



HAL
open science

**Recherche du boson de Higgs du Modèle Standard dans
le canal de désintégration $ZH \rightarrow \nu \nu b\bar{b}$ sur le
collisionneur Tevatron dans l'expérience D0.
Développement d'une méthode d'étiquetage des jets de
quark b avec des muons de basses impulsions transverses**
D. Jamin

► **To cite this version:**

D. Jamin. Recherche du boson de Higgs du Modèle Standard dans le canal de désintégration $ZH \rightarrow \nu \nu b\bar{b}$ sur le collisionneur Tevatron dans l'expérience D0. Développement d'une méthode d'étiquetage des jets de quark b avec des muons de basses impulsions transverses. Physique des Hautes Energies - Expérience [hep-ex]. Université de la Méditerranée - Aix-Marseille II, 2010. Français. NNT : . tel-00557839

HAL Id: tel-00557839

<https://theses.hal.science/tel-00557839v1>

Submitted on 20 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE LA MÉDITERRANÉE AIX-MARSEILLE II
FACULTÉ DES SCIENCES DE LUMINY
163 avenue de Luminy
13288 MARSEILLE Cedex 09

THESE DE DOCTORAT

Spécialité : Physique et Sciences de la Matière

Mention : Physique des Particules et Astroparticules

présenté par

David Jamin

en vue d'obtenir le grade de docteur de l'Université de la Méditerranée

**Recherche du boson de Higgs du Modèle Standard
dans le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$
sur le collisionneur Tevatron dans l'expérience DØ**

**Développement d'une méthode d'étiquetage
des jets de quarks b
avec des muons de basses impulsions transverses**

soutenue le 30 Septembre 2010 devant le jury composé de

Dr.	D. Bloch	Rapporteur
Dr.	A. Duperrin	Directeur de thèse
Prof.	F. Feinstein	Rapporteur
Dr.	J.F. Grivaz	Président du jury
Dr.	E. Kajfasz	Co-Directeur de thèse

UNIVERSITÉ DE LA MÉDITERRANÉE AIX-MARSEILLE II
FACULTÉ DES SCIENCES DE LUMINY
163 avenue de Luminy
13288 MARSEILLE Cedex 09

THESE DE DOCTORAT

Spécialité : Physique et Sciences de la Matière

Mention : Physique des Particules et Astroparticules

présenté par

David Jamin

en vue d'obtenir le grade de docteur de l'Université de la Méditerranée

**Recherche du boson de Higgs du Modèle Standard
dans le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$
sur le collisionneur Tevatron dans l'expérience DØ**

**Développement d'une méthode d'étiquetage
des jets de quarks b
avec des muons de basses impulsions transverses**

soutenue le 30 Septembre 2010 devant le jury composé de

Dr.	D. Bloch	Rapporteur
Dr.	A. Duperrin	Directeur de thèse
Prof.	F. Feinstein	Rapporteur
Dr.	J.F. Grivaz	Président du jury
Dr.	E. Kajfasz	Co-Directeur de thèse

Table des matières

Remerciements	1
Introduction	3
1 Cadre Théorique	5
1.1 Le Modèle Standard	5
1.2 La recherche du boson de Higgs	10
1.3 Conclusion	18
2 Cadre Expérimental	19
2.1 Le système d'accélération	19
2.2 Le détecteur DØ	22
2.3 Objets physiques utilisés	33
2.4 La chaîne de simulation	38
2.5 Qualité des données	38
3 Soft Lepton Tagging	41
3.1 Préambule : NN	41
3.2 Motivations de la méthode SLTNN	47
3.3 Procédure	49
3.4 Etiquetage de muons dans les jets	50
3.5 Entraînement du Réseau de Neurone	58
3.6 Résultats avec les données	65
3.7 Conclusion	81
4 Recherche du boson de Higgs	83
4.1 Lot de données	84
4.2 Bruits de fond	85
4.3 Simulation du signal	92

4.4	Présélection	92
4.5	Sélection	93
4.6	Étiquetage des jets issus de quark b avec NN	105
4.7	Étiquetage des jets issus de quark b avec SLTNN	111
4.8	Conclusion	130
Conclusion		133
A Réseau de neurones		135
A.1	Structure du réseau de neurones	135
A.2	Apprentissage du réseau de neurones	137
A.3	Arbres de décision	138
B Distributions supplémentaires du développement de la méthode d'étiquetage SLTNN		141
B.1	Coefficients de corrélation	141
B.2	Variables d'entrées de SLTNN dans les lots de bruit de fond simulés	147
B.3	Variables d'entrées de SLTNN dans les lots de données et de signal simulé	149
B.4	Variables d'entrées de SLTNN dans les lots de données et de signal plus bruit de fond simulé	151
C Autres combinaisons d'étiquetage dans l'analyse $ZH \rightarrow \nu\bar{\nu}b\bar{b}$		153
D Distributions des variables d'entrées du SLTNN dans l'analyse $ZH \rightarrow \nu\bar{\nu}b\bar{b}$		157
D.1	Variables d'entrées de SLTNN dans l'analyse avant <i>b-tagging</i>	157
D.2	Variables d'entrées de SLTNN dans l'analyse après <i>b-tagging</i>	160
Bibliographie		163

Remerciements

Je remercie tout d'abord Eric Kajfasz pour m'avoir accueilli au CPPM et pour sa participation au jury de thèse.

Je voudrais remercier particulièrement Arnaud Duperrin, mon directeur de thèse, et ce pour plusieurs raisons. Toute cette histoire a commencé suite à un entretien téléphonique lors de ma recherche de stage de fin de M2 où nous avons eu un échange sincère dans une période difficile pour moi. Ensuite, je n'oublierai jamais les efforts consentis pour me permettre d'obtenir une bourse de thèse et me permettre d'accomplir tout ce que l'on pourra lire par la suite dans ce mémoire. Merci de m'avoir soutenu durant ces trois dernières années et d'avoir toujours défendu mes intérêts.

Je souhaite remercier Daniel Bloch et Fabrice Feinstein, mes rapporteurs, pour la lecture de ce manuscrit et les commentaires qui ont contribué à améliorer ce document. Je tiens à remercier aussi Jean-François Grivaz d'avoir accepté de participer et de présider mon jury de thèse, ainsi que de l'aide et de l'intérêt apportés à mes travaux en tant que collaborateur.

J'aimerais aussi remercier les membres du groupe DØ à Marseille : Anne-Fleur Barfuss, Betty Calpas, Samuel Calvet, Marie-Claude Cousinou, Marc Escalier, Smaïn Kermiche, Steve Muanza et Elemer Nagy qui ont, tout au long de ces trois ans et demi, partagé de leur temps et de leur expérience.

Je remercie aussi les thésards et post-docs DØ France et de la collaboration, pour les moments agréables et enrichissants partagés avec eux lors de nos différentes rencontres à Fermilab où durant les nombreux meetings de collaboration.

Je remercie aussi les thésards et autres du CPPM, CPT et LAM : Nicolas Picot-Clément, Guillaume Lambard, Pierre-Simon Mangeard, Benjamin Clément, Mathieu Beau, Baptiste Savoie (bien), Imen Al-Samarai, ... la liste étant non-exhaustive ; pour tous les moments partagés, autant au labo (genre problèmes de codes ou de physique) que en dehors (genre toujours problèmes de codes ou de physique ou existentiel avec en plus la propagation de la science au

reste du monde). Ca a été drôle de se connaître et l'histoire n'est pas finie ...

Je voudrais remercier l'ensemble du personnel du CPPM qui font de ce laboratoire un lieu agréable à vivre.

Je remercie tous mes amis de toujours : le grumeau des potes du collège, celui du lycée et celui des années universitaires qui ont su apprécier les rares moments que j'ai pu leur accorder. Je te remercie particulièrement Mikhail pour le fait que l'on se soit rencontré, pour notre amitié qui a su renaître et pour tout ce que nous réserve l'avenir en bien (j'ai déjà des souvenirs mémorables) et en mal.

Je remercie aussi mes parents et ma famille pour leur soutien. Je tiens à leur dire, malgré que ce ne soit pas mon fort qu'ils sont important pour moi et que je leur serai toujours reconnaissant de m'avoir permis d'en arriver là à travers toutes les concessions et toute leur affection.

Introduction

Le Modèle Standard de la physique des particules permet de décrire les particules fondamentales ainsi que leurs interactions. Il a connu de grand succès et grâce à son pouvoir prédictif, il a permis, par exemple, la découverte expérimentale du quark top à la masse attendue ($m_{\text{top}} = 173.3 \pm 1.1 \text{ GeV}$ [1]). Cependant, ce modèle connaît quelques défauts. Le principal est qu'il ne peut expliquer l'origine de la masse des particules. Pour remédier à ce problème, il est possible d'introduire le mécanisme de Higgs. Ce mécanisme nécessite l'existence d'une nouvelle particule : le boson de Higgs, dont la masse reste un paramètre libre. Les expériences du LEP ont tenté de le découvrir sans succès et il est toujours l'objet d'actives recherches au collisionneur Tevatron ou au naissant LHC. La recherche de cette particule peut se faire dans divers canaux de désintégration et le sujet de la thèse ci-présente porte sur la recherche du boson de Higgs à basse masse ($< 135 \text{ GeV}$) dans le canal où celui-ci se désintègre en paire $b\bar{b}$ et est produit en association avec un boson Z se désintégrant en deux neutrinos. Le détail du mécanisme, les contraintes existantes sur la masse du boson de Higgs ainsi que la justification du canal de désintégration choisi seront explicités dans le premier chapitre.

Les données issues du collisionneur de proton-antiproton Tevatron à Fermilab seront utilisées pour cette recherche. L'énergie dans le centre de masse des collisions est de 1.96 TeV. L'expérience $D\bar{O}$ se situe sur l'un des deux points où s'effectuent les collisions. Le deuxième chapitre porte sur le principe de fonctionnement du Tevatron, du détecteur $D\bar{O}$ et de la description des objets nécessaires à cette analyse.

Le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ comporte dans l'état final deux neutrinos non-détectés produisant de l'énergie transverse manquante et deux quarks b qui vont s'hadroniser pour former des jets. Il est vital pour la sensibilité de l'analyse de pouvoir distinguer les jets issus de quarks b . Cette technique est appelée *b-tagging*. Il existe dans la collaboration un outil standard d'identification de ces quarks (NN) basé sur les traces de grandes impulsions transverse à l'intérieur des jets et la reconstruction de vertex secondaires déplacés (caractéristique des quarks b). Il est possible de compléter ce dernier en utilisant les désintégrations

semi-leptoniques des quarks b. A partir de cette propriété, un nouvel outil d'identification des quarks b a été développée : SLTNN. Le chapitre 3 fera le point sur ces deux méthodes d'étiquetage des jets issus de quark b : NN et SLTNN.

Le chapitre 4 portera sur la recherche du boson de Higgs du Modèle Standard dans le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$. Nous passerons en revue les défis que doit surmonter l'analyse : réduction, maîtrise et modélisation des bruits de fonds. Nous étudierons aussi l'efficacité de notre nouvelle méthode d'étiquetage des jets issus de quark b pour rechercher le boson de Higgs à basse masse.

Cadre Théorique

Ce manuscrit de thèse traite de la recherche du boson de Higgs du Modèle Standard dans le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ dans l'expérience $D\bar{O}$ au Tevatron. Le but de ce chapitre est de poser les bases du Modèle Standard de la physique des particules (noté SM pour *Standard Model*) et de la nécessité d'y introduire le boson de Higgs. Par la suite, nous verrons l'intérêt du canal de désintégration choisi.

1.1 Le Modèle Standard

Le Modèle Standard propose une description des constituants élémentaires de la matière appelées particules. Ces particules peuvent se séparer en deux groupes caractérisés par la valeur de leur spin.

1.1.1 Les Fermions

Ce sont des particules de spin demi-entier qui obéissent à la statistique de Fermi-Dirac. Les fermions sont donc soumis au principe d'exclusion de Pauli, qui leur impose de ne pouvoir se trouver dans le même état quantique. Les fermions sont les particules constituantes de la matière.

On peut distinguer deux types de fermions :

- les quarks : ils sont au nombre de 6 et sont nommés u (*up*), d (*down*), c (*charm*), s (*strange*), b (*bottom*), t (*top*). La découverte de ce dernier en 1995 au Tevatron [2] [3] est un symbole du pouvoir prédictif du Modèle Standard. Les quarks, ne pouvant exister à l'état libre, s'assemblent et forment les hadrons. Il existe deux types de hadrons (observés) : les baryons (combinaison de trois quarks) et les mésons (paire quark anti-quark) ;
- les leptons : aussi au nombre de six, ils comprennent e^- (électron), μ^- (muon), τ^- (tau) et trois neutrinos associés : ν_e , ν_μ et ν_τ .

On peut ranger ces particules en trois familles (comme le montre le tableau 1.1). Par exemple, la première famille comprend les quarks u et d , ainsi que e^- et ν_e . En passant de la famille i à $i + 1$, la masse des particules augmente et donc les particules de la première famille sont les plus stables. En effet, on observe que la matière ordinaire est constituée des fermions de la première famille. Les particules des familles d'ordre supérieure sont issues de processus astrophysiques ou peuvent être générées dans les collisionneurs de particule.

Le Modèle Standard permet aussi l'existence d'anti-particules. Ces anti-particules sont les reflets de la particule correspondante (même masse, spin) à la différence près que leur charge électrique est opposée.

Fermions	1 ^{ère} Famille		2 ^{ème} Famille		3 ^{ème} Famille		Charge
	Nom	Masse (GeV)	Nom	Masse (GeV)	Nom	Masse (GeV)	
Quarks	u	$(1.55-3.3) \times 10^{-3}$	c	1.27	t	171.2	2/3
	d	$(3-6) \times 10^{-3}$	s	$(70-130) \times 10^{-3}$	b	4.2-4.7	-1/3
Leptons	e⁻	3.3×10^{-3}	μ⁻	105.7×10^{-3}	τ⁻	1.77	-1
	ν_e	$< 2 \times 10^{-9}$	ν_μ	$< 0.19 \times 10^{-6}$	ν_τ	$< 18.2 \times 10^{-6}$	0

TAB. 1.1: L'organisation des fermions en trois familles avec leur masse et charge respectives. L'antiparticule associée à chaque fermion possède une charge opposée. Les fermions ont pour valeur de spin $\frac{1}{2}$.

1.1.2 Les Bosons

Ce sont des particules de spin entier qui obéissent à la statistique de Bose-Einstein. L'échange d'un boson entre deux particules élémentaires de spin quelconque constitue une interaction.

Il existe 4 types d'interactions fondamentales :

- l'électromagnétisme : elle agit entre toutes les particules ayant une charge électrique par le biais du photon (noté γ). Elle permet par exemple, la cohésion des atomes en liant les électrons et le noyau des atomes ;
- l'interaction faible : elle agit sur tous les fermions à l'aide des "courants" chargés W^+ , W^- (notés W^\pm) et le "courant" neutre Z^0 . Elle est responsable de la radio-activité β qui transforme un neutron en proton ;
- l'interaction forte : elle agit sur les particules ayant une charge de couleur (rouge, bleu ou vert) : les quarks et les gluons. Ces derniers, au nombre de 8, sont les bosons d'échange de l'interaction forte. Cette interaction est la cause du confinement des quarks dans les hadrons. En effet la force "forte" augmente lorsque les particules "colorées" s'éloignent. L'habillage des quarks dans une interaction donnée (appelé hadronisation) entraîne des états neutres de couleur et ce phénomène est à l'origine des "jets" de particules observés dans les détecteurs de physique des particules ;
- la gravitation : elle agit sur toutes les particules massives. Cette interaction a pour boson d'échange le graviton (particule de masse nulle, de charge électrique nulle et de spin

2) mais il n’y a toujours pas de preuve expérimentale directe de son existence. De plus, c’est la seule interaction qui ne soit pas encore incluse dans le Modèle Standard.

Le Modèle Standard nécessite l’existence d’une particule supplémentaire : le boson de Higgs que nous présenterons plus en détail dans les paragraphes qui suivent.. Cette particule fait l’objet de recherches actives dans les collisionneurs de particules. Le tableau 1.2 résume l’ensemble des bosons du Modèle Standard.

Bosons			Charge	Spin
Nom	Masse (GeV)	Interaction		
photon γ	0	Electromagnétique	0	1
W^\pm, Z^0	80.403, 91.188	Faible	$\pm 1, 0$	1
8 gluons g	0	Forte	0	1
Higgs H	$>114.4 @ 95\% C.L.$ $\notin [158, 175] @ 95\% C.L.$	Mécanisme de Higgs	0	0

TAB. 1.2: Les bosons associés aux trois interactions décrites par le Modèle Standard de la physique des particules avec leur masse et charge respectives. Ces bosons sont de spin 1. Le boson de Higgs induit par le mécanisme de Higgs est de spin nul et sa masse reste à définir. Les contraintes indiquées sont celles obtenues expérimentalement et la mention C.L. signifie *Confidence Level* (niveau de confiance).

1.1.3 Le modèle

Le Modèle Standard de la physique des particules repose sur une théorie quantique des champs. Les interactions entre les particules ainsi que la nature des bosons d’échange sont décrits par la théorie de jauge :

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y \quad (1.1)$$

- $U(1)_Y$ est le groupe d’hypercharge ;
- $SU(2)_L$ est le groupe d’isospin et décrit l’interaction faible ;
- $SU(2)_L \otimes U(1)_Y$ est l’unification des deux groupes précédents pour former la théorie électrofaible ;
- $SU(3)_c$ est le groupe de la chromodynamique quantique QCD (*Quantum ChromoDynamics*) et décrit l’interaction forte. L’indice c fait référence à la charge de couleur des quarks et des gluons.

A l’aide de ce cadre, les particules et les interactions sont bien décrites. De plus les fermions et les bosons doivent être de masse nulle afin de rendre la théorie renormalisable ¹. Cependant, les fermions (hormis les neutrinos) et les bosons W^\pm et Z^0 ont été découverts et leur masse

¹seules les théories renormalisables sont considérées comme valides. La renormalisation est l’opération qui permet, par exemple, de rendre finie la charge de l’électron.

a été mesurée avec précision. Une solution à ce problème de la masse est proposée par le mécanisme de Higgs.

1.1.4 Le mécanisme de Higgs

Le Modèle Standard montre que la masse des particules n'est pas une de leurs propriétés intrinsèques, mais plutôt la conséquence d'une interaction. En rajoutant un champ de Higgs qui interagit avec les fermions et les bosons W^\pm et Z^0 , le Lagrangien du Modèle Standard peut générer la masse de ces dernières. Les termes de masse apparaissent avec la brisure de symétrie de l'invariance locale de jauge des champs de Higgs et électrofaibles ($SU(2)_L \otimes U(1)_Y$). On appelle ce processus le mécanisme de Higgs [4] [5] [6] [7] [8].

Le champ de Higgs est un doublet de champs scalaires complexes :

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad (1.2)$$

et le Lagrangien associé :

$$\mathcal{L}_H = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \quad (1.3)$$

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (1.4)$$

Ce Lagrangien comprend deux termes :

- $(D_\mu \phi)^\dagger (D^\mu \phi)$ correspond à la partie cinétique du champ scalaire. Le terme D_μ représente la dérivée covariante permettant d'imposer l'invariance de jauge locale. Elle s'exprime comme suit :

$$D_\mu = \partial_\mu + ig_2 \frac{\tau^a}{2} W_\mu^a + ig_1 \frac{Y}{2} B_\mu \quad (1.5)$$

où $W^{\mu a}$ représente les bosons de $SU(2)_L$, g_2 sa constante de couplage et τ^a ses générateurs (matrices de Pauli avec $a=1, 2, 3$) ; B_μ représente le boson de $U(1)_Y$, g_1 sa constante de couplage et Y l'hypercharge ;

- $V(\phi)$ est un potentiel scalaire définit de telle sorte qu'il soit invariant sous $SU(2)_L$ et renormalisable. La figure 1.1 montre le potentiel $V(\phi)$ à deux dimensions, avec μ^2 négatif et λ positif. Le potentiel possède alors une infinité de minimas tels que : $|\phi|^2 = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}$. L'état de vide v est l'une de ces positions et est noté *vev* (*vacuum expectation value*). Après avoir effectué un changement de coordonnées, on obtient :

$$\phi = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \quad (1.6)$$

En développant autour du minimum du potentiel $V(\phi)$, le champ scalaire devient :

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix} \quad (1.7)$$

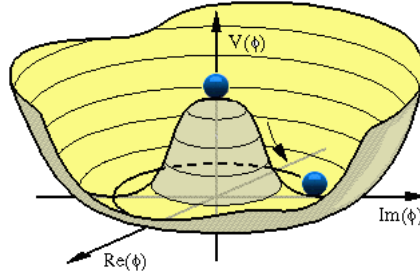


FIG. 1.1: Représentation du potentiel utilisé dans le mécanisme de Higgs.

où h est un champ nul dans le vide.

Le Lagrangien du modèle Standard pour le secteur électrofaible est le suivant :

$$\begin{aligned} \mathcal{L}_{\text{SU}(2)_L \times \text{U}(1)_Y} = & -\frac{1}{4} W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \\ & + \bar{L}_i i D_\mu \gamma^\mu L_i + \bar{e}_{R_i} i D_\mu \gamma^\mu e_{R_i} + \bar{Q}_i i D_\mu \gamma^\mu Q_i + \bar{u}_{R_i} i D_\mu \gamma^\mu u_{R_i} + \bar{d}_{R_i} i D_\mu \gamma^\mu d_{R_i} \end{aligned} \quad (1.8)$$

où :

- $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$;
- $W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_2 f_{abc} W_\mu^b W_\nu^c$ et f_{abc} sont les constantes du groupe de symétrie considéré définies comme $[T^a, T^b] = i f_{abc} T^c$ avec T^a les générateurs de ce groupe ;
- γ^μ sont les matrices 4×4 de Dirac : $\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ et $\gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}$ avec σ^i les matrices de Pauli ;
- L_i (resp. Q_i) représentent les doublets de leptons (resp. de quarks) de chiralité gauche. e_{R_i} (resp. u_{R_i} et d_{R_i}) représentent les singlets de leptons (resp. de quarks) de chiralité droite.

Les bosons physiques médiateurs de l'interaction faible (W^\pm et Z) et électromagnétique A_μ (photon γ) sont définis à partir des champs de jauge de la manière suivante :

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \pm W_\mu^2) \quad (1.9)$$

$$Z_\mu = -B_\mu \sin\theta_w + W_\mu^3 \cos\theta_w \quad (1.10)$$

$$A_\mu = B_\mu \cos\theta_w + W_\mu^3 \sin\theta_w \quad (1.11)$$

où θ_w est l'angle de mélange de Weinberg qui lie les constantes de couplages de l'interaction faible et électromagnétique de la manière suivante :

$$\sin\theta_w = \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \quad \cos\theta_w = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} \quad (1.12)$$

Les combinaisons linéaires des champs de jauge sont définies de manière à rendre le terme de masse du photon nul.

En reportant le Lagrangien associé au champ de Higgs dans le Lagrangien du Modèle Standard, nous obtenons les termes de masse des bosons :

$$M_{W^\pm} = \frac{g_2 v}{2}, \quad M_{Z^0} = \frac{v}{2} \sqrt{g_1^2 + g_2^2}, \quad M_\gamma = 0, \quad M_H = \sqrt{-2\mu^2} = \sqrt{2\lambda v^2} \quad (1.13)$$

Il est possible de générer la masse des fermions (sauf les neutrinos qui n'ont pas de composante chirale droite) en utilisant le Lagrangien de Yukawa (pour chaque famille de fermions), le champ scalaire ϕ et le champ $\phi_c = i\tau_2 \phi^*$ (τ_2 est l'une des trois matrices de Pauli). Ce Lagrangien est invariant sous $SU(2)_L \times U(1)_Y$:

$$\mathcal{L}_{\text{Yukawa}} = -\lambda_e \bar{L}_i \phi e_{R_i} - \lambda_d \bar{Q}_i \phi d_{R_i} - \lambda_u \bar{Q}_i \phi_c u_{R_i} + \text{conjugué hermitien} \quad (1.14)$$

La masse des fermions obtenue est la suivante :

$$m_f = \lambda_f \frac{v}{\sqrt{2}} \quad (1.15)$$

où λ_i sont appelés couplages de Yukawa et représentent l'intensité du couplage entre les fermions et le boson de Higgs.

1.2 La recherche du boson de Higgs

On a vu que l'introduction du champ de Higgs h permet de générer la masse des bosons de $SU(2)_L \times U(1)_Y$ ainsi que des fermions (sauf les neutrinos) par le processus de brisure de symétrie électrofaible. Il entraîne aussi l'existence d'un boson de Higgs dont la masse dépend de μ (voir équation 1.13) et reste donc indéterminée par la théorie.

1.2.1 Les contraintes sur la masse du boson de Higgs

Afin de réduire le champ de recherche du boson de Higgs, nous avons à notre disposition plusieurs contraintes sur sa masse.

Les contraintes théoriques

Une contrainte supérieure sur la masse du boson de Higgs peut être déduite de contraintes d'unitarité [9] pour les processus de diffusion des di-bosons (comme $W^+W^- \rightarrow W^+W^-$) à grande énergie ($s \gg M_W^2$). L'amplitude de cette diffusion augmente proportionnellement à l'inverse de m_H^2 . Il est possible de définir la section efficace de diffusion en terme d'ondes partielles a_l et l'on peut montrer la relation suivante :

$$a_0 \xrightarrow{s \gg M_W^2} -\frac{M_H^2}{8\pi v^2} \quad (1.16)$$

La condition d'unitarité impose $|\text{Re}(a_1)| < 1/2$ et on trouve la limite suivante : $M_H < 870$ GeV. En utilisant les ordres supérieurs des arbres dans ce calcul, on obtient une limite supérieure d'environ 1 TeV.

On peut déterminer une autre contrainte sur la masse du boson de Higgs à partir du paramètre Λ , qui représente la frontière en énergie de validité du Modèle Standard. Au-delà de cette valeur la théorie des perturbations ne peut plus être appliquée et les prédictions théoriques sont dominées par des phénomènes non-perturbatifs. Cette contrainte est obtenue en regardant l'évolution de la constante d'auto-couplage du boson de Higgs λ_H pour une échelle d'énergie basse Q_0 (souvent prise égale à v) et une échelle d'énergie quelconque Q :

$$\frac{1}{\lambda_H(Q)} = \frac{1}{\lambda_H(Q_0)} - \frac{3}{4\pi^2} \ln \frac{Q^2}{Q_0^2} \quad (1.17)$$

Afin de conserver $\lambda(Q) > 0$ lorsque $Q \rightarrow \infty$, il est nécessaire que $\lambda(Q_0) \rightarrow 0$ (ce qui est équivalent à une théorie de champs libres). On obtient ainsi une borne supérieure sur la masse du boson de Higgs pour la limite d'énergie Λ :

$$\lambda_v = \frac{M_H^2}{2v^2} \leq \frac{4\pi^2}{3 \ln \frac{\Lambda^2}{v^2}} \quad (1.18)$$

Il est aussi possible de fournir une contrainte inférieure sur cette masse. Afin de satisfaire les conditions potentiel borné par le bas et constante d'auto-couplage du Higgs pour une échelle d'énergie donnée positive, on obtient une borne inférieure sur la masse du boson de Higgs :

$$M_H^2 > \frac{v^2}{8\pi^2} \left[-12 \frac{M_t^4}{v^4} + \frac{3}{16} (2g_2^4 + (g_2^2 + g_1^2)^2) \right] \ln \frac{Q^2}{v^2} \quad (1.19)$$

Ainsi nous avons assuré la stabilité du vide [9].

Le résultat des deux contraintes précédentes est reporté sur la figure 1.2.

Les contraintes expérimentales directes

Le boson de Higgs est recherché de manière active auprès de collisionneurs de particules. Sa non-découverte au jour d'aujourd'hui, a permis l'exclusion de certains domaines de masse du boson de Higgs (voire figure 1.3).

Le premier résultat provient des 4 expériences du LEP : ALEPH, DELPHI, L3 et OPAL. Le LEP était un collisionneur e^+e^- avec une énergie dans le centre de masse allant de 91 à 209 GeV. Dans cette configuration le canal de production du boson de Higgs principal est : $e^+e^- \rightarrow Z^* \rightarrow HZ$. Les canaux étudiés ont utilisé toutes les désintégrations du boson Z et les désintégrations du boson de Higgs H en $b\bar{b}$ et $\tau^+\tau^-$. Les deux périodes de prise de données (LEP1 et LEP2) entre 1989 et 2000 correspondent à une luminosité intégrée de 2461 pb⁻¹. La recherche du boson de Higgs a abouti à une limite inférieure à 95% de niveau de confiance :

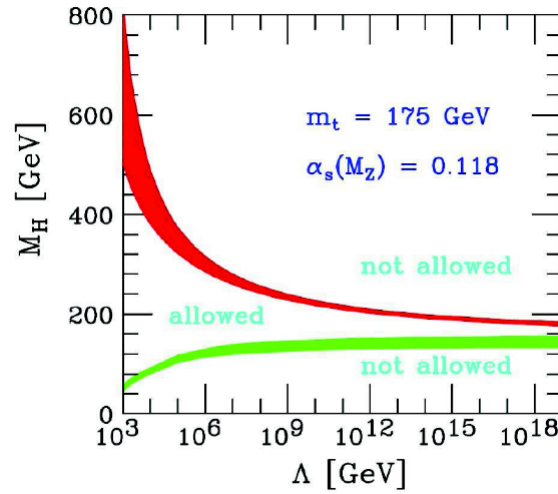


FIG. 1.2: Contraintes théoriques sur la masse du boson de Higgs (M_H) en fonction de l'échelle Λ : frontière en énergie de validité du Modèle Standard. Les limites supérieures (frontière du régime perturbatif) et inférieures (stabilité du vide) sont reportées en rouge et vert respectivement.

$m_H > 114.4$ GeV. Finalement le LEP s'est arrêté en 2000 pour permettre de voir le jour au LHC qui a démarré en 2009.

Depuis 2002, les expériences du Tevatron (DØ et CDF) ont pris la relève et ont intensifié leurs efforts dans ce domaine. Le Tevatron est un collisionneur $p\bar{p}$ avec une énergie dans le centre de masse de 1.96 TeV. Les données du *RunII* du Tevatron délivrées entre 2002 et aujourd'hui correspondent à 9 fb^{-1} par expérience. La combinaison de plusieurs canaux de recherche de Higgs et des résultats des deux expériences (luminosité intégrée jusqu'à environ : 6 fb^{-1}) ont permis d'exclure (à l'été 2010) les masses du boson de Higgs avec un niveau de confiance à 95% dans les domaines [100, 109 GeV] et [158, 175 GeV] (voir référence [10] pour plus de détails).

Les contraintes expérimentales indirectes

A partir des résultats expérimentaux sur les mesures de précision électrofaible, il est possible de contraindre la masse du boson de Higgs. La masse du W^\pm est sensible à la correction des boucles d'ordre supérieur des masses du quark top et du boson de Higgs. Les corrections sur la masse du W^\pm dépend de la masse du top de manière quadratique et de la masse du boson de Higgs de manière logarithmique. Ces relations permettent de contraindre le domaine possible pour la masse de boson de Higgs (voir la figure 1.4).

Toujours à partir des mesures de paramètres électrofaibles dépendants de la masse du boson de Higgs (dont M_{W^\pm} et M_{top}), on peut définir $\Delta\chi^2$ entre les valeurs de χ^2 théorique

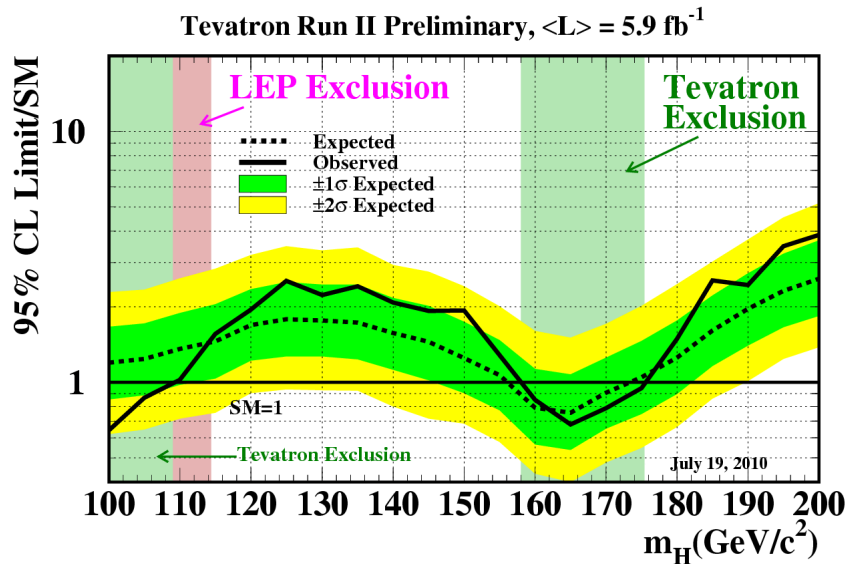


FIG. 1.3: Evolution de la limite sur la section efficace de production du boson de Higgs en fonction de sa masse, obtenue en combinant l'ensemble des résultats des expériences $D\bar{O}$ et CDF du Tevatron. Les lignes noires et pointillées représentent respectivement les limites observées et attendues divisées par la section efficace prédite par le Modèle Standard. L'erreur associée à la limite attendue à 1 et 2 σ est symbolisée par les bandes vertes et jaunes. La bande verticale rose représente les masses de boson de Higgs exclues par le LEP ($m_{\text{Higgs}} < 114.4$ GeV) et les bandes vert-claires correspondant à celles exclues par le Tevatron ($100 < m_{\text{Higgs}} < 109$ GeV et $158 < m_{\text{Higgs}} < 175$ GeV) à 95% C.L. [10].

et mesuré expérimentalement. L'ajustement du $\Delta\chi^2$ donne une masse de boson de Higgs la plus probable (minimum de l'ajustement) de 89^{+35}_{-26} GeV à 68% de niveau de confiance et une borne supérieure sur cette même masse de 158 GeV à 95% de niveau de confiance (voir la figure 1.5) [11].

1.2.2 Le boson de Higgs au Tevatron

A la vue des contraintes obtenues précédemment, il est légitime de rechercher un boson de Higgs de masse inférieure à 200 GeV. Pour de telles valeurs de masse, la figure 1.6 de gauche, nous montre que le processus dominant de production du boson de Higgs est la fusion de gluon (voir la figure 1.7). La figure 1.6 de droite, met en évidence deux modes privilégiés de désintégration du boson de Higgs suivant sa masse : à basse masse ($m_H < 135$ GeV) désintégration en $b\bar{b}$ et à haute masse désintégration en W^+W^- . Le cas de figure $gg \rightarrow H \rightarrow b\bar{b}$ (section efficace de production pb) ne peut pas être étudié à basse masse étant donné que

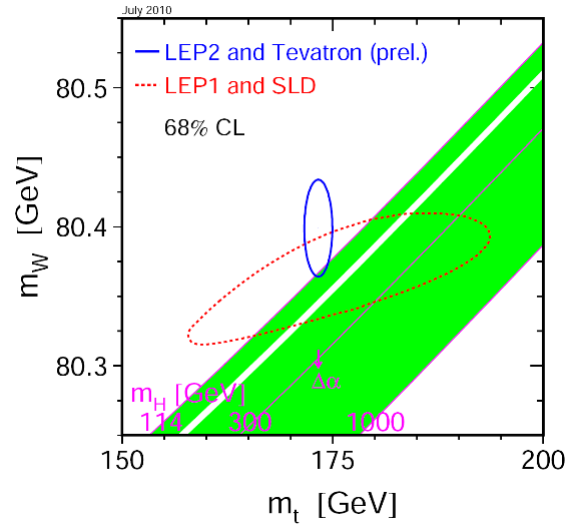


FIG. 1.4: Masse du boson W en fonction de celle du quark top. Les zones bleues et rouge représentent les contours obtenus à 68% C.L. respectivement par les mesures directes du Tevatron et de LEP2, et les mesures indirectes de LEP1 et SLD. La zone verte montre les valeurs autorisées par le Modèle Standard et pour différentes masses de boson de Higgs du couple (m_W, m_t) [11].

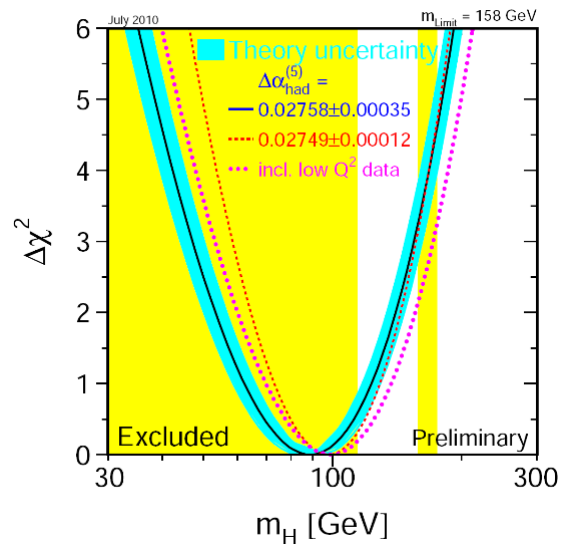


FIG. 1.5: $\Delta\chi^2$ de l'ajustement des mesures de précision électro-faibles (ligne noire) et son incertitude associée (bande bleue) en fonction de la masse du boson de Higgs. Les zones en jaune représentent les domaines de masse exclus par le LEP ($m_{\text{Higgs}} < 114.4$ GeV) et le Tevatron ($158 < m_{\text{Higgs}} < 175$ GeV) à 95% C.L. [11].

le bruit de fond principal dans les collisionneurs hadroniques est du type $q\bar{q}$ (bruit de fond QCD) dont la section efficace de production μb et constitue par conséquent un bruit de fond gigantesque.

A basse masse, il convient donc de rechercher les autres processus de production du boson de Higgs. Les autres processus possibles pour un boson de Higgs de basse masse sont les productions associées avec un boson W ou Z . La figure 1.7 montre les trois canaux d'études principaux pour la recherche d'un boson de Higgs de basse masse. Dans le cas de la production associée avec un boson Z , où celui-ci se désintègre en neutrinos, la non-interaction des neutrinos dans le détecteur crée de l'énergie transverse manquante (notée par la suite \cancel{E}_T). Ce canal d'étude reste cependant très intéressant car le rapport de branchement $Z \rightarrow \nu\bar{\nu}$ est de 20% (pas de distinction du type de neutrinos). Ce rapport de branchement est à comparer à celui du canal plus facilement identifiable $Z \rightarrow l^+l^-$ avec $l^- = e^-$ ou μ^- ou τ^- qui est seulement de 3%.

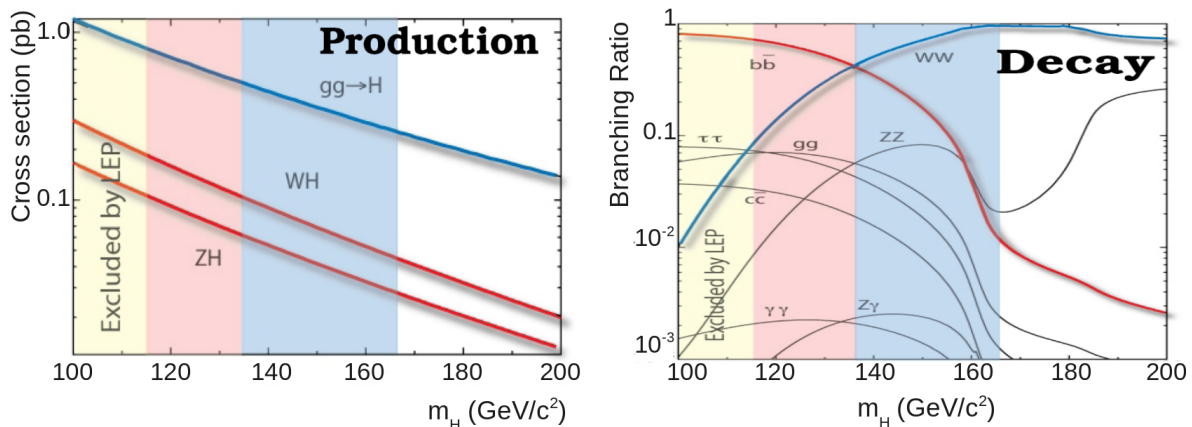


FIG. 1.6: La figure de gauche montre la section efficace de production du boson de Higgs (en pb) et la figure de droite son rapport de branchement. Ces deux figures sont paramétrées en fonction de la masse du boson de Higgs (en GeV) et sont définies à partir des caractéristiques du Tevatron (collisions $p\bar{p}$). La bande jaune symbolise les masses exclues par le LEP ($m_H > 114.4$ GeV). La bande rouge représente l'échelle de masse de boson de Higgs dite de basse masse et la bande bleue de haute masse. La limite entre ces deux bandes est située à environ 135 GeV, en-dessous de cette valeur le boson de Higgs se désintègre principalement en $b\bar{b}$ et au-dessus en WW . [9]

Le résultat le plus récent publié sur la limite attendue (observée) sur la section efficace divisée par la valeur du Modèle Standard obtenu pour le canal $ZH \rightarrow \nu\bar{\nu}b\bar{b}$, pour une luminosité intégrée de 5.2 fb^{-1} et pour une masse de boson de Higgs de 115 GeV est de 4.6 (3.7) [12]. Ce résultat ainsi que ceux qui seront présentés par la suite ont été évalués à 95% de niveau de confiance. Ce résultat a été mise à jour avec une luminosité intégrée de 6.4 fb^{-1} et a pour valeur 4.2 (3.4) [13] pour une masse de boson de Higgs de 115 GeV. On peut comparer ce résultat à la limite attendue (observée) sur la section efficace divisée par la valeur du Modèle

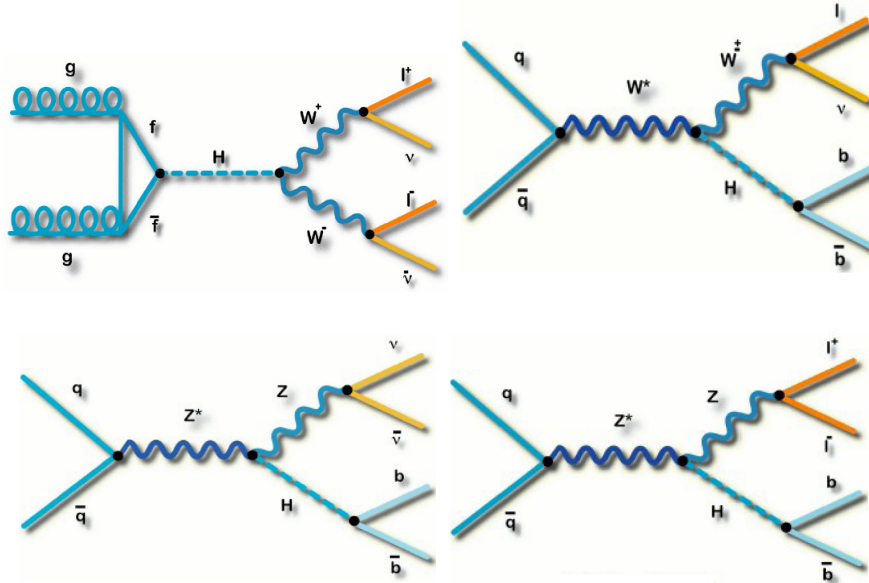


FIG. 1.7: Diagrammes de Feynman à l'ordre de l'arbre des canaux majeurs de recherche du boson de Higgs à haute masse (fusion de gluons) à basse masse (production associée avec un boson Z ou W) au Tevatron. De haut en bas et de gauche à droite, nous avons les canaux $H \rightarrow W^+W^- \rightarrow l^+ \nu l^- \bar{\nu}$, $WH \rightarrow l^\pm \nu b \bar{b}$, $ZH \rightarrow \nu \bar{\nu} b \bar{b}$ et $ZH \rightarrow l^+ l^- b \bar{b}$. Les canaux contenant un lepton (l) dans l'état final sont analysés indépendamment pour les cas électron, muon ou tau.

Standard obtenue pour les principaux autres canaux de recherche de boson de Higgs dans l'expérience DØ :

- le canal $ZH \rightarrow l^+ l^- b \bar{b}$ a obtenu pour un boson de Higgs de masse 115 GeV et une luminosité intégrée de $6.2 \text{ fb}^{-1} : 5.7 (8.0)$ [14];
- le canal $WH \rightarrow l^\pm \nu b \bar{b}$ a obtenu pour un boson de Higgs de masse 115 GeV et une luminosité intégrée de $5.3 \text{ fb}^{-1} : 4.8 (4.1)$ [15];
- le canal $H \rightarrow W^+W^- \rightarrow l^+ \nu l^- \bar{\nu}$ a obtenu pour un boson de Higgs de masse 115 GeV et une luminosité intégrée de $5.4 \text{ fb}^{-1} : 14.9 (20.8)$ [16]. Ce canal étant plus sensible à un boson de Higgs de haute masse, il est intéressant de regarder ce résultat pour un boson de Higgs de masse 165 GeV : 1.36 (1.55).

On peut aussi comparer le résultat sur la limite obtenue dans le canal $ZH \rightarrow \nu \bar{\nu} b \bar{b}$ de DØ à celui de CDF. La stratégie d'analyse de CDF dans ce canal prend elle aussi en compte les événements $WH \rightarrow l^\pm \nu b \bar{b}$ où le lepton n'est pas identifié. La limite attendue (observée) sur la section efficace divisée par la valeur du Modèle Standard obtenue pour un boson de Higgs de 115 GeV et une luminosité intégrée de 5.7 fb^{-1} est de 4.0 (2.3) [17].

Les résultats combinés des différents canaux de recherche de boson de Higgs relatifs à l'été 2010 pour une masse de boson de Higgs de 115 GeV sont de 1.9 (1.79) pour CDF [18] et de

2.31 (2.65) pour $D\emptyset$ [19]. CDF a obtenu ce résultat avec des luminosités intégrées pour les différents canaux variant entre 2.3 et 5.9 fb^{-1} alors que $D\emptyset$ utilise une statistique variant entre 2.1 et 6.7 fb^{-1} .

Les résultats combinés de $D\emptyset$ et CDF pour une masse de boson de Higgs de 115 GeV sont de 1.50 (1.64) [10].

Le LHC est un collisionneur proton-proton construit pour avoir une énergie dans le centre de masse de 14 TeV. A la fin de l'année 2011, il est prévu que les expériences ATLAS et CMS aient collecté 1 fb^{-1} de données avec des collisions de 7 TeV. Avec cette énergie dans le centre de masse et des collisions proton-proton, les processus dominants de production du boson de Higgs du Modèle Standard sont la fusion de gluon et VBF $q\bar{q}H$ (*Vector Boson Fusion*) comme le montre la figure 1.8 [20]. Les processus de production associée avec un boson Z ou W ne seront pas utilisés pour la recherche de boson de Higgs étant donné leur faible section efficace de production par rapport à l'important bruit de fond. Les principaux canaux d'étude sont la production directe d'un boson de Higgs se désintégrant en WW (par suite deux leptons et deux neutrinos), ZZ (par suite quatre leptons) ou $\gamma\gamma$. Les résultats attendus pour chaque expérience avec 1 fb^{-1} sont les suivants [21] :

- exclusion à 95% de niveau de confiance sur la fenêtre de masse 140-185 GeV pour le canal WW ;
- la limite de production de boson de Higgs du canal $\gamma\gamma$ attendue est 5 fois supérieure à celle prédite par le Modèle Standard dans la fenêtre de masse 110-140 GeV ;
- la limite de production de boson de Higgs divisée par celle prédite par le Modèle Standard pour la combinaison des canaux WW , ZZ et $\gamma\gamma$ est résumée sur la figure 1.9.

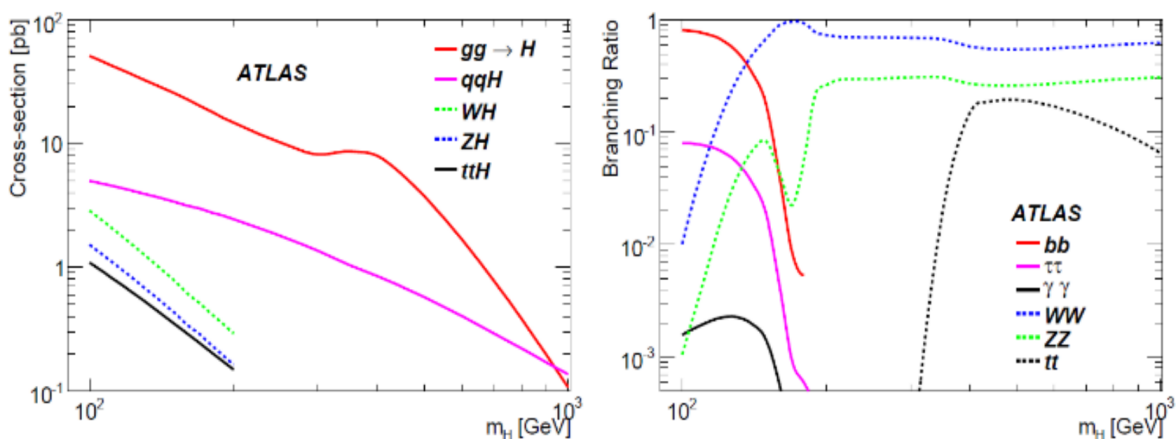


FIG. 1.8: La figure de gauche montre la section efficace de production du boson de Higgs (en pb) et la figure de droite son rapport de branchement avec une énergie dans le centre de masse de 14 TeV [20]. Ces deux figures sont paramétrées en fonction de la masse du boson de Higgs (en GeV) et sont définies à partir des caractéristiques du LHC (collisions p p).

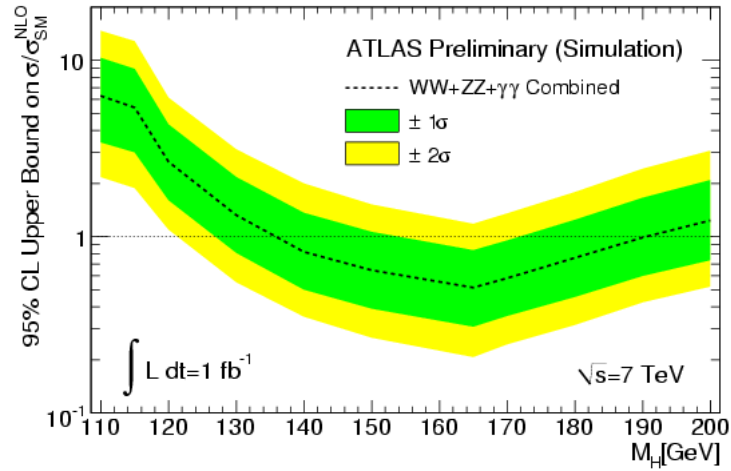


FIG. 1.9: Evolution de la limite sur la section efficace de production du boson de Higgs en fonction de sa masse, prévue pour la seule expérience ATLAS du LHC avec 1 fb^{-1} de données (fin 2011) et une énergie dans le centre de masse de 7 TeV [21]. La ligne noire en pointillés représente la limite attendue divisée par la section efficace prédite par le Modèle Standard. L'erreur associée à cette limite attendue à 1 et 2σ est symbolisée par les bandes vertes et jaunes.

1.3 Conclusion

Le Modèle Standard permet d'expliquer à l'aide de particules élémentaires la structure de la matière (fermions) ainsi que leurs interactions véhiculées par d'autres particules élémentaires (bosons). Ce modèle nécessite l'introduction d'un nouveau champ scalaire afin de pouvoir expliquer l'origine de la masse des particules élémentaires. Ce champ interagit avec les particules par le biais du boson de Higgs, dont la masse n'est pas prédite par le modèle. Plusieurs contraintes théoriques et expérimentales permettent de réduire le champ d'investigation de ce boson. Il fait l'objet de nombreuses recherches auprès des collisionneurs de particules comme au Tevatron et désormais au LHC. Les canal d'étude $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ étudié par la suite est l'un des plus sensibles pour la recherche du boson de Higgs de basse masse ($m_H < 135 \text{ GeV}$) au Tevatron.

Cadre Expérimental

Ce chapitre a pour objectif de présenter succinctement le cadre expérimental de l'analyse en commençant par la description du système de création et d'accélération des faisceaux du Tevatron [22]. Ensuite nous verrons le fonctionnement du détecteur DØ. Finalement nous décrirons les objets détectés, la chaîne d'acquisition des données et la génération des événements simulés nécessaires aux analyses.

2.1 Le système d'accélération

Dans ce manuscrit, nous utilisons les données issues du collisionneur proton-antiproton de Fermilab. La figure 2.1 présente les différents systèmes qui le composent. Les caractéristiques et le rôle de ces derniers seront détaillés par la suite.

2.1.1 Cycle des protons

Le processus de fabrication du faisceau de protons commence par l'ionisation d'atomes d'hydrogène dans une cavité baignée dans un champ électrostatique. Ensuite, attirés par le champ électrique, les protons résultants s'accumulent sur une plaque de césium. Le cycle d'arrivée de protons continuant, ces nouveaux protons libèrent ceux déjà présents sur la plaque. En se libérant, ils arrachent deux électrons au césium et le champ électrique expulse les ions H^- ainsi formés (figure 2.2 de gauche). Les ions atteignent une énergie de 750 keV grâce à un accélérateur de type Cockroft – Walton [23].

