrounds will be located.

All events used for training are satisfying the WH selection criteria and all previously discussed corrections have been applied. The list of the 20 input variables used in the Random Forest is given in Table 5.1. Variables having a discriminating shape are chosen, with the condition of having a good agreement between data and simulation.

The aplanarity [115] reflects the isotropy of an event and its value ranges between 0 and 0.5. Large values correspond to spherical events, while small values correspond to planar events. While $t\bar{t}$ final states are more spherical, as is typical for decay of massive objects, W +jets and multijet events tend to be more planar. It can be interpreted as the transverse momentum component out of the event plane.

To define it mathematically, let us start from the normalized momentum tensor, defined as:

$$\mathcal{M}_{ij} = \frac{\sum_k p_i^k p_j^k}{\sum_k |\vec{p}^k|^2} \quad (5.1)$$

where k runs over the jets and charged lepton in the event, and p_i^k is the i -th 3-momentum compo-

5.1. IMPROVING SENSITIVITY USING A RANDOM FOREST

Variable	Definition
$p_T(j_1)$	Leading jet p_T
$p_T(j_2)$	Sub-leading jet p_T
$E(j_2)$	Sub-leading jet energy
$\Delta R(j_1, j_2)$	ΔR between jets
$\Delta\phi(j_1, j_2)$	$\Delta\phi$ between jets
$\Delta\phi(j_1, \ell)$	$\Delta\phi$ between lepton and leading jet
$p_T(\text{dijet system})$	p_T of dijet system
m_{jj}	Dijet invariant mass
$p_T(\ell - \cancel{E}_T \text{ system})$	p_T of W candidate
\cancel{E}_T	Missing transverse energy
\mathcal{A}	Aplanarity of the event, defined from the normalized momentum tensor (Eq. 5.1)
$\sqrt{\hat{s}}$	Invariant mass of the $\nu + \ell + \text{dijet}$ system
$\Delta R(\text{dijet}, \cancel{E}_T)$	ΔR between the dijet system and W candidate
M_T^W	Lepton- \cancel{E}_T transverse mass
H_T	Scalar sum of the transverse momenta of all jets in the event
H_Z	Scalar sum of the longitudinal momenta of all jets in the event
$\cos\theta^*$	Cosine of angle between W candidate and beam direction in the zero-momentum frame [114] (see Figure 5.2)
$\cos\chi$	Cosine of angle between the lepton and the rotated vectorial sum of b quarks in the production plane [114] (see Figure 5.2)

Table 5.1: List of RF input variables, where j_1 (j_2) refers to the jet with the highest (second highest) p_T .

ment of the k -th physics object. The diagonalization of \mathcal{M} yields three eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$, with $\lambda_1 + \lambda_2 + \lambda_3 = 1$. The aplanarity is then defined as $\mathcal{A} = \frac{3}{2}\lambda_3$.

For both $\sqrt{\hat{s}}$ and $\Delta R(\ell - \cancel{E}_T)$ there are two inputs for each event, corresponding to the two solutions for the neutrino z -momentum $p_{z\nu}$. The longitudinal component of the neutrino momentum is reconstructed by imposing $M_W = 80.398$ GeV. Both real solutions for the p_z of the neutrino are considered and complex solutions are discarded. These solutions are obtained by solving this second order polynomial equation :

$$\begin{aligned}
 & p_{T\ell}^2 p_{z\nu}^2 \\
 & - (M_W^2 + 2(p_{x\ell}p_{x\nu} + p_{y\ell}p_{y\nu})) p_{z\ell} \\
 & + p_{T\ell}^2 p_{T\nu}^2 + p_{z\ell}^2 p_{T\nu}^2 - \frac{1}{4} (M_W^2 + 2(p_{x\ell}p_{x\nu} + p_{y\ell}p_{y\nu}))^2 = 0.
 \end{aligned} \tag{5.2}$$

The $\cos\theta^*$ and $\cos\chi$ variables are aimed to discriminate pairs of b jets decaying from spin-0 or spin-1 particles, providing discrimination between $Wb\bar{b}$ and WH processes. It make use of spin angular correlations observables of the final state particles. The angle θ^* is the angle between the u -quark and the W candidate in the zero momentum frame. The angle χ is the angle between the rotated $b\bar{b}$ system in the production plane and the lepton in the W candidate rest frame (see Figure 5.2). More details can be found in [114].

The reader can refer to Appendix D to see the input variables distributions to the RF. The

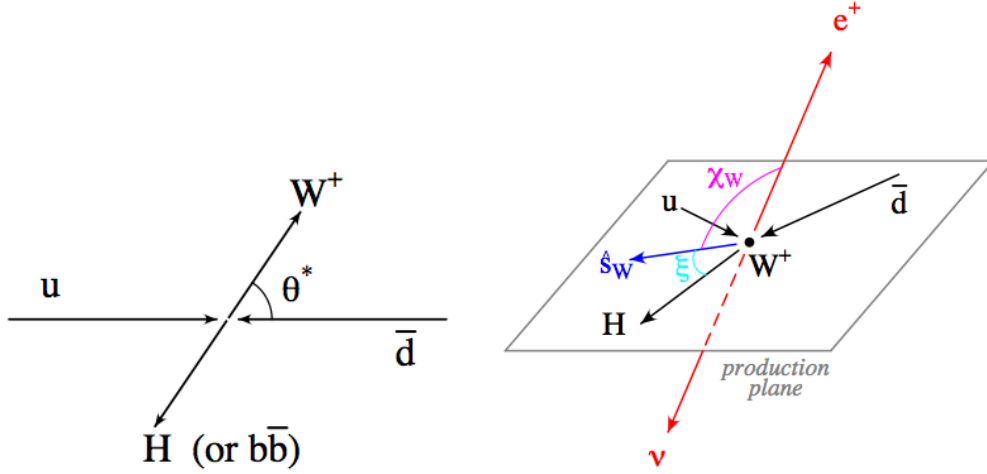


Figure 5.2: Schematic view of the angle θ^* (left figure) in the zero momentum frame and χ in the production plane (right figure). Figures are taken from [114].

choice of the input variables is made on their discriminant power and whether the simulation describes well the data distribution in order to avoid undesirable artifacts at the limit setting level.

The Random Forest is trained separately for each Higgs Boson mass hypothesis, from $M_H = 100$ GeV to $M_H = 150$ GeV by steps of 5 GeV since the kinematic properties of the signal vary as function of its mass (The heavier the Higgs boson, the harder its decay products will be). For the 2-jet channels, electron and muon channels and RunIIa and RunIIb periods are trained separately. For the 3-jet channels, RunIIa and RunIIb samples are combined in order to have sufficient statistics in the training samples to perform the training of the RF. The training parameters n and ℓ are given in Table 5.2, S being set to 13 for all channels. This tune of parameters is chosen after optimization.

Sample	Single tag		Double tag	
	n	ℓ	n	ℓ
RunIIa e 2-jet	70	1000	100	300
RunIIa μ 2-jet	70	500	200	300
RunIIb e 2-jet	50	400	50	150
RunIIb μ 2-jet	50	300	50	150
RunIIa+RunIIb e 3-jet	70	1500	70	850
RunIIa+RunIIb μ 3-jet	70	1500	70	750

Table 5.2: RF training parameters : Number of trees n and minimum leaf size ℓ for all channels.

5.1.3 Rebinning of the Random Forest output

To avoid statistical fluctuations in the limit setting procedure, a procedure is introduced in order ensure that enough Monte Carlo events populate each bin of the RF distributions. This ensures that the background modeling describes well the final discriminant output. Histograms are initially made with 1000 bins and are then rebinned to 24 bins such that each bin contains a statistically significant estimate of the data.

The current rebinning algorithm combines bins from the highest RF output down until these conditions are satisfied:

$$\begin{aligned}\mu_B &> 0.01 \\ \mu_{S+B} &> 0.1 \\ \sigma_B/\mu_B &< 0.28 \\ \sigma_{S+B}/\mu_{S+B} &< 0.2\end{aligned}\tag{5.3}$$

where μ_B (μ_{S+B}) is the total expected background (signal + background) yield and σ_B (σ_{S+B}) is the statistical uncertainty of that background (signal + background) expectation, calculated as $\sqrt{\sum w_i^2}$ over the weights, w_i , for contributing events. After the highest bin is calculated in this manner, the remaining bins in the original RF output distribution are divided equally as a function of RF output into the other 23 bins of the final discriminant.

The value of σ_B/μ_B (σ_{S+B}/μ_{S+B}) in Equation 5.3 can be understood as representing the inverse of the statistical significance of the background (signal plus background) away from zero, in standard deviations. The choice of μ_B and μ_{S+B} conditions in Equation 5.3 are made to ensure that a single high-statistics background sample does not dominate the choice of binning by giving a high statistical significance to a very low expected yield. Separate conditions on the background-only and signal-plus-background samples allows for the possibility of having a high signal-to-background ratio while maintaining a reasonable statistical understanding of the background. The choice of rebinning parameters introduces no bias in the final limit, as shown in Figure 5.3. One can see that the expected limit is stable according to the rebinning parameter choice.

Single and double tag Random Forest output distributions are shown for all channel in Figure 5.4 for RunIIa electron channel, Figure 5.5 for RunIIb electron channel, Figure 5.6 for RunIIa muon channel and Figure 5.7 for RunIIb muon channel. These figures are shown after passing the rebinning algorithm has been applied.

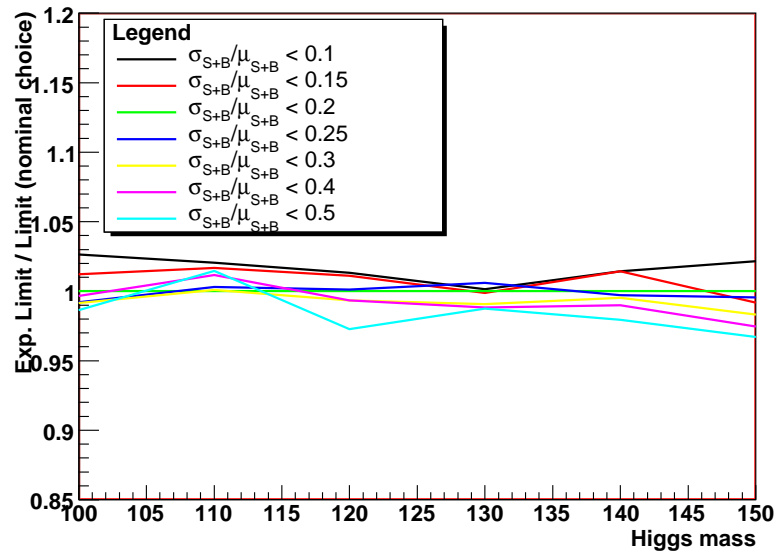


Figure 5.3: Effect of different choices of the σ_{S+B}/μ_{S+B} RF rebinning parameter on the expected limit as function of M_H . Variation of the expected limit (y-axis) is expressed as the ratio of the expected limit without systematics obtained from an RF distribution rebinned according to the σ_{S+B}/μ_{S+B} value shown in the legend to the expected limit obtained from rebinning the RF distribution using the nominal σ_{S+B}/μ_{S+B} choice of 0.2 (green line). Different choices of rebinning parameters lead to a similar result within 5%.

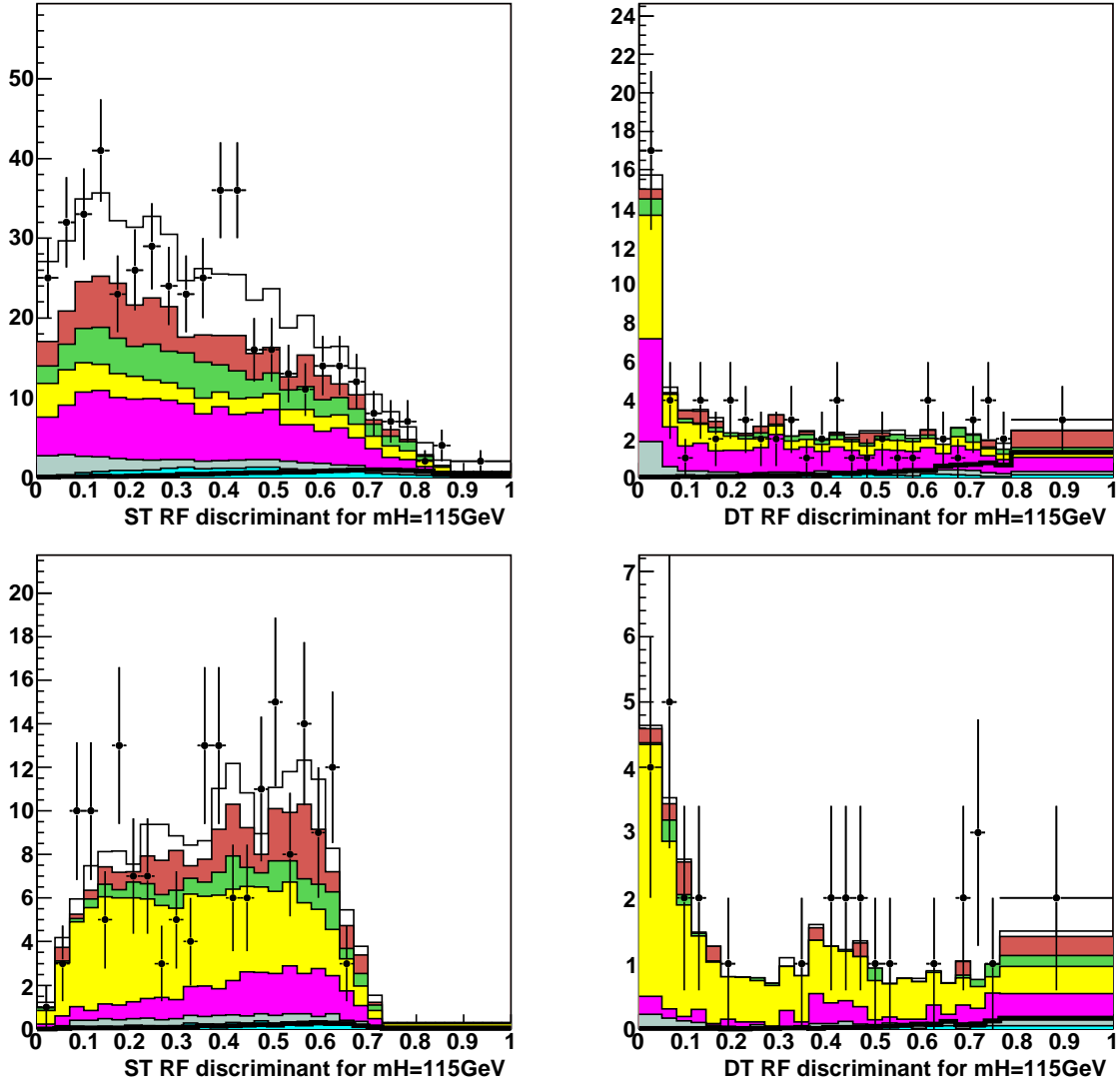


Figure 5.4: Random Forest output distributions shown separately for RunIIa electron, with 1 b -tagged jet(left figures) and 2 b -tagged jets (right figures), for the 2-jets (upper row) and 3-jets channel (bottom row).

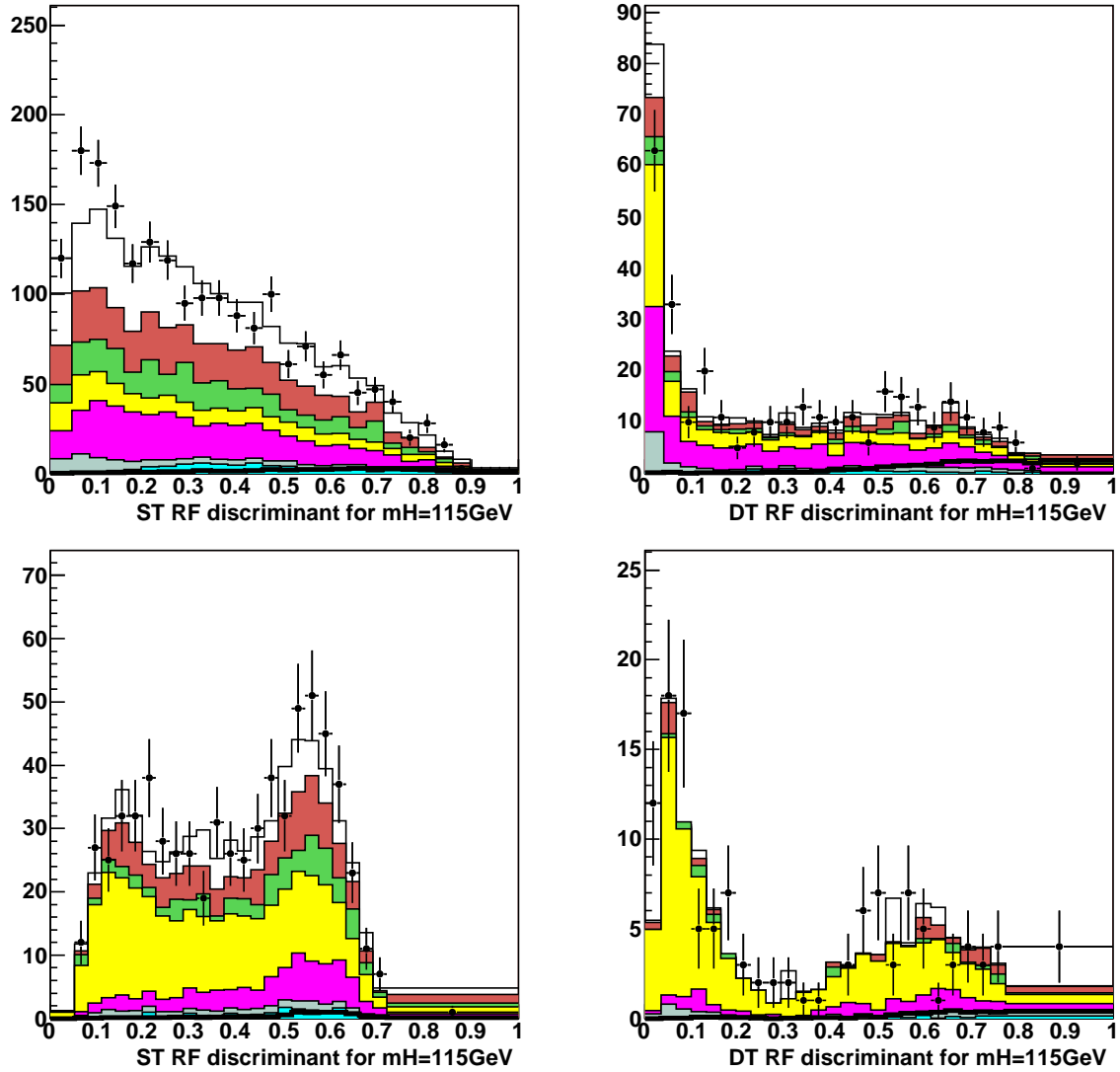


Figure 5.5: Random Forest output distributions shown separately for RunIIb electron, with 1 b -tagged jet(left figures) and 2 b -tagged jets (right figures), for the 2-jets (upper row) and 3-jets channel (bottom row).

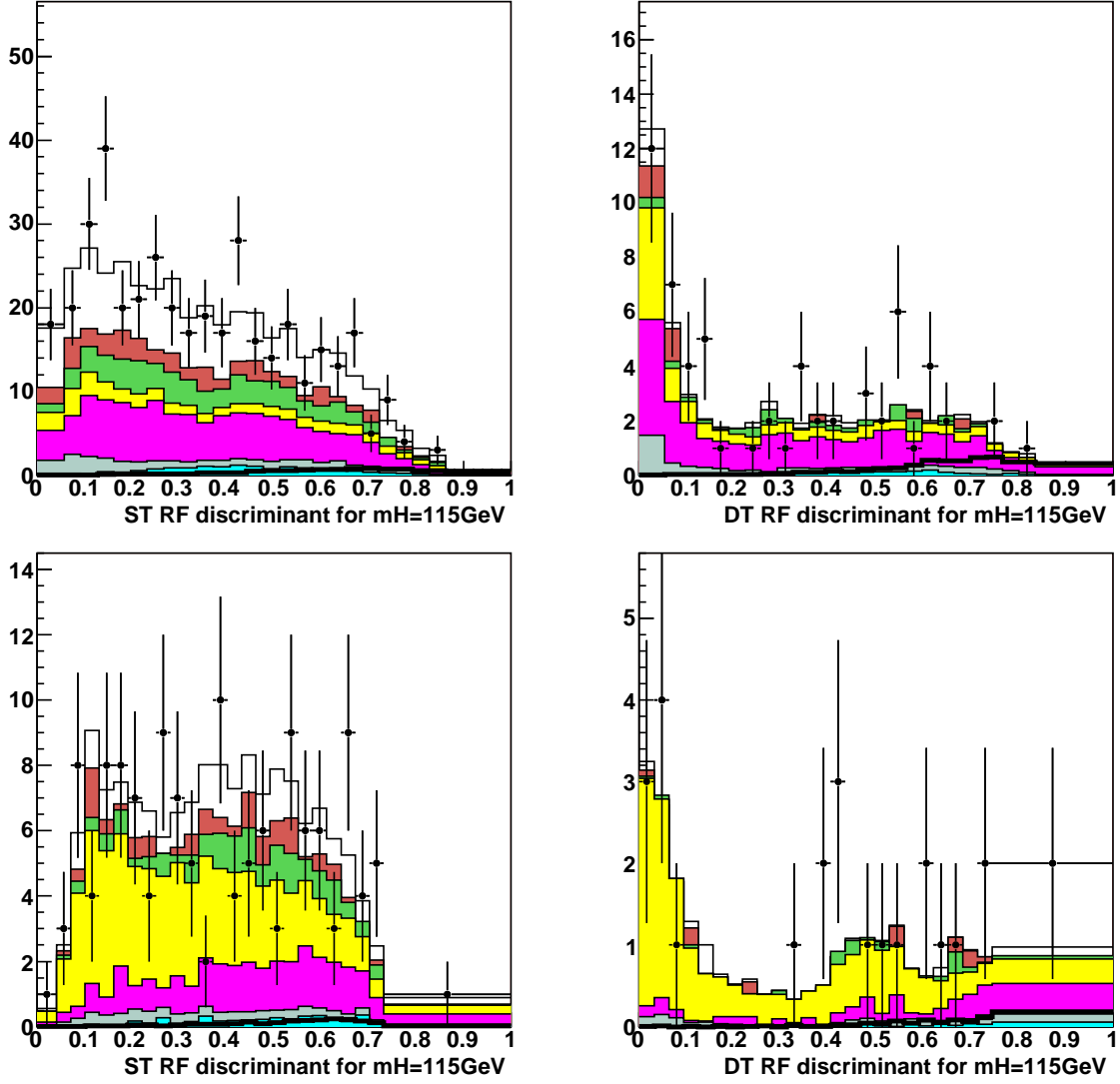


Figure 5.6: Random Forest output distributions shown separately for RunIIa muon, with 1 b -tagged jet(left figures) and 2 b -tagged jets (right figures), for the 2-jets (upper row) and 3-jets channel (bottom row).

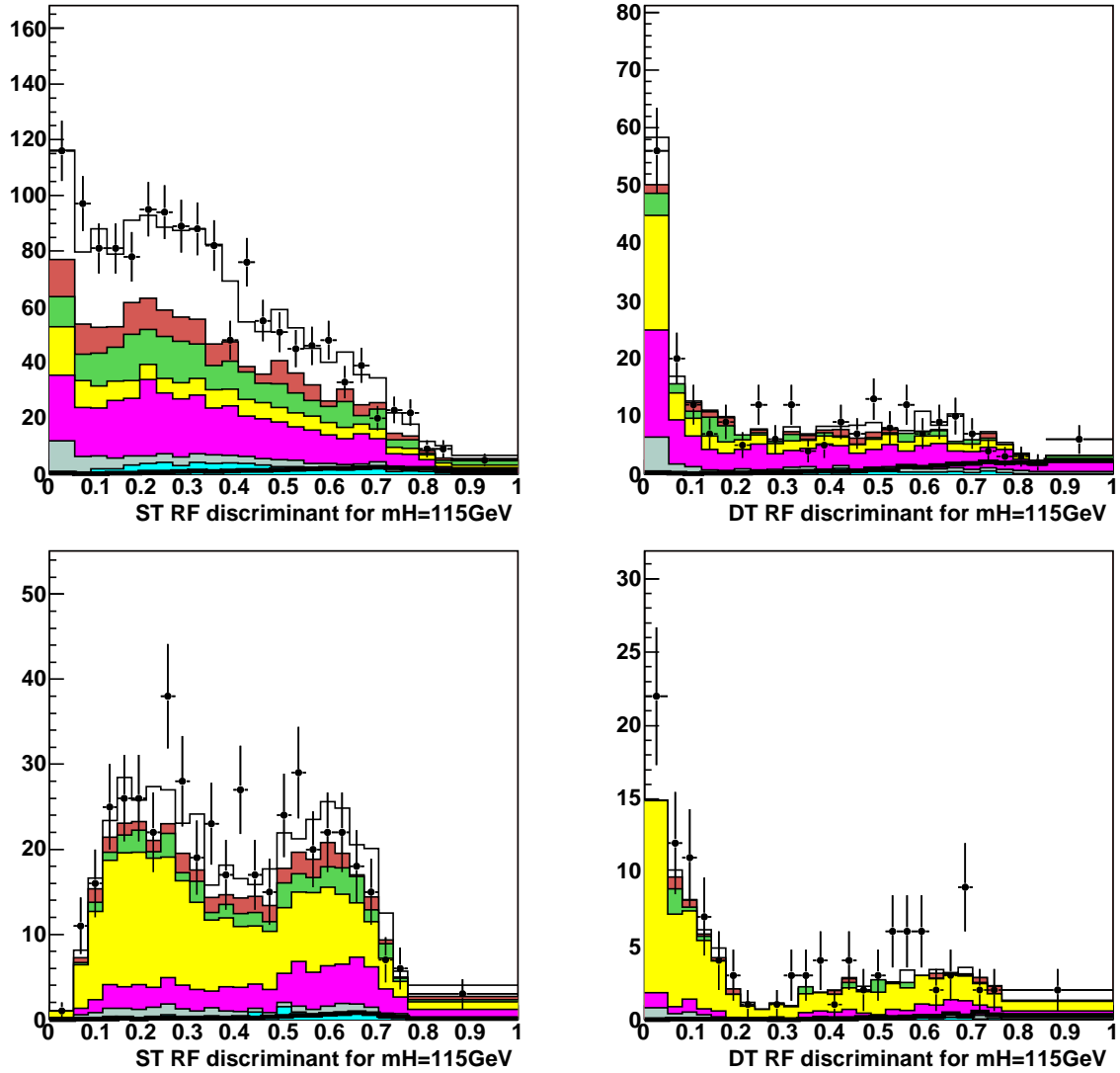


Figure 5.7: Random Forest output distributions shown separately for RunIIb muon, with 1 b -tagged jet(left figures) and 2 b -tagged jets (right figures), for the 2-jets (upper row) and 3-jets channel (bottom row).

5.1.4 Random Forest distributions

The final distribution for the RF discriminant for $M_H = 115$ GeV, with combined RunIIa and RunIIb, electron and muon channels, is shown on Figure 5.8 for the 2-jet channel and on Figure 5.9 for the 3-jet channel. Distributions for all Higgs boson masses are given in Appendix E. On these figures, all bins of the final discriminant for each channel have passed the rebinning procedure but this combination is presented with equal bin width. However, the content of each bin is set to the same as the outcome of the rebinning step. Further on, equal bin width is kept for all final discriminant distributions, as well as distributions where systematic variations are evaluated, as it will be seen in 5.2.2.

One can see, especially on Figure 5.8, the last bin of the distribution which is a very high WH signal density region, where S/B is about 1/10 (distributions having equal bin size). Discriminant distributions are then constructed, where most of the available event information, either kinematic or topological, has been exploited. No significant excess compatible with the presence of signal being observed, the RF output is used to calculate the limit on the WH associated production.

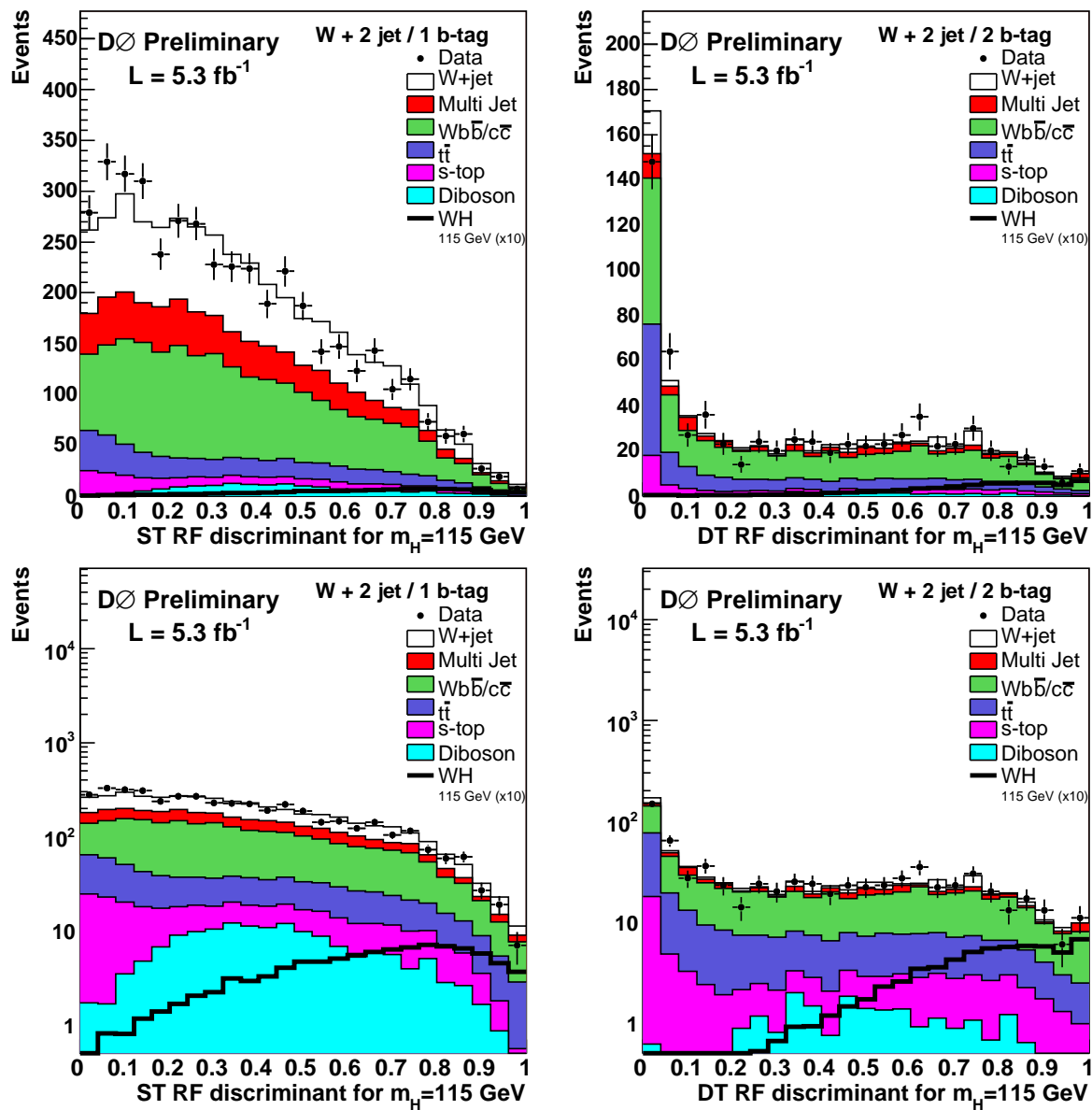


Figure 5.8: Random Forest output distributions for the combination of all 2-jet channels (e, μ , RunIIa, RunIIb) for single tag (left) and double tag (right) in linear (upper row) and logarithmic scale (bottom row).

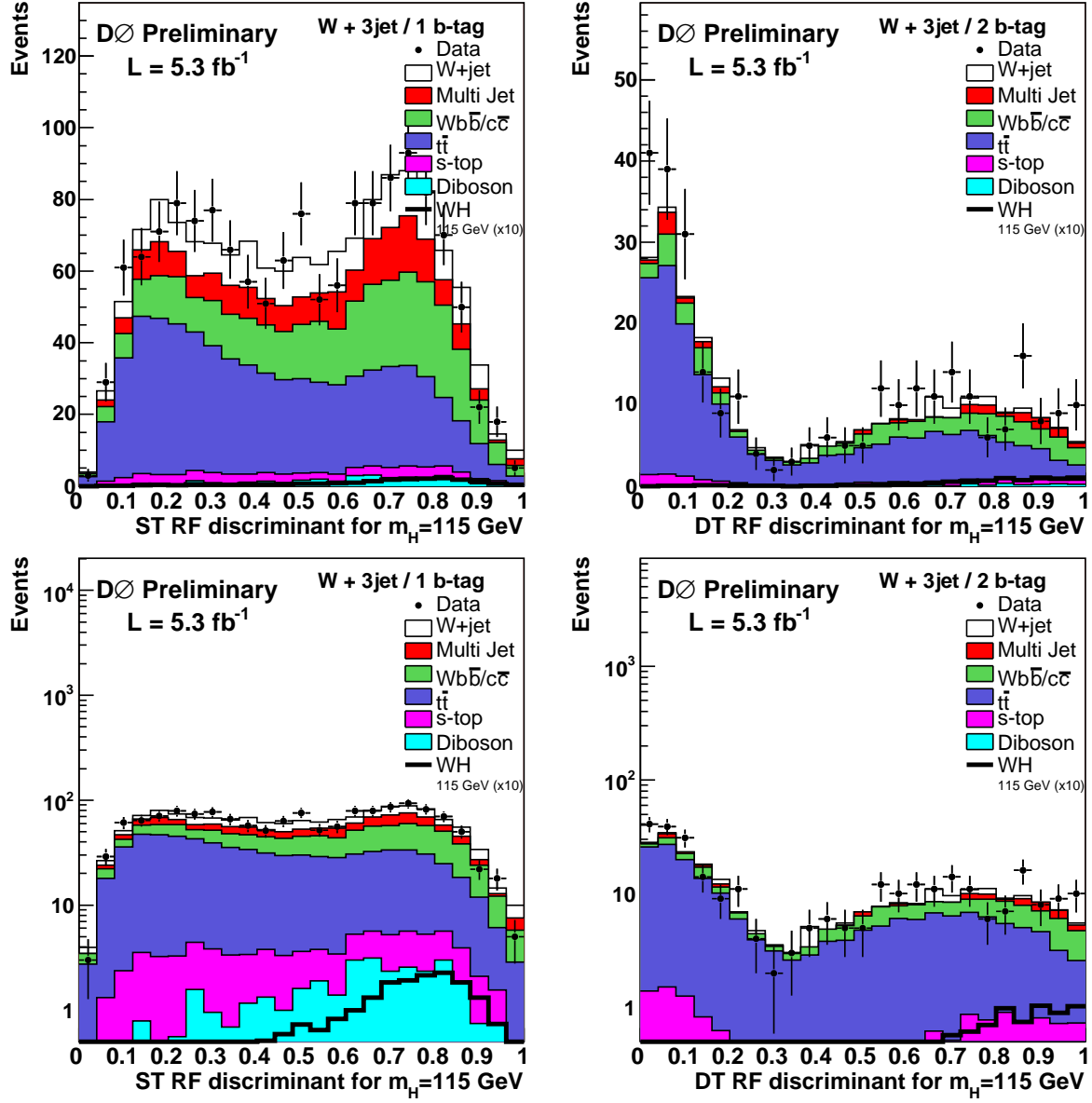


Figure 5.9: Random Forest output distributions for the combination of all 3-jet channels (e , μ , RunIIa, RunIIb) for single tag (left) and double tag (right) in linear (upper row) and logarithmic scale (bottom row).

5.2 The CL_s method

As seen in the previous section, the achieved MC background modeling is accurate, but unfortunately, the expected $WH \rightarrow \ell\nu b\bar{b}$ signal yield is low after all selection criteria applied, S/B being of the order of 1/100 after b -tagging in the most sensitive channel of the analysis. As detailed earlier, a discriminant method is used in order to exploit all event informations to differentiate signal from background events.

As no excess consistent with signal is observed, the shape of signal, background and data are used to set expected limits and observed limits from data, after combining all 16 orthogonal channels (e/μ , 2-jet/3-jet, single/double b -tagged jets, RunIIa/RunIIb).

The limit setting procedure, based on a semi-frequentist approach, is presented in Section 5.2.1. In Section 5.2.2, the sources and determination of systematic uncertainties will be detailed as well as their treatment in this last step of the analysis. Finally, results obtained for the search of the $WH \rightarrow \ell\nu b\bar{b}$ signal will be presented in Section 5.3.

5.2.1 Principle of the method

Sensitivity estimator

All Higgs analyses and analyses aimed on searches of New Physics signatures at the DØ experiment make use of the Collie software suite [116]. This package provides the tools to evaluate cross sections or limits, based on the CL_s method [117, 118]. This semi-frequentist method, designed to simulate outcomes of repeated experiments, was introduced at LEP experiments and used for their Higgs search program [119]. Its purpose is hypothesis testing, for two types of assumptions, either data satisfies the background-only (or Test) hypothesis or the signal+background (Null) hypothesis.

In order to evaluate and quantify which hypothesis is favoured for a given dataset, a test statistic based on the likelihood ratio of both hypotheses $Q = L(s)/L(s+b)$ is introduced. This sensitivity estimator is aimed to exploit as best as possible the sensitivity power of a given distribution (*i.e.* a final discriminant distribution). These two hypotheses, drawn randomly from a Poisson parent distribution are characterized by their expected numbers of signal (s), background (b) and observed number of data or pseudo-data (d). The likelihood ratio can be written as

$$Q(s, b, d) = \frac{e^{-(s+b)}(s+b)^d/d!}{e^{-b}b^d/d!} \quad (5.4)$$

In this definition, s and b are the expected number of signal and background from MC and d is either the observed data or pseudo-data.

This test statistic can be easily expanded to combine separate channels or bins of an histogram used as final variable of an analysis by multiplying each contribution:

$$Q' = \prod_{i=1}^{N_{bins}} \prod_{j=1}^{N_{channels}} Q_{ij} \quad (5.5)$$

$$= \prod_{i=1}^{N_{bins}} \prod_{j=1}^{N_{channels}} \frac{e^{-(s_{ij}+b_{ij})} (s_{ij} + b_{ij})^{d_{ij}} / d_{ij}!}{e^{-b_{ij}} b_{ij}^{d_{ij}} / d_{ij}!} \quad (5.6)$$

$$= \prod_{i=1}^{N_{bins}} \prod_{j=1}^{N_{channels}} e^{-s_{ij}} \left(\frac{s_{ij} + b_{ij}}{b_{ij}} \right)^{d_{ij}} \quad (5.7)$$

A more usual form can be used by transforming Q' to the negative log-likelihood ratio Γ , also denoted LLR :

$$\Gamma = -2 \ln(Q') \quad (5.8)$$

$$= 2 \sum_{i=1}^{N_{bins}} \sum_{j=1}^{N_{channels}} (s_{ij} - d_{ij} \ln(1 + s_{ij}/b_{ij})) \quad (5.9)$$

In order to give a frequentist interpretation, many trials are repeated using pseudo-data for the Test and Null hypotheses. Pseudo-data are generated by throwing random trials for a Poisson distribution with a mean value d_{ij} given by the expected background or signal+background hypothesis, where $d_{ij} = s_{ij} + b_{ij}$ for the Test hypothesis and $d_{ij} = b_{ij}$ for the Null hypothesis, for a given bin i of the channel j . Probability Density functions (PDFs), denoted as P_H , are then constructed for a given hypothesis H from the sampling of the pseudo-data. It is done separately for the Test (P_{S+B}) and Null (P_B) hypotheses.

Confidence levels, interpreted as the probability to observe a deviation from a reference (*i.e.* measured in data), are derived according to the chosen test statistic. Confidence levels for both hypotheses are computed as follows:

$$CL_{S+B} = \wp_{S+B}(\Gamma \geq \Gamma_d) = \int_{\Gamma_d}^{+\infty} \frac{\partial P_{S+B}}{\partial \Gamma} d\Gamma \quad (5.10)$$

$$CL_B = \wp_B(\Gamma \geq \Gamma_d) = \int_{\Gamma_d}^{+\infty} \frac{\partial P_B}{\partial \Gamma} d\Gamma \quad (5.11)$$

An example of the representation of the PDFs and confidence levels can be seen in Figure 5.10.

In order to prevent biases that could arise by choosing the CL_{S+B} to evaluate exclusion limits¹, the signal confidence level is estimated as:

$$CL_S = \frac{CL_{S+B}}{CL_B}. \quad (5.12)$$

The CL_S estimator can be defined for two cases:

¹for instance if the data fluctuates down significantly below the background prediction, an example is given in Reference ??

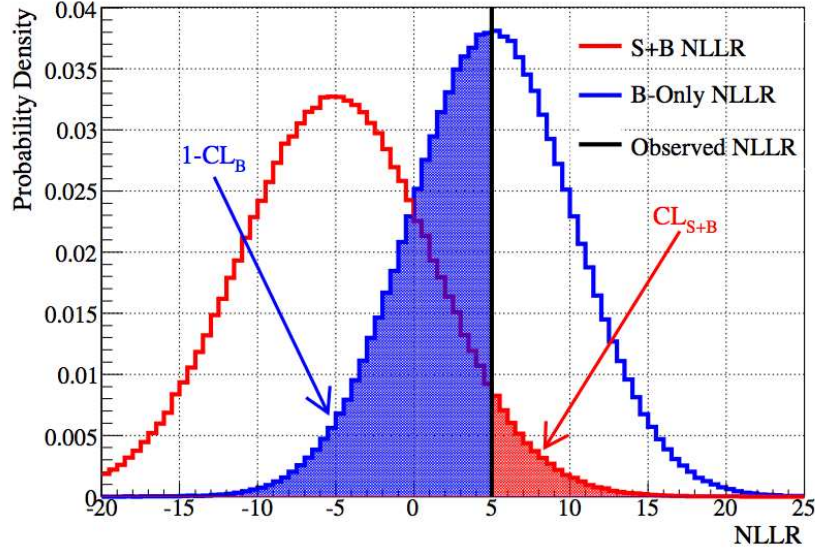


Figure 5.10: Distributions of the PDFs P_{s+b} and P_b (respectively red and blue histograms) obtained by replacing the observed data with the median background expectation and Γ_{obs} (black line). The probability to observe a signal+background fluctuation (or p -value for the S+B hypothesis) corresponds to the integral of the shaded red region (CL_{S+B}) and the p -value for a background-only observation corresponds to the integral of the shaded blue region ($1-CL_B$).

- CL_S^{exp} , obtained by making the assumption that the “data” corresponds exactly to the background-only hypothesis (Γ_d in Equation 5.11 is equal to the median value of the LLR_{S+B})
- CL_S^{obs} , obtained from observation in data (Γ_d in Equation 5.11 is calculated from data observation, d_{ij} is the bin content from each data histogram)

This procedure can be repeated over all the model parameters (for each Higgs boson mass point sample in the case of the WH channel). An example of the representation of the median values of Γ_{s+b} and Γ_b (and its associated $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands) is shown on Figure 5.11. The separation between Γ_{s+b} and Γ_b can be interpreted as the discriminating power of the analysis to be able to separate the Test and Null hypotheses. Moreover, the value of Γ_{obs} relative to Γ_{S+B} and Γ_B indicates whether the data distribution appears to be more signal-like or background-like (in this case, data is visibly following the background-like hypothesis).

Handling of systematic uncertainties and profile likelihood ratio

The introduction of systematic errors is a parameter that has to be included in order to evaluate the final result in a proper manner. These uncertainties, also referred to as *nuisance parameters*, are taken into account due to models having a limited accuracy of their description, as it will be described later in Section 5.2.2.

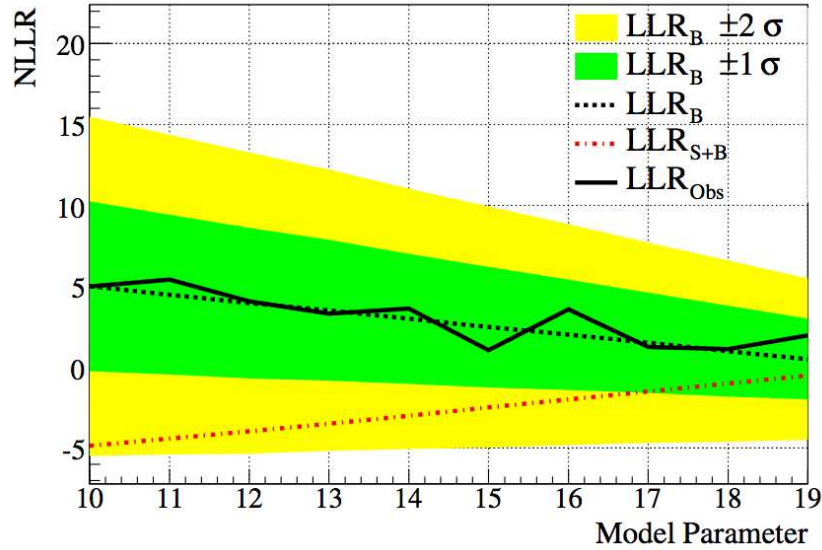


Figure 5.11: Example of the distributions of the median values of Γ_{s+b} , Γ_b with its associated $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands and Γ_{obs} (denoted as LLR_{s+b} , LLR_b and LLR_{obs} on the figure), for different parameters of the model. On this figure, the data is the median of the background-only model

Systematic errors undergo a bayesian treatment which can interpreted as a gaussian smearing of the prediction. This is due to the fact that the predicted number of events per bin (for a given contributing event source, either signal or one of the backgrounds) is modified as the following:

$$p'_i = p_i \prod_{k=0}^{N_{systematics}} (1 + \sigma_i^k S_k) \quad (5.13)$$

where p_i is the nominal prediction, σ_i^k is the fractional change in the number of events for the specified nuisance parameter k (or uncertainty scale) and S_k representing a stochastic variable. In order to constrain to physical values, p'_i has a lower bound of 0.

LLR distributions will become broader with a width proportional to the quadrature sum of Poisson uncertainty and systematics uncertainty ($\sigma_{tot}^2 \propto \sigma_{Poisson}^2 + \sigma_{systematics}^2$). This will have the effect of reducing the sensitivity of the analysis and can impact on the limit extraction.

Systematic errors can be of two kinds:

Flat systematics : the error is uniform across all bins of the final distribution (σ_i^k is constant for every bin i), only affecting the rate of a process (for example the luminosity uncertainty),

Shape systematics : the scale of the uncertainty σ_i^k is different for each bin of the final distribution, therefore introducing a shape dependence.

In order to reduce the impact of nuisance parameters that could be overestimated (hence reducing the sensitivity of the analysis), one can perform a fit to the data for each hypothesis, by

letting float each event contribution of the analysis (either signal or background separately) within its uncertainties. Systematic uncertainties are also floated within their gaussian prior. This method is called *profiling*. A modified χ^2 function is introduced for the application of this method A simple derivation is given here, starting from the general χ^2 definition:

$$\chi^2 = \sum_i^{N_{bins}} \frac{(p_i - d_i)^2}{\sigma_i^2} \cong -2 \ln \left(\frac{\mathcal{L}}{\mathcal{L}_0} \right) \quad (5.14)$$

$$= -2 \sum_i^{N_{bins}} \left[(p_i - d_i) - d_i \ln \left(\frac{p_i}{d_i} \right) \right] \quad (5.15)$$

$$(5.16)$$

with

$$\mathcal{L} = \prod_i^{N_{bins}} \frac{p_i^{d_i} e^{-p_i}}{d_i!}, \quad \mathcal{L}_0 = \prod_i^{N_{bins}} \frac{d_i^{d_i} e^{-d_i}}{d_i!}$$

The profiling method consists in including gaussian priors for the systematic uncertainties by rewriting Equation 5.14 as:

$$\chi^2 = -2 \ln \left(\frac{\mathcal{L}}{\mathcal{L}_0} \times \frac{\mathcal{L}^{syst}}{\mathcal{L}_0^{syst}} \right), \quad (5.17)$$

with

$$\frac{\mathcal{L}^{syst}}{\mathcal{L}_0^{syst}} = \prod_k^{N_{systematics}} e^{-\frac{s_k^2}{2}} \quad (5.18)$$

. A full derivation can be found in Reference [120] Hence a general χ^2 formula is obtained for a given hypothesis as:

$$\chi^2(H) = -2 \ln P(data|H, \theta) = 2 \sum_i^{N_{bins}} \left[(p(H)_i' - d_i) - d_i \left(\frac{p(H)_i'}{d_i} \right) \right] + \sum_k^{N_{systematics}} S(H)_k^2 \quad (5.19)$$

where $p(H)_i'$ is the expected number of events for a given hypothesis H in bin i (either $s_i + b_i$ for the Test hypothesis or b_i for the Null hypothesis), d_i is either from observed data or the pseudo-data and θ represent the set of nuisance parameters associated to the hypothesis H .

Th $S(H)_k^2$ term enforces a Gaussian constraint on nuisance parameter values (see Equation 5.17).

The LLR test statistic is equivalent to the $\Delta\chi^2$ after fitting to data, defined as

$$LLR = -2 \ln \left(\frac{P(data|H_1, \hat{\theta}_1)}{P(data|H_0, \hat{\theta}_0)} \right) \quad (5.20)$$

$$= -2 \ln P(data|H_1, \hat{\theta}_1) + 2 \ln P(data|H_0, \hat{\theta}_0) \quad (5.21)$$

$$= \Delta\chi^2 \quad (5.22)$$

$$= \chi_{min}^2(H_1) - \chi_{min}^2(H_0) \quad (5.23)$$

$$= 2 \sum_i^{N_{bins}} \left[(\hat{p}(H_1)_i' - \hat{p}(H_0)_i') - d_i \ln \left(\frac{\hat{p}(H_1)_i'}{\hat{p}(H_0)_i'} \right) \right] \\ + \sum_k^{N_{systematics}} \left(\hat{S}(H_1)_k^2 - \hat{S}(H_0)_k^2 \right) \quad (5.24)$$

after the χ^2 for both hypotheses has been minimized independently by varying the nuisance parameters (namely S_k defined in Equation 5.13) to give the best fit of the model to the data.

θ_1 represents the set of nuisance parameters for H_1 , θ_0 represents the set of nuisance parameters for H_0 , $\hat{\theta}_1$ represents the set of nuisance parameter values that maximize the likelihood for H_1 , and $\hat{\theta}_0$ represents the set of nuisance parameters that maximize the likelihood for H_0 . The parameters $\hat{p}(H_1)_i'$ and $\hat{p}(H_1)_0'$ represent the predicted numbers of events in bin i given $\hat{\theta}_1$ and $\hat{\theta}_0$, respectively. The parameters $\hat{S}(H_1)_1'$ and $\hat{S}(H_0)_0'$ represent the central values of the nuisance parameters as defined by $\hat{\theta}_1$ and $\hat{\theta}_0$, respectively.

Systematic errors correlations among channels and samples are properly taken into account during this fitting procedure.

Limit calculation

This method provides a computation of the confidence level for observing a signal excess in the data. In order to quote a "95% confidence level" (C.L.) limit as it is usual in high energy physics is to give, the Collie software will iteratively scale the signal rate $s \propto \sigma_s L \epsilon_s$ by increasing the signal SM cross section σ_s by steps to reach $\sigma_s^{95\%}$. Since the amount of signal impacts the CL_S value, this process will end when the condition

$$1 - CL_S \geq 95\% \quad (5.25)$$

is fulfilled, where CL_S can be either CL_S^{exp} (obtained by replacing the observed data with the median background expectation) or CL_S^{obs} . Figure 5.12 shows the CL_S value obtained as function of the signal scaling factor, quoted as the 95% confidence level limit in the analysis, after all channels being combined.

5.2.2 Systematic uncertainties : sources and treatment

As emphasized in Section 5.2.1, systematic errors can be either flat along the final distribution used for limit setting or shape dependent. So as to get shape systematics, the whole analysis is repeated

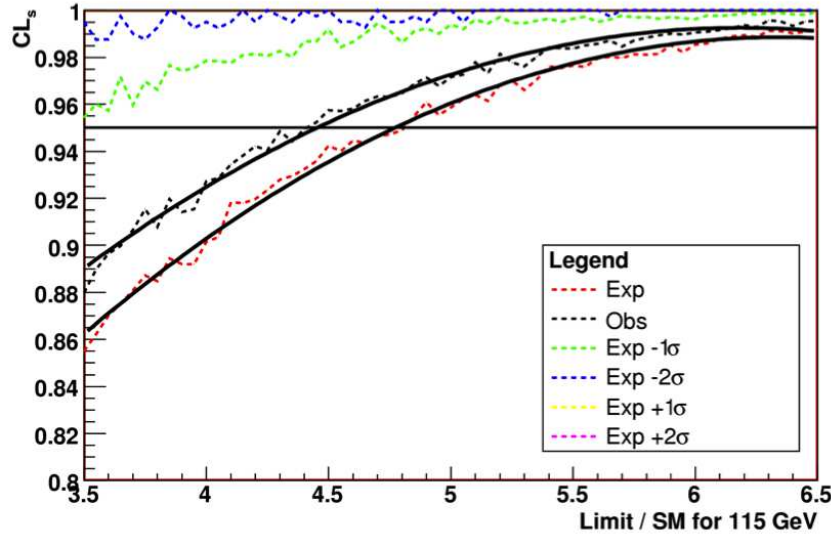


Figure 5.12: CL_s values obtained as function of the signal scaling factor. The observed (expected) value is shown as the intersection of the black (red) fitted dashed lines with $CL_s = 0.95$. The $\pm 1\sigma$ and $\pm 2\sigma$ expected CL_s curves are also obtained by varying the signal scaling factor, the intersection with $CL_s = 0.95$ being out of range of the plot.

by varying each source of uncertainty by $\pm 1\sigma$, where σ is the size of the uncertainty. For every variation, the total normalization of backgrounds is preserved to be the same as in the nominal (or no variation) case, letting the W +jets be adjusted to data as described in 4.7.2.

The Collie software takes as input the absolute value of the positive and negative fractional deviations to the nominal distributions, Δ_i^+ and Δ_i^- , thus defined for a given bin i as

$$\Delta_i^+ = \frac{N(+1\sigma) - N(0\sigma)}{N(0\sigma)} \quad (5.26)$$

$$\Delta_i^- = \frac{N(0\sigma) - N(-1\sigma)}{N(0\sigma)} \quad (5.27)$$

where $N(0\sigma)$ is the nominal case and $N(\pm 1\sigma)$ the $\pm 1\sigma$ fluctuation.

These variations are determined independently for each data-taking period, lepton channel and number of jets in the final state. The motivation to use shaped systematic errors is to allow the fitting in the profile likelihood technique only in a reasonable amount, in order to not over-evaluate the impact of systematic errors. It is also the most precise evaluation of systematic uncertainties, compared to a flat systematic which may cover the high signal density region.

Systematic uncertainties have different sources:

Luminosity Two sources of uncertainty are considered for luminosity: one coming from the uncertainty on the delivered integrated luminosity by the Tevatron (correlated with CDF) and

another contribution from the $D\bar{O}$ luminosity measurement. Both are considered as flat uncertainties.