Ces ions sont ensuite portés jusqu'à une énergie de 400 MeV à l'aide de cavités radiofréquences. Cette opération est réalisée par l'accélérateur linéaire LINAC [24] long de 130 mètres. Les ions H^- arrivent ensuite dans le premier synchrotron appelé *booster* [25] (figure 2.2 de droite). Ce dernier a un rayon de 475 m et a pour objectif de transformer les ions H^- en protons et ensuite de les accélérer. Les protons sont obtenus en arrachant les électrons des ions H^- lors de leur passage à travers une feuille de carbone. Ensuite les cavités radiofréquences

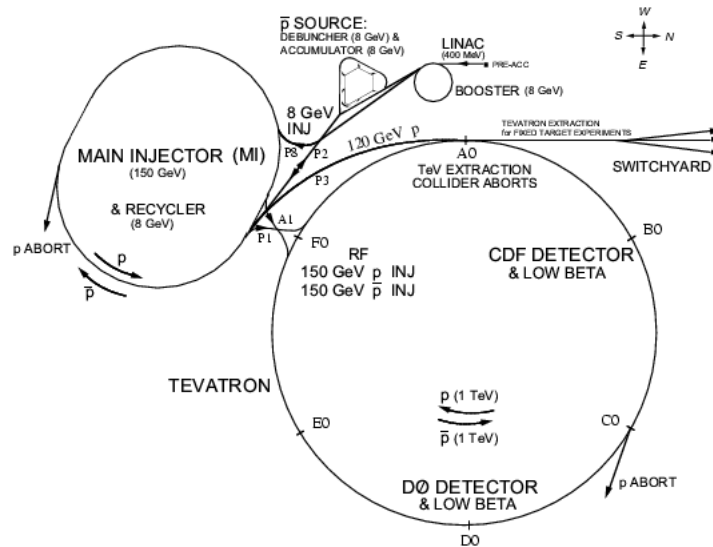


FIG. 2.1: Ensemble des systèmes de création et d'accélération des faisceaux de proton et d'antiprotons du Tevatron à Fermilab.

du *booster* accélèrent les protons jusqu'à une énergie de 8 GeV.

L'injecteur principal (*Main Injector* [26] [27]) prend le relais. Cet anneau de 3.3 km de circonférence, accélère les protons jusqu'à une énergie de 150 GeV et les regroupe par paquets de $3 \cdot 10^{11}$ particules. Par la suite, ces paquets sont envoyés dans l'anneau principal appelé Tevatron.

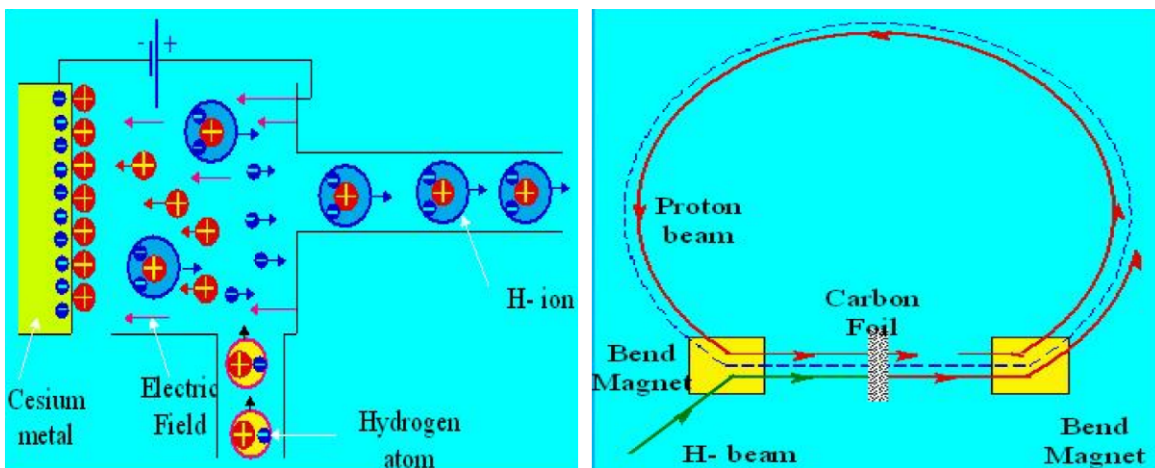


FIG. 2.2: Figure de gauche : cycle de production des ions H^- dans le Cockcroft-Watson. Figure de droite : les ions H^- injectés dans le *booster*.

2.1.2 Cycle des antiprotons

Pour générer les antiprotons [28], on utilise le faisceau de protons de l'injecteur principal porté à une énergie de 120 GeV et on l'envoie sur une cible de Nickel (figure 2.3 de gauche). Le taux de production d'antiprotons pour un tel processus est de $1/10^5$. Les particules résultantes sont focalisées par une lentille au Lithium et on utilise des aimants dipolaires afin de séparer les antiprotons obtenus des autres particules. La structure en paquets d'antiprotons est préservée mais a subi une dispersion en énergie et en direction.

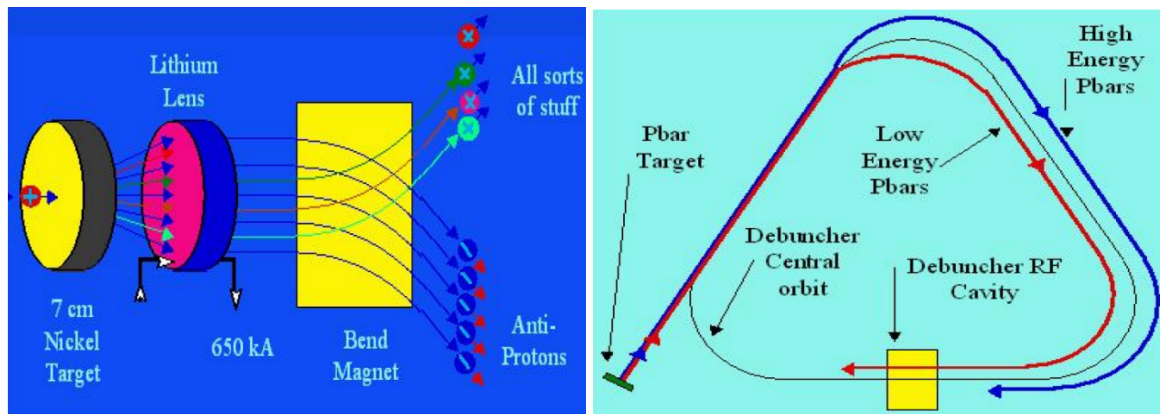


FIG. 2.3: Figure de gauche : cycle de production des antiprotons sur la cible de Nickel. Figure de droite : production et accélération des antiprotons dans le *debuncher*.

Le faisceau produit passe ensuite dans un synchrotron triangulaire *Debuncher* (figure 2.3 de droite). Son rôle est d'uniformiser le faisceau en énergie en faisant parcourir une plus grande distance aux particules de plus grandes impulsions. De plus la cavité radiofréquence fournit une accélération plus grande aux antiprotons lents. Le faisceau est aussi uniformisé en direction par un système de refroidissement stochastique (figure 2.4 de gauche). Ce faisceau arrive ensuite dans l'accumulateur (figure 2.4 de droite) où il subit de nouveau un refroidissement stochastique. Le faisceau uniformisé à une énergie de 8 GeV est envoyé dans le *Recycler*. Ce dernier est situé dans le même tunnel que l'injecteur principal et stocke les antiprotons.

Lorsque le faisceau contient suffisamment d'antiprotons, il les envoie dans l'injecteur principal. Les antiprotons tournent dans le sens inverse des protons et sont accélérés jusqu'à une énergie de 150 GeV.

2.1.3 Le Tevatron

Dernier élément de la chaîne d'accélération des faisceaux de protons et d'antiprotons injectés en sens inverses depuis l'injecteur principal. C'est un synchrotron de 1 km de rayon qui accélère les faisceaux jusqu'à une énergie de 980 GeV grâce à des cavités radiofréquences placées au point F0 (figure 2.1). Les faisceaux sont répartis en 3 super-paquets espacés de 2.6

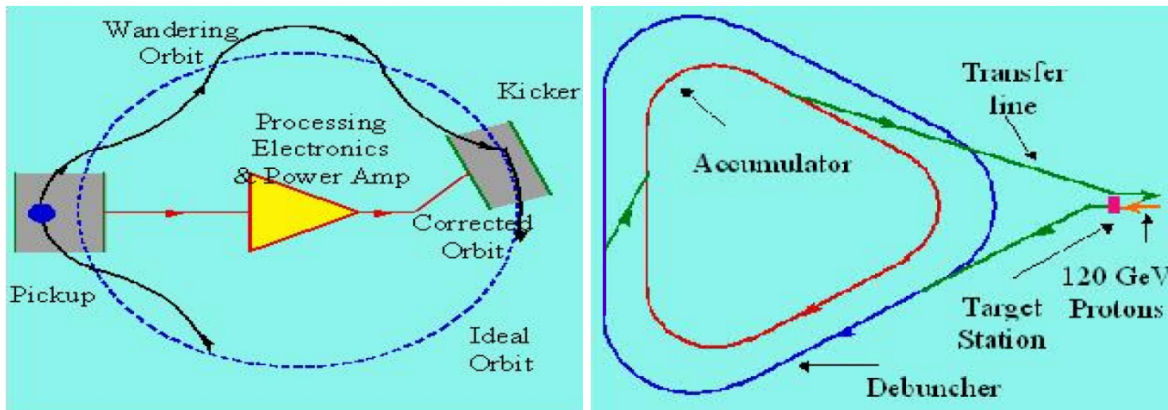


FIG. 2.4: Figure de gauche : cycle de refroidissement stochastique. Figure de droite : stockage des antiprotons dans l'accumulateur.

μs . Ces super-paquets contiennent 12 paquets séparés de 396 ns.

Les faisceaux sont guidés et focalisés dans le Tevatron par le biais d'aimants supraconducteurs. Ce système permet de générer les collisions des faisceaux en deux points : B0 et D0, qui accueillent respectivement les expériences CDF et DØ. Les collisions obtenues ont une énergie dans le centre de masse de 1.96 TeV.

2.2 Le détecteur DØ

2.2.1 Vue d'ensemble

Les domaines de physique des particules étudiés par le détecteur DØ [29] sont généralistes : mesures de précision du Modèle Standard (masse du quark top ou du boson W par exemple), recherche du boson de Higgs ou de nouvelle physique au-delà du Modèle Standard.

Le détecteur DØ est de forme cylindrique (voir la figure 2.5) et est constitué de plusieurs sous-détecteurs qui seront détaillés par la suite. La localisation de ces objets mesurés par le détecteur utilise comme système de coordonnées (figure 2.6 de gauche) un repère orthonormé centré sur le centre du détecteur. L'axe x est dirigé horizontalement vers l'extérieur du Tevatron, l'axe y vers le haut et l'axe z suivant l'axe du faisceau de protons. Etant donné la géométrie du détecteur, on introduit la distance $r = \sqrt{x^2 + y^2}$ et aussi les deux angles : θ et ϕ qui sont respectivement l'angle polaire et azimutal. Il est utile de définir la rapidité :

$$y = \frac{1}{2} \ln \left[\frac{E + P_z}{E - P_z} \right] \quad (2.1)$$

qui est approximée par la pseudo-rapacité lorsque la masse de la particule devient négligeable devant son énergie :

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (2.2)$$

On utilisera par la suite la pseudo-rapacité η plutôt que l'angle polaire. La pseudo-rapacité peut être définie par rapport au centre du détecteur ou par rapport au point d'interaction (figure 2.6 de droite). Nous les nommerons respectivement η_{det} et η_{phy} . On peut ainsi définir des cônes de rayon $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta_{det}^2}$.

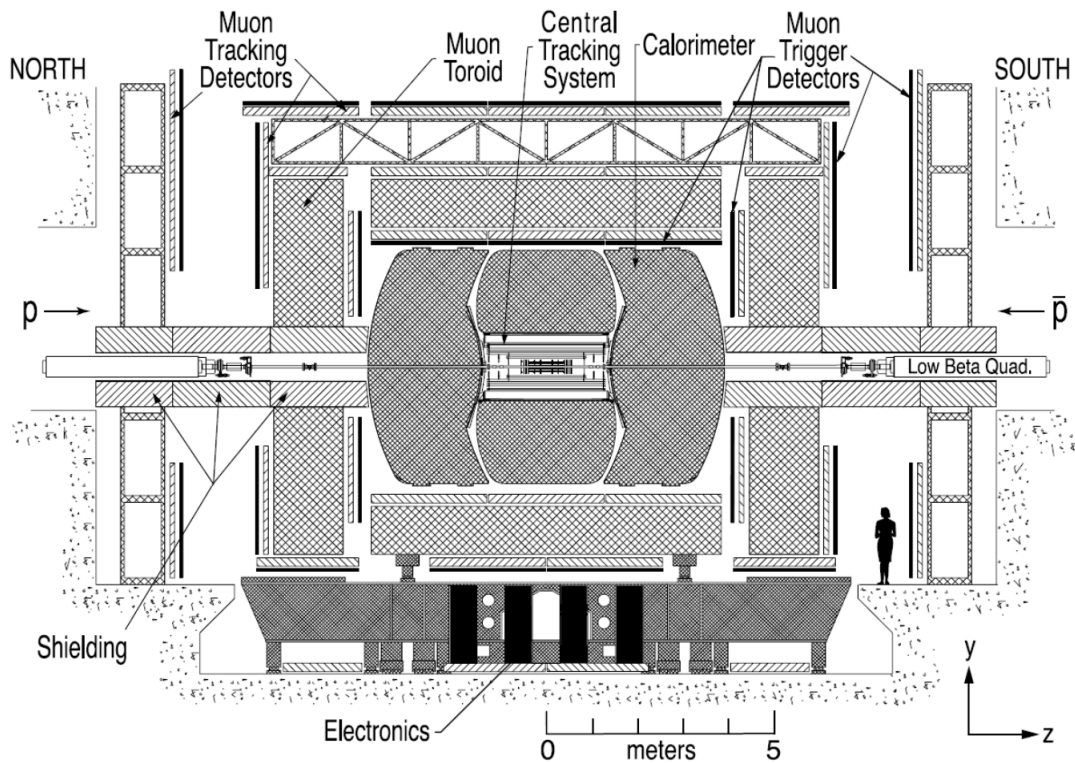


FIG. 2.5: Vue d'ensemble du détecteur DØ du RunII.

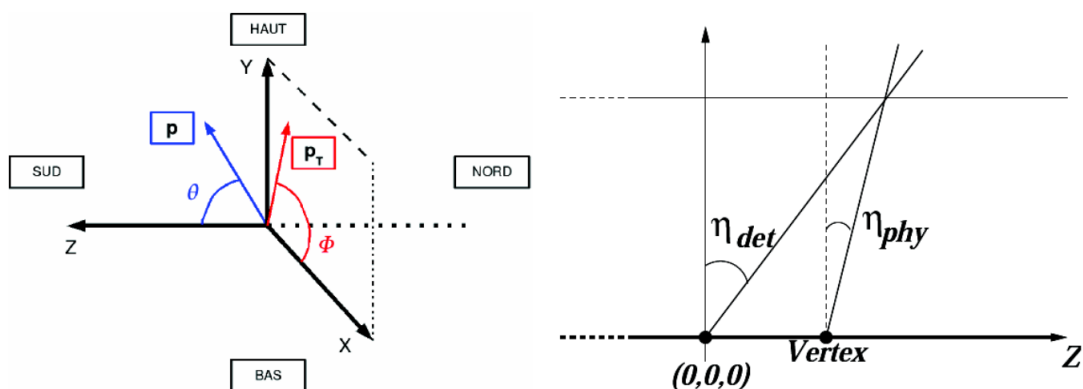


FIG. 2.6: Représentation des coordonnées utilisées dans le détecteur (à gauche) et des pseudo-rapidités (à droite).

Il sera nécessaire pour l'analyse d'utiliser les caractéristiques dans le plan transverse des

objets mesurés. En effet avec la loi de conservations de l'énergie, les quantités dans le plan orthogonal au faisceau de proton-antiproton sont conservées. C'est de cette manière que l'on évalue l'énergie transverse manquante (E_T). L'énergie E et l'impulsion P d'une particule dans le plan transverse sont définies par :

$$E_T = E \times \sin\theta, \quad P_T = P \times \sin\theta \quad (2.3)$$

Le Tevatron et l'expérience DØ ont connu plusieurs périodes de prises de données. Ces différentes périodes ont vu une augmentation de la luminosité instantanée L fournie par le Tevatron que l'on définit comme suit :

$$L(t) = \frac{f B N_p N_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)} \cdot F \quad (2.4)$$

avec f la fréquence de révolution des paquets, B le nombre de paquets des faisceaux, N_p ($N_{\bar{p}}$) le nombre de protons (antiprotons) des paquets, σ_p ($\sigma_{\bar{p}}$) la dimension transverse du paquet de protons (antiprotons) au point d'interaction et F est un facteur de forme.

On caractérise ainsi la luminosité intégrée :

$$\mathcal{L} = \int L(t) dt \quad (2.5)$$

La figure 2.7 montre l'évolution de la luminosité intégrée délivrée par le Tevatron et collectée par le détecteur DØ depuis avril 2002. A l'été 2010, l'expérience DØ a collecté 8 fb^{-1} de données avec une efficacité de plus de 90%. On distingue deux grandes périodes de prises de données le *RunIIa* (2002-2006) et le *RunIIb* (commencé depuis juin 2006) qui a vu le maximum de luminosité instantanée fournie par le Tevatron passer de 40.10^{30} à $400.10^{30} \text{ cm}^{-2}\text{s}^{-1}$.

L'information sur la luminosité intégrée nous permet d'obtenir le nombre d'évènements N attendus pour un processus ayant une section efficace σ :

$$N = \epsilon \cdot \mathcal{L} \cdot \sigma \quad (2.6)$$

où ϵ est le facteur correctif qui contient les efficacités expérimentales.

2.2.2 Le trajectographe interne

Le trajectographe (figure 2.8) est composé de sous-détecteurs qui sont les plus proches des faisceaux. Ils ont pour vocation de reconstruire les traces des particules et aussi d'identifier les vertex d'interaction des collisions. Ces deux caractéristiques sont essentielles pour l'identification des quarks b comme nous pourrons le voir dans les chapitres suivants.

SMT

Le premier de ces sous-détecteurs est le détecteur de traces à micropistes de Silicium SMT [31] (*Silicon Microstrips Tracker*). Il permet de reconstruire les traces des particules jusqu'à une valeur de η_{det} égale à 3. Le passage d'une particule génère une paire électron-trou

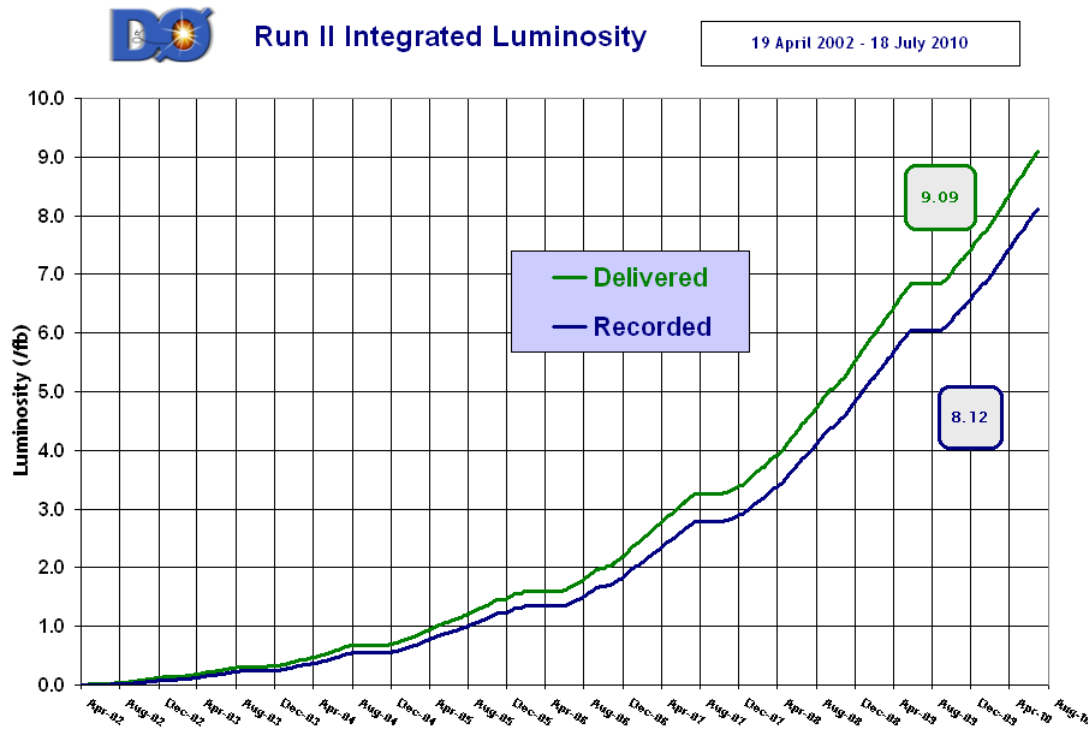


FIG. 2.7: Luminosité intégrée du *RunII* entre avril 2002 et l'été 2010. La courbe verte (supérieure) montre la luminosité intégrée délivrée par le Tevatron et la courbe bleue (inférieure) celle qui a été enregistrée par le détecteur DØ. Le plateau d'avril 2006 correspond à la transition du *runIIa* et *runIIb*. Les plateaux suivants séparent les périodes de prise de données que l'on notera *runIIb1*, *runIIb2* et *runIIb3*.

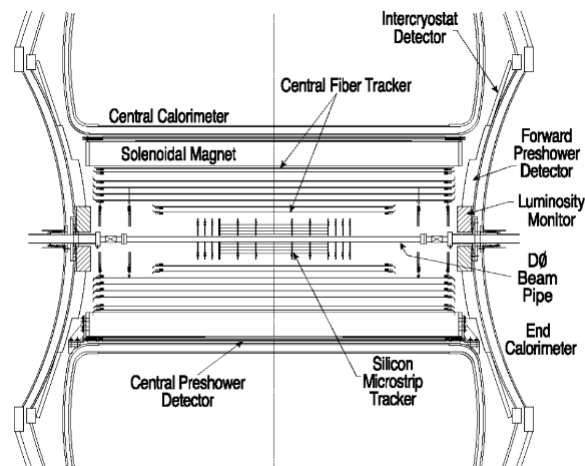


FIG. 2.8: Le trajectographe interne du détecteur DØ.

dans la partie active du SMT constituée de Silicium dopé. La structure du SMT est la suivante (figure 2.9) :

- trois barillets de rayon intérieurs et extérieurs 2.7 et 9.4 cm. Ces derniers sont constitués de quatre couches et sont longs de 12 cm. Afin de maximiser la couverture en ϕ , les couches comportent plusieurs modules répartis en quinconce (figure 2.10). Chaque couche comprend une série de modules sur la face interne et externe de la couche. Les couches deux et quatre (faces interne et externe) sont composées de capteurs double face dont les pistes forment un angle de 2° . Les modules face interne des couches un et trois sont composées de capteurs double face dont les pistes forment un angle de 90° . Alors que les modules face externe de ces mêmes couches sont composés de capteurs simple face. Les capteurs double face permettent la reconstruction en trois dimensions du point de passage de la particule ;
- douze disques F dont six sont placés au niveau des barillets ($|z| = 12.5, 25.3$ et 38.2 cm) et six sont placés après ($|z| = 43.1, 48.1$ et 58.1 cm). Ils ont un rayon interne et externe de 2.57 et 9.96 cm et sont constitués de capteurs de forme trapézoïdale double face dont les pistes forment un angle stéréo de 30° ;
- quatre disques H sont placés à $|z| = 110$ et 120 cm. Ils ont un rayon interne et externe de 9.5 et 26 cm et sont constitués de capteurs de forme trapézoïdale simple face. Ces capteurs sont montés dos-à-dos par paire et forment un angle stéréo de 15° . Ce sont ces disques qui permettent la reconstruction des traces ayant de grand η_{det} .

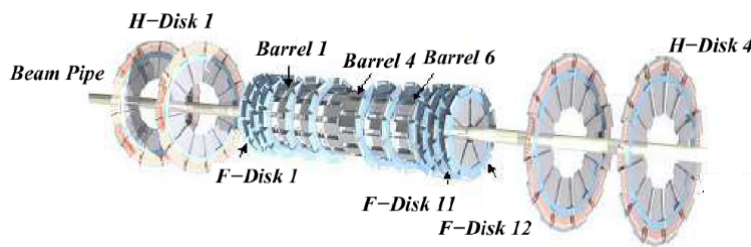


FIG. 2.9: Structure du SMT.

Les barillets et les disques sont soutenus par une structure en Beryllium. Afin d'assurer une température homogène dans le SMT et inférieure à 5° , on le refroidit par un mélange d'eau et d'éthylène glycol à 30% de -10° .

Le SMT a subi une amélioration lors du *RunIIb* par l'ajout de la couche 0 au plus près du faisceau [30]. Cette couche montée sur le tube à vide est composée de détecteurs simple face. Cela permet d'améliorer la précision sur la mesure du paramètre d'impact et la position des vertex secondaires.

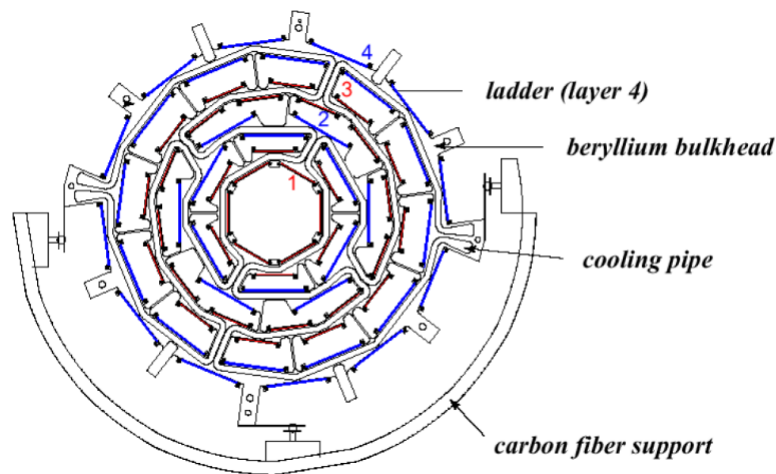


FIG. 2.10: Coupe transversale du SMT.

CFT

L'autre sous-détecteur utilisé pour la reconstruction des traces est le détecteur à fibre scintillante CFT [32] (*Central Fiber Tracker*). Le CFT est constitué de huit cylindres de fibres scintillantes concentriques ayant un rayon compris entre 20 et 52 cm. Les deux premiers cylindres en partant du centre ont une longueur de 1.66 m et les autres 2.52 m. Un cylindre comprend deux doublets de fibres alignés avec l'axe du faisceau et formant un angle stéréo de $\pm 3^\circ$. L'ensemble comprend 76800 fibres de 835 μm de diamètre. Une fibre est un mélange de polystyrène, de paraterphényl (1% en poids) et de 3-hydroxy-flavone (1500 ppm).

L'énergie déposée par le passage d'une particule dans la fibre est réémise en photons de longueur d'onde de 530 nm. On récupère les photons du côté de la fibre où se trouve un guide d'onde long de 7.8 à 11.9 m (l'autre côté de la fibre recouvert d'aluminium réfléchit les photons). Ce guide d'onde transmet les photons au compteur de photons visibles VLPC (*Visible Light Photon Counter*). Ce dernier est un photodétecteur à avalanche qui a une réponse rapide, une bonne efficacité quantique (supérieure à 80%) et un gain élevé.

L'aimant solénoïdal

Le dernier élément du trajectographe est l'aimant solénoïdal qui permet de mesurer l'impulsion et la charge des particules issues des collisions. C'est un aimant supraconducteur de 2.73 m de long et de 1.42 m de diamètre qui permet de créer un champ magnétique uniforme de 2T dans le trajectographe (figure 2.11). Il fonctionne à une température proche de 10 K et est refroidi avec de l'Helium liquide.

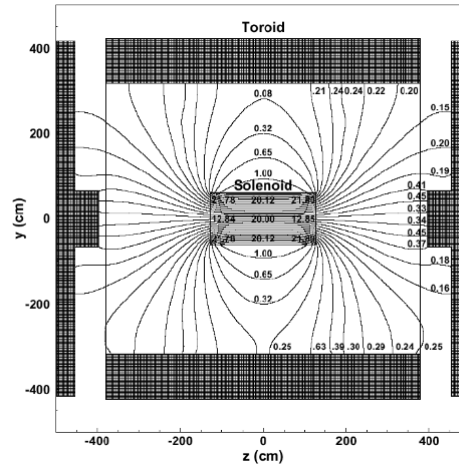


FIG. 2.11: Champ magnétique dans le détecteur DØ. On peut distinguer au centre le solénoïde et à l'extérieur le toroïde. Les valeurs de champ magnétique portés sont en kG.

2.2.3 Les détecteurs de pied de gerbe

L'aimant solénoïdal ayant une longueur de radiation équivalente à $0.8 X_0$, il permet aux gerbes de pouvoir se développer avant le calorimètre. Il est donc nécessaire d'avoir des détecteurs pied de gerbe (*Preshowers*) qui vont avoir le rôle intermédiaire de détecteur de traces et de calorimètre.

Ils sont constitués de couches de scintillateurs de section triangulaire permettant un recouvrement maximal. Ces scintillateurs sont composés de polystyrène (dopé avec 1% de p-terphényl et de 150 ppm de diphényl stilbène) et sont isolé optiquement. Ils sont reliés à des VLPC communs au CFT. On trouve deux types de détecteurs pied de gerbe : le CPS [33] (*Central Preshowers*) devant le calorimètre central et les FPS [34] (*Forward Preshowers*) devant les bouchons du calorimètre.

Le CPS correspond à la région centrale d'acceptance $|\eta_{\text{det}}| < 1.3$. La localisation de la gerbe se fait par l'intermédiaire de trois couches dont les angles par rapport au faisceau sont de 0 et $\pm 23^\circ$. Après avoir passé le CPS, les particules ont traversé une épaisseur de $2 X_0$ (4 pour les grands angles).

Les FPS couvre la région $1.5 < |\eta_{\text{det}}| < 2.5$ et sont fixés sur les parois du bouchon du calorimètre. Ils sont composés de deux couches de scintillateurs séparées par un absorbeur en plomb de $2 X_0$.

2.2.4 Le Calorimètre

Le calorimètre [35] a pour rôle d'identifier et de mesurer l'énergie des objets électromagnétiques (photons et électrons) et des jets.

Le calorimètre comprend trois parties cylindriques distinctes : une au centre CC (*Central*

Calorimeter) et deux bouchons EC (*End Cap*). Ces trois parties sont dans des cryostats séparés. Chaque partie comprend trois modules (*electromagnetic EM, fine hadronic FH et Coarse Hadronic CH*). Entre le CC et EC, il existe des zones inactives qui sont partiellement compensées avec les détecteurs inter-cryostats ICD (*InterCryostat Detector*) et MG (*Massless Gap*).

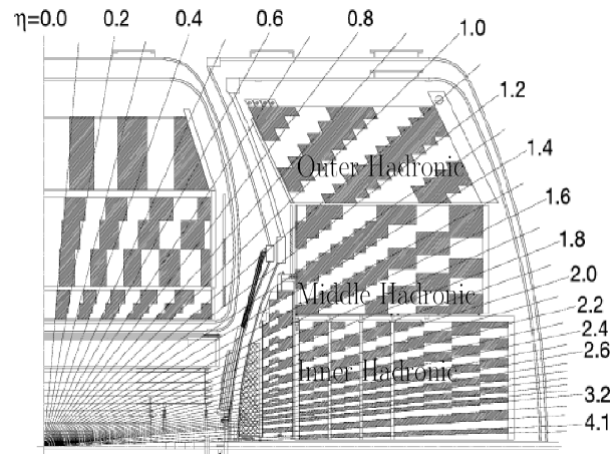


FIG. 2.12: Le calorimètre de DØ. L'arrangement des cellules permet de mettre en évidence la structure pseudo-projective du calorimètre.

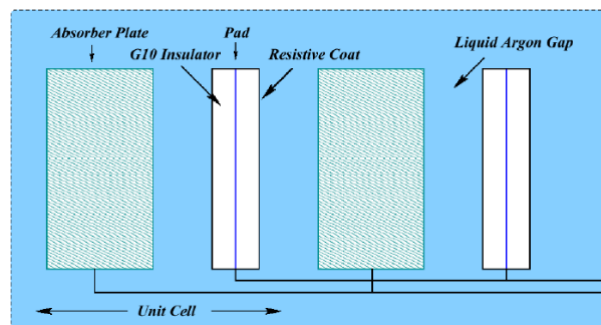


FIG. 2.13: Schéma d'une cellule du calorimètre.

Les cellules des trois modules sont agencées de tel sorte que la géométrie du calorimètre soit pseudo-projective en η_{det} (figure 2.12). Une cellule (figure 2.13) comporte une paire d'électrodes de lecture entre deux plaques d'absorbeurs, le tout baignant dans de l'Argon liquide.

Le Calorimètre Central CC

Les modules électromagnétiques du CC comprennent quatre couches qui sont nommés EM1, EM2, EM3 et EM4 (en s'éloignant du faisceau). Ces couches sont constituées d'absorbeurs en Uranium appauvri, couvrent la région en $|\eta_{\text{det}}| < 1.1$ et ont une longueur de

1.4, 2.0, 6.8 et 9.9 X_0 (respectivement). Leur granularité en $\eta_{\text{det}} \times \phi = 0.1 \times \pi/32$ (sauf EM3 : $0.05 \times \pi/64$).

Les couches hadroniques fines sont au nombre de trois et sont nommées FH1, FH2 et FH3. Elles sont constituées d'absorbeurs de 6 mm en Uranium appauvri/Niobium (2%). La couche hadronique grossière CH est quant à elle constituée d'absorbeurs de 46.5 mm en Cuivre. Les longueurs d'absorption des couches FHi (i croissant) sont de 1.3, 1.0 et 0.76 λ_A et pour la couche CH : 3.2 λ_A . La granularité des couches FHi et CH en $\eta_{\text{det}} \times \phi = 0.1 \times \pi/32$.

Les bouchons du Calorimètre EC

Ils sont complètement symétriques l'un par rapport à l'autre. Les modules électromagnétiques comprennent ici aussi quatre couches qui sont nommées EM1, EM2, EM3 et EM4 (en s'éloignant du centre du détecteur). Ces couches sont constituées d'absorbeurs en Uranium appauvri, couvrent la région $1.4 < |\eta_{\text{det}}| < 4$ et ont une longueur de 1.6, 2.6, 7.9 et 9.3 X_0 (respectivement). Leur granularité en $\eta_{\text{det}} \times \phi = 0.1 \times \pi/32$ (sauf EM3 : $0.05 \times \pi/64$).

La partie hadronique de EC est différente de CC. Comme nous le montre la figure 2.12, elle est composée de trois parties distinctes : intérieure (*Inner Hadronic*), intermédiaire (*Middle Hadronic*) et extérieure (*Outer Hadronic*).

La partie intérieure de EC contient quatre couches FH avec des absorbeurs en Uranium appauvri ($\lambda_A = 4 \times 1.1$) et la couche CH avec des absorbeurs en Inox ($\lambda_A = 4.1$). La partie intermédiaire est similaire à la partie intérieure, à la taille des absorbeurs près : $\lambda_A = 4 \times 0.9$ pour les couches FH et 4.4 pour la couche CH. La partie extérieure est uniquement composée de couches CH avec des absorbeurs en Inox ($\lambda_A = 6.0$).

Les détecteurs intercryostat ICD

Les détecteurs intercryostat sont placés sur les faces externes des cryostats des bouchons et couvrent la région $1.1 < |\eta_{\text{det}}| < 1.4$. Un ICD comprend 16 tuiles de granularité $\eta_{\text{det}} \times \phi = 0.3 \times 0.4$ et une tuile est composée de 12 scintillateurs. Ces derniers sont reliés à des photomultiplicateurs à l'aide de guides d'ondes.

Les *Massless Gap* MG

Les *Massless Gap* sont composés de cellules identiques au calorimètre et sont plongés dans l'Argon liquide des cryostats du CC ($0.8 < |\eta_{\text{det}}| < 1.2$) et EC ($1.0 < |\eta_{\text{det}}| < 1.3$). La granularité des cellules est de $\eta_{\text{det}} \times \phi = 0.1 \times \pi/32$.

2.2.5 Le spectromètre à muon

Ce système mesure la trajectoire et l'impulsion des muons dans la zone de pseudo-rapacité $|\eta_{\text{det}}| < 2$. Il est placé sur la partie externe du détecteur car les muons sont des particules qui interagissent très peu avec la matière.

Il est constitué de trois éléments :

- des aimants toroïdaux (figure 2.11) pour courber les trajectoires des muons (nécessaire pour la mesure de leur impulsion) ;
- des scintillateurs (figure 2.14 du bas) pour le déclenchement ;
- des chambres à dérive (figure 2.14 du haut) pour le déclenchement et les mesures.

Le spectromètre possède trois couches : notées A, B et C (en s'éloignant de centre du détecteur). Chaque couche est constituée de scintillateurs et de chambres à dérive. Finalement, les aimants toroïdaux sont placés entre les couches A et B.

Le spectromètre comprend une partie centrale WAMUS [36] (*Wide Angle MUon System*) et une partie avant et arrière FAMUS [37] (*Forward Angle MUon System*). Ils couvrent respectivement la région en pseudo-rapacité : $|\eta_{\text{det}}| < 1$ et $1 < |\eta_{\text{det}}| < 2$. Le WAMUS se compose de chambres à dérive PDT (*Proportional Drift Tubes*) et le FAMUS de tubes d'Iarocci MDT (*Mini Drift Tubes*). Le système de déclenchement utilise une (pour WAMUS) et trois (pour FAMUS) couches de scintillateurs. De plus, les fibres scintillantes qui entourent le WAMUS, permettent la réduction des bruits de fond cosmique.

2.2.6 Mesure de la luminosité

Cette mesure est effectuée par deux chaînes de 24 scintillateurs (*Luminosity Monitor* [38]) qui sont placés sur les bouchons du calorimètre (voir figure 2.8). Les scintillateurs couvrent la région en pseudo-rapacité $2.7 < |\eta_{\text{det}}| < 4.4$ et reçoivent les débris des collisions inélastiques qui permettent la mesure de la luminosité. Les données sont stockées en bloc de luminosité défini auxquels sont attribués des numéros identificateurs LBN (*Luminosity Block Number*).

2.2.7 Le système de déclenchement

Le grand nombre de collisions produites par le Tevatron et la faible section efficace des types d'évènements recherchés par DØ (boson de Higgs, particules au-delà du Modèle Standard, ...) ne permet pas de pouvoir enregistrer toutes les informations reçues. Il est donc nécessaire de les trier en intégrant un système de déclenchement (*trigger*). Il est bâti sur plusieurs niveaux ayant des critères de plus en plus restrictifs (voir figure 2.15) :

- niveau 0 : sélection d'évènements ayant une collision inélastique $p\bar{p}$. On les appelle évènements de biais minimum et sont mesurés par les détecteurs de luminosité. La fréquence de sortie du niveau 0 est de 1.7 MHz ;
- le niveau 1 : il utilise les informations de l'électronique de lecture du CFT, des détecteurs pied de gerbe, du calorimètre et du spectromètre à muon. La fréquence de sortie est réduite à 2 kHz pour un temps de décision de 3.5 μs ;
- le niveau 2 : il utilise les informations des pré-processeurs des détecteurs du niveau 1 auxquels on a rajouté celles du SMT. Cela permet d'obtenir des informations plus pré-

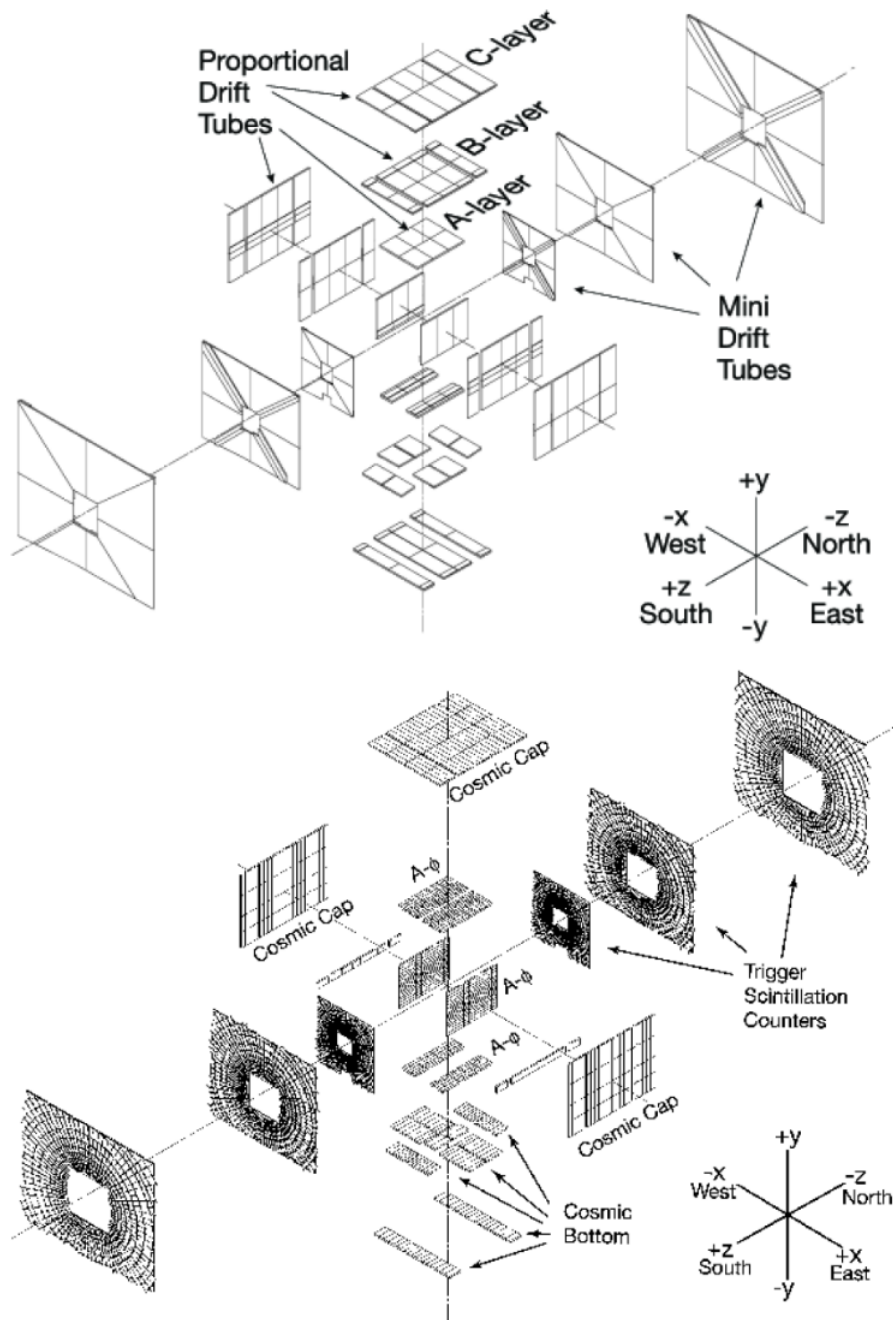


FIG. 2.14: Structure en couche des chambres à dérive (en haut) et des scintillateurs (en bas) à muon.

- cises sur les objets physiques (traces, muons, ...) dans les évènements. La fréquence de sortie est de 1 kHz pour un temps de décision de $100 \mu\text{s}$;
- le niveau 3 : il permet une reconstruction partielle des évènements issus du niveau 2 dans une ferme d'ordinateurs et ce à l'aide d'algorithmes. La fréquence de sortie est de

100 Hz ;

Les niveaux 1, 2 et 3 possèdent différents critères sur les événements considérés qui, une fois combinés forment le *trigger*. Il existe des *triggers* pour différentes topologies d'évènements. Le trigger utilisé dans l'analyse $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ est décrit dans la section 4.1.1.

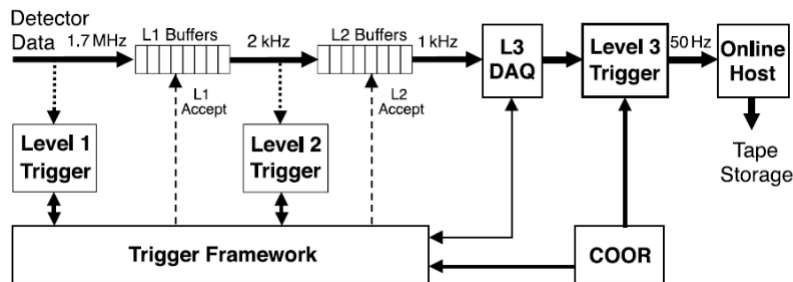


FIG. 2.15: Vue d'ensemble du système de déclenchement de DØ.

2.3 Objets physiques utilisés

2.3.1 Les traces et les vertex

Les traces

Les informations issues du CFT et SMT permettent la reconstruction des traces et suivant la région où se trouvent ces traces :

- $|\eta_{\text{det}}| < 1.7$: information maximale fournie par le trajectographe ;
- $1.7 < |\eta_{\text{det}}| < 2$: moins bonne reconstruction car la trace traverse moins de couches de CFT ;
- $|\eta_{\text{det}}| > 2$: seul SMT est utilisé.

Les vertex

Les vertex sont construits en utilisant des traces ayant $p_T > 0.5$ GeV et qui ont au moins deux coups dans le SMT. À l'aide de ces critères on définit des pseudo-vecteurs qui sont séparés d'au moins 2 cm selon l'axe z et on calcule la valeur du χ^2 associée à partir des traces. On retire les traces qui contribuent le plus au χ^2 du pseudo-vertex jusqu'à ce que le χ^2 par degré de liberté (χ^2/ddl) soit inférieur à 10. On crée finalement une liste de vertex primaires pour l'évènement avec les pseudo-vertex ayant au moins deux traces.

Finalement, on calcule pour chaque vertex la probabilité qu'il provienne d'une collision secondaire. La détermination de cette probabilité est basée sur l'impulsion transverse des traces. Par exemple, dans le cas d'une collision dure les traces auront des impulsions transverses plus

grandes que si elles provenaient d'une collision secondaire. Le vertex principal de l'évènement (PV_0) est celui qui a la plus petite probabilité.

2.3.2 Les objets électromagnétiques

Les objets électromagnétiques (EM) sont les électrons ou les photons γ . Ils ont principalement déposé leur énergie dans les couches électromagnétiques et FH1 du calorimètre.

On les reconstruit en utilisant l'algorithme de simple cône [40] qui procède de la manière suivante :

- recherche de la tour calorimétrique ayant le plus grand p_T (au moins supérieur à 0.5 GeV) afin de définir la direction en (η_{det}, ϕ) de référence ;
- on forme un amas de tours à partir des tours se trouvant dans un rayon $\Delta R = 0.4$ autour de la direction de référence. Cet amas a une valeur de p_T , η_{det} et ϕ qui lui est propre ;
- on répète ces deux opérations sur toutes les autres tours calorimétriques. Si les amas obtenus ont $p_T > 1.5$ GeV, ils seront candidats objets électromagnétiques.

Les critères suivants permettent de discriminer les objets électromagnétiques des gerbes hadroniques :

- la fraction d'énergie déposée dans la partie électromagnétique plus FH1, notée EM fraction ;
- isolation calorimétrique définie par :

$$\text{isolation} = \frac{E_{\text{Tot}}(\Delta R < 0.4) - E_{\text{EM}}(\Delta R < 0.2)}{E_{\text{EM}}(\Delta R < 0.2)} \quad (2.7)$$

avec $E_{\text{Tot}}(\Delta R < 0.4)$ (resp. $E_{\text{EM}}(\Delta R < 0.2)$) représente l'énergie déposée dans le calorimètre (resp. les couches électromagnétiques) dans un cône de rayon 0.4 (resp. 0.2) autour de l'objet électromagnétique ;

- matrice HMatrix8 définit le χ^2 d'une matrice à huit variables qui compare la forme de la gerbe de l'objet électromagnétique à la forme attendue pour un candidat électromagnétique. Les variables utilisées sont la position suivant z du vertex, la largeur des gerbes en η et ϕ , le logarithme de l'énergie et les fractions d'énergie déposées dans les quatre premières couches du calorimètre. La matrice HMatrix7 n'utilise pas la variable largeur en η ;
- vraisemblance électromagnétique (*Likelihood*) définie à partir de variables liées à l'objet électromagnétique.

On peut définir plusieurs critères de qualité de l'objet électromagnétique suivant les valeurs appliquées aux critères ci-dessus. On pourra différencier (en première approximation) un photon d'un électron en utilisant le fait que ce dernier ait une trace associée.

2.3.3 Les jets

L'hadronisation des quarks et des gluons crée des amas d'énergie (*jet*) dans le calorimètre (figure 2.16).

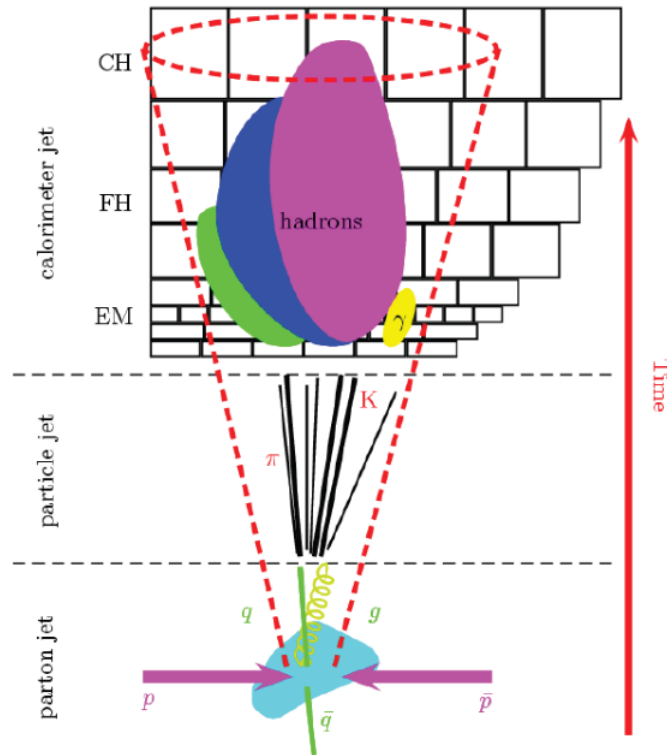


FIG. 2.16: Définition d'un jet selon le niveau de reconstruction.

La reconstruction d'un jet s'effectue en trois étapes. Tout d'abord on construit des tours calorimétriques à partir des cellules du calorimètre. À l'aide de ces cellules, on calcule le quadri-vecteur associé à la tour reconstruite défini par la somme des quadri-vecteurs des cellules (E_{cell} , \vec{p}_{cell}).

Ensuite on forme des pré-amas à partir des tours calorimétriques à l'aide de l'algorithme de simple cône (voir 2.3.2). Les tours sont classées par impulsion décroissante et on forme des pré-amas dans un cône de rayon $\Delta R = 0.3$. Ce pré-amas est conservé si son énergie est supérieure à 1 GeV et est déposée dans au moins deux tours calorimétriques.

Finalement on définit des proto-jets à partir des pré-amas à l'aide de l'algorithme de cône du *RunII* [41]. L'algorithme détermine si le pré-amas est inclus ou non dans un jet en comparant la distance entre les deux objets considérés et $\frac{R_{\text{cone}}}{2}$ ($R_{\text{cone}} = 0.5$ ou 0.7 , pour l'analyse présentée dans ce mémoire on utilisera 0.5). Si ce pré-amas n'est pas considéré comme étant à l'intérieur du jet, ce dernier sera considéré comme un nouveau jet. La reconstruction des jets utilise aussi comme point de départ des points appelés médians. Ils sont placés entre chaque paire de jets

reconstruits si la distance entre ces deux jets est comprise entre R_{cone} et $2R_{\text{cone}}$. L'algorithme de formation des jets (à partir des pré-amas ou des points médians) combine les tours qui sont dans le cône autour du point considéré. Lors de cette étape une tour peut être comprise dans plusieurs jets. Dans le cas où plus de la moitié de l'énergie un jet est comprise dans un autre jet, ces derniers sont combinés en un seul jet (*merging*) sinon on construit deux jets distincts (*splitting*). On gardera uniquement les jets reconstruits ayant une énergie supérieure à 6 GeV.

On identifie les jets reconstruits comme provenant d'un parton, en utilisant des coupures d'identification [42]. Les candidats ne passant pas ces coupures seront identifiés comme étant de "mauvais" jets. Quelques exemples de coupures d'identification :

- $0.05 < \text{EM fraction} < 0.95$, la borne inférieure rejette les candidats de bruits de fond et la borne supérieure rejette les candidats électromagnétiques ;
- limitation du bruit électronique dans les couches grossières (CH) du calorimètre : fraction d'énergie déposée dans les couches CH inférieure à 0.4 ;
- rejeter les tours bruyantes : au moins deux tours doivent contenir 90% de l'énergie du jet ;
- confirmation du jet au niveau 1 du système de déclenchement : le rapport p_T du jet et énergie des tours de déclenchement ($\Delta R = 0.5$ autour du centre du jet) doit être supérieur à 0.5. Les signaux de sortie du calorimètre sont envoyés au *trigger* de niveau 1 et à une électronique de lecture de précision. Cette dernière possède un grand gain et peut permettre à un bruit électronique d'apparaître comme un dépôt calorimétrique.