Jet Energy Scale The JES uncertainty is evaluated by scaling the jet energy by the JES correction shifted by $\pm 1\sigma$. Since the jet energy changes, the jet acceptance is modified and events can either pass or not the jet cut in the analysis depending if the jets energy is shifted up or down. The final distribution shape can therefore be modified. Moreover, events can migrate between jet multiplicity channels. Let us consider an event with two jets passing the $p_T > 20$ GeV criteria and a third which does not fulfill this requirement. If the energy of the third jet is shifted, the event can potentially migrate from the 2 jets channel to the the three jets channel.

Jet Resolution As shown in Section 3.5.4, the jet resolution in Monte Carlo is adjusted to match the data. The JSSR parameters are derived but are subject to uncertainties arising from the procedure, such as error on fits from statistical limitations or event selection applied to extract the parameters. Similarly to the JES uncertainty, event migration can occur by varying the parameters by $\pm 1\sigma$, hence it is a shaped systematic.

Jet identification For this type of error, only the negative fractional change is evaluated. Since the random removal procedure is applied, jet scale factors can not be greater than 1. Indeed, in case the case of scale factors greater than 1 if shifted up, jets can not be added in an event (also called "jet promotion"). Therefore, only negative systematic variations are considered. Hence, for the treatment of the positive fractional change of this systematic error in the Collie software is made by symmetrization of the negative systematic error. It is a shaped systematic uncertainty for the same reason mentioned for the JES and jet resolution uncertainties.

Jet Vertex Confirmation Similarly to the jet identification systematic error, only the negative shaped fractional change of the systematic error is evaluated and symmetrized to evaluate the positive fractional change.

Lepton identification Systematic errors for electron and muon identification are of different origins since their identification relies on different detector properties.

Electron identification Systematic errors for electron identification, reconstruction efficiency and energy smearing is obtained by varying by $\pm 1\sigma$ systematic uncertainties for each component, leading to a shaped systematic error.

Muon identification The muon identification systematic error has three components: uncertainty on the muon identification efficiency, uncertainty on the track reconstruction efficiency, and the uncertainty on the muon isolation efficiency. Separate uncertainties in the RunIIa [121] and RunIIb [45] samples are applied.

- *Identification efficiency uncertainty:* A 0.8% uncertainty in RunIIa and a 1.2% uncertainty in RunIIb is applied. For RunIIb, an additional 2% systematic is applied for events with muon $p_T < 20$ GeV per the recommendation of the Muon ID group.

- *Track reconstruction efficiency uncertainty:* A 2.3% uncertainty in RunIIa and a 1.4% uncertainty in RunIIb is applied.
- *Isolation efficiency uncertainty:* A 3.8% uncertainty in RunIIa and a 0.9% uncertainty in RunIIb is applied.

A flat uncertainty is used.

ALPGEN reweightings Several systematic uncertainties arise from ALPGEN inaccuracy to model the dominant V +jets backgrounds:

- A systematic uncertainty is evaluated for each kinematic reweighting for W/Z +jets samples only. A function is fitted and the uncertainty from the fit (using the correlation matrix from the fit) is taken as systematic uncertainty for the η reweightings of the leading and second leading jets as well as the reweighting of the ΔR between the two leading jets (for details on reweighting see Section 4.5). In addition, standard $D\bar{D} \rightarrow p_T^V$ ($V = W, Z$) have a shaped uncertainty assigned.
- ALPGEN MLM matching parameters are varied. These changes impact the shape of the dijet invariant mass distribution and are taken as $\pm 1\sigma$ uncertainty bands. This shaped systematic error impacts only W/Z +light jets samples. $t\bar{t}$ samples, also produced with ALPGEN, are not affected by this uncertainty even though they are produced using the same generators. The reason is that jets from W decays in $t\bar{t}+0/p^1$ do not pass the process of matching since they do not originate from a radiation. However, $t\bar{t}+n/p$ ($n=1,2$) could be affected by such systematic treatment, but the effect is expected to be small. More details on the determination of this uncertainty can be found in Reference [122].
- $D\bar{D}$ analyses use the same tuning of ALPGEN parameters at the parton-level. Systematic uncertainties related to the ALPGEN event scale (k_T and Q^2) and the underlying event modeling are introduced. The scale uncertainty is evaluated separately for W/Z +light partons and W/Z +heavy partons. The underlying uncertainty is evaluated on W/Z +light partons, and applied to all W/Z +jets. More details on the determination of these uncertainties can be found in Reference [122].

The extraction of ALPGEN tunes related errors is detailed in [122].

Taggability As explained in 4.9.1, taggability scale factors are derived in the WH analysis. Uncertainties from the fit of the parametrization (shown as blue bands on Figures ??–4.15) are propagated to the final discriminant, giving a shaped uncertainty.

b -tagging efficiency The systematic uncertainty on b -tagging efficiency is evaluated by using the $\pm 1\sigma$ deviation from the Tagging scale factor (TSF). In this analysis, direct tagging to the MC is applied, hence introducing a scale factor which is the ratio of the data TRF and the MC TRF. This systematic uncertainty is evaluated separately for light and for heavy flavor jets. On average, 3%, 2.5%, 1.4% for the tagging efficiency are observed in single tag events

¹ $/p$ stands for light partons originating from additional radiations.

(respectively for b -jets, c -jets and light jets). For double b -tagged events, the amplitude of the fractional changes are twice higher. Being dependent on the jet kinematics, a shape uncertainty is obtained.

Trigger In the electron channel, a 2% uncertainty from trigger efficiency is applied.

In the muon channel, a different treatment is adopted (see Section 4.3.3). Since all triggers are used (inclusive MU trigger) a larger systematic for the normalization change from the correction described in 4.3.3 is applied. To confirm that the muon trigger efficiency is 100%, this result is compared with a result triggered with the Single Muon triggers.

The change in surface normalization is 2% between the inclusive MU trigger and the Single Muon trigger result. In order to check for a shape uncertainty from a trigger effect, ratios for Data/MC (inclusive trigger - single mu trigger) over Data/MC (single mu trigger) and also separately the ratios Data/Data and MC/MC (inclusive)/(single mu) are built. A fit function (sigmoid + exponential decay) is derived as a function of p_T^{jj} for the Data/MC (inclusive trigger - single mu trigger) over Data/MC (single mu trigger) double ratio.

The function is derived in the pretag sample and apply it to both single and double b -tagged samples (on signal and all MC backgrounds) as well. Figure 5.13 shows the function obtained from a fit. The value of the function is applied as a multiplicative factor on P_{corr} (see Section 4.3.3) and take the difference with the nominal as the systematic on the muon trigger (normalization included.). The muon systematic is typically 1-3% .

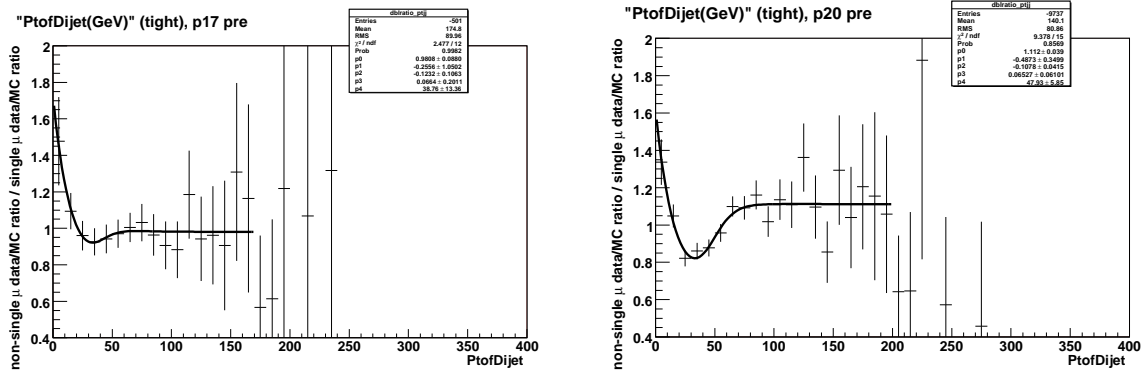


Figure 5.13: Fit function (sigmoid + exponential) on p_T^{jj} for the Data/MC (inclusive trigger - single mu trigger) over Data/MC (single mu trigger) double ratio used as the systematic uncertainty on the muon trigger for RunIIa (left) and RunIIb (right) before b -tagging being applied.

Multijet background modeling and normalization of the W +jet background The multijet modeling is impacted by two sources of systematic uncertainties : one from the $\pm 1\sigma$ uncertainties of the lepton efficiency (shown in Figure 4.9 for the muon channel) and the other from jet fake rate parametrization (shown in Figure ?? for the electron channel) as explained in 4.6.

The error obtained from the fit of the W +jet sample is propagated as a flat systematic error (and taken as a "cross section" uncertainty, as well as the 20% uncertainty from the S_{HF} factor for W +heavy flavor).

The normalization of the multijet sample is anti-correlated with the normalization of the W +jets (light and heavy flavor) sample, therefore taken into account in the limit setting procedure.

Cross sections A variation in the cross section value only affects the rate of events, not their shape in the final distribution. Uncertainties for processes used in the WH analysis are:

Process	Cross section uncertainty
WH	6%
$t\bar{t}$	10%
Single-top	12%
W/Z +light jets	6%
$W/Z + c\bar{c}, b\bar{b}$	20%
WW, WZ, ZZ	7%

Table 5.3: Cross section uncertainties from theory for WH , $t\bar{t}$, Single-top and WW, WZ, ZZ processes. W/Z +light jets uncertainty is from the normalization to data procedure and from S_{HF} for $W/Z + c\bar{c}, b\bar{b}$

Parton Density Functions Cross sections uncertainties detailed earlier take into account PDF uncertainties. However the PDF uncertainty could impact both the cross section and the kinematic acceptance of a MC process. Only the changes due to kinematic acceptance are retained as part of the PDF systematic. At the generator level, the normalization shouldn't be affected. Hence, the overall yield is kept same as the nominal case in order to no impact the cross section of a process. The resultant shaped-only systematic uncertainty is propagated to the analysis. After selection cuts applied, the residual normalization effect comes from the acceptance change. This procedure prevents us to not double count cross section uncertainties for limit settings.

Due to low statistics in W +light jets and diboson MC samples after b -tagging, jet-related systematic errors are treated differently, namely Jet Energy Scale, jet resolution, jet identification and jet vertex confirmation. Without any particular treatment, large fluctuation in the $\pm 1\sigma$ variations arises from the ratio defined in Equation 5.27 for the samples mentioned above. For example, such large deviations can arise if in a given bin, the number of events is low and probably having large event weights. An example of such large deviations from the nominal value is shown in Figure ??.

Fluctuations introduced by the low statistics of these samples can be reduced by using an algorithm that can smooth the shape of the systematic variations. For a given bin, a new value can be assigned by evaluated the value of a function fitted across 5 neighbouring bins (2 before and 2 after the studied bin). Such algorithm performs in a way that fluctuations are averaged.

However, correlations between bins introduce biases in the shape of the distributions. Features that should initially be described in the variations can be completely smeared out and not be correct

to describe the behaviour of shifted errors. Moreover, even after several iterations of the algorithm, fluctuations can still be present and every bin has a strong correlation to any other bin of the entire distribution.

Hence, another strategy is adopted. Systematic variations for the 1 and 2 b -tagged jets samples is determined from pretag events. A full comparison can be found in [123], along with cross-checks with higher statistics samples

Distributions showing the $\pm 1\sigma$ variations are given in Appendix F for shaped systematic errors. In each figure, the $+1\sigma$ variation is shown as a red line, while the corresponding -1σ variations is shown in blue. The shape of nominal random forest output distributions are also included in the figures as dashed black lines, with an arbitrary vertical scale.

Table 5.4 summarizes the systematic errors used in the WH analysis for each channel.

	electron		electron		electron		electron		muon		muon		muon		muon	
	2jet-1b	1la 1lb	2jet-2b	1la 1lb	3jet-1b	1la 1lb	3jet-2b	1la 1lb	2jet-1b	1la 1lb	2jet-2b	1la 1lb	3jet-1b	1la 1lb	3jet-2b	1la 1lb
DZero_Lumi	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%	4.6% 4.6%
Lumi	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%	4.0% 4.0%
EMID/EMTrigger	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
JES	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
JSSR	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
JetID	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
VCJet	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Tagga	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
bTag_HF	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
bTag_LF	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
PDF 1-20	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
Bkgd_Xsec_EW	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%
ALP_JTE	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
ALP_JTH	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
ALP_DR	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
ALP_LPE	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
ALP_ULE	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
ALP_SCA	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
ALP_MLM	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
ALP_ZPT	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
Bkgd_Xsec_Top	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%	10% 10%
Bkgd_Xsec_singletop	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%	12% 12%
Bkgd_Xsec_HF	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%	20% 20%
Bkgd_Xsec_LF	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%	6% 6%
MUID																
MUTrigger_WH																
Bkgd_QCDev_fake	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
Bkgd_QCDev_eff	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S	S S
Bkgd_QCDmv_fake									S S	S S	S S	S S	S S	S S	S S	S S
Bkgd_QCDmv_eff									S S	S S	S S	S S	S S	S S	S S	S S

Table 5.4: Summary of the systematics uncertainties across the sixteen channels. The values given here correspond to the change in normalization from a flat systematic, SS standing for "shape systematic" errors. The errors having a name starting with Bkgd apply only to the background, the other to Background and Signal.

5.3 Limits obtained in the WH analysis

As seen earlier in 5.1.4, no significant signal excess in data is visible at the final stage of the analysis, after all selection cuts applied. Following the method described above and using the obtained systematic uncertainties, limits on $\sigma(p\bar{p} \rightarrow WH) \times BR(H \rightarrow b\bar{b})$ are set using the CL_S method. 95% confidence level expected and observed limits are set for each Higgs boson mass hypothesis. Table 5.5 summarizes the limits obtained for the combination of all channels, for the electron and muon separately as well as the full combination, used in the WH analysis. Expected and observed limits are given in multiples of ratios of the absolute cross section limit divided by the expected Standard Model Higgs boson production cross section. Limits can also be visualized on Figure 5.16.

Electron channel Results (RunIIa+RunIIb combined)											
Mass (GeV)	100	105	110	115	120	125	130	135	140	145	150
Expected Limit $/\sigma_{SM}$	4.39	4.78	5.57	6.13	7.53	9.09	11.46	14.99	21.44	31.09	66.91
Observed Limit $/\sigma_{SM}$	4.40	4.89	5.09	6.11	9.28	8.55	9.19	12.35	17.77	24.42	64.18
Muon channel Results (RunIIa+RunIIb combined)											
Mass (GeV)	100	105	110	115	120	125	130	135	140	145	150
Expected Limit $/\sigma_{SM}$	4.68	5.03	5.98	6.86	7.90	9.64	12.46	16.57	23.11	33.20	75.10
Observed Limit $/\sigma_{SM}$	4.12	5.65	5.60	5.78	6.29	12.31	10.46	13.86	21.23	25.69	75.10
Full Combined Results											
Mass (GeV)	100	105	110	115	120	125	130	135	140	145	150
Expected Limit $/\sigma_{SM}$	3.54	3.72	4.44	4.94	6.01	7.05	9.09	12.00	17.28	24.60	38.15
Observed Limit $/\sigma_{SM}$	3.40	4.77	3.70	4.62	6.13	7.46	7.86	8.55	12.78	14.00	41.81

Table 5.5: Expected and observed limits at 95% C.L. as a ratio to the standard model cross section for all considered Higgs boson masses. Electron and muon channels are shown separately as well as the full combination of all channels of the WH analysis.

The LLR distributions over all Higgs boson mass hypotheses is shown in Figures 5.14 for the electron and muon channel separately and in Figure 5.15 for the combination of all channels in the WH analysis. One can observe that with the amount of analyzed integrated luminosity and techniques employed in the analyses, the separation between the Signal+Background and Background hypotheses is less than 1σ . A combination between all channels in $D\bar{O}$ and CDF is necessary to achieve the sensitivity to Standard Model Higgs production.

The limit ratio obtained translates into an observed cross section limit

$$\sigma(p\bar{p} \rightarrow WH) \times B(H \rightarrow b\bar{b}) < 0.585 \text{ pb at 95\% C.L.} \quad (5.28)$$

for a Higgs boson mass of 115 GeV. The cross section that is probed is approximately twice smaller than the single-top theoretical cross section times leptonic branching ratio, which is one of the lowest for a $\ell\nu jj$ final state process. The corresponding expected upper limit is 0.624 pb, to be compared to a standard model cross section expectation of 0.13 pb.

In Figure 5.17 is given the ratio of limits obtained without systematics to limits obtained with systematics after constraint to data. The estimated loss in sensitivity is about 25%.

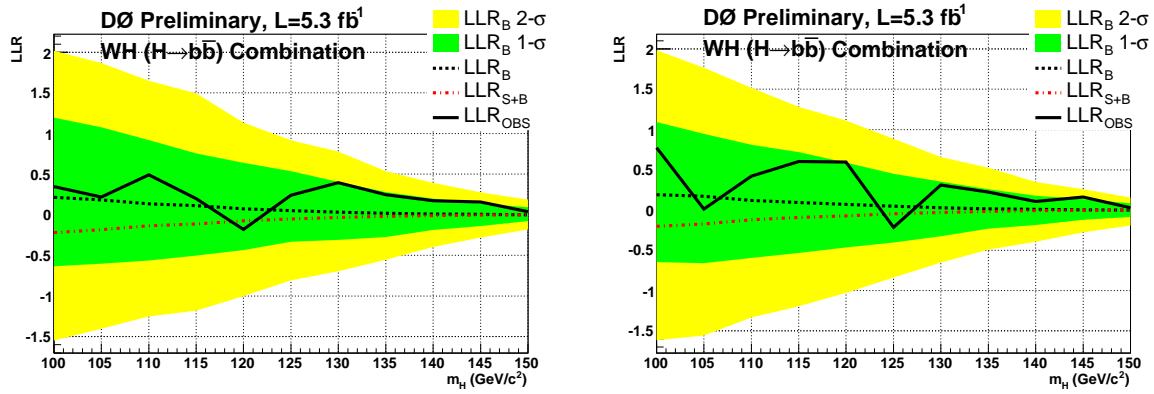


Figure 5.14: Log-likelihood ratio (LLR) distribution as a function of the Higgs boson mass, after combination of RunIIa and RunIIb data, 2 and 3 jets channels, separately for the electron channel (left figure) and the muon channel (right figure).

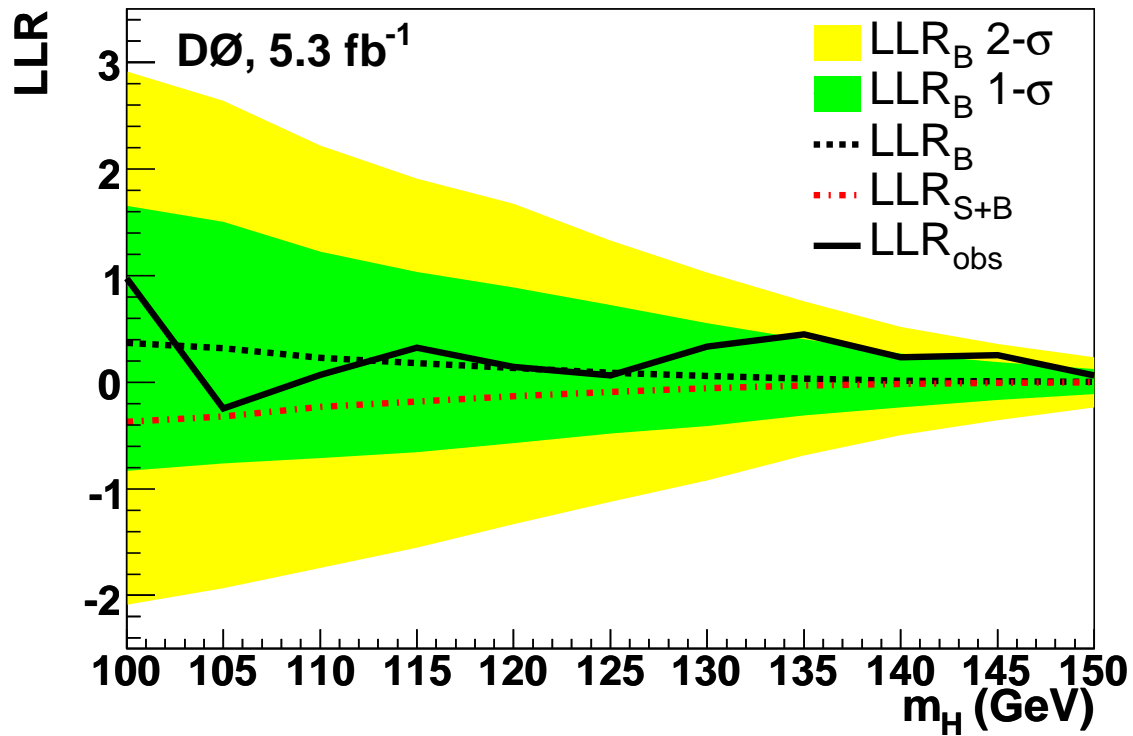


Figure 5.15: Log-likelihood ratio (LLR) distribution as a function of the Higgs boson mass, for all channels combined.

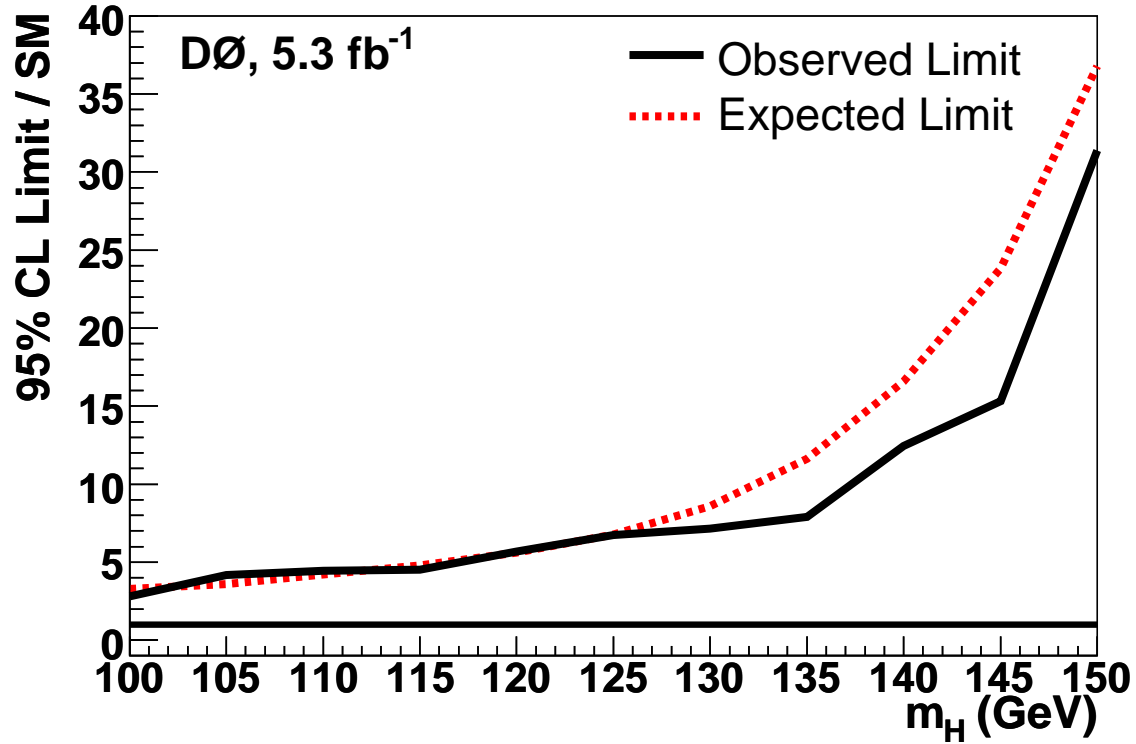


Figure 5.16: Ratio of the 95% C.L. limit cross section times the of $H \rightarrow b\bar{b}$ branching ratio to the SM prediction as a function of the Higgs boson mass, for all channels combined. The black curve corresponds to the observed limit, obtained from data, and the dotted red curve corresponds to the expected limit, obtained by replacing the observed data with the median background expectation. The straight line represents the SM predicted value. By crossing this line at a given mass value, a Higgs boson would be excluded at 95% C.L. For $M_H = 115$ GeV, an upper observed (expected) limit is set to 4.62 (4.94) on the ratio.

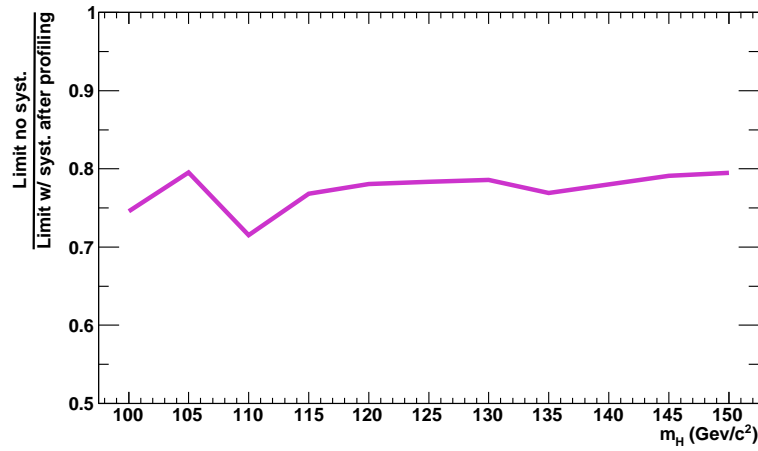


Figure 5.17: Ratio of limits obtained without and with systematic uncertainties (using the profile likelihood technique) as function of the Higgs Boson mass. The impact of systematics on the sensitivity of the analysis is of the order of 25%.

The background-subtracted data points for the RF discriminant for $M_H = 115$ GeV, for all channels combined, are shown with their systematic uncertainties in Figure 5.18. It is clearly visible that the constraint to data during the limit setting provides a significant decrease of the overall uncertainty. Finally, Figure 5.19 shows the results of the best fit to data for the different systematic uncertainties and the deviation from the *a priori* assigned values.

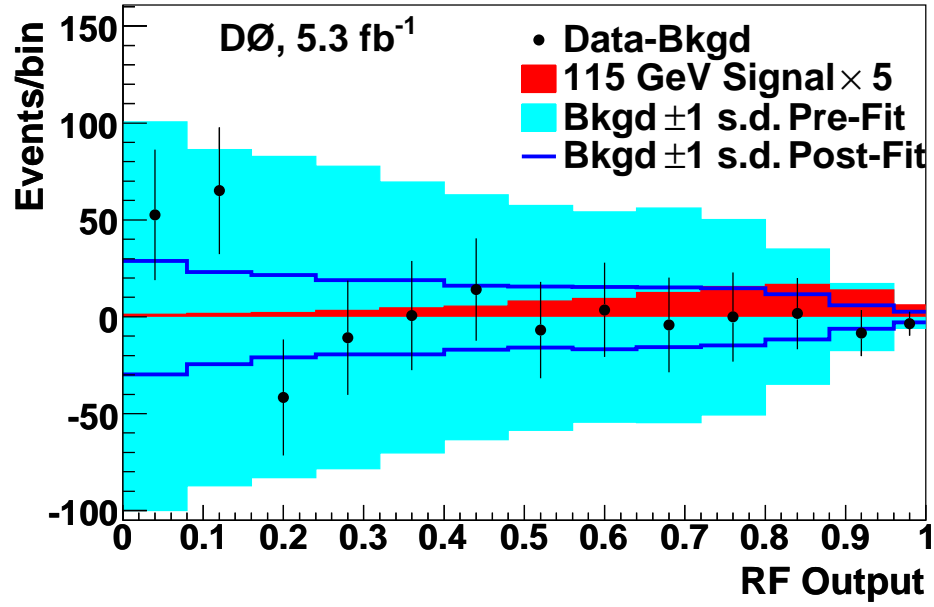


Figure 5.18: Distribution in the output of the RF discriminant for $M_H = 115$ GeV, for the difference between data and background expectation, combined for all channels (both e and μ , one and two b -tagged jets, and 2-jet and 3-jet channels), shown with statistical uncertainties. The lightly-shaded region represents the total systematic uncertainty before using constraints from data (referred to as “Pre-Fit” in the legend), while the solid lines represent the total systematic uncertainty after constraint on data (“Post-Fit” in the legend). The darker shaded region represents the SM Higgs signal expectation scaled up by a factor of 5.

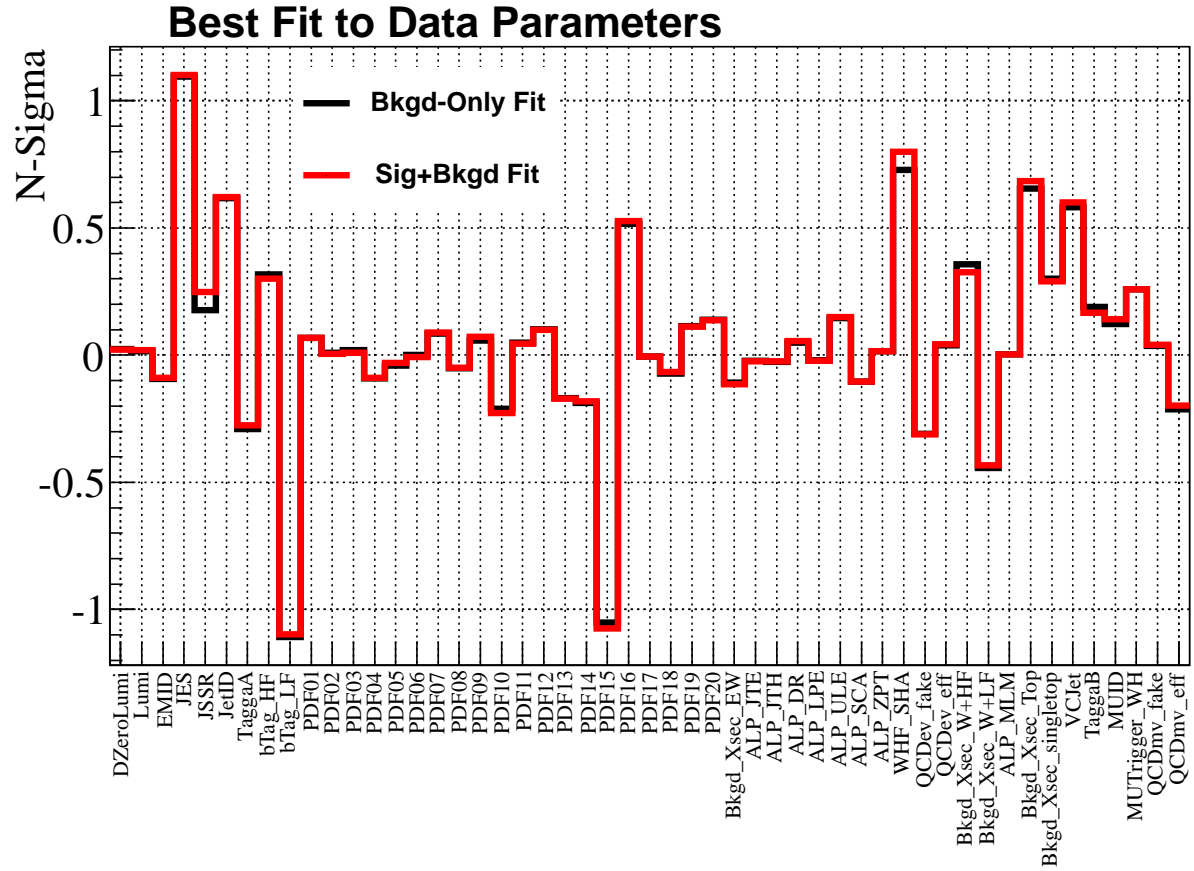


Figure 5.19: Deviation of each systematic uncertainty from its nominal value in units of standard deviations. The black line denotes the fit result for the Background-only hypothesis and the red line denotes the fit result assuming the Signal+Background hypothesis.

6

Future improvements and prospects

Now that the Tevatron gets close to its final shutdown, every improvement is important in order to gain sensitivity. In this section, different improvements in the WH analysis will be discussed. Since the dijet invariant mass is the single most powerful variable of the analysis, some studies have been conducted to increase its discrimination between the signal which is a resonance and backgrounds.

Also, many changes are currently developed at the selection level of the analysis, as well as new algorithms used to isolate as much as possible signal from backgrounds such as multivariate analysis techniques and b -tagging.

The contribution of the author was to explore potential sensitivity improvements. A set of jet energy corrections developed within the DØ collaboration have been implemented in the WH analysis. The validation is based whether the limits visibly improves by using them in the analysis. Also, a novel technique has been derived and tested, based on the jet shape to improve the resolution.

In parallel of analysis developments, studies on optimization of multivariate analysis techniques has been carried. After the Random Forest technique being introduced (see Section 5.1), a new technique from the TMVA package has been tested and significant improvement has been confirmed. Others aspects on the optimization such as statistics enhancement or $Wb\bar{b}$ and $t\bar{t}$ dedicated trainings have been explored and are described in Section 6.1.2.

After the result presented earlier have been published [1], the following iteration of the analysis is intended to be based on this result, used as “reference”. One major change comes from new b -tagger algorithm provided by the collaboration. The result of the study shown in this section indicates that a significant improvement can be achieved, by using the b -tagger output as an MVA input variable.

An overview of jet corrections will be first given, as well as the new method based on jet shapes. Improvement from optimizing multivariate discriminants in the analysis will be presented, along with

the Summer 2011 preliminary results of the WH analysis, where some of these improvements have been implemented. Finally, prospects at the Tevatron and LHC will be presented.

6.1 Improvements in the WH analysis

6.1.1 Jet Energy Resolution

As mentioned before, the dijet invariant mass is the most discriminating variable in the WH analysis and more generally in low mass Higgs boson searches, with two jets produced from the resonant decay $H \rightarrow b\bar{b}$. Since its dijet mass distribution is a narrow resonance located around the value at which the signal has been generated, its resolution is a powerful discriminant against the non resonant backgrounds.

However, as shown in Figure 6.1, one can observe that the mean value of the signal distribution is closer to 100 GeV than the true value (115 GeV in Figure 6.1), hence confronted to a bigger amount of background. Moreover, instead of being a narrow resonance, the signal displays a broad dispersion coming from experimental and physics effects:

- the calorimeter has a finite energy resolution (see Section ??),
- part of the jet energy can escape with a neutrino produced in a semi-leptonic decay $b \rightarrow \ell \nu X$ as previously mentioned in Section 3.5.3,
- b -jets have a larger spatial spread, therefore using a wider jet cone in the reconstruction may be more efficient to cluster all the energy of the jet and recover from energy leakage outside of $\Delta R = 0.5$ jet cone.
- one of the jet from the decay of the Higgs boson can radiate an energetic gluon, which can initiate a third jet. Not taking into account this additional jet can lead to a decrease in the dijet mass resolution and lower the reconstructed invariant mass.

Different attempts have been made to improve the jet resolution. The dijet invariant mass peak for the signal would thus be narrower and shifted closer to the expected value.

A set of corrections provided from the collaboration have been tested in the WH analysis. Another correction based on the jet width information has been derived and future improvements will be described.

Dijet mass corrections

Several corrections have been explored by the $D\bar{0}$ collaboration, aimed in particular to be applied in low mass Higgs analyses. Other informations can be added to reconstructed jets in order to improve the dijet mass resolution. The investigated corrections are the following:

- The energy of the CPS can be used in order to recover energy losses in dead material of the detector. More details can be found in [124] about the CPS calibration for jets.

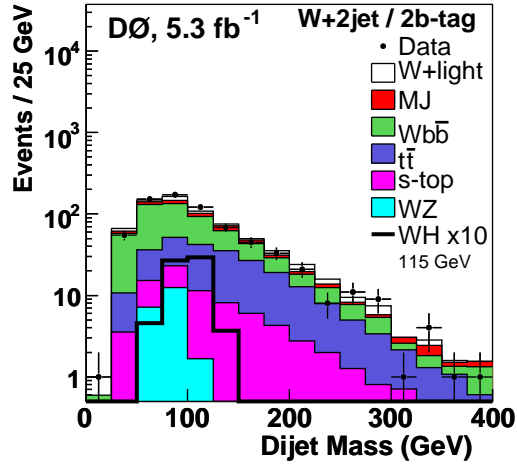


Figure 6.1: Dijet invariant mass distribution with 2 b -tagged jets for electron and muon channels combined in logarithmic scale.

- The so-called TrackCal jet correction is aimed to take into account tracks from jets in the energy calculation. If a track is associated to an energy cluster in the calorimeter, the calorimeter energy is subtracted and the associated track momentum is added. Such correction is derived independently for data and MC.
- In order to calculate the Higgs invariant mass at the parton level where detector effects would not be present (and not oversmearing the dijet invariant mass), a method relying on a multivariate technique (H-Matrix) has been developed [125]. By exploiting correlations between the parton energy and a set of observables, it is possible to provide a new estimate of the parton energy. This correction is derived from MC only and is tuned separately for light and heavy flavor jets.
- An extension of the Jet Energy Scale correction exists for the case of muons detected inside a jet. A new correction has been developed to correct jets when no muon decay has been identified. This correction E_{parton}/E_{JES} is parametrized as function of the hadronic mass of jets.
- Finally, a combination of all previous corrections are combined in a neural network (briefly described in Section 3.7.4) and is tuned separately on light and heavy flavor jets.

The above corrections have been implemented and tested within the WH analysis. To best exploit the effect of these corrections, a multivariate analysis technique has been trained in order to quantify the effect on the limits.

The improved dijet invariant mass distributions are shown on Figure 6.2. However, by comparing the limits obtained using these modifications, no significant improvement is found in the final result, therefore the decision to not use these modified variables in the WH analysis has been taken. Figure 6.3 depicts the relative change in the limit in various configurations of RF training.

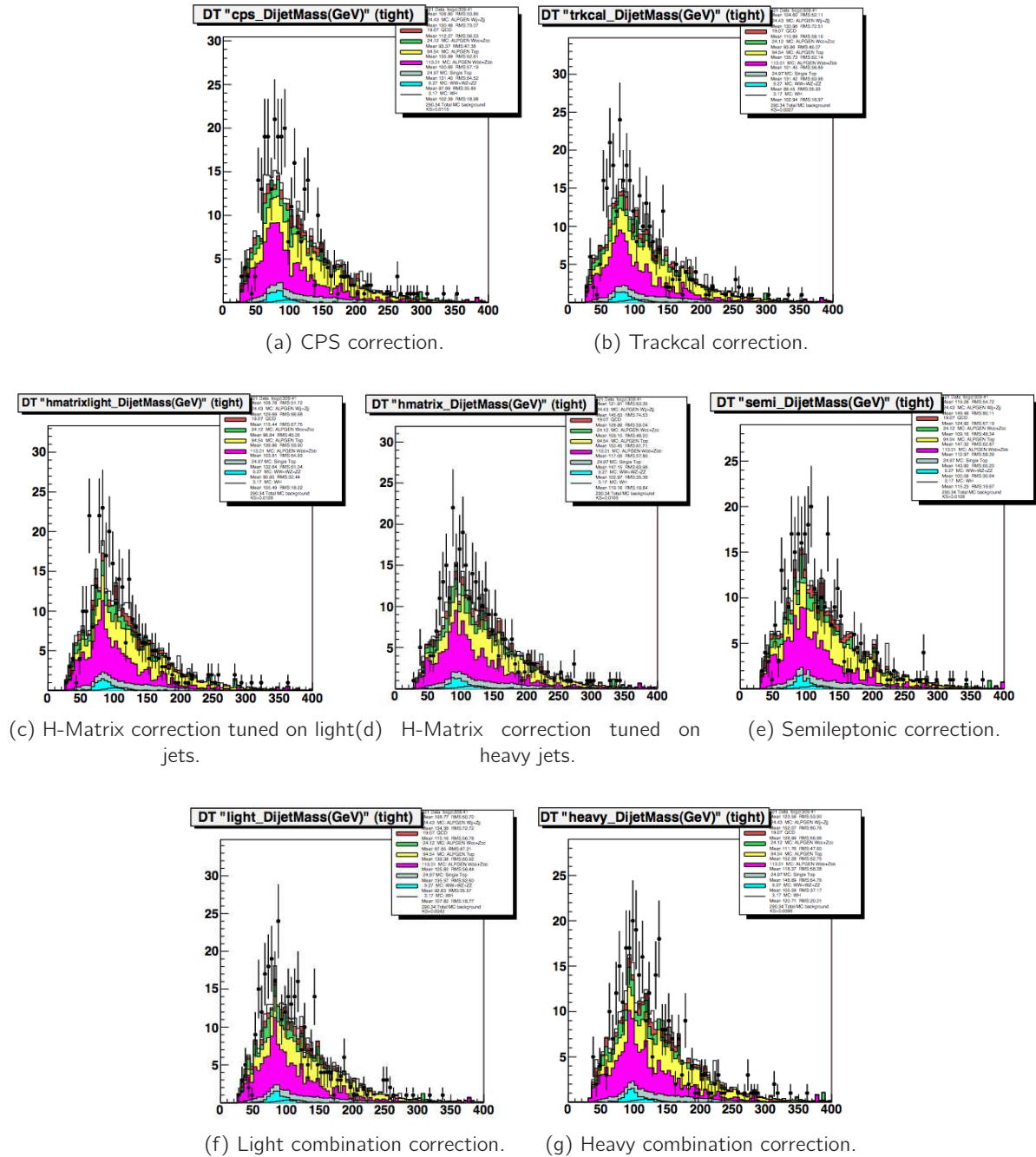


Figure 6.2: Dijet invariant mass distributions with jet resolution corrections applied, with 2 b -tagged jets.

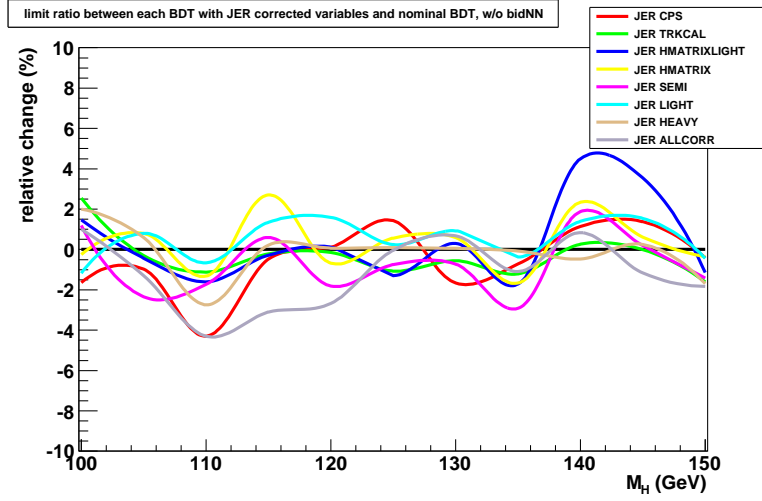


Figure 6.3: Relative change on the limit by adding one corrected dijet invariant mass as input variable, and all the 7 corrected dijet invariant masses (grey curve).

Jet width information

An additional correction related to the jet shape can be derived on top of the Jet Energy Scale [126] since it does not include any dependence on the jet width. Previous studies [127] with dijet events showed that the response of the calorimeter can be different according to the jet width, defined as:

$$w_{jet} = \sqrt{\left(\frac{\sum \phi_{cell} \times E_{cell}}{E_{jet}}\right)^2 + \left(\frac{\sum \eta_{cell} \times E_{cell}}{E_{jet}}\right)^2}, \quad (6.1)$$

where E_{cell} , ϕ_{cell} and η_{cell} are defined as the energy, azimuthal angle and pseudorapidity of a calorimetric cell contained within a jet.

The method described here is aimed to improve the jet energy resolution $\frac{\sigma(p_T)}{p_T}$.

Let $a = \frac{p_T^1 - p_T^2}{p_T^1 + p_T^2}$. The variance of a can be expressed as $\sigma_a^2 = \left|\frac{\partial a}{\partial p_T^1}\right| \sigma_{p_T^1}^2 + \left|\frac{\partial a}{\partial p_T^2}\right| \sigma_{p_T^2}^2$. For two jets in the same range of p_T and pseudo-rapidity,

$$RMS(a) = \frac{\sqrt{(RMS(p_T^1))^2 + (RMS(p_T^2))^2}}{\langle p_T^1 + p_T^2 \rangle} = \frac{\sigma_{p_T}}{\sqrt{2} p_T}. \quad (6.2)$$

In order to deal with gaussian distributions, an additional term is introduced. Therefore the jet energy resolution can be estimated as function of the variance of the dijet asymmetry variable A :

$$\frac{\sigma(p_T)}{p_T} = \sigma(A), \quad A = \sqrt{2} \cdot \text{sign}(\eta_1 - \eta_2) \cdot \frac{p_T^1 - p_T^2}{p_T^1 + p_T^2} \quad (6.3)$$

where $p_T^{1,2}$ is the transverse momentum of the leading (sub-leading) jet and $\eta^{1,2}$ is the detector pseudo-rapidity of the leading (sub-leading) jet. The randomness of $\text{sign}(\eta_1 - \eta_2)$ ensures that the distribution of A to be symmetrically distributed around 0.

The aim of this method is to minimize $\sigma(A)$ with correction factors to the jets transverse momenta. The best correction factors are assumed to be the ones which minimize $\sigma(A)$.

Event selection For the purpose of this study, samples used to derive the correction factors are:

- Data triggered by jet inclusive triggers or zero bias (ZBMB) triggers corresponding to the RunIIb1, RunIIb2 and RunIIb3 data periods,
- Dijet Monte Carlo samples generated by PYTHIA

MC samples are generated in 8 \hat{p}_T bins¹. After processing, they are individually normalized to the LO cross section. Events are selected with the following requirements:

- 2 Vertex Confirmed jets,
- $|\Delta\phi(j,j)| \geq 3$, to select well balanced jet in the ϕ coordinate.

The analysis is performed in different bins:

- Both jets in the same calorimeter region, respectively CC, EC or ICR,
- 10 jet width bins from 0 to 0.45 and above, by steps of 0.05,
- 3 p_T^{mean} bins (defined as $p_T^{mean} = \frac{p_T^1 + p_T^2}{2}$), from 15 to 40 GeV, from 40 to 80 GeV and above 80.

For data, the ZBMB triggers are used for the lowest p_T^{mean} bin. The two other bins contain events fired respectively by the JT25 and JT45 triggers. In order to not introduce any trigger bias, both jets are required to match to the L1/L2/L3 triggers requirements.

Method The optimal jet correction factor k should be the one that minimizes $\sigma(A)$. One can rewrite Equation 6.3:

$$A = \sqrt{2} \cdot \text{sign}(\eta_1 - \eta_2) \cdot \frac{k(w_1) \cdot p_T^1 - k(w_2) \cdot p_T^2}{k(w_1) \cdot p_T^1 + k(w_2) \cdot p_T^2}, \quad (6.4)$$

where the $k(w_i)$ factor is parametrized as a function of the i^{th} jet width bin. For a given jet width bin, one of the jets or both can be used in the minimization, only jet in the bin will have $k(w_i)$ varied, the other is kept constant to 1.

In the case that both jets are within the same jet width bin, both jets' $k(w_i)$ will be varied. Note that this type of event are not contributing for the minimization of the dijet asymmetry since the same factor is multiplied to both jets in the numerator and denominator of Equation 6.4.

For each different bin in the $\{p_T^{mean}, w_i, \eta\}$ space, the $k(w_i)$ factor is varied from 0.5 to 2 with 400 steps. Such variation of the resolution (or $\sigma(A)$) can be seen in Figure 6.4, where the thickness of the red curve represents the error on the RMS of A .

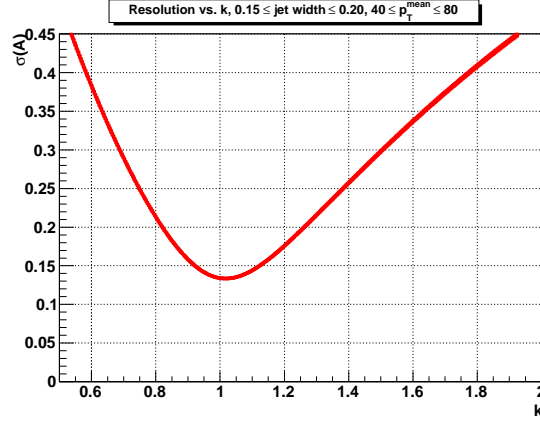


Figure 6.4: Variation of $\sigma(A)$ as function of k , for $40 \leq p_T^{mean} \leq 80$ GeV, $0.15 \leq w_i \leq 0.2$, in CC region.

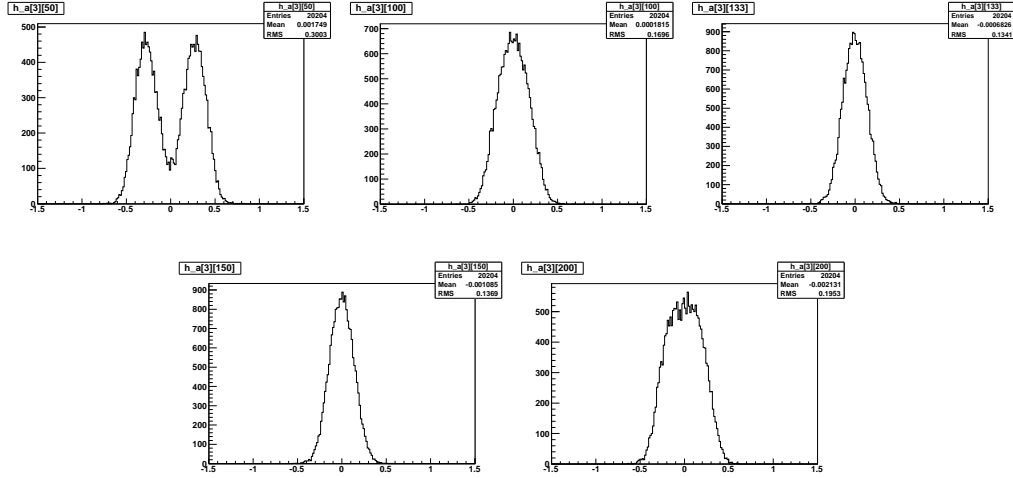


Figure 6.5: A in data by increasing k (0.6875, 0.875, 1 on the upper row, and 1.0625, 1.25 in the bottom row, from left to right), which corresponds to step 50, 100, 133, 150, 200. The non-gaussian behaviour is a consequence of using k values being far from the nominal case $k = 1$, forcing the jets to not be balanced in p_T .

Figure 6.5 shows the dijet asymmetry after applying different values of k (0.6875, 0.875, 1, 1.0625, 1.25) for data.

In order to find the minimum point k_0 , the red curve shown in Figure 6.4 is fitted by a second order polynomial. The correction factor is taken as the minimum of this function. f^\pm correspond to $\pm 1\sigma$ of the nominal function f , estimated from the fit to the obtained $\sigma(A)$. Associated errors

¹ \hat{p}_T is the center of mass transverse energy used to generate events. They are split as following: 10-20, 20-40, 40-80, 80-160, 160-320, 320-480, 480-640 and 640-980 GeV.

k_{min} and k_{max} are taken where $f^-(k_{min}) = f(k_0)$ (see Figure 6.6, blue line).

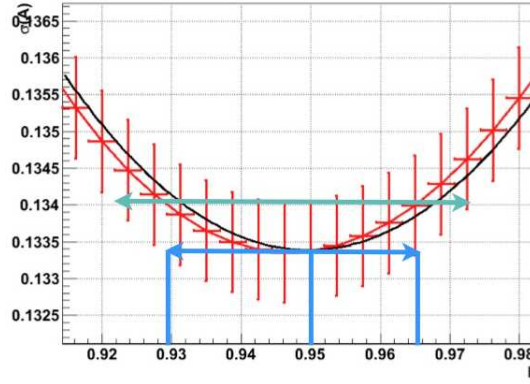


Figure 6.6: Zoom around minimum value of $\sigma(A)$ after scanning over the k parameter. The associated error on the correction is estimated from the region delimited by the blue arrow.

Results The method explained above is done separately for data and simulation, for all jet widths, p_T^{mean} and η_{det} bins. The estimated correction factors are showed as a function of jet width on Figure 6.7, where a continuous correction can be obtained by fitting the correction factors with a second order polynomial. Figure 6.8 shows the resulting k factor in the $\{p_T^{mean}, w_{jet}\}$ plane for data and Monte Carlo.

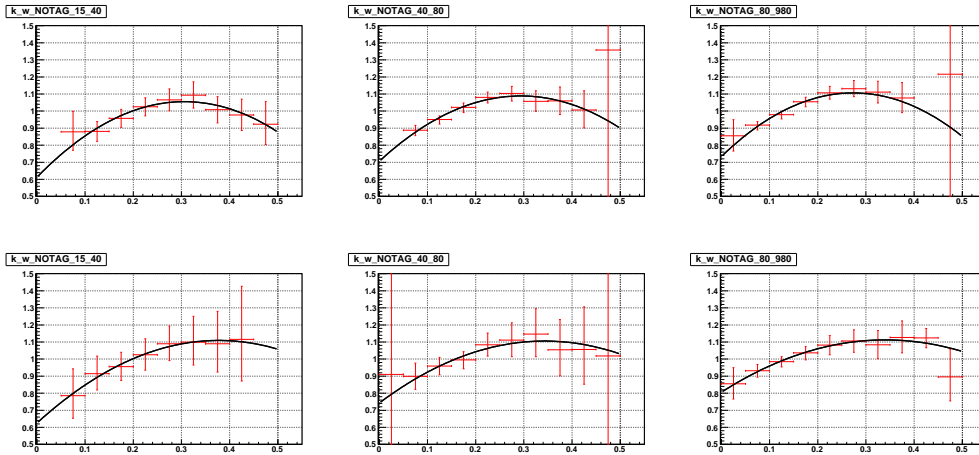


Figure 6.7: k factor obtained for data (upper row) and Monte Carlo (bottom row) in CC region, for $15 \leq p_T^{mean} \leq 40$ (left), $40 \leq p_T^{mean} \leq 80$ (middle) and $p_T^{mean} \geq 80$ GeV (right).

The method is validated by applying the correction factor to each jet and evaluating the improvement on $\sigma(A)$, as showed in Figure 6.9 for dijet events in data and in Figure 6.10 for Monte

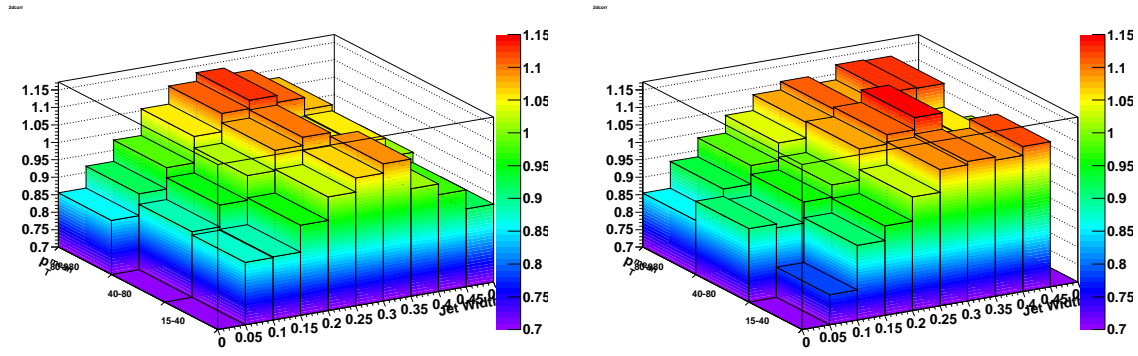


Figure 6.8: Corrections shown as function of the jet width and p_T^{mean} for data (left) and Monte Carlo (right).

Carlo. The black (red) histograms are before (after) applying the appropriate correction factors. A summary of $\sigma(A)$, before and after correction, is given in Table 6.1 for data and in Table 6.2 for MC.

		CC	ICR	EC
$15 < p_T^{mean} < 40$ GeV	before correction	0.2424	0.2566	0.2360
	after correction	0.2350	0.2504	0.2206
	relative change (%)	3.05	2.42	6.53
$40 < p_T^{mean} < 80$ GeV	before correction	0.1398	0.1494	0.1243
	after correction	0.1281	0.1394	0.1128
	relative change (%)	8.37	6.70	9.25
$p_T^{mean} > 80$ GeV	before correction	0.1174	0.1371	0.1112
	after correction	0.1028	0.1223	0.0973
	relative change (%)	12.44	10.80	12.50

Table 6.1: $\sigma(A)$ for data.

Cross check on Monte Carlo In the context of the search for a low mass Higgs Boson, the main analyses rely on the discrimination of b-jets to light quark/gluon jets. Usually, the final analysis is dominated by jets coming from b quarks. One needs to understand how the correction varies with the jet flavor. To give an estimate, the relative quark/gluon composition is showed in Figure 6.10 (bottom row).

The cross check is done in three cases:

- gg jets, where both jets are initiated from gluons,
- qg jets, where one jet is initiated from a gluon, the other one coming from a quark,
- qq jets, where both jets are initiated from quarks,

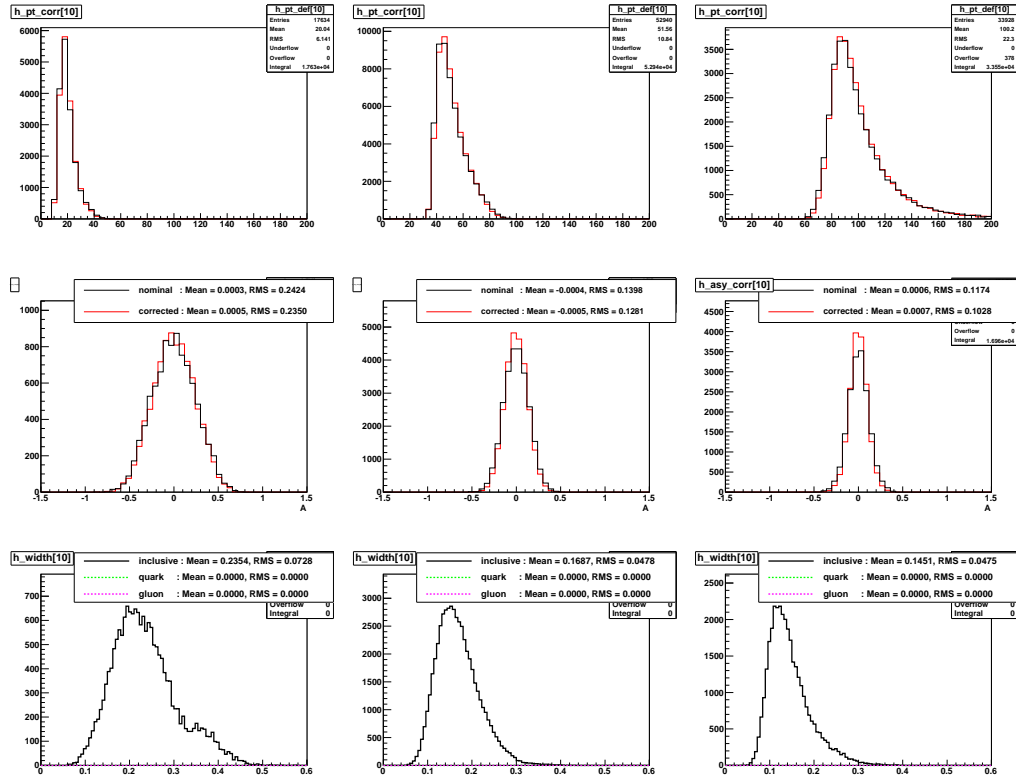


Figure 6.9: Upper row: The black (red) histograms are p_T^{jet} for data before (after) correction for $15 < p_T^{mean} < 40$ GeV (left), $40 < p_T^{mean} < 80$ GeV (center) and $p_T^{mean} > 80$ GeV (right). Middle row: The black (red) histograms are the asymmetry distributions for data before (after) correction for the same p_T^{mean} bins. Bottom row: Jet width in data for the same p_T^{mean} bins.

		CC	ICR	EC
$15 < p_T^{mean} < 40$ GeV	before correction	0.2494	0.2589	0.2122
	after correction	0.2430	0.2492	0.1937
	relative change (%)	2.57	3.75	8.72
$40 < p_T^{mean} < 80$ GeV	before correction	0.1680	0.1883	0.1539
	after correction	0.1595	0.1763	0.1563
	relative change (%)	5.06	6.37	-1.56
$p_T^{mean} > 80$ GeV	before correction	0.1173	0.1392	0.1123
	after correction	0.1087	0.1316	0.1637
	relative change (%)	7.33	5.46	-45.77

Table 6.2: $\sigma(A)$ for MC.

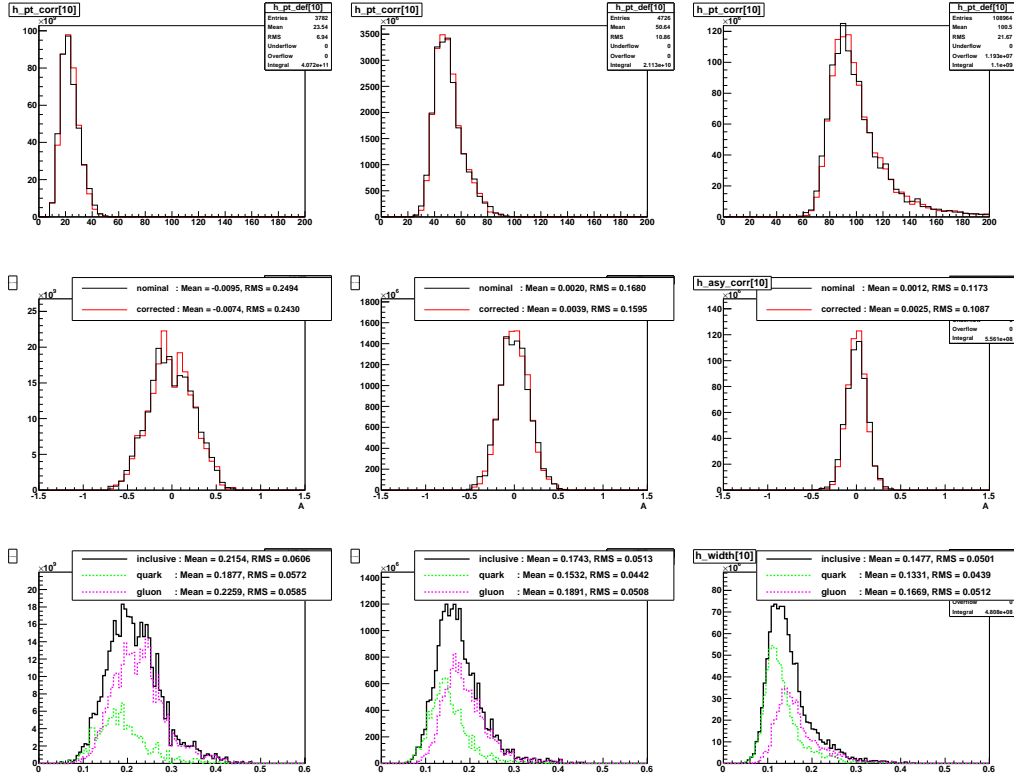


Figure 6.10: Upper row: The black (red) histograms are p_T^{jet} for MC before (after) correction for $15 < p_T^{mean} < 40$ GeV (left), $40 < p_T^{mean} < 80$ GeV (center) and $p_T^{mean} > 80$ GeV (right). Middle row: The black (red) histograms are the asymmetry distributions for MC before (after) correction for the same p_T^{mean} bins. Bottom row: Jet width in MC for quark (green), gluon (pink) and all (black) jets for the same p_T^{mean} bins. Jets are matched to quark and gluons within a ΔR cone of 0.5. Jets that are not matched are considered in the black histograms.

One can see that the resolution and the jet width vary with the jet flavor. Hence, the derived corrections could reduce this dependence, improving the jet energy resolution for both gluon and quark originated jets.

Application to the $WH \rightarrow \ell\nu b\bar{b}$ analysis In order to quantify the impact of this correction in the context of an analysis, the corrections are applied in the $WH \rightarrow \ell\nu b\bar{b}$ analysis.

In Figure 6.12 are shown the jet width distributions in the RunIIb 2 jets electron channel. Figures 6.13–6.15 show the leading, subleading jet p_T and dijet invariant mass before applying b -tagging, for the 1 b -tagged jet channel and for the 2 b -tagged jets channel.

The limit on $\sigma(p\bar{p} \rightarrow WH) \times B(H \rightarrow b\bar{b})$ is calculated using the dijet mass as the final distribution. The expected limits with and without correction applied are shown in Table 6.3, for $M_H = 115$ GeV.

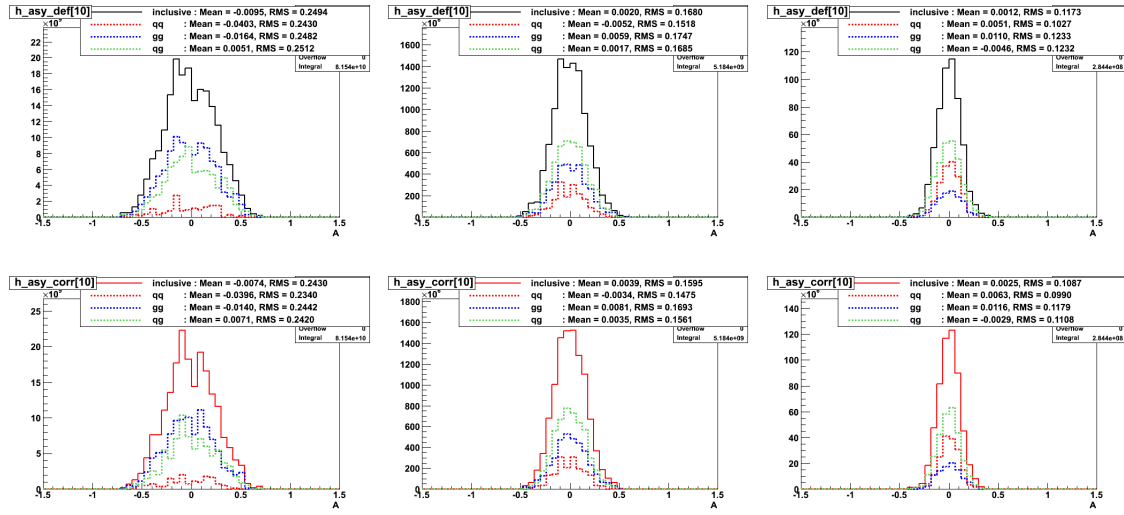


Figure 6.11: Dijet asymmetry for Monte Carlo, showing the qq , gg and $q\bar{q}$ contributions, respectively in dashed red, dashed blue and dashed green, for $15 \leq p_T^{mean} \leq 40$ (left), $40 \leq p_T^{mean} \leq 80$ (middle) and $p_T^{mean} \geq 80$ GeV (right).

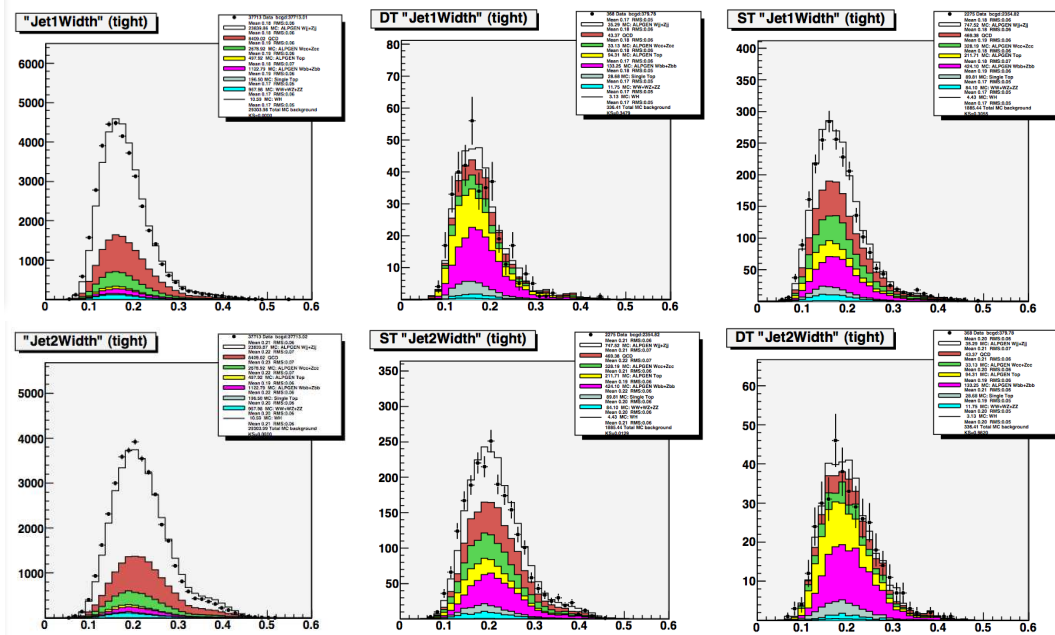


Figure 6.12: Leading jet width (upper row) and second leading jet width (bottom row), for the pretag, 1 and 2 b -tagged sample .

Conclusion Following the method described here, the improvement which can be achieved in dijet events is about 10%. Such improvement would have an important impact on the low mass

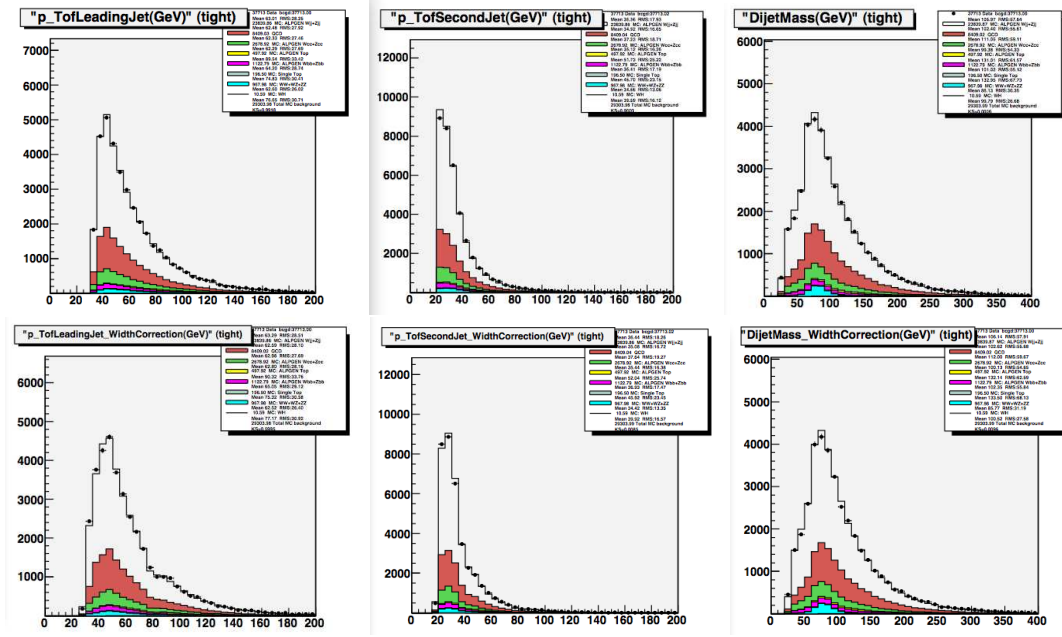


Figure 6.13: Leading jet p_T (left), second leading jet p_T (middle) and dijet invariant mass (right), before (upper row) and after (bottom row) applying the jet width correction, before applying b -tagging.

	Single tag	Double tag	Single tag + Double tag
without correction	16.343	8.337	7.435
with correction	17.395	8.933	7.949

Table 6.3: Expected limits for $M_H = 115$ GeV.

Higgs searches. Moreover, this method has a potential to improve the jet energy resolution for both gluon and quark initiated jets.

The method is applied to the WH search but the expected improvement is not observed. Cross checks are performed to check separately on selected backgrounds; the resolution is not visibly improved (see Figure 6.16). Some further studies would be needed to understand why the effect of such corrections is degrading the limits in the context of the WH analysis.

CHAPTER 6. FUTURE IMPROVEMENTS AND PROSPECTS

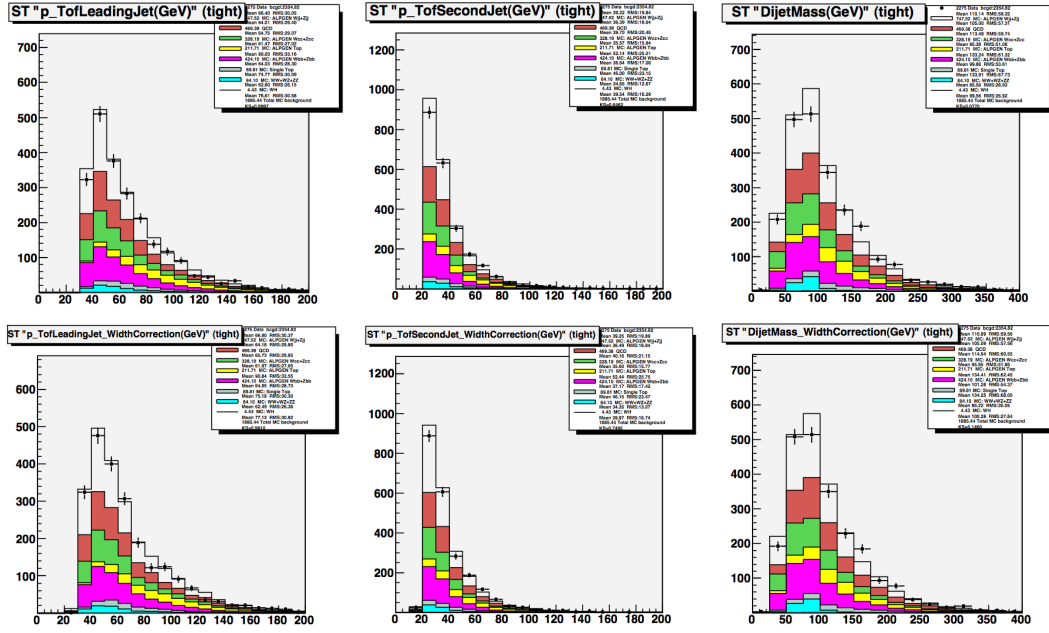


Figure 6.14: Leading jet p_T (left), second leading jet p_T (middle) and dijet invariant mass (right), before (upper row) and after (bottom row) applying the jet width correction, with 1 b -tagged jet.

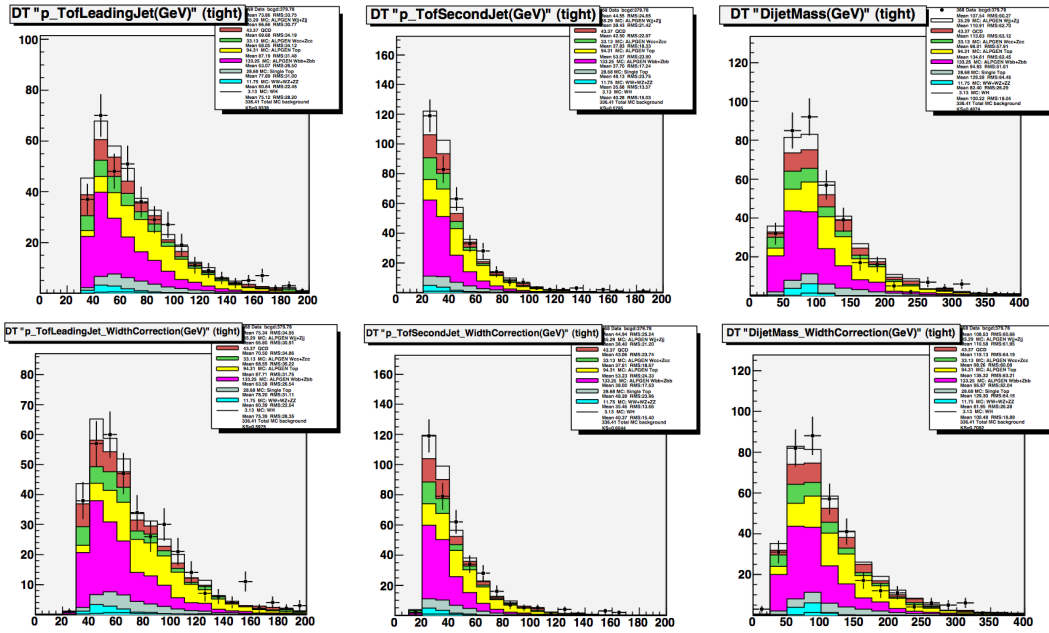


Figure 6.15: Leading jet p_T (left), second leading jet p_T (middle) and dijet invariant mass (right), before (upper row) and after (bottom row) applying the jet width correction, with 2 b -tagged jets.

6.1. IMPROVEMENTS IN THE WH ANALYSIS

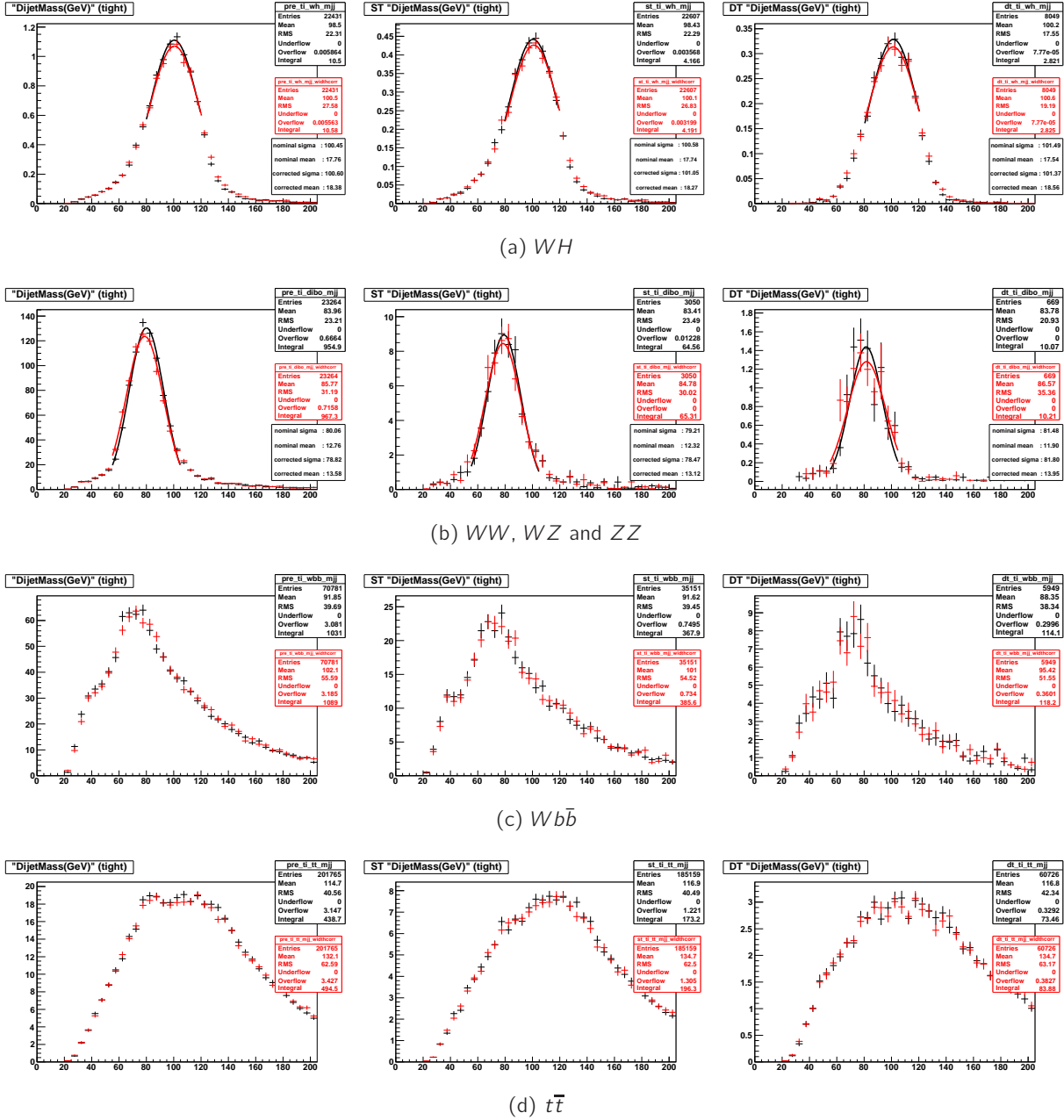


Figure 6.16: Dijet invariant mass distributions before (black histogram) and after (red histogram) correction for WH , diboson, $Wb\bar{b}$ and $t\bar{t}$ MC samples, before applying b -tagging (left), for the 1 b -tagged jet channel (middle) and for the 2 b -tagged jets (right).

Ongoing effort

Semi-leptonic corrections Used as default as Jet Energy Scale for jets containing muons, the JESMU correction introduced in Section 3.5.3, is under further development. This correction is based on simulation and is derived in order to take into account lost energy from semileptonic decays in jets, especially for b -jets.

This correction is now revisited. First, the muon-in-jet acceptance can be increased by relaxing the muon selection criteria, for example track requirements or cosmic veto, more jets can potentially be corrected. The increase in acceptance is estimated from MC truth information and to be around 10%. Moreover, instead of fitting the correction parametrization as function of E_μ , an event-by-event correction can be determined via a multivariate analysis technique¹ which response is the closest as possible to the true value. Hence, correlations between muon observables, jet observables, as well as their angular relations can be used. A few percent improvement on the dijet mass resolution are expected with respect to the current version of the algorithm.

Since it is established that JESMU performs well, the idea can be extended to recover energy losses from soft electrons in jets. However, since such object may not be well calibrated enough, the track associated to such electrons is used. Then, the parametrization of the correction, which is of the same form as the current JESMU correction, is parametrized in function of the track momentum associated to the reconstructed electron (instead of E_μ in correction formula given in Section 3.5.3), in different P_T^{rel} bins.

Preliminary results for the two types of correction of given in Figure 6.17.

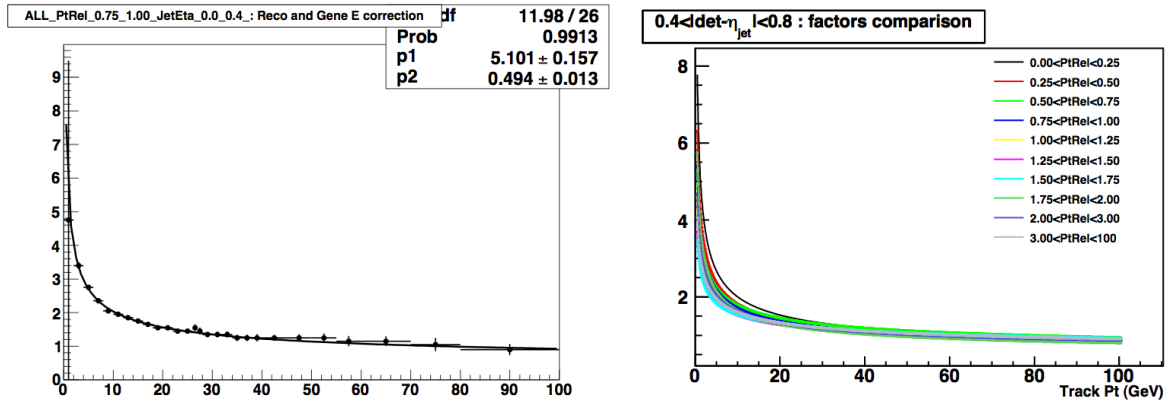


Figure 6.17: Correction factors derived for jets containing soft muons (left) or soft electrons (right).

Radiation recovery In the 3 jet channel, the dijet invariant mass is constructed by default from the two highest p_T jets. For signal events, the third jet would come either from Initial State Radiation (ISR) or Final State Radiation (FSR). FSR jets should be included in the calculation

¹also called regression analysis

of the Higgs boson candidate mass reconstruction, since it is indirectly a decay product. A new variable is defined in order to choose whether the third jet should be used to reconstruct the Higgs boson candidate, the minimum angle α_{min} between a jet and any other in the event. Studies showed that this variable can be used to discriminate ISR jets from FSR jets, given the event topology. Therefore, a decision can be taken to include or not the third jet in the Higgs boson mass reconstruction depending on the α_{min} value.

6.1.2 Changes in the analysis

After the results shown in Section 4–5 have been made public and published [1], important changes have been studied in order to increase the sensitivity. Some major changes will be detailed here.

Adding more data

As of summer 2011, the recorded luminosity is reaching 10 fb^{-1} as shown in Figure 6.18. The WH analysis will be conducted by analyzing 8.6 fb^{-1} of RunIIb data. The expected limit is scaled as $1/\sqrt{\mathcal{L}}$.

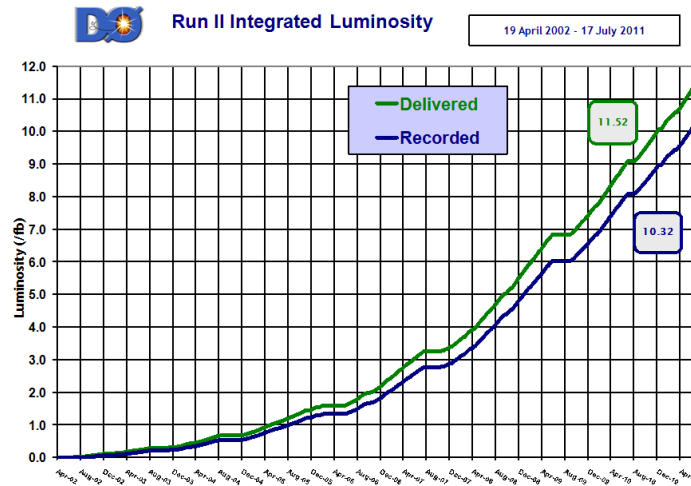


Figure 6.18: Integrated Luminosity as of Summer 2011.

Event selection

As seen in Section 4, a number of selection cuts are introduced to remove instrumental background and improve the data modeling. By relaxing these selection criteria, the signal acceptance can be increased. For example:

- The $\cancel{E}_T > 20 \text{ GeV}$ cut is relaxed to 15 GeV in the electron channel,
- Looser lepton identification criteria are used for both electron and muon selections.

- The triangular cut $M_W^T > -0.5 \cancel{E}_T + 40$ GeV is not used, a new method to reject multijet events more efficiently is introduced, see 6.1.2,
- The H_T cut is discarded,
- The primary vertex selection is extended from ± 40 cm to ± 60 cm,

These changes have the drawback of introduce a large amount of multijet events after selection.

QCD removal

In order to reduce the contribution of the multijet background especially in the electron channel, the triangle cut is replacing by a dedicated Multivariate analysis technique, aimed to reject such background while keeping as much as possible signal events.

Input variables are chosen according to their discriminating power between the signal and this particular background. Events used in the training of the so-called *MVAQCD* are events not satisfying the tight lepton requirements, therefore they will not be used in the final sample used to derive limits and avoiding biases from overtraining.

The output distribution of the MVAQCD corresponding to the RunIIb period is shown in Figure 6.19. For the same background rejection obtained from the triangle cut, the signal acceptance is increased by 18%, by selecting events with MVAQCD output greater than 0. This discriminant is trained separately for each channel and Higgs boson mass hypothesis.

Distributions after the MVAQCD cut is applied are shown in Figure 6.20–6.27. The signal contributions are multiplied by 500 in the following figures.

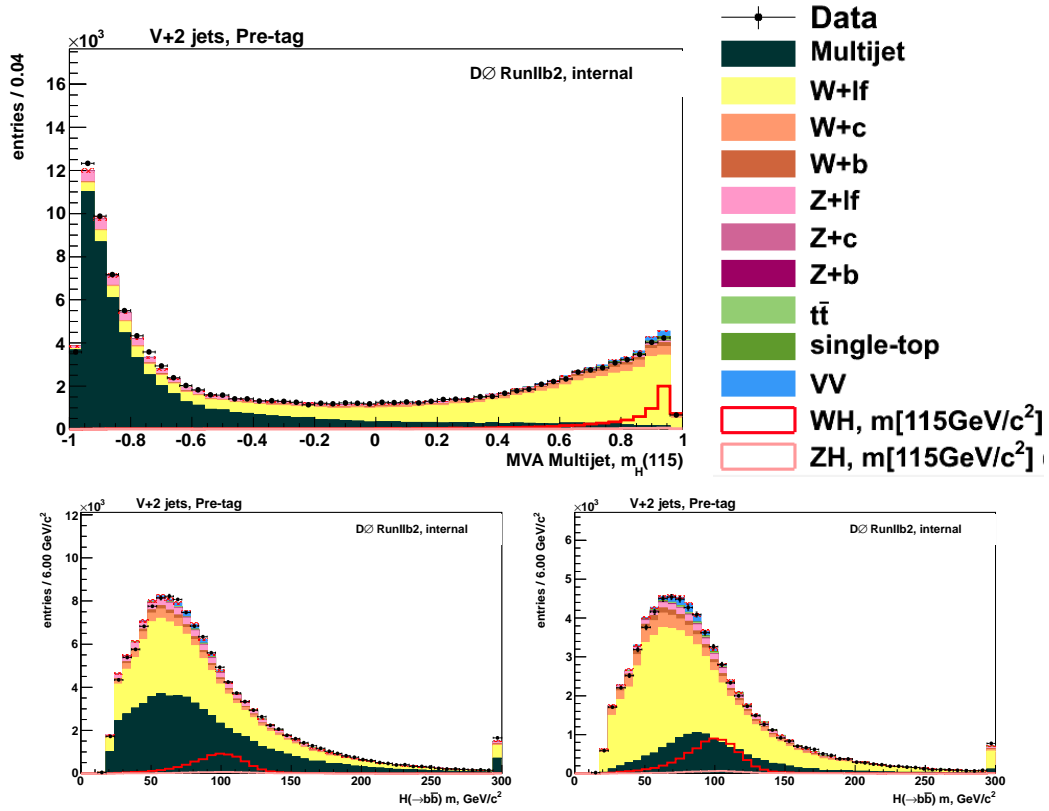


Figure 6.19: Top figure: MVAQCD output distribution. The color legend for the following plot is presented on the left. Bottom figures: Dijet invariant mass distributions before (left) and (right) cut the MVAQCD output. The reduction of the multijet component is clearly visible.

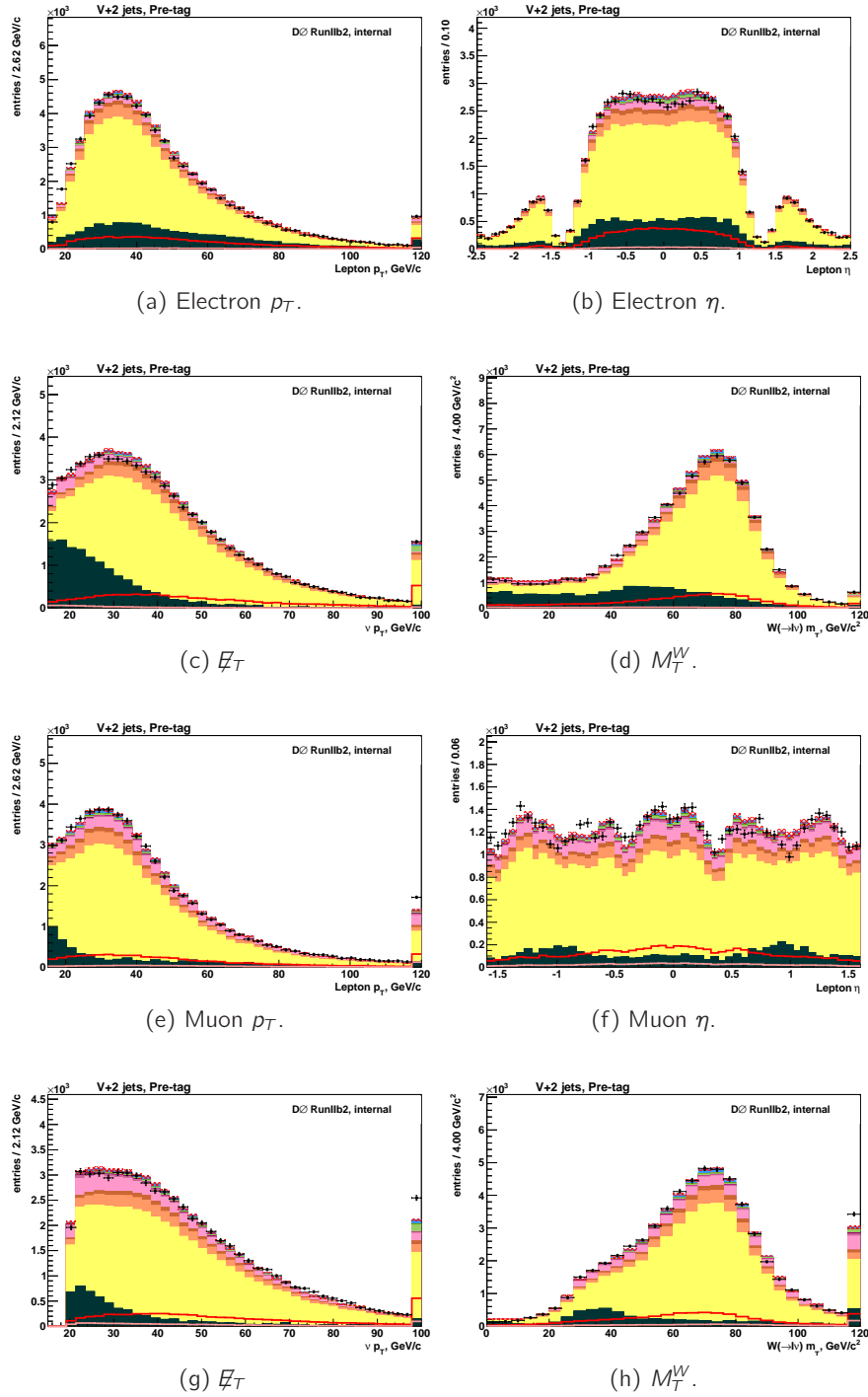


Figure 6.20: Lepton related variables for the electron (top figures) and muon channel (bottom figures).

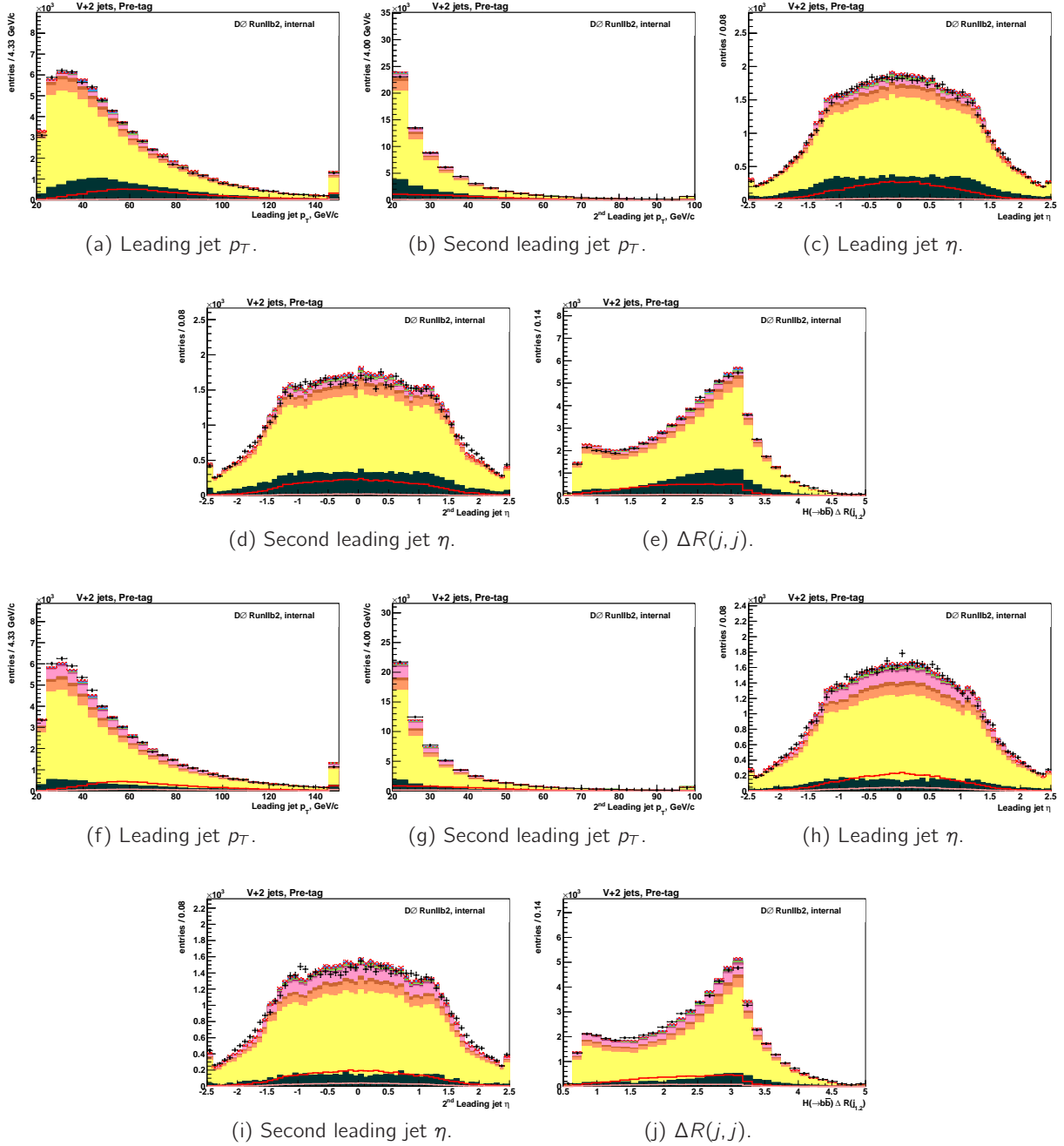


Figure 6.21: Jet related variables distributions for the electron (top figures) and muon channel (bottom figures).

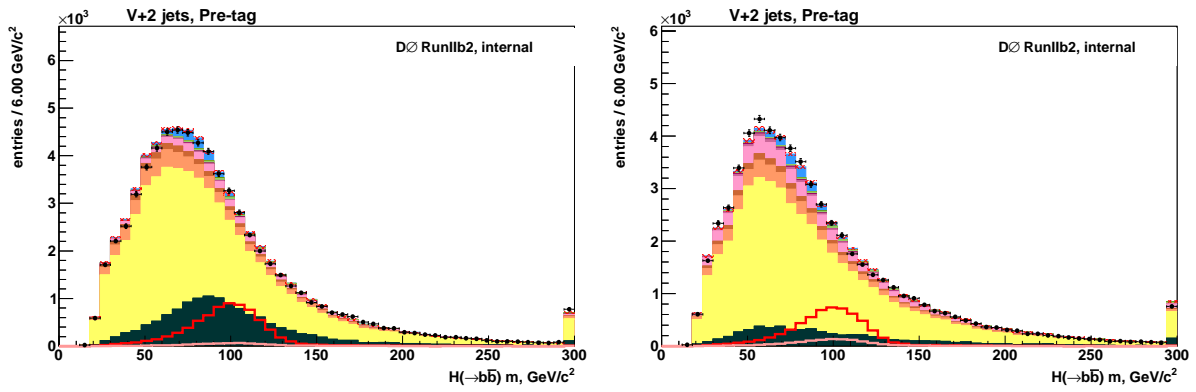


Figure 6.22: Dijet invariant mass distributions for the electron (left) and muon channel (right).

New b -tagging tools

New algorithms to identify b jets have been developed and can be now used:

- The *MVA BL* is used to discriminate b jets from light jets,
- The *MVA BC* is trained to differentiate b from c -like quarks, and could potentially help to reduce the $Wc\bar{c}$ background, which is still present after b -tagging (see Table 4.16),
- The *MVA BB* is built in order to differentiate pairs of b quarks merged in a single jet from b jets induced by a single parton, the latter being more likely to occurs in $H \rightarrow b\bar{b}$ decays.

Instead of using a neural network technique, a combination of Random Forests is employed for the discriminants. For now, only the MVA BL will be used in the WH analysis. As shown in Figure 6.23, an increase in b -jet efficiency can be reached for the same fake rate.

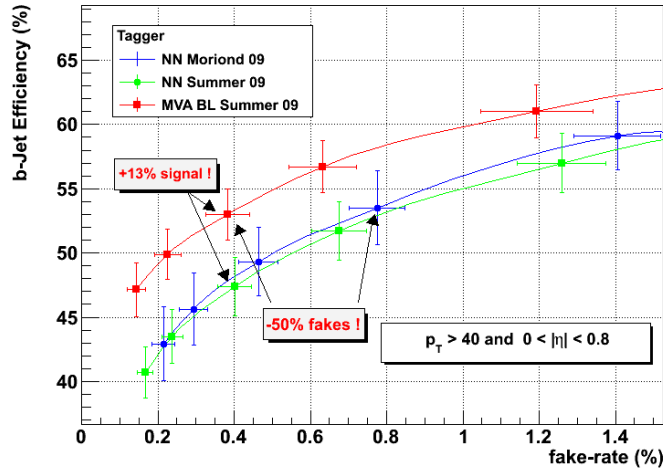


Figure 6.23: The performance of the NN and MVA BL taggers for the tighter operating points.

Following the will of increasing acceptance as much as possible, the loosest MVA BL operating points are used. The background rate is higher as expected, compared to the operating points described in 4.9. However, since the MVA BL output is well modeled (see Figure 6.24), it can be introduced to the final discriminant, thus providing the best possible discrimination between light and b jets.

Distributions obtained after requiring one L6 or two L6 b -tagged jets and a cut on the MVAQCD output are shown in Figures 6.25 and 6.27. The signal contributions are multiplied by a factor of 100 and 50 respectively for figures with one L6 or two L6 b -tagged jets.

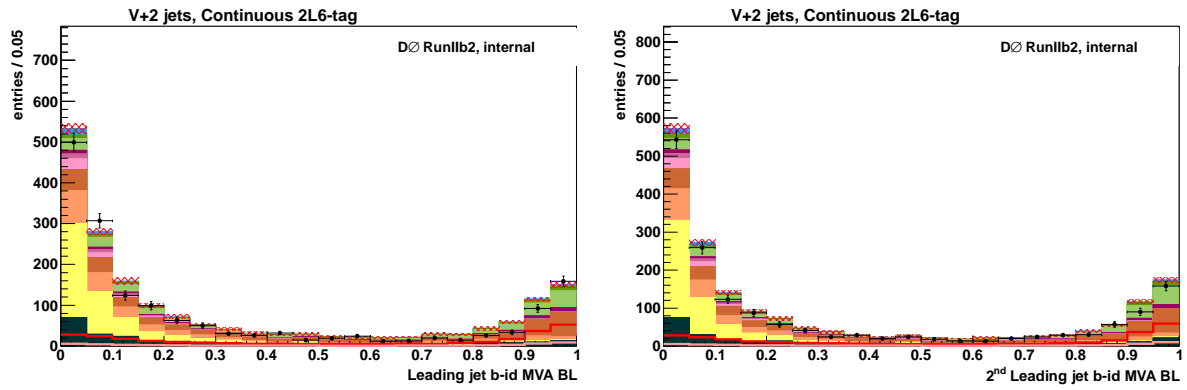


Figure 6.24: MVA BL output distributions for the leading (left) and sub-leading jet (right).

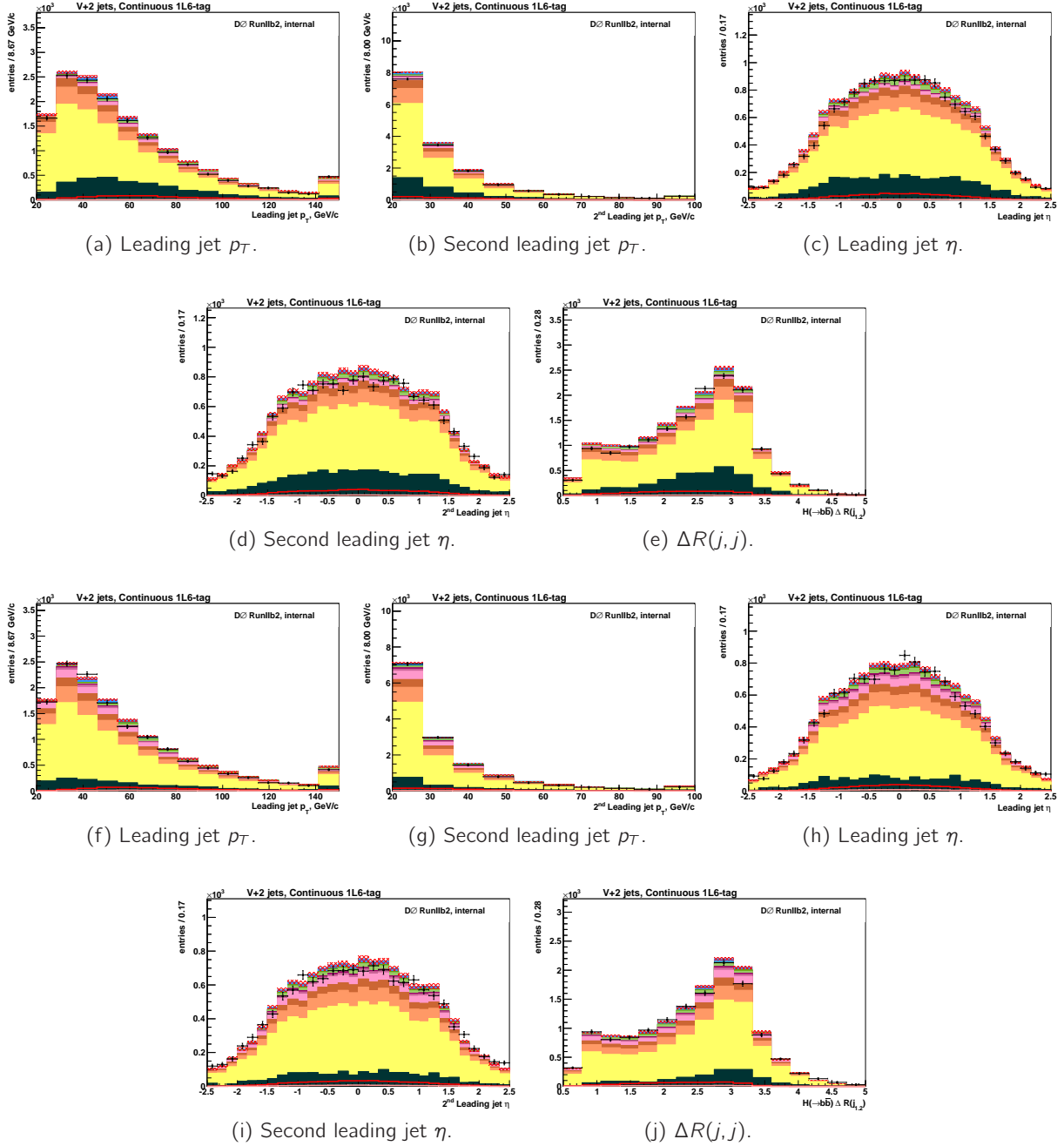


Figure 6.25: Jet related variables distributions for the electron (top figures) and muon channel (bottom figures) with 1 L6 b -tagged jet.

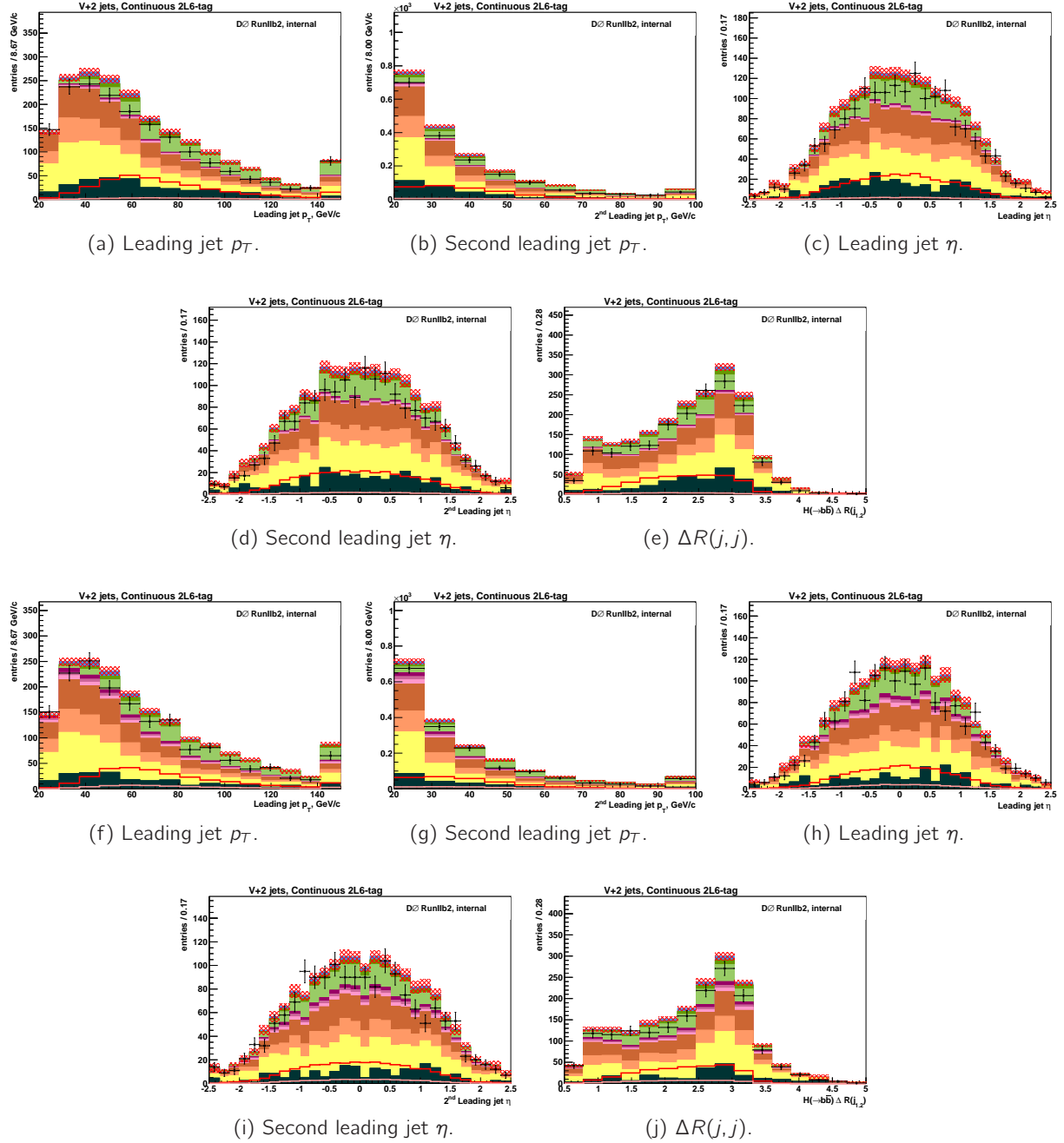


Figure 6.26: Jet related variables distributions for the electron (top figures) and muon channel (bottom figures) with 2 L6 b -tagged jets.