Afin de prendre en compte les perturbations dans la mesure des jets et donc de ramener cette énergie à la valeur du jet de particules, on calibre l'énergie de jets notée JES (*Jet Energy Scale*) [43] [44] définie par :

$$E_{\text{jet}}^{\text{part}} = \frac{E_{\text{jet}}^{\text{det}} - O}{R \times S} \quad (2.8)$$

avec :

- $E_{\text{jet}}^{\text{part}}$ l'énergie du jet au niveau particule ;
- $E_{\text{jet}}^{\text{det}}$ l'énergie mesurée ;
- O la correction de seuil, correspondant à l'énergie dans le cône du jet qui n'est pas associée à l'interaction dure (par exemple la radiation résiduelle de l'Uranium dans le calorimètre). Cette correction varie avec le rayon du cône et η_{det} du jet ;
- R la réponse corrigée du calorimètre, prend en compte la perte d'énergie dans certaines parties du calorimètre. Cette correction varie avec l'énergie et η_{det} du jet ;
- S la correction de gerbe, prend en compte la fraction d'énergie déposée en dehors du cône par les particules issues du jet de particules, mais aussi l'énergie déposée dans le cône par des particules qui n'appartiennent pas au jet de particules.

On peut prendre en compte un dernier effet pour cette calibration en énergie, les désintégrations semi-muoniques pour définir JESMU [45]. Le muon n'interagissant que très peu avec

le détecteur va créer de l'énergie transverse manquante. Cette correction est importante pour les jets issus de quark b car ce type de désintégration est présent dans 10% des désintégrations de quarks b.

2.3.4 Les muons

Les spectromètres à muon et les détecteurs de traces permettent la reconstruction des muons jusqu'à $\eta_{\text{det}} < 2$. Suivant le nombre de coups comptés dans le spectromètre à muons : n_{seg} , on crée deux définitions de muon [46] : *Loose* ou *Medium*.

- $n_{\text{seg}} = 1$: coup à l'intérieur du toroïde (couche A) ;
- $n_{\text{seg}} = 2$: coup à l'extérieur du toroïde (couche B/C) ;
- $n_{\text{seg}} = 3$: coup dans les couches A et B/C ;
- $n_{\text{seg}} > 0$: si la trace reconstruite dans le spectromètre à muons correspond à la trace reconstruite dans les détecteurs de traces. Si ce n'est pas le cas, $n_{\text{seg}} < 0$.

Si $n_{\text{seg}} = \pm 3$:

- le muon est *Medium* si :
 - au moins 2 coups dans les chambres à dérive de la couche A ;
 - un coup dans les scintillateurs de la couche A ;
 - au moins 2 coups dans les chambres à dérive de la couche B/C ;
 - au moins 1 coup dans les scintillateurs de la couche B/C ;
- le muon est *Loose* si : une des conditions ci-dessus n'est pas respectée.

Si $n_{\text{seg}} = 2$:

- le muon est *Loose* si :
 - au moins 1 coup dans les scintillateurs de la couche B/C ;
 - au moins 2 coups dans les chambres à dérive de la couche B/C ;
- le muon est *Medium* si : les conditions ci-dessus sont respectées et si le muon est dans la partie basse du détecteur.

Si $n_{\text{seg}} = 1$:

- le muon est *Loose* si :
 - un coup dans les scintillateurs de la couche A ;
 - au moins 2 coups dans les chambres à dérive de la couche A ;
- le muon est *Medium* si : les conditions ci-dessus sont respectées et si le muon est dans la partie basse du détecteur.

Les muons ont aussi trois définitions de traces :

- *Loose* : si la trace n'a pas de coups dans le SMT, on demande une distance minimale d'approche $d_{\text{ca}} < 0.2$ cm sinon $d_{\text{ca}} < 0.02$ cm ;
- *Medium* : si c'est une trace *Loose* et χ^2/ddl de la trace est inférieure à 4 ;
- *Tight* : si c'est une trace *Medium* et qu'au moins un coup dans le SMT est attaché à la trace.

De plus, il existe les critères d'isolation suivant :

- CAL isolation : la somme des énergies transverses des cellules dans un cône de rayon 0.4 autour de la direction du muon divisé par le p_T du muon, auquel on soustrait l'énergie calorimétrique dans un cône de rayon 0.1 autour de la direction du muon. Ce calcul ne prend pas en compte l'énergie des couches grossières CH du calorimètre ;
- track isolation : la somme scalaire des p_T des traces à l'intérieur d'un cône de rayon 0.5 autour de la trace associée au muon divisé par le p_T du muon.

2.3.5 L'énergie transverse manquante

Un neutrino n'interagit pas avec le détecteur. La seule possibilité de voir l'énergie emportée par le neutrino est d'utiliser le principe de conservation de l'énergie dans le plan transverse. L'énergie du neutrino est donc la contribution manquante du bilan énergétique : l'énergie transverse manquante [47], notée \cancel{E}_T :

$$\vec{\cancel{E}}_T = - \sum^{\text{cellules}} \vec{p}_T \quad (2.9)$$

Elle dépend des corrections apportées par la calibration des jets, il faut donc lui appliquer les corrections de JES. Dans le cas de JESMU, l'impulsion du muon est mesurée par le trajectographe, il faut donc soustraire le p_T du muon du calcul de \cancel{E}_T .

2.4 La chaîne de simulation

La simulation permet de vérifier les prédictions théoriques dans nos analyses en les confrontant aux données collectées. La simulation des données se fait en plusieurs étapes : génération du processus à étudier (plusieurs possibilités : ALPGEN [48], PYTHIA [49], COMPHEP [50]), hadronisation des partons (PYTHIA), interaction des particules avec le détecteur (d0gstar) dont la simulation est basée sur GEANT-3 [51]. Les données et les simulations passent par la même chaîne de reconstruction des événements (d0reco).

2.5 Qualité des données

Des critères sont définis afin de sélectionner les événements dont la qualité est jugée bonne.

Le premier critère consiste à faire une sélection des événements par période de prise de données en enlevant les périodes de disfonctionnement du détecteur. Par exemple les périodes durant lesquelles se déclare une cellule chaude dans le calorimètre (dépôt d'énergie moyen important) sont facilement identifiables.

Il est aussi possible de faire une sélection d'un événement ayant un problème connu. Le bruit appelé *Spanish fan* est un exemple de l'un des principaux bruits identifiés au niveau du calorimètre. Il apparaît sur les cellules du calorimètre qui sont dans la région $0.7 < |\eta_{\text{det}}| < 0.8$

en formant deux demi-roues. Il crée de "faux" jets et de l'énergie transverse manquante. On l'identifie à l'aide des fractions électromagnétiques et hadroniques des jets. Entre juin 2006 et l'automne 2007, ce bruit a constitué 0.0001% des évènements de biais minimum.

Discrimination des jets de quark b avec des muons de basse impulsion transverse

De nombreux canaux de physique possèdent un quark b dans l'état final (Higgs, top, nouvelle physique). Afin d'augmenter la sensibilité des analyses, on peut discriminer les quarks b des quarks légers grâce à des méthodes multivariées.

Ce chapitre rappelle les principes de base des techniques d'identification des quarks b (*b-tagging*). Nous décrivons en détail un nouvel outil de *b-tagging* développé au cours de la thèse basé sur l'identification des muons de basse impulsion (SLTNN).

3.1 Préambule : NN

L'expérience DØ a développé une méthode d'étiquetage de jets issus de quarks b (*b-tagging* en anglais) reposant sur des réseaux de neurones (*Neural Network* en anglais) et les caractéristiques des jets de quarks b. Cette méthode d'étiquetage sera nommée par la suite NN [52].

Le quark b a une durée de vie plus grande que les quarks légers. Cela permet aux hadrons B de "voler" sur une distance de l'ordre de 3 mm pour une impulsion de 40 GeV avant désintégration, ce qui induit l'existence d'un vertex secondaire déplacé (voir figure 3.1) et de traces issues de vertex ayant un grand paramètre d'impact¹ par rapport au vertex d'interaction. Les caractéristiques du sous-détecteur SMT (présenté dans la section 2.2.2) rendent possible la mesure de ce vertex secondaire. On identifie un jeu de variable qui exploite ces propriétés et permettent de discriminer les jets issus de quarks b de ceux issus des jets de quarks légers (u, d et s ou de gluon). Ces variables seront utilisées comme variables d'entrée d'un réseau de neurones qui renvoie un nombre compris entre 0 et 1. Plus ce nombre est proche de 1, plus il

¹le paramètre d'impact est la distance minimale entre le vertex primaire estimé et la trajectoire de la trace

est probable que le jet étudié soit issu d'un quark b .

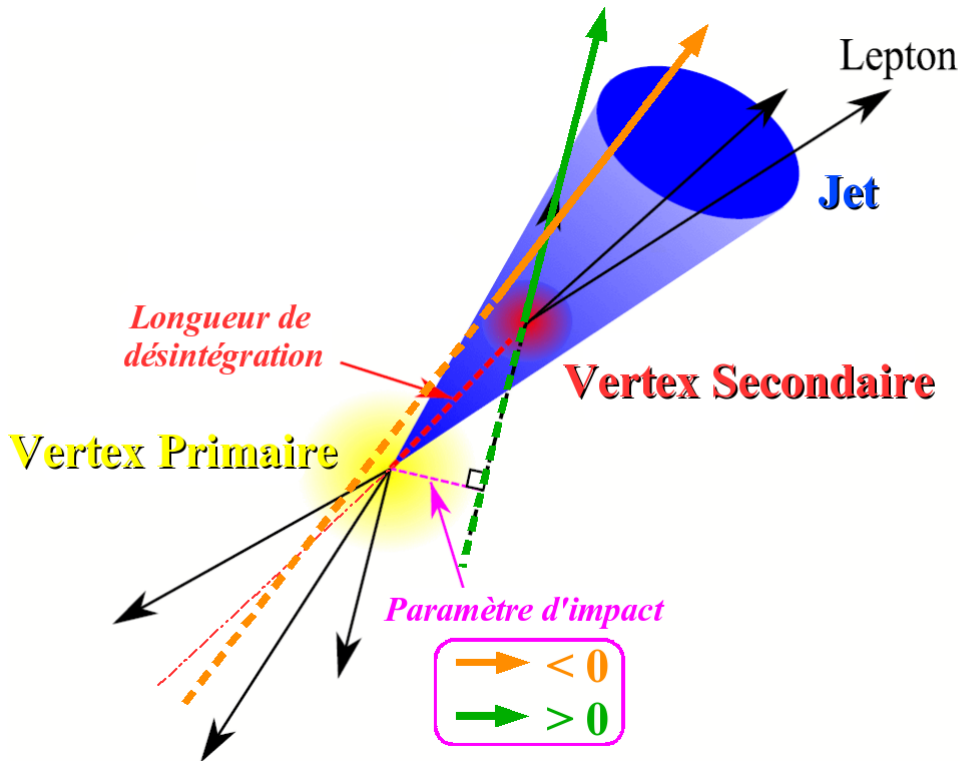


FIG. 3.1: Topologie d'un évènement ayant un quark b . Le hadron B "vole" avant de se désintégrer (longueur de désintégration en pointillé rouge) et entraîne l'existence d'un vertex secondaire déplacé. On peut voir un exemple de traces associées au jet ayant un paramètre d'impact négatif (trace orange) ou positif (trace verte).

3.1.1 Variables discriminantes du réseau de neurones

Les variables discriminantes utilisées en entrée du réseau de neurones reposent sur trois algorithmes d'étiquetage de quark b :

- JLIP [53] (*Jet Lifetime Probability*) utilise l'information sur le paramètre d'impact des traces dans les jets pour définir une unique variable qui représente la probabilité que toutes les trace du jet soient issues du vertex primaire. Plus cette probabilité est proche de 0, plus la probabilité que le jet soit issu d'un quark b est grande ;
- CSIP [54] (*Counting Signed Impact Parameter*) utilise l'information sur le paramètre d'impact en signifiacance² (noté IPsig). Un jet sera étiqueté comme étant issu d'un quark b

²le paramètre d'impact en signifiacance représente le rapport entre le paramètre d'impact et son erreur. Cette grandeur donne une information sur la résolution du paramètre d'impact.

s'il contient au moins 2 traces dont $IP_{sig} > 3$ ou s'il contient au moins 3 traces dont $IP_{sig} > 2$;

- SVT [55] (*Secondary Vertex Tagging*) repose sur le long temps de vol des hadrons B. Les vertex secondaires reconstruits doivent correspondre avec le jet calorimétrique reconstruit ($\Delta R(\text{jet}, \text{vertex secondaire}) < 0.3$). On peut ainsi définir la longueur de désintégration en signification (*Decay Length Significance* noté DLS) du vertex secondaire qui est utilisé pour étiqueter le jet. Le critère d'étiquetage le plus sévère (*Tight*) de SVT demande $DLS > 0.7$.

Le réseau de neurones de l'algorithme NN utilise les 7 variables suivantes en entrée (figure 3.2) :

- JLIPprob : probabilité que le jet provienne du vertex primaire (vu précédemment) ;
- CSIPcomb : combinaison pondérée de IP_{sig} des traces :

$$CSIP_{comb} = 6 \times CSIP_{3s} + 4 \times CSIP_{2s} + 3 \times CSIP_{3w} + 2 \times CSIP_{2w} \quad (3.1)$$

$CSIP_n a$ représente le nombre de traces dont $IP_{sig} > n$. Si $a = w$ (resp. s), on prend en compte les traces qui ont $IP_{sig} < 0$ (resp. > 0) et on demande (uniquement si $a = w$) $\Delta\Phi(\text{traces}, \text{jet}) < 1.15$ (angle azimutal entre les traces du jet et le jet considéré). $CSIP_n a$ est une variable discrète répartie sur une petite échelle de valeur. Ceci ne permet pas une utilisation optimale du réseau de neurones. Avec $CSIP_{comb}$, on a construit une variable unique qui a une plus grande échelle de valeur ;

- $SVT_{SL}DLS$: longueur de désintégration en signification du vertex secondaire ;
- $SVT_{SL}X_{ddl}^2$: χ^2 par degré de liberté du vertex secondaire ;
- SVT_{LN}_{Tracks} : nombre de traces utilisées pour reconstruire le vertex secondaire ;
- $SVT_{SL}Mass$: masse du vertex secondaire ;
- $SVT_{SL}Num$: nombre de vertex secondaires reconstruits pour le jet considéré.

Pour les variables liées au vertex secondaire, la mention SL (*SuperLosse*) ou L (*Loose*) représente le critère d'étiquetage utilisé pour l'algorithme SVT. Il est utile pour le réseau de neurones d'avoir le maximum d'information possible, les critères sont donc choisis de manière à être les moins restrictifs possibles. Le critère *SuperLosse* requiert $DLS > 0$ et *Loose* $DLS > 5$. Le critère *Loose* a été préféré pour la variable SVT_{LN}_{Tracks} afin d'augmenter son pouvoir discriminant (comme illustré par la figure 3.3).

3.1.2 Structure et fonctionnement du réseau de neurones

Ce jeu de variables défini, nous pouvons les utiliser en entrée d'un réseau de neurones (voir figure 3.4). Un réseau de neurones permet de définir une nouvelle variable discriminante (sa valeur de sortie) qui est combinaison linéaire de fonction des variables d'entrée. La valeur de sortie est continue et est comprise entre 0 et 1. Plus cette est valeur est proche de 1 et plus le jet considéré est identifié comme étant issu d'un quark b.

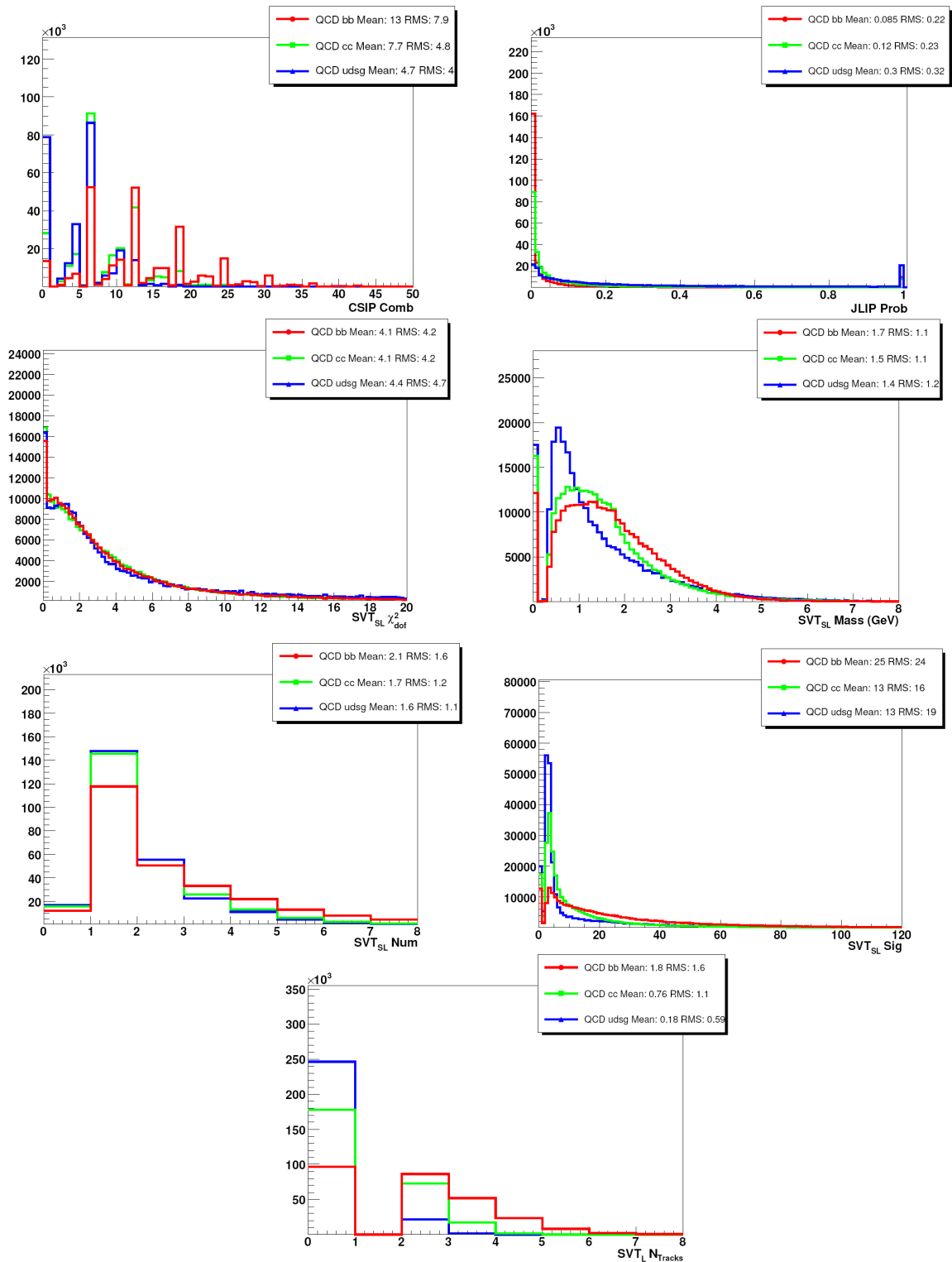


FIG. 3.2: Variables d'entrée du réseau de neurones de l'algorithme NN [57] dans des lots de simulation. De haut en bas et de gauche à droite : CSIPcomb, JLIPprob, SVT_{SL}X²_{ddl}, SVT_{SL}Mass, SVT_{SL}Num, SVT_{SL}DLS et SVT_{SL}N_{Tracks}. Le bruit de fond (courbe bleue) correspond à des jets de quarks u,d, s et gluons. Le signal (courbe rouge) correspond à des jets de quark b. La courbe verte montre les résultats correspondants aux jets de quarks c.

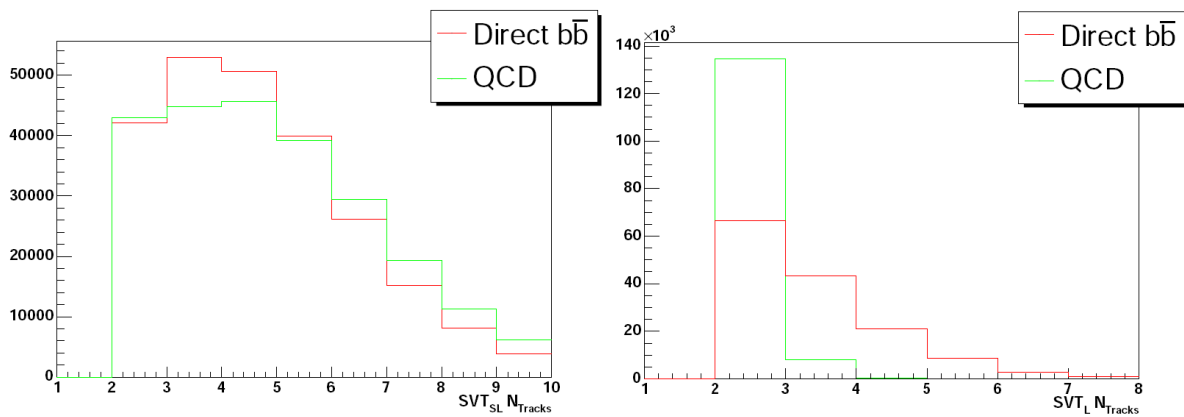


FIG. 3.3: Nombre de traces utilisées pour reconstruire le vertex secondaire lorsque l'on utilise l'algorithme SVT avec le critère SL (à gauche) ou L (à droite) sur des événements simulés [52]. Le bruit de fond (courbe verte) correspond à des jets de quarks u,d, s et gluons. Le signal (courbe rouge) correspond à des jets de quark b.

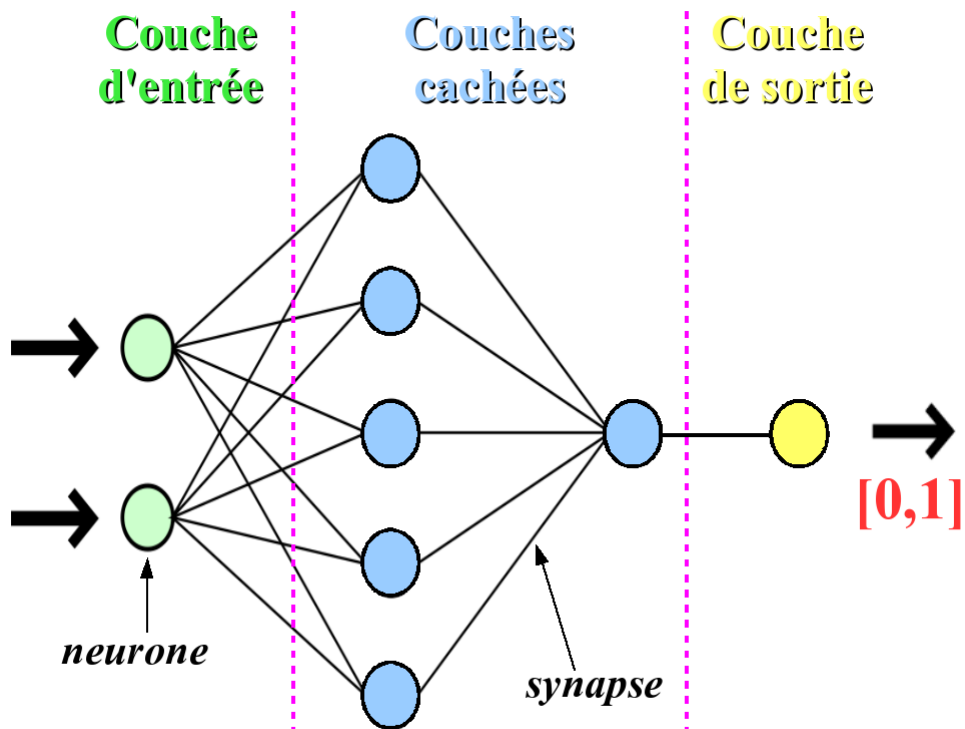


FIG. 3.4: Structure d'un réseau de neurone.

L'entraînement du réseau de neurones a pour objectif de déterminer les coefficients de pondération de la combinaison linéaire. Il est constitué de neurones répartis sur différentes couches : la couche d'entrée composée d'autant de neurones que de variables d'entrée, une ou plusieurs couches cachées et finalement une couche de sortie (dans notre cas un seul neurone). Le nombre de neurones et de couches cachées sont des paramètres libres. La configuration optimale trouvée est la suivante : deux couches cachées, dont la première comprend 24 neurones et la seconde 1 neurone (ce dernier permet juste de normaliser la valeur de sortie du réseau de neurones entre 0 et 1). La liaison entre chaque neurone d'une couche et les neurones de la couche précédente est appelée synapse. Cette synapse correspond à un poids qui est appliqué à la valeur de sortie du neurone de la couche précédente. Ce poids est initialisé par un nombre aléatoire compris entre -0.5 et 0.5.

Le réseau de neurones est entraîné et testé avec des événements simulés de signal (sortie attendue égale à 1) et de bruit de fond (sortie attendue égale à 0). Dans notre cas, le signal est modélisé par un processus du type $q\bar{q} \rightarrow g \rightarrow b\bar{b}$ et le bruit de fond par un processus de type QCD (on construit une collection de quarks u, d et s et de gluons). Chacun des lots d'événements est préalablement séparé en deux sous-lots servant distinctement à l'entraînement et au test du réseau de neurones. Lorsque tous les événements des deux lots sont utilisés, on a effectué ce que l'on appelle une époque. Le poids des synapses sont réévalués au cours de chaque époque afin de minimiser la différence entre la sortie du réseau et la valeur attendue. A chaque époque, l'ordre des événements dans le lot est redéfini de manière aléatoire. Le nombre d'époque est un paramètre libre et la valeur retenue est égale à 400 après optimisation. L'annexe A fournit des compléments d'informations sur le fonctionnement des réseaux de neurones.

3.1.3 Résultats du réseau de neurones

L'algorithme NN a été construit pour les données du *RunIIa* et du *RunIIb* indépendamment et la figure 3.5 montre la sortie obtenue pour ces deux cas. On peut voir une très nette séparation du signal et du bruit de fond. En coupant sur cette sortie, 12 points de fonctionnement notés OP (pour *Operating Point*) ont été certifiés par la collaboration. Ces OP sont résumés dans le tableau 3.1.

OP	<i>MegaTight</i>	<i>UltraTight</i>	<i>VeryTight</i>	<i>Tight</i>	<i>Medium</i>	<i>oldLoose</i>
Coupure	>0.925	>0.9	>0.85	>0.775	>0.65	>0.5
OP	<i>Loose</i>	<i>L2</i>	<i>L3</i>	<i>L4</i>	<i>L5</i>	<i>L6</i>
Coupure	>0.45	>0.325	>0.25	>0.2	>0.15	>0.1

TAB. 3.1: Définition des 12 points de fonctionnement (OP) de l'algorithme de *b-tagging* NN. Les valeurs correspondent aux coupures sur la sortie du réseau de neurones.

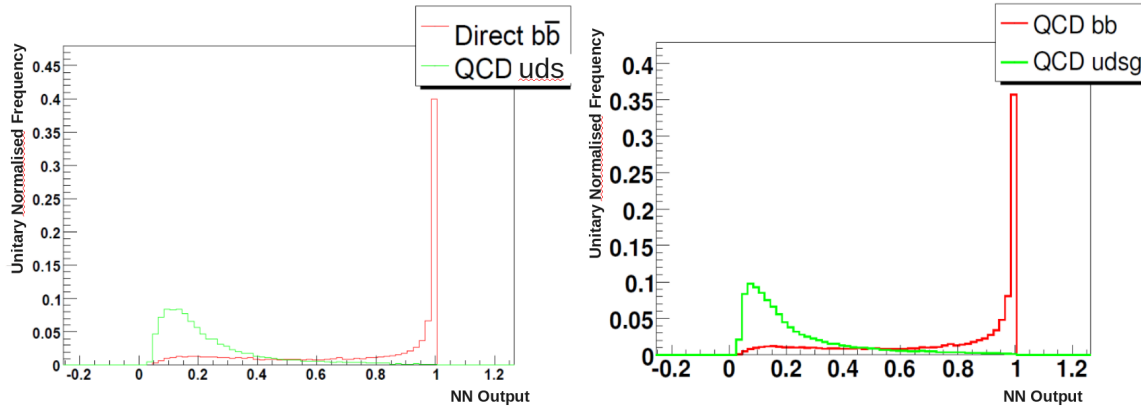


FIG. 3.5: Sortie du réseau de neurones de l’algorithme de b -tagging NN dans le cas où il est entraîné avec des évènements simulés correspondant au RunIIa [56] (à gauche) ou au RunIIb [57] (à droite). Le bruit de fond (courbe verte) correspond à des jets de quarks u, d, s et gluons. Le signal (courbe rouge) correspond à des jets de quark b .

La figure 3.6 montre l’efficacité d’étiquetage (b -Jet Efficiency) en fonction du taux de mauvaise identification ($Fake Rate$) pour chaque points de fonctionnement de NN. La figure de gauche compare les résultats obtenus avec la simulation (signal : $Z \rightarrow b\bar{b}$ et bruit de fond : $Z \rightarrow q\bar{q}$) dans les cas où le réseau de neurones a été entraîné pour les données du *RunIIa* (p17) ou pour les données du *RunIIb* (p20). On remarque que les performances sont identiques. La figure de droite compare les résultats obtenus avec JLIP et NN pour les données du *runIIa*. On peut voir que dans le meilleur des cas :

- à taux de mauvaise identification fixée : NN améliore l’efficacité d’étiquetage de 46% ;
- à efficacité d’étiquetage fixée : NN diminue le taux de mauvaise identification de 28%.

Les méthodes utilisées pour définir l’efficacité d’étiquetage, le taux de mauvaise identification et les erreurs systématiques seront expliqués dans la partie SLTNN. En effet pour des raisons de cohérence SLTNN a été développé de la même manière que NN.

3.2 Motivations de la méthode SLTNN

La méthode d’étiquetage SLTNN (*Soft Lepton Tagging Neural Network*) repose sur la désintégration semi-leptonique du hadron B . L’identification d’un électron à l’intérieur d’un jet demande à savoir distinguer deux objets calorimétriques et est donc plus délicate. Le nouvel algorithme d’étiquetage SLTNN ne prendra en compte que les muons.

Dans 20% des cas un jet issu de quark b contiendra un muon, comme le montrent les rapports de branchement : $b \rightarrow \mu X = 10.95\%$ et $b \rightarrow c \rightarrow \mu X = 9.58\%$ [58]. De plus, on impose les conditions suivantes sur le jet :

- rayon d’ouverture du cône $\Delta R = 0.5$;
- $p_T > 15 \text{ GeV}$;

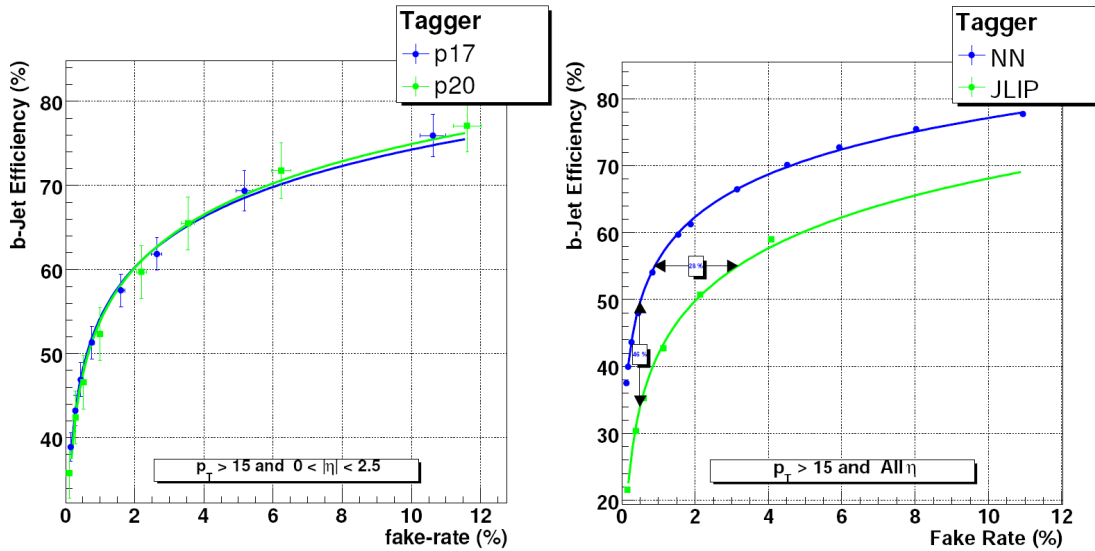


FIG. 3.6: Efficacité d'étiquetage (*b-Jet Efficiency*) en fonction du taux de mauvaise identification (*Fake Rate*) pour chaque point de fonctionnement de NN. La figure de gauche [57] compare les résultats obtenus avec la simulation (signal : $Z \rightarrow b\bar{b}$ et bruit de fond : $Z \rightarrow q\bar{q}$) dans les cas où le réseau de neurone a été entraîné pour les données du *RunIIa* (p17, courbe bleue) ou pour les données du *RunIIb* (p20, courbe verte). La figure de droite [56] compare les résultats obtenus avec JLIP (courbe verte) et NN (courbe bleue) pour les données du *runIIa*. Les barres d'erreur comprennent les erreurs statistiques et systématiques.

- $|\eta| < 2.5$;
- et être étiquetable : $p_T > 15$ GeV et au moins deux traces dans le jet ayant produit au moins trois coups dans le CFT. Pour la partie centrale du calorimètre (CC) on demande aux traces d'avoir produit au moins sept coups dans le CFT.
- confirmé par les vertex (c.à.d. au moins 2 traces associées avec le vertex primaire);
- corrigé de l'échelle d'énergie des jets JESMU;
- rejet des candidats électrons et "mauvais" jets (voir section 2.3.3).

Le vertex primaire est reconstruit avec l'algorithme *Adaptive PV algorithm* (section 2.3.1).

Afin d'augmenter la bonne qualité du vertex reconstruit, on demande également :

- $|z_{PV}| < 60$ cm (zone d'acceptance totale du SMT);
- au moins trois traces attachées au vertex.

Finalement on demande aux muons :

- $p_T > 4$ GeV (*Soft Lepton*);
- $|\eta| < 2.0$;
- $\Delta R(\text{muon}, \text{jet}) < 0.5$: jet contenant (au moins) un muon que l'on appellera par la suite jet "muonique" ou *Soft Lepton Tagging* (SLT pour des muons).

On utilisera trois qualités de muon différentes dont les définitions certifiées [46] ("*Loose*" et "*Medium*") sont fournies par le groupe dédié à leur identification :

- *loose* : muon de définition "*Loose*" ayant $n_{seg} > 0$;
- *med* : muon de définition "*Medium*" ayant $n_{seg} > 0$;
- *med3* : muon de définition "*Medium*" ayant $n_{seg} = 3$.

La variable n_{seg} est basée sur le nombre de coups dans les différentes couches du système de détection des muons (voir la section 2.2.5) :

- $n_{seg} = 1$: coup à l'intérieur du toroïde (couche A) ;
- $n_{seg} = 2$: coup à l'extérieur du toroïde (couche B/C) ;
- $n_{seg} = 3$: coups dans les couches A et B/C.

La correspondance entre la trace (de qualité *Loose* voir section 2.3.4) dans le détecteur de traces et de muon doit avoir $\chi^2 < 100$.

3.3 Procédure

La méthode d'étiquetage SLTNN repose sur deux étapes : on demande au jet de contenir un muon et ensuite on entraîne un réseau de neurones spécifique à cette méthode. Les résultats relatifs à ces deux étapes seront développés à l'aide d'évènements simulés. Ensuite nous comparerons les performances de SLTNN dans les données et la simulation.

Nous allons tout d'abord créer une collection de jets ayant les caractéristiques présentées dans la section 3.2 (la condition de muon dans les jets n'est pas vérifiée à ce stade), que l'on appellera $N_{jets}^{incl.}$.

3.3.1 Lots de données

Nous utilisons les données du *RunIIb* enregistrées entre juin 2006 et septembre 2008. Deux lots sont définis sur ces données : *MUinclusive* (ensemble de conditions de *triggers* demandant des évènements ayants au moins 1 muon) et QCD (ensemble de conditions de *triggers* demandant des évènements ayants au moins 1 jet).

Les évènements qui ont déclenché au niveau 3 les *triggers* basés sur l'identification de traces ayant un grand paramètre d'impact ne seront pas utilisés afin d'éviter un biais provenant de la sélection en ligne sur l'estimation de l'efficacité de *b-tagging* dans les données.

Après avoir passé toutes ces conditions, le lot *MUinclusive* contient 398 millions d'évènements et le lot QCD 19 millions.

3.3.2 Lots simulés

Les différents processus utilisés durant le développement de SLTNN sont résumés dans le tableau 3.2. Les processus associés aux jets issus de quarks b ou c sont :

- direct $b\bar{b}$: $q\bar{q} \rightarrow g \rightarrow b\bar{b}$ ¹ ;
- $t\bar{t}$: $q\bar{q} \rightarrow g \rightarrow t\bar{t}$;

¹produit avec le générateur PYTHIA

- bb+jets : $b\bar{b} + i \rightarrow b\bar{b} + ii^2$ où $i = 0, 1$ ou 2 partons légers et $gg \rightarrow b\bar{b} h \rightarrow b\bar{b}b\bar{b}$ (MSSM);
 - direct $c\bar{c}$: $q\bar{q} \rightarrow g \rightarrow c\bar{c}$;
 - cc+jets : $c\bar{c} + i \rightarrow c\bar{c} + i$ où $i = 0, 1$ ou 2 partons légers ;
- et les processus pour le bruit de fond (des jets issus de quarks u, d, s ou gluons) :
- QCD $q\bar{q}$: $q\bar{q} \rightarrow q\bar{q}$;
 - γ +jets : $qg \rightarrow \gamma q$.

Nom du processus simulé	$N_{\text{évènements}}$	$N_{\text{jets étiquetables}}$
$t\bar{t}$	24 798 732	82 000 587
bb+jets	7 630 338	67 901 965
cc+jets	5 120 553	46 269 135
direct $b\bar{b}$	4 572 668	7 840 058
direct $c\bar{c}$	1 008 570	1 737 759
QCD $q\bar{q}$	11 057 360	20 186 812
γ +jets	3 621 779	3 859 520

TAB. 3.2: Liste de processus simulés utilisés lors du développement de SLTNN avec le nombre d'évènements et de jets étiquetables.

3.4 Etiquetage de muons dans les jets

On définit trois types de jets contenant des muons à partir des trois qualités de muons définies dans la section 3.2. Le nombre de jets contenant un muon (N_{jets}^{μ}) dans les lots de simulation et de données sont montrés dans les tableaux 3.3 et 3.4.

Nom du lot de simulation	N_{jets}^{μ}		
	<i>loose</i>	<i>med</i>	<i>med3</i>
$t\bar{t}$	6 447 328	4 387 491	3 968 315
bb+jets	3 437 830	2 577 080	2 391 910
cc+jets	1 289 030	754 310	658 675
direct $b\bar{b}$	956 754	669 694	612 262
direct $c\bar{c}$	160 657	78 066	62 249
QCD $q\bar{q}$	1 027 812	351 395	225 907
γ +jets	235 325	81 037	51 308

TAB. 3.3: Nombre de jets contenant un muon (N_{jets}^{μ}) dans les lots de simulation.

²produit avec le générateur ALPGEN + PYTHIA

Nom du lot de données	N_{jets}^{μ}		
	<i>loose</i>	<i>med</i>	<i>med3</i>
MUinclusive	87 568 923	84 499 024	84 117 440
QCD	1 166 203	431 659	334 159

TAB. 3.4: Nombre de jets contenant un muon (N_{jets}^{μ}) dans les lots de données.

Les efficacités de reconstruction d'un muon (ε_{μ}) dans les données et la simulation sont différentes. Cette efficacité comprend deux effets : l'identification du muon (ε_{ID}) et l'identification de la trace associée au muon ($\varepsilon_{\text{corr. trace}}$).

$$\varepsilon_{\mu} = \varepsilon_{\text{ID}} \times \varepsilon_{\text{corr. trace}} \quad (3.2)$$

Le groupe de la collaboration travaillant sur l'identification des muons [46] nous fournit ces efficacités (ainsi que les erreurs statistiques et systématiques associées) pour les données et la simulation. L'efficacité ε_{ID} est paramétrée en fonction de ϕ et η du muon, alors que $\varepsilon_{\text{corr. trace}}$ est paramétrée en fonction de la position z et de η^{CFT} de la trace du muon. Le rapport de ε_{μ} dans les données et la simulation, nous donne le facteur de correction (SF_{μ}) à appliquer à la simulation.

Nous pouvons ainsi définir l'efficacité d'étiquetage d'un muon dans un jet issu de quark b (ou c) par :

$$\varepsilon_{\mu}^{\text{SLT}} = \frac{N_{\text{jets}}^{\mu} \times \text{SF}_{\mu}}{N_{\text{jets}}^{\text{incl.}}} \quad (3.3)$$

où $\varepsilon_{\mu}^{\text{SLT}}$ prend en compte le rapport de branchement $b \rightarrow \mu$ de 20% (voir 3.2) et l'efficacité de reconstruction d'un muon dans les données. Il est important de noter que les corrections utilisées pour définir SF_{μ} sont définies pour des muons isolés de $p_{\text{T}} > 15$ GeV, ce qui ne correspond pas aux critères des muons utilisés par SLT³.

On paramétrise $\varepsilon_{\mu}^{\text{SLT}}$ en fonction du p_{T} du jet à l'aide de la combinaison de trois processus simulés : $t\bar{t}$, direct $b\bar{b}$ (direct $c\bar{c}$) et bb +jets (cc +jets). La figure 3.7 met en évidence que la forme de la distribution obtenue est différente suivant les régions du calorimètre (dont les définitions se trouvent dans le tableau 3.5) et dépend du processus de simulation utilisé.

On ajuste $\varepsilon_{\mu}^{\text{SLT}}$ par une fonction du type $a + \frac{b}{\log x} + c \times \log x$. A l'aide de cet ajustement et de la matrice de covariance associée, on définit la fonction d'erreur absolue ($\Delta^{\text{stat.}} \varepsilon_{\mu}^{\text{SLT}}$) qui est elle-même ajustée par un polynôme d'ordre 3.

³ L'évaluation de SF_{μ} pour les muons utilisés par SLT n'existe pas à l'heure actuelle et représente un travail conséquent qui dépasse le cadre du développement de la méthode d'étiquetage présentée. Résoudre cette inconsistance est une perspective d'amélioration de SLTNN.

Région du calorimètre	η physique
Calorimètre central (CC)	$0 < \eta < 1.2$
Inter-Cryostat (ICR)	$1.2 < \eta < 1.8$
Bouchons (EC)	$1.8 < \eta < 2.5$

TAB. 3.5: Définition des régions en η du calorimètre.

3.4.1 Erreurs systématiques

Il y a deux sources d'erreurs systématiques liées à l'évaluation de $\varepsilon_{\mu}^{\text{SLT}}$. La première prend en compte la dépendance du lot de simulation utilisé sur l'efficacité d'étiquetage du muon dans le jet. Pour ce faire nous allons comparer le rapport de l'efficacité obtenue directement (sur les lots de processus simulés $t\bar{t}$, direct $b\bar{b}$ et bb +jets) à celle donnée par l'ajustement de $\varepsilon_{\mu}^{\text{SLT}}$ (sur le lot $t\bar{t}$) en fonction du p_T du jet pour les trois régions du calorimètre. Ce rapport est ajusté par une constante. La valeur de l'ajustement est utilisé pour obtenir une estimation de cette erreur systématique qui est reportée dans le tableau 3.6. L'ajustement par une droite n'étant pas représentatif dans bien des cas, on utilisera l'information *bin par bin* de la distribution pour obtenir l'erreur systématique totale. Les figures 3.8 et 3.9 montrent le résultat de cette systématique ainsi que l'ajustement obtenu.

Qualité du muon	jets issus de quark b			jets issus de quark c		
	bb +jets	$t\bar{t}$	direct $b\bar{b}$	cc +jets	$t\bar{t}$	direct $c\bar{c}$
<i>loose</i>	7.5%	0.3%	11.1%	1.8%	0.2%	0.5%
<i>med</i>	7.9%	0.2%	5.7%	11.4%	0.1%	15.1%
<i>Med3</i>	14.4%	0.3%	12.9%	17.3%	0.1%	23.0%

TAB. 3.6: Erreur systématique liée à l'évaluation de $\varepsilon_{\mu}^{\text{SLT}}$. Les valeurs présentées sont les moyennes issues du *fit* par une constante.

La deuxième source d'incertitude systématique rend compte de l'efficacité de reconstruction des muons ε_{μ} . Pour évaluer cette systématique, on fait varier les facteurs de correction ε_{ID} et $\varepsilon_{\text{corr. trace}}$ de la valeur de leur incertitude $\pm 1\sigma$, à l'aide de la combinaison des trois processus : $t\bar{t}$, direct $b\bar{b}$ (direct $c\bar{c}$) et bb +jets (cc +jets). La systématique associée est globale (i.e. indépendante du p_T du jet, voir figure 3.10). Le tableau 3.7 résume l'erreur systématique obtenue pour toutes les qualités de muon.

On détermine l'erreur systématique totale d'étiquetage de muons dans les jets ($\Delta^{\text{sys.}} \varepsilon_{\mu}^{\text{SLT}}$) en additionnant de manière quadratique les deux sources d'erreurs vues précédemment. Cette systématique est fonction du p_T du jet, de la région du détecteur et de la saveur du quark dont est issu le jet (voir figures 3.11 et 3.12). L'erreur systématique totale est de 20% en moyenne.

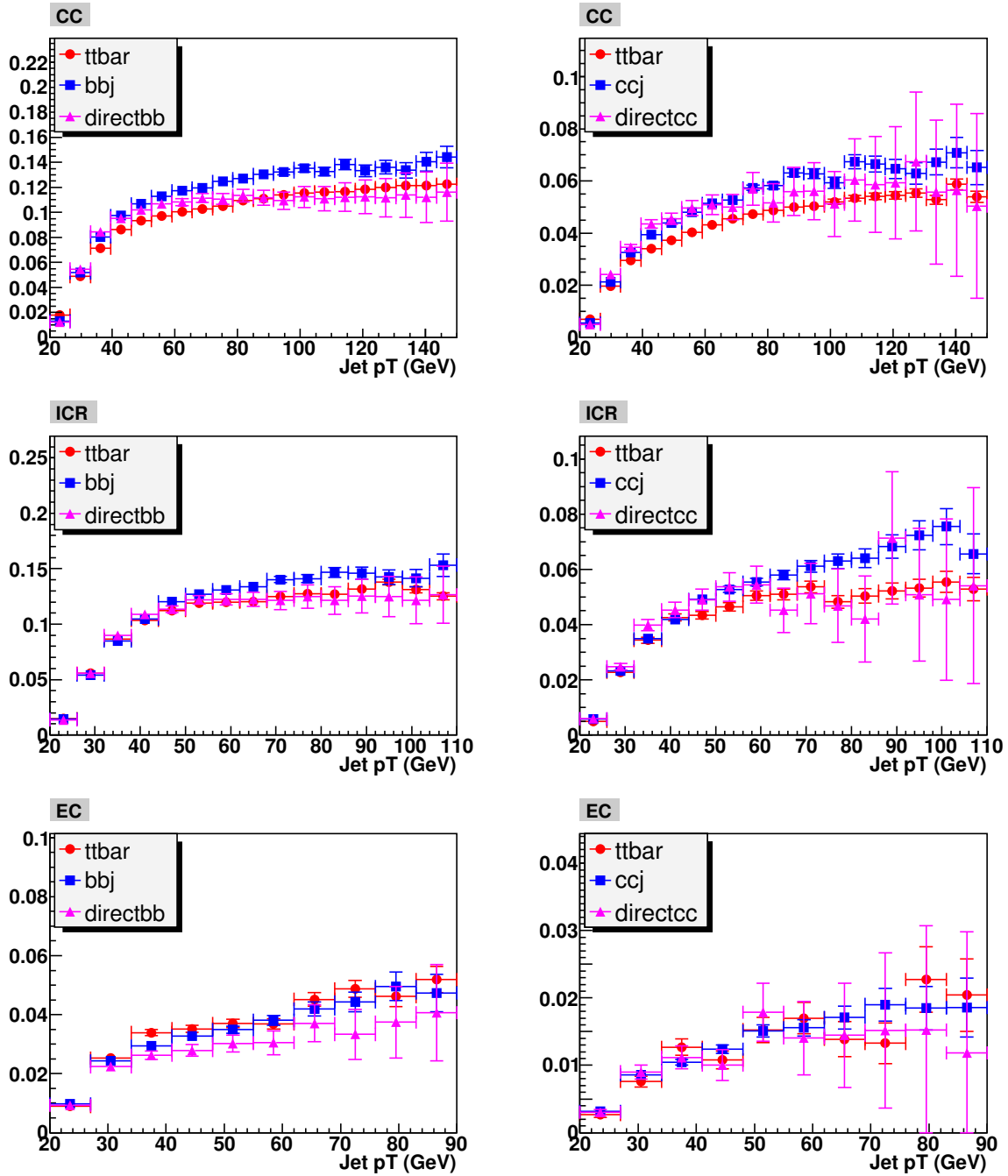


FIG. 3.7: Efficacité (corrigée de celle des données) d'étiquetage du muon de qualité *med* dans le jet issus de quark b (à gauche) ou c (à droite) en fonction du p_T du jet. En partant du haut vers le bas, on voit cette efficacité en fonction des trois zones du calorimètre : CC, ICR, EC. Cette efficacité a été obtenue pour plusieurs processus simulés : les points rouges pour $t\bar{t}$, les points bleux pour bb +jets (cc +jets) et les points roses pour direct $b\bar{b}$ (direct $c\bar{c}$). Les erreurs portées sont statistiques.

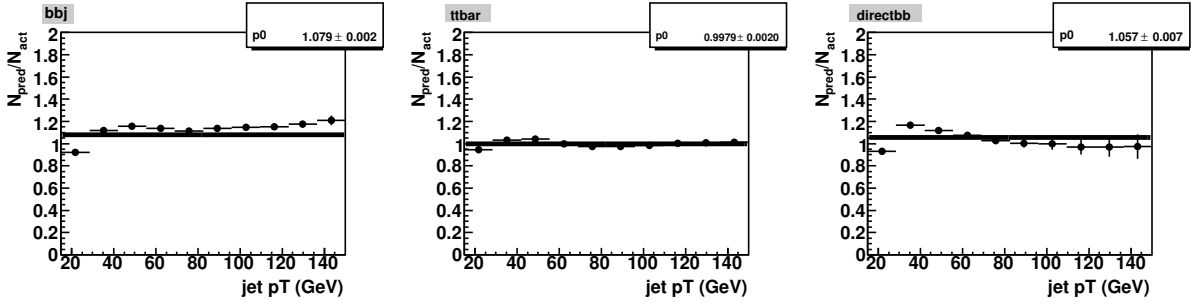


FIG. 3.8: Erreur systématique relative à la dépendance du lot de simulation utilisé sur l'efficacité d'étiquetage du muon dans les jets issus de quark b : rapport de l'efficacité obtenue directement (sur les lots de processus simulés bb +jets à gauche, $t\bar{t}$ au milieu et direct $b\bar{b}$ à droite) à celle donnée par l'ajustement de $\varepsilon_{\mu}^{\text{SLT}}$ (sur le lot $t\bar{t}$) en fonction du p_T du jet pour un muon de qualité *med*. Le fit par une constante est utilisé pour obtenir une valeur moyenne de l'erreur systématique.

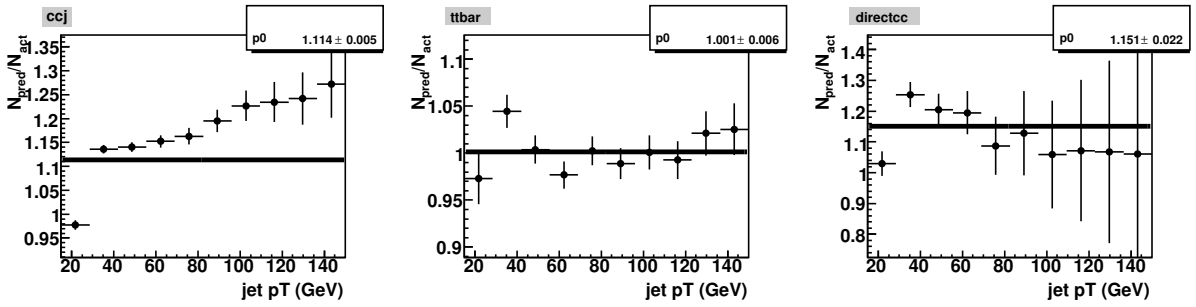


FIG. 3.9: Erreur systématique relative à la dépendance du lot de simulation utilisé sur l'efficacité d'étiquetage du muon dans les jets issus de quark c : rapport de l'efficacité obtenue directement (sur les lots de processus simulés cc +jets à gauche, $t\bar{t}$ au milieu et direct $c\bar{c}$ à droite) à celle donnée par l'ajustement de $\varepsilon_{\mu}^{\text{SLT}}$ (sur le lot $t\bar{t}$) en fonction du p_T du jet pour un muon de qualité *med*. Le fit par une constante est utilisé pour obtenir une valeur moyenne de l'erreur systématique.

Qualité du muon	jets issus de quark b		jets issus de quark c	
	ε_{ID}	$\varepsilon_{\text{corr. trace}}$	ε_{ID}	$\varepsilon_{\text{corr. trace}}$
<i>loose</i>	0.69%	0.42%	0.70%	0.42%
<i>Med</i>	1.23%	0.42%	1.24%	0.42%
<i>med3</i>	1.29%	0.42%	1.30%	0.42%

TAB. 3.7: Erreur systématique associée aux facteurs de correction ε_{ID} et $\varepsilon_{\text{corr. trace}}$ du muon.

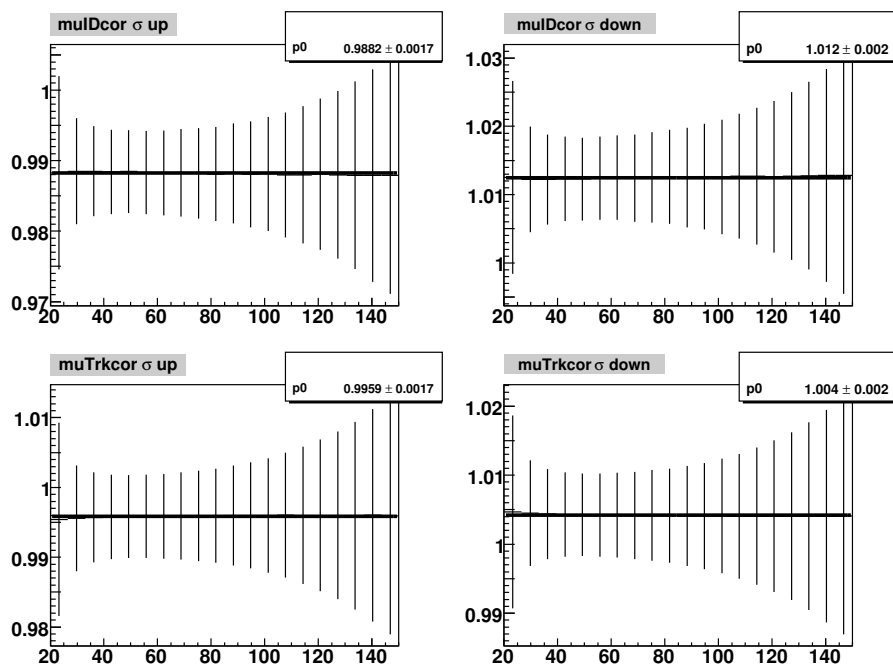


FIG. 3.10: Erreur systématique associée à la variation de $\pm 1\sigma$ (+ : à gauche et - : à droite) des facteurs de correction ϵ_{ID} (en haut) et $\epsilon_{corr. trace}$ (en bas) du muon de qualité *med*. Le *fit* par une constante est utilisé pour obtenir une valeur moyenne fiable de l'erreur systématique.

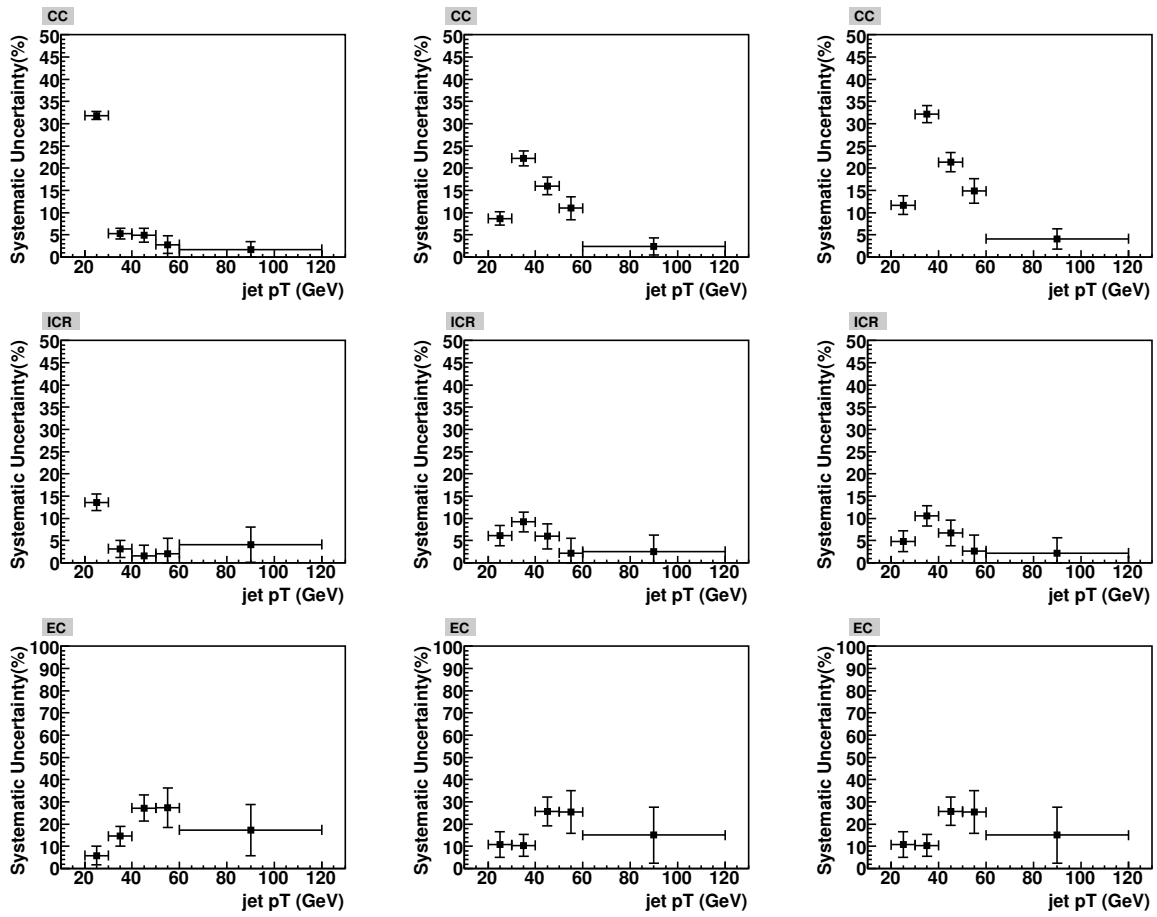


FIG. 3.11: Erreur systématique totale d'étiquetage de muons dans les jets issus de quark b en fonction du p_T du jet, de la région du détecteur (CC, ICR et EC de haut en bas) pour les trois qualités de muons (*loose*, *med* et *med3* de gauche à droite).

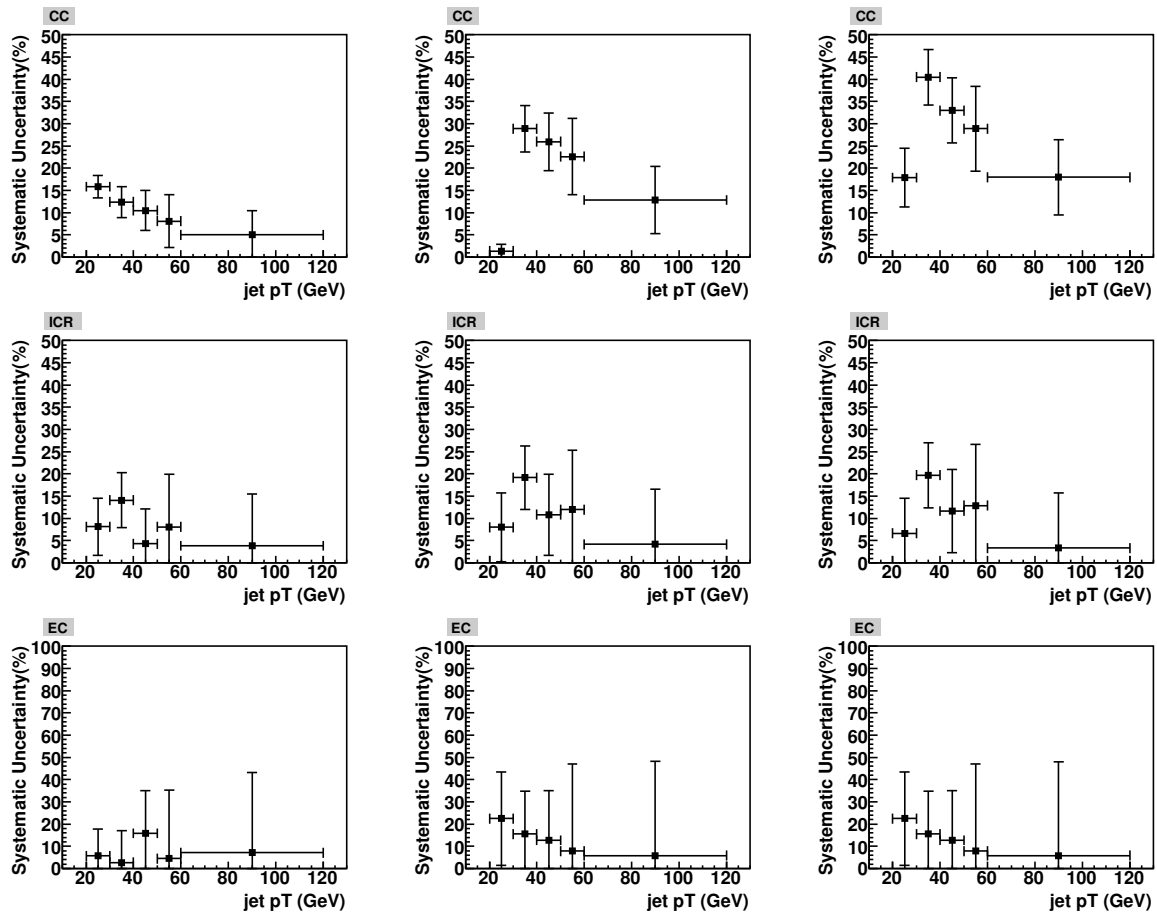


FIG. 3.12: Erreur systématique totale d'étiquetage de muons dans les jets issus de quark c en fonction du p_T du jet, de la région du détecteur (CC, ICR et EC de haut en bas) pour les trois qualités de muons (*loose*, *med* et *med3* de gauche à droite).

3.5 Entraînement du Réseau de Neurone

On a vu que l'efficacité d'identification d'un muon dans un jet issu de quark b de 40 GeV dans le CC est d'environ de 10% (figure 3.7). L'objectif de la méthode SLTNN est maintenant d'affiner cette efficacité d'étiquetage des jets muoniques tout en rejetant un maximum de bruit de fond (i.e. de minimiser le taux de mauvaise identification). Pour y parvenir, nous allons construire un réseau de neurones spécifique à cette méthode.

Nous utilisons un réseau de neurones Perceptron multi-couches qui comprend :

- 11 neurones (variables) dans la couche d'entrée ;
- 2 couches cachées : 24 neurones composent la première couche et 1 unique neurone la deuxième⁴ ;
- 1 neurone dans la couche de sortie ;
- 2000 époques d'entraînement.

Les caractéristiques ci-dessus du réseau de neurones ont été obtenues après optimisation.

Le réseau a été entraîné avec les processus simulés suivants :

- pour le signal : direct $b\bar{b}$;
- pour le bruit de fond : QCD $q\bar{q}$ et γ +jets. Les figures B.6, B.7 (en annexe B) montrent les distributions des 11 variables d'entrée utilisées pour ces deux lots. Nous pouvons voir que globalement la forme des distributions pour $q\bar{q}$ et γ +jets sont les mêmes. Ces lots seront combinés par la suite afin d'augmenter la statistique de l'échantillon de bruit de fond.

Les variables d'entrée du réseau de neurones utilisées peuvent se séparer en deux catégories : celles relatives aux caractéristiques du muon à l'intérieur du jet (au nombre de 5) et celles relatives aux caractéristiques du vertex secondaire reconstruit (au nombre de 6). Les variables issues de l'algorithme SVT nécessitent que le vertex secondaire reconstruit utilise la trace du muon. La liste des 11 variables utilisées est :

- p_T du muon ;
- p_T^{rel} du muon ⁵. Il correspond au p_T du muon relatif à l'impulsion totale du muon et du jet ;
- $\Delta R(\text{muon}, \text{jet})$ (on a déjà imposé la condition $\Delta R(\text{muon}, \text{jet}) < 0.5$) ;
- χ^2 du muon ;
- IP_{sig} : le paramètre d'impact en signification du muon ;
- $CSIP_{\text{comb}}$;
- SVT_{SLDLS} : longueur de désintégration en signification du vertex secondaire ;
- $SVT_{\text{SLX}_{\text{ddl}}^2}$: χ^2 par degré de liberté du vertex secondaire ;
- $SVT_{\text{SLN}_{\text{Tracks}}}$: nombre de traces utilisées pour reconstruire le vertex secondaire ;
- SVT_{SLMass} : masse du vertex secondaire ;

⁴le rôle du neurone unique est de normaliser la sortie du réseau de neurones entre 0 et 1, voir annexe A.

⁵ $p_T^{\text{rel}} = \sqrt{\frac{\sum_{i=1}^3 (p_{x_i}^{\mu} p_{x_{i+1}}^{\mu+\text{jet}} - p_{x_i}^{\mu+\text{jet}} p_{x_{i+1}}^{\mu})^2}{\sum_{i=1}^3 (p_{x_i}^{\mu+\text{jet}})^2}}$ avec $p_{x_i}^{\mu+\text{jet}} = p_{x_i}^{\mu} + p_{x_i}^{\text{jet}}$, $x_1 = x_4 = x$, $x_2 = y$ et $x_3 = z$

- $SVT_{SL}Num$: pour cette variable, il suffit qu'au moins un vertex secondaire soit reconstruit en utilisant la trace du muon. Cette condition passée, on dénombre les vertex secondaires reconstruits pour le jet considéré.

On demande à l'algorithme SVT de travailler avec le critère SL (*SuperLoose*) et par conséquent la variable $SVT_{SL}N_{Tracks}$ aussi. Sur ce point SLTNN diffère de la méthode NN qui utilise le critère L (*Loose*) pour cette variable. En effet, l'optimisation du réseau de neurones utilisé par SLTNN a montré de meilleurs résultats avec ce critère SL. On peut vérifier sur la figure 3.13 que les deux variables $SVT_{SL}N_{Tracks}$ et $SVT_L N_{Tracks}$ ont un bon pouvoir discriminant. Cependant, comme $SVT_{SL}N_{Tracks}$ a un critère plus lâche sur la sélection du vertex secondaire, cela nous permet d'avoir plus d'évènements disponibles et donc plus d'information.

Les figures 3.14 et 3.15 montrent les variables d'entrées du SLTNN pour le signal et le bruit de fond.

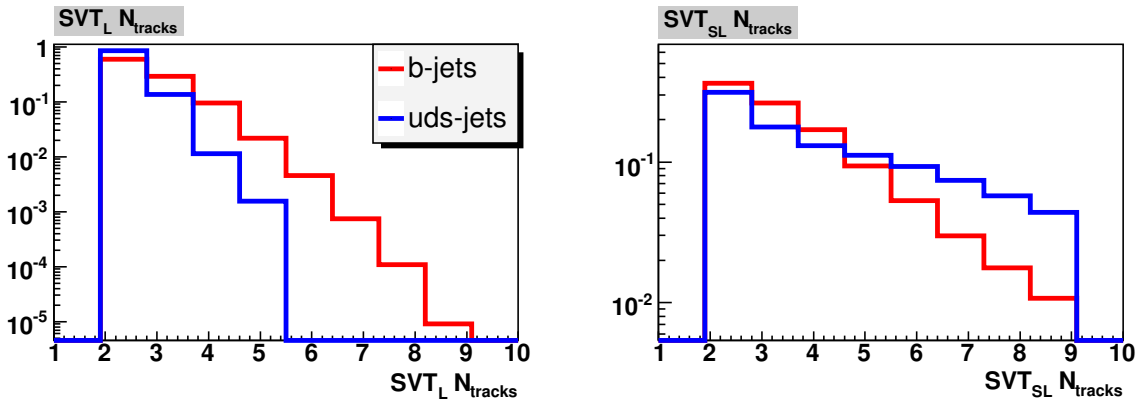


FIG. 3.13: Comparaison de $SVT_L N_{Tracks}$ (droite) et $SVT_{SL} N_{Tracks}$ (gauche) pour le signal (rouge) et le bruit de fond combiné (bleu) et pour un muon de qualité *med*. Ces distributions sont normalisées à l'unité.

Points de fonctionnement	<i>loose</i>	<i>med</i>	<i>med3</i>
<i>LOOSE</i>	0.094	0.036	0.031
<i>MEDIUM</i>	0.151	0.068	0.064
<i>TIGHT</i>	0.352	0.231	0.209

TAB. 3.8: Valeur des coupures sur la sortie du réseau de neurones pour les trois points de fonctionnement (*LOOSE*, *MEDIUM* et *TIGHT*) et pour les trois qualités de muon (*loose*, *med* et *med3*).

La figure 3.16 présente la sortie du réseau de neurones obtenue. On peut voir une bonne séparation du signal et du bruit de fond. Il nous est maintenant possible de convertir ce résultat en terme :

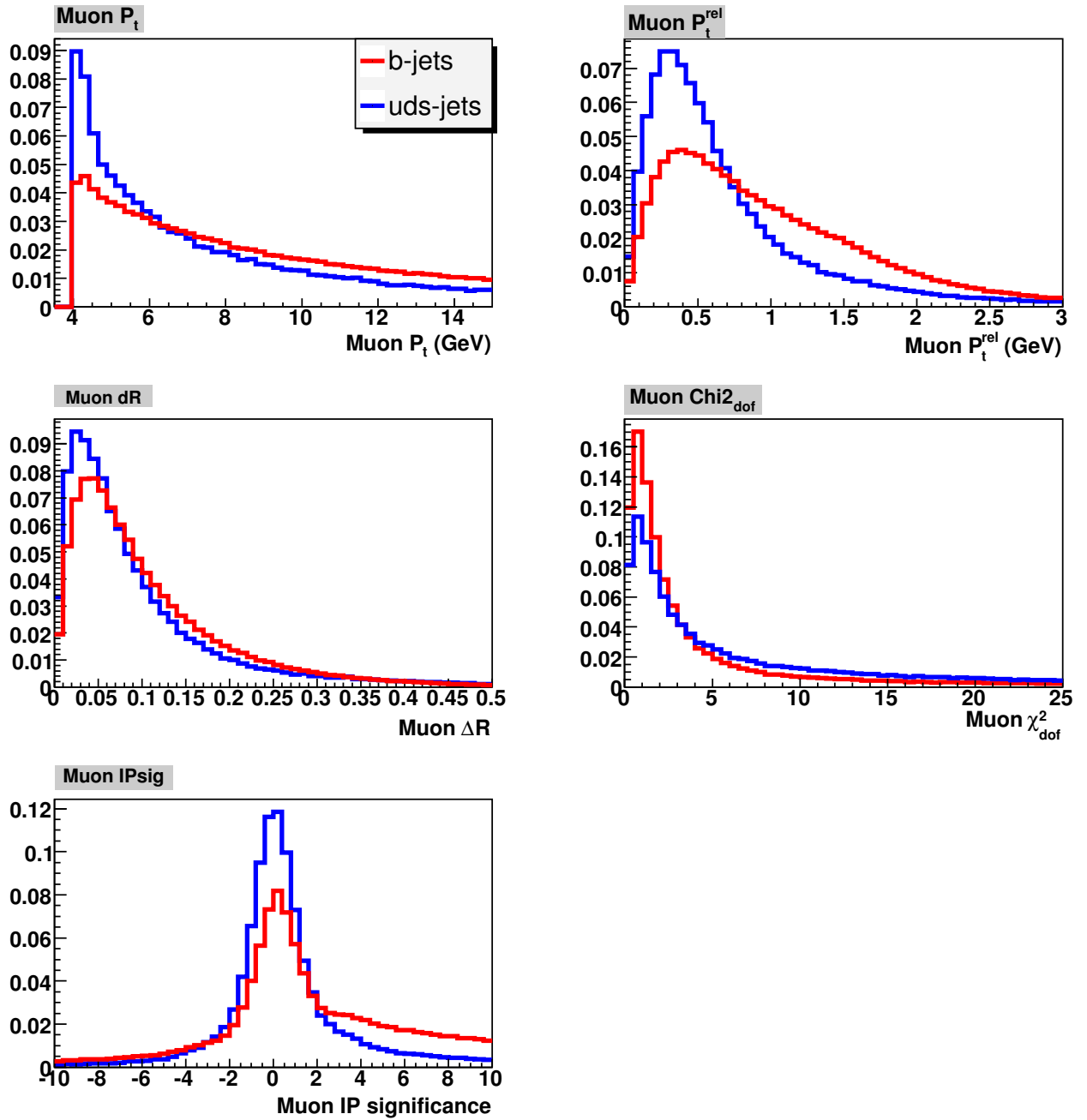


FIG. 3.14: Variables d'entrée du réseau de neurones relatives au muon de qualité *med* pour le signal (rouge) et le bruit de fond combiné (bleu). De haut en bas et de gauche à droite : p_T du muon, p_T^{rel} du muon, $\Delta R(\text{muon}, \text{jet})$, χ^2_{ddl} du muon et IP_{sig} du muon. Ces distributions sont normalisées à l'unité.

- d'efficacité d'étiquetage : nombre de fois où un jet issu de quark b est identifié comme tel;

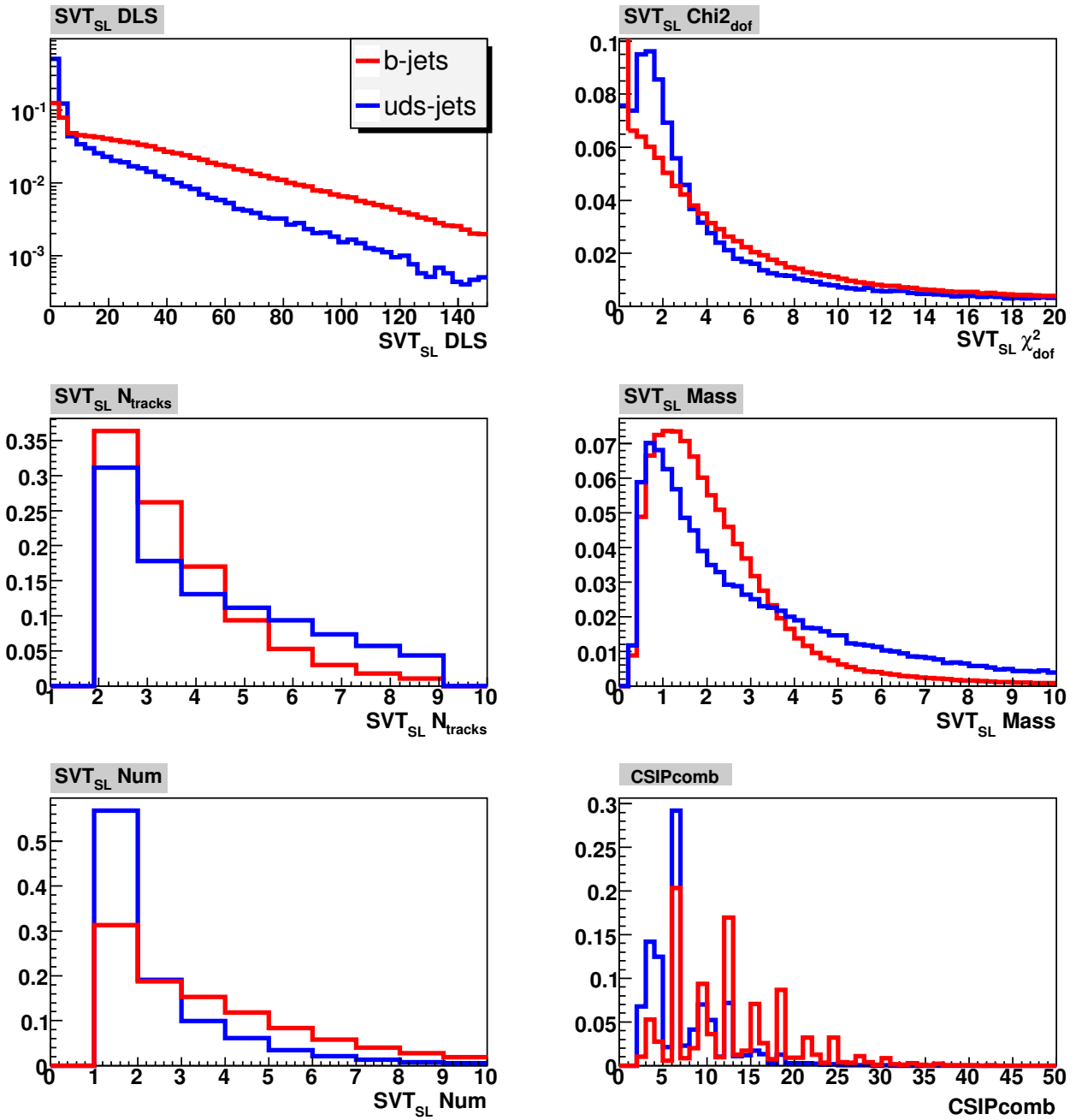


FIG. 3.15: Variables d'entrée du réseau de neurones relatives au vertex secondaire pour le signal (rouge) et le bruit de fond combiné (bleu) pour un muon de qualité *med*. De haut en bas et de gauche à droite : $SVT_{SL} DLS$, $SVT_{SL} \chi_{dof}^2$, $SVT_{SL} N_{tracks}$, $SVT_{SL} Mass$, $SVT_{SL} Num$ et $CSIPcomb$. Ces distributions sont normalisées à l'unité.

- de taux de mauvaise identification (*Fake Rate*) : nombre de fois où un jet issu de quark léger est identifié comme étant issu d'un quark b.

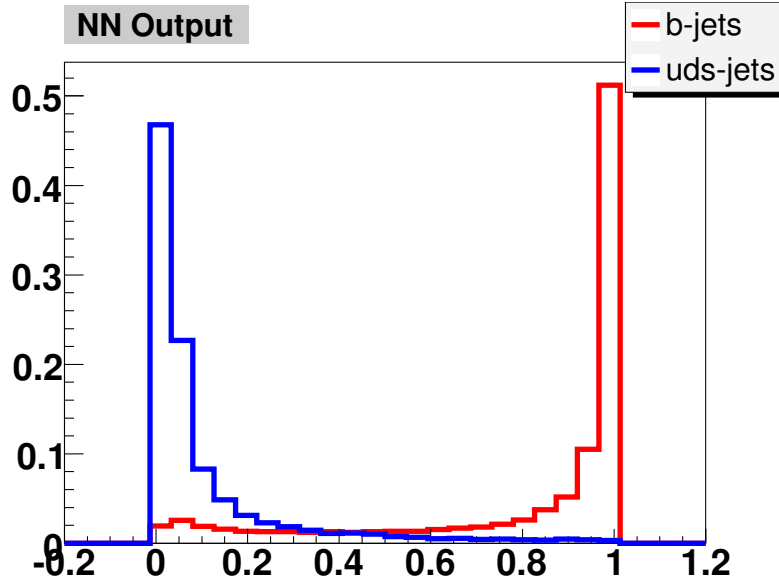


FIG. 3.16: Sortie du réseau de neurones pour le signal (rouge) et pour le bruit de fond combiné (bleu) pour un muon de qualité *med*. Ces distributions sont normalisées à l'unité.

Le choix de la coupure sur la sortie du réseau de neurones est un compromis entre l'efficacité et le taux de contamination. Plusieurs points de fonctionnement ayant des critères plus ou moins sévères sont fournis aux utilisateurs. Trois points de fonctionnement ayant un taux de mauvaise identification de 0.15% (*LOOSE*), 0.1% (*MEDIUM*) et 0.04% (*TIGHT*) ont été choisis pour maintenir une efficacité d'identification suffisamment élevée. Les valeurs de coupure sur la sortie du réseau de neurone correspondantes sont indiquées dans le tableau 3.8.

La figure 3.17 du haut montre l'efficacité d'étiquetage d'un jet issu de quark b en fonction du taux de mauvaise identification du réseau de neurones ($\epsilon_{\text{NN}}^{\text{SLTNN}}$) pour un lot de jets ayant déjà passé la contrainte $\Delta R(\text{muon}, \text{jet}) < 0.5$. La figure 3.17 du bas convolue l'efficacité d'étiquetage du muon dans le jet dans les données (section 3.4) avec l'efficacité du réseau de neurones ($\epsilon^{\text{SLTNN}} = \epsilon_{\mu}^{\text{SLT}} \times \epsilon_{\text{NN}}^{\text{SLTNN}}$). On peut voir que pour un muon de qualité *loose* et le point de fonctionnement *MEDIUM*, on obtient une efficacité d'étiquetage de la méthode SLTNN (ϵ^{SLTNN}) des jets issus de quark b d'environ 5%. Cette efficacité prend déjà en compte le rapport de branchement des désintégrations semi-leptoniques des jets issus de quark b (20%), de l'efficacité de reconstruction d'un muon dans les jets (figure 3.7) et de l'efficacité du réseau de neurones (>80% figure 3.17 du haut). A ce stade, seule l'efficacité d'un muon dans les jets est corrigée de celle obtenue dans les données. L'efficacité associée au réseau de neurone SLTNN $\epsilon_{\text{NN}}^{\text{SLTNN}}$ ne provenant que de la simulation, il convient d'apporter une correction supplémentaire pour refléter l'efficacité des données.

Sans utiliser le réseau de neurones, l'efficacité d'identification d'un muon de qualité *med* pour un jet b de 20 GeV dans le CC était d'environ 2% (figure 3.7) pour un taux de mauvaise

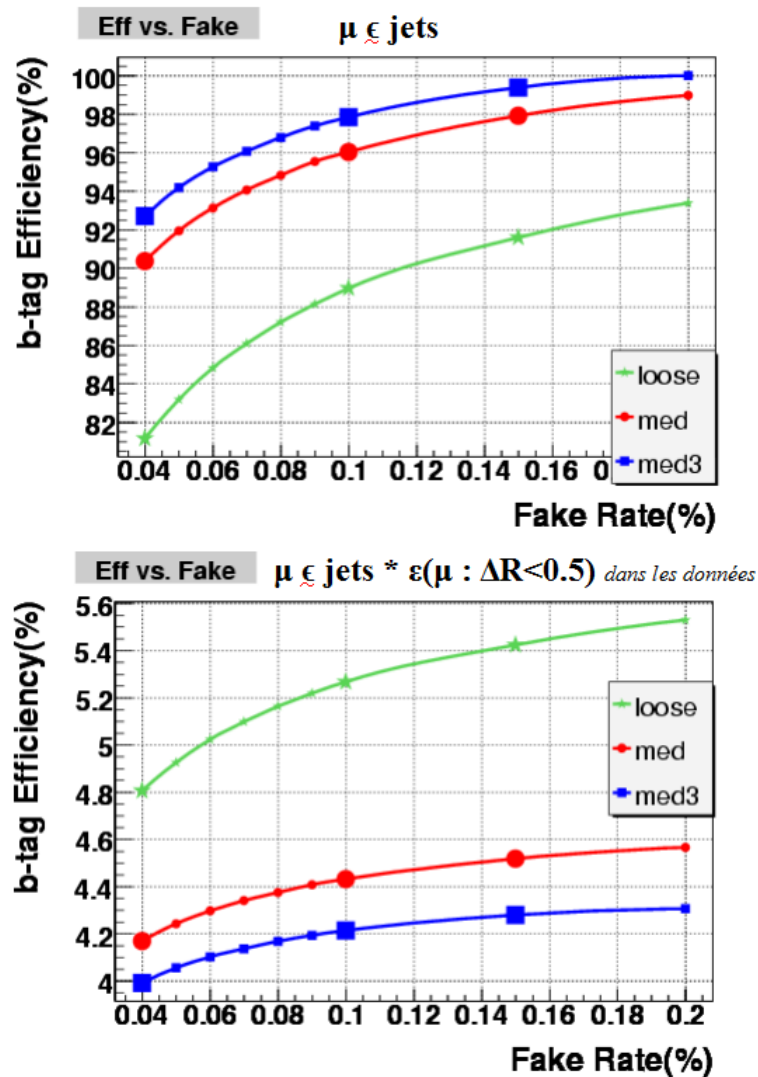


FIG. 3.17: Efficacité d'étiquetage des jets issus de quark b en fonction du taux de mauvaise identification pour les qualités de muon *loose* (vert), *med* (rouge) et *med3* (bleu). La distribution du haut prend uniquement en compte l'efficacité du réseau de neurone (ϵ_{NN}^{SLTNN}) alors que celle du bas inclut en plus l'efficacité d'étiquetage des muons dans les jets corrigés des données (ϵ_{μ}^{SLT}) et le rapport de branchement $b \rightarrow \mu$ qui est de 20%. Ces efficacités ont été obtenues sur des lots de simulation.

identification dans un jet léger d'environ 0.15% (voir plus loin la figure 3.18). En utilisant le réseau de neurones, l'efficacité d'étiquetage d'un jet b muonique est d'environ 4.5% pour le même niveau de mauvaise identification et qualité de muon. On a donc plus que doublé l'efficacité à l'aide du réseau de neurones à taux de mauvaise identification fixée.

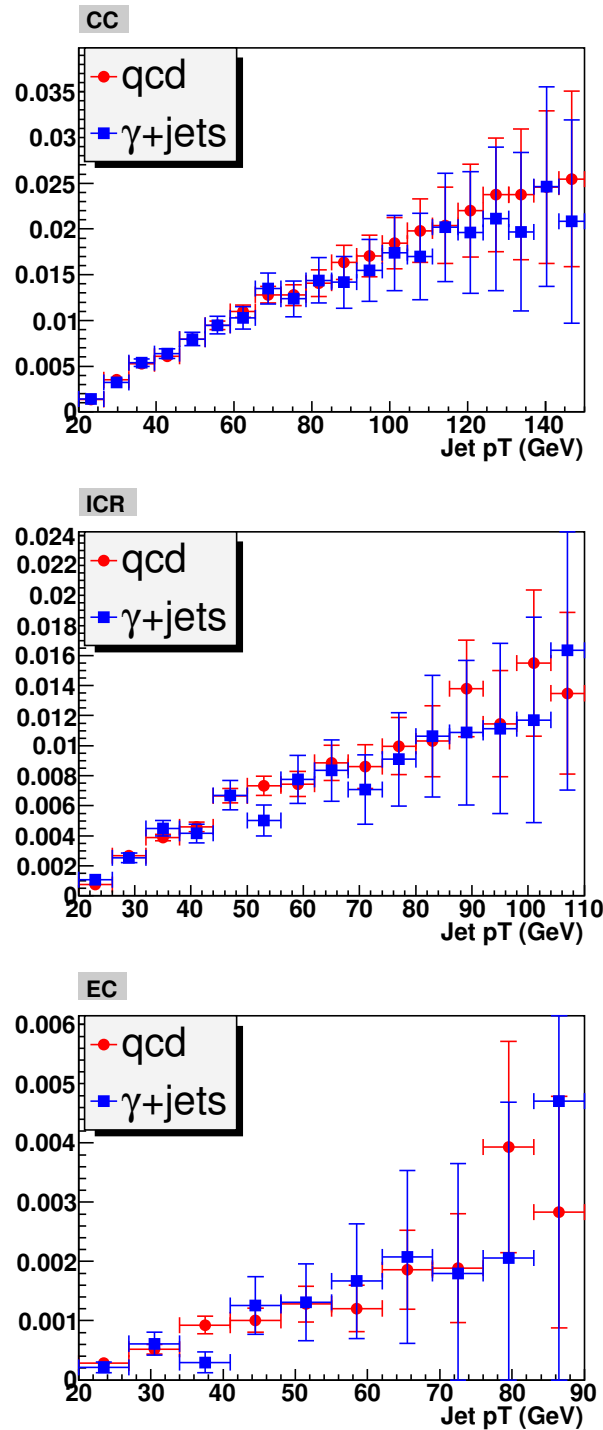


FIG. 3.18: Efficacité (corrégée de celle des données) d'étiquetage du muon de qualité *med* dans le jet issu de quark légers en fonction du p_T du jet. En partant du haut vers le bas, on voit cette efficacité en fonction des trois zones du calorimètre : CC, ICR, EC. Cette efficacité a été obtenue pour plusieurs processus simulés : les points rouges pour $q\bar{q}$ et les points bleus pour γ +jets. Les erreurs portées sont statistiques.

3.6 Résultats avec les données

Nous allons maintenant corriger les valeurs de l'efficacité d'étiquetage et le taux de mauvaise identification avec celle obtenue dans les données. Dans la section 3.5, nous avons travaillé avec les lots de simulation. Nous avons donc accès à l'information sur la saveur des quarks. Avec les données cette information n'est plus disponible et il nous faut maintenant trouver des méthodes pour quantifier ces efficacités.

3.6.1 Efficacité d'étiquetage

L'efficacité ϵ_{NN}^{SLTNN} est évaluée à partir du lot *MUinclusive* et à l'aide de la méthode *System8* [59] [60]. Cette méthode permet de déterminer les fractions d'évènements de signal (jets issus de quark b) et de bruit de fond (jets issus de quark c ou légers : noté cl), ainsi que les efficacités dans les données. Elle consiste à résoudre un système de 8 équations à 8 inconnues grâce à deux algorithmes d'étiquetage indépendants et à deux lots de données différents.

Les algorithmes d'étiquetage utilisés sont SLTNN et JLIP (voir section 3.1.1). Afin de renforcer l'indépendance, on modifie l'algorithme JLIP. On définit la variable *modJLIP* comme étant la probabilité que les traces du jet proviennent du vertex primaire à l'aide de leur paramètre d'impact (JLIP) en excluant la trace reliée au muon. La figure 3.19 montre la distribution de *modJLIP*. Le *System8* sera résolu pour les trois points de fonctionnement de SLTNN et avec la contrainte : probabilité de *modJLIP* < 0.1 (pour enrichir l'échantillon en jets issus de quark b).

La méthode NN, quant à elle, utilise les algorithmes d'étiquetage NN et SLT pour résoudre *System8*. Cela crée une corrélation entre NN et SLTNN.

L'algorithme SLT est aussi utilisé pour la détermination des efficacités d'étiquetage dans les données dans l'expérience ATLAS au LHC. L'algorithme dans leur cas utilise aussi les muons et est nommé SMT [20] (*Soft Muon Tagging*). ATLAS n'a pas pour l'instant prévu de l'utiliser comme méthode d'étiquetage en complément des autres outils qu'ils ont à disposition.

Le premier lot de données est composé de jets muoniques. Le second contient deux jets muoniques dos-à-dos ($|\Delta\phi(\text{jet1}, \text{jet2}) - 180^\circ| < 22.5^\circ$) dont l'un des deux jets passe le critère : probabilité JLIP < 0.01 . Ce second lot est donc un sous-lot du premier avec une composition en jets issus de quarks de saveurs lourdes plus grande. On l'appellera lot enrichi en b.

Le *System8* utilise les paramètres suivants :

- n : est le nombre de jets dans le lot muonique ;
- p : est le nombre de jets dans le lot enrichi en b ;
- $n^{SLTNN} (p^{SLTNN})$: est le nombre de jets étiquetés par SLTNN dans le lot muonique (enrichi en b) ;
- $n^{\text{modJLIP}} (p^{\text{modJLIP}})$: est le nombre de jets étiquetés par *modJLIP* dans le lot muonique (enrichi en b) ;
- $n^{\text{modJLIP}, SLTNN} (p^{\text{modJLIP}, SLTNN})$: est le nombre de jets étiquetés par *modJLIP* et SLTNN

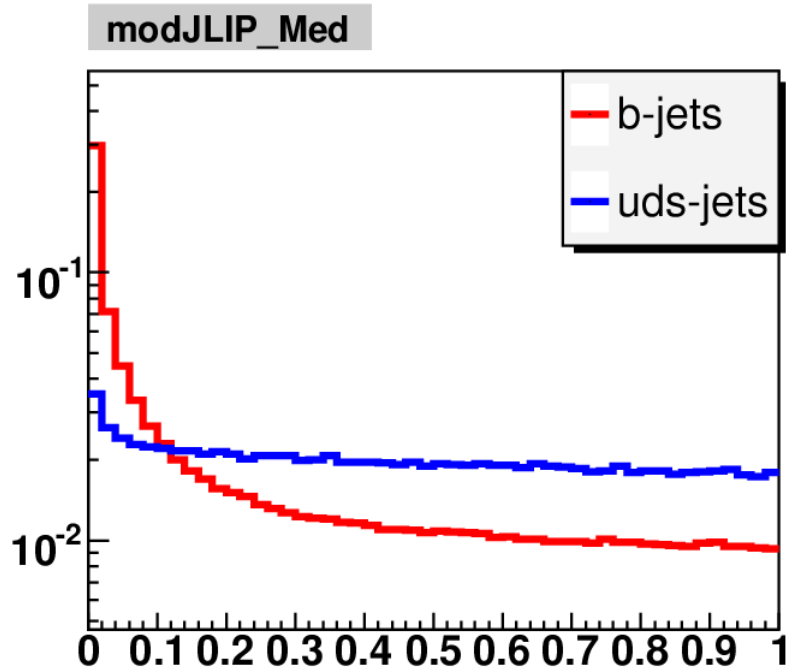


FIG. 3.19: Distribution de modJLIP pour les qualités de muon *med* pour les jets issus de quark b (en rouge, lot simulé direct $b\bar{b}$) ou de quarks légers (en bleu, lots simulés QCD $q\bar{q}$ et γ +jets).

dans le lot muonique (enrichi en b) ;

- n_b ou n_{cl} (p_b ou p_{cl}) : sont le nombre de jets issus de quark b ou cl dans le lot muonique (enrichi en b) ;
- ε_b^{SLTNN} ou ε_{cl}^{SLTNN} ($\varepsilon_b^{\text{modJLIP}}$ ou $\varepsilon_{cl}^{\text{modJLIP}}$) : sont les efficacités d'étiquetage des jets issus de quark b ou cl de SLTNN (modJLIP) ;

ainsi que les six coefficients de corrélation entre l'efficacité d'étiquetage de modJLIP et SLTNN :

- $\kappa_b = \frac{\varepsilon_b^{\text{modJLIP,SLTNN}}}{\varepsilon_b^{\text{modJLIP}} \times \varepsilon_b^{\text{SLTNN}}}$ (figure 3.20) ;
- $\kappa_{cl} = \frac{\varepsilon_{cl}^{\text{modJLIP,SLTNN}}}{\varepsilon_{cl}^{\text{modJLIP}} \times \varepsilon_{cl}^{\text{SLTNN}}}$ (figure B.1 en annexe B).
- $\delta = \frac{\varepsilon_b^{\text{modJLIP}} \text{ dans le lot enrichi}}{\varepsilon_b^{\text{modJLIP}} \text{ dans le lot muonique}}$ (figure B.2 en annexe B) ;
- $\gamma = \frac{\varepsilon_{cl}^{\text{modJLIP}} \text{ dans le lot enrichi}}{\varepsilon_{cl}^{\text{modJLIP}} \text{ dans le lot muonique}}$ (figure B.3 en annexe B) ;
- $\beta = \frac{\varepsilon_b^{\text{SLTNN}} \text{ dans le lot enrichi}}{\varepsilon_b^{\text{SLTNN}} \text{ dans le lot muonique}}$ (figure B.4 en annexe B) ;
- $\alpha = \frac{\varepsilon_{cl}^{\text{SLTNN}} \text{ dans le lot enrichi}}{\varepsilon_{cl}^{\text{SLTNN}} \text{ dans le lot muonique}}$ (figure B.5 en annexe B).

Voici les équations du *System8* :

$$n = n_b + n_{cl} \quad (3.4)$$

$$p = p_b + p_{cl} \quad (3.5)$$

$$n^{\text{modJLIP}} = \varepsilon_b^{\text{modJLIP}} \cdot n_b + \varepsilon_{cl}^{\text{modJLIP}} \cdot n_{cl} \quad (3.6)$$

$$p^{\text{modJLIP}} = \delta \cdot \varepsilon_b^{\text{modJLIP}} \cdot p_b + \gamma \cdot \varepsilon_{cl}^{\text{modJLIP}} \cdot p_{cl} \quad (3.7)$$

$$n^{\text{SLTNN}} = \varepsilon_b^{\text{SLTNN}} \cdot n_b + \varepsilon_{cl}^{\text{SLTNN}} \cdot n_{cl} \quad (3.8)$$

$$p^{\text{SLTNN}} = \beta \cdot \varepsilon_b^{\text{SLTNN}} \cdot p_b + \alpha \cdot \varepsilon_{cl}^{\text{SLTNN}} \cdot p_{cl} \quad (3.9)$$

$$n^{\text{modJLIP, SLTNN}} = \kappa_b \cdot \varepsilon_b^{\text{modJLIP}} \cdot \varepsilon_b^{\text{SLTNN}} \cdot n_b + \kappa_{cl} \cdot \varepsilon_{cl}^{\text{modJLIP}} \cdot \varepsilon_{cl}^{\text{SLTNN}} \cdot n_{cl} \quad (3.10)$$

$$p^{\text{modJLIP, SLTNN}} = \kappa_b \cdot \delta \cdot \beta \cdot \varepsilon_b^{\text{modJLIP}} \cdot \varepsilon_b^{\text{SLTNN}} \cdot p_b + \kappa_{cl} \cdot \gamma \cdot \alpha \cdot \varepsilon_{cl}^{\text{modJLIP}} \cdot \varepsilon_{cl}^{\text{SLTNN}} \cdot p_{cl} \quad (3.11)$$

A partir des huit valeurs connues (les termes de gauche dans les équations), on veut déterminer les huit inconnues :

- les nombres de jets n_b , n_{cl} , p_b et p_{cl} ;
- les efficacités d'étiquetage $\varepsilon_b^{\text{SLTNN}}$, $\varepsilon_{cl}^{\text{SLTNN}}$, $\varepsilon_b^{\text{modJLIP}}$ et $\varepsilon_{cl}^{\text{modJLIP}}$.

On détermine la valeur des coefficients de corrélation à l'aide d'évènements simulés. Si la valeur de ces coefficients de corrélation est proche de l'unité, on aura vérifié que les algorithmes d'étiquetage modJLIP et SLTNN sont indépendants. On utilise la combinaison des processus $bb+\text{jets}$ et $t\bar{t}$ (resp. $cc+\text{jets}$ et direct $c\bar{c}$) pour les coefficients reliés aux efficacités d'étiquetage des jets issus de quark b (resp. cl). Il a été vérifié que le coefficient de corrélation obtenu est le même quelque soit le processus utilisé (par exemple $\frac{\kappa_b(bb+\text{jets})}{\kappa_b(t\bar{t})} \sim 1$). Les valeurs obtenues sont résumées dans le tableau 3.9. On remarque que tous les coefficients de corrélation sont proches de l'unité sauf κ_{cl} qui diffère fortement. Dans une moindre mesure κ_b diffère aussi de l'unité. Ces écarts à l'unité seront pris en compte lors de l'estimation des erreurs systématiques. Les coefficients de corrélation sont aussi paramétrisés en fonction du p_T (tableau 3.10) et η (tableau 3.11) du jet (voir figures 3.20 à B.5).

Muon	OP	κ_b	κ_{cl}	β	α	δ	γ
loose	LOOSE	1.067±0.001	1.416±0.009	1.000±0.001	0.961±0.009	0.993±0.002	0.936±0.022
	MEDIUM	1.086±0.001	1.498±0.010	0.999±0.001	0.962±0.010	0.993±0.002	0.924±0.023
	TIGHT	1.137±0.001	1.687±0.012	0.998±0.001	0.959±0.013	0.990±0.002	0.922±0.026
med	LOOSE	1.019±0.001	1.129±0.007	1.000±0.000	0.987±0.005	0.996±0.003	0.977±0.025
	MEDIUM	1.034±0.001	1.225±0.009	1.000±0.001	0.983±0.007	0.996±0.003	0.969±0.026
	TIGHT	1.075±0.001	1.436±0.012	0.999±0.001	0.983±0.011	0.995±0.003	0.956±0.029
med3	LOOSE	1.007±0.001	1.044±0.006	1.000±0.000	0.994±0.003	0.996±0.003	0.989±0.026
	MEDIUM	1.021±0.001	1.135±0.008	1.000±0.000	0.984±0.006	0.995±0.003	0.982±0.027
	TIGHT	1.060±0.001	1.357±0.012	1.000±0.001	0.974±0.011	0.995±0.003	0.955±0.030

TAB. 3.9: Valeurs des coefficients de corrélation du *System8* en fonction de la qualité du muon (*loose*, *med* et *med3*) et du point de fonctionnement de SLTNN (*LOOSE*, *MEDIUM* et *TIGHT*). Les erreurs reportées sont uniquement statistiques.

Muon	OP	κ_b	κ_{cl}	β	α	δ	γ
loose	LOOSE	1.065±0.003	1.408±0.030	1.002±0.002	0.962±0.031	1.009±0.006	0.970±0.064
	MEDIUM	1.083±0.003	1.493±0.031	1.002±0.002	0.963±0.036	1.009±0.006	0.970±0.064
	TIGHT	1.132±0.003	1.683±0.034	1.002±0.003	0.959±0.047	1.009±0.006	0.970±0.064
med	LOOSE	1.018±0.003	1.125±0.034	1.000±0.001	0.990±0.019	1.010±0.006	0.974±0.082
	MEDIUM	1.032±0.003	1.218±0.038	1.001±0.001	0.986±0.026	1.010±0.006	0.974±0.082
	TIGHT	1.072±0.003	1.426±0.043	1.001±0.002	0.985±0.039	1.010±0.006	0.974±0.082
med3	LOOSE	1.006±0.003	1.044±0.032	1.000±0.001	0.995±0.011	1.010±0.007	0.983±0.086
	MEDIUM	1.019±0.003	1.132±0.037	1.001±0.001	0.987±0.021	1.010±0.007	0.983±0.086
	TIGHT	1.057±0.003	1.346±0.045	1.001±0.002	0.978±0.036	1.010±0.007	0.983±0.086

TAB. 3.10: Valeurs des coefficients de corrélation du *System8*, paramétrisés en fonction du p_T du jet (voir figures 3.20 et B.1 à B.5), de la qualité du muon (*loose*, *med* et *med3*) et du point de fonctionnement de SLTNN (*LOOSE*, *MEDIUM* et *TIGHT*). Les erreurs reportées sont uniquement statistiques.

Muon	OP	κ_b	κ_{cl}	β	α	δ	γ
loose	LOOSE	1.067±0.002	1.414±0.030	1.000±0.001	0.962±0.030	0.997±0.005	0.968±0.062
	MEDIUM	1.086±0.002	1.497±0.031	0.999±0.002	0.962±0.034	0.997±0.005	0.968±0.062
	TIGHT	1.137±0.003	1.684±0.033	0.998±0.002	0.959±0.044	0.997±0.005	0.968±0.062
med	LOOSE	1.019±0.002	1.129±0.034	1.000±0.001	0.988±0.019	0.999±0.006	0.974±0.079
	MEDIUM	1.034±0.002	1.225±0.037	1.000±0.001	0.984±0.026	0.999±0.006	0.974±0.079
	TIGHT	1.076±0.003	1.437±0.042	0.999±0.002	0.983±0.038	0.999±0.006	0.974±0.079
med3	LOOSE	1.007±0.002	1.044±0.031	1.000±0.001	0.995±0.011	0.999±0.006	0.981±0.082
	MEDIUM	1.021±0.003	1.136±0.037	1.000±0.001	0.985±0.021	0.999±0.006	0.981±0.082
	TIGHT	1.060±0.003	1.359±0.044	1.000±0.002	0.975±0.034	0.999±0.006	0.981±0.082

TAB. 3.11: Valeurs des coefficients de corrélation du *System8*, paramétrisés en fonction de η du jet (voir figures 3.20 et B.1 à B.5), de la qualité du muon (*loose*, *med* et *med3*) et du point de fonctionnement de SLTNN (*LOOSE*, *MEDIUM* et *TIGHT*). Les erreurs reportées sont uniquement statistiques.

Nous pouvons maintenant résoudre *System8* dont les deux dernières équations ne sont pas linéaires. Nous allons donc utiliser *TMinuit* [61] pour y parvenir. Nous pouvons maintenant comparer les efficacités obtenues avec les données et la simulation, paramétrées en fonction du p_T et η du jet. Comme nous le montre la figure 3.21, on peut voir qu'il y a une différence de quelques pourcents dans la réponse obtenue dans les deux cas. On définit donc un facteur de correction aux événements simulés, noté SF (pour *Scale Factor*). On l'obtient en faisant le rapport de l'efficacité d'étiquetage dans les données et la simulation, et en ajustant cette distribution par un polynôme d'ordre 4 (figure 3.21).