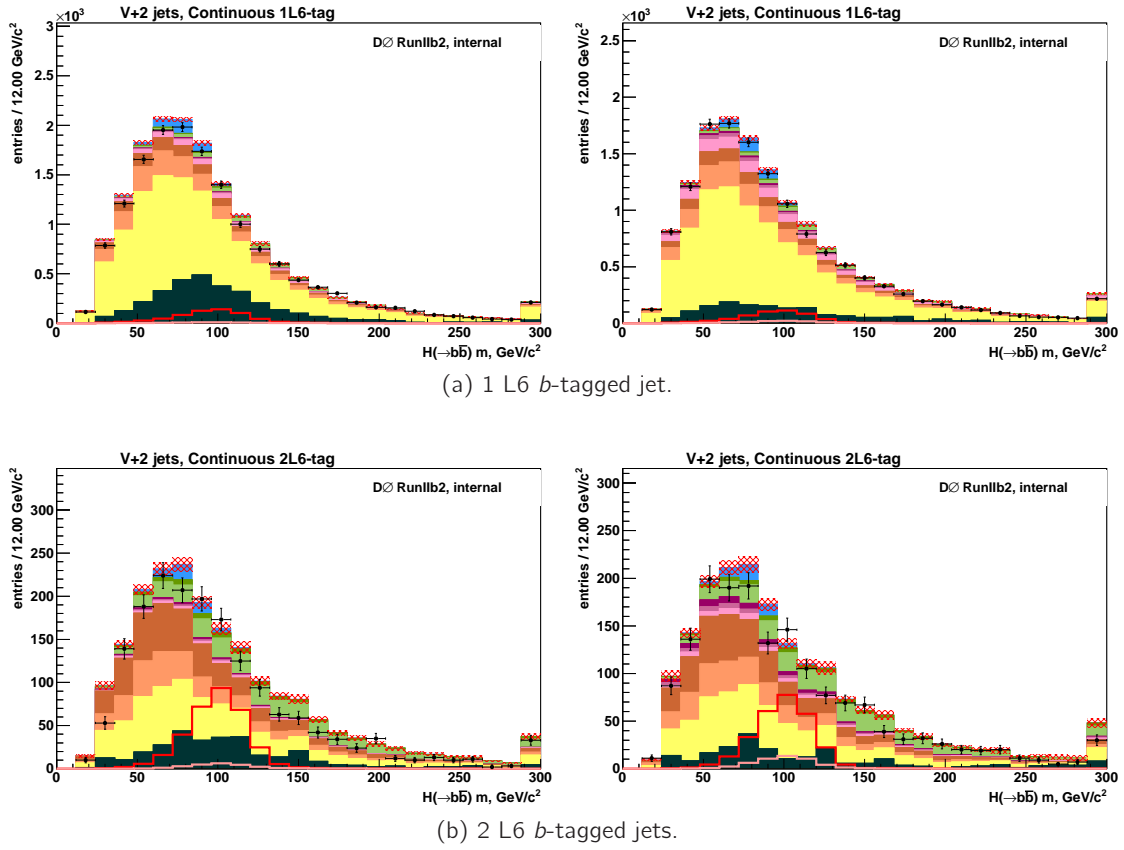


Figure 6.27: Dijet invariant mass distributions for the electron (left) and muon channel (right).

Multivariate analysis technique

Since limits are extracted from the final discriminant distribution, improvements on this aspect can lead to better discrimination between backgrounds and signal, thus a better sensitivity. Studies conducted in parallel of analysis development show that some gain is possible.

In 2009, limits were derived using a neural network discriminant used to discriminate the main background $Wb\bar{b}$ from the signal WH . Another discriminant has been developed to discriminate signal from the second dominant background, which is $t\bar{t}$. By combining both discriminants in another neural network, it has been shown that a significant improvement can be gained. Figure 6.28 shows the sensitivity gain compared to the limit obtained from the dijet invariant mass distribution can be on the order of 30%, especially at higher mass. This comes from the fact that $t\bar{t}$ is a more energetic process, therefore lying at higher values of H_T and M_{jj} , which is where a higher mass Higgs boson is expected. Hence, this increases the discrimination with the signal WH , the effect being visible above $M_H = 130$ GeV.

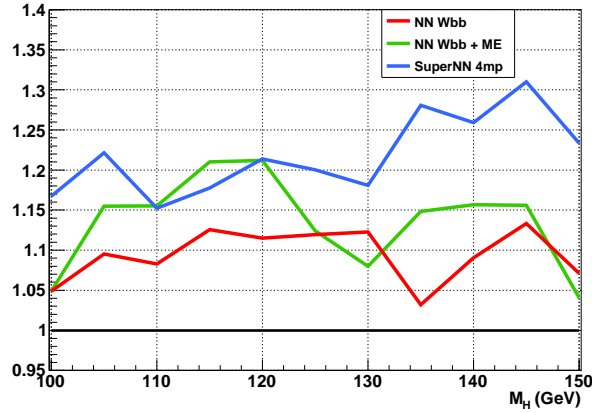


Figure 6.28: Sensitivity improvement factor compared to the limit obtained with the dijet invariant mass. The red curve is obtained with a neural network trained on $Wb\bar{b}$ and signal. The green curve is by adding the Matrix Elements discriminant as an input variable of the neural network, used in [128]. The blue curve is from the combination of the neural networks trained to discriminate $Wb\bar{b}$ and $t\bar{t}$ backgrounds from signal.

In order to use as many events as possible during the training of the neural network, and not removing events used for the training in the analysis, 4 Higgs mass points are used in the training¹. Since the $M_H = 115$ GeV signal events have not been used, there is no need to remove events at the analysis level, as explained in Section 5.1.1.

With this usage of additional signal samples in the training, they are normalized to the same cross section relatively to each other since lower mass Higgs bosons have higher cross sections, thus favoring the training to lighter Higgs bosons. Moreover, since 4 neighbouring mass points are used, only 2 or 3 signal samples are used for the 100, 105, 145 and 150 GeV mass points, which is less optimal. As depicted in Figure 6.29, this treatment brings about 2% improvement in the limit.

¹i.e. 105, 110, 120 and 125 GeV for the training at $M_H = 115$ GeV

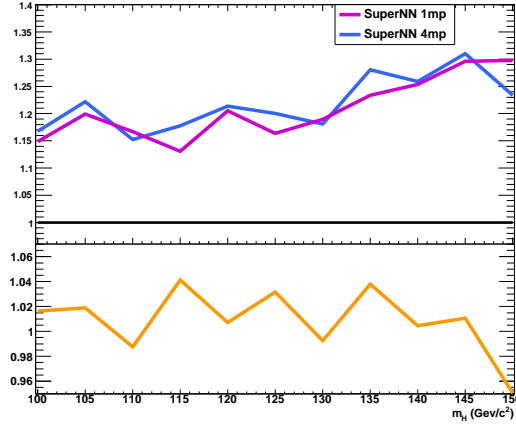


Figure 6.29: Sensitivity gain by using the combination of the $Wb\bar{b}$ and $t\bar{t}$ neural networks, using 1 (purple) and 4 signal mass points (blue) in the training. The bottom figure shows the improvement when using 4 mass point relative to 1 mass point, which is about 2% averaged over the 100–150 GeV mass range.

Preliminary studies suggest that up to 20% improvement can be achieved by using the TMVA package [129], as shown in Figure 6.30. A new algorithm, called Boosted Decision Trees with Stochastic Gradient Boost, is used. The boost weight for each tree is determined by minimizing a loss function in a similar way as performed for neural networks training (steepest-descent approach by computing the gradient of the loss function). More details can be found in the TMVA documentation [130] about this new training method. The average output of all trees is then taken as the final output. The main feature of this method is that the forest of decision trees can be large ($\mathcal{O}(100\text{-}1000)$) and each tree can contain a small number of nodes (between 5 and 10).

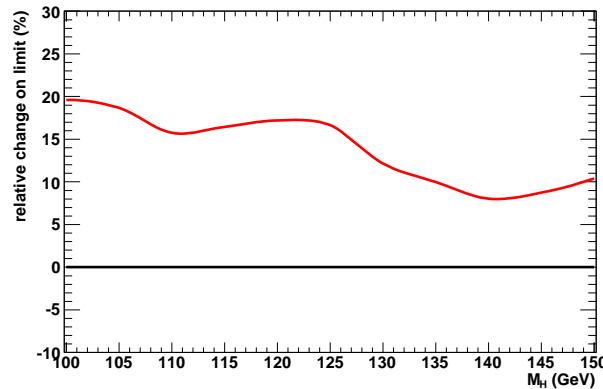


Figure 6.30: Sensitivity improvement using the Boosted Decision Trees method from the TMVA package, relative to the method employed in 5.1.

As emphasized earlier, the b -tagging discriminants can be fed in the final discriminant as an input variable in order to give information about the jet flavor and provide better discrimination. A comparison of limits obtained with 2 L6 b -tagged jets, with and without b -tagging discriminant used as an input to the final discriminant is shown in Figure 6.31. As expected, using L6 jets in the analysis leads to a decrease in sensitivity (up to 15% degradation in the limit compared to the result presented in Section 5.3), since more light flavor jets backgrounds is introduced. However, the addition of the MVA BL output for the 2 jets allows to improve the final sensitivity between 10 and 20% across the mass range, since the jet flavor information is added in order to discriminate signal from other backgrounds.

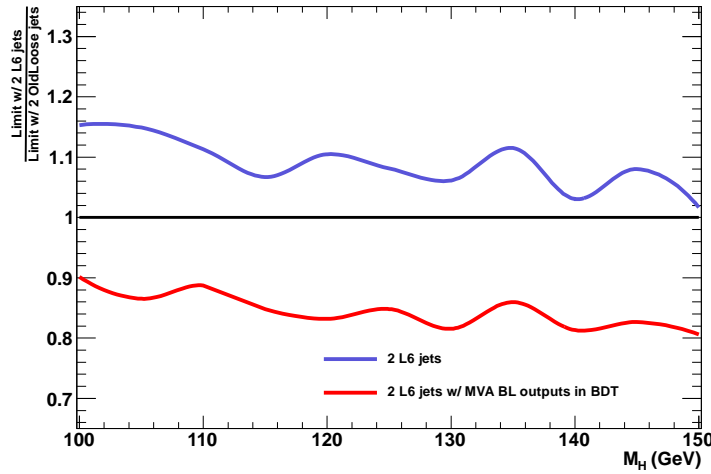


Figure 6.31: Limits obtained with 2 L6 b -tagged jets, with and without b -tagging discriminants used as an input to the MVA. Results are shown as the ratio to the limit obtained with 2 OldLoose b -tagged jets (see Section 4.9 and Section 5.3).

6.1.3 Summary of potential improvements

Ongoing works on dijet mass and jet resolution are developed in order to enhance the sensitivity to a Higgs boson decaying to a pair of jets. Several aspects are investigated and an expected improvement up to 10% is expected. In the WH analysis, several tools and techniques have been introduced to increase the signal acceptance (*i.e.* relaxed selection cuts and b -tagging) and reject as much background as possible (*i.e.* MVAQCD and final discriminant optimizations) on top of previous iterations of the WH analysis. An overall sensitivity improvement of 60% could be reached in the case that each expected improvement do not cancel with others.

6.2 Prospects at the Tevatron and LHC

As presented in Section 5, no excess consistent with the expected signal is observed in data, and the WH analysis is only sensitive to the production which would be about 5 times higher than the Standard Model expectation. In order to have the highest sensitivity, all search channels for the Standard Model are combined together with both CDF and DØ results.

The Tevatron is not anymore the most energetic collider since the Large Hadron Collider (LHC) at CERN [131] is fully operating since November 2009. Originally designed to collide protons at a center-of-mass energy of 14 TeV, it is currently running at 7 TeV.

6.2.1 Prospects at the Tevatron

Current status at the Tevatron

In order to improve the sensitivity for the Higgs boson production, all channels are combined together as a final step. The CL_S method, explained earlier in 5.2, is used. It was shown that several bins of a distribution can be used, as well as several channels for a given analysis. The idea behind the combination is identical, where this procedure is extended and repeated across several analyses, for both CDF and DØ.

The Tevatron Higgs combination incorporates all analyses dealing in total with 129 different analysis sub-channels from both experiments (56 for CDF and 73 for DØ) for the last combination performed in Summer 2010 for the 100–200 GeV mass range [132]. With analyses using up to 6.7 fb^{-1} of Tevatron data, an exclusion at the 95% C.L. for a Higgs boson with $158 < M_H < 175 \text{ GeV}$ and $100 < M_H < 109 \text{ GeV}$ is set, region already excluded at LEP in 2001. The expected and observed limits are shown in Figure 6.32, and the observed and expected LLR values in Figure 6.33.

Mass (GeV)	100	105	110	115	120	125	130	135	140	145	150
Expected	1.20	1.24	1.36	1.45	1.69	1.78	1.76	1.73	1.57	1.45	1.25
Observed	0.64	0.87	1.02	1.56	1.95	2.54	2.23	2.41	2.07	1.92	1.93
Mass (GeV)	155	160	165	170	175	180	185	190	195	200	
Expected	1.07	0.79	0.76	0.91	1.04	1.25	1.61	1.96	2.31	2.58	
Observed	1.28	0.85	0.68	0.79	0.95	1.49	2.55	2.44	3.49	3.87	

Table 6.4: Ratios of expected and observed 95% C.L. limit to the SM cross section for the combined CDF and DØ analyses as a function of the Higgs boson mass in GeV.

An update was made in Winter 2011 [133], only for the “High mass” analyses, where the main sensitivity is carried by $H \rightarrow W^+W^-$ channels. The $158 < M_H < 173 \text{ GeV}$ range is excluded at the 95% C.L.

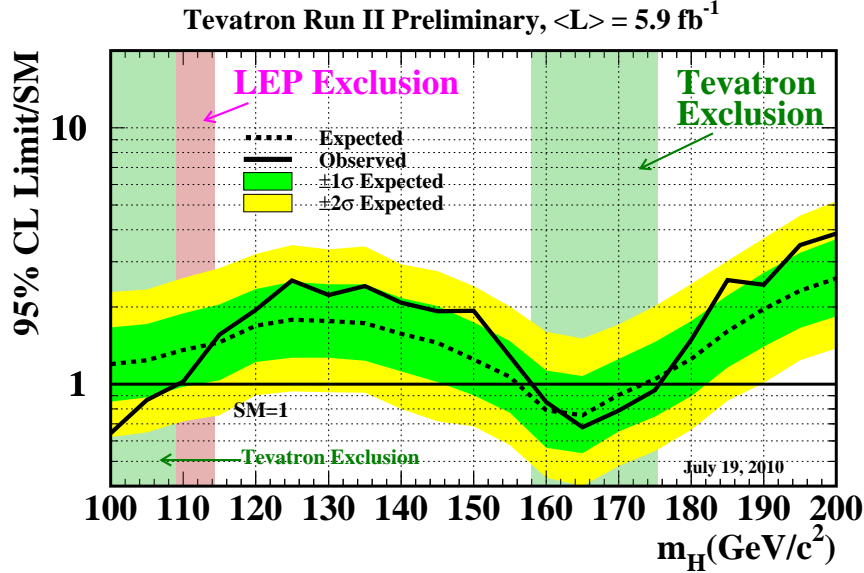


Figure 6.32: Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on the ratios to the SM cross section, as functions of the Higgs boson mass for the combined CDF and DØ analyses. The limits are expressed as a multiple of the SM prediction for test masses (every 5 GeV) for which both experiments have performed dedicated searches in different channels. The points are joined by straight lines for better readability. The bands indicate the 68% and 95% probability regions where the limits can fluctuate, in the absence of signal.

Prospects at the Tevatron

Although the decision to shut down the Tevatron has been taken and will be effective at the end of September 2011, thorough sensitivity studies have been conducted beforehand in order to evaluate the feasibility of a discovery or exclusion of the Higgs boson [134].

Detector aging has been modeled to take into account radiation damages from higher luminosities, particularly important for the b -tagging performance for “low mass” Higgs boson searches. As shown in Figure 6.34, it has been found to have a small impact (about 2% loss) on the sensitivity at the Tevatron. Given that about 11 fb^{-1} are expected to be analyzed after the Tevatron shutdown, one can see that almost the whole mass range could potentially be excluded at 95% C.L. In order to probe evidence of the Higgs boson, more luminosity is required for the region around 130 GeV and above 185 GeV.

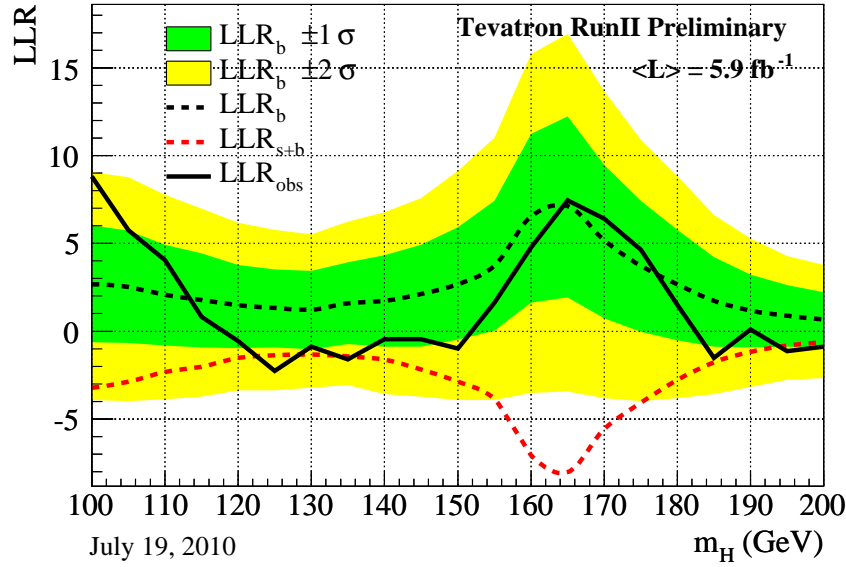


Figure 6.33: Distributions of the log-likelihood ratio (LLR) as a function of Higgs mass obtained with the CL_s method for the combination of all CDF and DØ analyses.

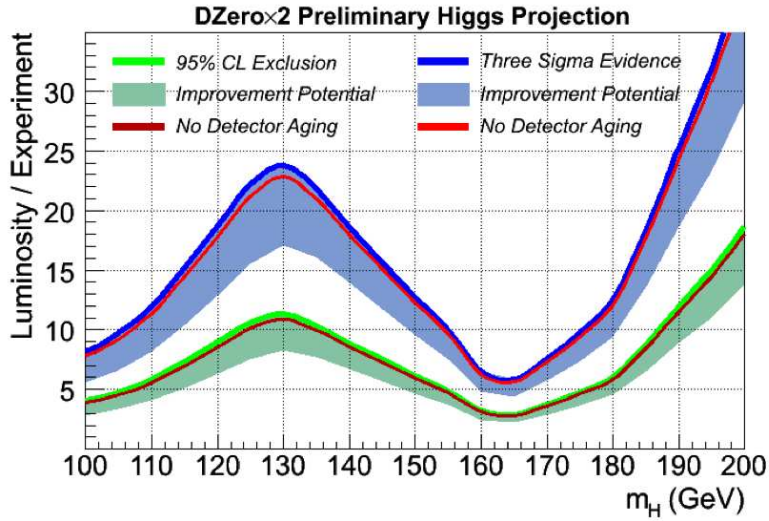


Figure 6.34: Expected luminosity per experiment to achieve a 95% C.L. exclusion (green line) or 3σ evidence (blue line), with detector aging taken into account. The shaded region represents the improvement potential, therefore decreasing the needed luminosity. Red curves represents the expected luminosities without including detector aging.

6.2.2 Prospects at the LHC

Although both CDF and DØ are able to push the Standard Model to its limits, experiments from the LHC are expected to provide very stringent limits with a relatively low luminosity, as they are

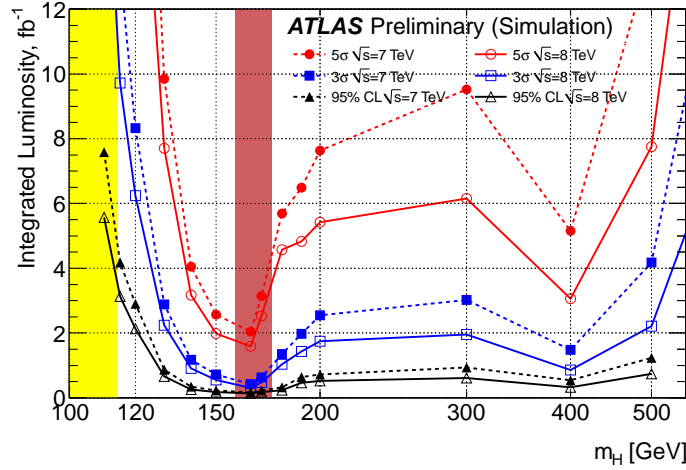


Figure 6.35: The luminosity required for the ATLAS experiment, as a function of M_H , to give 95% C.L. exclusion, evidence or discovery sensitivity for a SM Higgs at $\sqrt{s} = 7$ or 8 TeV. The shaded regions are the regions excluded by LEP, (yellow) and the Tevatron (brown).

starting to accumulate data. Studies based on simulation [135] show that even with a center of mass energy of 7 TeV, the exclusion or discovery potential can be reached with less than 2 fb^{-1} of analyzed data for $M_H \gtrsim 130 \text{ GeV}$ (see Figure 6.35). However, below this range of mass, more integrated luminosity will be needed and channels with $H \rightarrow \tau\tau$ or $H \rightarrow \gamma\gamma$ will play a major role to reach the SM sensitivity.

Figure 6.36 shows the limits evaluated from simulated data events corresponding to the combination of the most sensitive channel with 1 fb^{-1} .

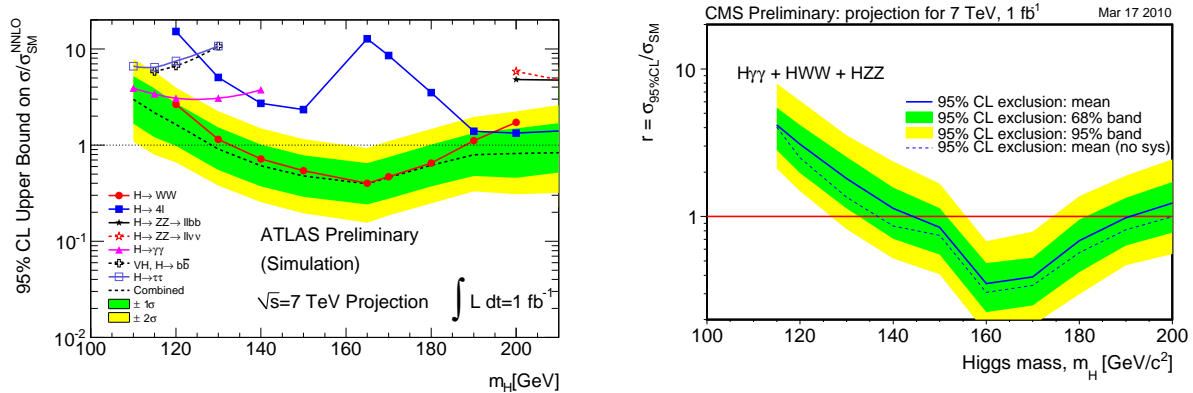


Figure 6.36: Ratio between expected cross sections limits and Standard Model cross sections for the combination of the most sensitive channel at ATLAS (left) and CMS (right), taken from [136] and [137].

As an example, limits obtained in the $H \rightarrow \gamma\gamma$ analysis from the ATLAS collaboration are

showed in 6.37 and the sensitivity reached is below 10 times the Standard Model prediction, almost across the whole mass range.

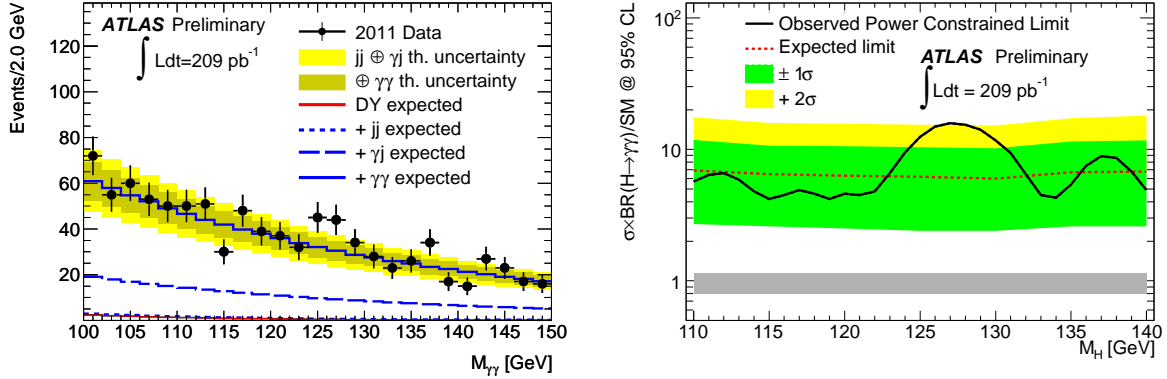


Figure 6.37: Left: Diphoton invariant mass distribution for data and the predictions of the Drell-Yan, dijet, photon-jet and diphoton components of the background. The two yellow bands depict the total uncertainty on the prediction and the uncertainty on the reducible background component only. Right: Exclusion limits on the production cross section relative to the Standard Model cross section as a function of the Higgs boson mass hypothesis.

6.2.3 Conclusion

During the last few years, the Tevatron has been able to set an exclusion on the Higgs boson for the first time since the LEP experiments shut down. Exclusion was achieved also by CDF and DØ alone in February 2011 [138] [139]. However, the Tevatron will shut down in September 2011 with an expected analyzable luminosity of about 11 fb^{-1} . With such amount of data, almost the entire mass range could be excluded at the 95% C.L., except maybe the region close to $M_H = 130 \text{ GeV}$.

In the meantime, the LHC started to collect data at a center-of-mass energy of 7 TeV. At this energy regime, the cross section for the Higgs boson production is much higher and with 1 fb^{-1} , the exclusion of the $M_H > 130 \text{ GeV}$ region can be reached. Channels that cannot be exploited at the Tevatron will have a significant impact on this search, such as channels where the Higgs boson is produced by vector boson fusion for low mass Higgs searches. However, the $H \rightarrow b\bar{b}$ mode will be difficult to exploit due to the boost of the dijet system. We are looking forward to the final Tevatron results at low mass, but in any case, the experience gained at the Tevatron for this particular channel (analysis techniques, objects identification) will be helpful to reach higher sensitivity for lower Higgs boson masses.

Conclusion

The search for the Higgs boson in the $WH \rightarrow \ell\nu b\bar{b}$ at the DØ experiment has been presented. This channel is one of the two most sensitive for a low mass Higgs boson search and one of its feature is that most of the DØ detector parts are used to reconstruct events.

Indeed, isolated electrons from the W boson decay and jets are reconstructed mainly from their characteristic signature in the calorimeter. Muons are detected in the muon spectrometer since they are minimum ionizing particle, thus only depositing a small fraction of their energy in the calorimeter. Finally, the neutrino from the W decay does not interact in the detector but its presence is inferred by an energy imbalance in the transverse plane, relying on the measurement of all other objects present in the calorimeter. In order to reconstruct objects interacting in the detector, a efficient tracking device is used, since jet, muon and electron identification also rely on detection of tracks produced from $p\bar{p}$ collisions.

Several powerful techniques are used in order to reach the highest sensitivity in the WH analysis, such as identification of jets coming from b quarks by looking at displaced vertices consistent with the presence of b hadrons or multivariate analysis techniques to increase the discrimination between backgrounds and the WH signal.

A crucial point in this search is to obtain a good description of the background processes involved in the analysis, knowing the present limitation of Monte Carlo generators and imperfect detector simulation. First, all identified objects from simulated events are corrected in order to account for reconstruction and identification differences in data and Monte Carlo. Additionally, a set of corrections is applied in order to obtain a good modeling of the dominant V +jets background. Based on the same reason, the instrumental background is inferred from a side-band region in data, using a precise method to accurately model distribution shapes and normalization.

Since a good agreement between data and simulation is obtained, and no signal excess is observed in data, upper observed (expected) limits (for $M_H = 115$ GeV) of 0.601 pb (0.642 pb) are set at 95% confidence level on the ratio of the WH cross section multiplied by the branching ratio of $H \rightarrow b\bar{b}$ to its Standard Model prediction, which represent 4.6 (4.9) times the Standard Model expectation. This procedure is performed using a semi-frequentist approach in order to efficiently take into account systematic uncertainties, which are found to degrade the sensitivity by about 20% once taken into account and after minimizing their impact by constraining them to data.

The presented result, based on an integrated luminosity of 5.3 fb^{-1} has been published in Physics Letter Review B in February 2011 [1] and is included in the Tevatron combination. Since the Tevatron is still operating and will accumulate data until September 2011, this analysis is not

CONCLUSION

the final result for this search at DØ . Based on the strategy documented here and good understanding of the data, updated results of this analysis will employ refined techniques established from this published result.

The recorded integrated luminosity is expected to be around 11 fb^{-1} when the data-taking will end. By combining results from all Higgs search analyses at the Tevatron, based on the whole dataset which will be analyzed, projections indicate that a 3σ evidence for the SM Higgs boson is achievable except in the mass range between 120 and 145 GeV. Since this region will be difficult to probe, every potential improvement is investigated in order to obtain the best sensitivity. Considering that the final state contains a $b\bar{b}$ resonant decay, a good resolution in the distribution obtained from the invariant mass of the two jets from the Higgs boson decay is a crucial aspect and this area is actively explored.

As data is accumulating and analysis techniques are refined, even more exciting results may be ahead of us. Hints about the validity of the Standard Model, as it is actually formulated, may be seen in a close future at the Tevatron. The LHC, which benefits from higher instantaneous luminosity and center-of-mass energy, may give a clear answer in a relative short time scale, using $H \rightarrow \gamma\gamma$, $H \rightarrow WW$ and $H \rightarrow ZZ$ channels. However, if the Higgs boson is observed or excluded at low mass, it will need to be validated in the $H \rightarrow b\bar{b}$ decay mode. Therefore, the experience gained at the Tevatron measurements will definitely bring crucial contribution to the understanding of the Higgs boson properties.

Bibliography

- [1] V.M. Abazov et al. Search for WH associated production in 5.3 fb^{-1} of $p\bar{p}$ collisions at the Fermilab Tevatron. *Phys. Lett.*, B698:6–13, 2011. [arXiv:1012.0874](#), [doi:10.1016/j.physletb.2011.02.036](#).
- [2] K. Nakamura et al. Review of particle physics. *J. Phys.*, G37:075021, 2010. [doi:10.1088/0954-3899/37/7A/075021](#).
- [3] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13(16):508–509, Oct 1964. [doi:10.1103/PhysRevLett.13.508](#).
- [4] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13(9):321–323, Aug 1964. [doi:10.1103/PhysRevLett.13.321](#).
- [5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13(20):585–587, Nov 1964. [doi:10.1103/PhysRevLett.13.585](#).
- [6] Benjamin W. Lee, C. Quigg, and H. B. Thacker. Weak interactions at very high energies: The role of the higgs-boson mass. *Phys. Rev. D*, 16(5):1519–1531, Sep 1977. [doi:10.1103/PhysRevD.16.1519](#).
- [7] A combination of preliminary electroweak measurements and constraints on the standard model, 2006. [oai:cds.cern.ch:1016509](#). Technical Report [hep-ex/0612034](#). ALEPH-2006-001 PHYSICS-2006-001. CERN-L3-310. CERN-PH-EP-2006-042. DELPHI-2006-014 PHYS-948. L3-Note-2833. LEPEWWG-2006-01. OPAL-PR-419, CERN, Geneva, Dec 2006. Preprint not submitted to publication.
- [8] M. Davier and H. Nguyen Ngoc. An unambiguous search for a light Higgs boson. *Phys.Lett.*, B229:150, 1989. [doi:10.1016/0370-2693\(89\)90174-3](#).
- [9] A. Snyder, W.N. Murray, G.S. Abrams, Chris Adolphsen, Carl Akerlof, et al. Search for B decay to Higgs bosons for Higgs boson masses between $50\text{-MeV}/c^2$ and $210\text{-MeV}/c^2$. *Phys.Lett.*, B229:169, 1989. [doi:10.1016/0370-2693\(89\)90177-9](#).
- [10] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys. Lett.*, B565:61–75, 2003. [arXiv:hep-ex/0306033](#), [doi:10.1016/S0370-2693\(03\)00614-2](#).
- [11] V.M. Abazov et al. Search for the standard model Higgs boson in the $ZH \rightarrow \nu\nu b\bar{b}$ channel in 5.2 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$. *Phys. Rev. Lett.*, 104:071801, 2010. [arXiv:0912.5285](#), [doi:10.1103/PhysRevLett.104.071801](#).

BIBLIOGRAPHY

- [12] V.M. Abazov et al. Search for $ZH \rightarrow \ell^+\ell^-b\bar{b}$ production in 4.2 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$. *Phys. Rev. Lett.*, 105:251801, 2010. arXiv:1008.3564, doi:10.1103/PhysRevLett.105.251801.
- [13] V.M. Abazov et al. Search for Higgs boson production in dilepton and missing energy final states with 5.4 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$. *Phys. Rev. Lett.*, 104:061804, 2010. arXiv:1001.4481, doi:10.1103/PhysRevLett.104.061804.
- [14] V.M. Abazov et al. Search for the Standard Model Higgs boson in $\mu + \tau_{had} + 1 \text{ jet}$ final state with 7.3 fb^{-1} of data. DØ Note 6179. Available from: <http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/HIGGS/H100/H100.pdf>.
- [15] V.M. Abazov et al. Search for the Standard Model Higgs Boson in the $H \rightarrow WW \rightarrow \ell\nu q'\bar{q}$ Decay Channel. *Phys. Rev. Lett.*, 106:171802, 2011. arXiv:1101.6079, doi:10.1103/PhysRevLett.106.171802.
- [16] V.M. Abazov et al. Search for the Standard Model Higgs Boson in $\gamma\gamma + X$ final states at DØ using 8.2 fb^{-1} data. DØ Note 6177. Available from: <http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/HIGGS/H99/H99.pdf>.
- [17] V.M. Abazov et al. Search for the Standard Model Higgs boson the $\tau^+\tau^- + 2 \text{ jets}$ final state. DØ Note 6172. Available from: <http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/HIGGS/H101/H101.pdf>.
- [18] V.M. Abazov et al. Search for associated Higgs boson production using like charge dilepton events in $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$. 2011. arXiv:1107.1268.
- [19] S. W. Herb et al. Observation of a dimuon resonance at 9.5 GeV in 400 GeV proton - nucleus collisions. *Phys. Rev. Lett.*, 39:252–255, 1977. doi:10.1103/PhysRevLett.39.252.
- [20] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys.Rev.Lett.*, 74:2626–2631, 1995. arXiv:hep-ex/9503002, doi:10.1103/PhysRevLett.74.2626.
- [21] S. Abachi et al. Observation of the top quark. *Phys.Rev.Lett.*, 74:2632–2637, 1995. arXiv:hep-ex/9503003, doi:10.1103/PhysRevLett.74.2632.
- [22] B. Baller et al. Direct observation of the tau neutrino. *Nucl.Phys.Proc.Suppl.*, 98:43–47, 2001. doi:10.1016/S0920-5632(01)01193-8.
- [23] C. M. Bhat and C. A. Rodriguez. Pbar Acceleration in the Main Injector for Run II: ESME Simulations. Beams-doc-354-v1, 2000.
- [24] Available from: <http://www-bd.fnal.gov/public/maininj.html>.
- [25] Gerald P. Jackson. Stochastic Cooling of the Recycler Antiproton Stack Momentum Spread Including Intrabeam Scattering and Barrier Buckets. Beams-doc-172-v1, 2002.
- [26] S. Abachi et al. The DØ Detector. *Nucl.Instrum.Meth.*, A338:185–253, 1994. doi:10.1016/0168-9002(94)91312-9.

- [27] V.M. Abazov et al. The Upgraded DØ detector. *Nucl.Instrum.Meth.*, A565:463–537, 2006. arXiv:physics/0507191, doi:10.1016/j.nima.2006.05.248.
- [28] R. Angstadt et al. The Layer 0 Inner Silicon Detector of the DØ Experiment. *Nucl. Inst. Meth. A*, arXiv:0911.2522v1, 2009.
- [29] Available from: http://pdg.lbl.gov/2011/AtomicNuclearProperties/HTML_PAGES/289.html.
- [30] Available from: http://pdg.lbl.gov/2011/AtomicNuclearProperties/HTML_PAGES/092.html.
- [31] C. W. Fabjan and F. Gianotti. Calorimetry for particle physics. *Rev. Mod. Phys.*, 75(CERN-EP-2003-075):1243–1286. 96 p, Oct 2003.
- [32] M. Voutilainen. Jet p_T resolution for RunIIa final JES (v7.2) with dijet J4S jet corrections. DØ Note 5499.
- [33] P. Gris. Electron Smearing Studies with RunIIa Data. DØ Note 5400.
- [34] T. Edwards and others. Fermilab-TM-2278-E, 2004. Available from: <http://lss.fnal.gov/archive/test-tm/2000/fermilab-tm-2278-e.pdf>.
- [35] Available from: <http://www-d0.fnal.gov/Run2Physics/cs/caf>.
- [36] Alexander Khanov. Htf: histogramming method for finding tracks. the algorithm description. DØ Note 3778.
- [37] G. Borissov. Ordering a Chaos or... Technical Details of AA Tracking. Presentation at the *All DØ Meeting*, 28 February 2003.
- [38] H. Greenlee. The DØ Kalman Track Fit. DØ Note 4303.
- [39] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987. doi:10.1016/0168-9002(87)90887-4.
- [40] K. Bloom and A. Dominguez. DØ Tracking Performance at High Luminosity. DØ Note 4980.
- [41] D. Bandurin et al. Electron and Photon Identification with p20 data. DØ Note 5761, 2008.
- [42] Masato Aoki. Electron Likelihood in p20. DØ Note 5675.
- [43] J. Hays, J. Mitrevski, C. Schwanenberger, and T. Toole. Single Electron Efficiencies in p17 Data and Monte-Carlo Using d0correct from Release p18.05.00. DØ Note 5105.
- [44] P. Calfayan, T. Gadford, G. Hesketh, V. Lesne, M. Owen, R. Stroehmer, V. Sharyy, and B. Tuchming. Muon identification certification for p17 data. DØ Note 5157.

BIBLIOGRAPHY

- [45] O. Brandt et al. Muon Identification Certification for the Summer 2009 Extended Dataset (Run IIb-1 and -2). DØ Note 6025.
- [46] M. Arthaud, F. Deliot, B. Tuchming, V. Sharyy, and D. Vilanova. Muon Momentum Over-smearing for p17 Data. DØ Note 5444.
- [47] M. Arthaud, F. Deliot, B. Tuchming, V. Sharyy, and D. Vilanova. Muon Momentum Over-smearing for p20 Data. DØ Note 5449.
- [48] E. Busato and B. Andrieu. Jet Algorithms in DØ RunII Software: Description and User's Guide. DØ Note 4457.
- [49] G. Bernardi, B. Olivier, B. Knuteson, and M. Strovink. NADA: A New Event by Event Hot Cell Killer. DØ Note 3687.
- [50] G. Bernardi and S. Trincaz-Duvold. Improvement of the NADA Algorithm: Hot Cell Killing in DØ Run II Data. DØ Note 4057.
- [51] U. Bassler and G. Bernardi. Towards a coherent treatment of calorimeter energies. DØ Note 4124, 2002.
- [52] J.R. Vlimant et al. Technical description of the t42 algorithm for the calorimeter noise suppression. DØ Note 4146, 2003.
- [53] G. Bernardi, E. Busato, and J.R. Vlimant. Improvements from the t42 algorithm on calorimeter objects reconstruction. DØ Note 4335, 2004.
- [54] Gerald C. Blazey et al. Run II jet physics. 2000. [arXiv:hep-ex/0005012](https://arxiv.org/abs/hep-ex/0005012).
- [55] Bernard Andrieu and Amnon Harel. Jet Efficiency Measurement and Monte Carlo Correction - Lessons and Options. DØ Note 5169.
- [56] Amnon Harel and Robert Wagner. Improved L1 Confirmation. DØ Note 4932.
- [57] Plot From Murilo Rangel.
- [58] Available from: http://www-d0.fnal.gov/phys_id/jes/public_RunIIa/.
- [59] B. Abbott et al. Determination of the absolute jet energy scale in the DØ calorimeters. *Nucl. Instrum. Meth.*, A424:352–394, 1999. [arXiv:hep-ex/9805009](https://arxiv.org/abs/hep-ex/9805009), [doi:10.1016/S0168-9002\(98\)01368-0](https://doi.org/10.1016/S0168-9002(98)01368-0).
- [60] Dag Gillberg and Dugan O'Neil and Yann Coadou and Brendan Pass and Herve Choi and Michel Vetterli. Measuring Jet Response Using the Missing E_T Projection Fraction Method in γ + jet Events. DØ Note 4571.
- [61] Jérémie Lellouch. Corrections for muonic b -jets in P17. DØ Note 5486.
- [62] N. Makovec and J.F. Grivaz. The relative data-mc jet energy scale. DØ Note 4807, 2005.

-
- [63] N. Makovec and J.F. Grivaz. Shifting, smearing and removing simulated jets. DØ Note 4914, 2005.
- [64] C. Ochando and J.F. Grivaz. Ssr for p17. DØ Note 5609, 2008.
- [65] Christophe Ochando. Slides from the JES meeting on 2008/5/15. Available from: http://www-d0.hef.ru.nl/askArchive.php?base=agenda&categ=a08924&id=a08924s1t3/transparencies/ochando_jes_15may08_updated.pdf.
- [66] Samuel Calvet, Patrice Verdier, and Eric Kajfasz. Towards MissingET Certification and Unclustered Energy Studies. DØ Note 4927.
- [67] V.M. Abazov et al. *Nucl. Inst. Meth. Phys. Rev.*, A620, 2010.
- [68] Alexander Khanov. Measurement of the $t\bar{t}$ production cross-section at $\sqrt{s} = 1.96$ TeV using lifetime tagging. FERMILAB-THESIS-2004-15.
- [69] Ariel Schwartzman and Meenakshi Narain. b -quark jet identification via Secondary Vertex Reconstruction. DØ Note 4080.
- [70] L. Feligioni, M. Narain, P. Schieferdecker, and A. Schwartzman. Update on b -quark jet identification with Secondary Vertex Reconstruction using Dreco version p14. DØ Note 4414.
- [71] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003. [arXiv:hep-ph/0206293](https://arxiv.org/abs/hep-ph/0206293).
- [72] C. Clément et al. Measurement of the $t\bar{t}$ cross section using Lifetime Tagging. DØ Note 4900, 2005.
- [73] Gustavo J. Otero y Garzon, Cecilia E. Gerber, and Elizaveta Shabalina. Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV using Secondary Vertex b-Tagging. 2006. DØ Note 5113.
- [74] D.K. Cho, U. Heintz, M. Narain, S. Park, R. Patridge, M. Pleier, S. Sumowidagdo, M. Weber, and H. Yoo. Measurement of the $t\bar{t}$ pair production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV in the lepton+jets final state using lifetime tagging on 900 pb⁻¹ of DØ data. DØ Note 5355, 2007.
- [75] Tim Scanlon. A neural network b-tagging tool. 2005. DØ Note 4889.
- [76] C. G. Broyden. The Convergence of a Class of Double-rank Minimization Algorithms. *Journal of the Institute of Mathematics and Its Applications*, 6, 1970.
- [77] R. Fletcher. A New Approach to Variable Metric Algorithms. *Computer Journal*, 13, 1970.
- [78] D. Goldfarb. A Family of Variable Metric Updates Derived by Variational Means. *Mathematics of Computation*, 24, 1970.

BIBLIOGRAPHY

- [79] D. F. Shanno. Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24, 1970.
- [80] V.M. Abazov et al. *Phys. Rev. Lett.*, 091802, 2005.
- [81] V.M. Abazov et al. *Phys. Rev. Lett.*, 051803, 2009.
- [82] M. Abolins et al. Design and Implementation of the New D Level-1 Calorimeter Trigger. *Nucl. Instrum. and Methods*, A 584, 2008.
- [83] A.Djouadi, J.Kalinowski, and M.Spira. HDECAY: a Program for Higgs Boson Decays in the Standard Model and its Supersymmetric Extension. *Comp. Phys. Commun.*, 108 C, 1998. hep-ph/9704448.
- [84] T. Sjöstrand et al. PYTHIA 6.3: Physics and manual. e-Print Archive: hep-ph/0308153, LU-TP-03-38, Lund, 2003.
- [85] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Phys. Rep.*, 97:31–145, July 1983. doi:10.1016/0370-1573(83)90080-7.
- [86] H.L. Lai et al. Improved parton distributions from global analysis of recent deep inelastic scattering and inclusive jet data. *Phys.Rev.*, D55, 1997. Available from: <http://hep.pa.msu.edu/people/wkt/cteq6/cteq6pdf.html>.
- [87] J. Pumplin et al. New generation of parton distributions with uncertainties from global qcd analysis. *JHEP07*, 012, 2002. arXiv:hep-ph/0201195v3.
- [88] Johan Alwall, Stefan Hoche, F. Krauss, N. Lavesson, L. Lonnblad, et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur.Phys.J.*, C53:473–500, 2008. arXiv:0706.2569, doi:10.1140/epjc/s10052-007-0490-5.
- [89] A. Pukhov et al. CompHEP, a package for evaluation of feynman diagrams and integration over multi-particle phase space. INP-MSU-98-41/542, Moscow, 1999.
- [90] T. Hahn et al. SM and MSSM Higgs Boson Production Cross Sections at the Tevatron and the LHC. arXiv:hep/ph:060308v2.
- [91] M. Begel, L. Christofek, and E. Shabalina. Determination of weighting factors for ALPGEN Monte Carlo signal and background samples. DØ Note 5016.
- [92] Y. Fisyak and J. Womersley. DØ Note 3191.
- [93] M. Goossens et al. GEANT, detector description and simulation tool, geant user’s guide. CERN, Geneva, 1994.
- [94] Top trigger efficiency measurements and the “top_trigger” package. DØ Note 4512.
- [95] Text and trigger tables for publications for the top group. DØ Note 4978, 2006.

- [96] Trigger efficiencies for the OR of single electron triggers in p17 data. DØ Note 5409, 2007.
- [97] `caf_trigger` package. http://www.phys.ualberta.ca/~kwchan/d0/caf_trigger/readme.html.
- [98] G. Bernardi et al. Improved limits on WH production using a neural net algorithm with 1 fb^{-1} of RunIIa data. DØ Note 5713.
- [99] Romain Madar and Fabrice Couderc. Tau lepton identification using MVA optimizations, preshower and bID tools. DØ Note 6061.
- [100] Daniel Boline. Jet Multiplicity Dependent Reweighting of the Z Boson pT in p20. DØ Note 5569.
- [101] Mansoor Shamim and Tim Bolton. Generator Level Reweighting of pT of Z Boson. DØ Note 5565.
- [102] G. Hesketh. W pT Re-weighting for Alpgen and Pythia. DØ Note 5786.
- [103] V. M. Abazov et al. A Search for associated W and Higgs Boson production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\text{-TeV}$. *Phys. Rev. Lett.*, 102:051803, 2009. [arXiv:0808.1970](https://arxiv.org/abs/0808.1970), doi:10.1103/PhysRevLett.102.051803.
- [104] Thomas Gadfort, Andy Haas, Dale Johnston, Don Lincoln, Tim Scanlon, and Sarah Schlobohm. Performance of the DØ NN b-tagging Tool on p20 Data. DØ Note 5554.
- [105] Slides from Y. Enari in the bID meeting on 2009/9/30. Available from: <http://www-d0.hef.kun.nl///fullAgenda.php?id=a091545#2009-09-30>.
- [106] G. Bernardi, D. Brown, J. Brown, M. Cooke, Y. Enari, S. Greder, H. Hegab, K. Herner, N. Huske, J. Lellouch, J. Qian, and C. Xu. Search for WH associated production with 5.3 fb^{-1} of RunII data. DØ Note 6168.
- [107] Morris Degroot. *Probability and Statistics*. Addison-Wesley, Boston, 2002.
- [108] Alexander Mood. *Introduction to the theory of statistics*. McGraw-Hill, New York, 1973.
- [109] D.K. Neal. Goodness of Fit Tests for Normality. *Mathematica Educ. Res*, 5, 1996.
- [110] I. Narsky. StatPatternRecognition: A C++ Package for Statistical Analysis of High Energy Physics Data. physics/0507143, 2005.
- [111] I. Narsky. Optimization of Signal Significance by Bagging Decision Trees. physics/0507157v1, 2005.
- [112] J. Sekaric W. Fisher, J. Haley. First evidence for diboson production in lepton plus jets decays. DØ Note 5544, 2008.
- [113] DØ Collaboration, V.M. Abazov and others. *Phys. Rev. Lett.*, 161801, 2009.
- [114] S. Parke and S. Veseli. *Phys. Rev. D*, 093003, 1999.

BIBLIOGRAPHY

- [115] V. D. Barger and R. J. N. Phillips. *Collider Physics*. Addison-Wesley Publishing Co., 1987.
- [116] Wade Fisher. Collie: A Confidence Level Limit Evaluator. DØ Note 5595.
- [117] A. Read. Presentation of search results: The CL(s) technique. *J. Phys.*, G 28, 2002.
- [118] T. Junk. *Nucl. Instrum. Meth.*, A434, 1999.
- [119] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys.Lett.*, B565:61–75, 2003. [arXiv:hep-ex/0306033](https://arxiv.org/abs/hep-ex/0306033), doi:10.1016/S0370-2693(03)00614-2.
- [120] Wade Fisher. Systematics and Limit Calculations. CollieChi2Formulism.pdf from the Collie Package.
- [121] T. Gadfort et al. Muonid certification for p17 data. DØ Note 5157.
- [122] D. Price W. Fisher, J. Haley. Studies of alpgen parameters, corrections and associated uncertainties. DØ Note 5966, 2009.
- [123] Systematic Comparison: Pretagged event ratios vs. Tagged event ratios. Available from: http://www-clued0.fnal.gov/~mpc/d0_private/wh/pretagSystematics/.
- [124] Murilo Rangel, Brendan Casey, and Aurelio Juste. Improving the Jet Energy Resolution with the CPS Detector. DØ Note 5767.
- [125] Emanuel Strauss, Paul Grannis, and John Hobbs. Jet Resolution Improvements Using an H-Matrix. DØ Note 5785.
- [126] Jonathan Brown and Murilo Rangel. Attempt to improve jet resolution by exploiting the jet width information. DØ Note 6214.
- [127] JES Group. Jet Four-vector Scale Determination for Dijets in D0 Run IIa (final p17 version). DØ Note 5550.
- [128] D Collaboration. Search for WH associated production using a combined Neural Network and Matrix Element approach with 2.7 fb^{-1} of Run II data. DØ Note 5828. Available from: <http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/HIGGS/H65/H65.pdf>.
- [129] P. Speckmayer, A. Höcker, J Stelzer, and H Voss. The toolkit for multivariate data analysis, TMVA 4. *Journal of Physics: Conference Series*, 219(3):032057, 2010. Available from: <http://stacks.iop.org/1742-6596/219/i=3/a=032057>.
- [130] A. Höcker, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss, A. Christov, S. Henrot-Versillé, M. Jachowski, A. Krasznahorkay, Y. Mahalalel, X. Prudent, and P. Speckmayer. TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide. Technical Report CERN-OPEN-2007-007, CERN, Geneva, Nov 2009.
- [131] Available from: <http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/>.

- [132] Combined CDF and D Upper Limits on Standard Model Higgs Boson Production with up to 6.7 fb^{-1} of Data. arXiv:1007.4587, 2010. arXiv:1007.4587.
- [133] Combined CDF and D Upper Limits on Standard Model Higgs Boson Production with up to 8.2 fb^{-1} of Data. arXiv:1103.3233, 2011. arXiv:1103.3233.
- [134] Physics Advisory Committee, presented by Dmitri Denisov, August 27, 2010. Available from: http://www.fnal.gov/directorate/program_planning/Aug2010PACPublic/Denisov_PACAug2010_HiggsQuestions.pdf.
- [135] ATLAS Collaboration. Further investigations of atlas sensitivity to higgs boson production in different assumed lhc scenarios. Technical Report ATL-PHYS-PUB-2011-001, CERN, Geneva, Jan 2011.
- [136] ATLAS Collaboration. ATLAS Sensitivity Prospects for 1 Higgs Boson Production at the LHC Running at 7, 8 or 9 TeV. Technical Report ATL-PHYS-PUB-2010-015, CERN, Geneva, Nov 2010.
- [137] The CMS physics reach for searches at 7 TeV. Technical Report CMS-NOTE-2010-008. CERN-CMS-NOTE-2010-008, CERN, Geneva, May 2010.
- [138] Search for $H \rightarrow WW$ Production at CDF Using 7.1 fb^{-1} of Data. CDF Note 10432. Available from: http://www-cdf.fnal.gov/physics/new/hdg/Results_files/results/hwwmenn_110304/cdf10432_HWW_71fb_public_note.pdf.
- [139] Combined Upper Limits on Standard Model Higgs Boson Production in the W^+W , $\tau\tau$ and $\gamma\gamma$ decay modes in up to 8.2 fb^{-1} of data from the DØ Experiment. DØ Note 6183. Available from: <http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/HIGGS/H105/H105.pdf>.

A

Summary of vetoes

As mentioned in Section 4.4.6, additional selection criteria are used to ensure orthogonality with other low mass Higgs boson search channels and reduce backgrounds containing a second lepton in the final state. Tables A.1–A.4 summarize the effect of vetoes in RunIIb data.

	Pre-tag			1 b -tagged jet			2 b -tagged jets		
	no veto yield	τ veto / no veto	$\tau + e/CR$ / no veto	no veto yield	τ veto / no veto	$\tau + e/CR$ / no veto	no veto yield	τ veto / no veto	$\tau + e/CR$ / no veto
data	38384.00	0.99	0.98	2044.00	0.98	0.98	336.00	0.97	0.97
all BG	38384.10	0.99	0.98	2024.26	0.98	0.98	354.97	0.96	0.96
WH	10.76	0.98	0.98	4.26	0.98	0.98	2.82	0.99	0.98
ZH	0.54	0.69	0.63	0.21	0.67	0.62	0.12	0.67	0.58
Vjj	24088.50	0.98	0.98	533.44	0.99	0.98	22.72	0.99	0.99
QCD	8935.43	0.99	0.99	469.27	0.99	0.99	55.58	0.99	0.99
$Vc\bar{c}$	2507.44	0.98	0.98	246.45	0.99	0.99	22.86	0.98	0.98
$t\bar{t}$	541.47	0.92	0.92	215.22	0.92	0.92	94.85	0.89	0.88
$Vb\bar{b}$	1192.87	0.98	0.98	417.16	0.99	0.98	124.54	0.99	0.99
stop	197.27	1.00	1.00	83.34	1.00	1.00	25.12	1.00	1.00
WW	783.42	0.99	0.99	44.45	0.99	0.99	1.63	0.98	0.98
WZ	129.73	0.97	0.97	13.94	0.98	0.98	7.35	0.99	0.99
WZ	7.90	0.71	0.67	1.00	0.70	0.66	0.32	0.78	0.72

Table A.1: RunIIb electron 2jet Event yields with no veto, yield ratio of τ veto to no veto, and yield ratio of τ veto + eICR event veto to no veto, for the pre-tag, single b-tag and double b-tag cases.