On définit une paramétrisation 2D du facteur de correction SF (voir figure 3.22) en factorisant les composantes p_T et η de l'efficacité d'étiquetage (la corrélation entre p_T et η est ignorée) :

$$\varepsilon(p_T, |\eta|) = \frac{1}{\varepsilon_{\text{all}}} \times (a + b \cdot p_T + c \cdot p_T^2 + d \cdot p_T^3) \times (e + f \cdot |\eta| + g \cdot |\eta|^2 + h \cdot |\eta|^3) \quad (3.12)$$

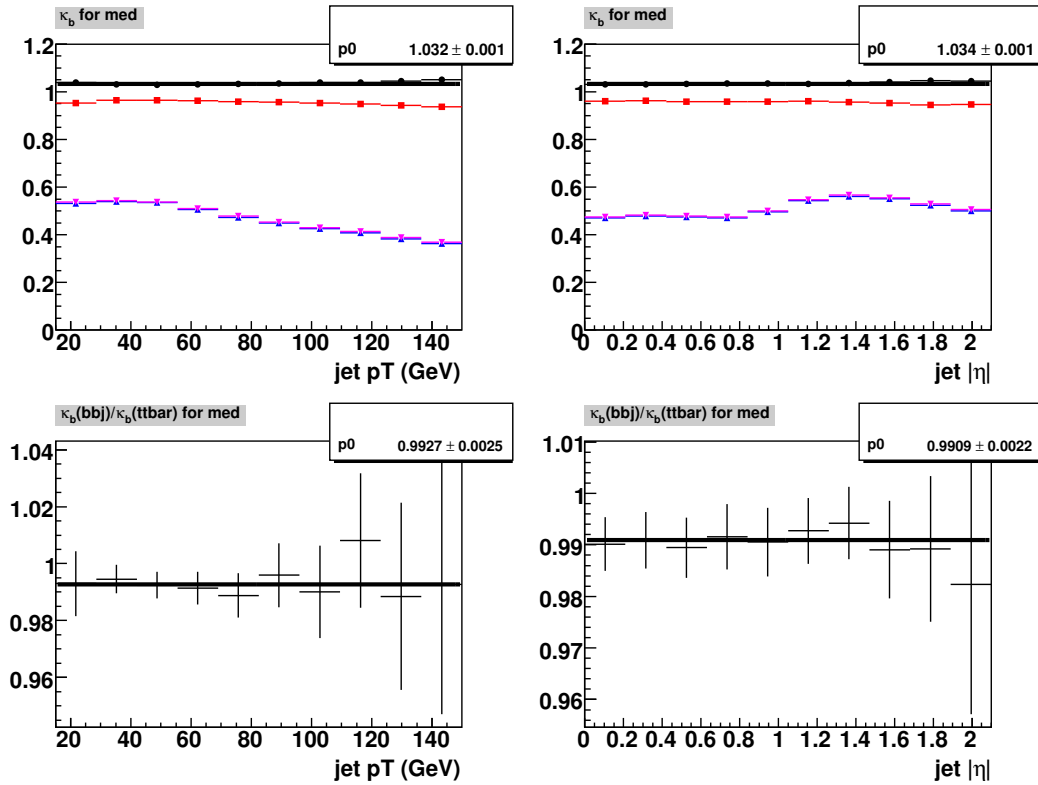


FIG. 3.20: Distributions du coefficient de corrélation κ_b pour un muon de qualité *med* et le point de fonctionnement de SLTNN *MEDIUM*. Le coefficient est paramétrisé en fonction du p_T (à gauche) ou de η (à droite) du jet. En bas, rapport entre κ_b obtenu avec le processus simulé bb +jets et celui obtenu avec $t\bar{t}$. En haut, κ_b est obtenu avec la combinaison des deux processus simulés (points noirs) ajusté par une constante ($\varepsilon_b^{\text{SLTNN}}$ en rouge, $\varepsilon_b^{\text{modJLIP}}$ en violet et $\varepsilon_b^{\text{SLTNN,modJLIP}}$ en bleu). Notons que $\varepsilon_b^{\text{modJLIP}}$ et $\varepsilon_b^{\text{SLTNN,modJLIP}}$ sont pratiquement superposés. Les erreurs sont uniquement statistiques. En faisant la somme quadratique des erreurs du coefficient de corrélation de la figure du haut et du bas, on détermine l'erreur reportée dans les tableaux 3.10 et 3.11.

Le terme ε_{all} représente $\varepsilon_{\text{NN}}^{\text{SLTNN}}$ mesurée indépendamment de p_T et η du jet. Cette efficacité est obtenue après avoir résolu *System8* dans les données (tableau 3.12) et la simulation (tableau 3.13). Les valeurs obtenues sont similaires dans les deux cas.

On calcule l'efficacité d'étiquetage inclusive des jets issus de quarks b (resp. c) dans les données :

$$\text{TRF}_{\text{NN}}^{\text{SLTNN}} = \varepsilon_{\text{NN}}^{\text{SLTNN}}(\text{données}) = \varepsilon_{\text{NN}}^{\text{SLTNN}}(\text{MC}) \times \text{SF} \quad (3.13)$$

On note cette quantité TRF (pour *Tag Rate Function*).

Muon	OP	$\epsilon_{\text{NN}}^{\text{SLTNN}}$
loose	LOOSE	0.914 ± 0.001
	MEDIUM	0.884 ± 0.001
	TIGHT	0.803 ± 0.001
med	LOOSE	0.976 ± 0.001
	MEDIUM	0.956 ± 0.001
	TIGHT	0.893 ± 0.001
med3	LOOSE	0.992 ± 0.001
	MEDIUM	0.974 ± 0.001
	TIGHT	0.918 ± 0.001

TAB. 3.12: $\epsilon_{\text{NN}}^{\text{SLTNN}}$ (indépendante de p_{T} et η du jet) mesurée avec la méthode *System8* dans les échantillons de données *MUinclusive* après avoir résolu *System8*.

Muon	OP	$\epsilon_{\text{NN}}^{\text{SLTNN}}$
loose	LOOSE	0.918 ± 0.001
	MEDIUM	0.891 ± 0.001
	TIGHT	0.812 ± 0.001
med	LOOSE	0.978 ± 0.001
	MEDIUM	0.959 ± 0.001
	TIGHT	0.902 ± 0.001
med3	LOOSE	0.992 ± 0.001
	MEDIUM	0.976 ± 0.001
	TIGHT	0.925 ± 0.001

TAB. 3.13: $\epsilon_{\text{NN}}^{\text{SLTNN}}$ (indépendante de p_{T} et η du jet) dans les échantillons de simulation combinés $b\bar{b}$ et $t\bar{t}$ après avoir résolu *System8*.

Erreurs systématiques

L'estimation des erreurs systématiques prend en compte deux effets.

Le premier effet correspond à l'erreur associée à l'évaluation de $\epsilon_{\text{NN}}^{\text{SLTNN}}$ mesurée par *System8*. Pour ce faire, on fait varier :

- la valeur des coefficients de corrélation de $\pm 1\sigma$ (erreur statistique). Dans le cas de $\kappa_{c\bar{c}}$ qui est très éloigné de 1, on appliquera une variation de ± 1.0 ;
- la valeur arbitraire choisie pour la probabilité de $\text{modJLIP} < 0.1$ de ± 0.025 .

L'erreur systématique correspond à la valeur de l'erreur relative entre $\epsilon_{\text{NN}}^{\text{SLTNN}}$ obtenue initialement et cette même efficacité après variation des coefficients de corrélation et de modJLIP . On la calcule pour chaque coefficient de corrélation et pour la probabilité de modJLIP . Le tableau 3.14 montre ces erreurs ainsi que l'erreur totale (somme quadratique de toutes les contributions). On pourra noter que l'erreur systématique totale de la résolution de *System8* (δ_{sys8}) est inférieure à 5%.

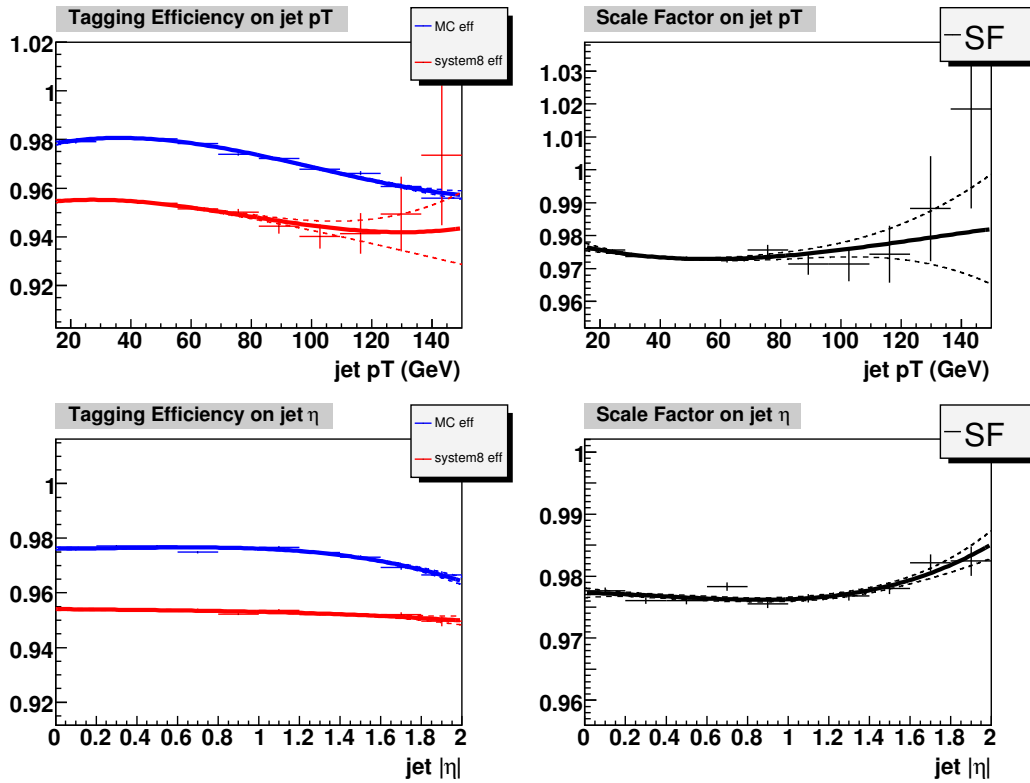


FIG. 3.21: Les figures de gauche représentent ε_{NN}^{SLTNN} dans les données (courbe rouge) et dans les lots de simulation (courbe bleue). Les figures de droite montrent les facteurs SF correspondants (rapport de ε_{NN}^{SLTNN} dans les données et dans les lots simulés). Ces distributions sont paramétrisées en fonction du p_T (en haut) et η (en bas) du jet. Elles ont été obtenues en utilisant un muon de qualité *med3* et le point de fonctionnement *MEDIUM*.

La variation de $\pm 1\sigma$ de κ_b n'est pas compatible avec l'unité. Le tableau 3.15 montre l'erreur systématique obtenue pour une variation de ± 0.1 . L'erreur systématique associée à ce coefficient augmente dans ce cas jusqu'à environ 9%. Ce résultat confirme que l'efficacité d'étiquetage est sensible au coefficient de corrélation κ_b .

Le deuxième effet vise à définir l'impact du lot de simulation utilisé pour l'évaluation des coefficients de corrélation en effectuant un test de fermeture (figure 3.23). L'erreur associée correspond à l'écart à 1 du rapport nombre de jets étiquetés de manière directe et de manière prédite. On entend par :

- étiquetage de manière directe : un jet passe la condition de coupure sur le réseau de neurone et est corrigé par la valeur associée du SF ;
- étiquetage de manière prédite : on associe au jet la probabilité qu'il soit étiqueté en utilisant la TRF.

Le tableau 3.16 (resp. 3.17) montre l'erreur faite sur l'étiquetage des jets issus de quark b (resp. c) pour chacun ou la combinaison ($\delta_{comb.}$) des lots de simulation bb +jets, $t\bar{t}$ et direct

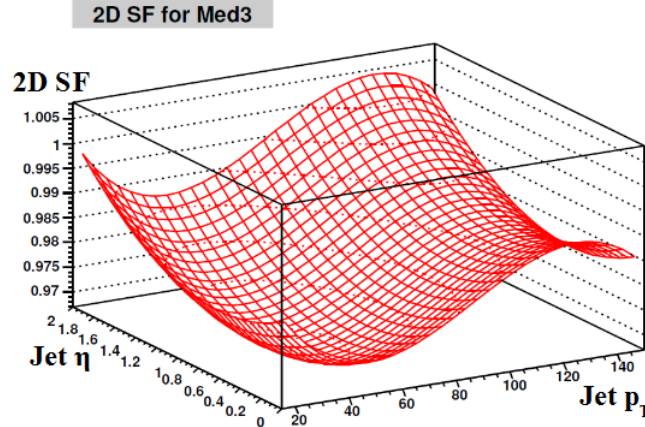


FIG. 3.22: Facteur SF 2D paramétrisé en fonction de p_T et η du jet pour un muon de qualité *med3* et le point de fonctionnement *MEDIUM*.

$b\bar{b}$ (resp. $cc+jets$, $t\bar{t}$ ou direct $c\bar{c}$). A partir de ces quatre erreurs, on choisira celle qui a la plus grande valeur ($\delta_{selec.}$).

On peut donc finalement associer une erreur systématique à SF et TRF. On les définit comme suit :

$$\delta_{SF} = \sqrt{(\delta_{sys8})^2 + (\delta_{selec.})^2} \quad (3.14)$$

$$\delta_{TRF_b} = \sqrt{(\delta_{SF})^2 + (\delta_{b\ comb.})^2} \quad (3.15)$$

$$\delta_{TRF_c} = \sqrt{(\delta_{SF})^2 + (\delta_{c\ comb.})^2} \quad (3.16)$$

et le tableau 3.18 montre le détail des erreurs systématiques obtenues.

Muon	OP	κ_b	κ_{cl}	β	α	δ	γ	modJLIP	δ_{sys8}
loose	LOOSE	$\pm 0.110\%$	$\pm 4.456\%$	$\pm 0.091\%$	$\pm 0.797\%$	$\pm 0.001\%$	$\pm 1.258\%$	$\pm 0.170\%$	$\pm 4.703\%$
	MEDIUM	$\pm 0.115\%$	$\pm 3.860\%$	$\pm 0.105\%$	$\pm 0.633\%$	$\pm 0.001\%$	$\pm 1.831\%$	$\pm 0.220\%$	$\pm 4.328\%$
	TIGHT	$\pm 0.128\%$	$\pm 1.952\%$	$\pm 0.142\%$	$\pm 0.483\%$	$\pm 0.001\%$	$\pm 3.339\%$	$\pm 0.405\%$	$\pm 3.924\%$
med	LOOSE	$\pm 0.114\%$	$\pm 1.363\%$	$\pm 0.053\%$	$\pm 1.235\%$	$\pm 0.000\%$	$\pm 0.247\%$	$\pm 0.035\%$	$\pm 1.861\%$
	MEDIUM	$\pm 0.118\%$	$\pm 3.215\%$	$\pm 0.071\%$	$\pm 1.267\%$	$\pm 0.000\%$	$\pm 0.591\%$	$\pm 0.063\%$	$\pm 3.509\%$
	TIGHT	$\pm 0.130\%$	$\pm 3.846\%$	$\pm 0.112\%$	$\pm 0.568\%$	$\pm 0.000\%$	$\pm 1.750\%$	$\pm 0.181\%$	$\pm 4.271\%$
med3	LOOSE	$\pm 0.115\%$	$\pm 0.427\%$	$\pm 0.034\%$	$\pm 0.457\%$	$\pm 0.000\%$	$\pm 0.078\%$	$\pm 0.013\%$	$\pm 0.642\%$
	MEDIUM	$\pm 0.119\%$	$\pm 1.766\%$	$\pm 0.058\%$	$\pm 1.185\%$	$\pm 0.000\%$	$\pm 0.319\%$	$\pm 0.055\%$	$\pm 2.155\%$
	TIGHT	$\pm 0.131\%$	$\pm 4.376\%$	$\pm 0.100\%$	$\pm 0.612\%$	$\pm 0.000\%$	$\pm 1.370\%$	$\pm 0.158\%$	$\pm 4.632\%$

TAB. 3.14: Erreur systématique sur l'efficacité d'étiquetage liée à la résolution de *System8* après avoir fait varier chaque coefficient de corrélation et la probabilité de modJLIP (voir texte).

Muon	OP	κ_b
loose	LOOSE	$\pm 9.279\%$
	MEDIUM	$\pm 9.273\%$
	TIGHT	$\pm 9.184\%$
med	LOOSE	$\pm 9.107\%$
	MEDIUM	$\pm 9.456\%$
	TIGHT	$\pm 9.422\%$
med3	LOOSE	$\pm 9.197\%$
	MEDIUM	$\pm 9.429\%$
	TIGHT	$\pm 9.457\%$

TAB. 3.15: Erreur systématique sur l'efficacité d'étiquetage liée à la résolution de *System8* après avoir fait varier le coefficient de corrélation κ_b de ± 0.1 .

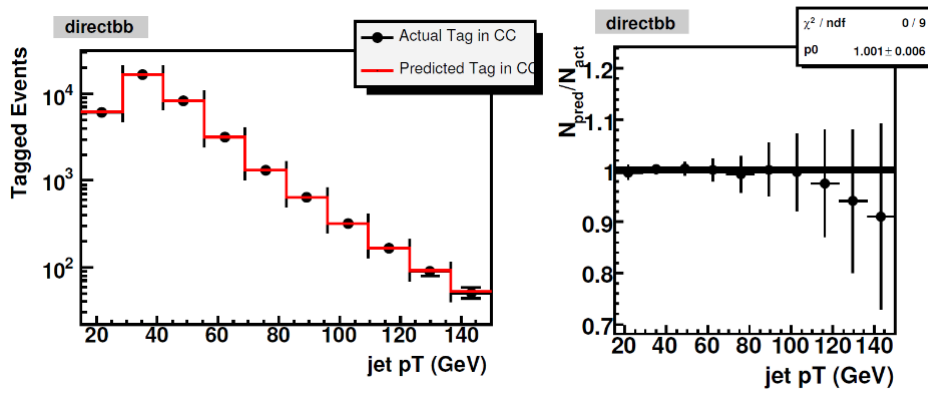


FIG. 3.23: Distributions relatives au test de fermeture afin de définir l'impact du lot de simulation utilisé pour évaluer les coefficients de corrélation. Seul le lot direct $b\bar{b}$ est présenté ici et l'on a utilisé un muon de qualité *med* et un point de fonctionnement de la méthode d'étiquetage SLTNN *LOOSE* pour des jets issus de quark b . La figure de gauche présente le nombre de jets étiqueté dans CC de manière directe (en noir) et de manière prédite (en utilisant la TRF, en rouge). La figure de droite montre le rapport des deux distributions précédentes avec son ajustement (i.e. *fit*) par une constante.

Muon	OP	bbj	$t\bar{t}$	directbb	δ_b comb.	$\delta_{selec.}$
loose	LOOSE	0.80%	0.67%	0.23%	0.01%	0.80%
	MEDIUM	1.00%	0.78%	0.13%	0.03%	1.00%
	TIGHT	1.55%	1.02%	0.21%	0.15%	1.55%
med	LOOSE	0.29%	0.28%	0.05%	0.02%	0.29%
	MEDIUM	0.48%	0.42%	0.01%	0.01%	0.48%
	TIGHT	0.98%	0.63%	0.13%	0.12%	0.98%
med3	LOOSE	0.16%	0.05%	0.02%	0.04%	0.16%
	MEDIUM	0.33%	0.12%	0.05%	0.08%	0.33%
	TIGHT	0.78%	0.38%	0.03%	0.15%	0.78%

TAB. 3.16: Erreurs systématiques relatives à la comparaison de l'étiquetage des jets issus de quark b de manière directe et de manière prédite (en utilisant la TRF).

Muon	OP	ccj	$t\bar{t}$	directcc	δ_c comb.
loose	LOOSE	0.21%	8.67%	5.97%	3.07%
	MEDIUM	0.18%	10.25%	5.85%	3.63%
	TIGHT	0.21%	14.57%	7.05%	5.12%
med	LOOSE	0.01%	2.93%	1.74%	1.03%
	MEDIUM	0.01%	4.66%	2.69%	1.65%
	TIGHT	0.11%	7.83%	5.23%	2.73%
med3	LOOSE	0.06%	0.65%	0.98%	0.21%
	MEDIUM	0.10%	2.00%	2.53%	0.68%
	TIGHT	0.13%	4.83%	5.02%	1.66%

TAB. 3.17: Erreurs systématiques relatives à la comparaison de l'étiquetage des jets issus de quark c de manière directe et de manière prédite (en utilisant la TRF).

Muon	TRF	Erreurs systématiques totales		
		SF	TRF_b	TRF_c
loose	LOOSE	4.77%	4.77%	5.67%
	MEDIUM	4.44%	4.44%	5.74%
	TIGHT	4.22%	4.22%	6.63%
med	LOOSE	1.88%	1.88%	2.15%
	MEDIUM	3.54%	3.54%	3.91%
	TIGHT	4.38%	4.38%	5.16%
med3	LOOSE	0.66%	0.66%	0.69%
	MEDIUM	2.18%	2.18%	2.28%
	TIGHT	4.70%	4.70%	4.98%

TAB. 3.18: Erreurs systématiques totales pour SF, TRF_b , TRF_c .

3.6.2 Taux de mauvaise identification

Après avoir défini l'efficacité d'étiquetage dans les données, il nous faut maintenant obtenir le taux de mauvaise identification des jets issus des quarks légers (u, d ou s). Pour ce faire, nous pouvons utiliser une caractéristique du paramètre d'impact en signification IP_{sig} du muon dans le jet. La figure 3.24 montre cette variable pour le lot simulé direct $b\bar{b}$ et pour le lot de données $MU_{inclusive}$ étiqueté *Tight* avec la méthode d'étiquetage NN décrite en section 3.1. On peut noter un excès de jet pour les valeurs positives de IP_{sig} et représente la contribution des jets issus de saveurs lourdes.

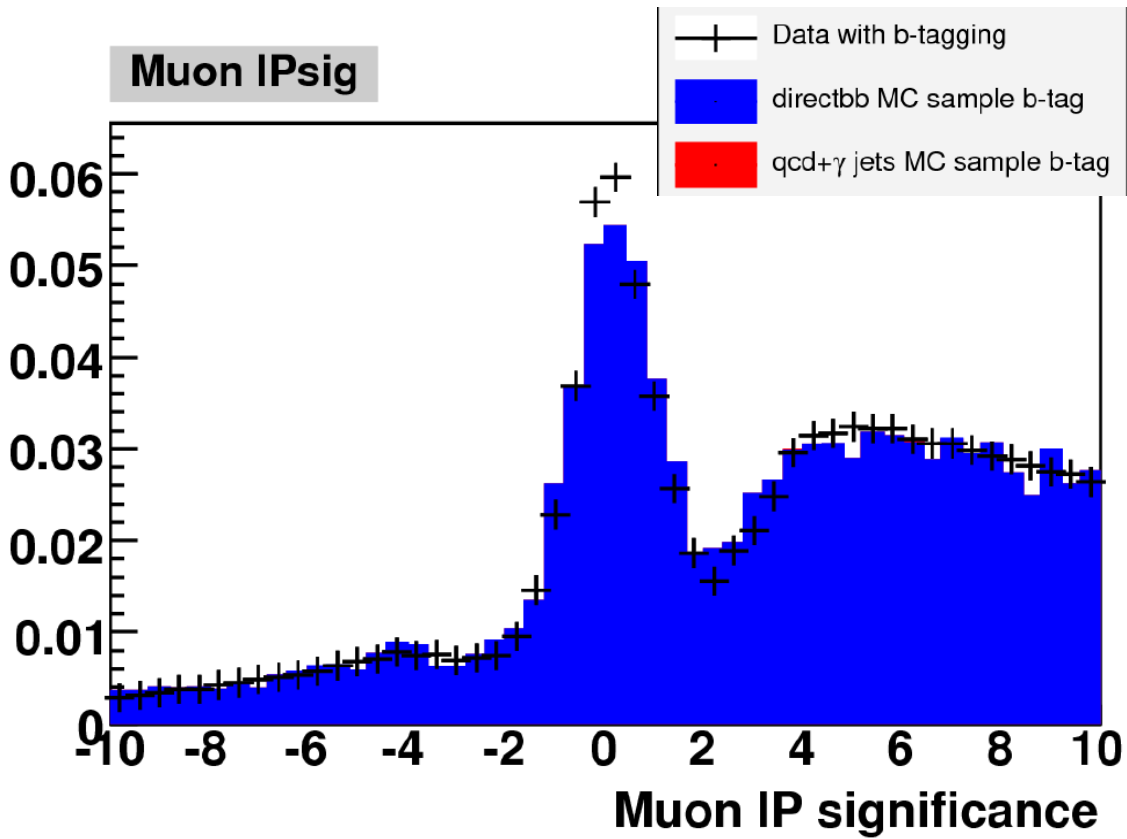


FIG. 3.24: Paramètre d'impact en signification IP_{sig} du muon de qualité *med* pour le lot simulé direct $b\bar{b}$ (courbe bleue) et pour le lot de données $MU_{inclusive}$ (courbe rouge). Les lots de simulation et de données sont étiquetés *Tight* avec la méthode NN. Les distributions sont normalisées à l'unité.

On définit donc la méthode du taux d'étiquetage négatif (notée NTR pour *Negative Tag Rate*), en demandant à un jet d'être étiqueté par SLTNN et d'avoir IP_{sig} du muon négatif. Le taux de mauvaise identification $\varepsilon_{léger}$ est défini comme :

$$\varepsilon_{léger} = \varepsilon_{données}^- \times F_{hf} \times F_{ll} \quad (3.17)$$

où :

- $\varepsilon_{\text{données}}^-$ représente le taux d'étiquetage négatif (IPsig <0) estimé sur le lot de données QCD ;
- $F_{\text{hf}} (= \frac{\varepsilon_{\text{jet léger}}^-}{\varepsilon_{\text{jet inclusif}}^-})$ représente le rapport entre la fraction de jets issus de quark léger étiquetés négatifs et tous les jets étiquetés négatifs. Si ce rapport est inférieur à 1, cela veut dire qu'il y a des jets issus de quark de saveur lourde dans le lot ;
- $F_{\text{ll}} (= \frac{\varepsilon_{\text{jet léger}}^+}{\varepsilon_{\text{jet léger}}^-})$ représente le rapport entre la fraction de jets issus de quark léger étiquetés positif et ceux issus de quark léger étiquetés négatifs. Ce facteur est sensible à la désintégration des hadrons ayant un long temps de vie dans les jets issus de quark léger.

Le produit $F_{\text{hf}} \times F_{\text{ll}}$, évalué uniquement avec des lots simulés, peut être vu comme un facteur SF d'étiquetage négatif et il sera nommé NT_SF. Les quantités F_{hf} et F_{ll} seront déterminées à partir des lots de simulation combinés QCD $q\bar{q}$ et γ +jets. L'efficacité $\varepsilon_{\text{léger}}$ peut être vue comme $\text{TRF}_{\text{léger}}$.

La figure 3.25 montre les quatre quantités de l'équation 3.17 pour un muon de qualité *med*, un point de fonctionnement *MEDIUM* et un jet dans la région CC du calorimètre. Le taux d'étiquetage négatif varie selon la région du calorimètre. $\varepsilon_{\text{données}}^-$ et $\varepsilon_{\text{léger}}$ seront paramétrisées en fonction du p_T du jet pour chacune des trois régions CC, ICR et EC dans les données. Pour des raisons de statistique dans les lots de simulation, les facteurs F_{hf} et F_{ll} seront déterminés sur la région complète en η (tableau 3.19).

Le tableau 3.20 résume les valeurs obtenues du taux de mauvaise identification pour les trois qualités de muon, les trois points de fonctionnements et les trois régions du détecteur. On notera que dans l'ensemble, il est inférieur à 1%.

Muon	OP	F_{hf} (%)	F_{ll} (%)
loose	LOOSE	0.528 ± 0.008	1.641 ± 0.048
	MEDIUM	0.448 ± 0.010	1.767 ± 0.064
	TIGHT	0.278 ± 0.012	2.128 ± 0.126
med	LOOSE	0.575 ± 0.008	1.245 ± 0.036
	MEDIUM	0.492 ± 0.009	1.346 ± 0.047
	TIGHT	0.320 ± 0.012	1.588 ± 0.090
med3	LOOSE	0.582 ± 0.008	1.194 ± 0.034
	MEDIUM	0.499 ± 0.009	1.280 ± 0.045
	TIGHT	0.321 ± 0.012	1.541 ± 0.087

TAB. 3.19: F_{hf} et F_{ll} mesurées à l'aide du lot de simulation combinés QCD $q\bar{q}$ et γ +jets. Il n'y pas de séparation des résultats suivant les trois régions du calorimètre pour des raisons de faible statistique dans le lot de simulation.

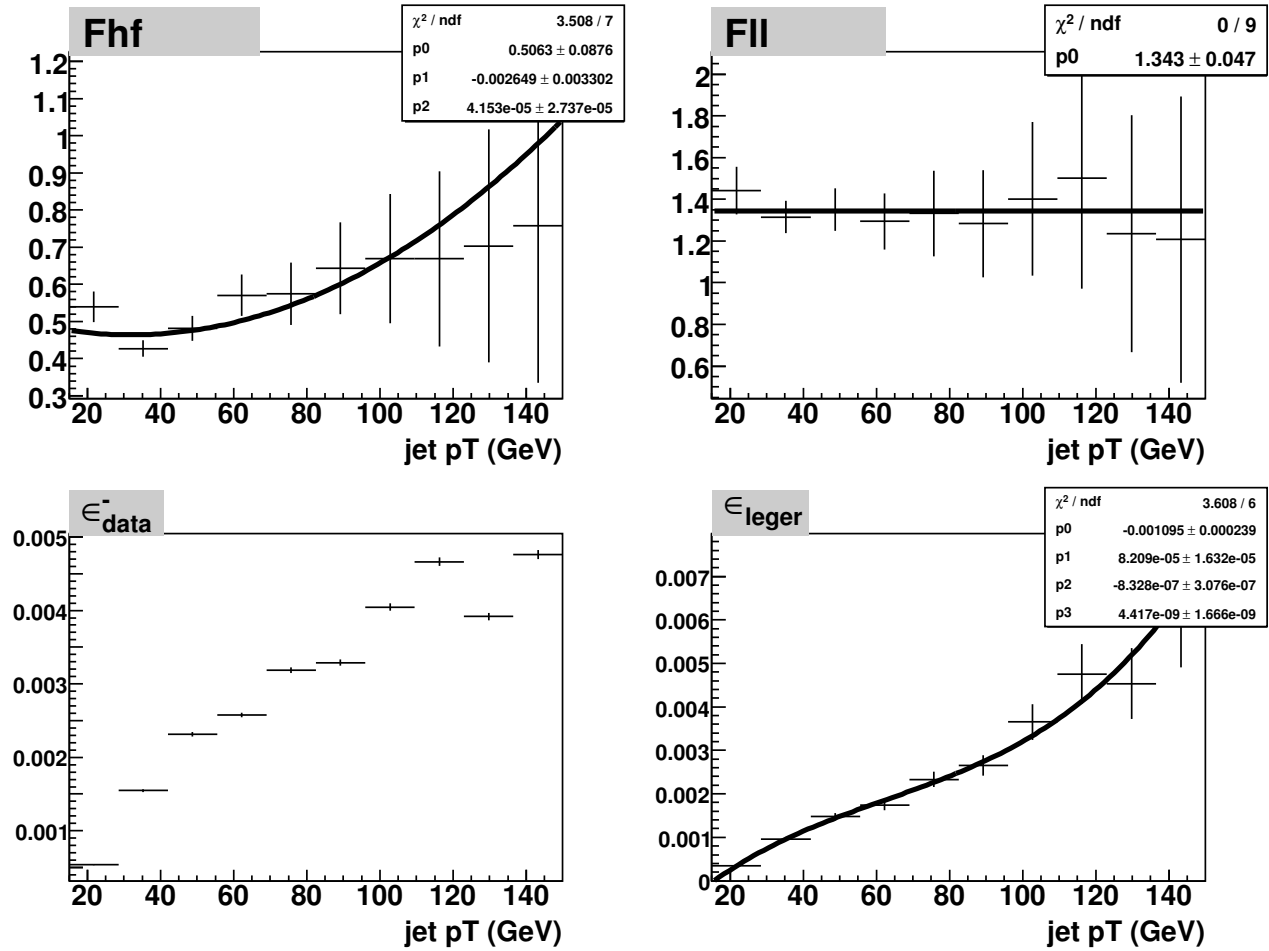


FIG. 3.25: Distributions des facteurs F_{hf} (en haut à gauche) et F_{ll} (en haut à droite), du taux d'étiquetage négatif $\varepsilon_{\text{données}}^-$ (en bas à gauche) et du taux de mauvaise identification $\varepsilon_{\text{leger}}$ (en bas à droite) définis dans l'équation 3.17. Ces distributions sont paramétrées en fonction du p_T du jet et sont représentées ici pour un muon de qualité *med*, un point de fonctionnement *MEDIUM* et dans la région CC du calorimètre. On obtient $\text{TRF}_{\text{leger}}$ en ajustant $\varepsilon_{\text{leger}}$ par un polynôme de degré 3.

3.6.3 Erreurs systématiques

L'évaluation de l'erreur systématique associée au taux de mauvaise identification a deux composantes.

La première est évaluée par un test de fermeture en comparant le nombre de jets étiquetés de manière directe (avec la méthode NTR) et de manière prédite (avec les $\text{TRF}_{\text{leger}}$). Le résultat de ce test apparaît dans le tableau 3.21 et est inférieure à 4% (ce résultat a été évalué pour chacune des trois régions du détecteur).

On détermine la deuxième erreur systématique en faisant varier la fraction en quark b ou c de $\pm 20\%$ (dans les lots simulés combinés QCD $q\bar{q}$ et γ +jets) et on regarde l'évolution du taux

Muon	OP	Taux de mauvaise identification (%)		
		CC	ICR	EC
loose	LOOSE	0.334 ± 0.011	0.153 ± 0.005	0.046 ± 0.002
	MEDIUM	0.237 ± 0.010	0.107 ± 0.005	0.030 ± 0.001
	TIGHT	0.098 ± 0.007	0.044 ± 0.003	0.011 ± 0.001
med	LOOSE	0.204 ± 0.007	0.150 ± 0.005	0.033 ± 0.001
	MEDIUM	0.152 ± 0.006	0.109 ± 0.004	0.023 ± 0.001
	TIGHT	0.071 ± 0.005	0.048 ± 0.003	0.009 ± 0.001
med3	LOOSE	0.181 ± 0.006	0.150 ± 0.005	0.039 ± 0.001
	MEDIUM	0.134 ± 0.005	0.108 ± 0.004	0.027 ± 0.001
	TIGHT	0.066 ± 0.004	0.050 ± 0.003	0.011 ± 0.001

TAB. 3.20: Taux de mauvaise identification mesurée à l'aide du lot de données QCD et dans les trois régions du calorimètre.

de mauvaise identification. Pour les quatre cas de figure (fraction de quarks b varié de +20%, ..., fraction de quarks c varié de -20%) on pondère le jet par 1.2 ou 0.8 et on obtient ainsi de nouvelles $TRF_{\text{léger}}$. L'erreur systématique est l'erreur relative entre ces nouvelles $TRF_{\text{léger}}$ et celle obtenue initialement. Cette variation agit sur les facteurs de NT_SF évalués sur la simulation. Comme ces facteurs ont été déterminés sur l'ensemble de la plage en η du jet, les erreurs systématiques associées ne distingueront pas les régions du calorimètre. Les erreurs systématiques associées sont reportées dans les tableaux 3.22 (variation de la fraction en quark b) et 3.23 (variation de la fraction en quark c). Cette erreur systématique est inférieure à 10%. Finalement l'erreur systématique totale correspond à la somme quadratique de celle décrite ci-dessus. On peut voir à l'aide du tableau 3.24 que l'erreur systématique totale associée au taux de mauvaise identification pour chacune des trois régions du calorimètre varie entre 6 et 11%.

Muon	OP	Erreur Systématique		
		CC	ICR	EC
loose	LOOSE	1.98%	3.05%	3.70%
	MEDIUM	1.75%	2.84%	3.56%
	TIGHT	1.09%	1.66%	3.85%
med	LOOSE	1.14%	2.98%	2.69%
	MEDIUM	0.98%	2.83%	1.99%
	TIGHT	0.45%	2.37%	2.69%
med3	LOOSE	1.17%	3.11%	2.26%
	MEDIUM	0.98%	2.90%	2.01%
	TIGHT	1.22%	2.23%	1.13%

TAB. 3.21: Erreur systématique obtenue avec le test de fermeture en comparant jets étiquetés de manière directe (avec la méthode NTR) et de manière prédite (avec les $\text{TRF}_{\text{léger}}$) pour chacune des trois régions du calorimètre.

Muon	OP	Erreur Systématique
loose	LOOSE	4.80%
	MEDIUM	5.99%
	TIGHT	8.85%
med	LOOSE	3.87%
	MEDIUM	5.02%
	TIGHT	7.93%
med3	LOOSE	3.69%
	MEDIUM	4.76%
	TIGHT	7.71%

TAB. 3.22: Erreur systématique obtenue en faisant varier la fraction en quarks b . Il n'y a pas de séparation des résultats suivant les trois régions du calorimètre pour des raisons de faible statistique dans le lot de simulation.

Muon	OP	Erreur Systématique
loose	LOOSE	4.59%
	MEDIUM	5.01%
	TIGHT	5.53%
med	LOOSE	4.61%
	MEDIUM	5.12%
	TIGHT	5.63%
med3	LOOSE	4.64%
	MEDIUM	5.23%
	TIGHT	5.85%

TAB. 3.23: Erreur systématique obtenue en faisant varier la fraction en quarks c . Il n'y pas de séparation des résultats suivant les trois régions du calorimètre pour des raisons de faible statistique dans le lot de simulation.

Muon	OP	Erreur Systématique		
		CC	ICR	EC
loose	LOOSE	6.93%	7.31%	7.60%
	MEDIUM	8.00%	8.31%	8.58%
	TIGHT	10.49%	10.57%	11.12%
med	LOOSE	6.12%	6.72%	6.59%
	MEDIUM	7.24%	7.71%	7.44%
	TIGHT	9.74%	10.01%	10.09%
med3	LOOSE	6.04%	6.70%	6.35%
	MEDIUM	7.14%	7.64%	7.35%
	TIGHT	9.75%	9.93%	9.74%

TAB. 3.24: Erreur systématique totale reliée au taux de mauvaise identification ($TRF_{\text{léger}}$) pour chacune des trois régions du calorimètre.

3.7 Conclusion

Nous avons vu en détail comment est développée la nouvelle méthode SLTNN [62] d'étiquetage de jets issus de désintégration semileptonique de quark b. Les techniques nécessaires à son élaboration sont communes avec la méthode standard d'étiquetage NN [57].

L'ajout de la méthode SLTNN dans les analyses vise à augmenter l'efficacité d'étiquetage par jets issus de quark b d'environ 5%. Les algorithmes NN et SLTNN utilisent un réseau de neurone qui combine plusieurs variables afin d'augmenter l'efficacité d'étiquetage et de diminuer le taux de mauvaise identification.

Ces deux caractéristiques ont été évaluées dans les lots de données et de simulation. Elles nécessitent de définir des corrections à appliquer aux lots de simulation. L'utilisation de ces méthodes dans une analyse se fera de la manière suivante :

- Etiquetage des jets dans un lot de données : on donne les caractéristiques du jet considéré en entrée du réseau de neurones et on obtient la valeur de sortie de ce réseau. Si cette sortie est supérieure à la valeur fixée par le point de fonctionnement choisi par l'utilisateur, le jet sera étiqueté comme étant issu d'un quark b ;
- Etiquetage des jets d'un lot de simulation :
 - Le jet provient d'un quark b ou c : l'utilisateur peut étiqueter le jet de manière directe (coupure sur la sortie du réseau de neurones et on utilise la correction SF) ou prédite (on utilise la TRF_b ou TRF_c) ;
 - Le jet provient d'un quark léger (u, d ou s) : identique au cas précédent en remplaçant SF par NT_SF, et TRF_b (ou TRF_c) par $TRF_{léger}$.

En résumé, sur un lot de jets inclusifs (désintégrations semi-leptoniques ou pas) :

- la méthode d'étiquetage NN a une efficacité d'étiquetage $\sim 70\%$ pour un taux de mauvaise identification $\sim 5\%$;
- la méthode d'étiquetage SLTNN a une efficacité d'étiquetage $\sim 5\%$ pour un taux de mauvaise identification $< 1\%$;
- les erreurs systématiques sont évaluées et ont le même ordre de grandeur pour les deux méthodes NN et SLTNN : $\sim 5\%$ pour l'efficacité d'étiquetage et $\leq 10\%$ pour le taux de mauvaise identification.

Pour un lot de jets muoniques (i.e. désintégrations semi-leptoniques) où les muons sont de qualité *loose* (figure 3.26) :

- la méthode d'étiquetage NN a une efficacité d'étiquetage 65% (85%) pour un taux de mauvaise identification 1% (30%) ;
- la méthode d'étiquetage SLTNN a une efficacité d'étiquetage 68% (95%) pour un taux de mauvaise identification 1% (30%) ;

Ces résultats ne tiennent pas compte de l'efficacité d'étiquetage du muon dans un jet (ϵ_{μ}^{SLT}). En conclusion, pour un taux de mauvaise identification fixé à 1%, SLTNN a une efficacité d'étiquetage meilleure que NN entre 5% et 10% en moyenne.

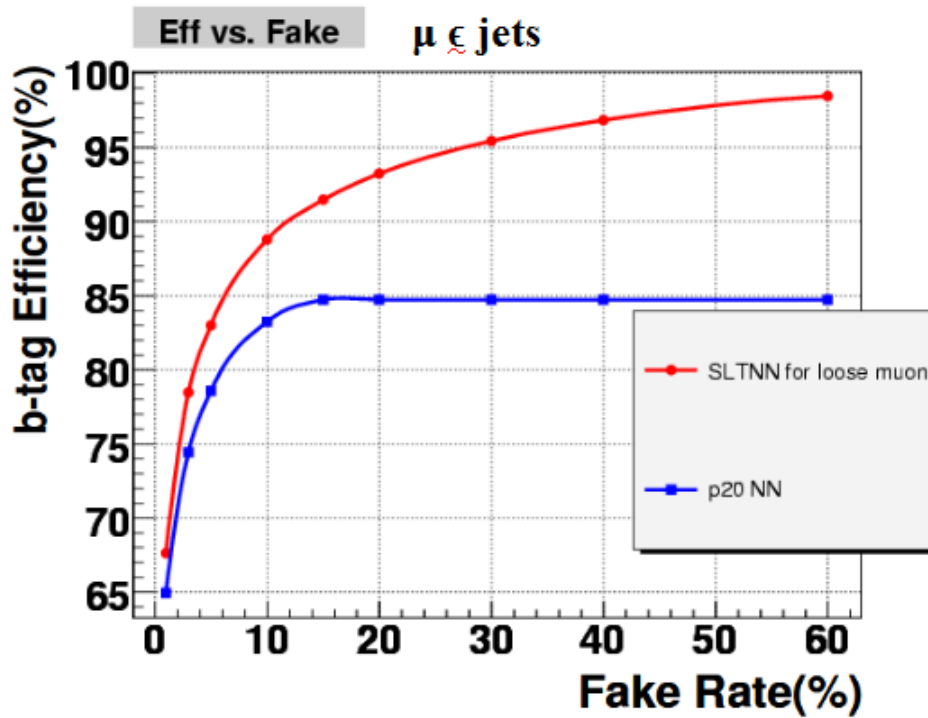


FIG. 3.26: Efficacité d'étiquetage des jets issus de quark b en fonction du taux de mauvaise identification pour des jets issus de désintégration semileptonique de quark b . On compare ces résultats pour la méthode d'étiquetage NN (en bleu) et SLTNN (en rouge). Ces efficacités ont été obtenues sur des lots de simulation (direct $b\bar{b}$ pour l'efficacité d'étiquetage, γ +jets et $q\bar{q}$ combinés pour le taux de mauvaise identification) composé de jets muoniques dont la qualité du muon est *loose*.

Nous verrons dans le prochain chapitre, les résultats obtenus avec les méthodes d'étiquetage NN et SLTNN dans l'analyse de recherche du boson de Higgs dans le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$.

Recherche du boson de Higgs dans le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$

Ce chapitre d'analyse détaille la recherche du boson de Higgs dans le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ avec l'échantillon de données collectées par le détecteur DØ présenté au chapitre 2. L'état final de ce canal se compose de deux jets de quarks b issus de la désintégration du boson de Higgs et de deux neutrinos issus de la désintégration du boson Z (voir figure 4.1). Les neutrinos, n'interagissant pas avec la matière du détecteur, contribueront sous forme d'énergie transverse manquante (notée \cancel{E}_T). Ce canal d'étude est aussi sensible au canal WH ($W \rightarrow l\nu$) lorsque le lepton n'est pas identifié.

L'analyse se fait en deux étapes. Il nous faut tout d'abord réduire les bruits de fond dominants (multijet et W/Z+jets) à l'aide d'une sélection cinématique des événements. Ensuite, nous utiliserons l'outil officiel de *b-tagging* de la collaboration DØ : le "Neural Net b-

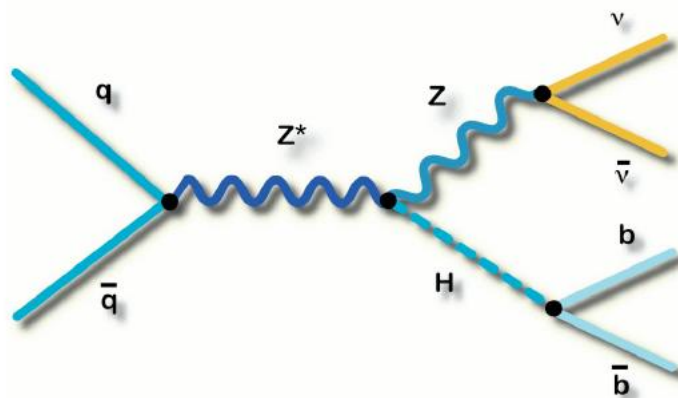


FIG. 4.1: Diagramme de Feynman du canal de production du boson de Higgs associé à un boson Z. Le boson de Higgs se désintègre en deux quarks b et le boson Z en deux neutrinos.

tagger" [57] (noté NN). L'objectif de ce chapitre est de comparer les performances du NN et du "soft muon tagging" développé sur les données du *RunIIb* (noté SLTNN [62]) décrit au chapitre 3 pour des jets issus de désintégration semi-leptonique de quark *b*, que l'on nommera jets muoniques.

4.1 Lot de données

Le lot de données utilisé (3 fb^{-1} : *RunIIb1* et une partie du *RunIIb2*) ne correspond qu'à une fraction des données du *RunIIb* (7 fb^{-1}) car le but est de faire une étude de performances de SLTNN sur un lot maîtrisé. L'analyse présentée utilise les données enregistrées entre juin 2006 et septembre 2008. Les données du Run IIa (1 fb^{-1}) ne sont pas utilisées. L'ensemble des données utilisées par notre analyse est regroupé dans des lots nommés :CSG_CAF_NP_PASS2_p21.10.00 et CSG_CAF_NP_PASS2_p21.10.00_p20.12.0x ($x = 0,1,2,3,4$). Le tableau 4.1 donne les luminosités intégrées enregistrées pour chacune des conditions de déclenchement utilisées (voir section 4.1.1).

Liste	Conditions de déclenchement	$\mathcal{L}_{\text{fournie}}$	$\mathcal{L}_{\text{enregistrée}}$	$\mathcal{L}_{\text{disponible}}$
v15.00-v15.20	JT1_ACO_MHT_HT	321.1	250.5	209.5
v15.20-v16.00	JT1_MET	1615.3	1535.3	1415.6
v16.00-v17.00	JT1_MET	1620.6	1555.9	1411.1
Total		3557.0	3341.7	3036.2

TAB. 4.1: Luminosités intégrées (en pb^{-1}) pour les différentes conditions de déclenchement utilisées dans l'analyse. $\mathcal{L}_{\text{fournie}}$ correspond à la luminosité intégrée fournie par le Tevatron, $\mathcal{L}_{\text{enregistrée}}$ celle enregistrée par le détecteur DØ et $\mathcal{L}_{\text{disponible}}$ celle disponible pour l'analyse après les coupures des critères sur la qualité des données.

4.1.1 Conditions de déclenchement

Les conditions de déclenchement (ou appelé *trigger* en anglais) se font sur trois niveaux. Pour les listes de déclenchement utilisées dans cette analyse, on prend en compte trois topologies correspondant aux conditions de déclenchement dites "jets+ \cancel{E}_T " : mono-jet, di-jet et multi-jet ([63], [64]). Afin d'augmenter l'efficacité du signal ZH, les conditions de déclenchement se font par l'intermédiaire d'un "OR-ing", c.à.d : déclenchement sur la topologie mono-jet OU di-jet OU multi-jet.

Les conditions de déclenchement utilisées pour notre lot de données sont :

- MJ_ACO_MHT_HT OU JT1_ACO_MHT_HT OU JT2_ACO_MHT_HT pour les listes de trigger de v15.00 à v15.20. Ces conditions de déclenchements sont relatives aux données prises pendant les premiers mois du Run IIb :

- au niveau 1 :
 - * mono-jet : au moins 1 jet ayant $E_T > 30$ GeV et $|\eta_{\text{det}}| < 3.2$, $\cancel{E}_T > 24$ GeV ;
 - * di-jet : au moins 2 jets ayant $E_T > 8$ GeV et $|\eta_{\text{det}}| < 2.4$, $\cancel{E}_T > 24$ GeV, acceptant toute paire de jets ayant $4 < E_T < 20$ GeV qui ne sont pas dos-à-dos par rapport à l'angle azimutal ϕ . Avec l'augmentation de la luminosité instantanée, une nouvelle condition a été rajoutée : au moins 1 jet ayant $E_T > 20$ GeV et $|\eta_{\text{det}}| < 2.4$;
 - * multi-jet : au moins 3 jets ayant $E_T > 8$ GeV et $|\eta_{\text{det}}| < 3.2$, au moins 2 jets ayant $E_T > 15$ GeV et $|\eta_{\text{det}}| < 2.4$, au moins 1 jet ayant $E_T > 30$ GeV et $|\eta_{\text{det}}| < 2.4$;
 - au niveau 2 : au moins un jet ayant $E_T > 20$ GeV et $|\eta_{\text{det}}| < 2.4$, $\cancel{H}_T > 20$ GeV (\cancel{H}_T est la somme vectorielle des impulsions transverses des jets, calculée à partir de jets ayant $E_T > 10$ GeV), $H_T > 35$ GeV (H_T est la somme scalaire des impulsions transverses des jets, calculée à partir de jets ayant $E_T > 6$ GeV et $|\eta_{\text{det}}| < 2.6$), $\Delta\Phi(\text{jet1},\text{jet2}) < 168.75^\circ$ (différence d'angle azimutal entre les 2 jets de plus grand p_T que l'on appelle acoplanarité) ;
 - au niveau 3 : au moins un jet ayant $E_T > 9$ GeV, $\Delta\Phi(\text{jet1},\text{jet2}) < 170^\circ$, $H_T > 50$ GeV, $\cancel{H}_T > 30$ GeV, $\Delta\Phi_{\text{min}}(\text{jets},\cancel{H}_T) > 25^\circ$ (différence minimale d'angle azimutal entre les jets et \cancel{H}_T).
- MJ_MET OU JT1_MET OU JT2_MET à partir de la liste de trigger v15.20. Elles remplacent les conditions précédentes. En effet, l'un des objectifs du Run IIb était de pouvoir doubler (au moins) la luminosité instantanée. Les conditions de déclenchement spécifiques aux topologies "jets+ \cancel{E}_T " ont été revues au niveau 3 où \cancel{E}_T a été introduit. Les nouvelles conditions sont : au moins deux jets ayant $E_T > 9$ GeV ET $\cancel{H}_T > 25$ GeV ET $\cancel{E}_T > 25$ GeV ET $\Delta\Phi(\text{jet1},\text{jet2}) < 170^\circ$ ET $\Delta\Phi_{\text{min}}(\text{jets},\cancel{H}_T) > 25^\circ$. Les conditions au niveau 1 et 2 restent les mêmes que précédemment.

4.1.2 Critères de qualité des données

On conserve dans l'analyse uniquement les données jugées de bonne qualité. Ce qui veut dire que l'on retire les événements identifiés comme étant du bruit ou issus d'une mauvaise période de prise de données ou issus de mauvais blocs de luminosité *LBNs* (voir section 2.5).

4.2 Bruits de fond

Les bruits de fond à prendre en compte comportent deux sources principales : instrumental et physique.

4.2.1 Bruit de fond instrumental

Le bruit de fond instrumental se compose principalement d'événements mutijets produits par interaction forte (typiquement deux jets dos-à-dos) et nous l'appellerons par la suite bruit

de fond QCD. Une mauvaise mesure de l'énergie des jets entraîne la création d'énergie transverse manquante et nous ramène donc à une topologie identique à notre canal d'étude. De plus, la grande section efficace de production du processus QCD (de l'ordre du mb) fait de ce bruit de fond une proportion non-négligeable des événements au niveau de la présélection. Nous estimerons ce bruit de fond à partir des données à cause de la grande difficulté de le modéliser précisément. La section 4.5.2 décrit la modélisation de ce bruit de fond.

4.2.2 Bruit de fond physique

Le bruit de fond physique regroupe tous les types de processus du Modèle Standard conduisant au même état final que notre canal d'étude. Nous l'appellerons par la suite bruit de fond électrofaible. On prend en compte les productions des bosons W ou Z associés à des jets, des dibosons (WW , ZZ et WZ) ou du quark top (célibataire ou par paire). Nous pouvons ici utiliser la simulation qui a été calibrée par la collaboration pour reproduire ces différents bruits de fond.

Simulation du bruit de fond physique

Les bruits de fond W/Z +jets et $t\bar{t}$ sont simulés avec le générateur ALPGEN [48]. Pour simuler les radiations dans les états initiaux et finaux et aussi pour simuler l'hadronisation des partons, ALPGEN est interfacé avec PYTHIA [49]. Ces bruits de fond ont été produits pour différentes multiplicités de partons légers (noté lp). Les processus correspondant à la plus grande multiplicité de jets ont été produits de manière inclusive (3 ou 5 partons légers dépendant du processus).

Le générateur COMPHEP [50] a été utilisé pour la production de quark top célibataire et PYTHIA pour les processus dibosons.

ALPGEN fournit des sections efficaces calculées à l'ordre dominant de la théorie des perturbations (*Leading Order*, noté LO). Comme les corrections d'ordre supérieur (*Next-to-Leading Order*, noté NLO ou *Next-to-Next-to-Leading Order*, noté NNLO) sont importantes, il convient d'appliquer des facteurs de corrections sur ces sections efficaces LO. On appelle ces facteurs de correction, les K-facteurs définis de la manière suivante :

$$K - \text{facteur} = \frac{\sigma_{\text{ordre supérieur}}}{\sigma_{\text{LO}}} \quad (4.1)$$

Le détail des K-facteurs utilisés pour les différents processus se retrouve dans le tableau 4.2. Le générateur MCFM [65] permet d'obtenir les sections efficaces de processus avec des saveurs lourdes (ex : $W + b\bar{b}$). Les sections efficaces des processus top et dibosons sont données par la référence [66]. Elles ont été calculées au niveau NNLO. Les tableaux 4.3, 4.4 et 4.5 résument les sections efficaces (LO) et le nombre d'événements générés associés à chaque processus des bruits de fond "électrofaible" considérés.

Afin de pouvoir traiter les événements simulés de la même manière que les données, on leur applique une simulation du détecteur DØ basée sur GEANT3 [51].

Bruit de fond	K-facteurs	Bruit de fond	K-facteurs
W + saveurs légères	1.3	Z + saveurs légères	1.3
W + c	1.3 × 1.47	Z + c	1.3 × 1.67
W + b	1.3 × 1.47	Z + b	1.3 × 1.52
t \bar{t}	1.43	WW	1.0
t célibataire	0.99	WZ	1.06
		ZZ	1.03

TAB. 4.2: Les processus de bruit de fond simulés et leur K-facteur correspondant.

Les corrections apportées à la simulation

JSSR [67] [68] (*Jets Smearing, Shifting and Removing*) corrige la résolution en énergie des jets de la simulation qui est meilleure que dans les données pour ajuster leur énergie. On évalue cette correction avec des événements $\gamma + \text{jet}$ ou $Z + \text{jet}$ dans les données et la simulation. On demande au boson et au jet d'être dos-à-dos et on définit la différence entre leur p_T respectifs par la variable ΔS :

$$\Delta S = \frac{p_T^{\text{jet}} - p_T^{\gamma/Z}}{p_T^{\gamma/Z}} \quad (4.2)$$

On paramétrise la différence de ΔS obtenue dans les données et la simulation en fonction du p_T du boson. L'ajustement de ΔS nous donne la correction à appliquer aux jets ayant un $p_T > 15$ GeV. La figure 4.2 montre l'effet de cette correction.

Repondération du profil de luminosité instantanée [69] : cette correction prend en compte la différence de profil de luminosité instantanée des événements de simulation et de données.

Repondération du p_T du boson Z [70] : cette correction a pour effet de palier à la mauvaise modélisation du p_T du boson Z des générateurs ALPGEN+PYTHIA. La valeur de la correction prend en compte les différences entre données et simulation au niveau du générateur. La figure 4.3 montre la correction obtenue. D'autre part, une repondération du p_T du boson W est aussi effectuée.

Confirmation des jets par les vertex [71] : l'efficacité de la confirmation des jets par les vertex est différente entre les données et la simulation. On évalue cette correction avec des événements $\gamma + \text{jet}$ ou $Z + \text{jet}$ dans les données et la simulation et sont définis en fonction de la position suivant z du vertex primaire, de η et p_T du jet.

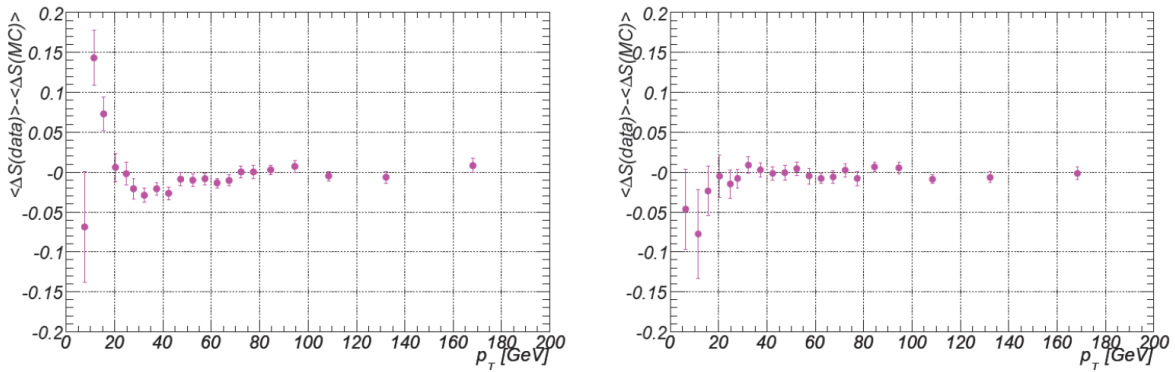


FIG. 4.2: Différence entre les valeurs de ΔS obtenues dans les données et la simulation avant (à gauche) et après (à droite) correction JSSR dans le calorimètre central CC [68].

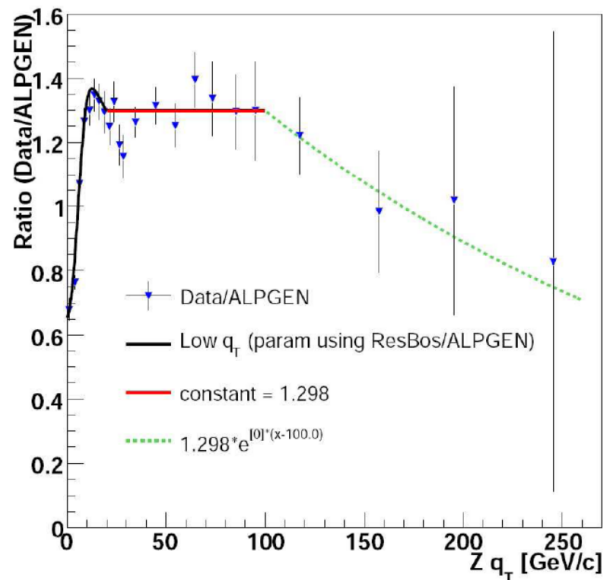


FIG. 4.3: Rapport données/simulation du p_T du boson Z [70].

Paramétrisation des conditions de déclenchement

(voir note d'analyse [74]) La paramétrisation des conditions de déclenchement est nécessaire car la simulation du système de déclenchement de $D\phi$ (*trigsim*) ne reproduit pas correctement les résultats obtenus dans les données. On la définit en utilisant des événements de type $Z \rightarrow \mu^+ \mu^- + \text{jets}$ dans les données et la simulation. On détermine l'efficacité de déclenchement d'évènements de type jets + \cancel{E}_T pour les trois niveaux de déclenchement distinctement (L1, L2 et L3). Finalement la paramétrisation obtenue est validée sur un lot d'évènements de type $W \rightarrow l\nu$.

Mode	σ (pb)	# d'événements
$W(\rightarrow l\nu)+0lp$ (excl.)	4510	46404567
$W(\rightarrow l\nu)+1lp$ (excl.)	1280	19898756
$W(\rightarrow l\nu)+2lp$ (excl.)	304	18087996
$W(\rightarrow l\nu)+3lp$ (excl.)	72.6	3754272
$W(\rightarrow l\nu)+4lp$ (excl.)	16.8	2602738
$W(\rightarrow l\nu)+5lp$ (incl.)	5.15	2044335
$W(\rightarrow l\nu)+2c+0lp$ (excl.)	24.5	934253
$W(\rightarrow l\nu)+2c+1lp$ (excl.)	13.5	738709
$W(\rightarrow l\nu)+2c+2lp$ (excl.)	5.5	554236
$W(\rightarrow l\nu)+2c+3lp$ (incl.)	2.53	469900
$W(\rightarrow l\nu)+2b+0lp$ (excl.)	9.37	1104413
$W(\rightarrow l\nu)+2b+1lp$ (excl.)	4.3	782487
$W(\rightarrow l\nu)+2b+2lp$ (excl.)	1.57	523717
$W(\rightarrow l\nu)+2b+3lp$ (incl.)	0.724	412747
WZ (\rightarrow incl.)	3.25	632296
WW (\rightarrow incl.)	11.6	709879
ZZ (\rightarrow incl.)	1.33	540273
$t+b \rightarrow e\nu b+b$	0.112	247517
$t+b \rightarrow \mu\nu b+b$	0.11	225286
$t+b \rightarrow \tau\nu b+b$	0.117	248722
$t+q+b \rightarrow e\nu b+qb$	0.243	272573
$t+q+b \rightarrow \mu\nu b+qb$	0.239	273354
$t+q+b \rightarrow \tau\nu b+qb$	0.254	246552
$t+t+0lp \rightarrow 2b+4lp$ (excl.)	1.42	793267
$t+t+1lp \rightarrow 2b+5lp$ (excl.)	0.576	456317
$t+t+2lp \rightarrow 2b+6lp$ (incl.)	0.281	277912
$t+t+0lp \rightarrow l\nu+2b+2lp$ (excl.)	1.4	777068
$t+t+1lp \rightarrow l\nu+2b+3lp$ (excl.)	0.577	457782
$t+t+2lp \rightarrow l\nu+2b+4lp$ (incl.)	0.267	321166
$t+t+0lp \rightarrow 2l+2\nu+2b+0lp$ (excl.)	0.352	749642
$t+t+1lp \rightarrow 2l+2\nu+2b+1lp$ (excl.)	0.142	452177
$t+t+2lp \rightarrow 2l+2\nu+2b+2lp$ (excl.)	0.267	281453

TAB. 4.3: Sections efficaces (σ en pb) et nombres d'événements des bruits de fond "électro-faibles" simulés (boson W , dibosons et quark top). Les sections efficaces (LO) des processus W ne tiennent pas compte du K -facteur. Les processus libellés "incl." ("excl.") ont été produits de manière inclusive (exclusive) en terme de partons.

Mode	σ (pb)	# d'événements
Z+0lp \rightarrow $\mu\mu$ +0lp (15 < m_Z < 75) (excl.)	344	1726264
Z+1lp \rightarrow $\mu\mu$ +1lp (15 < m_Z < 75) (excl.)	40.1	570408
Z+2lp \rightarrow $\mu\mu$ +2lp (15 < m_Z < 75) (excl.)	9.87	275077
Z+3lp \rightarrow $\mu\mu$ +3lp (15 < m_Z < 75) (incl.)	2.84	267966
Z+0lp \rightarrow $\mu\mu$ +0lp (75 < m_Z < 130) (excl.)	134	1514055
Z+1lp \rightarrow $\mu\mu$ +1lp (75 < m_Z < 130) (excl.)	41.4	604493
Z+2lp \rightarrow $\mu\mu$ +2lp (75 < m_Z < 130) (excl.)	9.91	400558
Z+3lp \rightarrow $\mu\mu$ +3lp (75 < m_Z < 130) (incl.)	3.25	146250
Z+0lp \rightarrow $\mu\mu$ +0lp (130 < m_Z < 250) (excl.)	0.887	351275
Z+1lp \rightarrow $\mu\mu$ +1lp (130 < m_Z < 250) (excl.)	0.359	170242
Z+2lp \rightarrow $\mu\mu$ +2lp (130 < m_Z < 250) (excl.)	0.0984	160267
Z+3lp \rightarrow $\mu\mu$ +3lp (130 < m_Z < 250) (incl.)	0.0335	141929
Z+0lp \rightarrow $\tau\tau$ +0lp (15 < m_Z < 75) (excl.)	338	1532702
Z+1lp \rightarrow $\tau\tau$ +1lp (15 < m_Z < 75) (excl.)	39.9	528075
Z+2lp \rightarrow $\tau\tau$ +2lp (15 < m_Z < 75) (excl.)	10	279564
Z+3lp \rightarrow $\tau\tau$ +3lp (15 < m_Z < 75) (incl.)	2.77	278238
Z+0lp \rightarrow $\tau\tau$ +0lp (75 < m_Z < 130) (excl.)	131	1516656
Z+1lp \rightarrow $\tau\tau$ +1lp (75 < m_Z < 130) (excl.)	40.3	562760
Z+2lp \rightarrow $\tau\tau$ +2lp (75 < m_Z < 130) (excl.)	9.99	274324
Z+3lp \rightarrow $\tau\tau$ +3lp (75 < m_Z < 130) (incl.)	3.1	173657
Z+0lp \rightarrow $\tau\tau$ +0lp (130 < m_Z < 250) (excl.)	0.922	358601
Z+1lp \rightarrow $\tau\tau$ +1lp (130 < m_Z < 250) (excl.)	0.375	171211
Z+2lp \rightarrow $\tau\tau$ +2lp (130 < m_Z < 250) (excl.)	0.0972	162320
Z+3lp \rightarrow $\tau\tau$ +3lp (130 < m_Z < 250) (incl.)	0.0362	157799
Z+0lp \rightarrow $\nu\nu$ +0lp (75 < m_Z < 130) (excl.)	805	2308113
Z+1lp \rightarrow $\nu\nu$ +1lp (75 < m_Z < 130) (excl.)	246	2449272
Z+2lp \rightarrow $\nu\nu$ +2lp (75 < m_Z < 130) (excl.)	61	590022
Z+3lp \rightarrow $\nu\nu$ +3lp (75 < m_Z < 130) (excl.)	14.1	166856
Z+4lp \rightarrow $\nu\nu$ +4lp (75 < m_Z < 130) (excl.)	3.28	82479
Z+5lp \rightarrow $\nu\nu$ +5lp (75 < m_Z < 130) (incl.)	0.936	38872

TAB. 4.4: Sections efficaces (" σ en pb) et nombres d'évènements des bruits de fond "électro-faibles" simulés (Z+saveurs légères). Les sections efficaces (LO) ne tiennent pas compte du K-facteur. Les processus libellés "incl." ("excl.") ont été produits de manière inclusive (exclusive) en terme de partons et ont été générés pour différentes masses m_Z du boson Z (en GeV).

Mode	σ (pb)	# d'événements
Z+cc+0lp \rightarrow cc+ $\mu\mu$ +0lp (15 < m_Z < 75) (excl.)	4.14	180580
Z+cc+1lp \rightarrow cc+ $\mu\mu$ +1lp (15 < m_Z < 75) (excl.)	0.953	93093
Z+cc+2lp \rightarrow cc+ $\mu\mu$ +2lp (15 < m_Z < 75) (incl.)	0.343	95436
Z+cc+0lp \rightarrow cc+ $\mu\mu$ +0lp (75 < m_Z < 130) (excl.)	0.932	193928
Z+cc+1lp \rightarrow cc+ $\mu\mu$ +1lp (75 < m_Z < 130) (excl.)	0.548	92744
Z+cc+2lp \rightarrow cc+ $\mu\mu$ +2lp (75 < m_Z < 130) (incl.)	0.281	51277
Z+cc+0lp \rightarrow cc+ $\mu\mu$ +0lp (130 < m_Z < 250) (excl.)	0.00756	79493
Z+cc+1lp \rightarrow cc+ $\mu\mu$ +1lp (130 < m_Z < 250) (excl.)	0.00439	45857
Z+cc+2lp \rightarrow cc+ $\mu\mu$ +2lp (130 < m_Z < 250) (incl.)	0.00283	47946
Z+cc+0lp \rightarrow cc+ $\tau\tau$ +0lp (15 < m_Z < 75) (excl.)	4.11	180024
Z+cc+1lp \rightarrow cc+ $\tau\tau$ +1lp (15 < m_Z < 75) (excl.)	1.05	181402
Z+cc+2lp \rightarrow cc+ $\tau\tau$ +2lp (15 < m_Z < 75) (incl.)	0.382	179172
Z+cc+0lp \rightarrow cc+ $\tau\tau$ +0lp (75 < m_Z < 130) (excl.)	0.898	260243
Z+cc+1lp \rightarrow cc+ $\tau\tau$ +1lp (75 < m_Z < 130) (excl.)	0.488	100802
Z+cc+2lp \rightarrow cc+ $\tau\tau$ +2lp (75 < m_Z < 130) (incl.)	0.298	50711
Z+cc+0lp \rightarrow cc+ $\tau\tau$ +0lp (130 < m_Z < 250) (excl.)	0.00743	91957
Z+cc+1lp \rightarrow cc+ $\tau\tau$ +1lp (130 < m_Z < 250) (excl.)	0.00441	48814
Z+cc+2lp \rightarrow cc+ $\tau\tau$ +2lp (130 < m_Z < 250) (incl.)	0.0025	46629
Z+cc+0lp \rightarrow cc+ $\nu\nu$ +0lp (75 < m_Z < 130) (excl.)	5.61	376456
Z+cc+1lp \rightarrow cc+ $\nu\nu$ +1lp (75 < m_Z < 130) (excl.)	3.00	199012
Z+cc+2lp \rightarrow cc+ $\nu\nu$ +2lp (75 < m_Z < 130) (incl.)	1.64	96147
Z+bb+0lp \rightarrow bb+ $\mu\mu$ +0lp (15 < m_Z < 75) (excl.)	0.509	182875
Z+bb+1lp \rightarrow bb+ $\mu\mu$ +1lp (15 < m_Z < 75) (excl.)	0.199	85812
Z+bb+2lp \rightarrow bb+ $\mu\mu$ +2lp (15 < m_Z < 75) (incl.)	0.0784	80002
Z+bb+0lp \rightarrow bb+ $\mu\mu$ +0lp (75 < m_Z < 130) (excl.)	0.424	205628
Z+bb+1lp \rightarrow bb+ $\mu\mu$ +1lp (75 < m_Z < 130) (excl.)	0.195	96232
Z+bb+2lp \rightarrow bb+ $\mu\mu$ +2lp (75 < m_Z < 130) (incl.)	0.099	44893
Z+bb+0lp \rightarrow bb+ $\mu\mu$ +0lp (130 < m_Z < 250) (excl.)	0.0034	88816
Z+bb+1lp \rightarrow bb+ $\mu\mu$ +1lp (130 < m_Z < 250) (excl.)	0.00184	44058
Z+bb+2lp \rightarrow bb+ $\mu\mu$ +2lp (130 < m_Z < 250) (incl.)	0.000884	41268
Z+bb+0lp \rightarrow bb+ $\tau\tau$ +0lp (15 < m_Z < 75) (excl.)	0.51	182703
Z+bb+1lp \rightarrow bb+ $\tau\tau$ +1lp (15 < m_Z < 75) (excl.)	0.189	89263
Z+bb+2lp \rightarrow bb+ $\tau\tau$ +2lp (15 < m_Z < 75) (incl.)	0.0801	80755
Z+bb+0lp \rightarrow bb+ $\tau\tau$ +0lp (75 < m_Z < 130) (excl.)	0.424	192733
Z+bb+1lp \rightarrow bb+ $\tau\tau$ +1lp (75 < m_Z < 130) (excl.)	0.197	98185
Z+bb+2lp \rightarrow bb+ $\tau\tau$ +2lp (75 < m_Z < 130) (incl.)	0.104	43850
Z+bb+0lp \rightarrow bb+ $\tau\tau$ +0lp (130 < m_Z < 250) (excl.)	0.0034	87909
Z+bb+1lp \rightarrow bb+ $\tau\tau$ +1lp (130 < m_Z < 250) (excl.)	0.00171	44571
Z+bb+2lp \rightarrow bb+ $\tau\tau$ +2lp (130 < m_Z < 250) (incl.)	0.00103	41018
Z+bb+0lp \rightarrow bb+ $\nu\nu$ +0lp (75 < m_Z < 130) (excl.)	2.52	367482
Z+bb+1lp \rightarrow bb+ $\nu\nu$ +1lp (75 < m_Z < 130) (excl.)	1.14	171942
Z+bb+2lp \rightarrow bb+ $\nu\nu$ +2lp (75 < m_Z < 130) (incl.)	0.617	82829

TAB. 4.5: Sections efficaces (σ en pb) et nombres d'événements des bruits de fond "électro-faibles" simulés (Z+saveurs lourdes). Les sections efficaces (LO) ne tiennent pas compte du K-facteur. Les processus libellés "incl." ("excl.") ont été produits de manière inclusive (exclusive) en terme de partons et ont été générés pour différentes masses m_Z du boson Z (en GeV).

4.3 Simulation du signal

On utilise PYTHIA pour générer les signaux de Higgs (HZ et HW). Les sections efficaces LO sont corrigées au niveau NNLO [72]. On considère dans l'analyse des masses de boson de Higgs entre 100 et 150 GeV par pas de 5 GeV. Le tableau 4.6 résume les sections efficaces (LO) et les nombres d'évènements générés associés aux différentes masses de signaux utilisés.

Mode	σ (pb)	# d'évènements
HZ(\rightarrow bb $\nu\nu$) ($m_H = 100$ GeV)	0.02709	246338
WH(\rightarrow bbl ν) ($m_H = 100$ GeV)	0.07528	320322
HZ(\rightarrow bb $\nu\nu$) ($m_H = 105$ GeV)	0.02278	235674
WH(\rightarrow bbl ν) ($m_H = 105$ GeV)	0.06271	293813
HZ(\rightarrow bb $\nu\nu$) ($m_H = 110$ GeV)	0.01899	320404
WH(\rightarrow bbl ν) ($m_H = 110$ GeV)	0.05183	316140
HZ(\rightarrow bb $\nu\nu$) ($m_H = 115$ GeV)	0.01563	235102
WH(\rightarrow bbl ν) ($m_H = 115$ GeV)	0.04225	279087
HZ(\rightarrow bb $\nu\nu$) ($m_H = 120$ GeV)	0.01256	328845
WH(\rightarrow bbl ν) ($m_H = 120$ GeV)	0.03368	321634
HZ(\rightarrow bb $\nu\nu$) ($m_H = 125$ GeV)	0.009830	133884
WH(\rightarrow bbl ν) ($m_H = 125$ GeV)	0.02611	278550
HZ(\rightarrow bb $\nu\nu$) ($m_H = 130$ GeV)	0.007420	318507
WH(\rightarrow bbl ν) ($m_H = 130$ GeV)	0.01956	553407
HZ(\rightarrow bb $\nu\nu$) ($m_H = 135$ GeV)	0.005369	133537
WH(\rightarrow bbl ν) ($m_H = 135$ GeV)	0.01404	470211
HZ(\rightarrow bb $\nu\nu$) ($m_H = 140$ GeV)	0.003720	319374
WH(\rightarrow bbl ν) ($m_H = 140$ GeV)	0.009639	495718
HZ(\rightarrow bb $\nu\nu$) ($m_H = 145$ GeV)	0.002429	133832
WH(\rightarrow bbl ν) ($m_H = 145$ GeV)	0.006259	446339
HZ(\rightarrow bb $\nu\nu$) ($m_H = 150$ GeV)	0.001479	331022
WH(\rightarrow bbl ν) ($m_H = 150$ GeV)	0.003769	320201

TAB. 4.6: Sections efficaces LO (σ en pb) et nombres d'évènements des signaux de Higgs simulés.

4.4 Présélection

Afin de réduire davantage le nombre d'évènements dans les lots initiaux de données (après les conditions de déclenchement et de qualité des données vues précédemment), on applique un ensemble de coupures minimales. On requiert :

- $N_{\text{jets}} \leq 3$: 2 ou 3 jets avec $p_T > 15$ GeV et $|\eta_{\text{det}}| < 3.2$;
- $N_{\text{jets confirmés}} \geq 2$: au moins 2 jets avec $p_T > 20$ GeV et $|\eta_{\text{det}}| < 2.5$, ils doivent être étiquetables (voir section 3.2) et confirmés par les vertex (au moins 2 traces du jet sont

- associées au vertex primaire). On classera ensuite les jets par p_T décroissant : le jet de plus haut p_T sera appelé jet principal (ou *Leading-jet*) ;
- Sélection des vertex : le vertex primaire doit être reconstruit dans la zone d'acceptance du SMT où il est le plus efficace ($|z_{PV}| < 40$ cm) et au moins 3 traces de particules chargées doivent provenir de ce vertex ;
 - $\cancel{E}_T \geq 20$ GeV (ici l'énergie transverse manquante est calculée uniquement à partir du calorimètre et elle n'est pas corrigée de la présence de muons : variable appelée METB-CorrCalo).

Le tableau 4.7 montre le nombre d'évènement subsistants après chaque coupure de présélection, ainsi que les efficacités absolues et relatives associées pour le lot de données et le signal ZH de référence ($m_H = 115$ GeV). Ce jeu de coupures nous permet de réduire très fortement le bruit de fond QCD dominant au départ (voir figure 4.4) tout en gardant une bonne efficacité sur notre signal. On définit ainsi un échantillon de données qui servira de base à nos différents lots de sélection.

Présélection	Données			ZH $\rightarrow \nu\bar{\nu}b\bar{b}$ MC (115 GeV)		
	Evts	Rel.(%)	Abs.(%)	Evts	Rel.(%)	Abs.(%)
Lot initial	84203433			235102		
Mauvais <i>Runs</i> & <i>LBNs</i>	72683613	86.32	86.32	234863	99.90	99.90
Qualité des évènements	66689396	91.75	79.20	227509	96.87	96.77
Triggers	19194892	28.78	22.80	-	-	-
$N_{jets} \leq 3$	16528585	86.11	19.63	193837	85.20	82.45
$N_{jets \text{ confirmés}} \geq 2$	11961224	72.37	14.21	159858	82.47	68.00
Simulation des Trigger	-	-	-	74890	46.85	31.85
Sélection des vertex	5906982	49.38	7.02	68609	91.61	29.18
$\cancel{E}_T \geq 20$ GeV	3304349	55.94	3.92	64407	93.88	27.40

TAB. 4.7: Liste des coupures de présélection avec les nombres d'évènements (Evts), les efficacités relatives (Rel.) et les efficacités absolues (Abs.) associées pour les données et pour un signal ZH ($m_H = 115$ GeV). Le nombre d'évènements de signal (simulé) correspond au nombre d'évènements générés.

4.5 Sélection

Après les coupures de sélection on obtient un échantillon qui sera appelé "échantillon de signal". Pour contrôler l'accord entre données et simulations dans notre lot de signal, nous allons définir deux autres lots : le premier sera enrichi en bruit de fond électrofaibles (échantillon dit de contrôle EW) et le deuxième sera enrichi en bruit de fond QCD (échantillon dit de contrôle QCD).

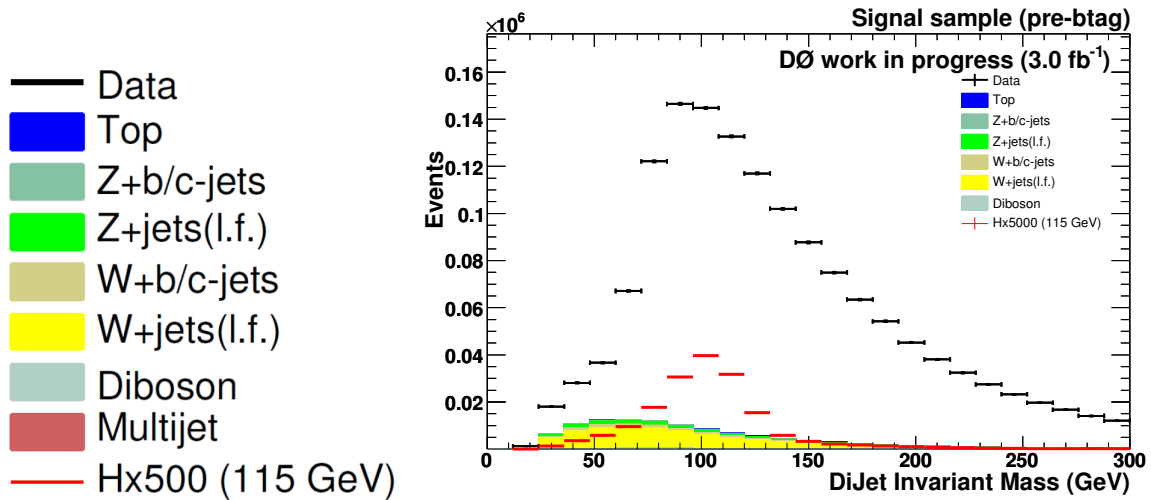


FIG. 4.4: La légende agrandie présente le jeu de couleur des histogrammes qui sera utilisé par la suite. Les points noirs représentent les données et les points rouges le signal de boson de Higgs simulé. Les histogrammes pleins représentent les bruits de fond simulés (électrofaibles) : le bleu foncé pour le top, le vert pâle pour le Z+saveurs lourdes (jets de b ou c), le vert vif pour le Z+saveurs légères (jets de u, d ou s), le jaune pâle pour le W+saveurs lourdes, le jaune vif pour le W+saveurs légères et le bleu pâle pour les dibosons (WW, WZ ou ZZ). La distribution représente la Masse Invariante des deux premiers jets de l'état final (racine carrée de la somme quadratique des vecteurs énergie-impulsion des jets considérés) après coupures de présélection (voir tableau 4.7). La contribution manquante (qui est à ce stade dominante) entre les données et les bruits de fond électrofaibles correspond au bruit de fond QCD. Après estimation sur les données, il sera représenté par un histogramme plein rouge.

4.5.1 Echantillon de controle Electrofaible

A partir de la présélection, on construit un lot enrichi en évènements W+jets. Pour ce faire, on demande que l'évènement contienne un muon isolé et ne contienne pas (i.e. "veto") d'électron isolé. En effet le muon ne déposant que très peu d'énergie dans le calorimètre, l'état final $W(\mu\nu) + \text{jets}$ correspondant (d'un point de vue calorimétrique) est similaire à notre canal d'étude $Z(\nu\nu) + \text{jets}$. D'autre part, ce lot de contrôle est orthogonal à notre lot de signal ZH dans lequel nous ferons un veto sur ces mêmes muons isolés.

Le muon utilisé pour le veto est de qualité *medium nseg3* (voir section 2.3.4) avec $p_T > 15$ GeV, $|\eta| < 2$ et ne doit pas être identifié comme étant un muon cosmique. La trace associée au muon doit être de qualité *medium*. Il faut aussi que le muon respecte les critères d'isolation suivants :

- CAL isolation < 0.08 ;
- track isolation < 0.06 ;
- $\Delta R(\mu, \text{jet}) > 0.5$.

L'électron utilisé pour le veto doit vérifier les conditions suivantes (voir section 2.3.2) :

- $p_T > 15$ GeV et $|\eta_{\text{det}}| < 1.1$ ou $1.5 < |\eta_{\text{det}}| < 2.5$;

- EM fraction > 0.9 ;
- HMatrix7 < 50 ;
- Likelihood > 0.85 ;

Ces définitions du muon et de l'électron ont été adoptées afin que le lot de signal soit orthogonal aux recherches de boson de Higgs dans le canal WH (qui sélectionne des leptons provenant du W). Cette condition nous permet ainsi d'éviter un double comptage des événements au moment de la combinaison des différents canaux de recherche du boson de Higgs.

Afin de réduire le bruit de fond QCD au maximum, on applique les coupures suivantes :

- \cancel{E}_T (corrigé des muons) > 20 GeV (\cancel{E}_T^{Mu}) ;
- la masse transverse du candidat W^1 > 30 GeV.

On demande aussi $\cancel{E}_T > 30$ GeV, afin de se placer dans les mêmes conditions que l'échantillon de contrôle QCD (qui est un lot préliminaire du lot de signal comme on le verra par la suite). Finalement on soustrait le vecteur impulsion transverse du muon à \cancel{E}_T dans les événements afin de se retrouver dans la topologie du canal d'étude.

La figure 4.5 montre quelques distributions dans ce lot de contrôle. Un bon accord en général entre les données et la simulation est constaté, non seulement sur la norme mais aussi sur la forme des distributions. Le rapport entre le nombre d'événements attendus (MC) et celui dans les données (DATA) est $\frac{DATA}{MC} = 0.97 \pm 0.01$. Ceci montre que les sections efficaces utilisées pour les bruits de fond électrofaibles correspondent à celles mesurées dans les données à 3% près. Ce facteur étant très proche de l'unité, nous le garderons égal à 1 pour le reste de l'analyse et la différence sera utilisée comme erreur systématique.

¹masse transverse du W = $\sqrt{2 \times p_T^\mu \times \cancel{E}_T^{Mu} (1 - \cos(\Delta\Phi(\cancel{E}_T^{Mu}, \mu)))}$

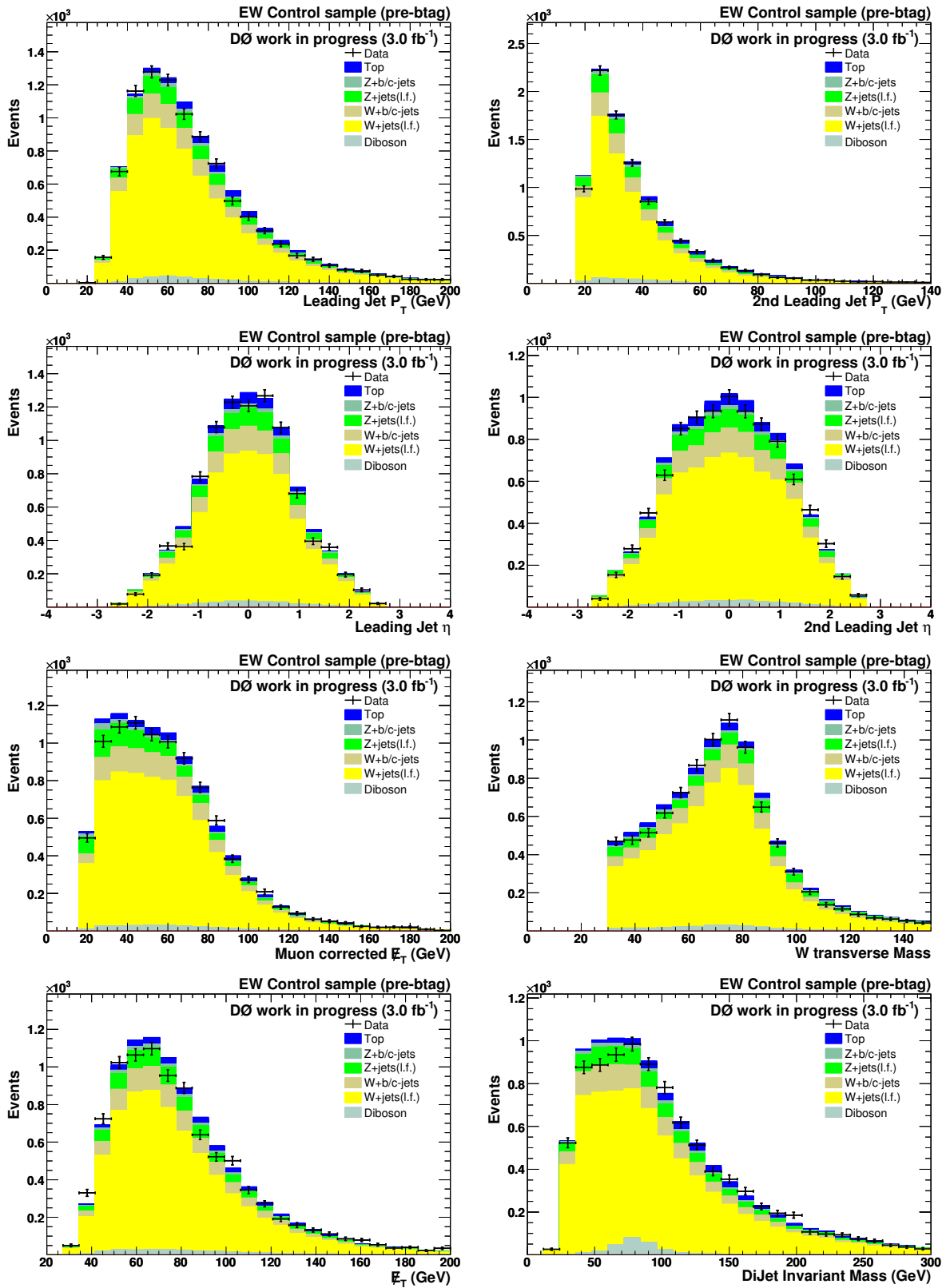


FIG. 4.5: Distributions dans l'échantillon de contrôle électrofaible : p_T du jet principal (en haut à gauche), p_T du jet secondaire (en haut à droite), η du jet principal (seconde ligne et à gauche), η du jet secondaire (seconde ligne et à droite), E_T corrigée de l'énergie des muons (troisième ligne à gauche), masse transverse du candidat W (troisième ligne à droite), E_T (en bas à gauche) et masse invariante des deux premiers jets (en bas à droite).

4.5.2 Estimation du bruit de fond QCD

A partir de la présélection, on construit un lot enrichi en évènements QCD à l'aide des coupures suivantes :

- véto sur les muons et électrons isolés avec les définitions utilisées dans la section 4.5.1 ;
- acoplanarité $< 165^\circ$;
- $E_T > 30$ GeV.

Le principe de modélisation du bruit de fond QCD repose sur le fait que dans le cas d'un évènement de type signal (à gauche dans la figure 4.6) E_T est alignée avec la somme de l'impulsion transverse des traces à l'intérieur des jets (on la note p_T^{trk}). On peut donc sélectionner des évènements majoritairement QCD, grâce à l'angle azimutal Φ entre E_T et p_T^{trk} . La figure 4.7 présente cette variable $\Delta\Phi(E_T, p_T^{\text{trk}})$. La coupure sur cette variable à $\pi/2$ permet de définir les évènements *sideband*. Au-delà de $\pi/2$ nous sommes dominés par les évènements QCD et la petite contribution des bruits de fond électrofaible est soustraite. En résumé, les évènements de données avec $\Delta\Phi(E_T, p_T^{\text{trk}}) > \pi/2$ seront utilisés pour modéliser la forme de notre bruit de fond QCD. La normalisation du bruit de fond QCD sera décrite ultérieurement.

Les critères suivants sur les traces ont été utilisés pour calculer p_T^{trk} :

- elles doivent provenir du vertex primaire (avec une distance minimale d'approche : $dca_r^1 < 2$ mm et $dca_z < 5$ mm) ;
- seules les traces ayant $p_T < 400$ GeV sont sélectionnées afin de rejeter les "fausses" traces qui sont caractérisées le plus souvent par une très grande impulsion transverse.

Nous cherchons à modéliser les évènements QCD pour lesquels E_T est alignée avec l'un des jets. La figure 4.8 montre la distribution de l'impulsion transverse du jet principal mesurée par le détecteur de traces (LJTrkPt) en fonction de celle mesurée par le calorimètre (LJetPt) pour des évènements de type signal ou QCD. On peut voir que notre échantillon QCD contient beaucoup d'évènements ayant des faibles valeurs de l'impulsion transverse mesurée par le détecteur de traces. En imposant la condition $3 \times \text{LJTrkPt} + 20 > \text{LJetPt}$, il nous est possible de supprimer ces évènements ayant des traces manquantes dans notre échantillon de contrôle QCD et ainsi de se rapprocher du type de bruit de fond QCD présent dans l'analyse. L'amélioration qu'apporte cette coupure sur la modélisation du bruit de fond QCD est visible dans la figure 4.9.

Les évènements de données *sideband* (avec $\Delta\Phi(E_T, p_T^{\text{trk}}) > \pi/2$) auxquels on a soustrait les bruits de fond électrofaibles et les évènements ayant des traces manquantes permettent de définir la forme du bruit de fond QCD dans toutes les distributions de l'analyse (i.e. les évènements $\Delta\Phi(E_T, p_T^{\text{trk}}) < \pi/2$). L'amplitude du bruit de fond QCD est obtenue en calculant le facteur de renormalisation suivant : $\frac{N_{\text{QCD}}^{<\pi/2}}{N_{\text{QCD}}^{>\pi/2}}$ avec

- $N_{\text{QCD}}^{<\pi/2}$ est le nombre d'évènements de données moins ceux de bruits de fonds électrofaibles dans la zone de signal : $\Delta\Phi(E_T, p_T^{\text{trk}}) < \pi/2$;

¹ $r = \sqrt{x^2 + y^2}$

- $N_{\text{QCD}}^{>\pi/2}$ est le bruit de fond QCD que nous avons simulé dans la zone *sideband* : $\Delta\Phi(\cancel{E}_T, \cancel{p}_T^{\text{trk}}) > \pi/2$.

On trouve ce facteur égal à 1.30 ± 0.01 (figure 4.9 à droite). La figure 4.10 montre quelques distributions dans ce lot de contrôle enrichi en QCD. Ces figures indiquent une bonne modélisation de ce bruit de fond par la méthode décrite ci-dessus.

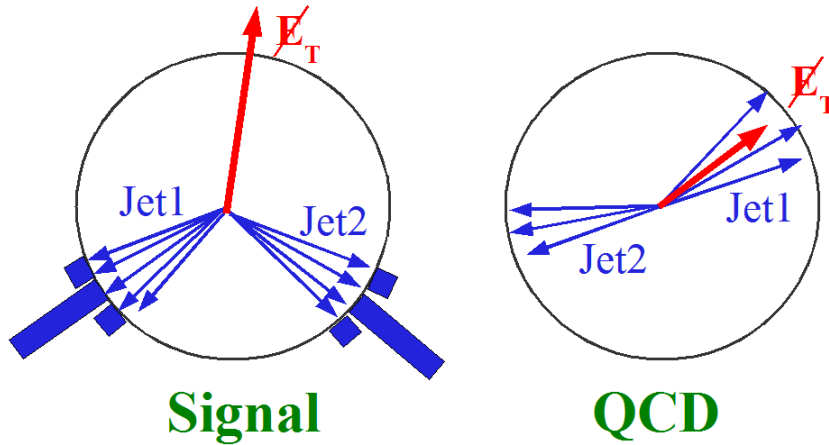


FIG. 4.6: Schématisation des jets et des traces des jets pour des événements de type signal (droite) ou QCD (gauche).

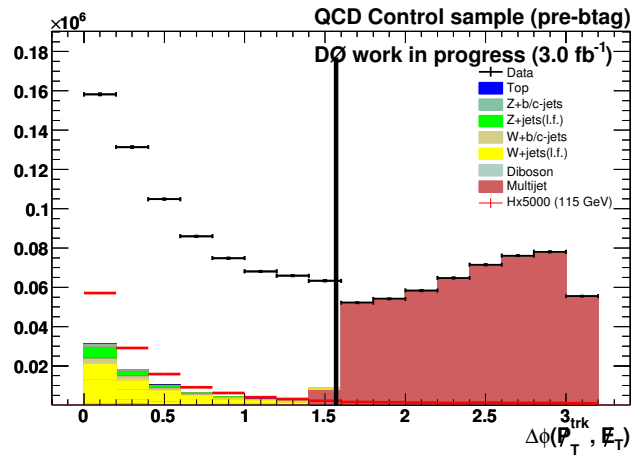


FIG. 4.7: Angle azimutal Φ entre \cancel{E}_T et $\cancel{p}_T^{\text{trk}}$ dans l'échantillon de contrôle QCD. La forme du bruit de fond QCD estimé (en rouge) est obtenu à partir des événements de données auquel on a soustrait les contributions des bruits de fonds électrofaibles dans la zone $\Delta\Phi(\cancel{E}_T, \cancel{p}_T^{\text{trk}}) > \pi/2$.

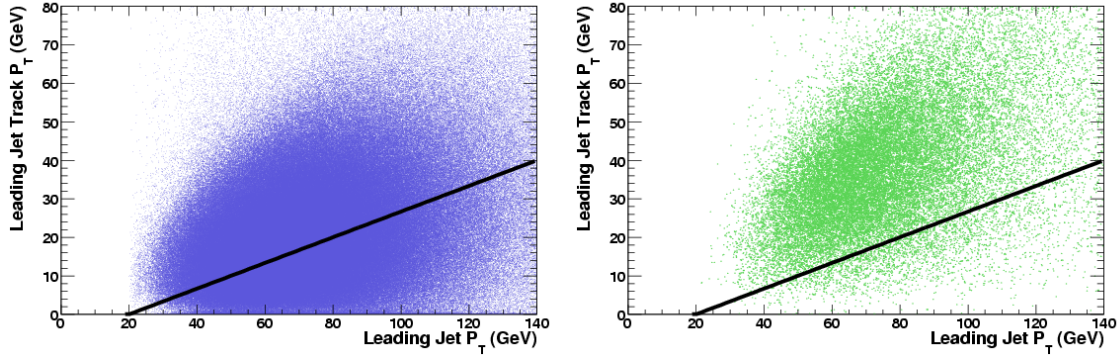


FIG. 4.8: Impulsion transverse du jet principal mesurée avec le détecteur de traces en fonction de celle mesurée avec le calorimètre. Cette distribution est représentée pour des événements "sideband" ($\Delta\Phi(\cancel{E}_T, p_T^{\text{trk}}) > \pi/2$) de type signal (droite) et de type QCD (gauche) dans l'échantillon de contrôle QCD. La ligne noire représente la coupure $3 * LJTrkPt + 20 > LJetPt$.

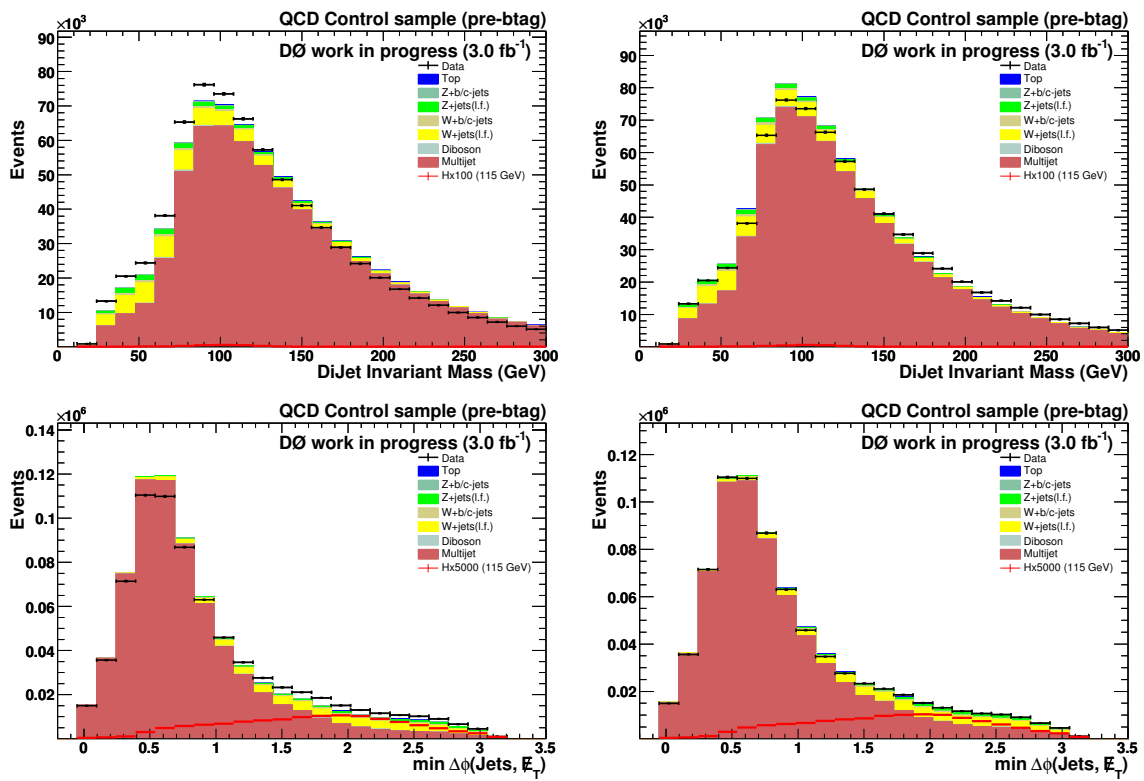


FIG. 4.9: Résultat de la modélisation du bruit de fond QCD dans l'échantillon de contrôle QCD. Cette modélisation a été obtenue sans (gauche) ou avec (droite) la coupure $3 * LJTrkPt + 20 > LJetPt$, avec un facteur de renormalisation de 0.77 et 1.30 respectivement. En haut : masse invariante des deux jets principaux, en bas : angle azimutal minimal entre le jet principal ou secondaire et \cancel{E}_T .

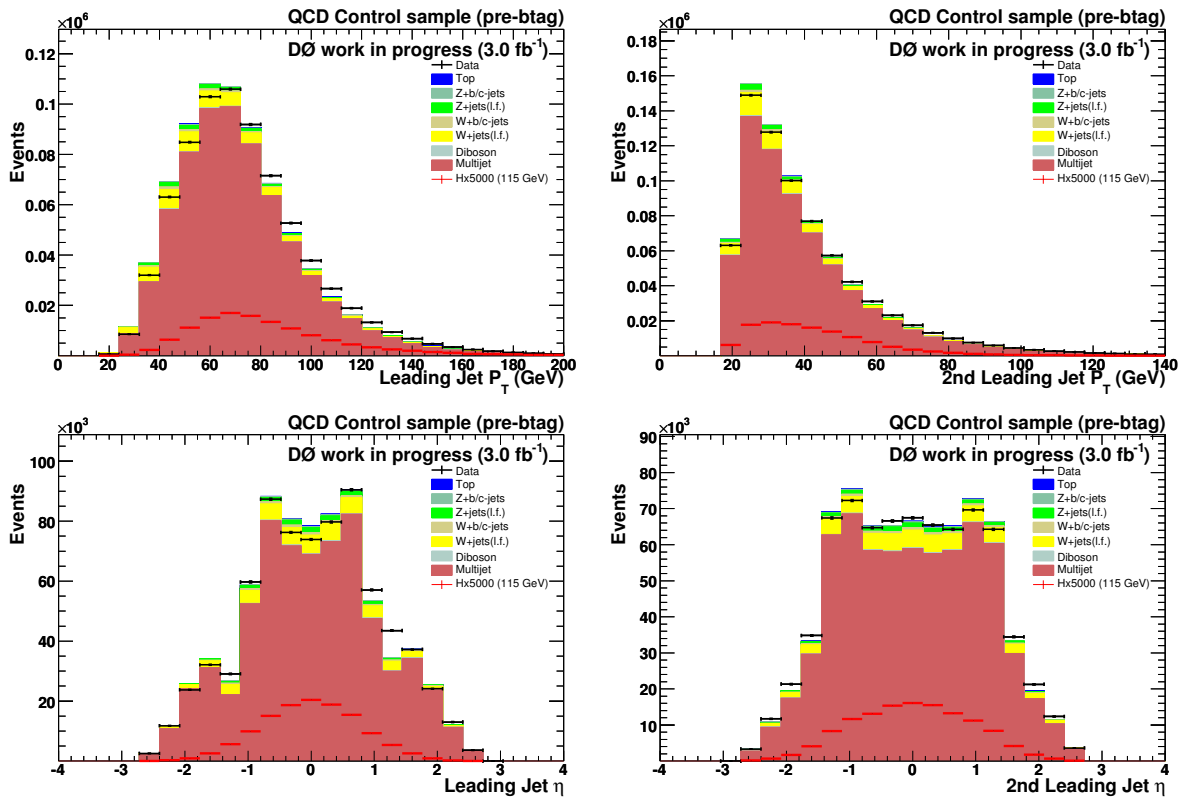


FIG. 4.10: Distributions dans l'échantillon de contrôle QCD : p_T du jet principal (en haut à gauche), p_T du jet secondaire (en haut à droite), η du jet principal (en bas à gauche), η du jet secondaire (en bas à droite).

4.5.3 Lot de signal

L'objectif est de définir un jeu de coupure visant à réduire les bruits de fonds principaux : QCD et W/Z +jets. Ces coupures ont été optimisées pour le signal ZH et une masse de boson de Higgs de 115 GeV. À partir de notre lot de présélection, nous définissons l'échantillon de signal à l'aide des critères suivants :

- véto sur les muons et électrons isolés avec les définitions utilisées dans la section 4.5.1 ;
- acoplanarité $< 165^\circ$ (figure 4.11 en haut à gauche) ;
- $E_T > 30$ GeV (figure 4.11 en haut à droite) et $\Delta\Phi(E_T, p_T^{\text{trk}}) < \pi/2$ (figure 4.7) ;
- E_T significance > 5 (figure 4.11 deuxième ligne à gauche). La significance de E_T [73] est une variable qui permet de vérifier si la E_T mesurée est conforme avec la fluctuation en résolution des objets mesurés dans la direction de la E_T observée. Cette résolution prend en compte plusieurs effets : la mauvaise identification du vertex primaire, la résolution en énergie des objets (jets, électrons et muons). Cette variable est comprise entre 0 et 10 par définition. Plus la significance en E_T est grande et plus nous sommes affranchis des effets de mauvaise mesure de nos objets servant à définir la E_T ;
- $E_T > 40$ GeV (figure 4.11 deuxième ligne à droite) ;
- véto sur les mauvais jets (voir section 2.3.3)
- nous avons vu en 4.5.2 que pour un événement de type QCD, nous attendons que la E_T soit alignée avec l'un des jets principaux. En regardant la distribution de E_T en fonction de l'angle azimutal minimum entre le jet principal ou secondaire et E_T , pour les événements de type signal et QCD, on peut définir la coupure triangulaire : $E_T > 80 - 40 \times \min\Delta\Phi(E_T, \text{jets})$ (figure 4.11 troisième ligne : événements QCD à gauche, signal à droite) ;
- $-0.1 < \mathcal{A} < 0.2$ avec l'assymétrie définie comme $\mathcal{A} = \frac{E_T - M_T}{E_T + M_T}$ (figure 4.11 en bas).

Le véto sur les électrons et muons isolés sert à réduire les bruits de fond W/Z +jets. Les autres coupures servent à réduire le bruit de fond QCD. La figure 4.11 montre les distributions des variables utilisées lors de cette sélection et la figure 4.12 montre l'évolution de la masse invariante des deux jets principaux lors des différentes coupures.