APPENDIX A. SUMMARY OF VETOES

	Pre-tag			1 b -tagged jet			2 b -tagged jets		
	no veto yield	τ veto / no veto	$\tau + e/CR$ / no veto	no veto yield	τ veto / no veto	$\tau + e/CR$ / no veto	no veto yield	τ veto / no veto	$\tau + e/CR$ / no veto
data	7737.00	0.98	0.98	685.00	0.98	0.98	132.00	0.95	0.95
all BG	7737.01	0.98	0.98	724.21	0.98	0.98	137.31	0.97	0.97
WH	2.22	0.97	0.96	0.89	0.97	0.97	0.38	0.97	0.97
ZH	0.21	0.67	0.62	0.08	0.75	0.62	0.02	1.00	1.00
Vjj	3619.72	0.97	0.97	83.97	0.98	0.99	4.72	0.99	0.99
QCD	2296.40	1.01	1.00	185.30	1.01	1.00	27.83	0.99	0.99
$Vc\bar{c}$	612.26	0.97	0.97	52.77	0.98	0.97	3.59	0.97	0.97
$t\bar{t}$	703.25	0.97	0.97	282.91	0.97	0.97	79.52	0.96	0.96
$Vb\bar{b}$	283.40	0.97	0.97	85.80	0.97	0.97	15.26	0.98	0.98
stop	53.61	1.00	0.99	21.43	1.00	1.00	5.58	1.00	1.00
WW	138.97	0.99	0.99	8.47	1.00	1.00	0.28	1.00	1.00
WZ	26.61	0.95	0.95	3.20	0.97	0.97	0.51	1.00	1.00
ZZ	2.80	0.70	0.65	0.35	0.77	0.74	0.03	1.00	1.00

Table A.2: RunIIb electron 3jet channel: Event yields with no veto, yield ratio of τ veto to no veto, and yield ratio of τ veto + eICR event veto to no veto, for the pre-tag, single b-tag and double b-tag cases.

	Pre-tag			1 b -tagged jet			2 b -tagged jets		
	no veto yield	τ veto / no veto	$\tau + \mu TRK$ / no veto	no veto yield	τ veto / no veto	$\tau + \mu TRK$ / no veto	no veto yield	τ veto / no veto	$\tau + \mu TRK$ / no veto
data	28289.00	0.96	0.94	1510.00	0.96	0.94	266.00	0.94	0.92
all BG	28289.00	0.96	0.94	1566.25	0.96	0.94	289.39	0.94	0.91
WH	8.57	1.00	0.95	3.42	1.00	0.95	2.37	1.00	0.95
ZH	0.83	0.98	0.52	0.34	0.97	0.50	0.21	1.00	0.48
Vjj	20888.70	0.97	0.94	446.44	0.96	0.95	29.13	0.97	0.95
QCD	2848.32	0.93	0.93	258.29	0.94	0.94	32.52	0.91	0.90
$Vc\bar{c}$	2204.97	0.97	0.94	216.69	0.97	0.94	21.64	0.96	0.94
$t\bar{t}$	375.34	0.91	0.89	151.00	0.91	0.88	68.61	0.87	0.83
$Vb\bar{b}$	1034.96	0.97	0.94	366.65	0.97	0.94	108.53	0.97	0.94
stop	165.82	1.00	1.00	71.23	1.00	1.00	20.68	1.00	1.00
WW	634.29	0.99	0.99	40.59	0.98	0.98	1.18	1.00	1.00
WZ	119.59	0.99	0.92	13.69	0.99	0.96	6.14	1.00	0.99
ZZ	17.00	0.97	0.52	1.68	0.99	0.52	0.95	0.97	0.46

Table A.3: RunIIb muon 2jet channel: Event yields with no veto, yield ratio of τ veto to no veto, and yield ratio of τ veto + μTRK event veto to no veto, for the pre-tag, single b-tag and double b-tag cases.

	Pre-tag			1 b -tagged jet			2 b -tagged jets		
	no veto yield	τ veto / no veto	$\tau + \mu TRK$ / no veto	no veto yield	τ veto / no veto	$\tau + \mu TRK$ / no veto	no veto yield	τ veto / no veto	$\tau + \mu TRK$ / no veto
data	5318.00	0.96	0.93	511.00	0.95	0.93	124.00	0.96	0.95
all BG	5318.00	0.96	0.93	559.07	0.95	0.94	114.06	0.94	0.94
WH	1.81	0.99	0.94	0.77	0.99	0.92	0.31	1.00	0.94
ZH	0.23	0.96	0.52	0.10	0.90	0.50	0.03	1.00	0.67
Vjj	3149.13	0.96	0.93	70.75	0.96	0.94	3.46	0.98	0.98
QCD	658.18	0.89	0.89	114.86	0.89	0.90	24.69	0.88	0.89
$Vc\bar{c}$	530.91	0.96	0.94	44.43	0.97	0.94	4.73	0.97	0.91
$t\bar{t}$	542.07	0.97	0.96	221.81	0.97	0.96	63.53	0.95	0.95
$Vb\bar{b}$	242.87	0.97	0.93	77.15	0.97	0.93	11.98	0.97	0.93
stop	46.92	0.99	0.99	19.11	1.00	0.99	4.80	1.00	1.00
WW	120.75	0.99	0.99	7.58	0.98	0.98	0.20	1.00	1.00
WZ	23.25	0.98	0.92	2.83	0.99	0.94	0.50	1.00	1.00
ZZ	3.93	0.96	0.58	0.56	0.98	0.66	0.16	1.00	0.62

Table A.4: RunIIb muon 3jet channel : Event yields with no veto, yield ratio of τ veto to no veto, and yield ratio of τ veto + μTRK event veto to no veto, for the pre-tag, single b-tag and double b-tag cases.

B

Electron fake rate parameterizations

APPENDIX B. ELECTRON FAKE RATE PARAMETERIZATIONS

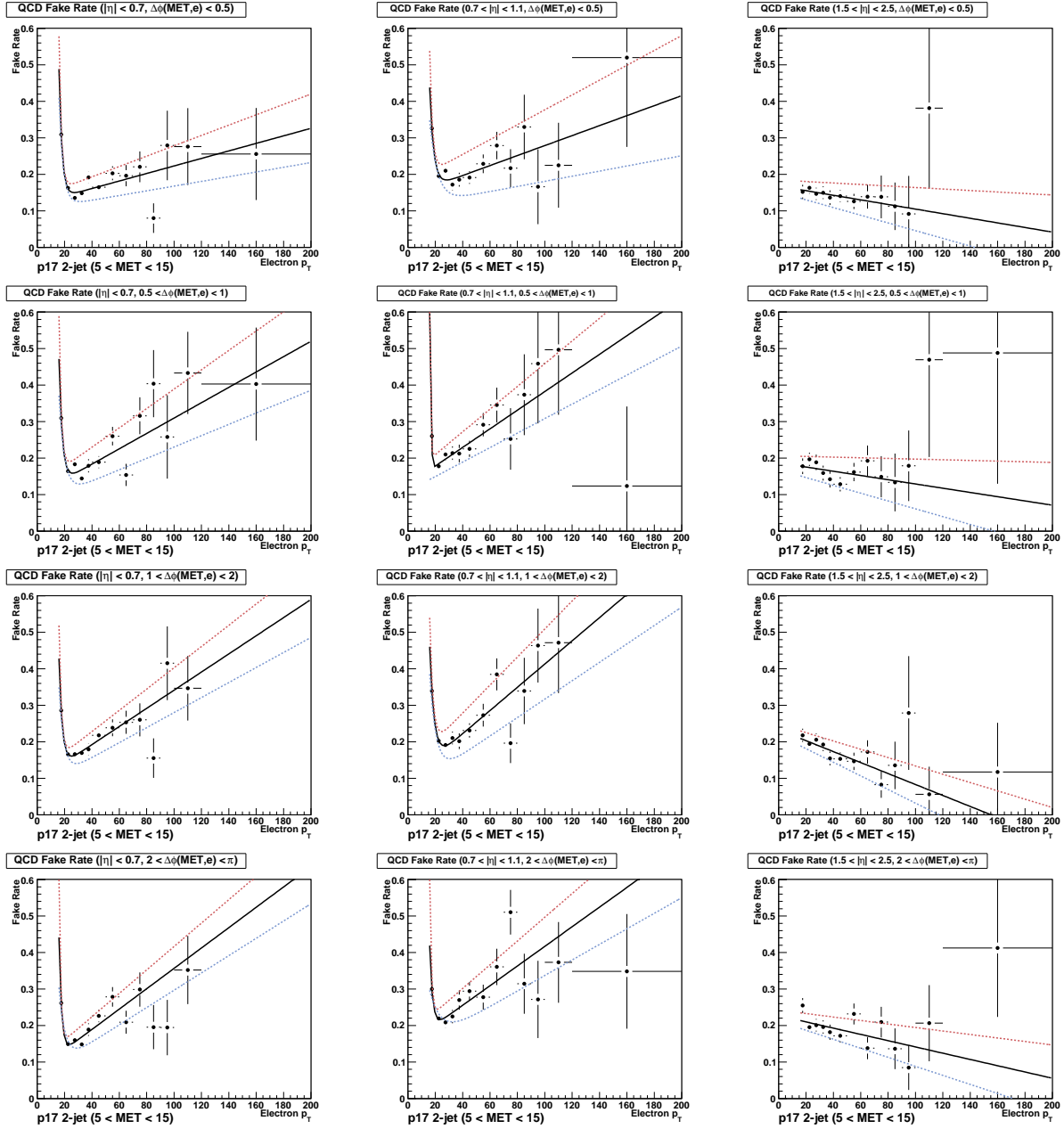


Figure B.1: The QCD fake rate for electrons in RunIIa 2-jet events as a function of $e p_T$ for each bin in $|\eta_{\text{CAL}}|$ vs. $\Delta\phi(\ell_T, e)$. The red dashed curve represents the $+1\sigma$ variation of the fit and the blue curve represents the -1σ variation.

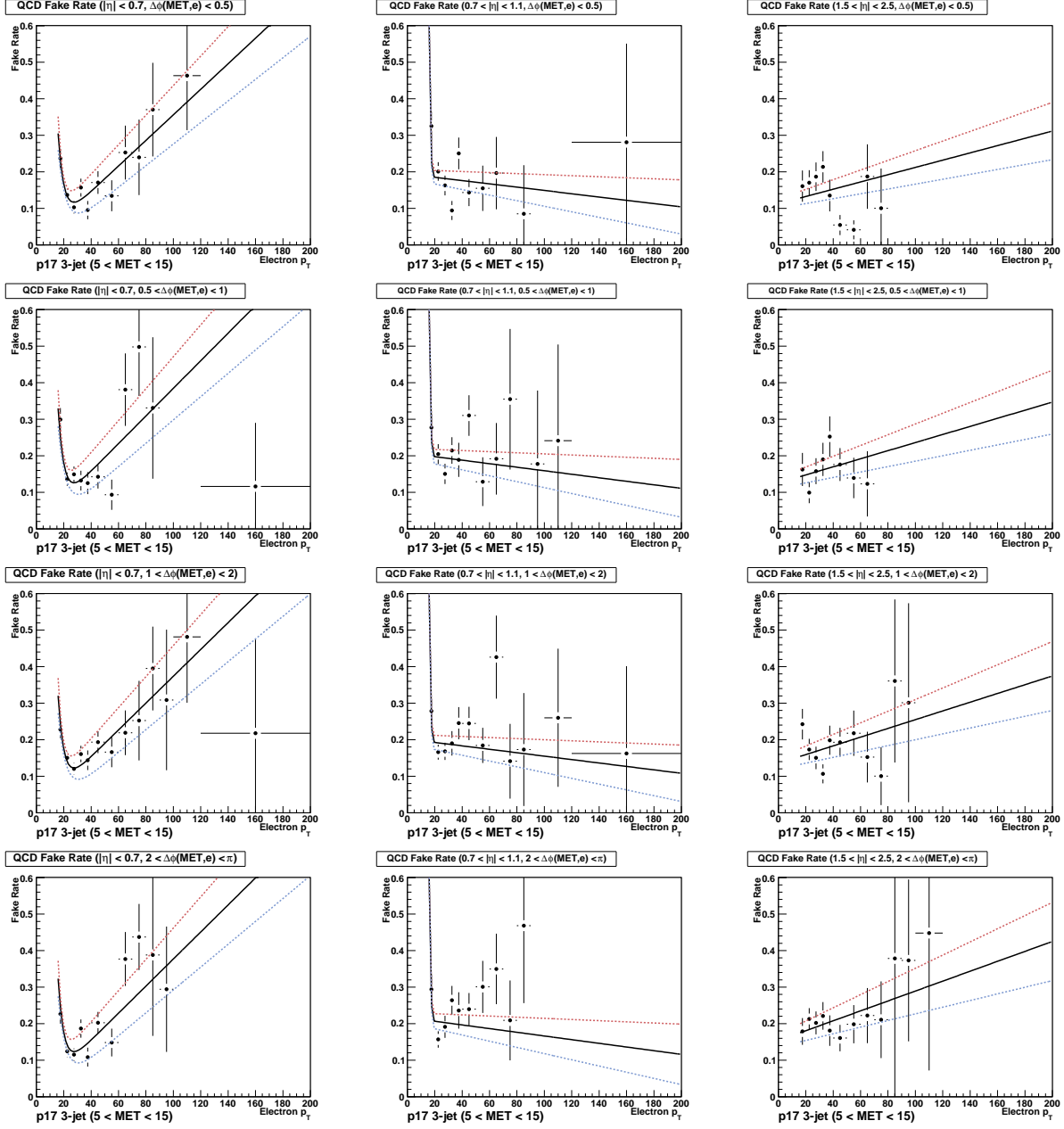


Figure B.2: The QCD fake rate for electrons in RunIIa 3-jet events as a function of $e p_T$ for each bin in $|\eta_{\text{CAL}}|$ vs. $\Delta\phi(\ell_T, e)$. The red dashed curve represents the $+1\sigma$ variation of the fit and the blue curve represents the -1σ variation.

APPENDIX B. ELECTRON FAKE RATE PARAMETERIZATIONS

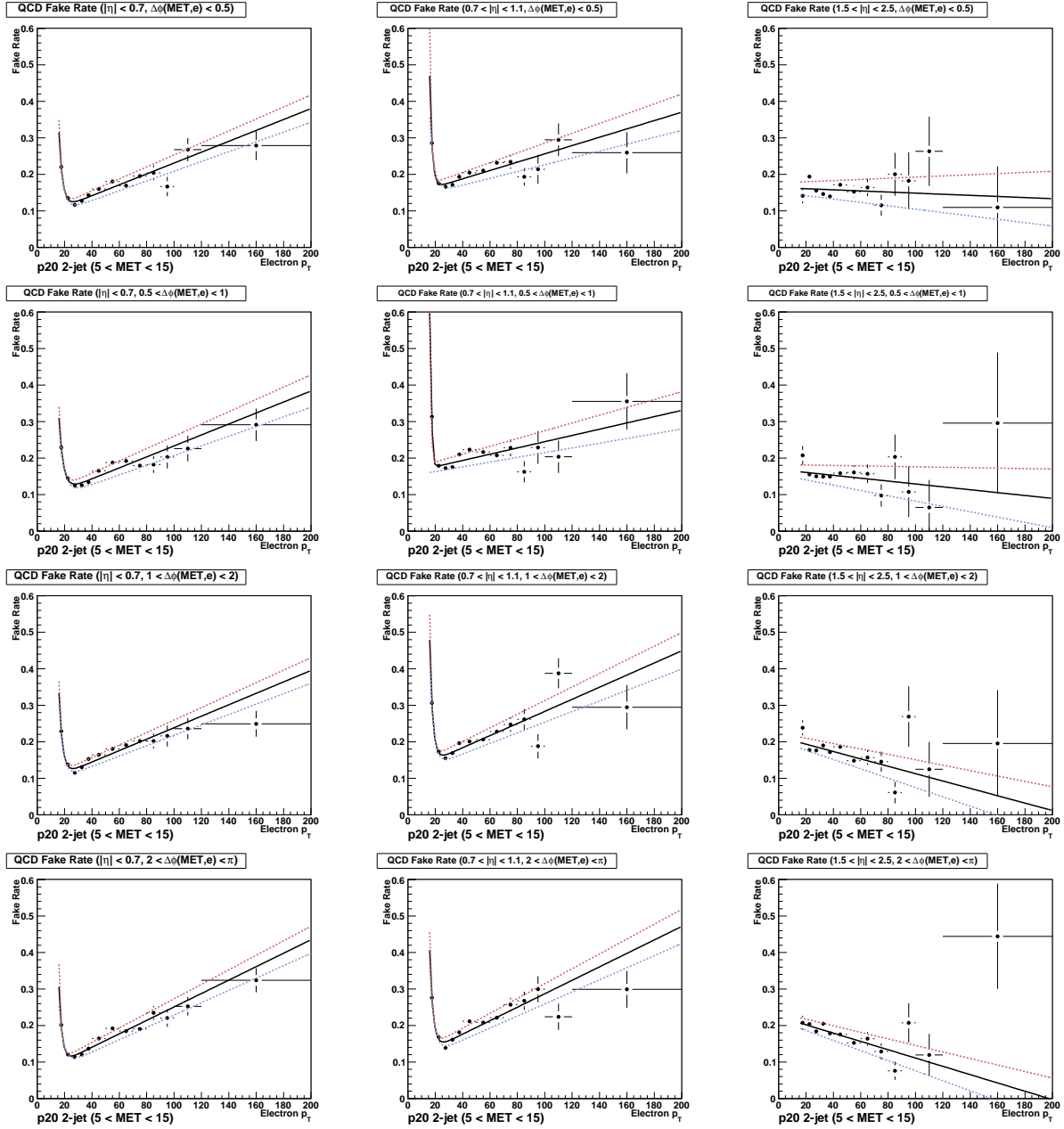


Figure B.3: The QCD fake rate for electrons in RunIIb 2-jet events as a function of $e p_T$ for each bin in $|\eta_{\text{CAL}}|$ vs. $\Delta\phi(\ell_T, e)$. The red dashed curve represents the $+1\sigma$ variation of the fit and the blue curve represents the -1σ variation.

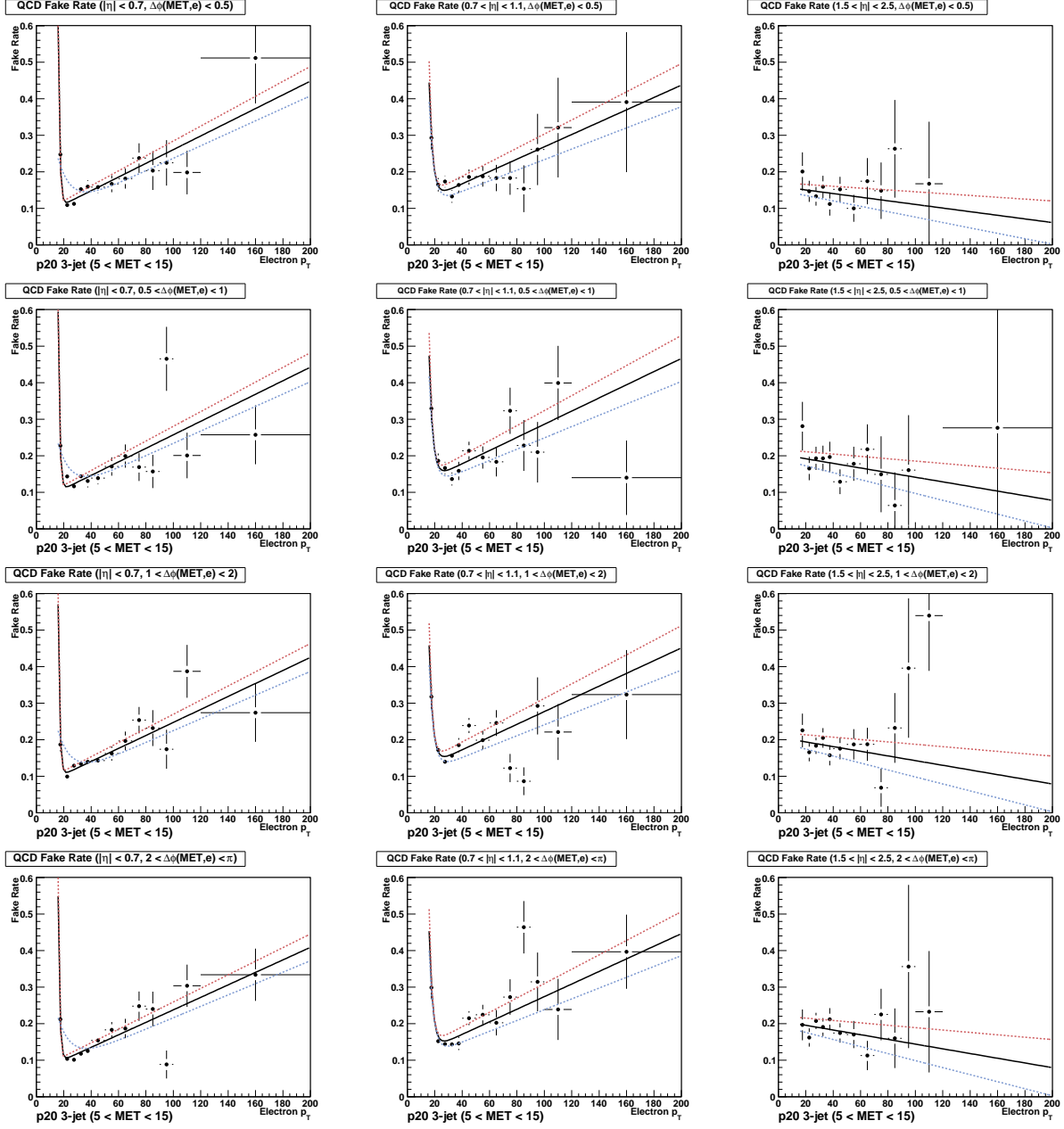
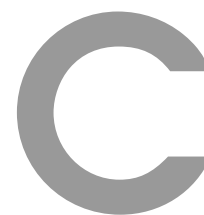


Figure B.4: The QCD fake rate for electrons in RunIIb 3-jet events as a function of $e p_T$ for each bin in $|\eta_{\text{CAL}}|$ vs. $\Delta\phi(\ell_T, e)$. The red dashed curve represents the $+1\sigma$ variation of the fit and the blue curve represents the -1σ variation.



RunIIa kinematic variables distributions

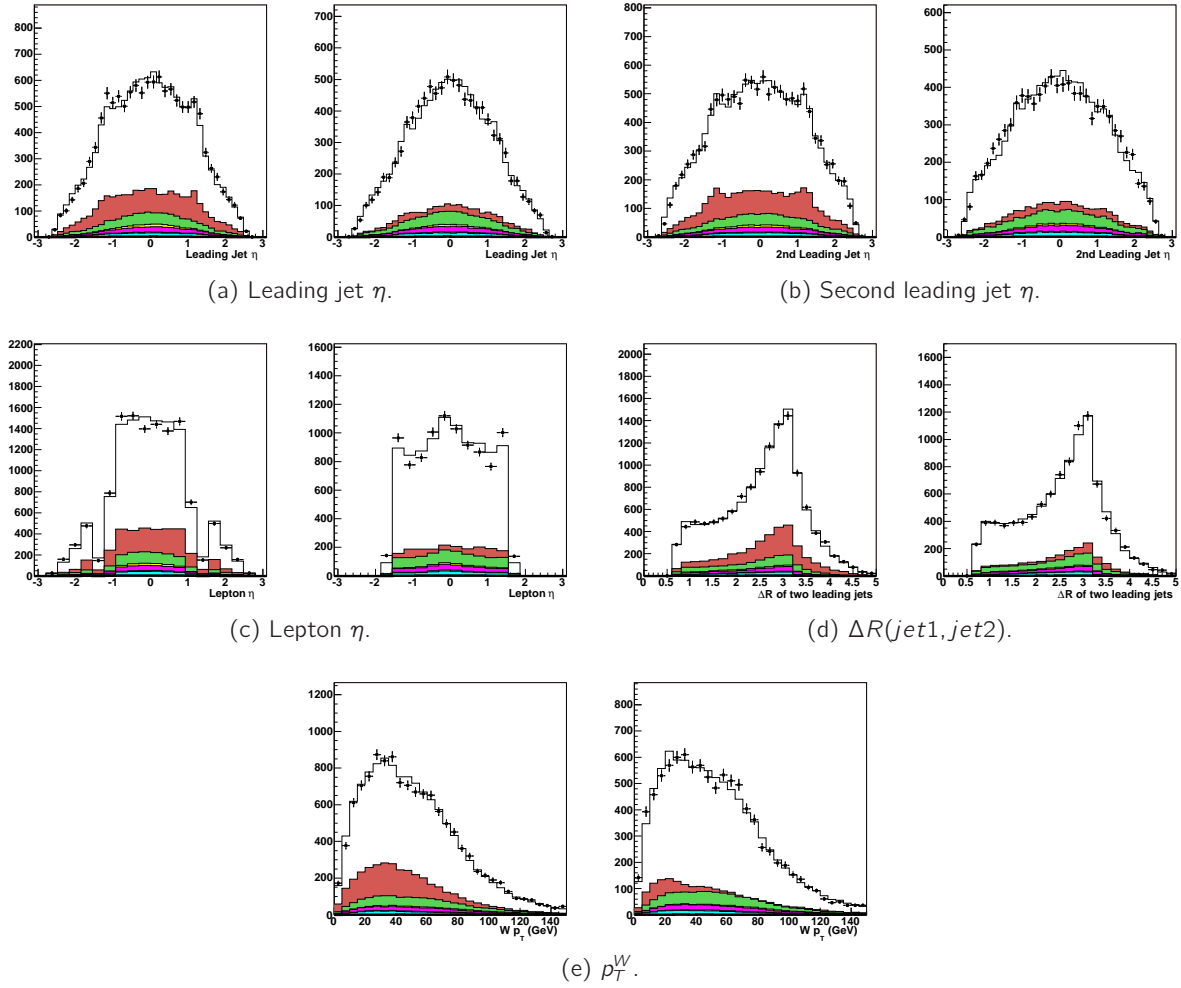


Figure C.1: η_{jet1} , η_{jet2} , η_{lep} , $\Delta R(jet1, jet2)$ and p_T^W distributions after all reweightings applied, for the RunIIa electron and muon channels (grouped in pairs, electron channel shown in the left figure and the muon channel in the right figure).

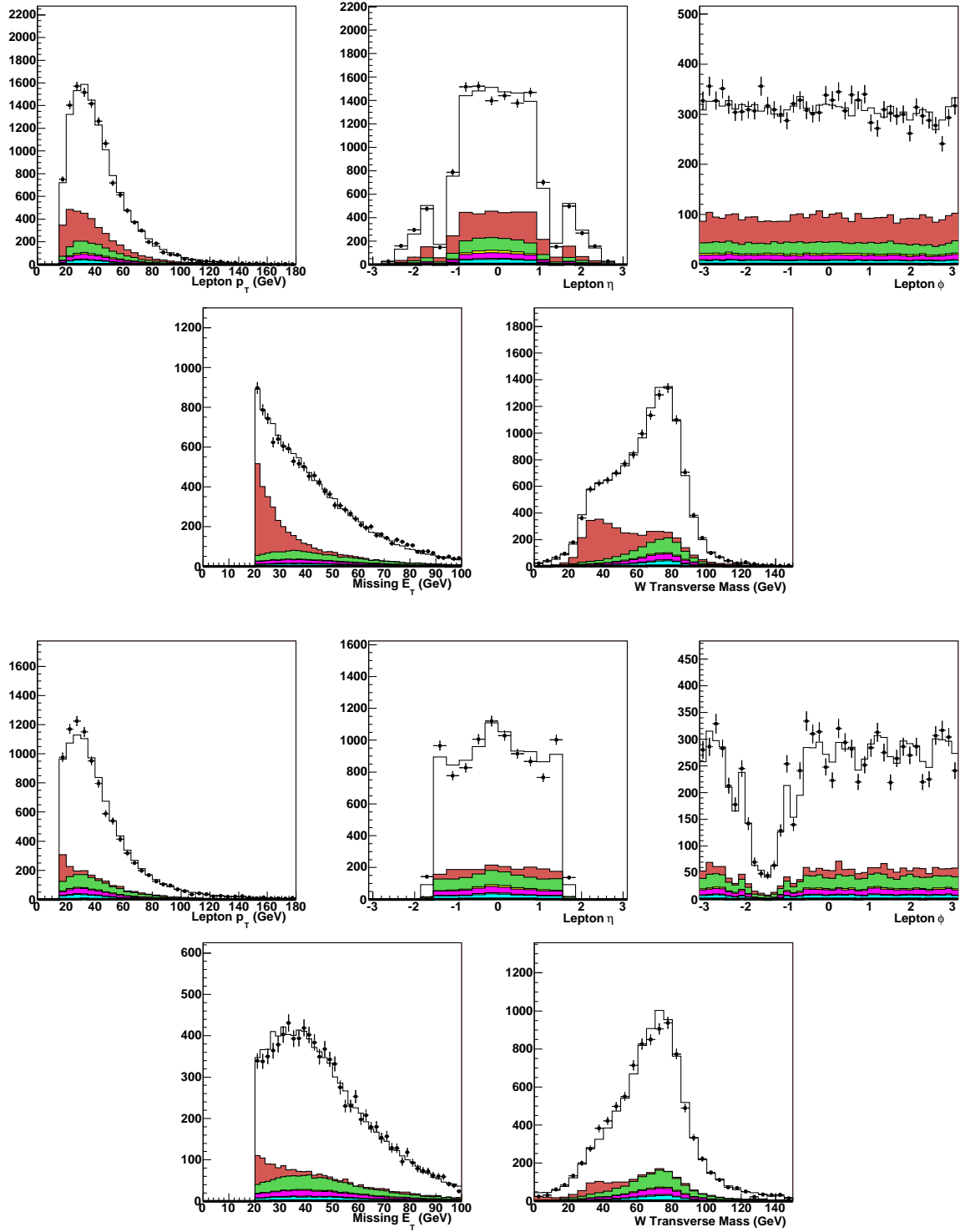


Figure C.2: The upper five figures show the lepton p_T , lepton η , lepton ϕ , \cancel{E}_T and M_T^W distributions, in the 2 jets RunIIa electron channel. The bottom five figures show the same variables in the RunIIa muon channel.

APPENDIX C. RUNIIA KINEMATIC VARIABLES DISTRIBUTIONS

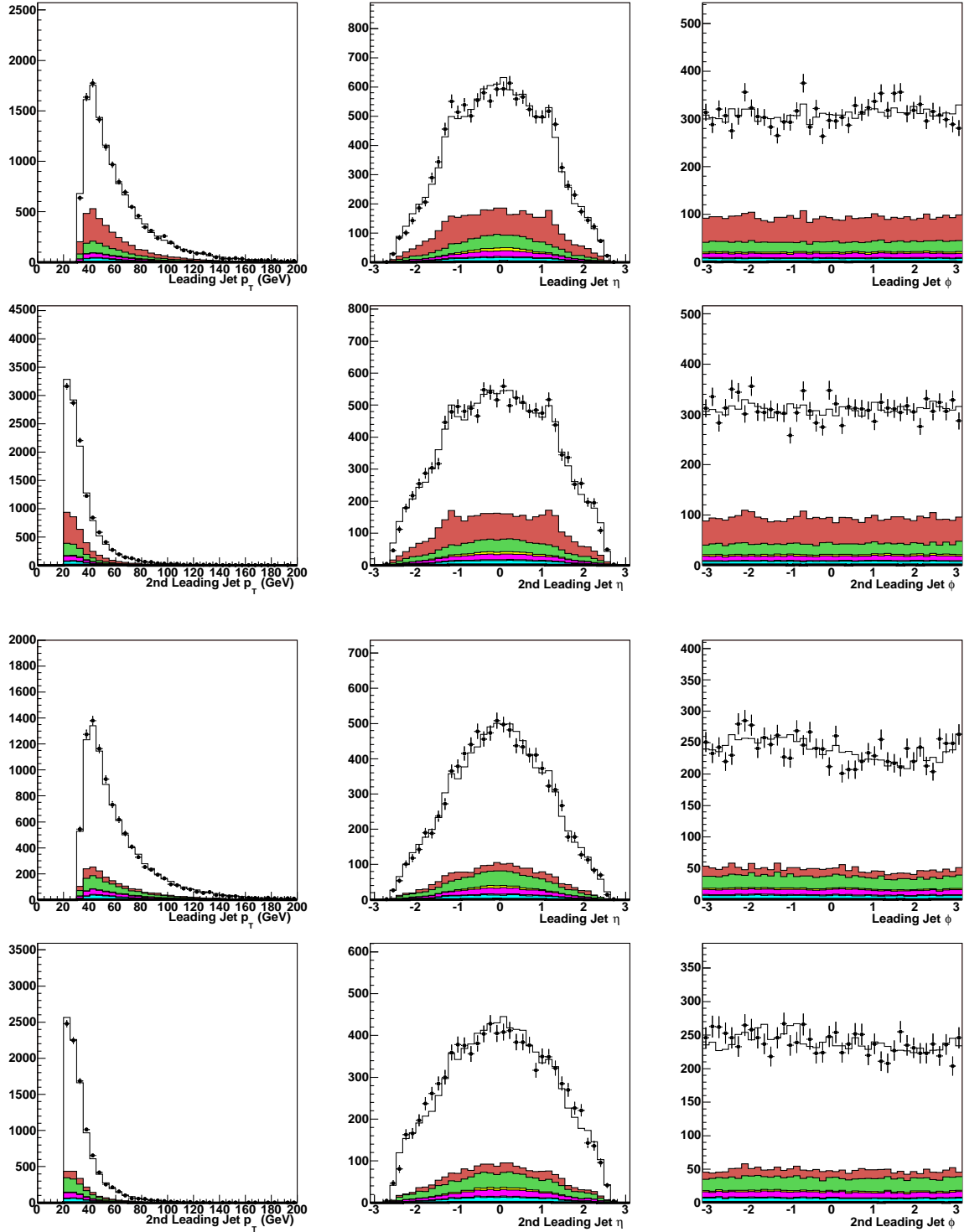


Figure C.3: The upper six figures show the p_T^{jet} , η^{jet} and ϕ^{jet} distributions for the leading and second leading jet, in the 2 jets RunIIa electron channel. The bottom six figures show the same variables in the RunIIa muon channel.

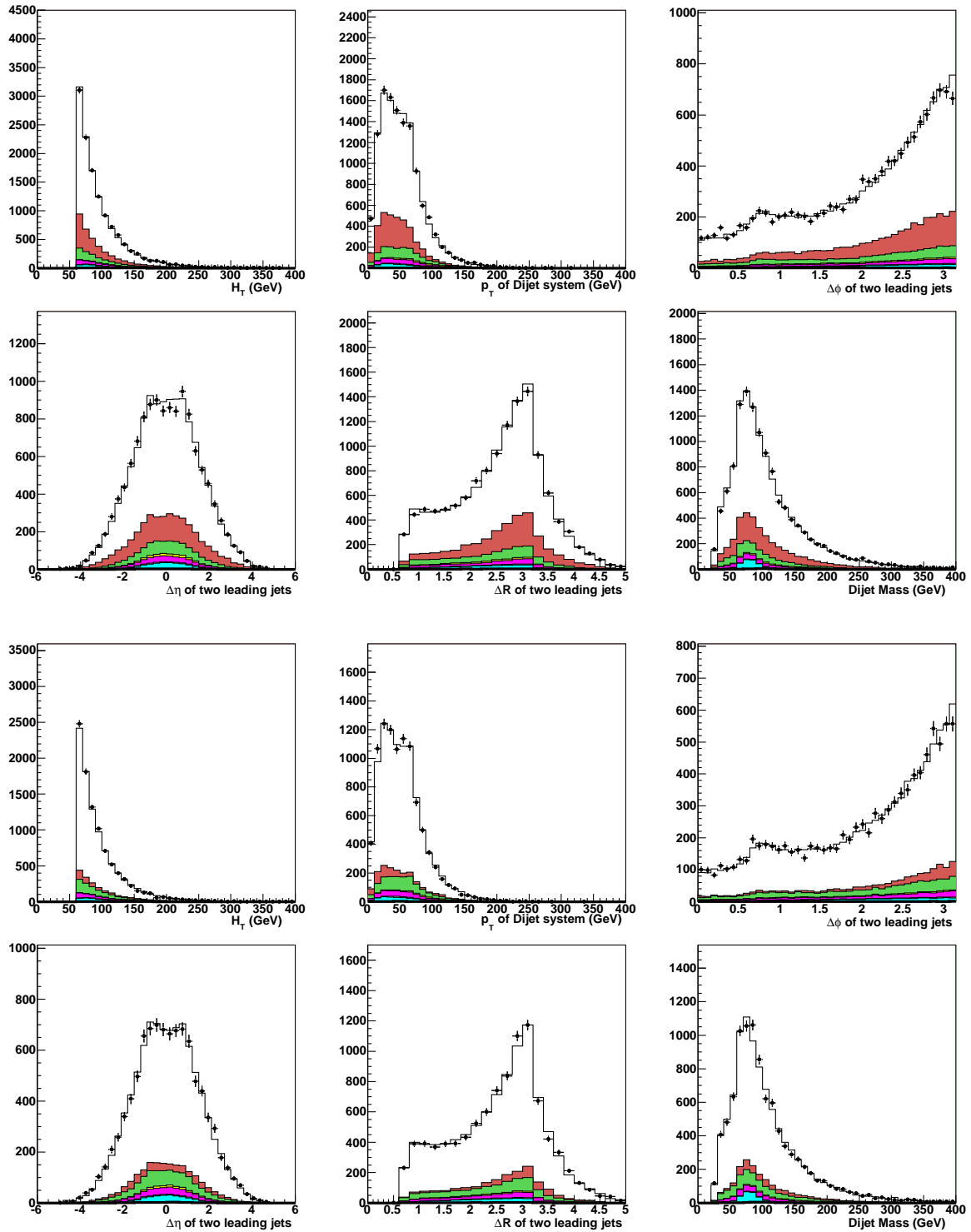


Figure C.4: The upper six figures show the H_T , $p_T(j,j)$, $\Delta\phi(j,j)$, $\Delta\eta(j,j)$, $\Delta R(j,j)$ and M_{jj} distributions for the leading and second leading jet, in the 2 jets RunIIa electron channel. The bottom six figures show the same variables in the RunIIa muon channel.

APPENDIX C. RUNIIA KINEMATIC VARIABLES DISTRIBUTIONS

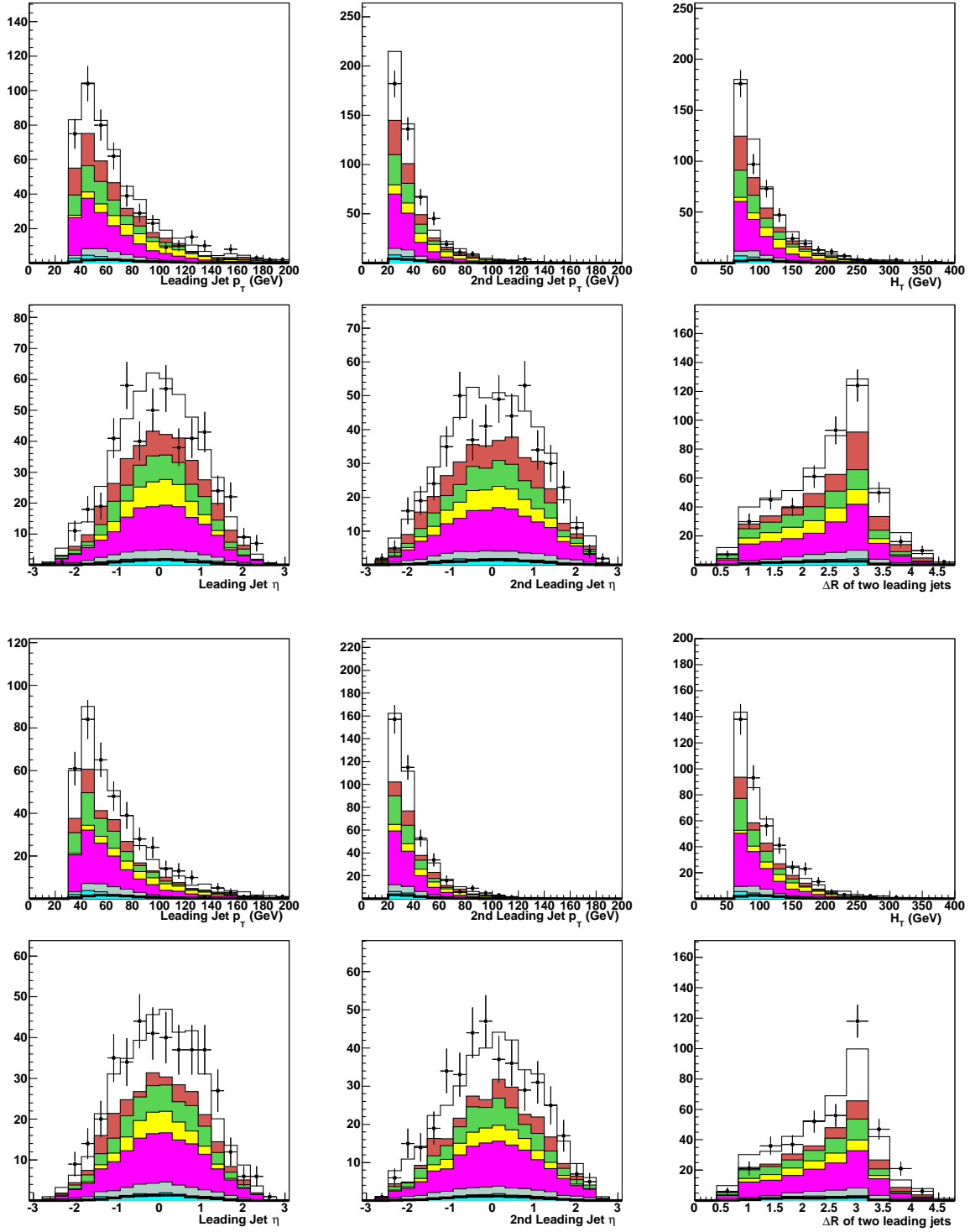


Figure C.5: The upper six figures show the $p_T^{jet_1}$, $p_T^{jet_2}$, H_T , η^{jet_1} , η^{jet_2} and $\Delta R(j, j)$ distributions in the 2 jets RunIIa electron channel with 1 b -tagged jet. The bottom six figures show the same variables in the muon channel.

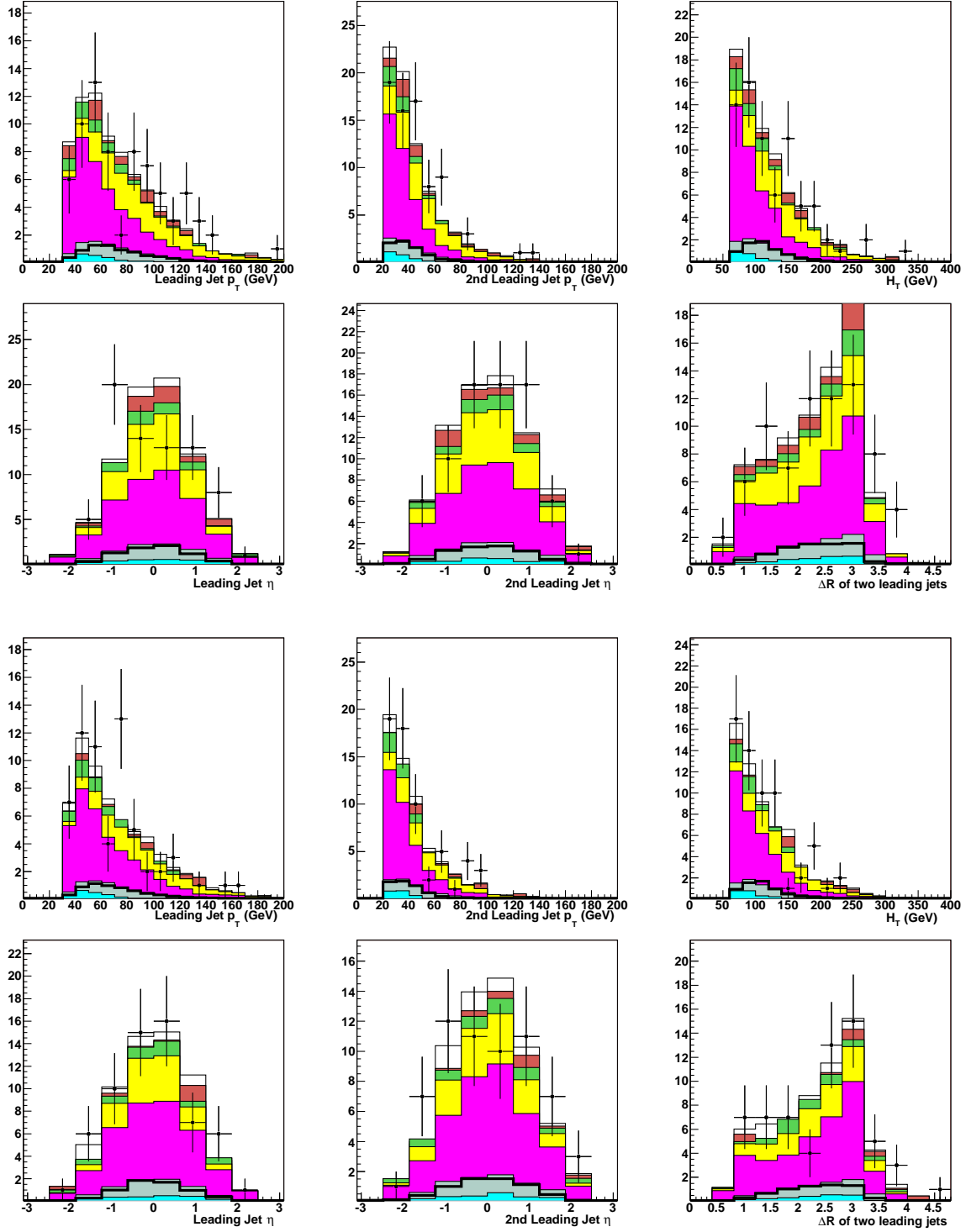


Figure C.6: The upper six figures show the $p_T^{jet_1}$, $p_T^{jet_2}$, H_T , η^{jet_1} , η^{jet_2} and $\Delta R(j,j)$ distributions in the 2 jets Run11a electron channel with 2 b -tagged jets. The bottom six figures show the same variables in the Run11a muon channel.

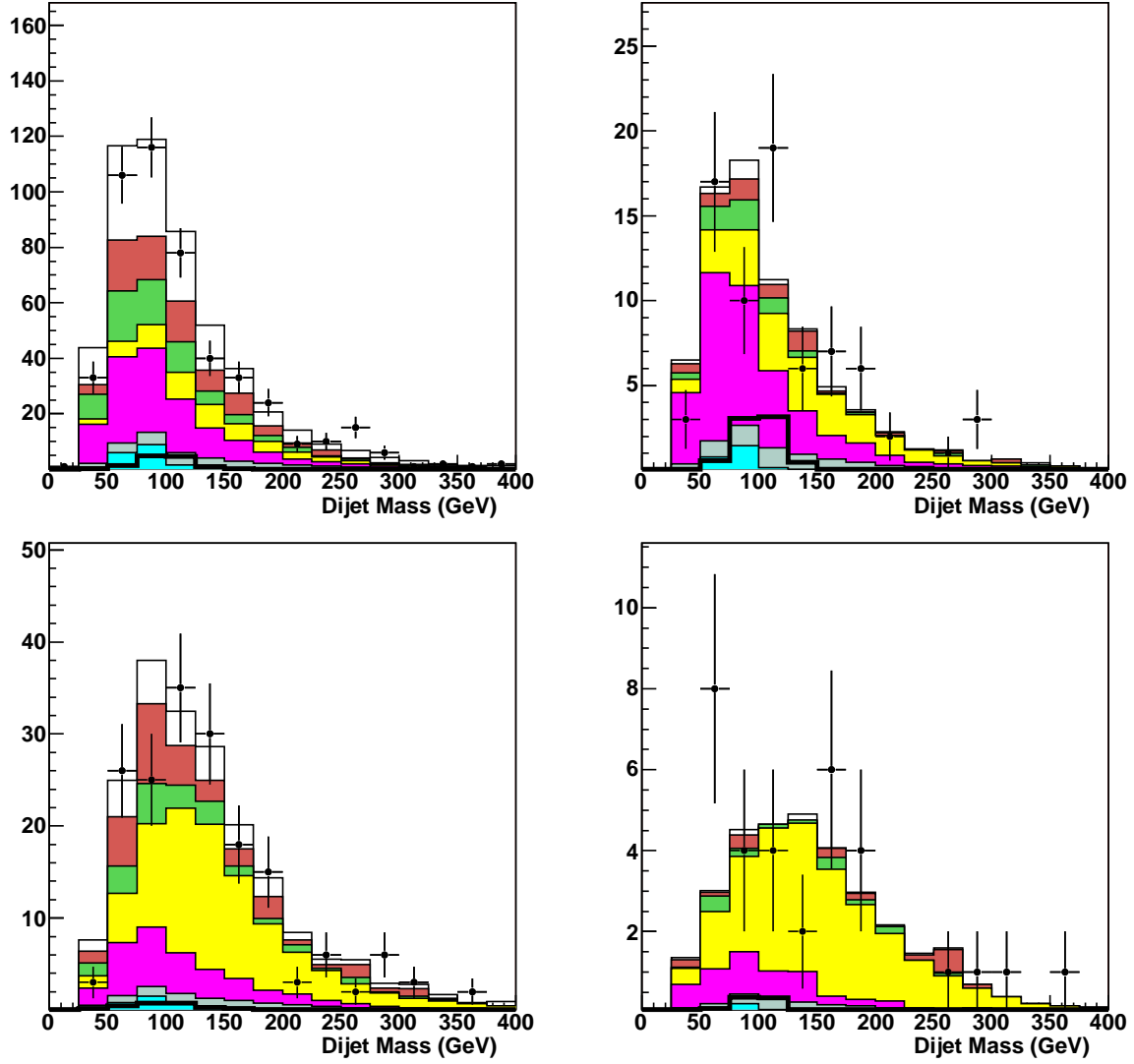


Figure C.7: Dijet invariant mass distributions in the RunIIa electron channel, for the 2 (top figures) and 3 jets (bottom plots) channels with one (left) or two (right) *b*-tagged jets.

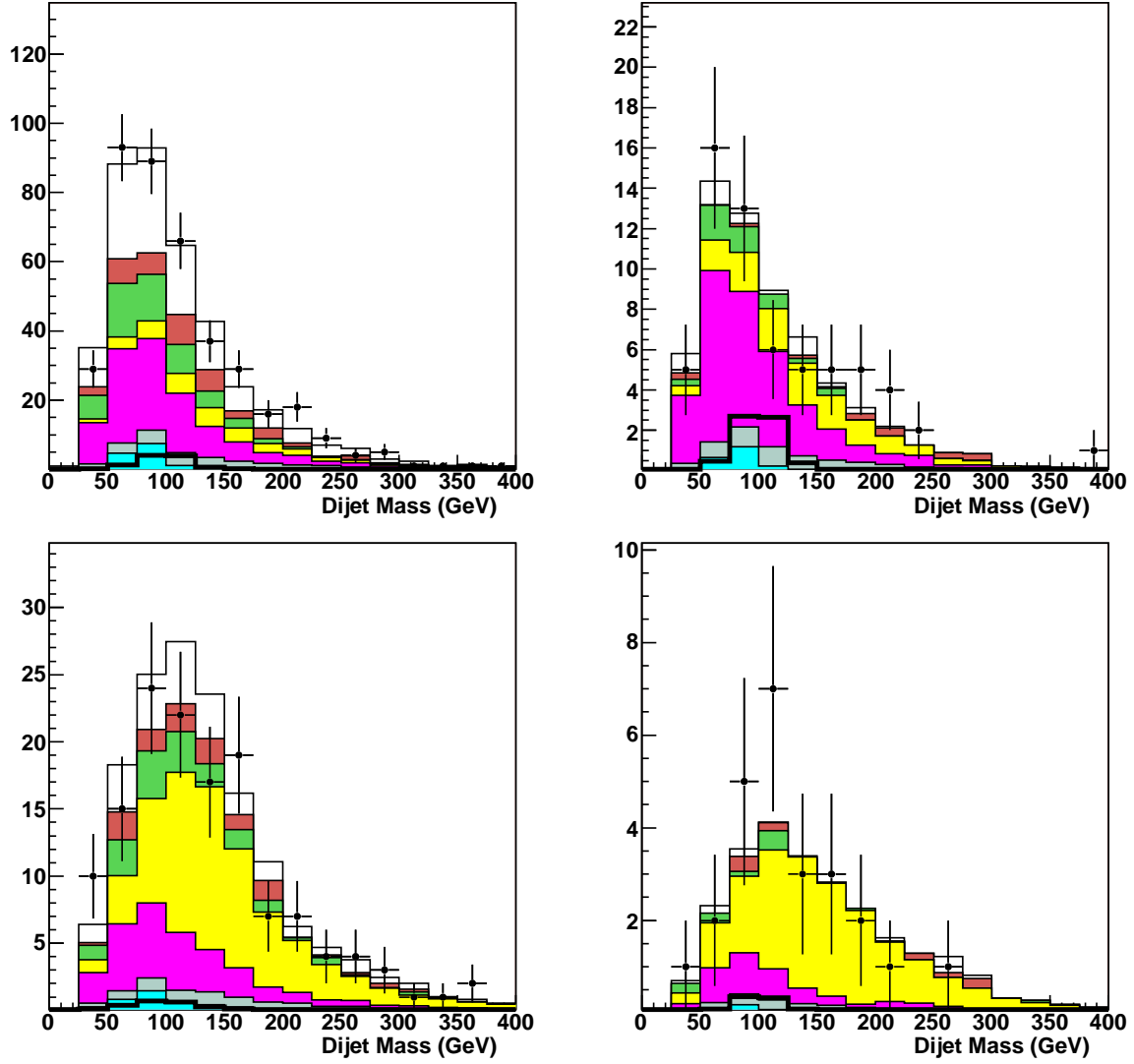


Figure C.8: Dijet invariant mass distributions in the RunIIa muon channel, for the 2 (top figures) and 3 jets (bottom plots) channels with one (left) or two (right) b -tagged jets.