À ce stade, le bruit de fond QCD a été fortement réduit et la normalisation de celui-ci est estimée en inversant la coupure $\Delta\Phi(E_T, p_T^{\text{trk}})$ en utilisons la méthode décrite en 4.5.2. Le facteur de normalisation obtenu est 1.01 ± 0.3 . La figure 4.13 montre quelques distributions représentatives dans ce lot de signal. Le détail du nombre d'événements de données et de signal après chaque coupure se trouve dans le tableau 4.8 (sélection standard correspond aux deux premières coupures dans la liste ci-dessus).

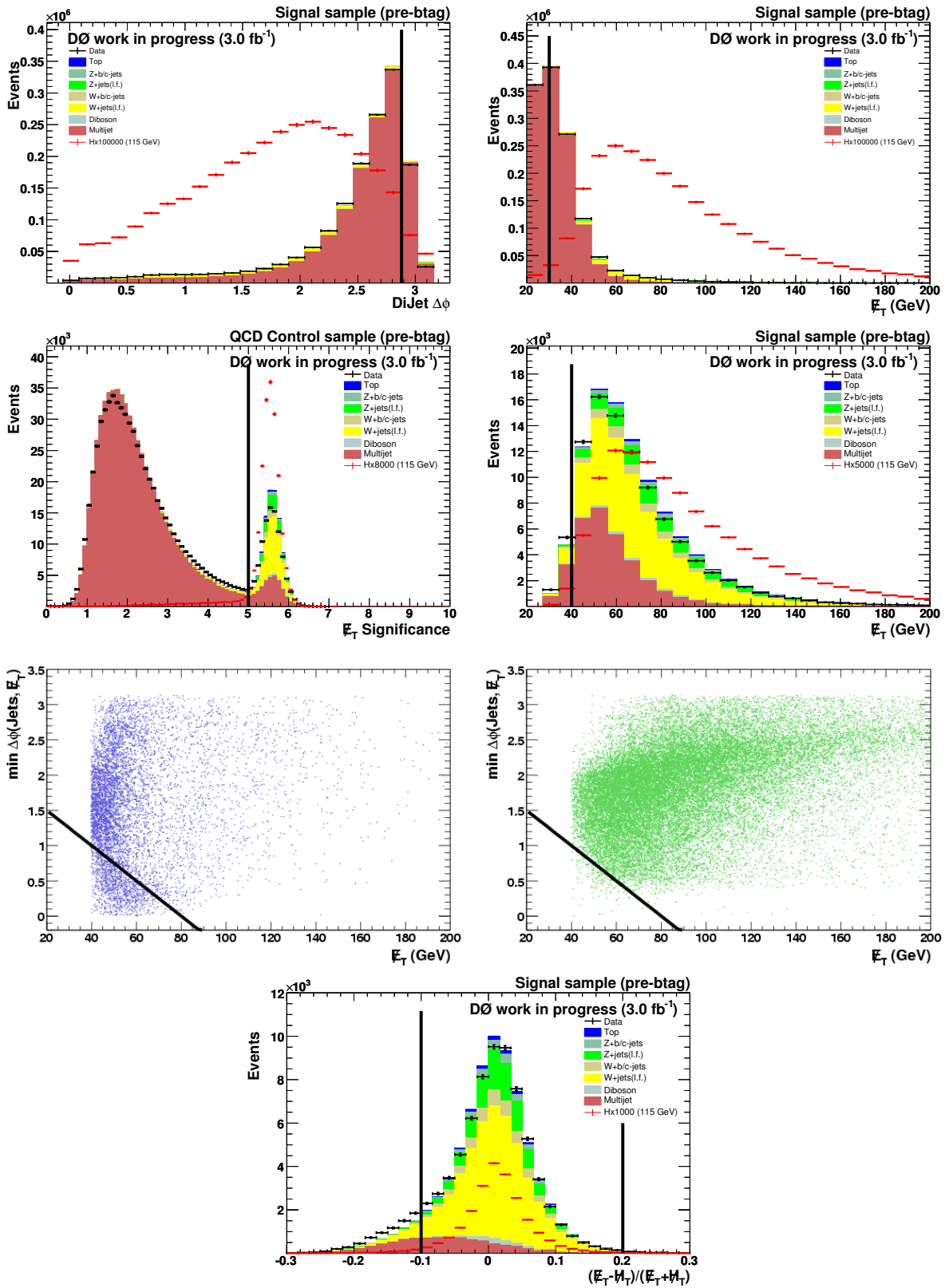


FIG. 4.11: Distributions des variables utilisées lors de la sélection. La ligne noire symbolise la coupure qui sera effectuée de manière séquentielle. De haut en bas et de gauche à droite nous trouvons les distributions suivantes : acoplanarité, E_T (coupure à 30 GeV), E_T significance, E_T (coupure à 40 GeV), coupure triangulaire, assymétrie.

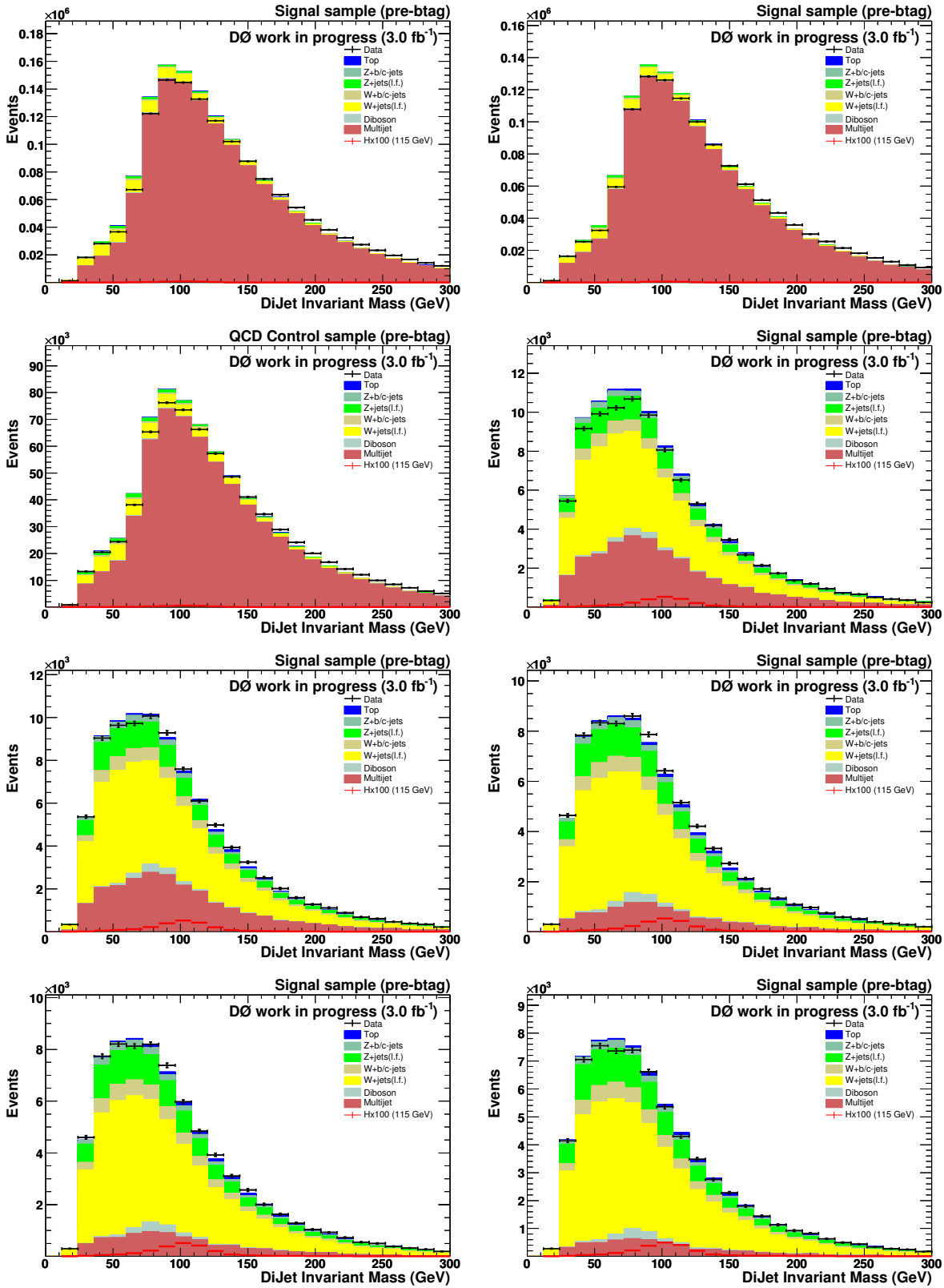


FIG. 4.12: Masse invariante des deux jets principaux lors des différentes coupures de sélection. De haut en bas et de gauche à droite, nous nous retrouvons à l'étape suivante : présélection (section 4.4), veto leptons isolés + $\mathcal{A}_{\text{coplanarité}} < 165^\circ$, $E_T > 30$ GeV + $\Delta\Phi(\vec{E}_T, \vec{p}_T^{\text{trk}}) < \pi/2$, E_T significativité > 5 , $E_T > 40$ GeV, veto sur les mauvais jets, coupure triangulaire, $-0.1 < \mathcal{A} < 0.2$.

Sélection	Données			ZH $\rightarrow \nu\bar{\nu}b\bar{b}$ MC (115 GeV)			S/ \sqrt{B} (10^{-3})
	Evts	Rel.(%)	Abs.(%)	Evts	Rel.(%)	Abs.(%)	
Présélection	3304349	100.00	100.00	11.95	100.00	100.00	6.57
Sélection standard	2789042	84.41	84.41	11.62	97.24	97.24	6.96
$E_T \geq 30$ GeV	745026	26.71	22.55	11.36	97.76	95.06	13.16
E_T significance ≥ 5	97532	13.09	2.95	10.87	95.69	90.96	34.81
$E_T \geq 40$ GeV	92719	95.07	2.81	10.82	99.54	90.54	35.53
Véto mauvais jets	79027	85.23	2.39	10.80	99.82	90.38	38.42
Coupeure Triangulaire	75783	94.99	2.29	10.71	99.17	89.62	38.90
$-0.1 < \mathcal{A} < 0.2$	68088	89.85	2.06	10.58	98.79	88.54	40.55

TAB. 4.8: Liste des coupures de sélection avec le nombre d'événements (Evts), l'efficacité absolue (Abs.), l'efficacité relative (Rel.) et la sensibilité (S/\sqrt{B}) associées pour les données et pour un signal ZH ($m_H = 115$ GeV).

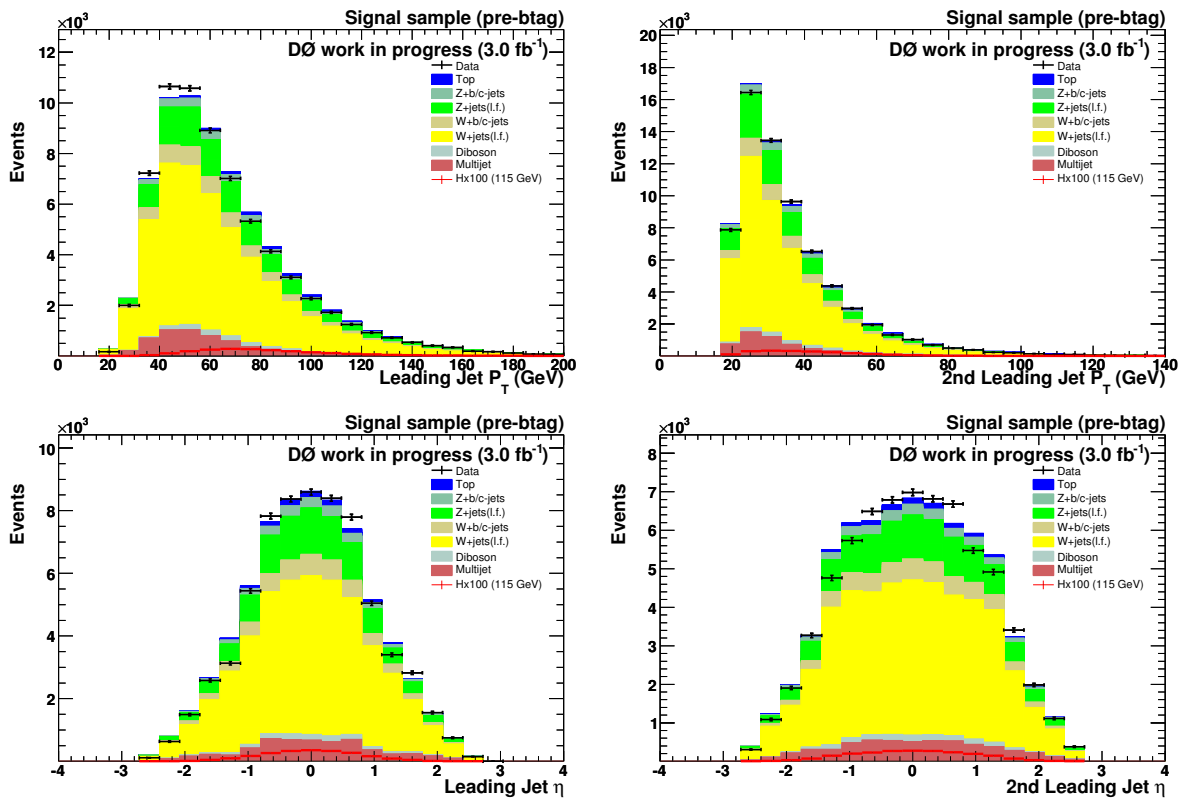


FIG. 4.13: Distributions dans l'échantillon de signal : p_T du jet principal (en haut à gauche), p_T du jet secondaire (en haut à droite), η du jet principal (en bas à gauche), η du jet secondaire (en bas à droite).

4.6 Etiquetage des jets issus de quark b avec NN

Avant de présenter les résultats obtenus avec l'algorithme de *b-tagging* que nous avons développés (SLTNN) [81], cette section reproduit l'analyse officielle qui est basée sur l'utilisation de la méthode d'étiquetage NN (voir section 3.1), qui identifie les jets issus de quark b à l'aide d'un réseau de neurones. En effet, après avoir réduit les contributions des bruits de fond QCD et W/Z +jets, on peut améliorer la sensibilité de l'analyse en demandant un ou deux jets issus de quark b.

4.6.1 Confirmation des jets par les vertex et notion d'étiquetabilité

Les méthodes d'étiquetages des jets issus de quark b requièrent deux conditions sur ces jets :

- ils doivent être confirmés par les vertex : au moins 2 traces du jet sont associées au vertex primaire ;
- et être étiquetable : $p_T > 15$ GeV et au moins deux traces dans le jet ayant produit au moins trois coups dans le CFT. Pour la partie centrale du calorimètre (CC) on demande aux traces d'avoir produit au moins sept coups dans le CFT.

Ces conditions ont été demandées au niveau de la présélection et permettent de rejeter une grande proportion d'évènements QCD comme le montre le tableau 4.7.

A ce stade il est nécessaire d'introduire un facteur de correction données/simulation car la confirmation des jets par les vertex est moins efficace dans les données que dans la simulation (voir la section 4.2.2). On définit ensuite l'étiquetabilité comme étant le rapport du nombre de jets étiquetables et le nombre de jets après confirmation par les vertex. Une fois de plus, l'efficacité de l'étiquetabilité est différente dans les données et la simulation. On effectue une paramétrisation de ce facteur de correction données/simulation à l'aide de l'échantillon de contrôle électrofaible (voir section 4.5.1). On utilise pour ce faire des variables bien reproduites par la simulation : η du jet et positionnement suivant z du vertex primaire (PVz). On peut voir sur la figure 4.14 l'étiquetabilité du jet en fonction de $\eta \times \text{sign}^1(\text{PVz})$ dans les données (distribution en haut à gauche) et dans la simulation (distribution en haut à droite). L'étiquetabilité a le même comportement suivant $\text{sign}(\text{PVz})$ (courbes rouge et noire), on peut sommer ces deux contributions afin d'augmenter la statistique. Le rapport données/simulation (distribution en bas à gauche) est ajusté par un polynôme d'ordre 5.

4.6.2 Résultats de la méthode d'étiquetage NN

A partir des 12 points de fonctionnement que possède l'algorithme d'étiquetage NN (voir section 3.1.3), nous définissons deux lots orthogonaux. Un lot dans lequel un des deux jets passe le critère d'étiquetage VT (*VeryTight*) et l'autre jet passe le critère L3 (*Loose3*) : ce lot sera

¹dans notre analyse PVz est défini entre ± 40 cm. La fonction sign renvoie le signe de PVz

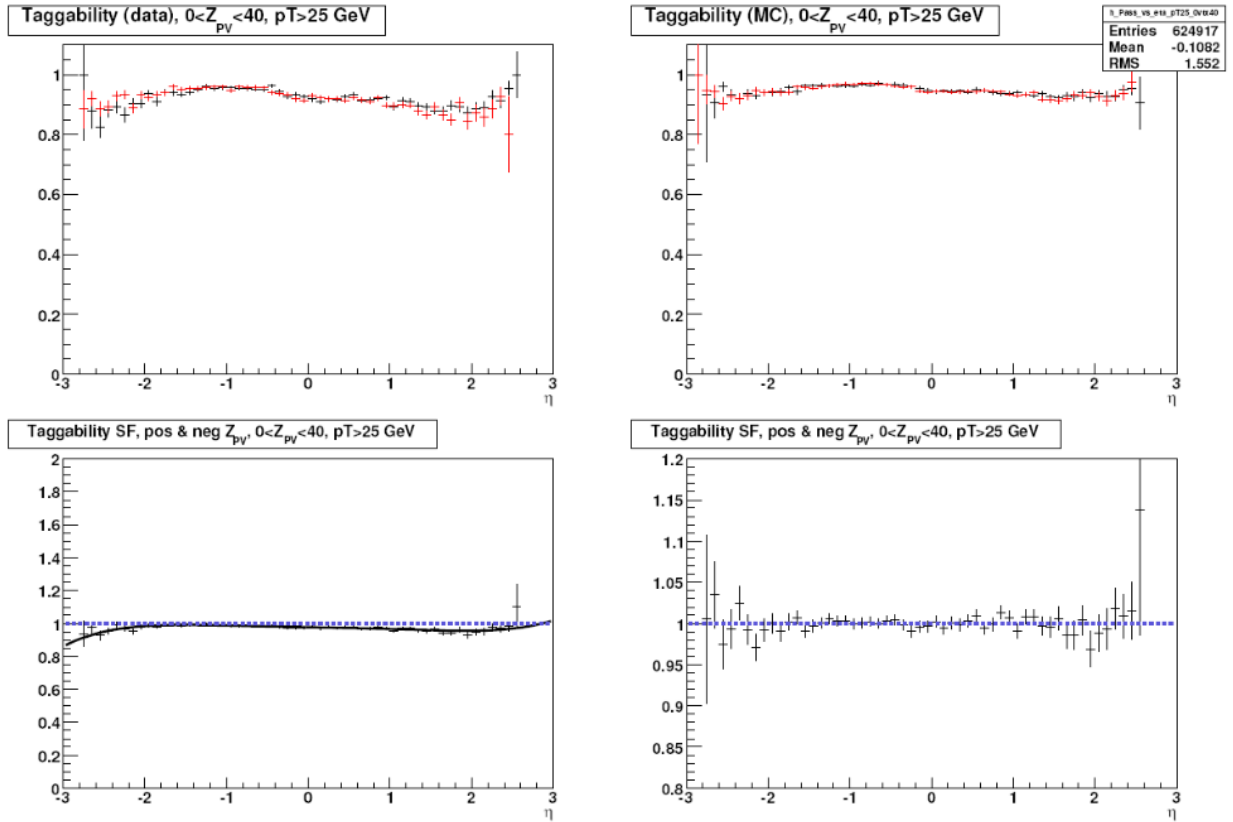


FIG. 4.14: Étiquetabilité du jet dans les données (en haut à gauche) et dans la simulation (en haut à droite) paramétrisée en fonction de $\eta^{\text{jet}} * \text{sign}(PVZ)$. La courbe noire représente les événements ayant $PVz > 0$ et la courbe rouge, les événements ayant $PVz < 0$. Rapport de l'étiquetabilité dans les données et la simulation (en bas à gauche) avec l'ajustement par un polynôme d'ordre 5. Rapport de l'étiquetabilité dans les données et la simulation (en bas à droite) en ayant appliqué le facteur de correction à la simulation. [74]

appelé lot d'étiquetage "asymétrique". Le deuxième lot requiert que l'un des deux jets passe le critère VT et que l'autre jet ne passe pas le critère L3 : ce lot sera appelé lot d'étiquetage *tight*. Nous avons vu dans le chapitre 3, que l'efficacité d'étiquetage est différente dans les données et la simulation. Nous allons utiliser les facteurs de correction d'étiquetage données/simulation, c'est-à-dire les *Scale Factors* (SF) fournis par l'algorithme NN pour chaque point de fonctionnement, pour définir des corrections à la simulation par événements :

$$SF_{\text{asymétrique}} = SF_{\text{jet étiqueté VT}} \times SF_{\text{jet étiqueté L3}} \quad (4.3)$$

$$SF_{\text{tight}} = SF_{\text{jet étiqueté VT}} \times SF_{\text{jet non-étiqueté L3}} \quad (4.4)$$

On définit $SF_{\text{jet non-étiqueté L3}}$, noté SF_{L3} . On peut relier le nombre d'événements étiquetés L3 (N_{L3}), le nombre d'événements non-étiquetés L3 ($N_{\neg L3}$) et le nombre total d'événements (N)

par :

$$N = N_{L3} + N_{!L3} \quad (4.5)$$

On connaît aussi la probabilité qu'un jet soit étiqueté L3 :

$$P_{L3} = \frac{N_{L3}}{N} \quad (4.6)$$

et cette probabilité nous est accessible grâce aux TRF fournies par l'algorithme NN. On peut donc écrire :

$$P_{!L3} = 1 - P_{L3} \quad (4.7)$$

$$\frac{N_{!L3}}{N} = 1 - \frac{N_{L3}}{N} \quad (4.8)$$

En sachant que :

$$SF_{L3} = \frac{p_{L3}^{\text{Données}}}{p_{L3}^{\text{Simulation}}} \quad (4.9)$$

on arrive finalement a :

$$SF_{!L3} = \frac{p_{!L3}^{\text{Données}}}{p_{!L3}^{\text{Simulation}}} \quad (4.10)$$

$$SF_{!L3} = \frac{1 - p_{L3}^{\text{Données}}}{1 - p_{L3}^{\text{Simulation}}} \quad (4.11)$$

$$SF_{!L3} = \frac{1 - p_{L3}^{\text{Simulation}} \times SF_{L3}}{1 - p_{L3}^{\text{Simulation}}} \quad (4.12)$$

$$SF_{!L3} = \frac{1 - TRF_{L3} \times SF_{L3}}{1 - TRF_{L3}} \quad (4.13)$$

Nous allons regarder les événements étiquetés *tight* ou asymétrique dans les échantillons de contrôle et de signal. L'algorithme d'étiquetage repose sur des quantités mesurées par le détecteur de traces. Malheureusement notre estimation du bruit de fond QCD reproduit mal certaines variables liées aux traces (voir figure 4.15 pour le p_T des traces). Il nous faut donc redéfinir le bruit de fond QCD après étiquetage : forme et facteur de renormalisation.

En supposant que l'efficacité d'étiquetage dans l'échantillon de contrôle QCD et dans l'échantillon de signal est la même, on peut relier les facteurs de renormalisation du bruit de fond QCD (F) avant et après étiquetage dans ces deux échantillons de la manière suivante :

$$F_{\text{après}}^{\text{signal}} = F_{\text{après}}^{\text{QCD}} \times \frac{F_{\text{avant}}^{\text{signal}}}{F_{\text{avant}}^{\text{QCD}}} \quad (4.14)$$

Le tableau 4.9 résume les valeurs des facteurs de renormalisation utilisés.

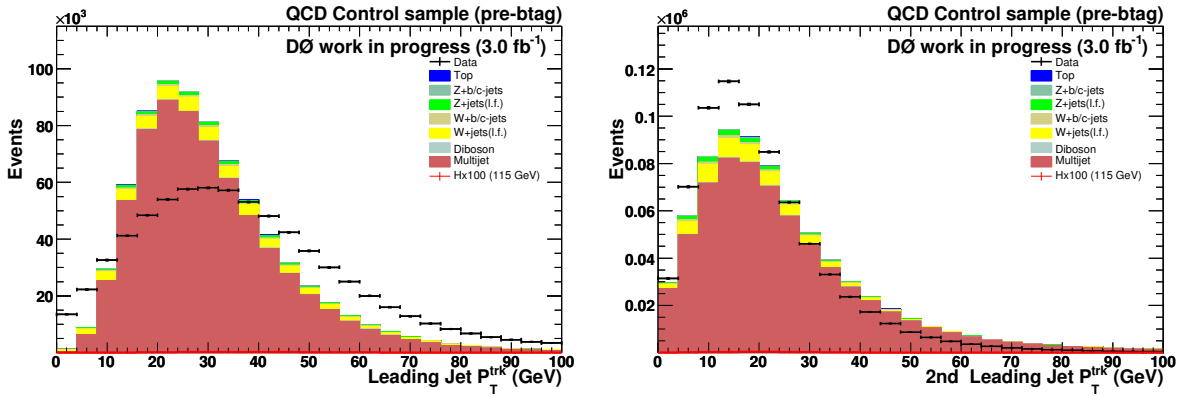


FIG. 4.15: p_T des traces du jet principal (gauche) et secondaire (droite) dans l'échantillon de contrôle QCD.

Niveau d'étiquetage	échantillon		
	contrôle électrofaible	contrôle QCD	signal
avant b -tagging	1.00	1.30 ± 0.00	1.01 ± 0.03
tight b -tagging	1.00	1.41 ± 0.00	1.09 ± 0.03
asymétrique b -tagging	1.00	1.44 ± 0.00	1.09 ± 0.04

TAB. 4.9: Résumé des facteurs de renormalisation (données/simulation) déterminés dans les échantillons de contrôle (électrofaible et QCD) et de signal, avant et après étiquetage avec NN (tight ou asymétrique).

La figure 4.16 montre les distributions de la masse invariante des deux jets principaux après étiquetage (*tight* ou asymétrique) dans les échantillons de contrôle (électrofaible ou QCD) et de signal après correction par les facteurs du tableau 4.9.

Le tableau 4.10 montre le détail du nombre d'évènements obtenus dans les données et la simulation. On peut voir un bon accord entre les simulations et les données ainsi que le gain en sensibilité (qui est conséquent dans le cas de l'étiquetage asymétrique).

Bruit de fond	Niveau d'étiquetage		
	avant	tight	asymétrique
Top	1220.0 ± 2.3	381.2 ± 0.7	209.1 ± 0.4
Di-boson	1461.8 ± 7.0	78.5 ± 0.3	25.2 ± 0.1
Zjj	11003.4 ± 57.7	134.5 ± 1.3	7.7 ± 0.1
Zbb/cc	2557.9 ± 15.5	373.4 ± 2.1	132.6 ± 1.5
Wjj	42023.5 ± 69.8	1463.2 ± 4.4	133.0 ± 0.4
Wbb/cc	5233.8 ± 20.0	705.8 ± 2.8	227.5 ± 1.1
Electrofaible	63500.5 ± 94.3	3136.7 ± 5.8	735.1 ± 1.9
QCD	4587.5 ± 14.3	390.5 ± 1.2	59.6 ± 0.2
Total	68088.0 ± 95.4	3527.2 ± 6.0	794.8 ± 1.9
Données	68088	3588	783
signal (115 GeV)	21.3 ± 0.1	6.0 ± 0.0	6.3 ± 0.0
S/√B	0.1	0.11	0.29

TAB. 4.10: Nombres d'évènements de bruit de fond, de signal (ZH et WH) et de données, avant et après étiquetage (tight ou asymétrique) dans l'échantillon de signal. La sensibilité (S/\sqrt{B}) a été déterminée dans une fenêtre de masse de ± 35 GeV autour de 115 GeV.

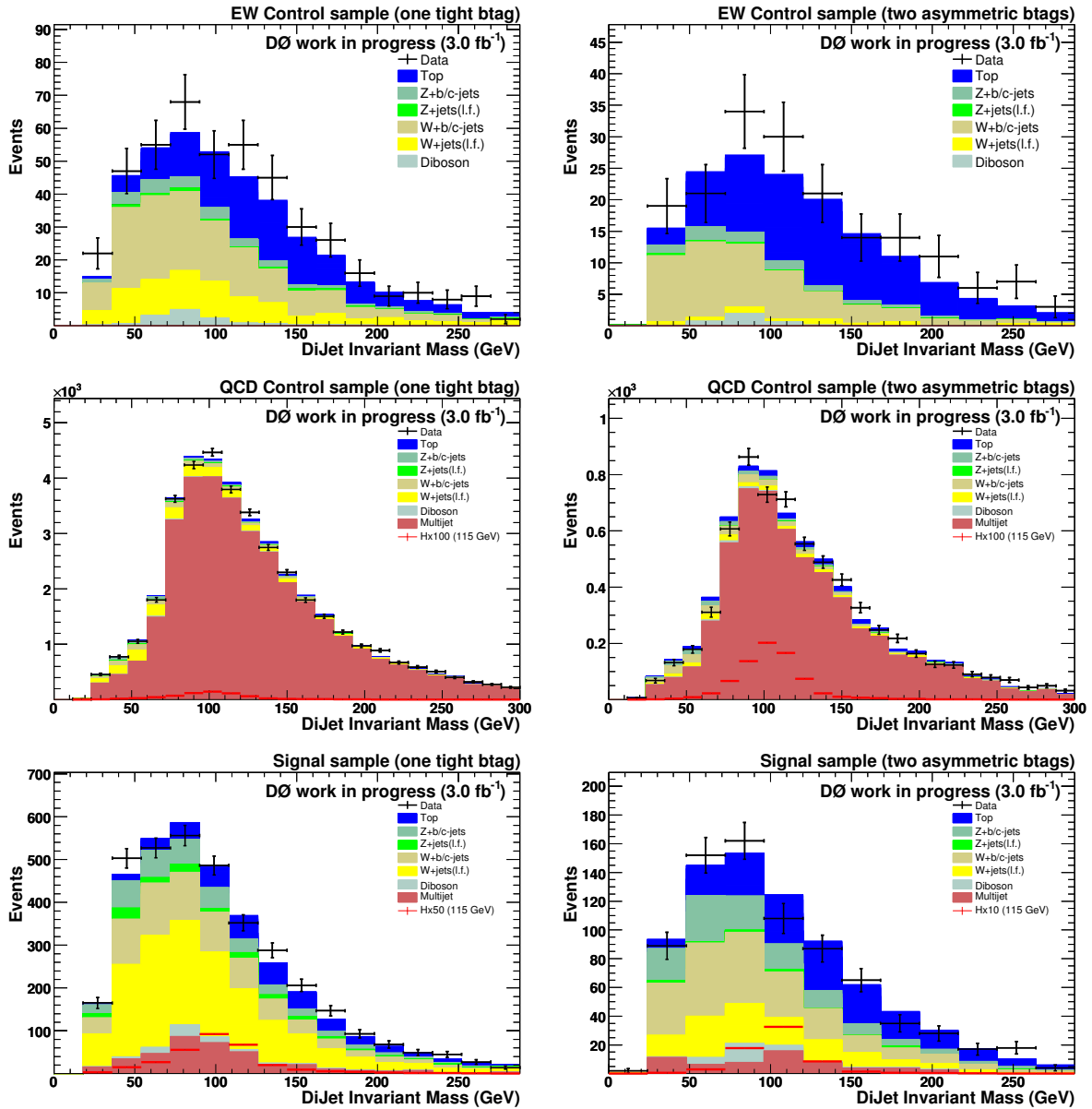


FIG. 4.16: Masse invariante des deux jets principaux après étiquetage tight (gauche) ou asymétrique (droite) dans l'échantillon de contrôle électrofaible (haut), échantillon de contrôle QCD (milieu) et échantillon de signal (bas).

4.7 Etiquetage des jets issus de quark b avec SLTNN

On utilise les mêmes corrections de confirmation par les vertex et d'étiquetabilité décrites précédemment. SLTNN utilise trois qualités de muons (*loose*, *medium* et *medium3*) et possède trois points de fonctionnement (*LOOSE*, *MEDIUM* et *TIGHT*). Il existe donc neuf possibilités pour étiqueter un jet avec SLTNN (voir tableau 4.11).

qualité de μ	points de fonctionnement		
	LOOSE	MEDIUM	TIGHT
loose	lL	lM	lT
medium	mL	mM	mT
medium3	m3L	m3M	m3T

TAB. 4.11: Les neuf possibilités d'étiquetage avec SLTNN.

L'efficacité d'étiquetage de SLTNN étant d'environ 5%, le cas où les deux jets sont étiquetés aura une efficacité très faible. Nous étudierons par conséquent seulement les cas où uniquement un des deux jets est étiqueté par SLTNN parmi les neuf possibilités (lL, lM, ..., m3T). Le cas où les deux jets sont étiquetés *looseLOOSE* (2lL) sera montré pour illustration. Les facteurs de correction d'étiquetage sont définis de la manière suivante :

$$SF_{lL} = SF_{\text{jet étiqueté } lL} \times SF_{\text{jet non-étiqueté } lL} \quad (4.15)$$

$$SF_{m3T} = SF_{\text{jet étiqueté } m3T} \times SF_{\text{jet non-étiqueté } lL} \quad (4.16)$$

$$SF_{2lL} = SF_{\text{jet étiqueté } lL} \times SF_{\text{jet étiqueté } lL} \quad (4.17)$$

Les facteurs de renormalisation QCD après étiquetage (tableau 4.12) sont obtenus de la même manière que pour l'étiquetage NN.

4.7.1 Résultats de la méthode d'étiquetage SLTNN

Les tableaux 4.13 et 4.14 récapitulent les nombres d'évènements de données et de bruit fond obtenus après étiquetage avec SLTNN pour tous les cas pris en compte. Ces résultats tiennent compte de l'efficacité de reconstruction du muon dans le jet. La manière dont on a procédé est détaillée dans la section 4.7.3.

Les figures 4.17 et 4.18 montrent les distributions de la masse invariante des deux jets principaux après étiquetage avec SLTNN lL, lM, lT et 2lL dans les lots de contrôle et de signal. Les distributions pour les autres combinaisons d'étiquetage sont dans l'annexe C.

On peut noter que la sensibilité de l'analyse diminue lorsque l'on augmente la qualité du muon et qu'elle augmente lorsque l'on augmente le point de fonctionnement. Nous avons la confirmation que dans le cas de l'étiquetage 2lL, nous avons une trop faible statistique pour exploiter ce cas.

Niveau d'étiquetage	échantillon		
	contrôle électrofaible	contrôle QCD	signal
SLTNN 1L	1.00	1.32 ± 0.01	1.02 ± 0.03
SLTNN 1M	1.00	1.34 ± 0.01	1.03 ± 0.03
SLTNN 1T	1.00	1.34 ± 0.01	1.03 ± 0.03
SLTNN mL	1.00	1.45 ± 0.01	1.12 ± 0.04
SLTNN mM	1.00	1.47 ± 0.01	1.13 ± 0.04
SLTNN mT	1.00	1.47 ± 0.01	1.13 ± 0.04
SLTNN m3L	1.00	1.63 ± 0.01	1.26 ± 0.05
SLTNN m3M	1.00	1.62 ± 0.01	1.25 ± 0.05
SLTNN m3T	1.00	1.64 ± 0.01	1.27 ± 0.05
SLTNN 2 1L	1.00	1.34 ± 0.01	1.04 ± 0.03

TAB. 4.12: Résumé des facteurs de renormalisation déterminés dans les échantillons de contrôle (électrofaible et QCD) et de signal, après étiquetage avec SLTNN.

On obtient un bon accord donnée/simulation uniquement dans les cas où l'on a étiqueté 1L et 1T. L'origine des désaccords observés pour les autres cas provient d'une mauvaise description des données pour de faibles valeurs de p_T des muons (figure 4.19) et muons centraux (figure 4.20), pour un muon à l'intérieur du jet principal ou secondaire étiqueté par SLTNN (1L, 1M et 1T). Ce phénomène est d'ailleurs plus visible dans le cas du jet secondaire et pourra être confirmé en regardant les distributions des variables d'entrée utilisée par SLTNN (annexe D).

Bruit de fond	Niveau d'étiquetage SLTNN		
	1L	1M	1T
Top	248.3 ± 0.5	211.5	221.6
Di-boson	73.0 ± 0.3	58.9	47.3
Zjj	122.5 ± 0.7	90.1	38.2
Zbb/cc	167.5 ± 0.9	135.1	130.2
Wjj	1224.4 ± 1.8	1052.0	712.5
Wbb/cc	398.9 ± 1.5	331.1	308.2
Phy. Bkgd	2234.6 ± 2.7	1878.7	1458.0
Instr. Bkgd	144.6 ± 0.5	143.9	86.8
Tot. Bkgd	2379.2 ± 2.8	2022.6	1544.8
Données	2466	2181	1549
signal	4.3 ± 0.0	3.7	3.9
S/\sqrt{B}	0.11	0.10	0.12

TAB. 4.13: Nombres d'évènements de bruit de fond, de signal et de données après étiquetage SLTNN (uniquement un jet étiqueté muon loose et les tous points de fonctionnement) dans l'échantillon de signal. La sensibilité (S/\sqrt{B}) a été déterminée dans une fenêtre de masse de ± 35 GeV autour de 115 GeV.

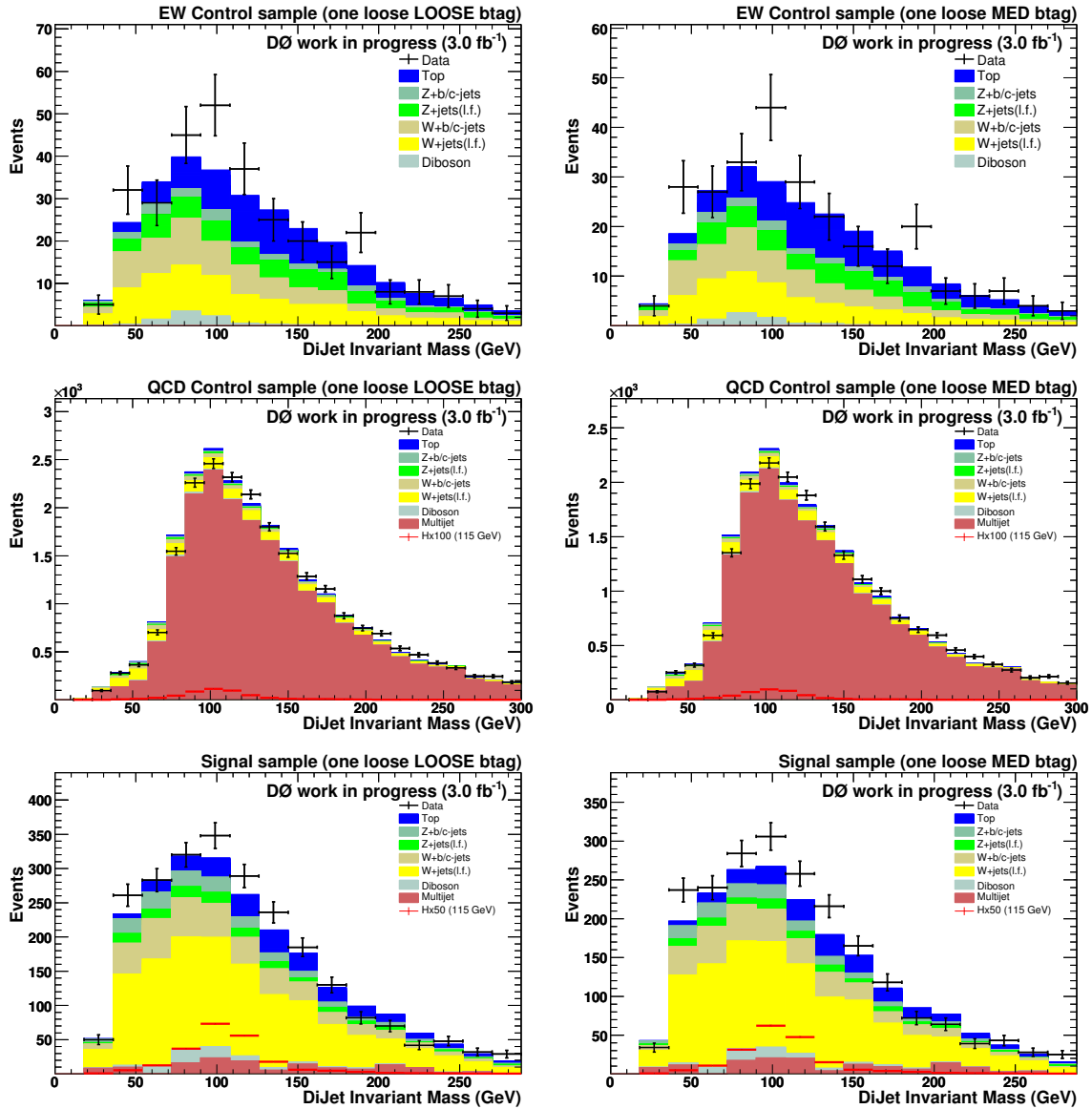


FIG. 4.17: Masse invariante des deux jets principaux dans l'échantillon de contrôle électro-faible (haut), dans l'échantillon de contrôle QCD (milieu) et dans l'échantillon de signal (bas) après étiquetage avec SLTNN dans le cas IL (à gauche) et LM (à droite).

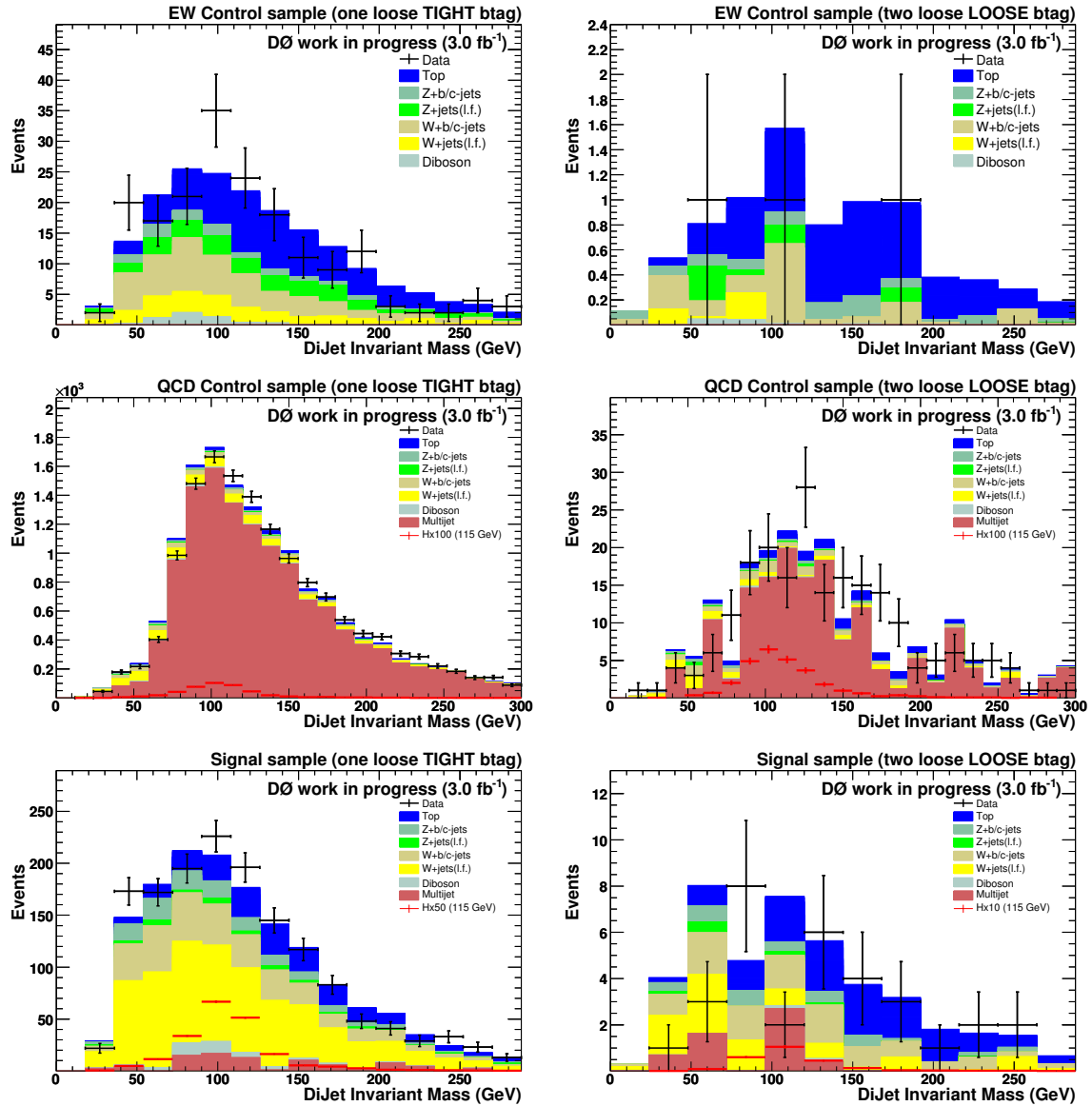


FIG. 4.18: Masse invariante des deux jets principaux dans l'échantillon de contrôle électro-faible (haut), dans l'échantillon de contrôle QCD (milieu) et dans l'échantillon de signal (bas) après étiquetage avec SLTNN dans le cas IT (à gauche) et 2IL (à droite).

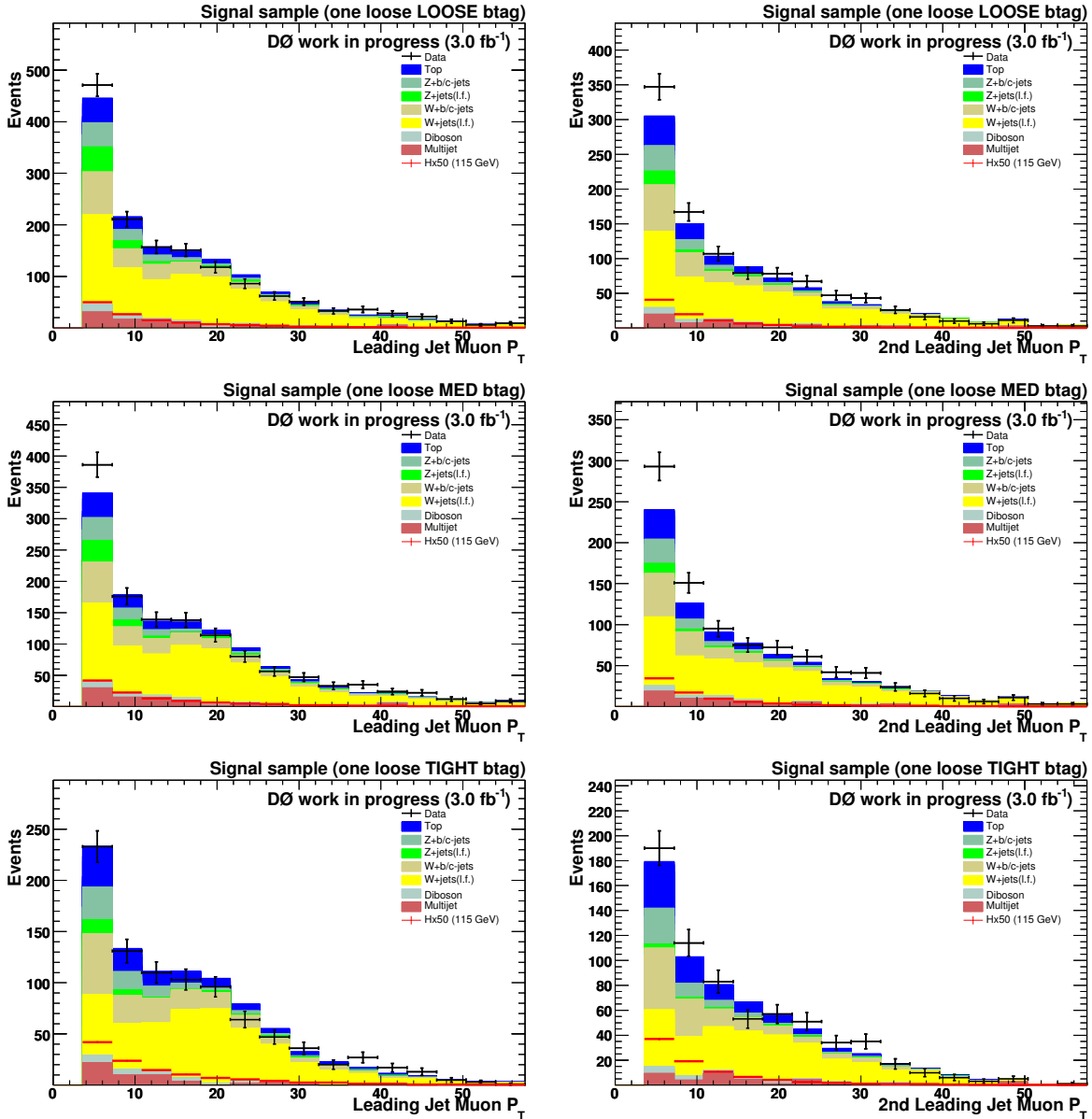


FIG. 4.19: p_T du muon à l'intérieur du jet principal (gauche) et secondaire (droite) après étiquetage SLTNN dans les cas LL (haut), LM (milieu) et LT (bas).

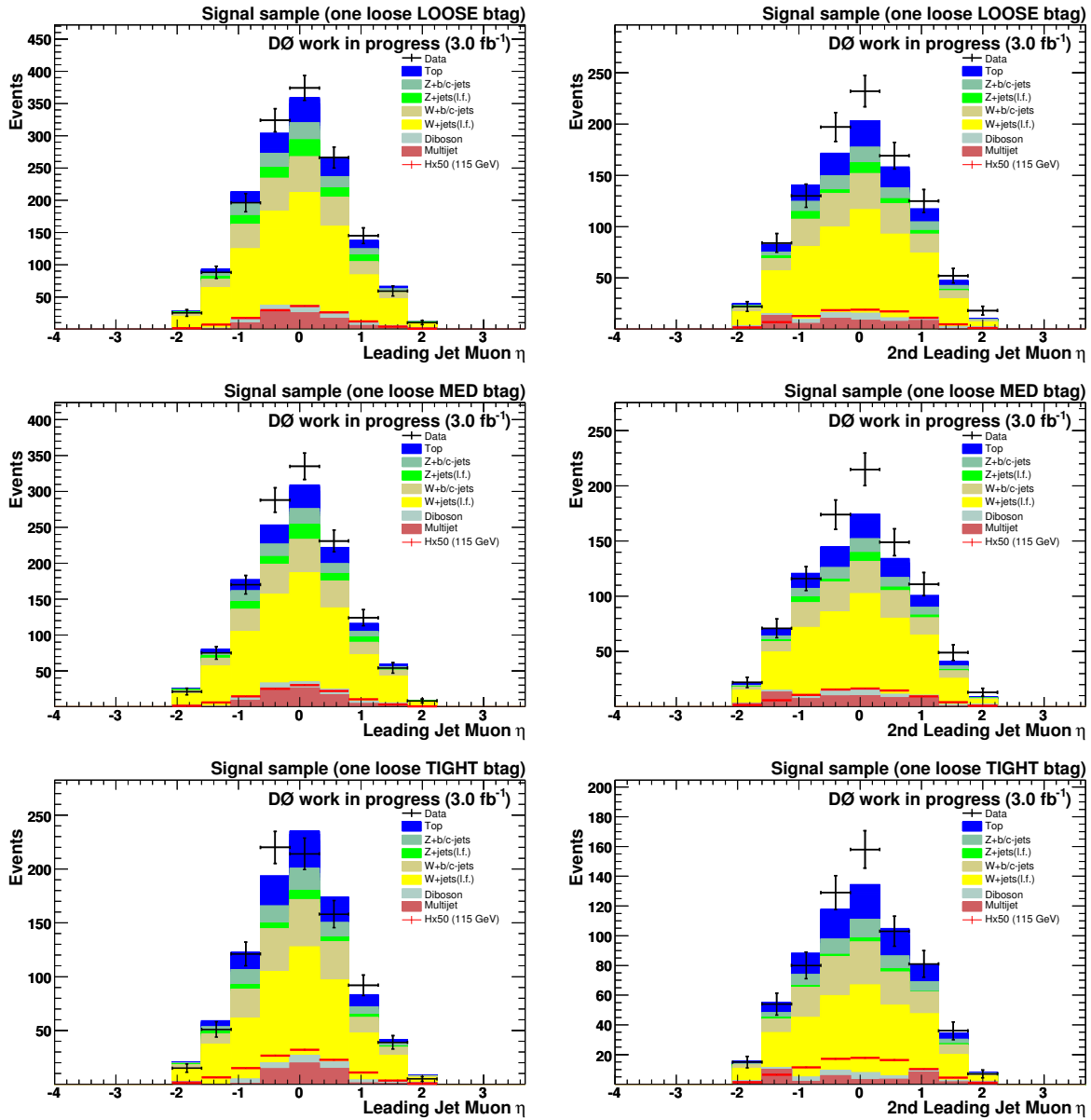


FIG. 4.20: η du muon à l'intérieur du jet principal (gauche) et secondaire (droite) après étiquetage SLTNN dans les cas LL (haut), LM (milieu) et LT (bas).