Random Forest input variables distributions

APPENDIX D. RANDOM FOREST INPUT VARIABLES DISTRIBUTIONS

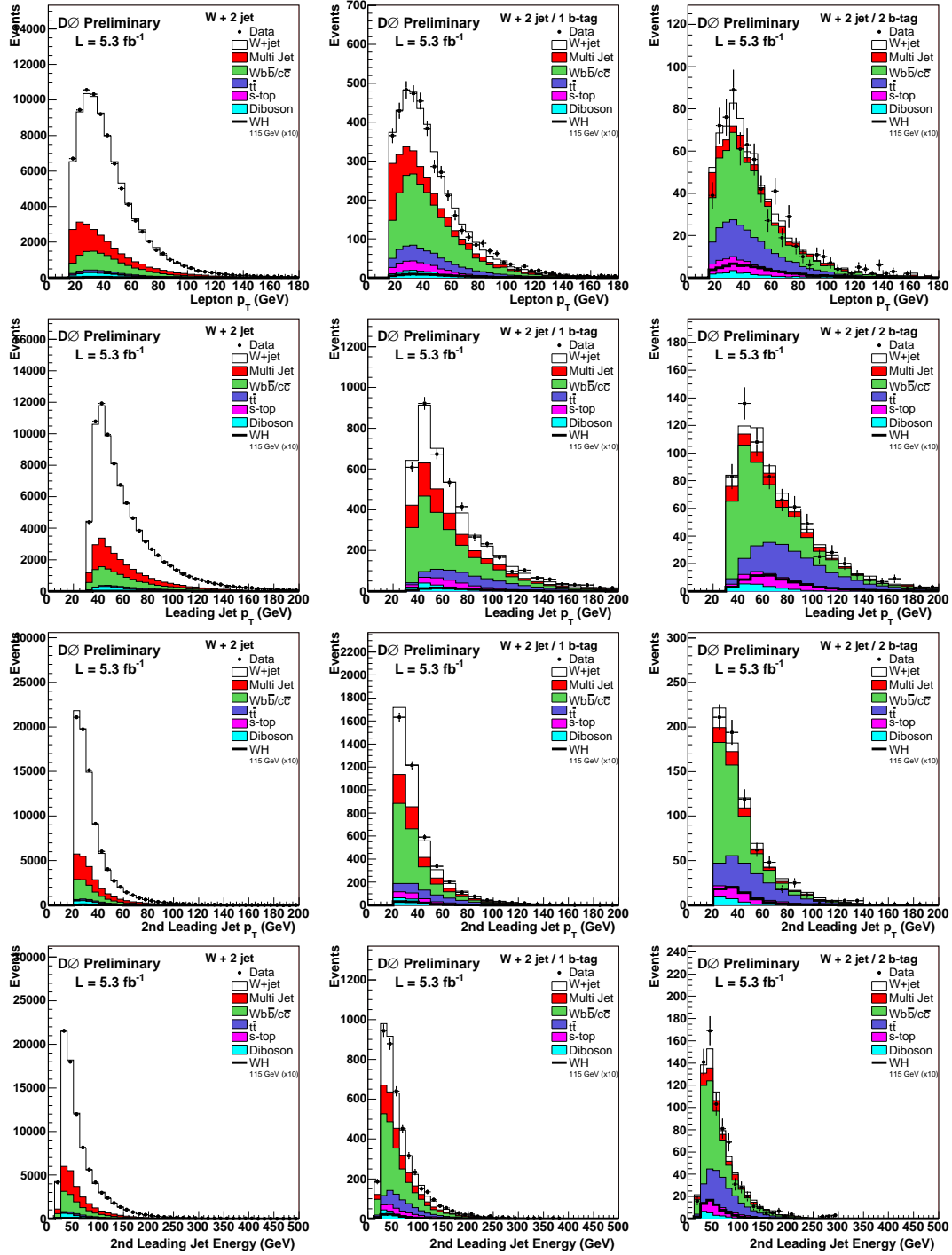


Figure D.1: Distribution in the 2 jets sample of lepton p_T (first row), leading jet p_T (second row), 2nd leading jet p_T (third row) and 2nd leading jet energy (bottom row). The left, center and right columns show the pretag, single b -tagged and double b -tagged samples, respectively.

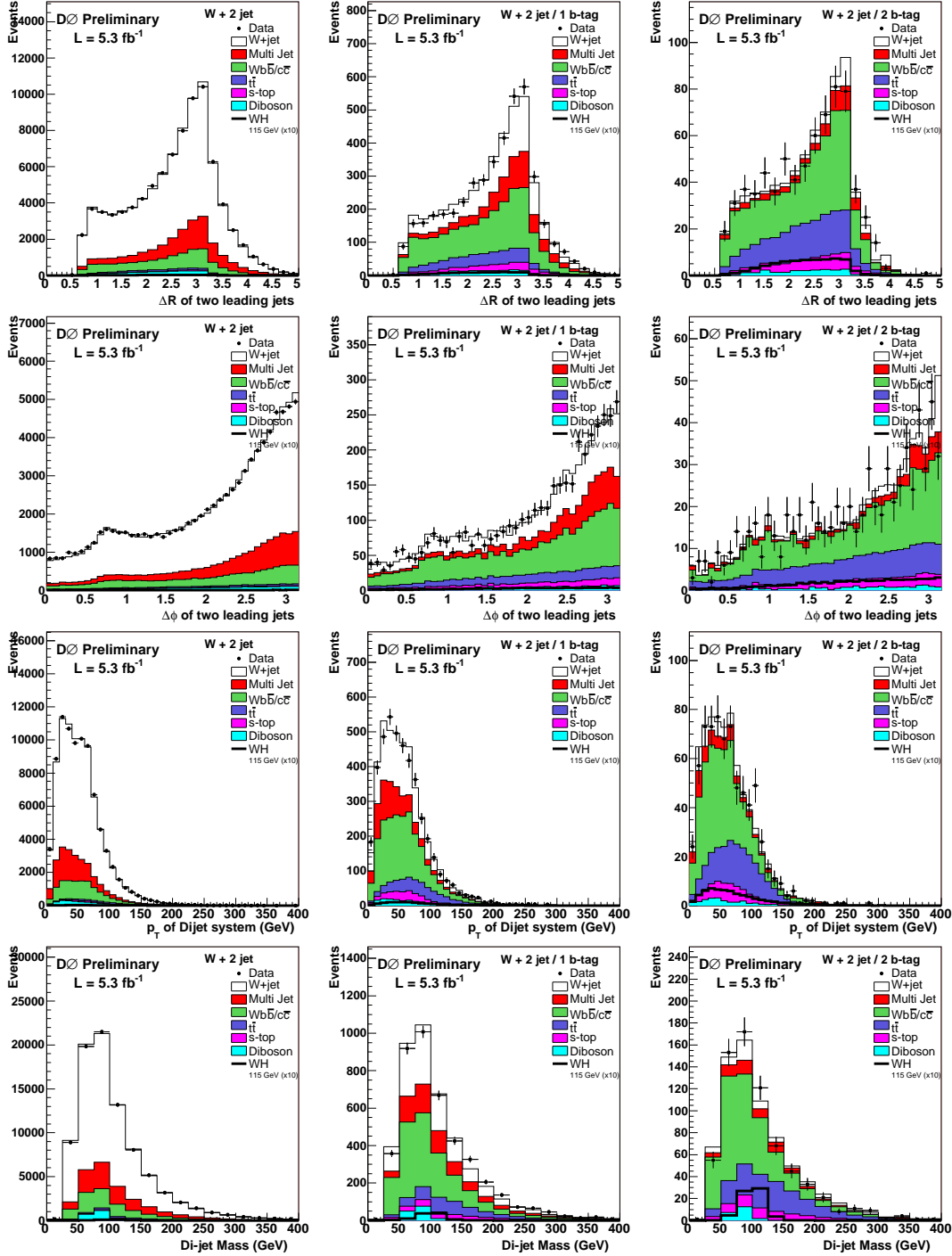


Figure D.2: Distribution in the 2 jets sample of $\Delta R(\text{two leading jets})$ (first row), $\Delta\phi(\text{two leading jets})$ (second row), $p_T(\text{dijet system})$ (third row) and dijet invariant mass (bottom row). The left, center and right columns show the pretag, single b -tagged and double b -tagged samples, respectively.

APPENDIX D. RANDOM FOREST INPUT VARIABLES DISTRIBUTIONS

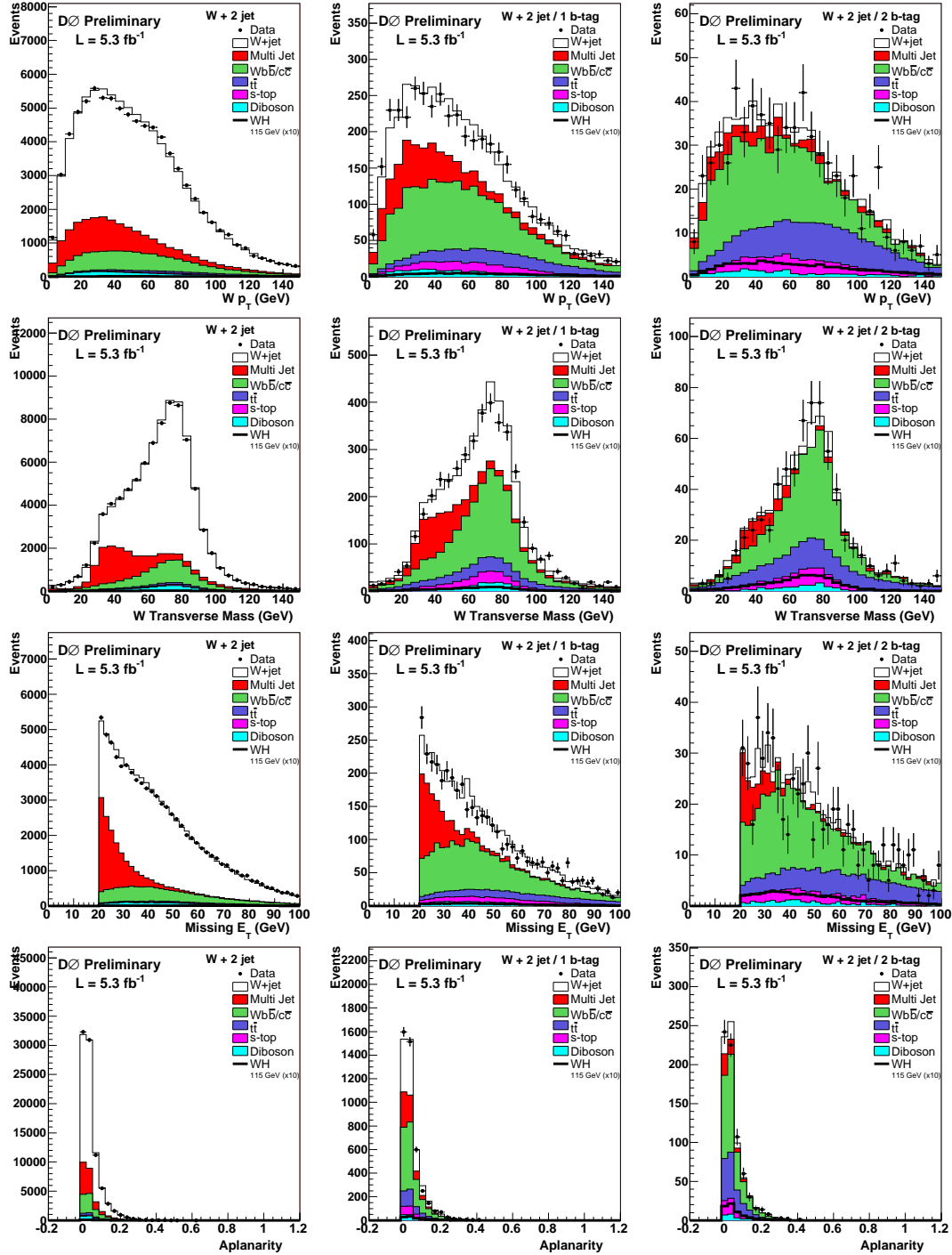


Figure D.3: Distribution in the 2 jets sample of p_T^W (first row), M_T^W (second row) and E_T^{miss} (third row) and aplanarity (bottom row). The left, center and right columns show the pretag, single b -tagged and double b -tagged samples, respectively.

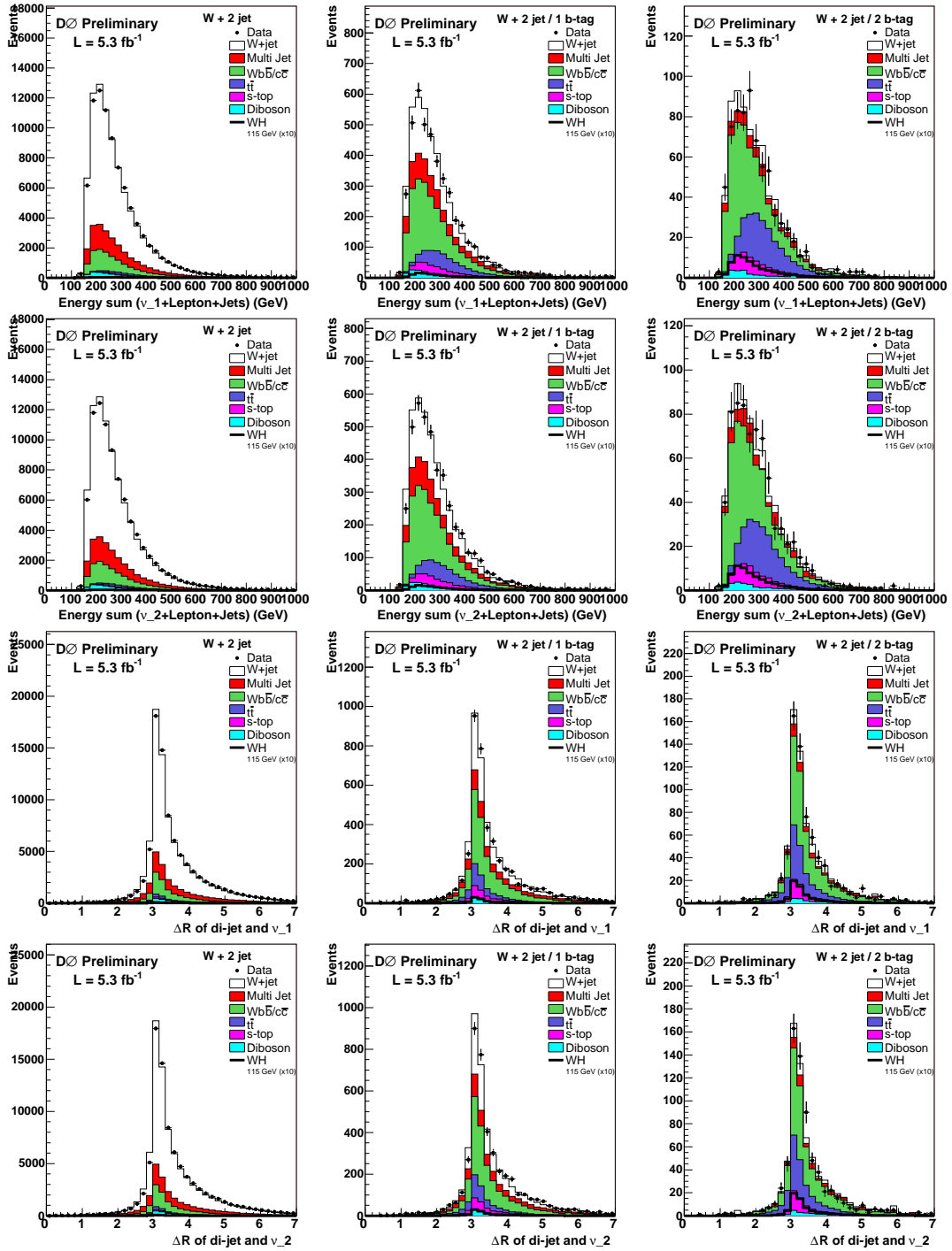


Figure D.4: Distribution in the 2 jets sample of the invariant mass of the $\ell\nu jj$ system using the first solution for p_Z^ν (first row), invariant mass of the $\ell\nu jj$ system using the second solution for p_Z^ν (second row), ΔR between the W and the dijet system using the first solution for p_Z^ν (third row) and ΔR between the W and the dijet system using the second solution for p_Z^ν (bottom row). The left, center and right columns show the pretag, single b -tagged and double b -tagged samples, respectively.

APPENDIX D. RANDOM FOREST INPUT VARIABLES DISTRIBUTIONS

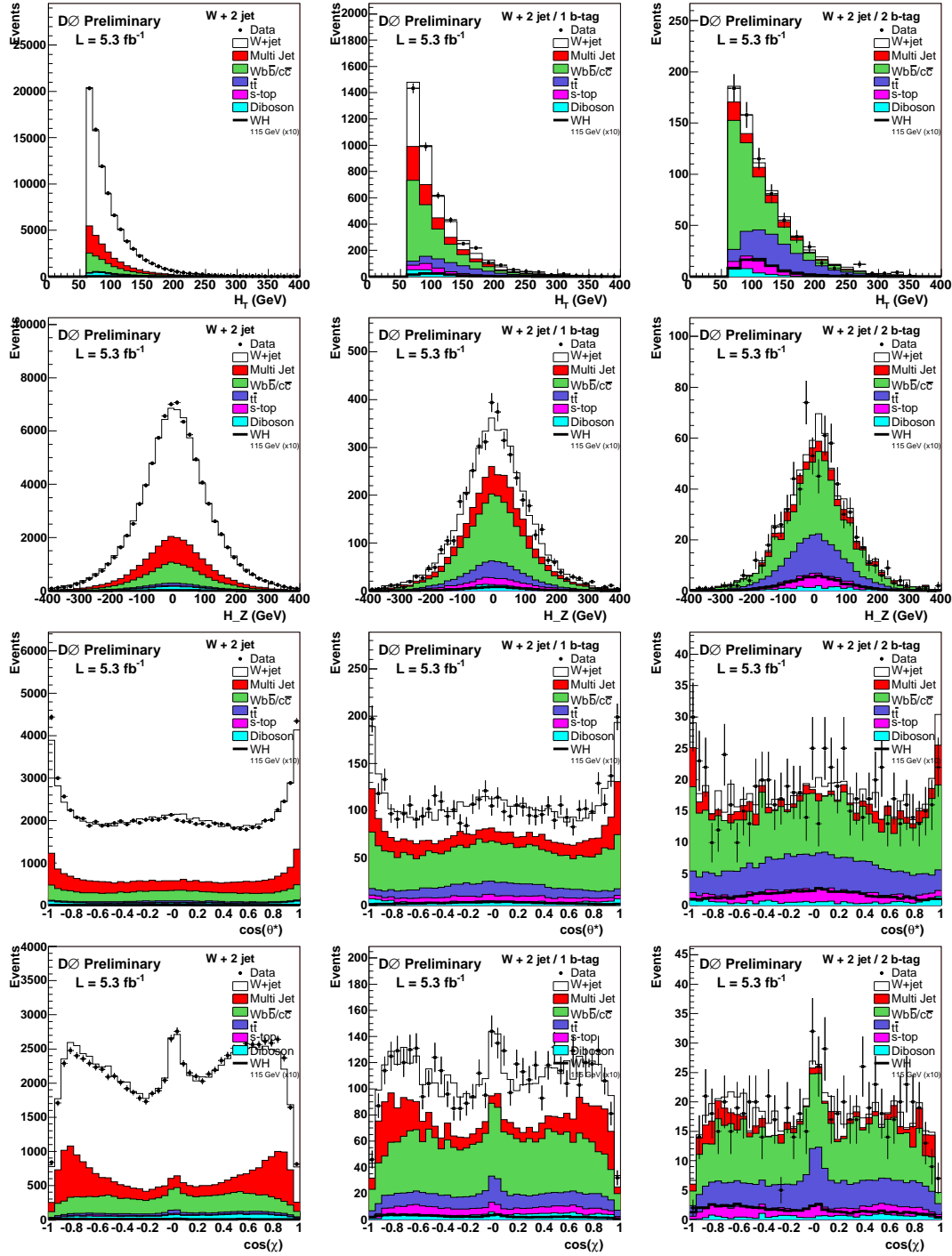


Figure D.5: Distribution in the 2 jets sample of H_T (first row), H_Z (second row), $\cos\theta^*$ (third row) and $\cos\chi$ (bottom row). The left, center and right columns show the pretag, single b -tagged and double b -tagged samples, respectively.

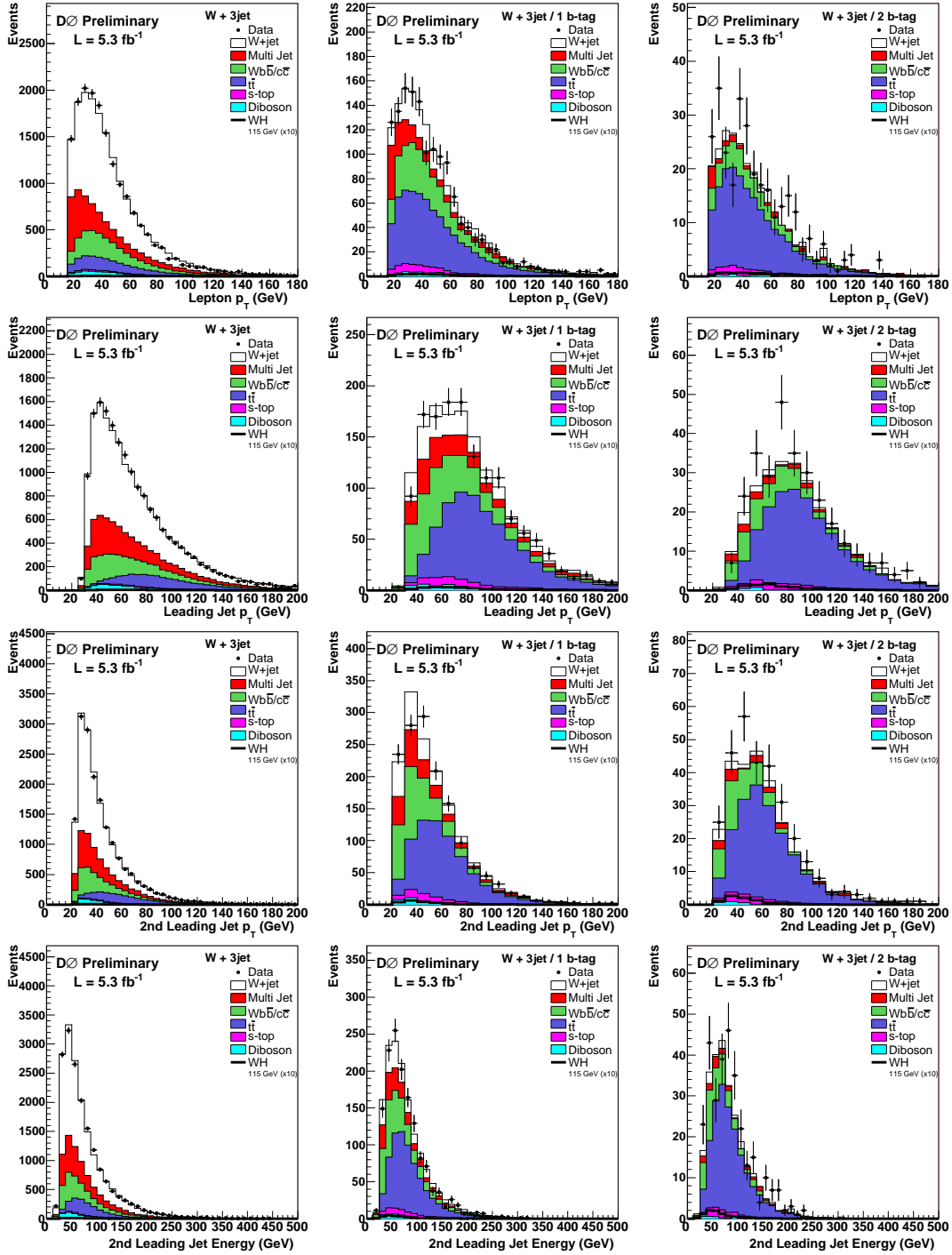


Figure D.6: Distribution in the 3 jets sample of lepton p_T (first row), leading jet p_T (second row), 2^{nd} leading jet p_T (third row) and 2^{nd} leading jet energy (bottom row). The left, center and right columns show the pretag, single b -tagged and double b -tagged samples, respectively.

APPENDIX D. RANDOM FOREST INPUT VARIABLES DISTRIBUTIONS

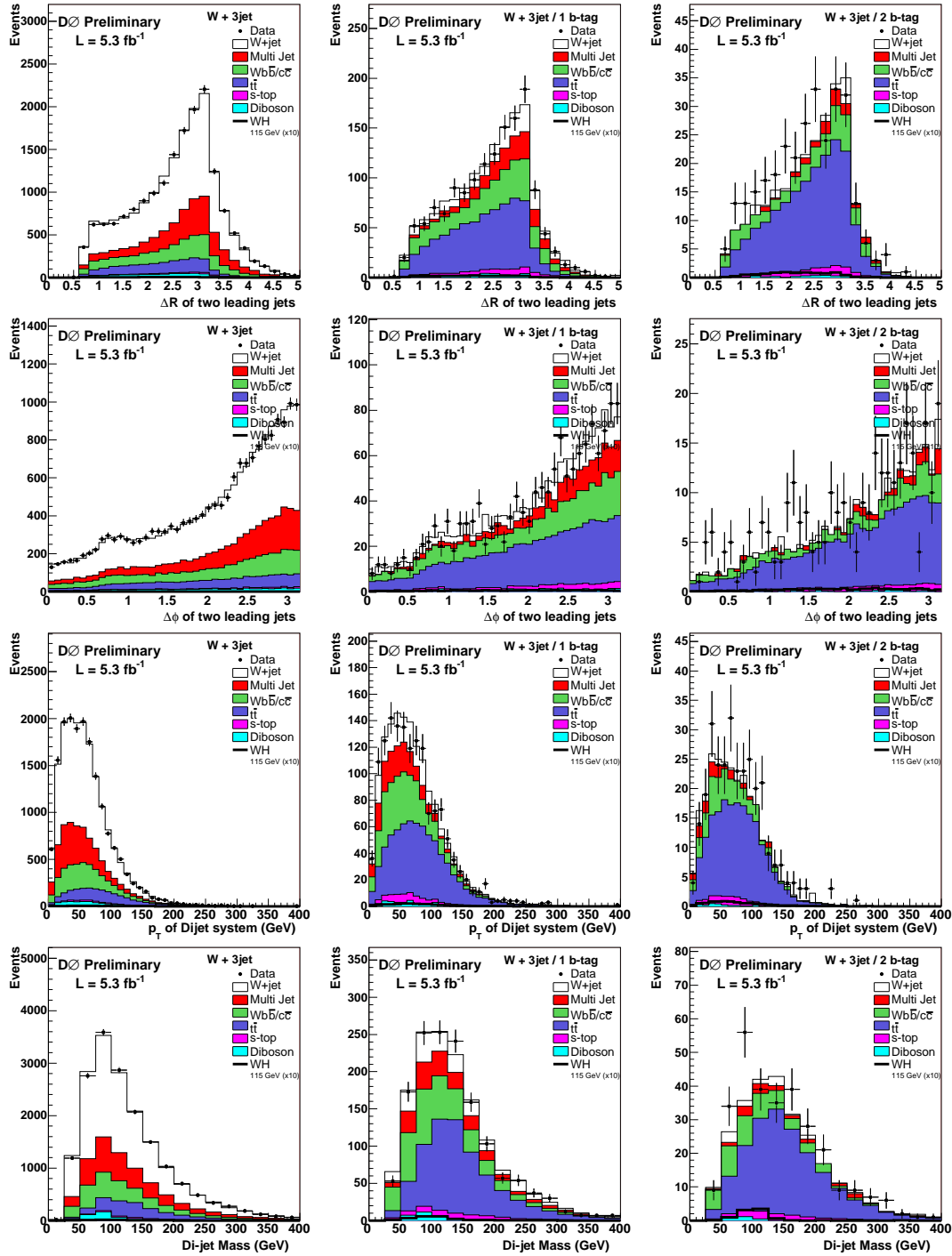


Figure D.7: Distribution in the 3 jets sample of ΔR (two leading jets) (first row), $\Delta \phi$ (two leading jets) (second row), p_T (dijet system) (third row) and dijet invariant mass (bottom row). The left, center and right columns show the pretag, single b -tagged and double b -tagged samples, respectively.

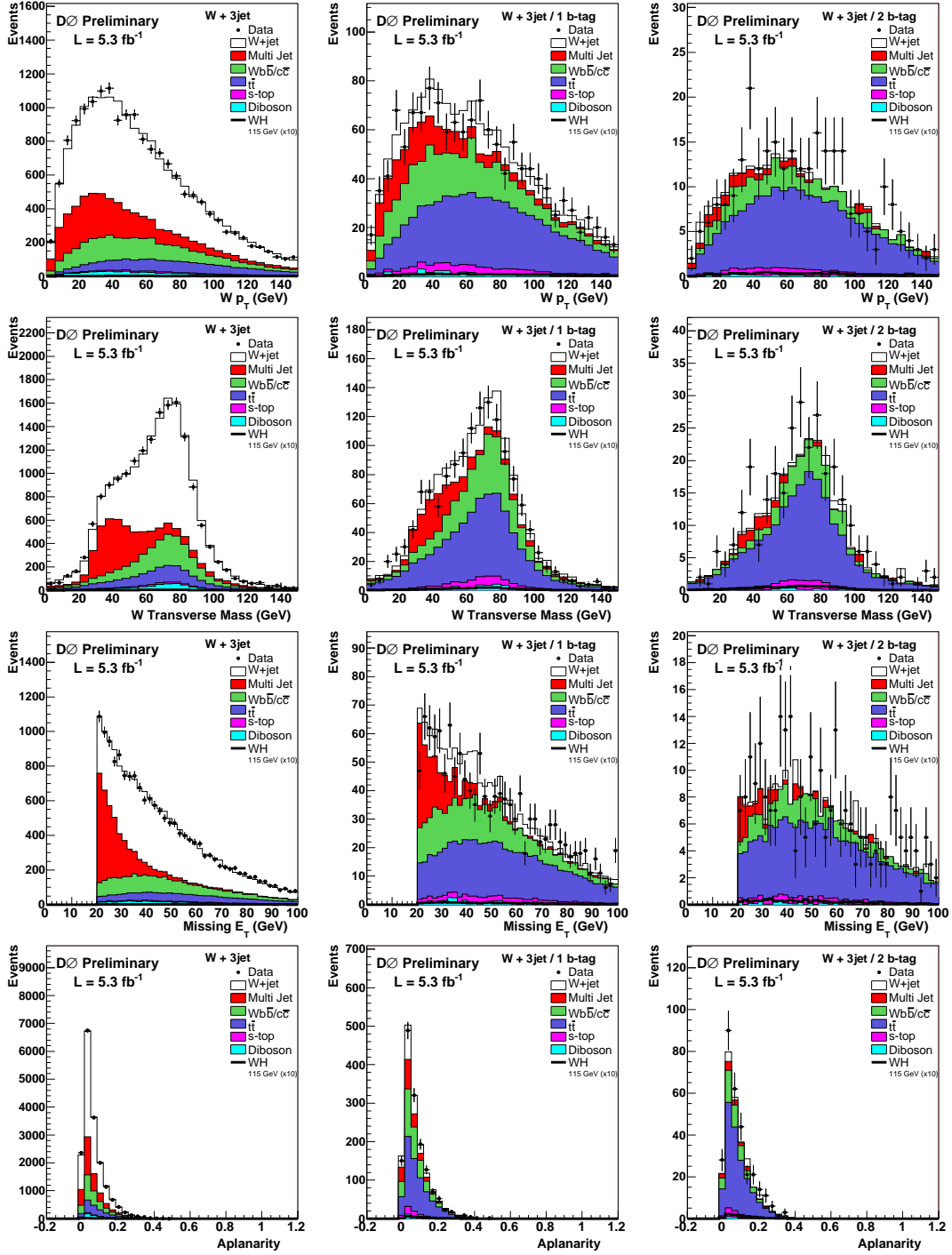


Figure D.8: Distribution in the 3 jets sample of p_T^W (first row), M_T^W (second row), E_T^{miss} (third row) and aplanarity (bottom row). The left, center and right columns show the pretag, single b -tagged and double b -tagged samples, respectively.

APPENDIX D. RANDOM FOREST INPUT VARIABLES DISTRIBUTIONS

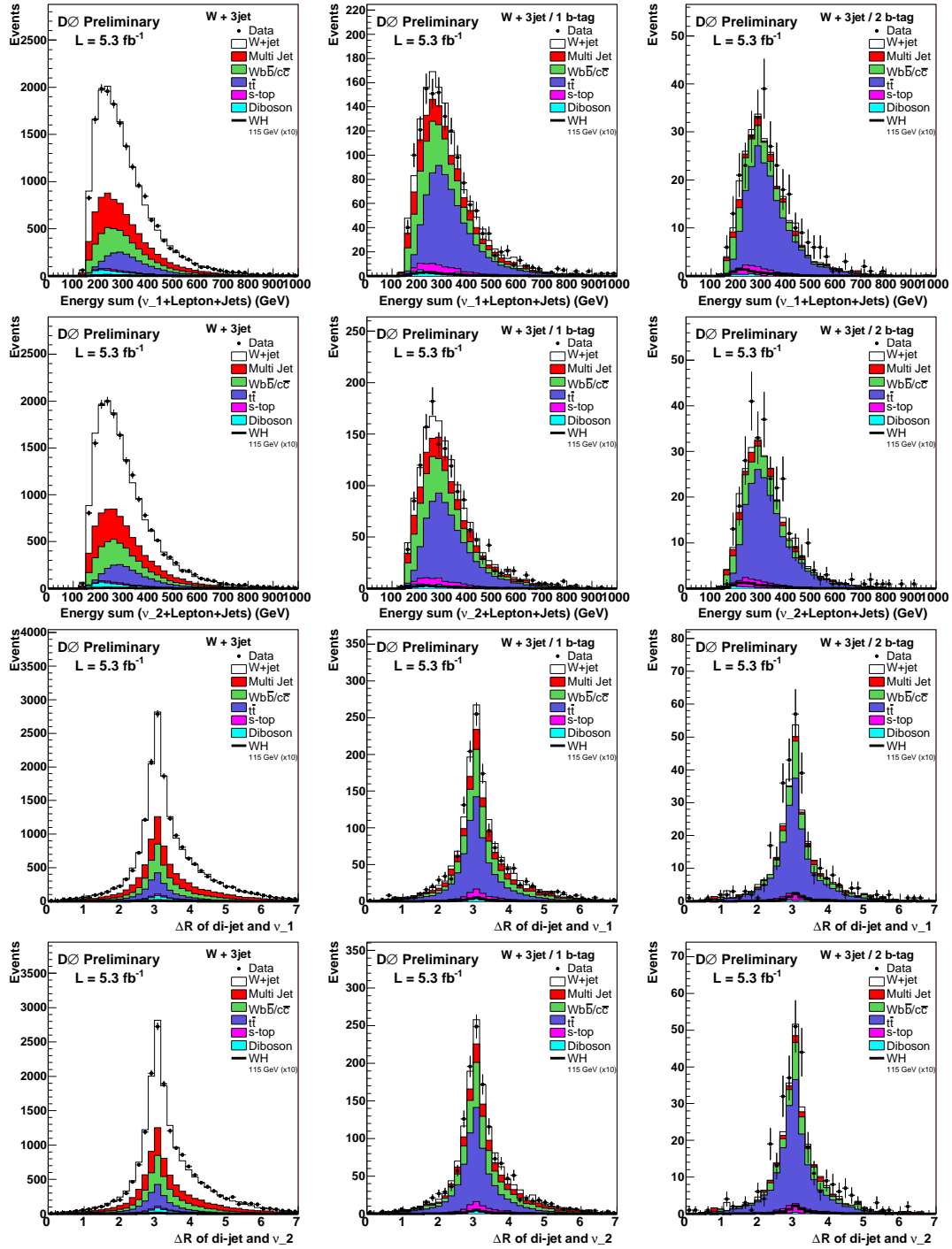
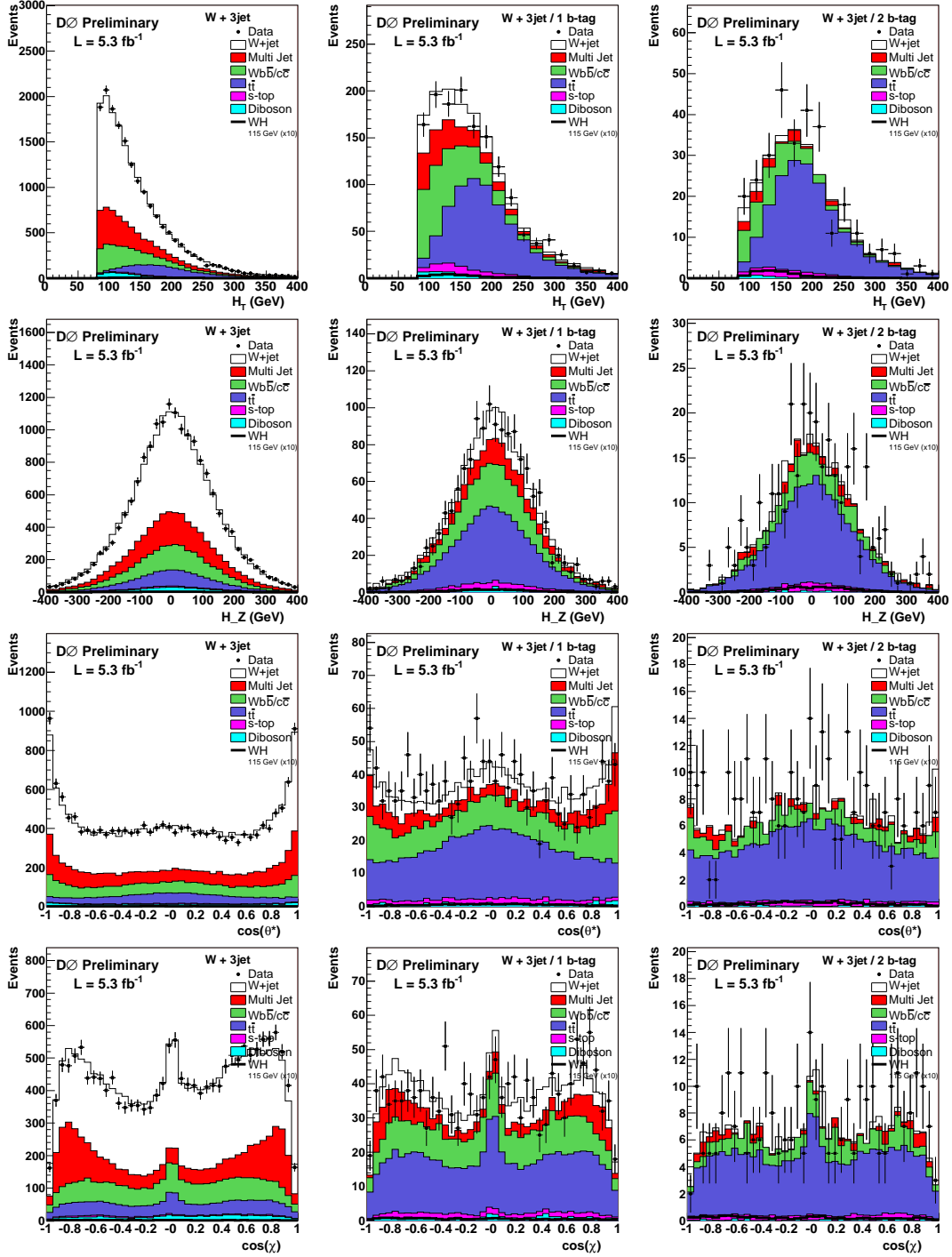


Figure D.9: Distribution in the 3 jets sample of the invariant mass of the $lvjj$ system using the first solution for p_z^ν (first row), invariant mass of the $lvjj$ system using the second solution for p_z^ν (second row), ΔR between the W and the dijet system using the first solution for p_z^ν (third row) and ΔR between the W and the dijet system using the second solution for p_z^ν (bottom row). The left, center and right columns show the pretag, single b -tagged and double b -tagged samples, respectively.





Random Forest output distributions

APPENDIX E. RANDOM FOREST OUTPUT DISTRIBUTIONS

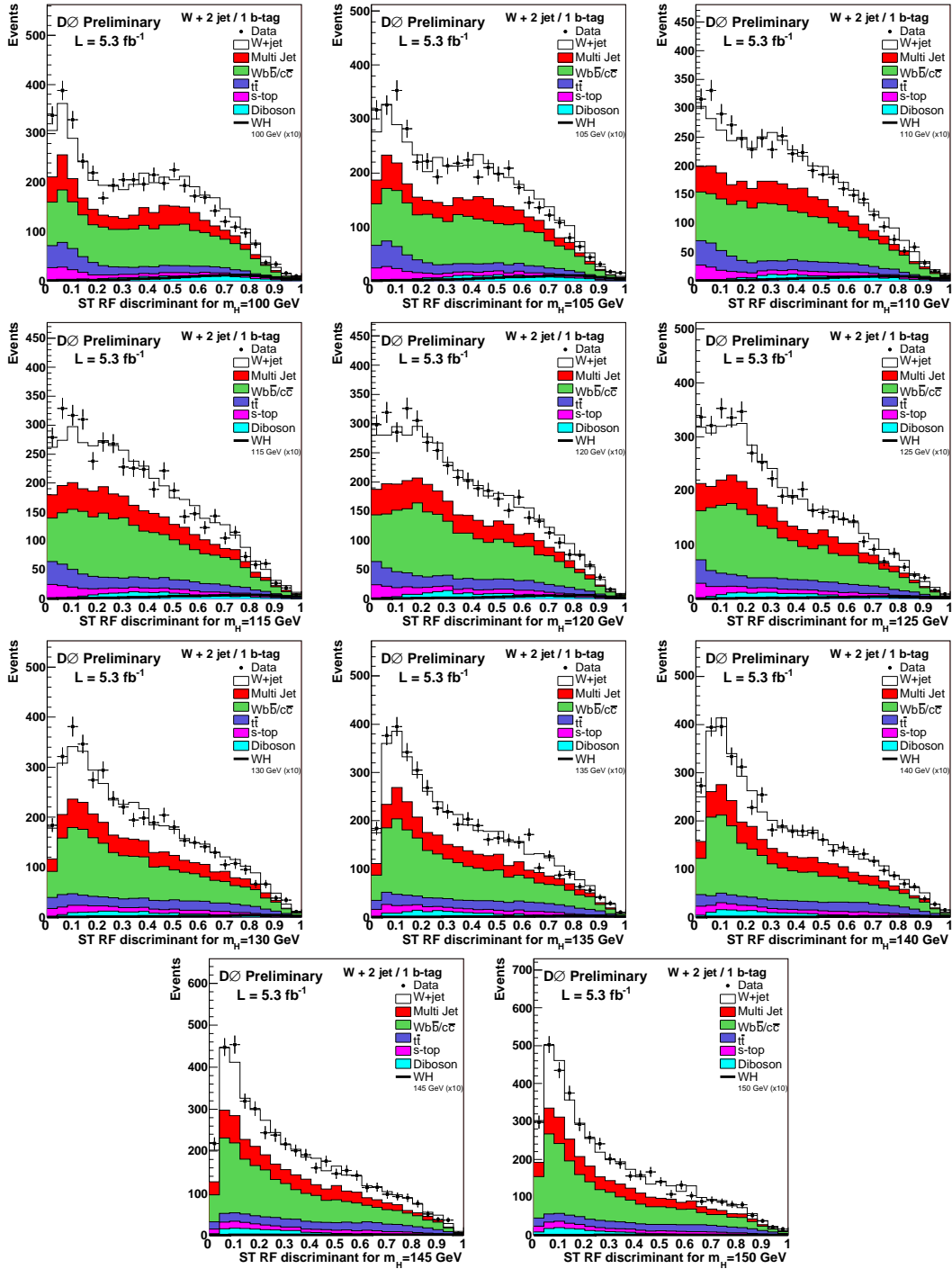


Figure E.1: RF output distributions for $M_H = 100$ GeV to $M_H = 150$ GeV, from top left to bottom right figures, for the combination of 2 jets single b -tagged events.

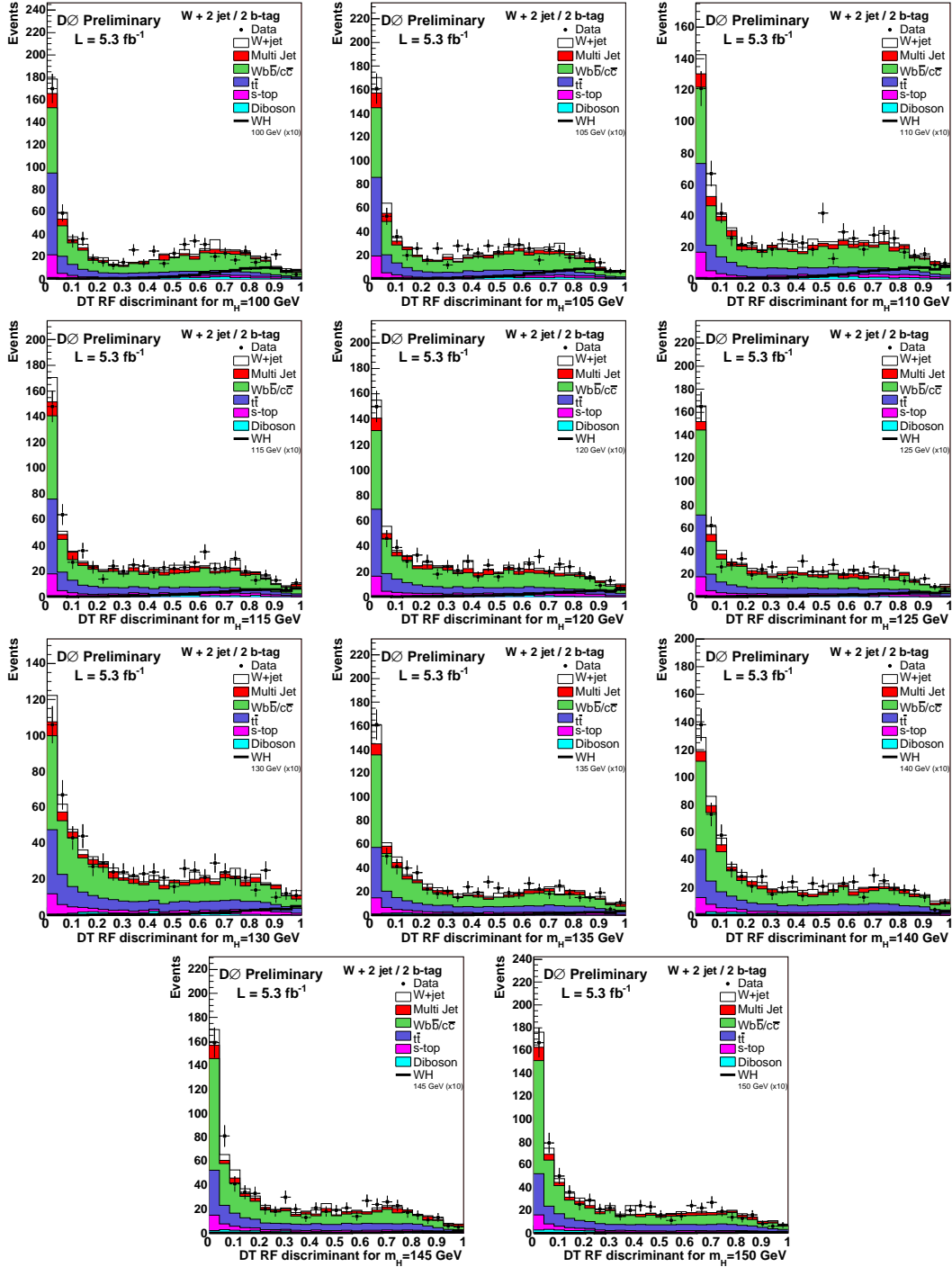


Figure E.2: RF output distributions for $M_H = 100$ GeV to $M_H = 150$ GeV, from top left to bottom right figures, for the combination of 2 jets double b -tagged events.

APPENDIX E. RANDOM FOREST OUTPUT DISTRIBUTIONS

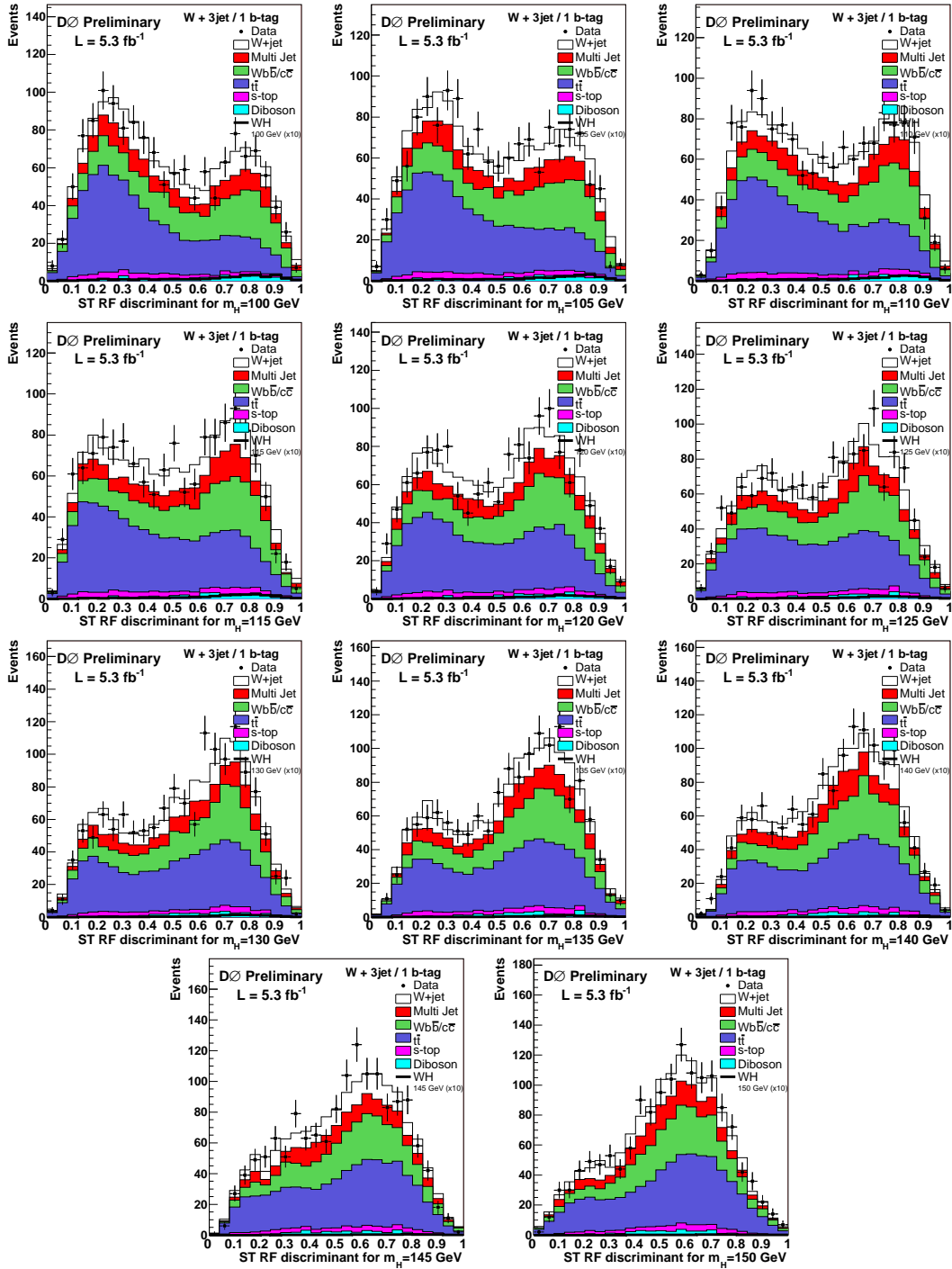


Figure E.3: RF output distributions for $M_H = 100$ GeV to $M_H = 150$ GeV, from top left to bottom right figures, for the combination of 3 jets single b -tagged events.

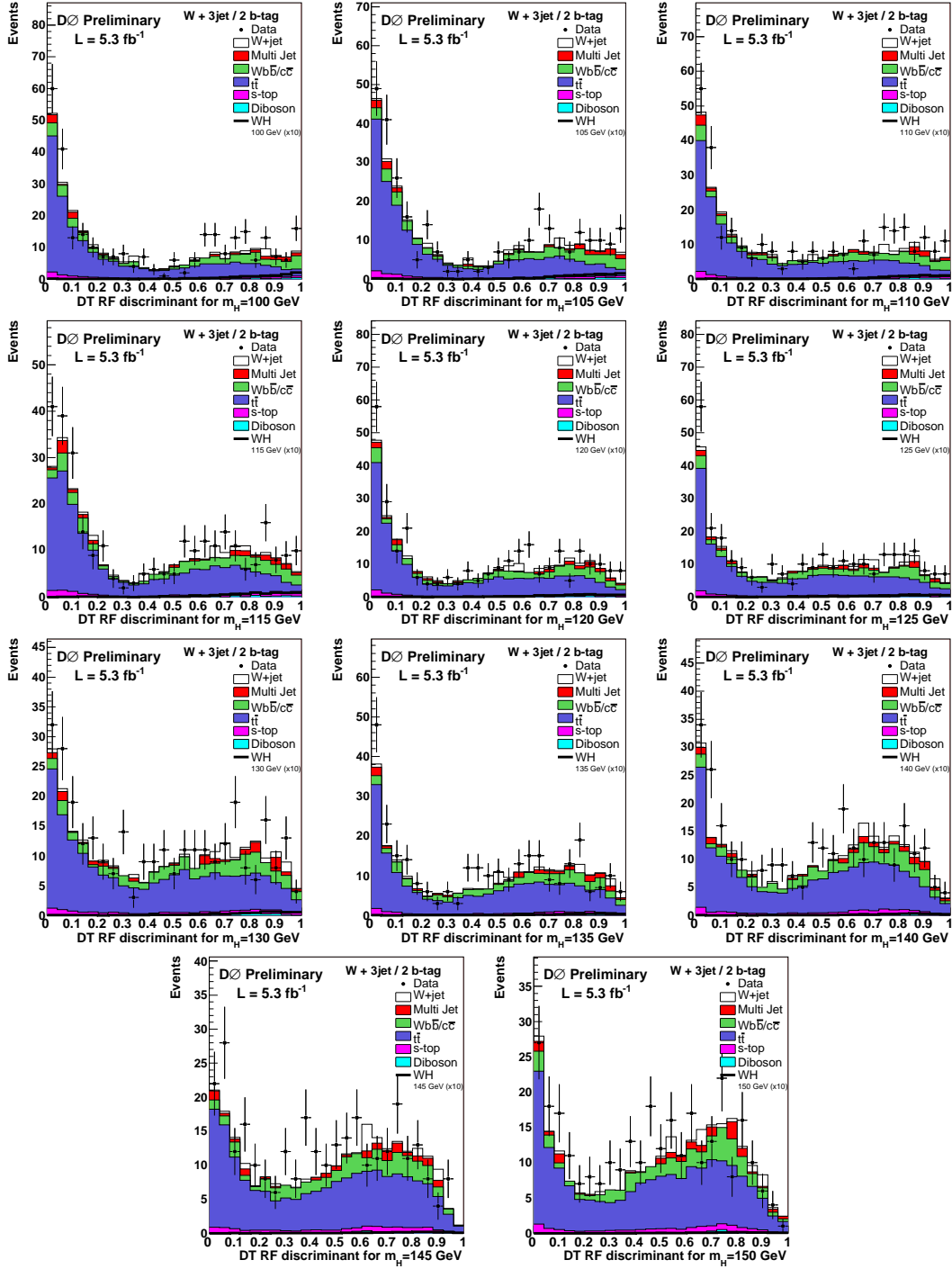


Figure E.4: RF output distributions for $M_H = 100$ GeV to $M_H = 150$ GeV, from top left to bottom right figures, for the combination of 3 jets double b -tagged events.



Systematic uncertainties distributions

As detailed in Section 5.2.2, the distributions for systematic uncertainties for each lepton flavor, number of jets in the final state and uncertainty source are shown in this section. In each figure, the $+1\sigma$ variation is shown as a red line, while the corresponding -1σ variation is shown in blue. The shape of nominal random forest output distributions are also included in the figures as dashed black lines, with an arbitrary vertical scale. Figures are shown only for the RunIIb data-taking period, in the 2 jets channel. Figures for the 3 jets channels are available in [106].

F.1 RunIIB electron 2 jets channel

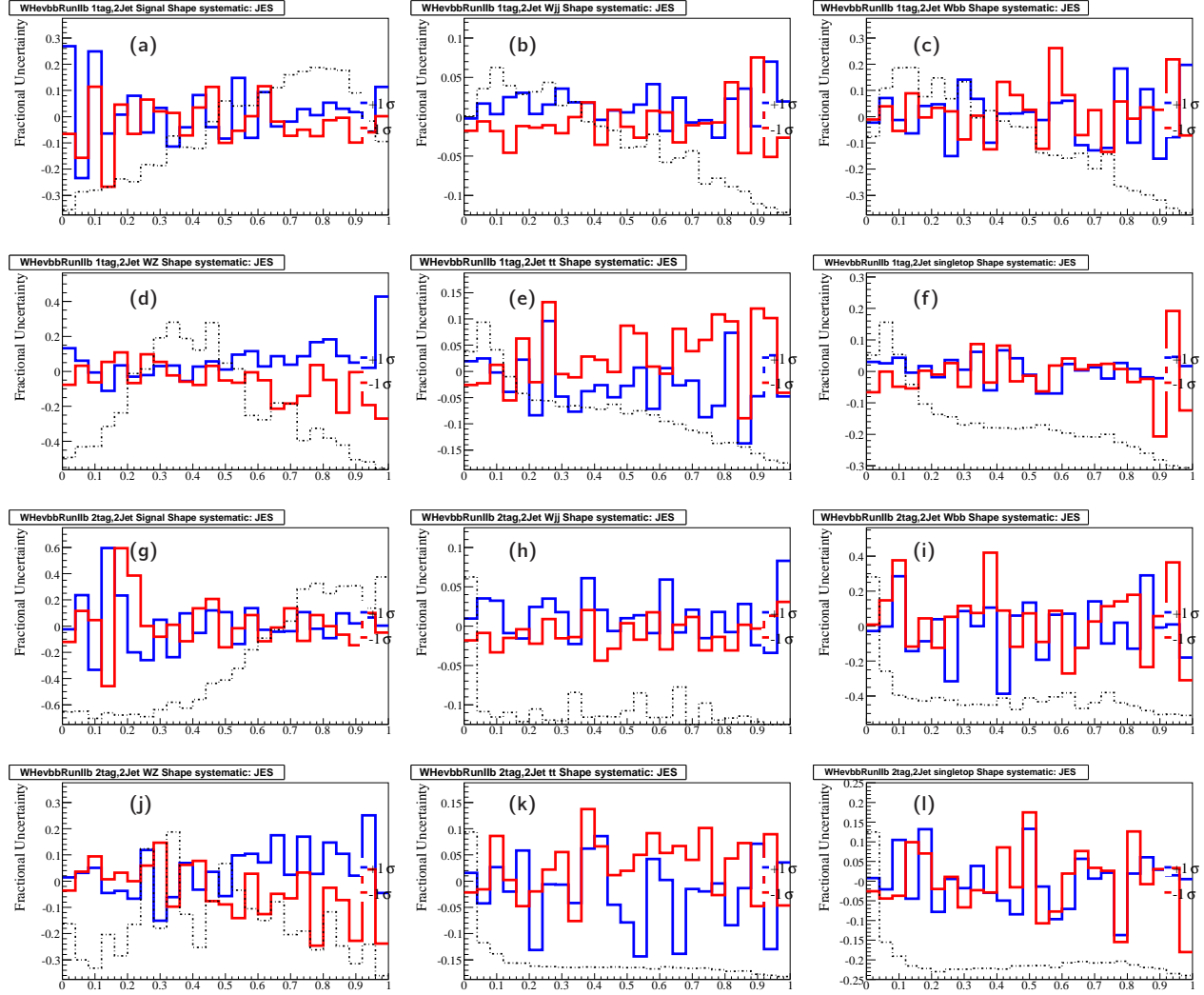


Figure F.1: **Electron Systematics (2-jet)** JES $\pm 1\sigma$ variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

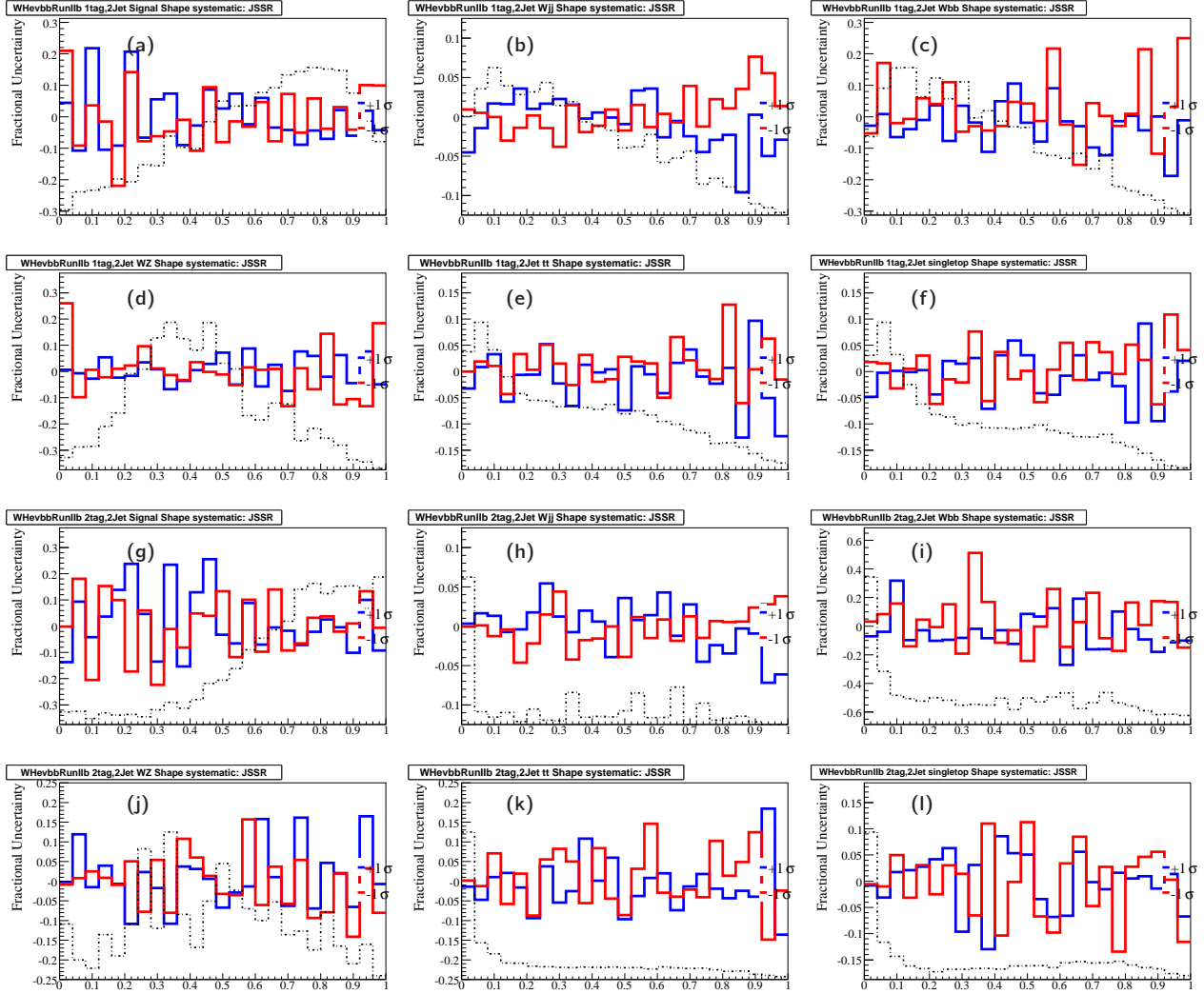


Figure F.2: **Electron Systematics (2-jet)** JSSR (res) $\pm 1\sigma$ variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) $W + \text{heavy flavor}$, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

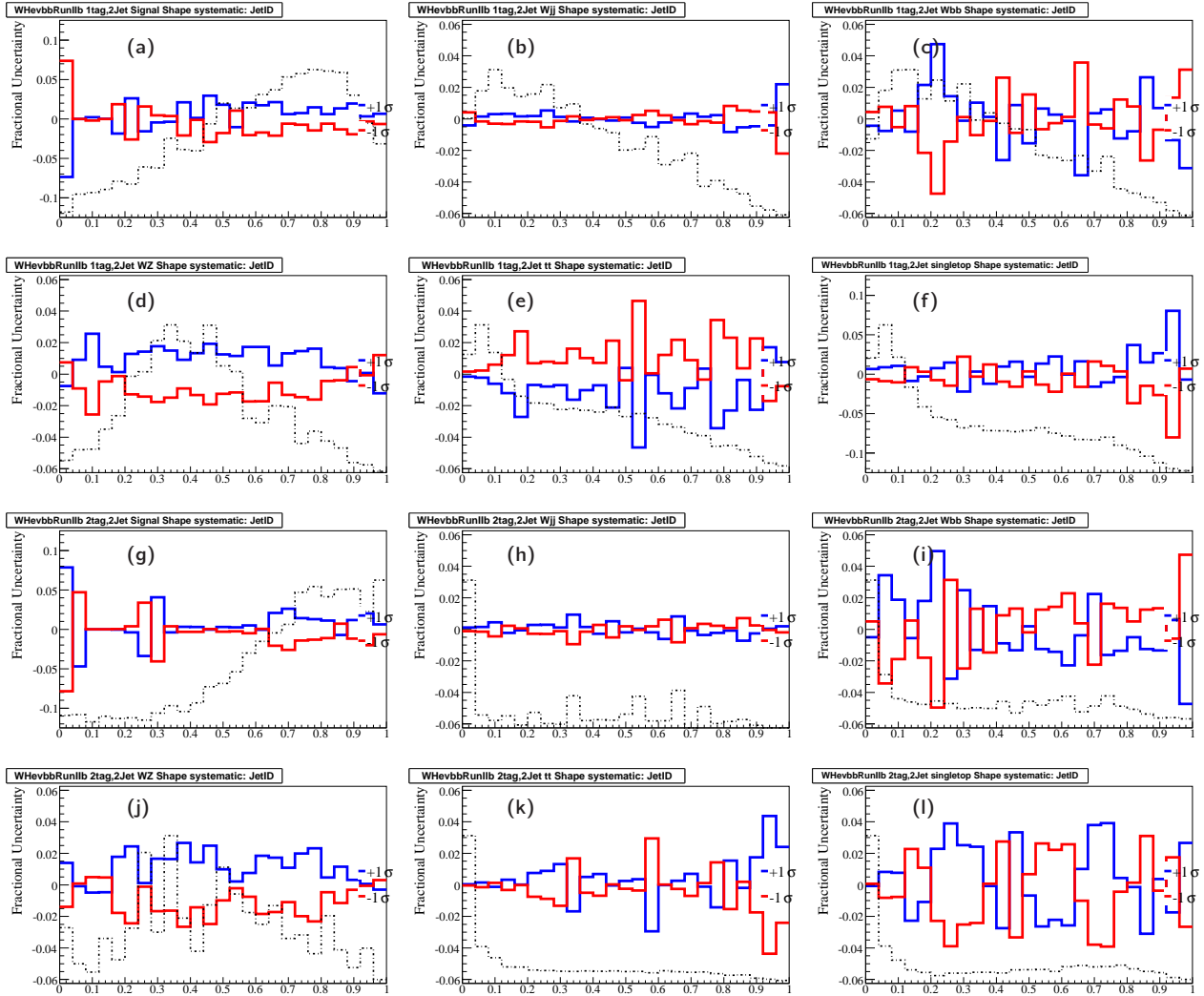


Figure F.3: **Electron Systematics (2-jet)** Jet-ID (efficiency) -1σ variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

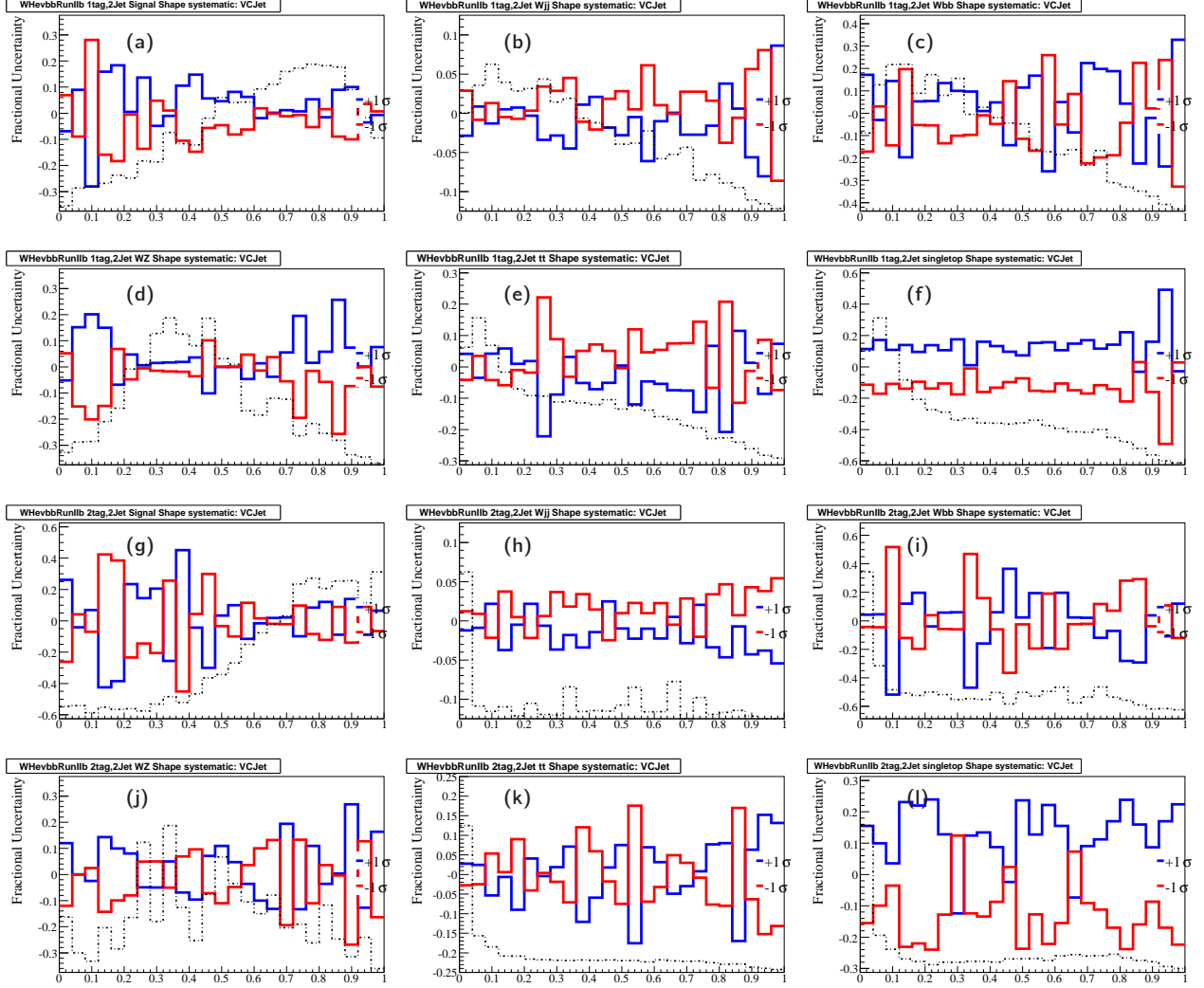


Figure F.4: **Electron Systematics (2-jet)** Jet Vertex Confirmation -1σ variation evaluated on the RF output ($+1\sigma$ variation taken as the symmetric to zero). Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) $W + \text{heavy flavor}$, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

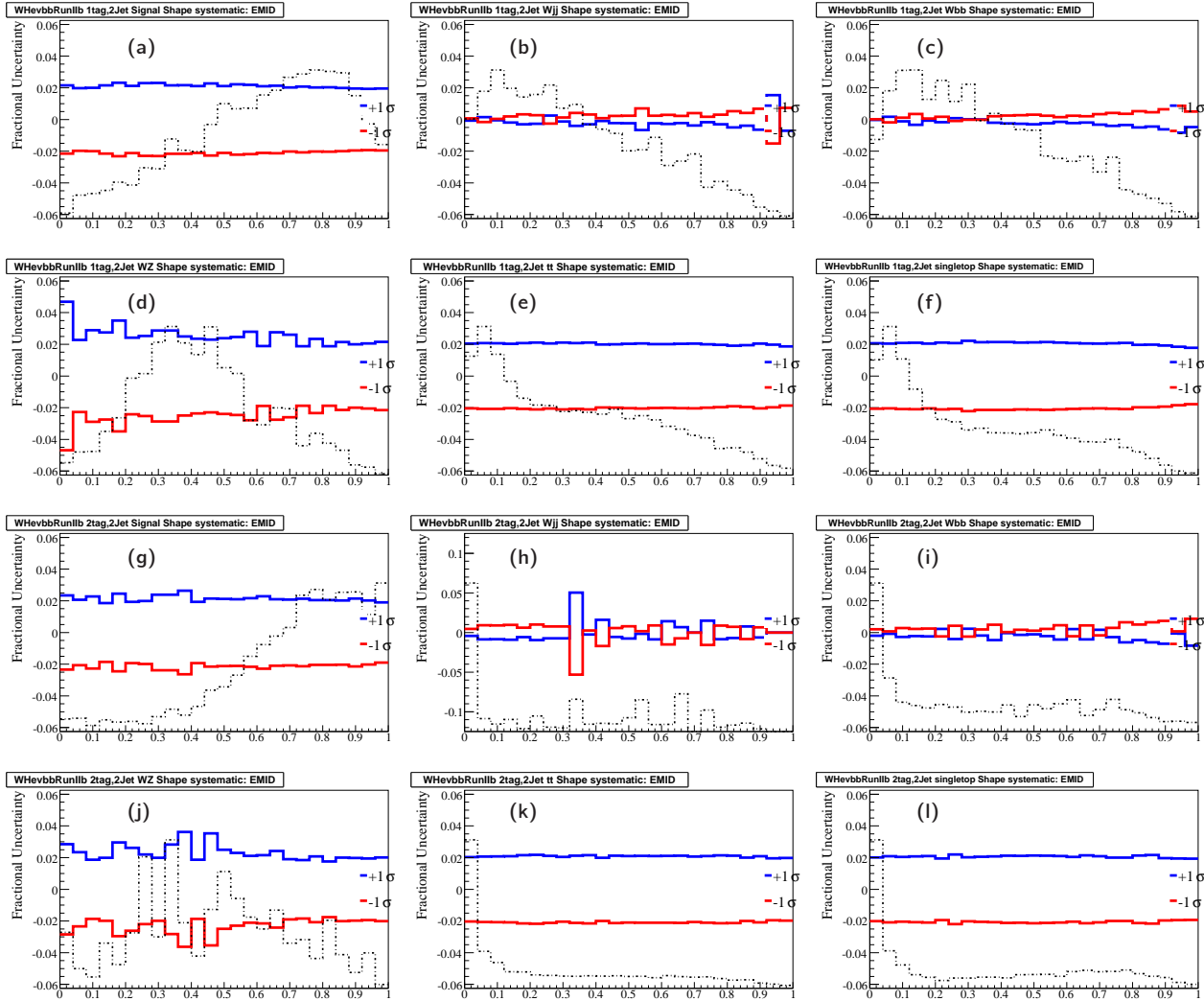


Figure F.5: **Electron Systematics (2-jet)** EMID $\pm 1\sigma$ variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) $W + \text{heavy flavor}$, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

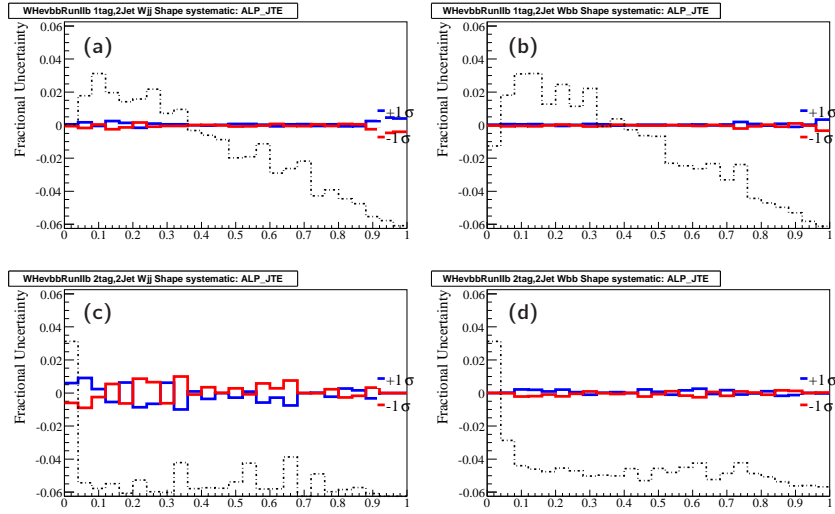


Figure F.6: **Electron Systematics (2-jet)** ALPGEN jet η reweighting $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

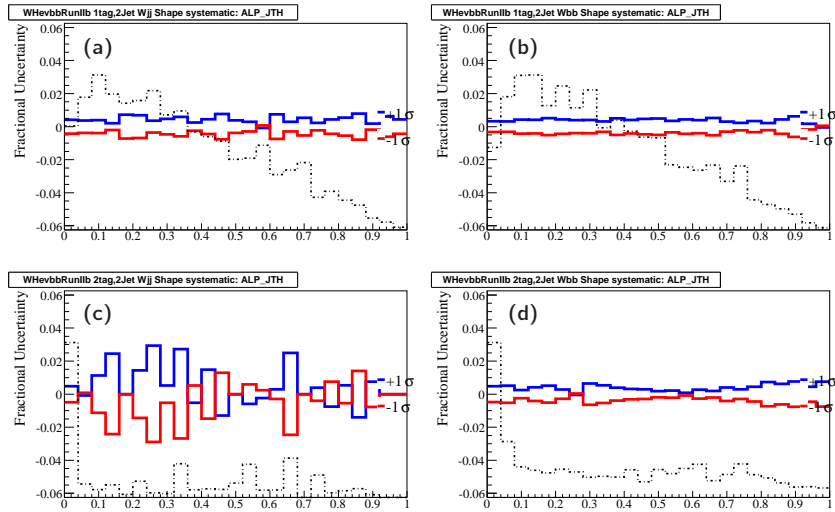


Figure F.7: **Electron Systematics (2-jet)** ALPGEN jet horns reweighting $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

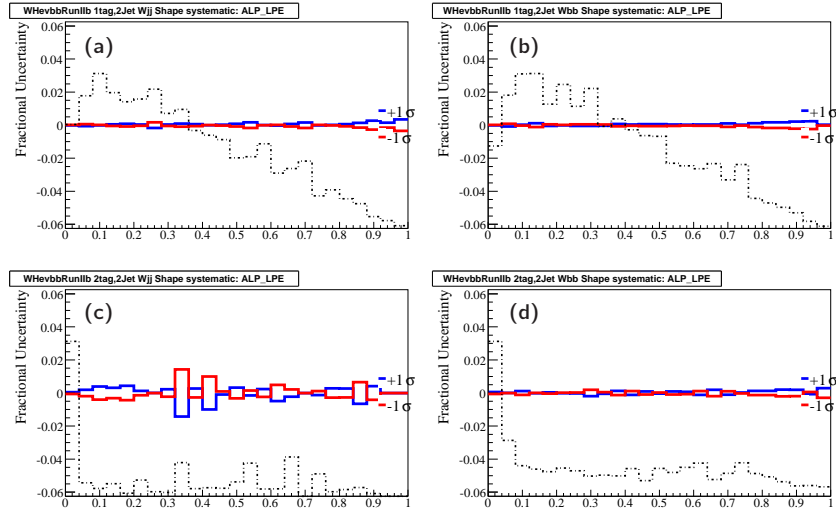


Figure F.8: **Electron Systematics (2-jet)** ALPGEN lepton η reweighting $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

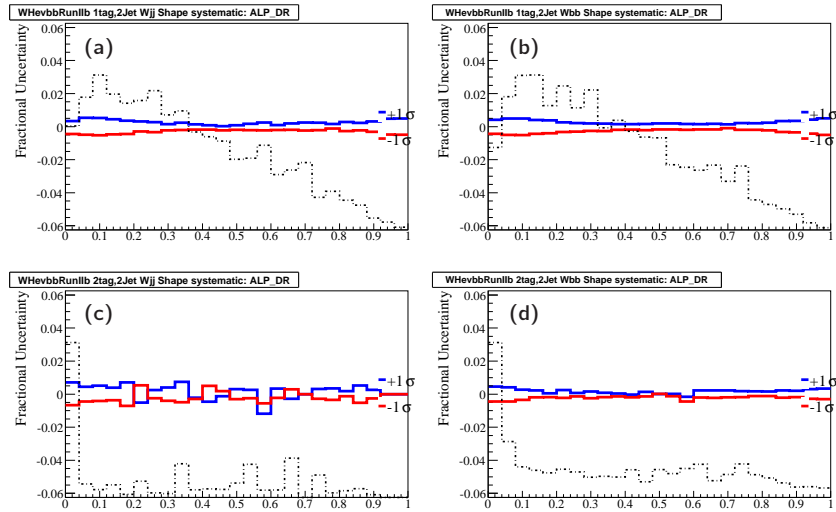


Figure F.9: **Electron Systematics (2-jet)** ALPGEN $\Delta R(j,j)$ reweighting $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

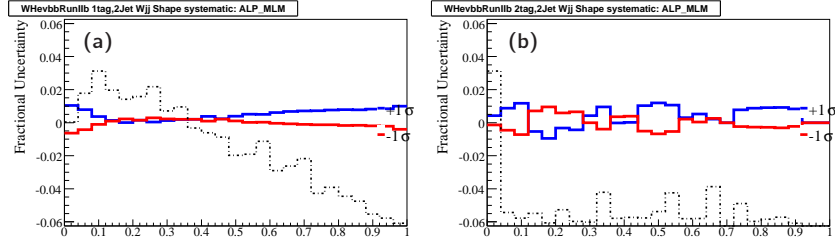


Figure F.10: **Electron Systematics (2-jet)** ALPGEN MLM $\pm 1\sigma$ variation for the RF output. Fig.a is for $W + jj$ single tag and Fig.b for $W + jj$ double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ variation in each sample.

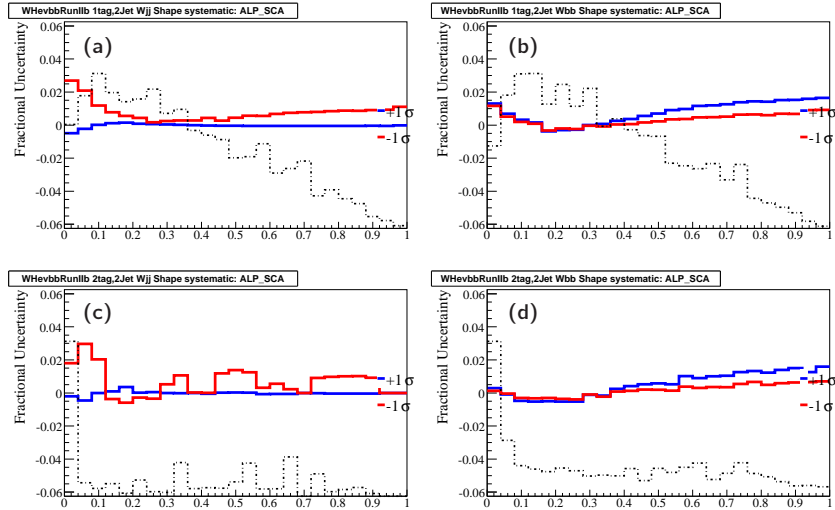


Figure F.11: **Electron Systematics (2-jet)** ALPGEN SCALE $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) $W + \text{heavy flavor}$.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

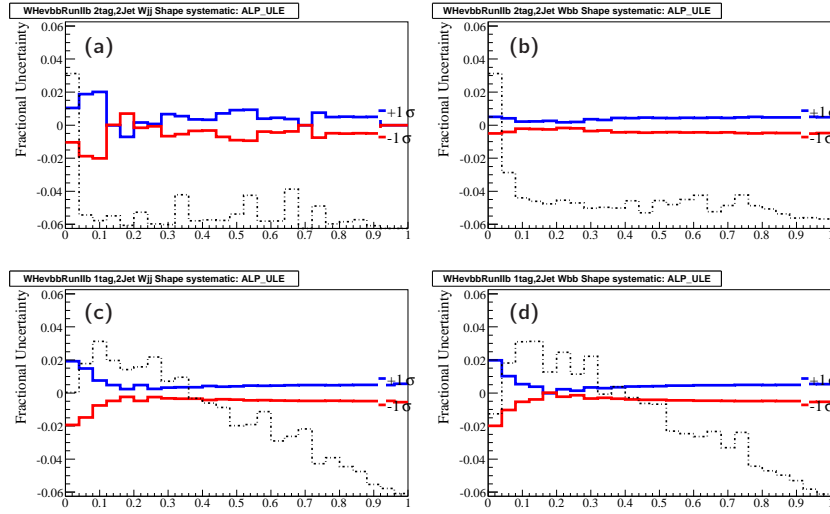


Figure F.12: **Electron Systematics (2-jet)** ALPGEN underlying event $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

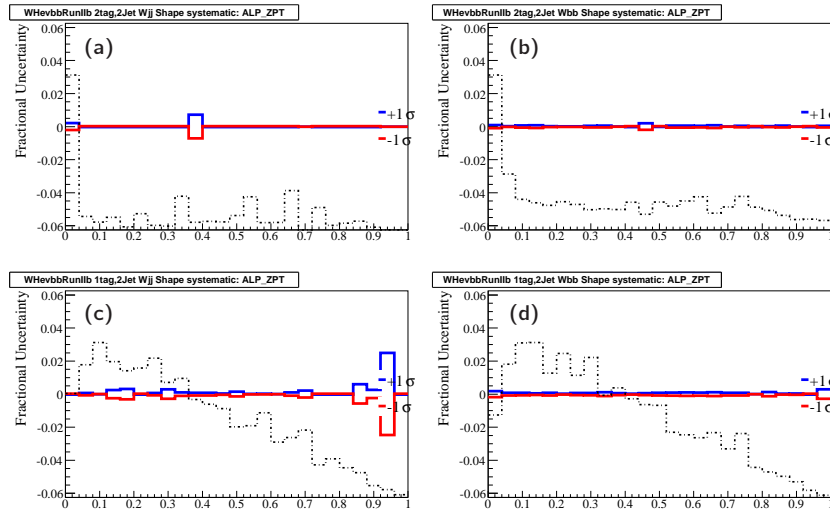


Figure F.13: **Electron Systematics (2-jet)** ALPGEN Zpt reweighting $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

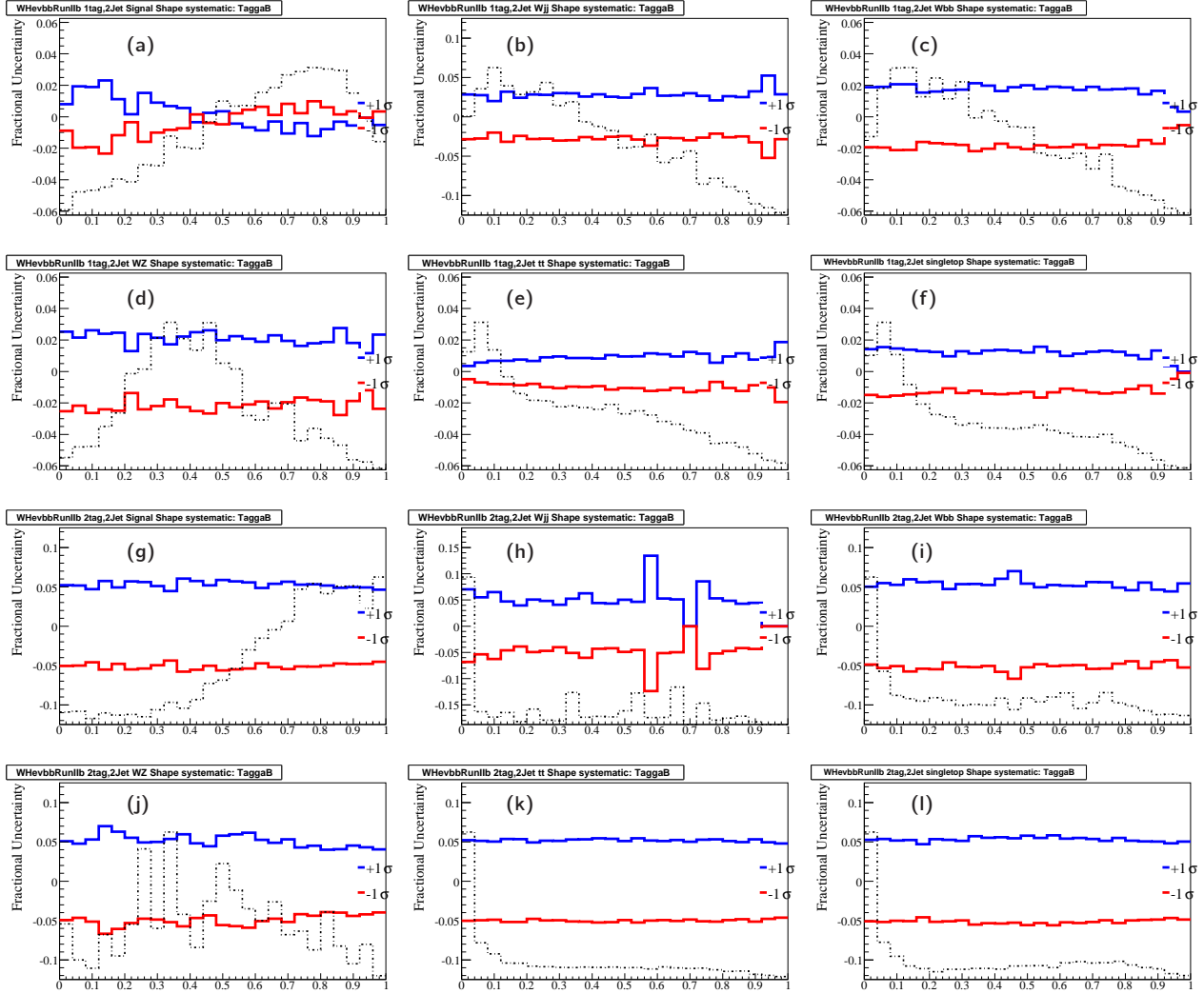


Figure F.14: **Electron Systematics (2-jet)** Taggability $\pm 1\sigma$ variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

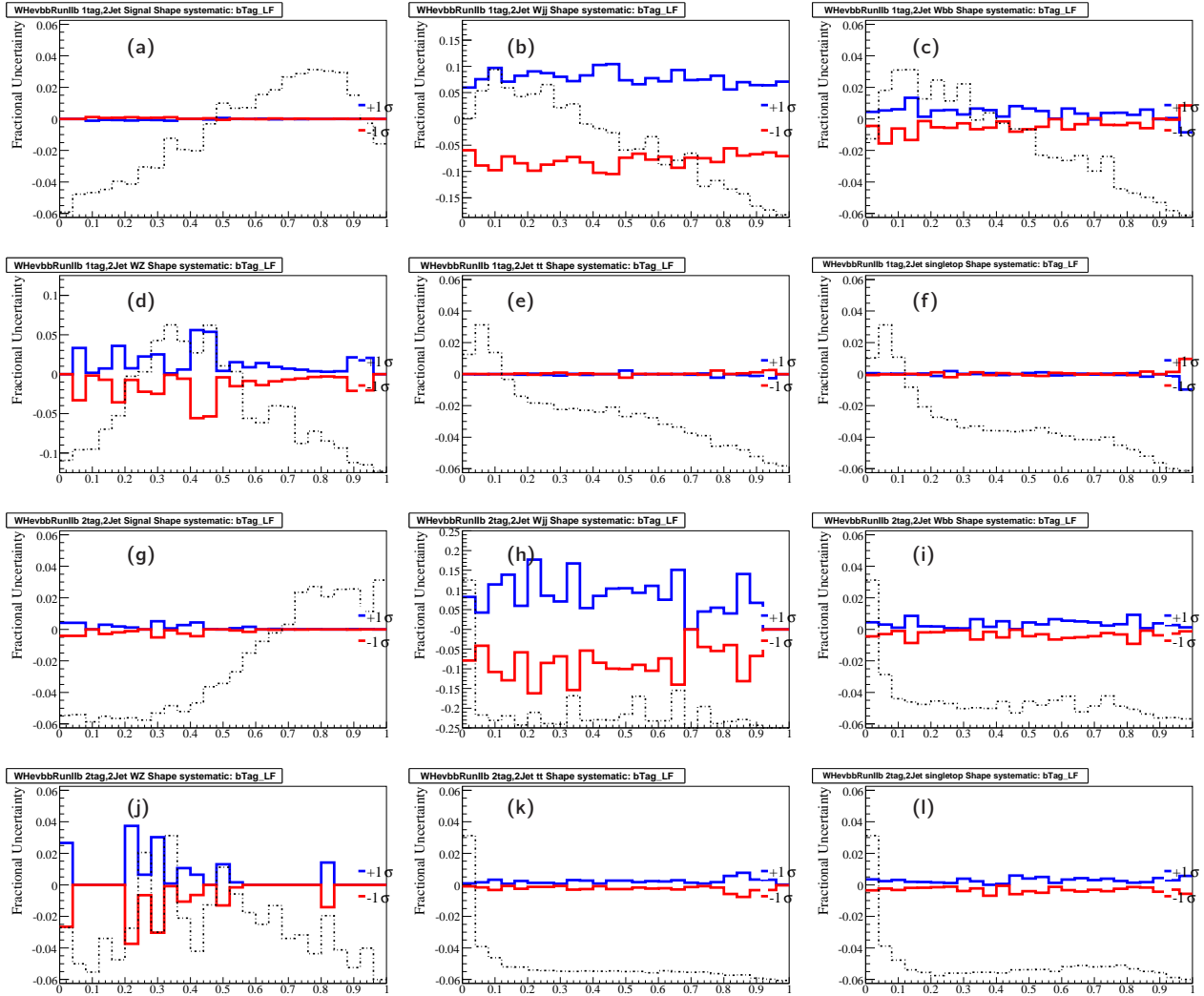


Figure F.15: **Electron Systematics (2-jet)** B-ID $\pm 1\sigma$ light jet variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

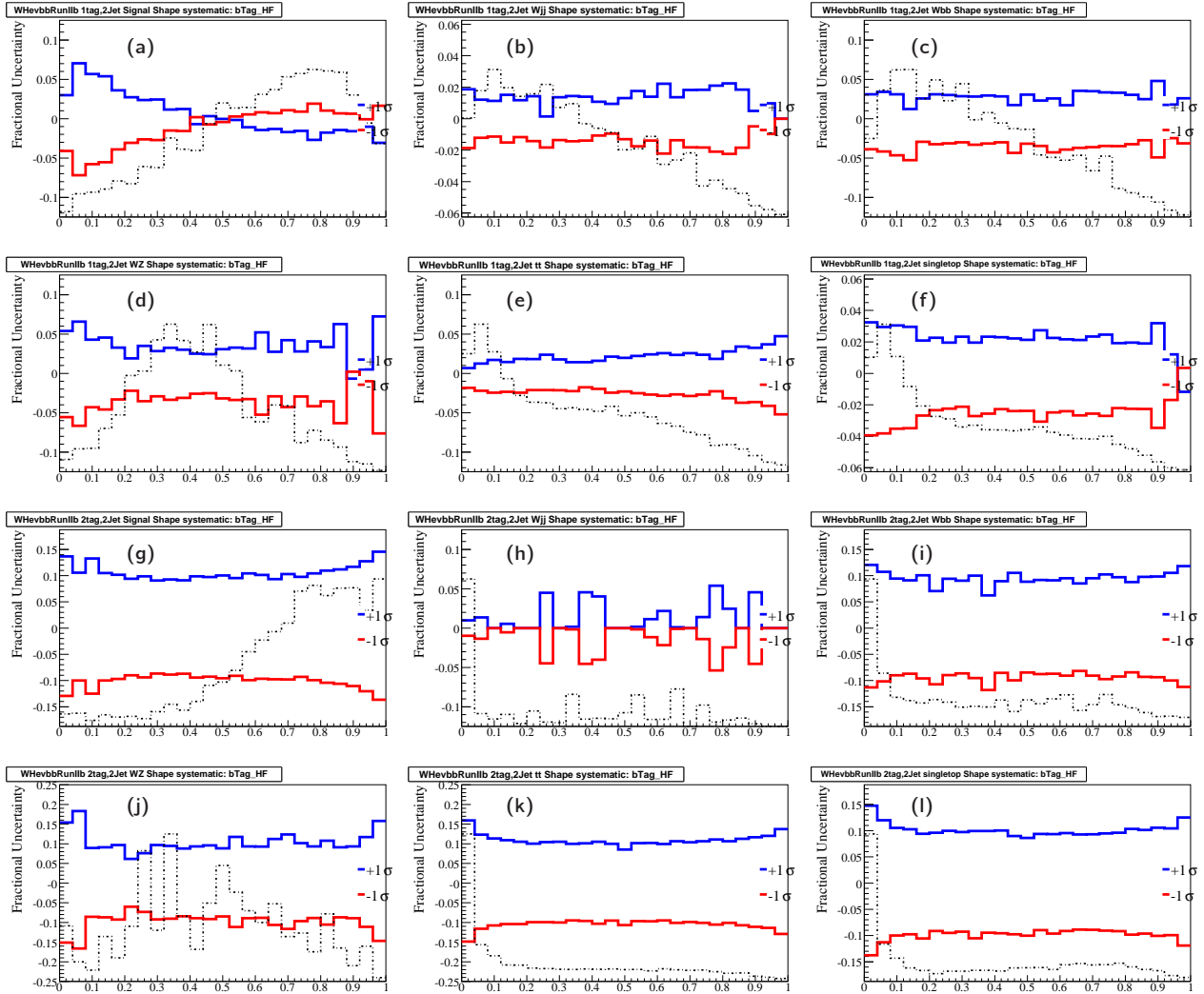


Figure F.16: **Electron Systematics (2-jet)** B-ID $\pm 1\sigma$ heavy jet variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

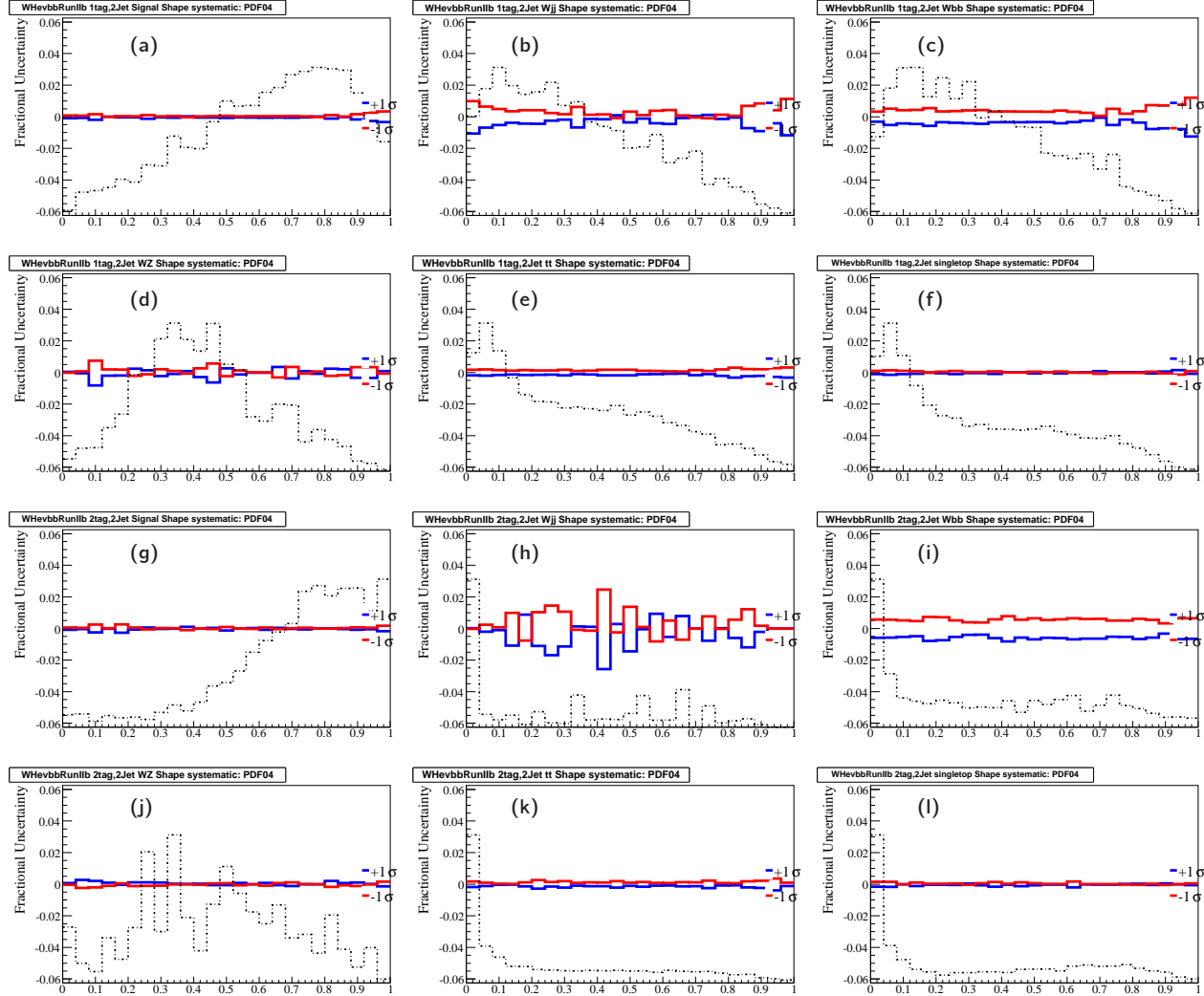


Figure F.17: **Electron Systematics (2-jet)** Variations of selected PDF uncertainties. A total of 40 PDF variations are made, each contributing to the final set of systematics. The 18th PDF variation for signal, the 15th PDF variation for BG (one of largest shape changes) are shown: Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

F.1. RUNIIB ELECTRON 2 JETS CHANNEL

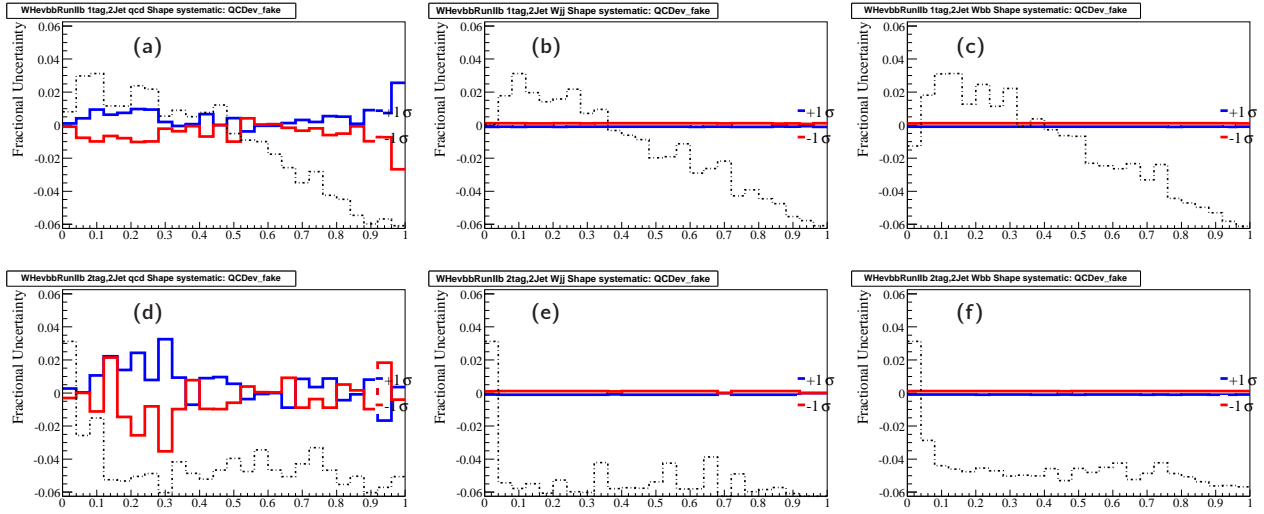


Figure F.18: **Electron Systematics (2-jet)** QCD Fake Rate $\pm 1\sigma$ variation for the RF output. Fig.a-c are for the single tag and d-f for the double tag sample. The red line shows $+1\sigma$, the blue line shows -1σ in each sample: Fig.(a,d) QCD, (b,e) $W + jj$, (c,f) $W + bb$.

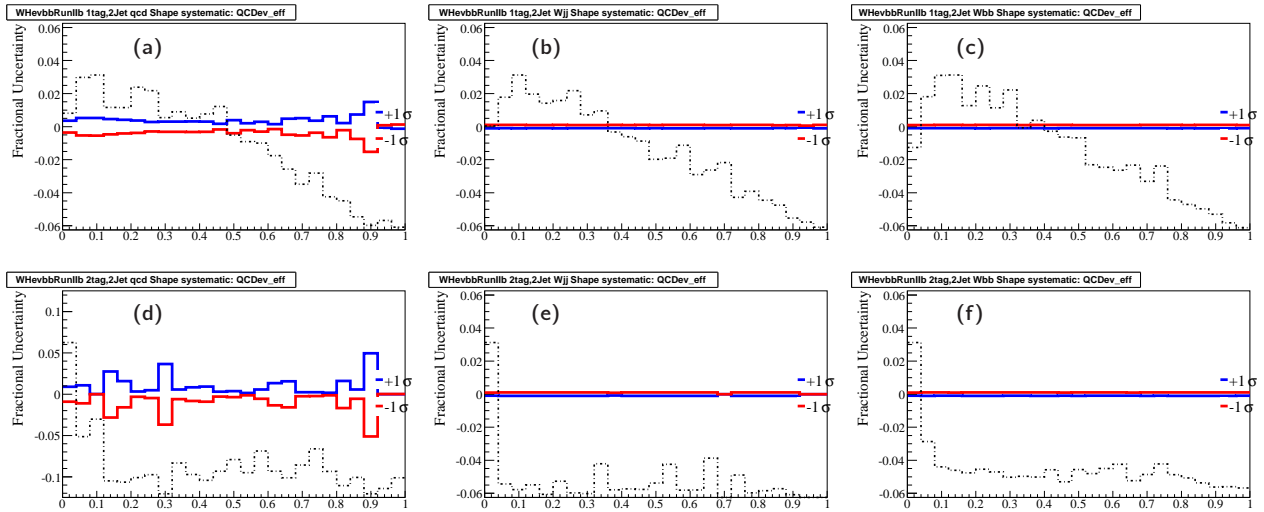


Figure F.19: **Electron Systematics (2-jet)** QCD electron efficiency $\pm 1\sigma$ variation for the RF output. Fig.a-c are for the single tag and d-f for the double tag sample. The red line shows $+1\sigma$, the blue line shows -1σ in each sample: Fig.(a,d) QCD, (b,e) $W + jj$, (c,f) $W + bb$.

F.2 RunI1b muon 2 jets channel

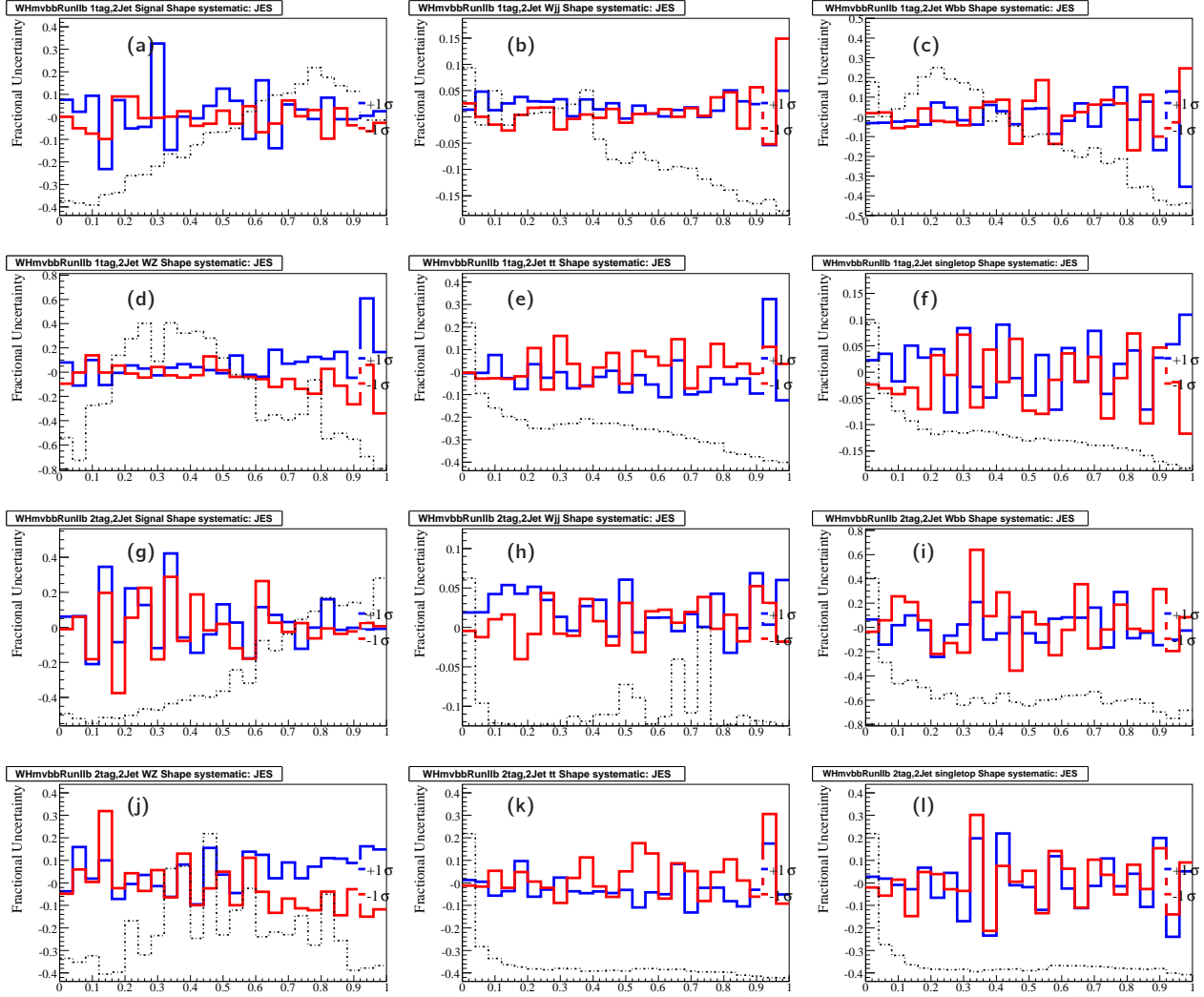


Figure F.20: **Muon Systematics (2-jet)** JES $\pm 1\sigma$ variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

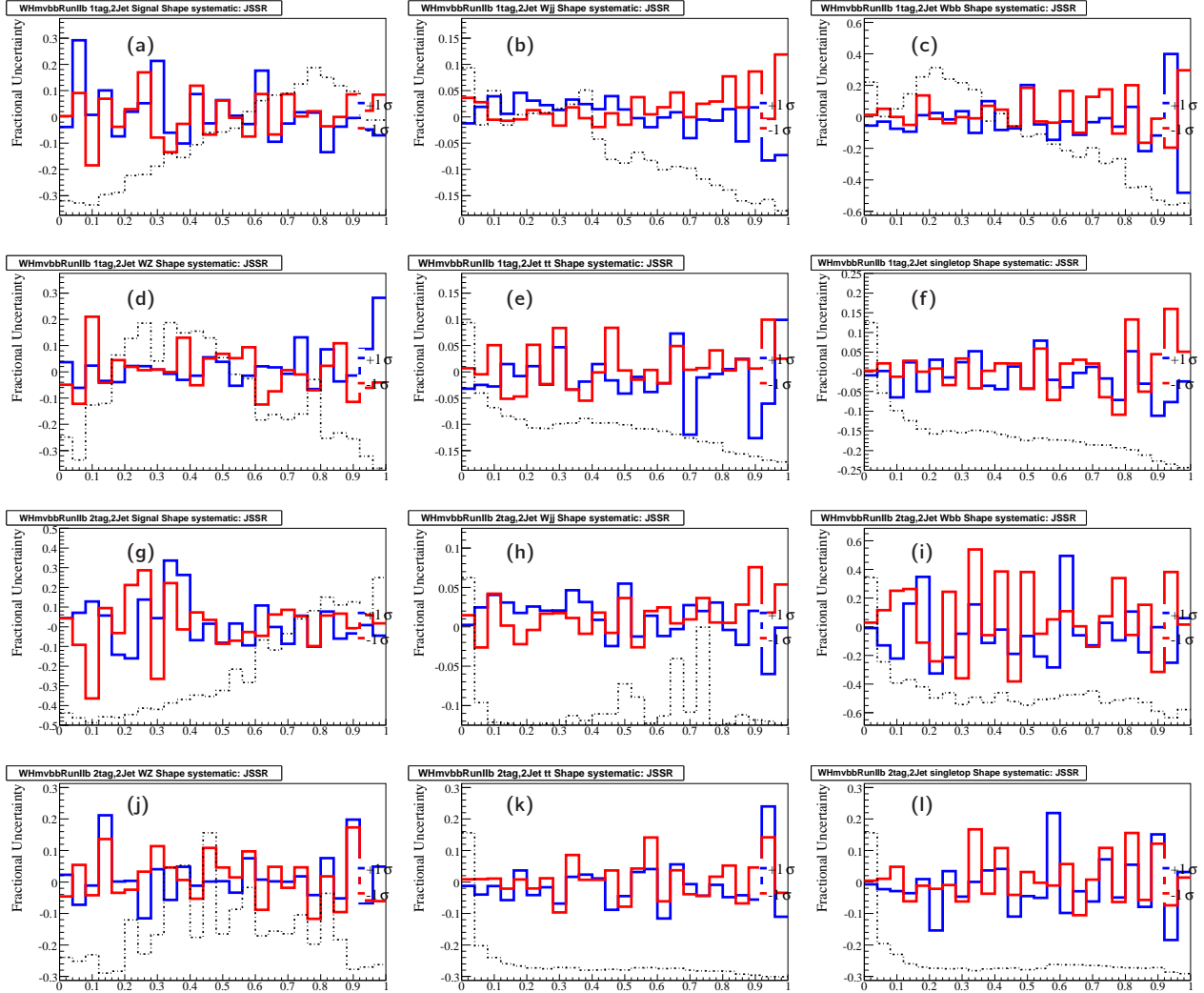


Figure F.21: **Muon Systematics (2-jet)** JSSR (res) $\pm 1\sigma$ variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

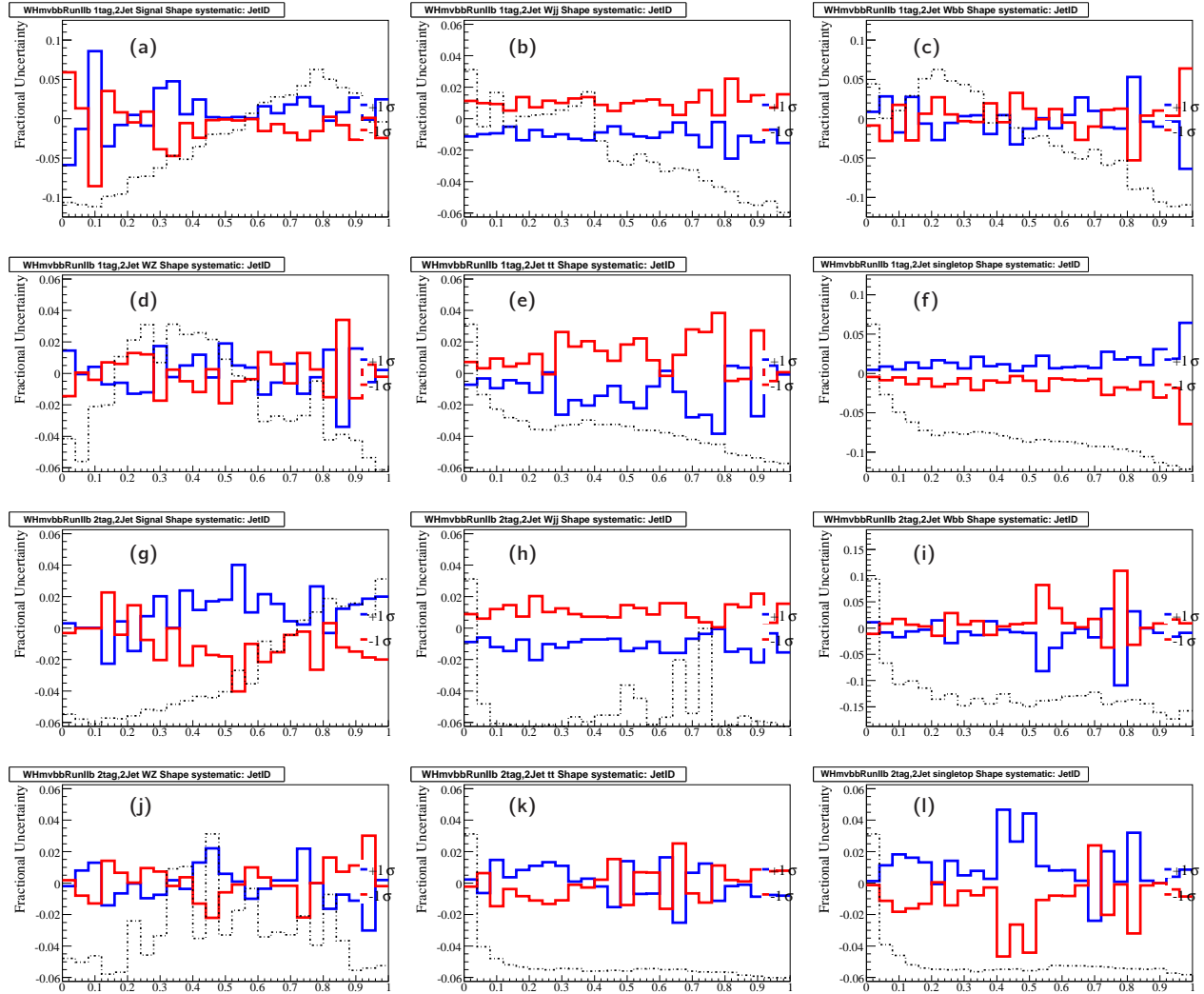


Figure F.22: **Muon Systematics (2-jet)** Jet-ID (efficiency) -1σ variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

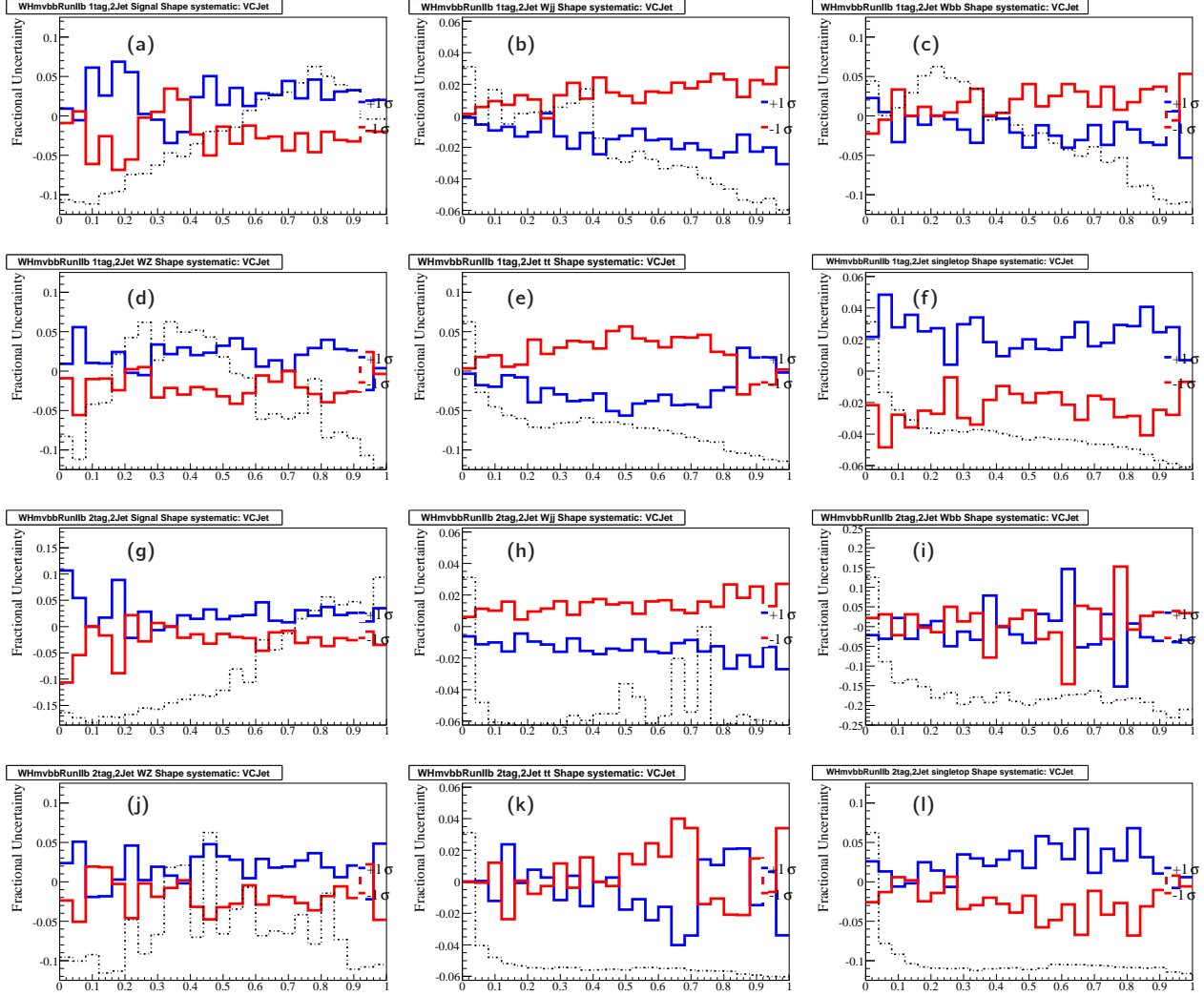


Figure F.23: **Muon Systematics (2-jet)** Jet Vertex Confirmation -1σ variation evaluated on the RF output ($+1\sigma$ variation taken as the symmetric to zero). Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) $W + \text{heavy flavor}$, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

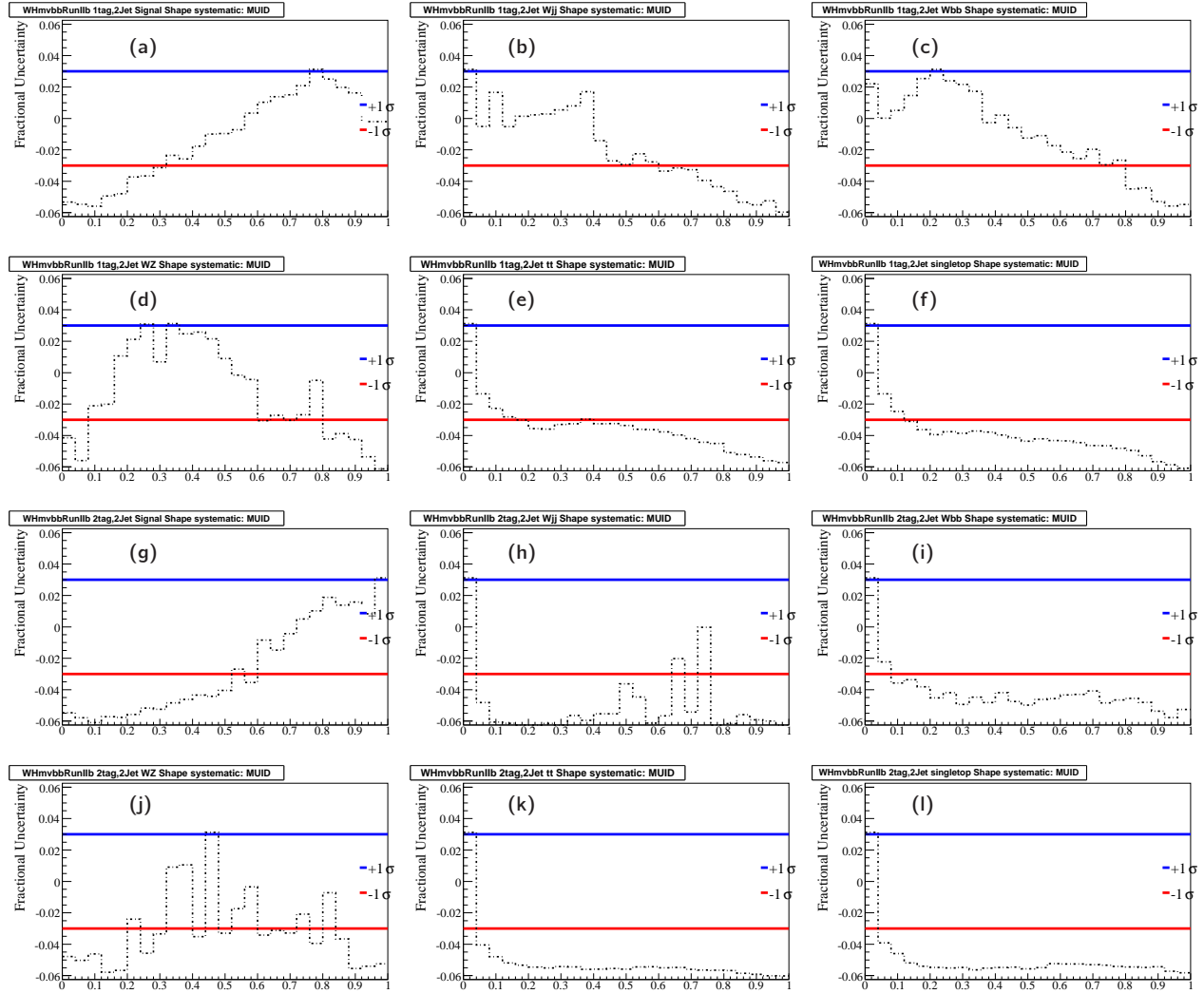


Figure F.24: **Muon Systematics (2-jet)** MUID $\pm 1\sigma$ variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

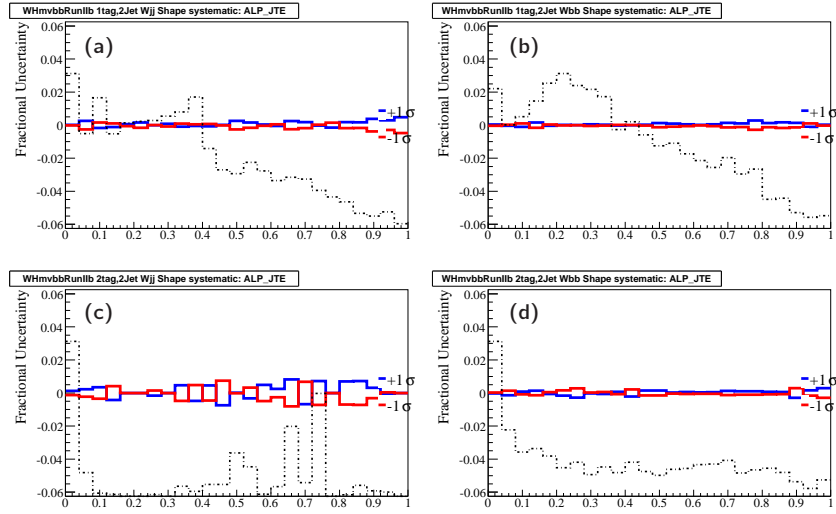


Figure F.25: **Muon Systematics (2-jet)** ALPGEN jet η reweighting $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

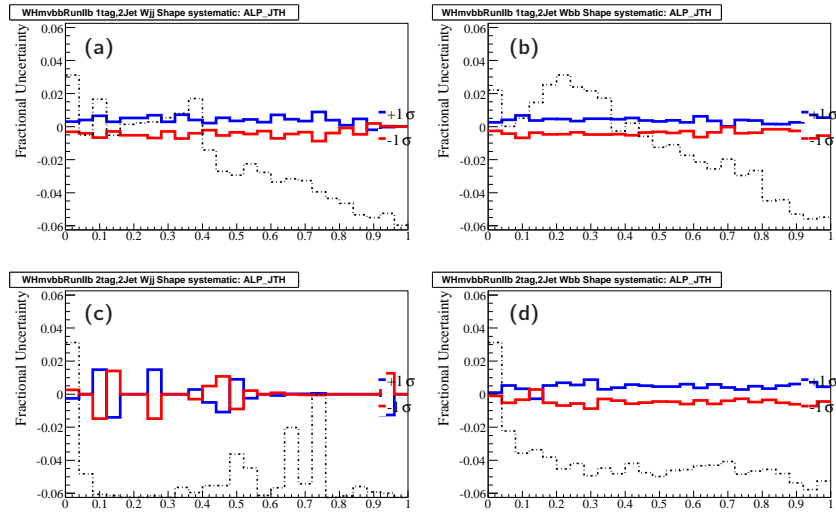


Figure F.26: **Muon Systematics (2-jet)** ALPGEN jet horns reweighting $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

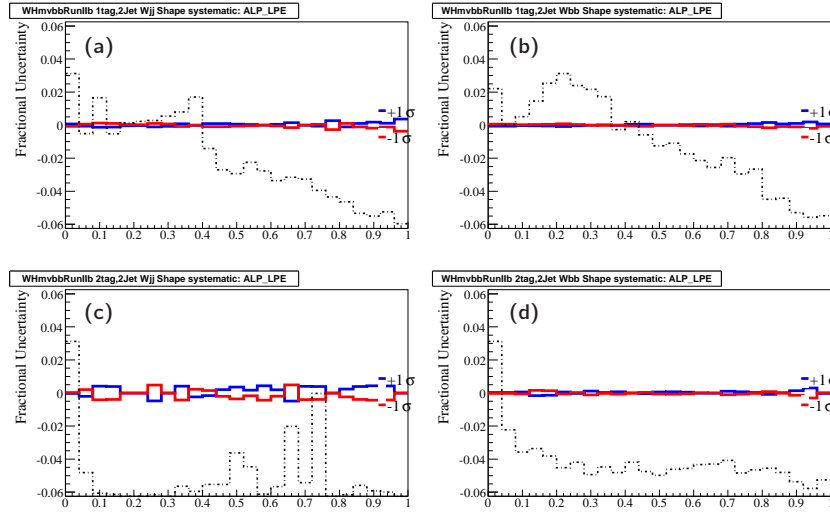


Figure F.27: **Muon Systematics (2-jet)** ALPGEN lepton η reweighting $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

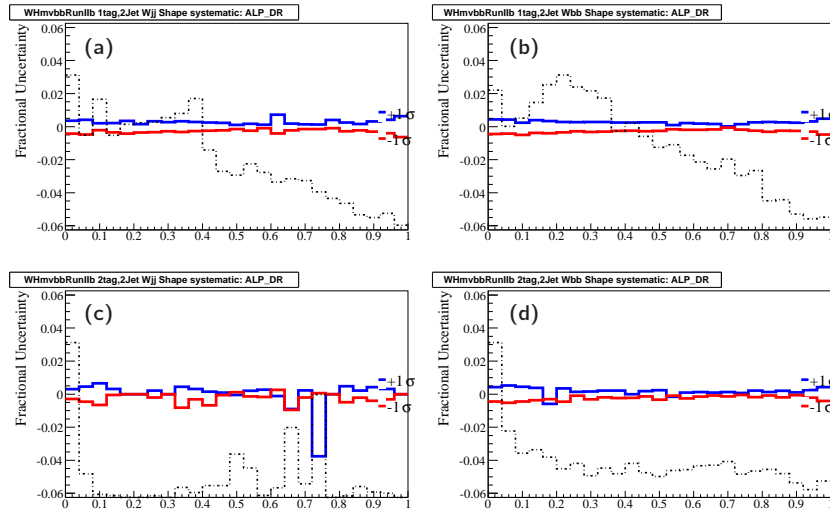


Figure F.28: **Muon Systematics (2-jet)** ALPGEN $\Delta R(j,j)$ reweighting $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

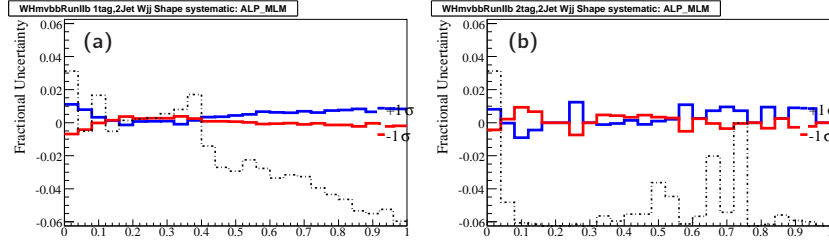


Figure F.29: **Muon Systematics (2-jet)** ALPGEN MLM $\pm 1\sigma$ variation for the RF output. Fig.a is for $W + jj$ single tag and Fig.b for $W + jj$ double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ variation in each sample.

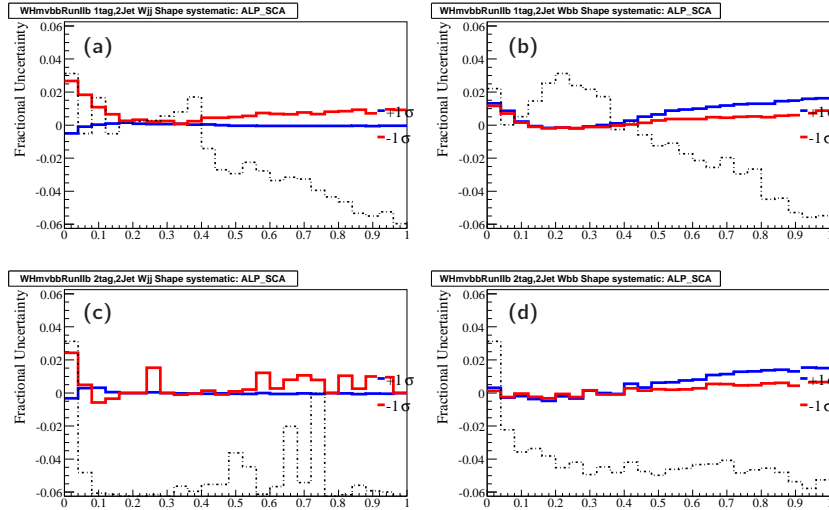


Figure F.30: **Muon Systematics (2-jet)** ALPGEN SCALE $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) $W + \text{heavy flavor}$.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

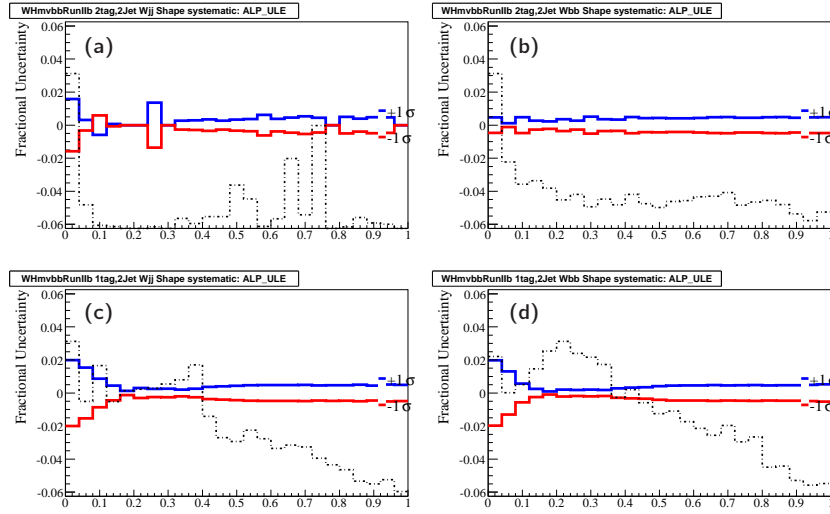


Figure F.31: **Muon Systematics (2-jet)** ALPGEN underlying event $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

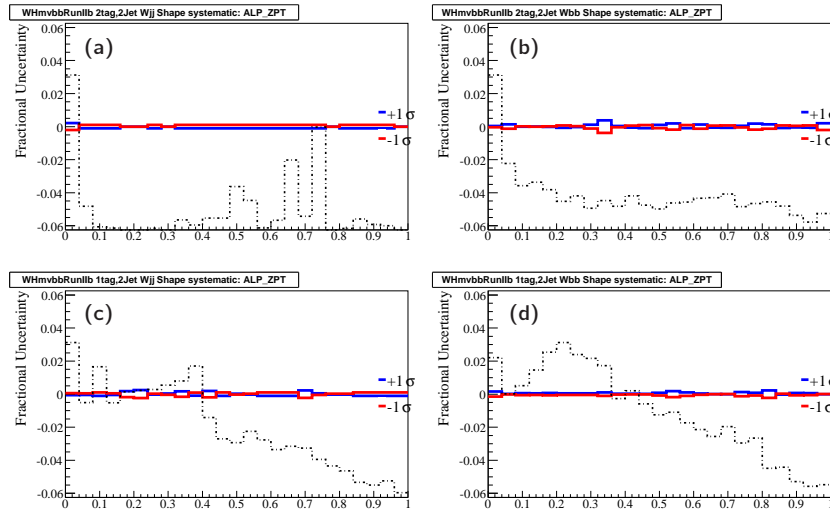


Figure F.32: **Muon Systematics (2-jet)** ALPGEN Zpt reweighting $\pm 1\sigma$ variation for the RF output. Fig. a-b are for the single tag and c-d for the double tag sample. The blue line shows the $+1\sigma$, the red line shows the -1σ in each sample: Fig.(a,c) $W + jj$, (b,d) W +heavy flavor.

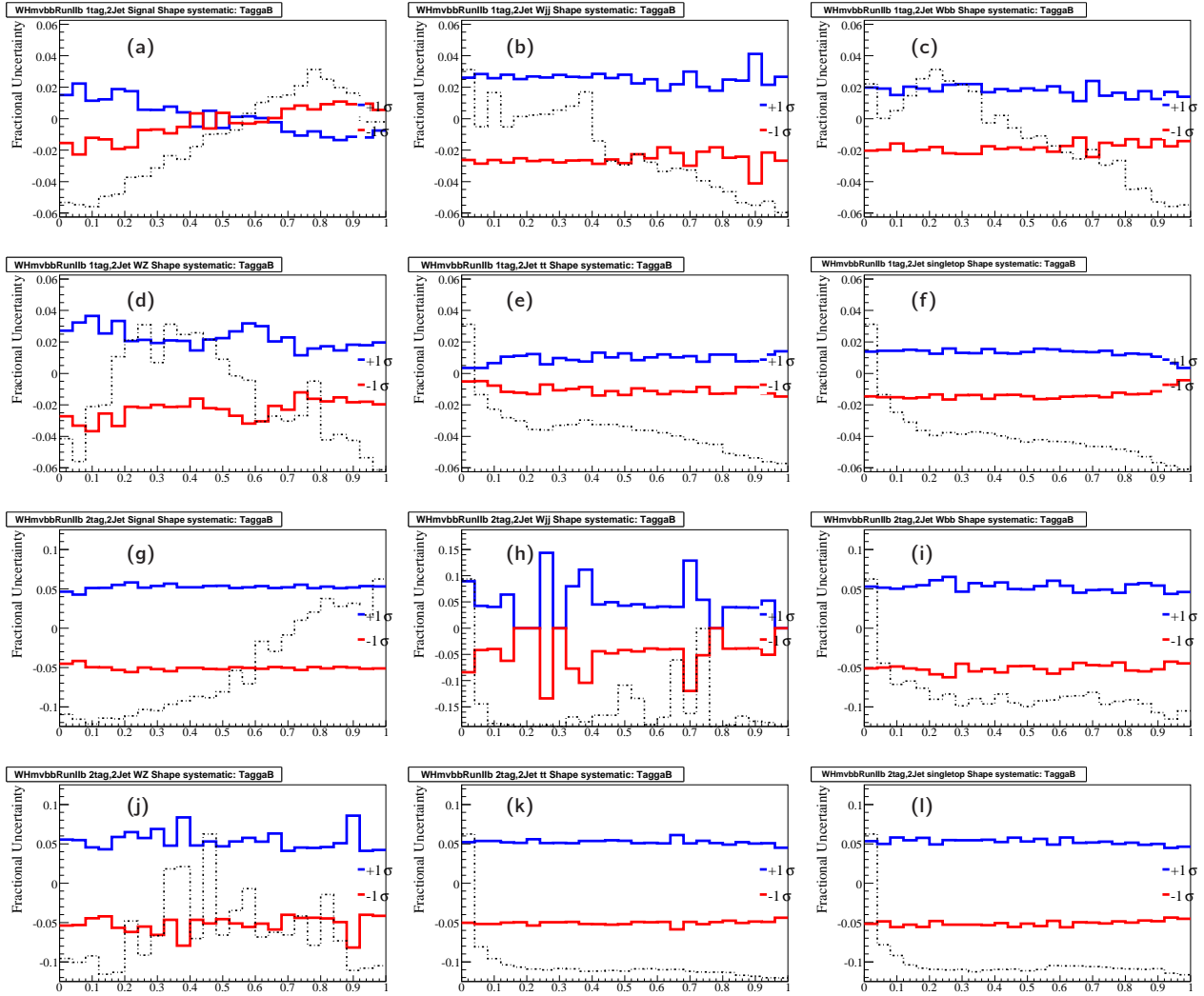


Figure F.33: **Muon Systematics (2-jet)** Taggability $\pm 1\sigma$ variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

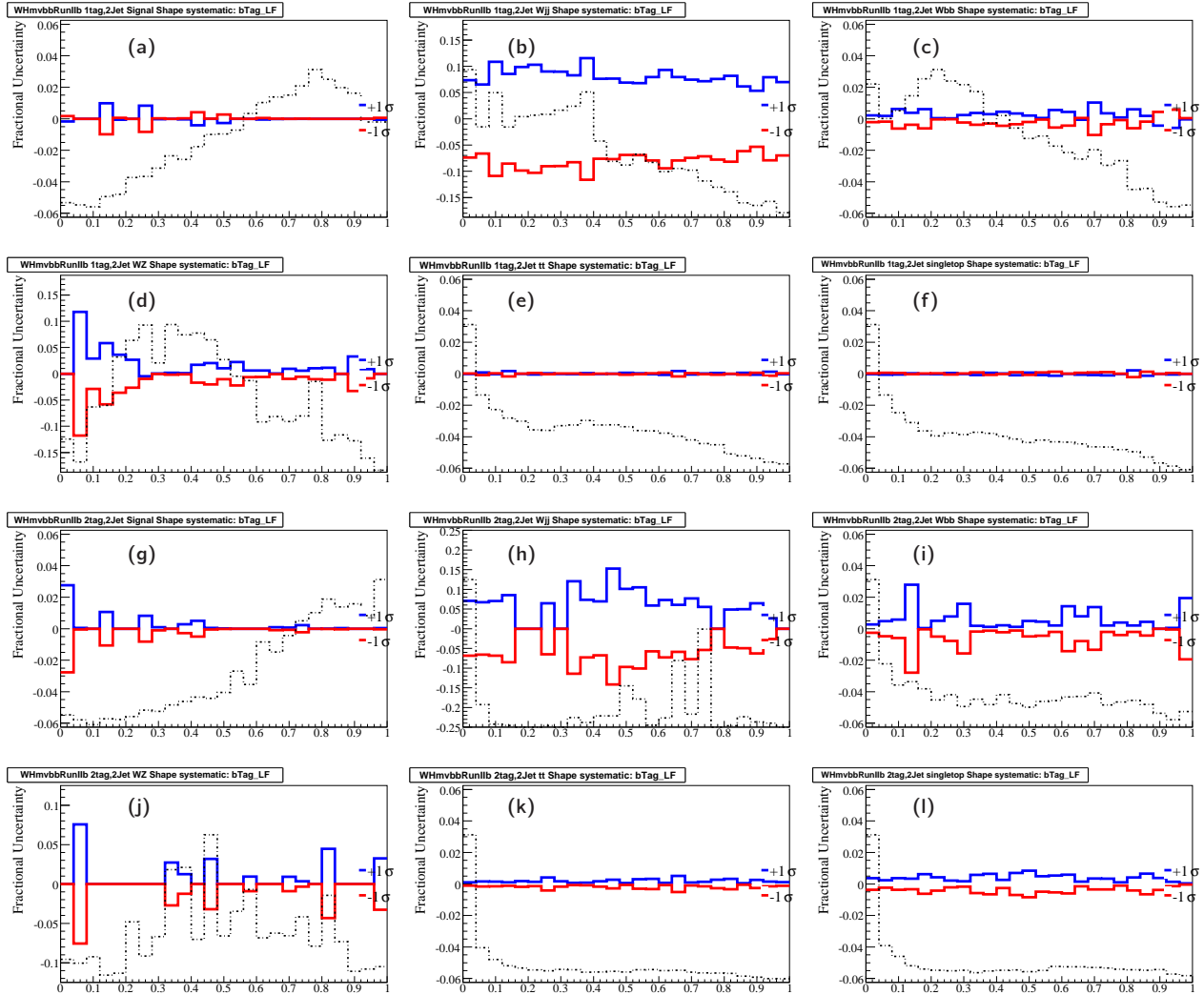


Figure F.34: **Muon Systematics (2-jet)** B-ID $\pm 1\sigma$ light jet variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

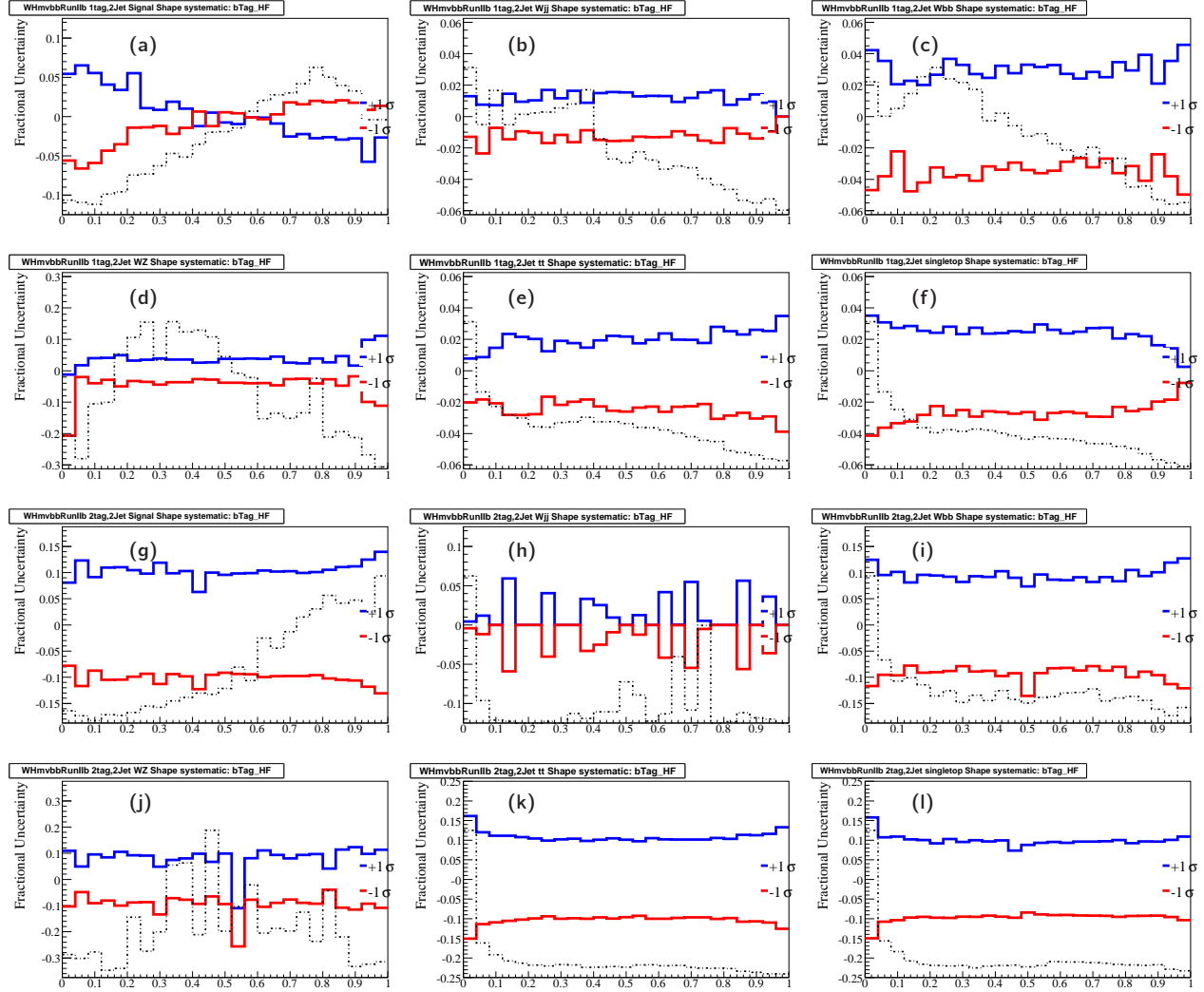


Figure F.35: **Muon Systematics (2-jet)** B-ID $\pm 1\sigma$ heavy jet variation evaluated on the RF output. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. The blue line shows the $+1\sigma$ variation, the red line shows -1σ variation in each sample: Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

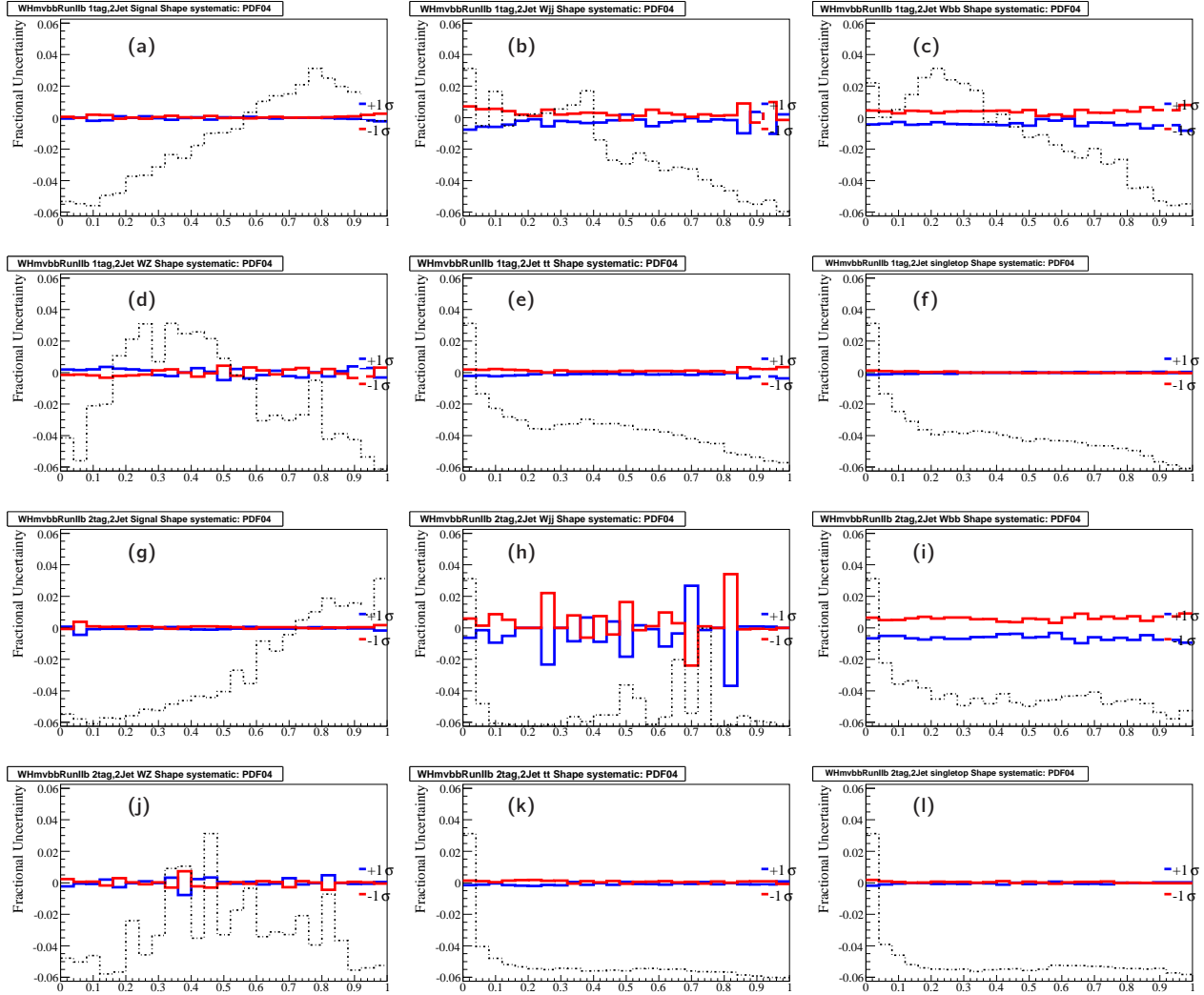


Figure F.36: **Muon Systematics (2-jet)** Variations of selected PDF uncertainties. A total of 40 PDF variations are made, each contributing to the final set of systematics. The 18th PDF variation for signal, the 15th PDF variation for BG (one of largest shape changes) are shown: Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

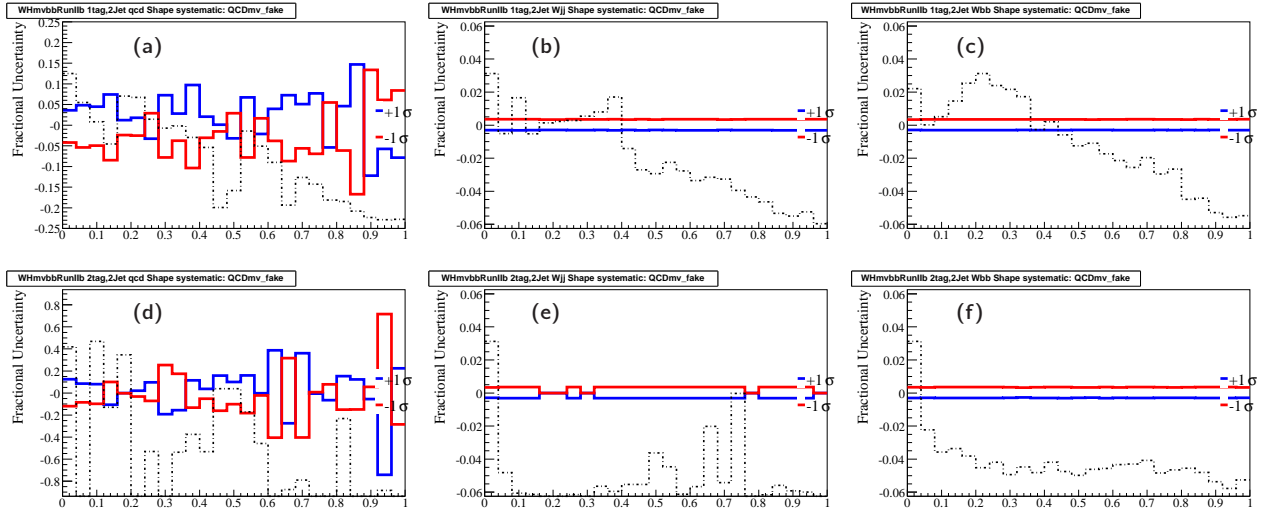


Figure F.37: **Muon Systematics (2-jet)** QCD Fake Rate $\pm 1\sigma$ variation for the RF output. Fig.a-c are for the single tag and d-f for the double tag sample. The red line shows $+1\sigma$, the blue line shows -1σ in each sample: Fig.(a,d) QCD, (b,e) $W + jj$, (c,f) $W+bb$.

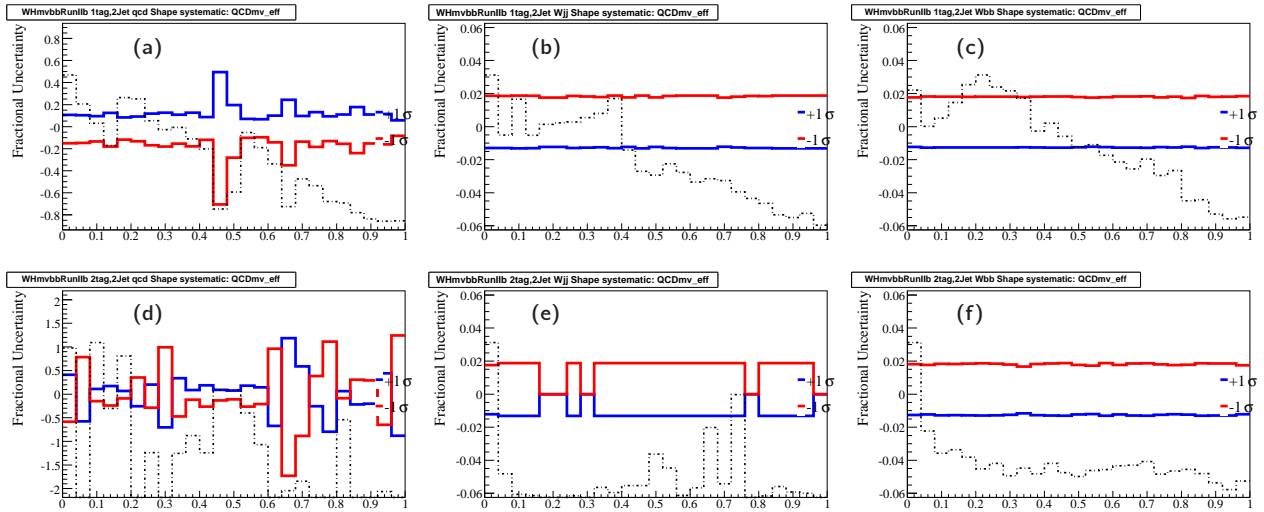


Figure F.38: **Muon Systematics (2-jet)** QCD muon efficiency $\pm 1\sigma$ variation for the RF output. Fig.a-c are for the single tag and d-f for the double tag sample. The red line shows $+1\sigma$, the blue line shows -1σ in each sample: Fig.(a,d) QCD, (b,e) $W + jj$, (c,f) $W+bb$.

APPENDIX F. SYSTEMATIC UNCERTAINTIES DISTRIBUTIONS

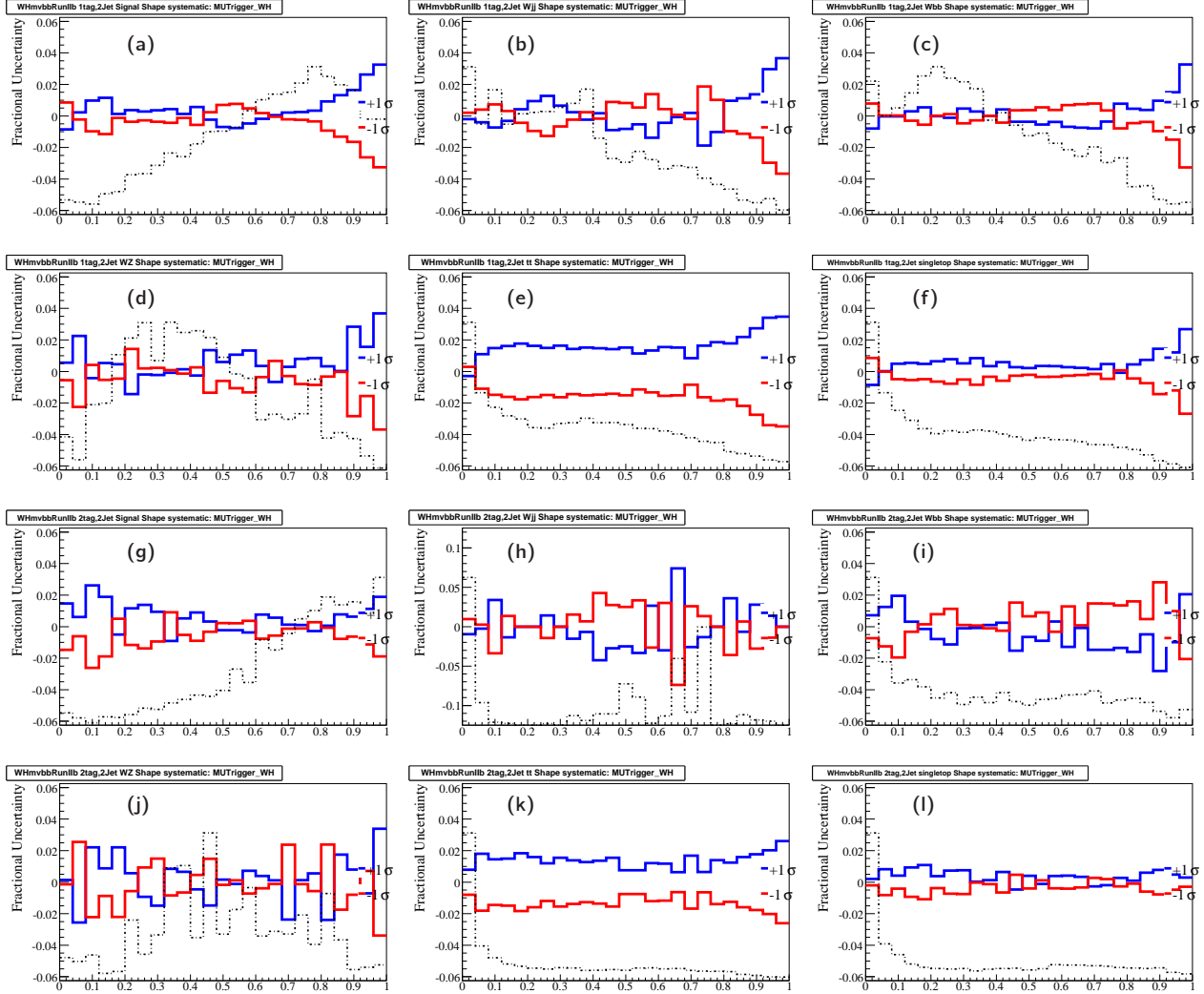


Figure F.39: **Muon Systematics (2-jet)** Variations of Muon trigger uncertainties. Fig. (a-f) are single tag samples, Fig. (g-l) are double tag samples. Fig.(a,g) WH , (b,h) $W + jj$, (c,i) W +heavy flavor, (d,j) Diboson, (e,k) $t\bar{t}$ and (f,l) single top.

Abstract

The Standard Model is the framework which allows to describe interactions between particles and their dynamics. The Higgs mechanism is a solution to naturally introduce a mass term in the theoretical description of this model. After electroweak spontaneous symmetry breaking, a new massive scalar particle is introduced, the Higgs boson.

Since it hasn't been discovered yet, the search for the Higgs boson is carried at the Tevatron, which is a $p\bar{p}$ collider at a center-of-mass of 1.96 TeV. For $M_H < 115$ GeV, the dominant decay mode is $H \rightarrow b\bar{b}$. The analysis presented in this document is focused on the $100 < M_H < 150$ GeV mass range, in the channel where the Higgs boson is produced in association with a W boson which decays either to an electron or muon and a neutrino. The study of this final state relies on informations collected from all parts of the DØ detector.

A result based on 5.3 fb^{-1} of RunII Tevatron collisions is presented here. In order to increase the sensitivity to the signal, the analysis is separated in different sub-channels according to the lepton flavour, number of jets in the final state, number of jets identified as originated from b quarks and data taking periods. After selecting events, a multivariate analysis technique is used to separate signal-like events from the expected physics and instrumental backgrounds. A good agreement between data and simulation is observed.

As no signal excess is observed in data, an observed (expected) upper limit of 4.5 (4.8) for $M_H = 115$ GeV is set on the ratio of the WH cross section multiplied by the $H \rightarrow b\bar{b}$ branching fraction to its standard model prediction, at 95% confidence level.

Since the final Tevatron dataset is soon to be analyzed, an effort is brought to achieve the maximum sensitivity. A preliminary analysis updated in Summer 2011 is presented as well as future improvements to be considered in the final publication for the search in the $WH \rightarrow \ell\nu b\bar{b}$ channel. Since the dijet invariant mass is the most discriminant single variable, a correction based on jet shapes is derived to improve the mass resolution, hence the analysis sensitivity. Moreover, a different usage of b -jets identification tools and optimization studies on the final discriminant yields a significant improvement. This result is part of the Tevatron Higgs combination.

Keywords

- Tevatron
- DØ
- Higgs Boson
- WH
- Multivariate analysis techniques
- Jet energy resolution

Résumé

Le Modèle Standard est le cadre qui permet de décrire les interactions entre les particules et leur dynamique. Le mécanisme de Higgs est une solution naturellement à introduire un terme de masse dans la description théorique de ce modèle. Après la rupture de symétrie électrofaible spontanée, une nouvelle particule scalaire massive est introduit, le boson de Higgs.

Comme il n'a pas encore été découverts, la recherche du boson de Higgs est effectué au Tevatron, qui est un collisionneur $p\bar{p}$ à une énergie au centre de masse de 1.96 TeV. Pour $M_H \lesssim 135$ GeV, le mode de désintégration dominant est $H \rightarrow b\bar{b}$. L'analyse présentée dans ce document est axé sur le domaine en masse $100 < M_H < 150$ GeV, dans le canal où le boson de Higgs est produit en association avec un boson W se désintégrant en un lepton chargé (électron ou muon) et un neutrino. L'étude de cet état final repose sur les informations recueillies auprès de toutes les parties du détecteur DØ.

Un résultat basé sur 5.3 fb^{-1} de collisions récoltées durant le RunII du Tevatron est présenté ici. Afin d'augmenter la sensibilité au signal, l'analyse est séparée en sous-canaux en fonction de la saveur du lepton considéré, le nombre de jets dans l'état final, le nombre de jets identifiés comme provenant de quarks b et par périodes de prise de données. Après sélection des événements, une technique d'analyse multivariée est utilisée pour séparer les événements de type signal du bruit de fond physique et instrumental attendu.

Comme un bon accord entre les données et simulation est observé, une limite supérieure observée (attendue) de 4.5 (4.8) est placée (pour $M_H = 115$ GeV) sur le rapport entre la section efficace $p\bar{p} \rightarrow WH$ multipliée par le rapport d'embranchement $H \rightarrow b\bar{b}$ et la prédiction du Modèle Standard, à 95% de niveau de confiance.

Alors que les dernière données du Tevatron sont en cours d'analyse, les perspectives de recherche au Tevatron sont exposées ici. Les améliorations futures pour la publication finale sur la recherche dans le canal $WH \rightarrow \ell\nu b\bar{b}$ sont présentées. La masse invariante dijet étant la variable la plus discriminante, l'amélioration de la résolution du pic de masse est un enjeu important, ainsi que l'optimisation de l'utilisation des outils d'identification de jets de quarks b et du discriminant final. Une amélioration significative est obtenue au niveau de la limite.

Mots-clés

- Tevatron
- DØ
- Boson de Higgs
- WH
- Techniques d'analyse multivariées
- Résolution en énergie des jets