Bruit de fond	Niveau d'étiquetage SLTNN						
	1mL	1mM	1mT	1m3L	1m3M	1m3T	2LL
Top	174.8	142.6	155.4	154.2	127.6	139.5	14.3 ± 0.0
Di-boson	57.8	46.6	38.7	49.9	41.3	35.0	0.8 ± 0.0
Zjj	103.5	72.3	38.9	94.3	68.9	36.4	0.8 ± 0.0
Zbb/cc	109.1	86.7	87.1	95.5	75.9	78.4	4.0 ± 0.0
Wjj	1069.8	938.4	727.7	932.6	833.5	626.8	9.9 ± 0.0
Wbb/cc	288.1	237.3	231.2	260.0	216.2	207.6	10.9 ± 0.1
Phy. Bkgd	1803.0	1524.1	1279.1	1586.6	1363.4	1123.7	40.7 ± 0.1
Instr. Bkgd	191.0	170.4	117.4	233.3	195.4	142.8	3.4 ± 0.0
Tot. Bkgd	1994.0	1694.4	1396.5	1819.9	1558.8	1266.6	44.1 ± 0.1
Données	2255	2031	1576	2029	1823	1409	32
signal	2.9	2.3	2.6	2.5	2.1	2.4	0.2 ± 0.0
S/\sqrt{B}	0.08	0.07	0.08	0.07	0.06	0.08	0.05

TAB. 4.14: Nombres d'évènements de bruit de fond, de signal et de données après étiquetage SLTNN (uniquement un jet étiqueté muon medium et medium3 et les tous points de fonctionnement, deux jets étiquetés muon loose et point de fonctionnement LOOSE) dans l'échantillon de signal. La sensibilité (S/\sqrt{B}) a été déterminée dans une fenêtre de masse de ± 35 GeV autour de 115 GeV.

4.7.2 Etude du désaccord donnée/simulation

On a observé précédemment que nous n'obtenions pas un accord donnée/simulation satisfaisant dans la plupart des cas d'étiquetage avec SLTNN. Ce désaccord apparaît clairement pour les évènements dont le jet secondaire est étiqueté.

En se limitant seulement au cas où le jet principal est étiqueté par SLTNN, les désaccords données/simulation peuvent être grandements améliorés comme illustré dans les figures 4.21, 4.22 et 4.23. Avec cette contrainte, 40% de la statistique disponible est cependant perdue. Il conviendrait donc pour les études futures d'améliorer la modélisation du jet muonique secondaire, c'est-à-dire la modélisation des muons de bas p_T non-isolés.

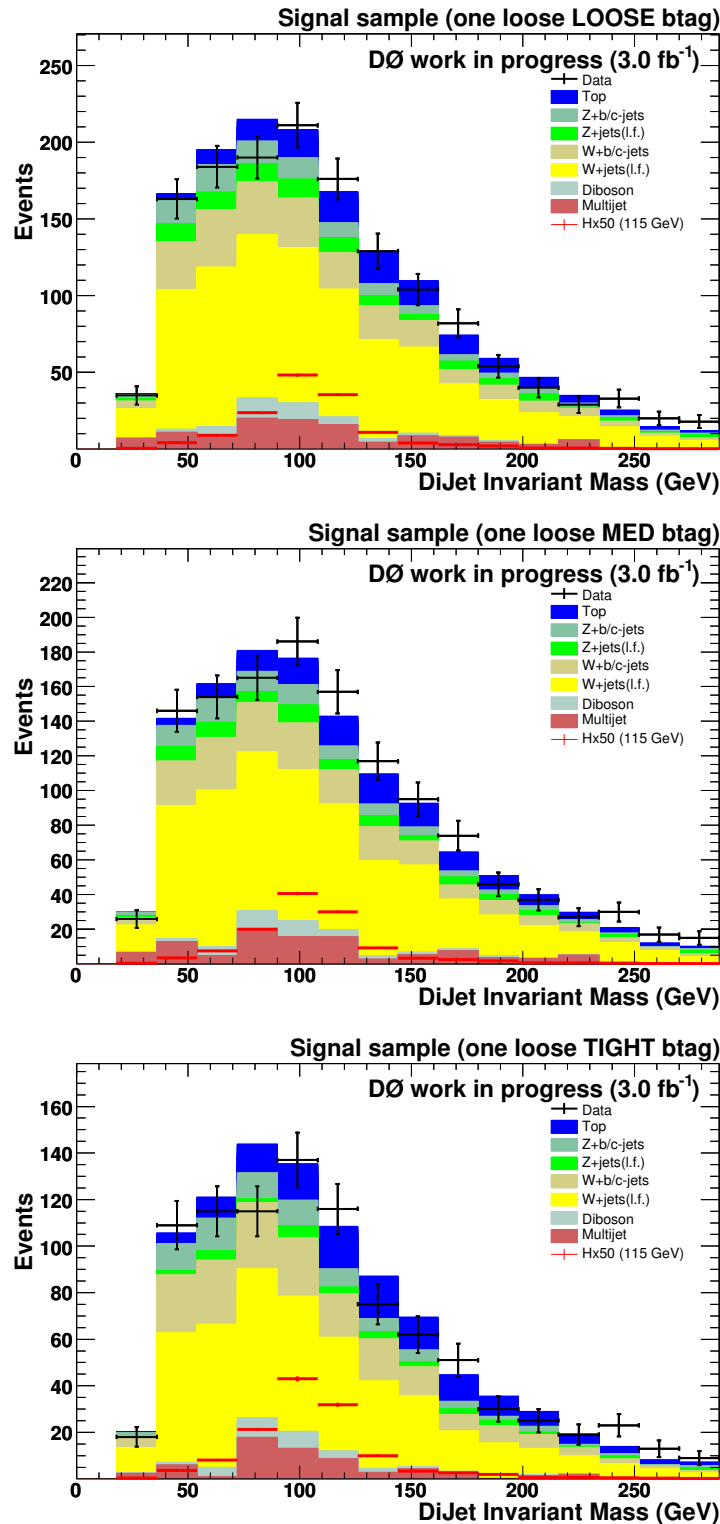


FIG. 4.21: Masse invariante des deux jets principaux lorsque l'on étiquette uniquement le jet principal avec SLTNN pour un muon de qualité *loose* et les points de fonctionnement *LOOSE* (haut), *MEDIUM* (milieu) et *TIGHT* (bas).

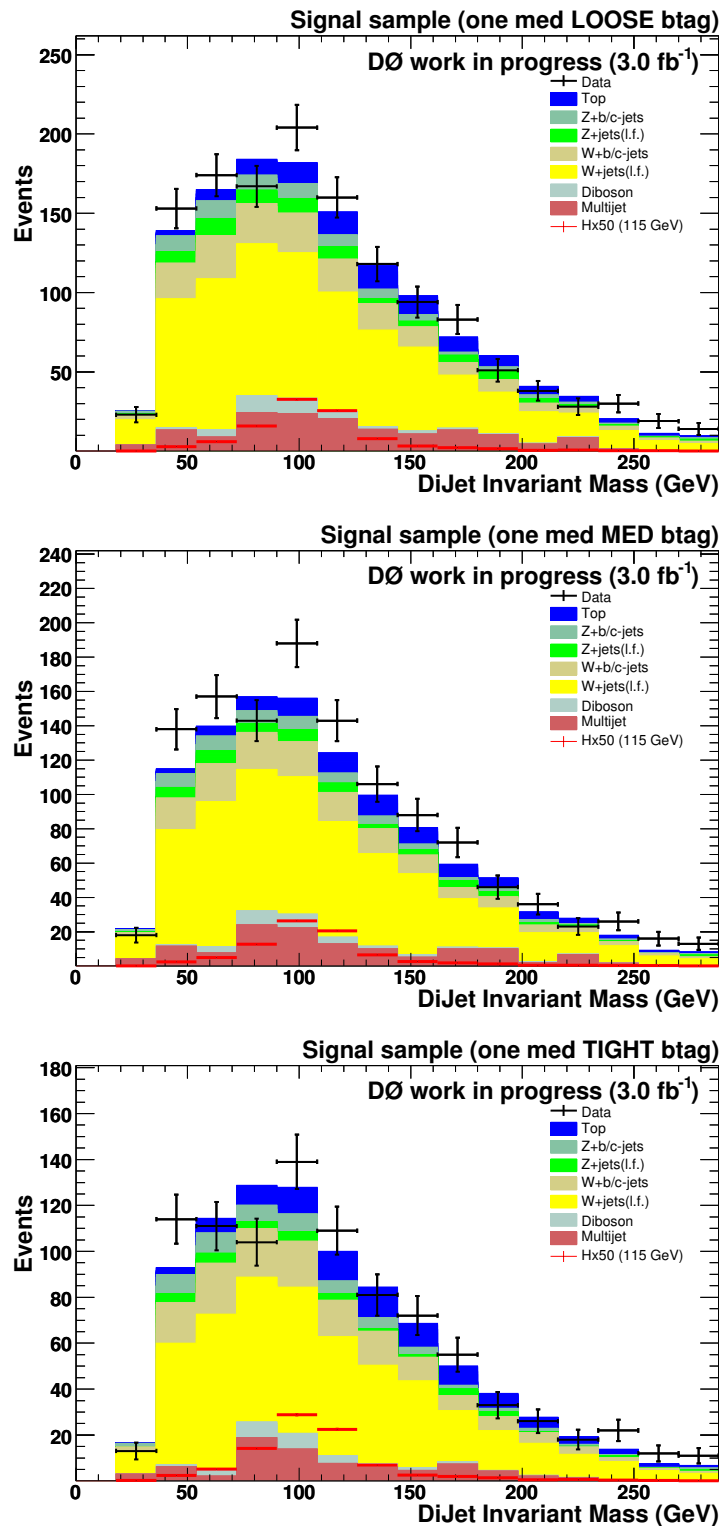


FIG. 4.22: Masse invariante des deux jets principaux lorsque l'on étiquette uniquement le jet principal avec SLTNN pour un muon de qualité *med* et les points de fonctionnement *LOOSE* (haut), *MEDIUM* (milieu) et *TIGHT* (bas).

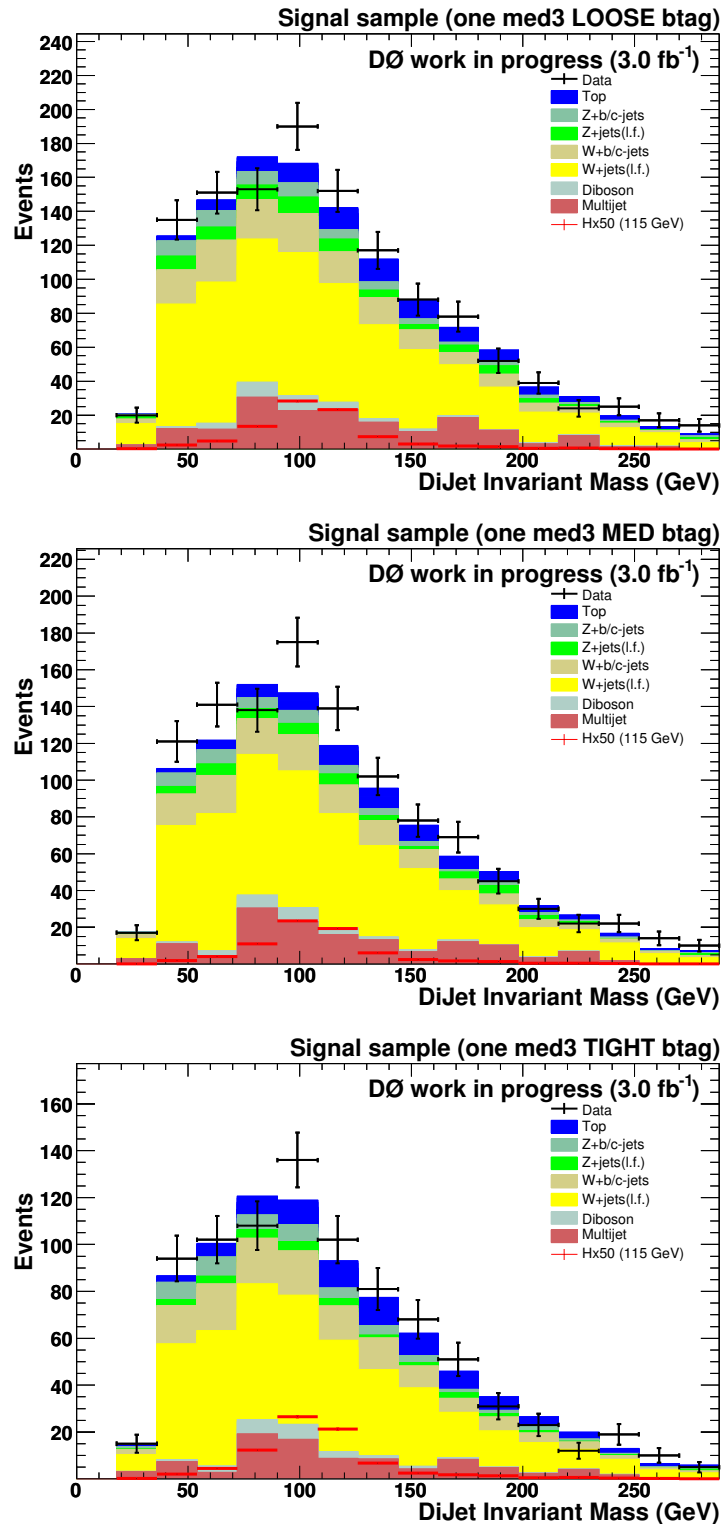


FIG. 4.23: Masse invariante des deux jets principaux lorsque l'on étiquette uniquement le jet principal avec SLTNN pour un muon de qualité $med3$ et les points de fonctionnement *LOOSE* (haut), *MEDIUM* (milieu) et *TIGHT* (bas).

4.7.3 Comparaison du NN et SLTNN dans l'échantillon muonique

L'objectif de cette partie est d'évaluer le gain que pourrait apporter SLTNN par rapport à une analyse faite uniquement avec NN (l'outil standard de la collaboration). Afin d'avoir une comparaison possible des deux méthodes, nous utiliserons des jets contenant un muon (jets "muoniques").

A la suite de la dernière coupure d'analyse sur l'assymétrie (21.3 évènements de signal pour 3 fb^{-1}), nous construisons deux échantillons :

- échantillon non-muonique : il n'y a pas de muons à l'intérieur des deux jets ($\Delta R(\mu, \text{jet}) > 0.5$);
- échantillon muonique : au moins un des deux jets contient un muon.

Seul NN peut étiqueter les jets dans l'échantillon non-muonique. La figure 4.24 montre les distributions de la masse invariante des deux jets principaux dans cet échantillon avant et après étiquetage *tight* et asymétrique. Comme le montre le tableau 4.15, l'échantillon non-muonique contient 16.6 évènements de signal dont 9.5 sont étiquetés par NN (4.7 dans le cas *tight* et 4.8 dans le cas asymétrique).

Bruit de fond	Niveau d'étiquetage		
	avant	tight	asymétrique
Top	933.7	292.3	151.6
Di-boson	1311.8	67.6	20.7
Zjj	10393.2	126.8	7.0
Zbb/cc	2300.8	320.2	109.2
Wjj	38875.4	1370.1	124.6
Wbb/cc	4624.4	603.2	186.5
Phy. Bkgd	58439.3	2780.1	599.8
Instr. Bkgd	5102.7	349.3	58.5
Tot. Bkgd	63542.0	3129.4	658.2
Données	63542	3186	646
signal (115 GeV)	16.6	4.7	4.8
S/\sqrt{B}	0.08	0.09	0.24

TAB. 4.15: Nombres d'évènements de bruit de fond, de signal et de données, avant et après étiquetage (*tight* ou asymétrique) dans l'échantillon non-muonique. La sensibilité (S/\sqrt{B}) a été déterminée dans une fenêtre de masse de $\pm 35 \text{ GeV}$ autour de 115 GeV .

Les mêmes distributions avant et après étiquetage *tight* et asymétrique dans l'échantillon muonique, se retrouvent dans la figure 4.25 pour NN. Le tableau 4.16 montre que l'échantillon muonique contient 4.7 évènements de signal dont 2.8 sont étiquetés par NN (1.3 dans le cas *tight* et 1.5 dans le cas asymétrique) avec une sensibilité (S/\sqrt{B}) de 0.08 (0.09 et 0.24 respectivement). Les résultats après étiquetage avec SLTNN sur ce même lot ont déjà été présentés dans la section 4.7.1. Pour rappel, SLTNN étiquette 4.3 évènements de signal dans le cas LL

avec une sensibilité de 0.11.

Comme nous travaillons avec au moins un des deux jets contenant un muon, il nous faut tenir compte de l'efficacité de reconstruction d'un muon dans la simulation de ces évènements (section 3.4).

On peut donc calculer les corrections (EvtCorr) que l'on va apporter à la simulation et ce, évènement par évènement, avant étiquetage et après étiquetage avec NN (les outils produits par SLTNN tiennent déjà compte de cette correction). On prend en compte les effets suivants :

- seulement un jet a un muon :

$$\text{EvtCorr} = \frac{\varepsilon_{\mu}^{\text{Données[jetA]}}}{\varepsilon_{\mu}^{\text{MC[jetA]}}} \quad (4.18)$$

- les deux jets ont un muon :

$$\text{EvtCorr} = \frac{\varepsilon_{\mu}^{\text{Données[jetA]}}}{\varepsilon_{\mu}^{\text{MC[jetA]}}} \times \frac{\varepsilon_{\mu}^{\text{Données[jetB]}}}{\varepsilon_{\mu}^{\text{MC[jetB]}}} \quad (4.19)$$

- les deux jets ont un muon, cependant un seul a été reconstruit. Nous utiliserons la correction moyenne suivante :

$$\text{EvtCorr} = \frac{1}{2} \times \left(\frac{\varepsilon_{\mu}^{\text{Données[jetA]}}}{\varepsilon_{\mu}^{\text{MC[jetA]}}} \times \frac{1 - \varepsilon_{\mu}^{\text{Données[jetB]}}}{1 - \varepsilon_{\mu}^{\text{MC[jetB]}}} + \frac{\varepsilon_{\mu}^{\text{Données[jetB]}}}{\varepsilon_{\mu}^{\text{MC[jetB]}}} \times \frac{1 - \varepsilon_{\mu}^{\text{Données[jetA]}}}{1 - \varepsilon_{\mu}^{\text{MC[jetA]}}} \right) \quad (4.20)$$

Bruit de fond	Niveau d'étiquetage		
	avant	tight	asymétrique
Top	286.3	88.9	57.5
Di-boson	150.1	10.9	4.4
Zjj	517.3	7.7	0.6
Zbb/cc	218.0	53.2	23.4
Wjj	2668.5	93.2	8.4
Wbb/cc	516.6	102.6	41.0
Phy. Bkgd	4356.7	356.6	135.4
Instr. Bkgd	266.3	44.5	3.6
Tot. Bkgd	4626.0	401.1	139.0
DATA	4626	407	138
signal (115 GeV)	4.7	1.3	1.5
S/\sqrt{B}	0.08	0.07	0.17

TAB. 4.16: Nombres d'évènements de bruit de fond, de signal et de données, avant et après étiquetage (tight ou asymétrique) dans l'échantillon muonique. La sensibilité (S/\sqrt{B}) a été déterminée dans une fenêtre de masse de ± 35 GeV autour de 115 GeV.

Nous allons regarder les trois cas possibles suivants à l'intérieur de l'échantillon muonique (voir figure 4.26) :

- cas A : les évènements sont étiquetés avec SLTNN (toutes les configurations possibles) et non avec NN (*tight* et asymétrique);
- cas B : les évènements sont étiquetés avec SLTNN (1L) et avec NN (*tight* : cas B1 et asymétrique : cas B2);
- cas C : les évènements ne sont pas étiquetés avec SLTNN (1L et 2L), ni NN (*tight* et asymétrique).

Cas A

Ce cas d'étude montre le gain qu'apporte SLTNN par rapport à l'analyse officielle faite uniquement avec NN. Le tableau 4.17 résume le détail du nombre d'évènements de données, de bruits de fond et de signal pour toutes les configurations d'étiquetage possibles avec SLTNN sur ce lot.

Sample	étiqueté uniquement avec SLTNN								
	1L	1LM	1LT	1mL	1mM	1mT	1m3L	1m3M	1m3T
Top	82.2	67.7	61.6	55.0	46.4	50.1	48.5	41.4	46.0
Di-boson	57.3	45.4	32.6	35.1	32.2	29.4	29.0	27.2	26.2
Zjj	117.5	85.5	34.7	46.3	46.2	36.5	37.0	37.7	32.7
Zbb/cc	83.8	63.4	49.1	43.6	36.7	37.3	36.2	30.8	34.4
Wjj	1169.0	1000.4	662.2	795.9	778.4	689.1	673.6	674.0	590.3
Wbb/cc	246.7	199.0	161.7	157.2	139.7	139.1	137.6	124.5	125.7
Phy. Bkgd	1756.5	1461.4	1002.0	1133.1	1079.6	981.5	961.8	935.6	855.3
Instr. Bkgd	120.5	112.4	66.7	108.7	111.2	88.8	119.9	118.9	112.5
Tot. Bkgd	1877.0	1573.8	1068.7	1241.8	1190.8	1070.3	1081.7	1054.5	967.8
Données	2050	1776	1175	1439	1420	1238	1238	1228	1106
signal (115 GeV)	1.0	0.9	0.8	0.7	0.6	0.6	0.6	0.5	0.6
S/\sqrt{B}	0.03	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.02

TAB. 4.17: Nombres d'évènements de bruit de fond, de signal et de données, avant et après étiquetage avec SLTNN dans l'échantillon muonique (cas A). La sensibilité (S/\sqrt{B}) a été déterminée dans une fenêtre de masse de ± 35 GeV autour de 115 GeV.

On peut voir qu'avec le critère d'étiquetage le plus lâche, il nous est possible de récupérer 1 évènement de signal de boson de Higgs (à comparer aux 12.3 évènements que l'on obtient avec une analyse basée sur NN). On peut aussi malheureusement remarquer que la sensibilité associée est très faible ($S/\sqrt{B} = 0.03$) car le bruit de fond reste important dans tous les cas de figure. Nous concluerons ce résultat en terme de gain sur la limite dans la section 4.7.4.

Cas B

Aussi bien SLTNN que NN ont retenu ces évènements. Il n'y a donc rien à gagner sur ces 3.4 évènements de signal. La figure 4.27 ainsi que le tableau 4.18 illustrent les cas B1 et B2.

Le cas B2 contient 2.1 évènements de signal de boson de Higgs et a une sensibilité de 0.21. Le cas B1 pour sa part, contient 1.2 évènements de signal de boson de Higgs pour une sensibilité de 0.07. On voit donc que sur les 4.3 évènements de signal de boson de Higgs dans le cas de l'étiquetage avec SLTNN LL (i.e. cas A + B), 3.4 évènements de signal de boson de Higgs sont aussi étiquetés par NN. L'algorithme NN est donc aussi très efficace pour les jets muoniques. L'étude du cas B1 est intéressante dans la mesure où l'on pourrait définir une méthode d'étiquetage plus robuste si NN et SLTNN n'étiquettent pas le même jet. Nous définissons donc un échantillon B1a qui correspond à cette nouvelle contrainte et un échantillon B1b qui ne la satisfait pas. Les résultats dans ce nouvel échantillon sont aussi regroupés dans la figure 4.27 et le tableau 4.18. On remarque que dans 80% des cas les deux méthodes d'étiquetage étiquettent le même jet et donc ce *tagging* "asymétrique" (NN + SLTNN) apporterait peu d'amélioration.

Bruit de fond	corrélation des méthodes d'étiquetage			
	cas B1	cas B2	cas B1a	cas B1b
Top	90.8	75.3	20.0	70.8
Di-boson	10.0	5.7	2.0	7.9
Zjj	4.6	0.4	0.6	3.9
Zbb/cc	53.1	30.6	7.5	45.6
Wjj	51.0	4.4	8.7	42.3
Wbb/cc	100.4	51.8	17.0	83.4
Phy. Bkgd	309.8	168.2	55.8	254.0
Instr. Bkgd	36.2	3.5	3.3	33.1
Tot. Bkgd	346.1	171.7	59.2	287.1
Données	326	124	69	257
signal (115 GeV)	1.2	2.1	0.3	0.9
S/\sqrt{B}	0.07	0.21	0.06	0.05

TAB. 4.18: Nombres d'évènements de bruit de fond, de signal et de données, avant et après étiquetage avec SLTNN dans l'échantillon muonique (cas B1, B2, B1a et B2b). La sensibilité (S/\sqrt{B}) a été déterminée dans une fenêtre de masse de ± 35 GeV autour de 115 GeV.

Cas C

La figure 4.28 et le tableau 4.19 montrent le nombre d'évènements qui ne sont ni étiquetés par NN, ni par SLTNN. On peut voir que dans ce cas les bruits de fond restant sont majoritairement constitués de jets de saveurs légères et que le nombre de signal de boson de Higgs perdu est très faible (0.3 évènements).

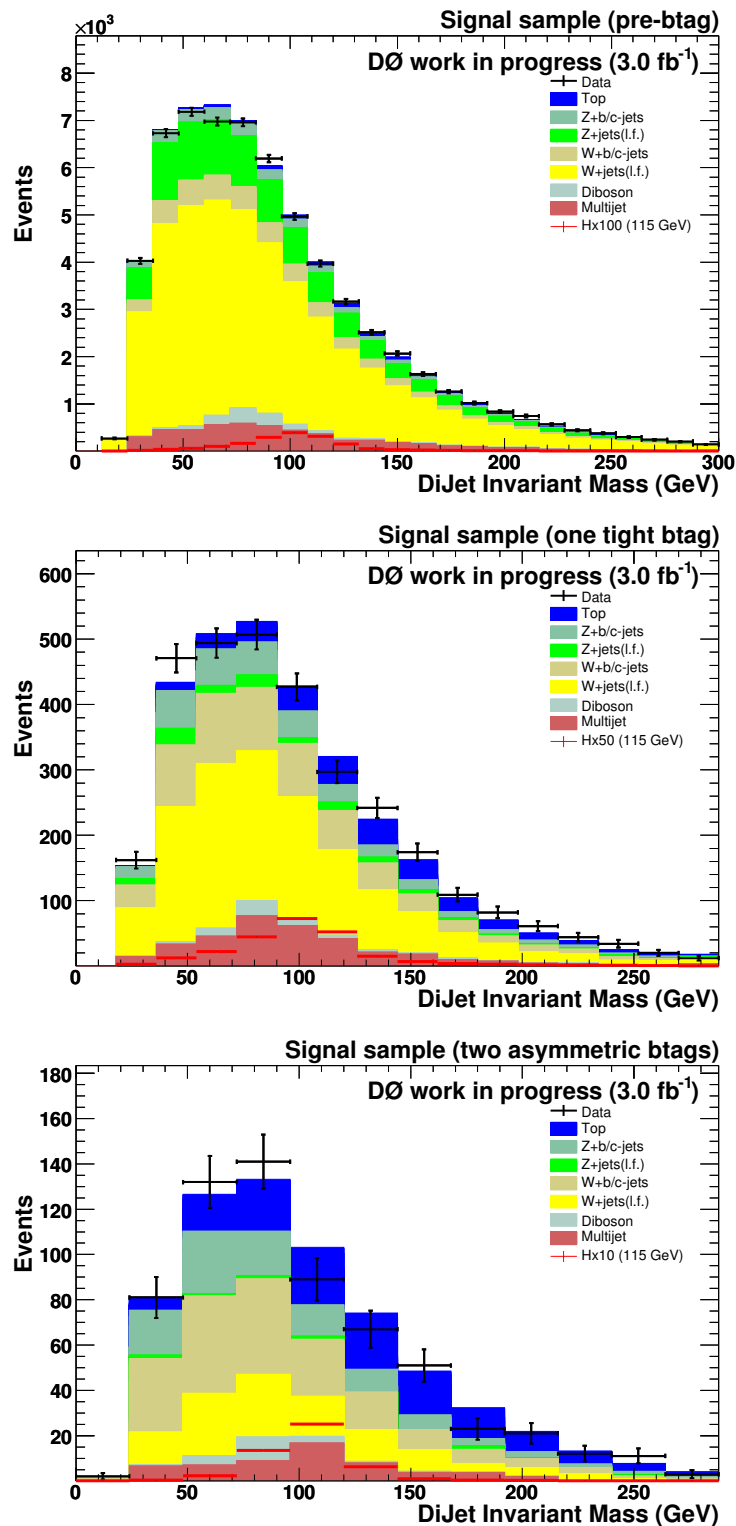


FIG. 4.24: Masse invariante des deux jets principaux avant (haut) et après étiquetage tight (milieu) ou asymétrique (bas) dans l'échantillon non-muonique à l'aide de l'algorithme de b -tagging NN.

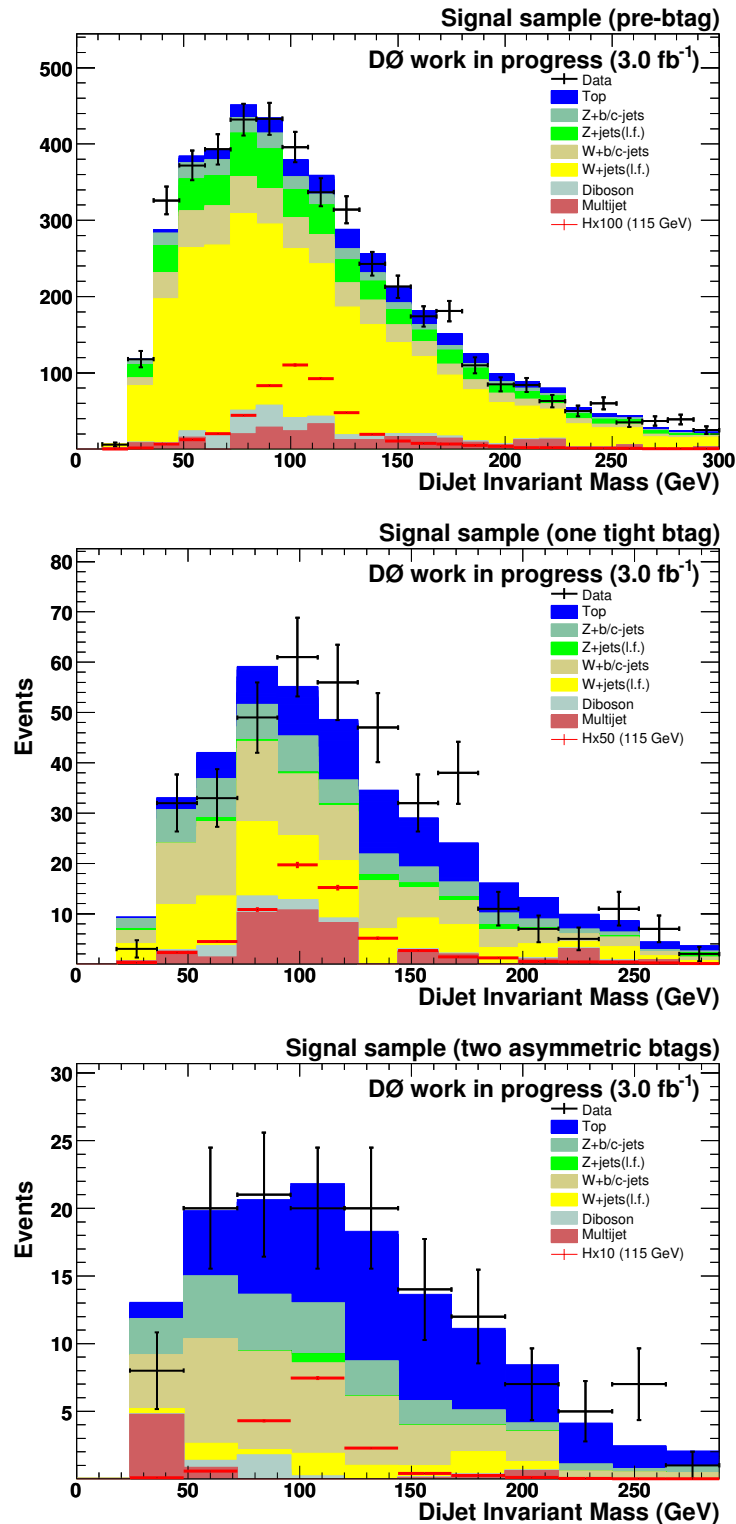


FIG. 4.25: Masse invariante des deux jets principaux avant (haut) et après étiquetage tight (milieu) ou asymétrique (bas) dans l'échantillon muonique à l'aide de l'algorithme de b -tagging NN.

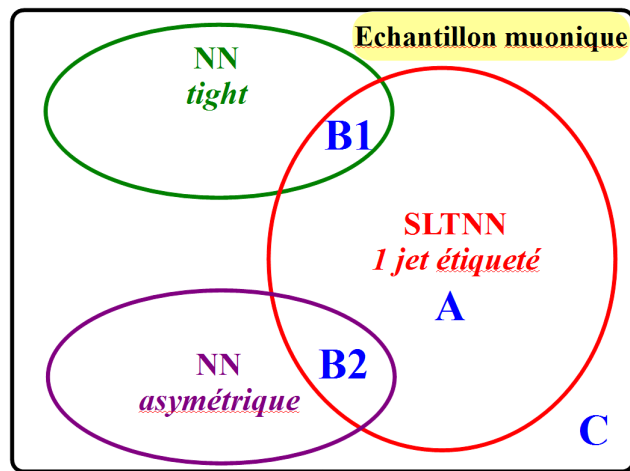


FIG. 4.26: Les différents cas étudiés dans l'échantillon muonique.

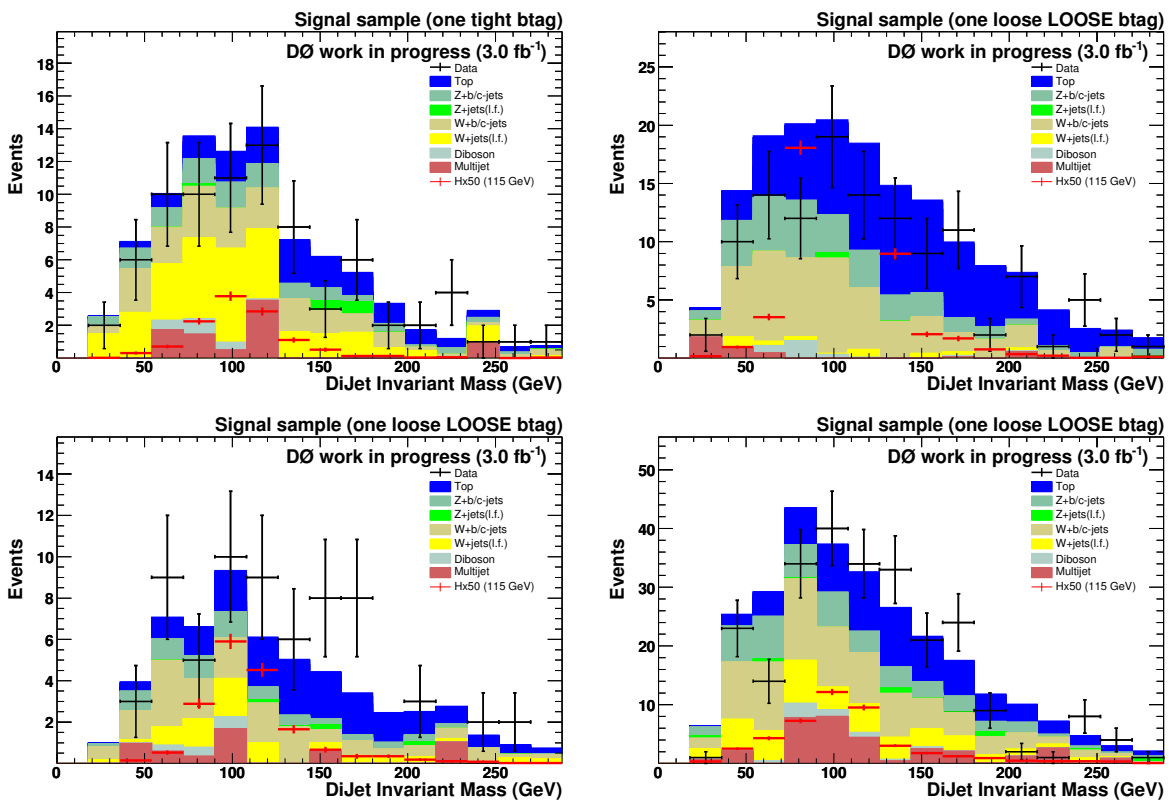


FIG. 4.27: Masse invariante des deux jets principaux après étiquetage pour le cas B1 (en haut à gauche), B2 (en haut à droite), B1a (en bas à gauche) et B1b (en bas à droite), dans l'échantillon muonique.

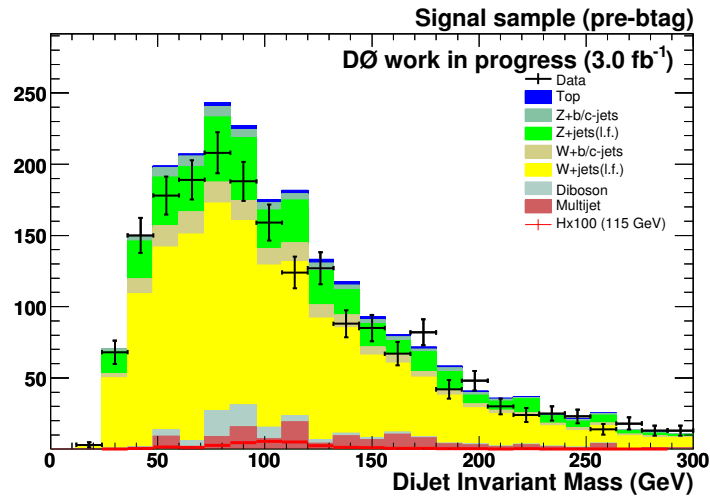


FIG. 4.28: Masse invariante des deux jets principaux après étiquetage pour le cas C, dans l'échantillon muonique.

Bruit de fond	cas C
Top	28.7
Di-boson	74.4
Zjj	405.2
Zbb/cc	72.2
Wjj	1433.4
Wbb/cc	155.7
Phy. Bkgd	2169.7
Instr. Bkgd	116.0
Tot. Bkgd	2285.7
Données	2014
signal (115 GeV)	0.3
S/\sqrt{B}	0.01

TAB. 4.19: Nombres d'évènements de bruit de fond, de signal et de données, avant et après étiquetage avec SLTNN dans l'échantillon muonique (cas C). La sensibilité (S/\sqrt{B}) a été déterminée dans une fenêtre de masse de ± 35 GeV autour de 115 GeV.

4.7.4 Evaluation du gain en terme de limite

L'analyse officielle $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ peut encore améliorer sa sensibilité sur les lots d'étiquetage *tight* et asymétrique (appelé par la suite lot "NN") en utilisant une méthode multivariée [83] : les arbres de décision DT (pour *Decision Tree*). Le principe de fonctionnement des arbres de décisions est brièvement expliqué en annexe A.3 et son utilisation dans la collaboration DØ en [76]. Cette méthode permet, comme le réseau de neurones, de définir une nouvelle variable discriminante à partir d'un jeu de variables préalablement sélectionnées. Cette variable de sortie permet de séparer le signal de boson de Higgs des bruits de fonds top, W/Z +jets et dibosons. Elle est finalement utilisée pour déterminer la limite sur la section efficace de production du boson de Higgs étudié : $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ (+ $WH \rightarrow l^\pm\nu b\bar{b}$).

L'objectif de ce chapitre est d'évaluer le gain rapporté par SLTNN dans l'analyse présentée. Il ne nous est donc pas utile d'utiliser les arbres de décision pour effectuer cette comparaison. Nous déterminerons donc la limite sur la section efficace de production de boson de Higgs à partir de la masse invariante des deux jets principaux dans le lot "NN" et dans le lot contenant les jets uniquement étiquetés par SLTNN (cas A).

On détermine une limite supérieure sur la section efficace de production de boson de Higgs [78] [79] [80] car la comparaison données/simulation n'a montré aucun excès d'évènements compatible avec le signal de Higgs considéré. On évalue cette limite en utilisant une approche fréquentiste modifiée [77] qui teste deux hypothèses :

- H_0 : l'observation est compatible avec une hypothèse de bruit de fond uniquement ;
- H_1 : une nouvelle particule est observée.

L'analyse $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ attend b évènements de bruit de fond, s évènements de signal et possède N évènements dans les données. N représente la variable aléatoire n suivant une loi de Poisson de paramètre b (pour l'hypothèse H_0) ou $s + b$ (pour l'hypothèse H_1). A partir de la définition de ces deux probabilités :

- CL_b : l'hypothèse H_0 est vérifiée dans les données. On peut donc estimer la non-compatibilité des données avec cette hypothèse par : $1-CL_b$;
- CL_{s+b} : l'hypothèse H_1 est vérifiée ;

on construit le niveau de confiance CL_s :

$$CL_s = \frac{CL_{s+b}}{CL_b} \quad (4.21)$$

On effectue une exclusion à 95% de niveau de confiance pour une valeur de CL_s de 5%.

Les hypothèses H_0 et H_1 sont testées par le biais des pseudo-expériences qui suivent une loi de Poisson dont la moyenne vaut b (pour H_0) ou $s + b$ (pour H_1). On utilise un rapport de vraisemblance logarithmique LLR (*Log-Likelihood Ratio*) comme test statistique sur les distributions de masse invariante des deux jets principaux. On définit aussi LLR_{obs} pour les données observées. La valeur de CL_s est évaluée en intégrant les distributions de LLR obtenues sur tous les résultats des pseudo-expériences.

L'évaluation de la limite nécessite une bonne reproduction des données par la simulation. Nous n'utiliserons donc que les cas où les jets sont étiquetés LL et LT par SLTNN car ils sont les seuls à remplir cette condition.

La figure 4.29 et le tableau 4.20 montrent les résultats de limites obtenus en fonction des masses de boson de Higgs prises en compte (entre 100 et 150 GeV par pas de 5 GeV). Ces résultats sont présentés en terme de section efficace exclue divisé par la section efficace du Modèle Standard (cette dernière tient compte du rapport de branchement du boson de Higgs se désintégrant en $b\bar{b}$). Ces limites ont été évaluées pour le lot "NN" seul et le lot "NN"+cas A (LL et LT). La différence entre ces dernières met en évidence le gain de l'ajout de la méthode d'étiquetage SLTNN dans l'analyse.

On voit une très faible amélioration ($\sim 1\%$) pour la limite attendue. Le résultat sur la limite observée n'est pas plus favorable et même parfois défavorable mais n'est pas significatif du point de vue de l'ajout de SLTNN dans l'analyse. Il est possible d'inclure les erreurs systématiques de l'analyse dans le calcul de limites en faisant une dégradation des densités de probabilité du signal et des bruits de fond. Cependant les résultats de limites obtenus ne sont pas suffisants pour avoir besoin de prendre en compte cet effet.

Echantillon	Lim.	Masse du boson de Higgs (GeV)										
		100	105	110	115	120	125	130	135	140	145	150
NN	Att.	3.97	4.31	4.84	5.56	6.52	7.97	10.11	13.30	18.36	27.14	42.57
	Obs.	4.26	4.18	4.35	4.72	5.47	6.70	8.98	12.67	18.78	30.29	51.17
NN+cas A (LL)	Att.	3.94	4.37	4.82	5.40	6.57	7.94	10.06	13.35	18.41	26.91	42.84
	Obs.	4.77	4.77	4.86	5.36	6.16	7.54	10.30	14.34	21.28	34.02	56.77
NN+cas A (LT)	Att.	3.94	4.28	4.82	5.52	6.52	7.88	10.04	13.27	18.21	27.14	42.32
	Obs.	4.64	4.64	4.78	5.22	6.09	7.42	9.92	13.93	20.42	33.08	55.09

TAB. 4.20: Résultat sur la limite (Lim.) attendue (Att.) et observée (Obs.) pour l'analyse NN (*tight* et asymétrique) et l'analyse NN + le cas A (LL ou LT) pour les masses de boson de Higgs comprises entre 100 et 150 GeV par pas de 5 GeV.

4.8 Conclusion

L'analyse de recherche de boson de Higgs dans le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ a été présentée. Elle utilise 3 fb^{-1} des données du *RunIII* et recherche les événements $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ et $WH \rightarrow l^\pm\nu b\bar{b}$ (dans le cas où le lepton n'est pas identifié). Les bruits de fond considérés sont les bruits de fond électrofaible (simulés par des méthodes *Monte-Carlo*) et instrumental (de type QCD, dérivé à partir des données).

La première étape de l'analyse consiste à vérifier la maîtrise de ces bruits de fond dans des lots de contrôle (nommés électrofaible et QCD) et de définir des facteurs de renormalisation de ces deux types de bruits de fond si nécessaire. Ensuite, nous réduisons les bruits de fond à l'aide de coupures sur des variables cinématiques ou caractéristiques du bruit de fond

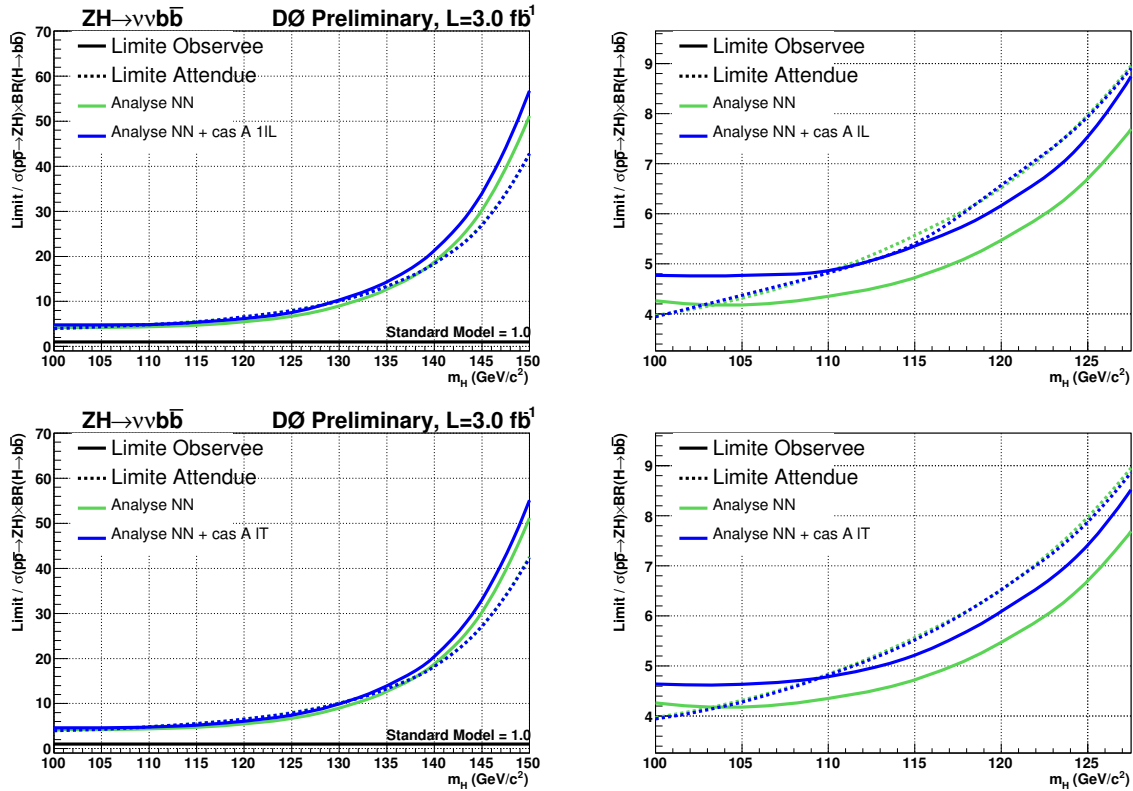


FIG. 4.29: Résultat sur la limite attendue (ligne pointillée) et observée (ligne continue) pour l'analyse NN (*tight* et asymétrique) (courbe verte) et l'analyse NN+le cas A (IL en haut ou IT en bas) (courbe bleue) pour les masses de boson de Higgs comprises entre 100 et 150 GeV par pas de 5 GeV. La limite est représentée pour l'ensemble des valeurs de masse de boson de Higgs (à gauche) et sur les zones les plus pertinentes (à droite).

considéré, en tenant compte des facteurs de renormalisation définis. Cette étape augmente la sensibilité de l'analyse (S/\sqrt{B}) de 0.007 à 0.082.

On utilise par la suite l'identification des jets issus de quark b avec la méthode NN afin d'accroître encore plus la sensibilité. Pour ce faire, on définit deux lots d'étiquetage orthogonaux : *tight* (un des deux jets est étiqueté VT et l'autre jet n'est pas étiqueté L3) et asymétrique (un des deux jets est étiqueté VT et l'autre est étiqueté L3) correspondant respectivement à une sensibilité de 0.11 et 0.29. On peut voir un bon accord entre données et simulation sur ces deux lots. Nous avons aussi regardé la réponse de la méthode SLTNN sur ce lot pour des événements dont un seul des deux jets est étiqueté, et ceci pour tous les points de fonctionnement (l'autre jet ne passant pas le critère IL). Il est apparu que l'accord données/simulations n'est bon que pour les jets principaux étiquetés. La sensibilité obtenue est de l'ordre de 0.1.

Finalement, nous avons estimé le gain qu'apporte SLTNN dans cette analyse en définissant un lot dont les jets sont uniquement étiquetés par SLTNN. Il apparaît que l'ajout de ces événements permet d'ajouter 10% de signal de boson de Higgs par rapport à l'analyse officielle

basée sur NN. Ce résultat a aussi été traduit en termes de gains sur la limite de la section efficace de production de boson de Higgs en prenant en compte les bruits de fond. En comparant la limite obtenue avec le lot "NN" (jets étiquetés tight et asymétrique par NN uniquement) et le lot où l'on ajoute les lots étiquetés par SLTNN uniquement, on voit une amélioration de la limite attendue de seulement 1%.

Conclusion

Le thème principal de cette thèse est la recherche du boson de Higgs du Modèle Standard dans le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$. Cette recherche a été effectuée avec les données fournies par le Tevatron et collectées par le détecteur DØ. La luminosité intégrée utilisée est de 3 fb^{-1} . Une nouvelle méthode d'étiquetage des jets issus de quark b (SLTNN [62]) a été développée afin d'améliorer la sensibilité de l'analyse aux jets provenant de désintégrations semi-leptoniques de quark b.

Dans 20% des cas, un jet issu d'un quark b contiendra un muon. Après avoir sélectionné ces jets, on utilise un réseau de neurone afin d'augmenter l'efficacité d'étiquetage des jets issus de quark b et de diminuer le taux de mauvaise identification des jets issus de quarks légers (u, d, s) ou gluon. Les efficacités ont été évaluées dans des lots de simulation et de données. L'efficacité de *b-tagging* est $\sim 5\%$ (ceci comprend le rapport de branchement $b \rightarrow c \rightarrow \mu$ et $b \rightarrow \mu$) pour un taux de mauvaise identification inférieur à 1%. Sur des jets issus de désintégration semi-leptoniques de quark b, SLTNN améliore de 10% l'efficacité de *b-tagging* par rapport à l'algorithme standard de la collaboration : NN.

SLTNN a été développé en suivant les procédures standard de la méthode NN. Les efficacités de *b-tagging* calibrées sur les données sont intégrées aux outils officiels du *b-tagging* de la collaboration. De plus les erreurs systématiques associées à ces corrections ont été estimées et sont de l'ordre de 5%. Cet algorithme a été revu par l'*editorial board* interne à DØ en charge du *b-tagging*.

L'algorithme SLTNN a été utilisé dans l'analyse de recherche de boson de Higgs dans le canal de désintégration $ZH \rightarrow \nu\bar{\nu}b\bar{b}$. Dans un premier temps, l'analyse officielle a été reproduite avec NN où un bon accord entre les données et la simulation a été observé. Une comparaison détaillée [81] entre NN et SLTNN a été effectuée. Des désaccords entre données et simulation ont été constatés pour ces événements semi-leptoniques contenant des muons isolés de bas p_T . Il serait également possible d'optimiser d'avantage l'entraînement du réseau de neurones pour améliorer son pouvoir discriminant ou de le remplacer par un

arbre de décision (une étude préliminaire a montré un gain supplémentaire de 10%).

Dans le meilleur des cas, 10% d'évènements de signal de boson de Higgs peuvent être ajoutés à l'analyse officielle qui est basée sur l'algorithme standard de *b-tagging* de la collaboration (NN). Cependant la sensibilité à un boson de Higgs du lot correspondant à ces évènements de signal ajouté reste marginale et nécessite d'avantage d'optimisation. Ceci se traduit en effet par un gain inférieur à 1% sur la limite de la masse du boson de Higgs. Enfin des développements plus récents sur l'identification des muons ou l'évaluation du taux de mauvaise identification de la méthode NN ont été effectués dans la collaboration, et ceci en parallèle au développement de SLTNN. Il conviendrait de porter ces améliorations au niveau de SLTNN.

D'autre part, l'analyse présentée a été effectuée sur les évènements déclenchés par les conditions "jets+ \cancel{E}_T ". Il serait possible d'étudier plus en profondeur les évènements déclenchés par les *triggers* " μ +jets". Ces efficacités de *triggers* ne sont pour l'instant disponibles que pour des muons isolés avec $p_T > 15$ GeV. Ce lot de données " μ +jets" permettrait aussi de s'affranchir des corrélations existantes délicates à évaluer entre les méthodes NN et SLTNN lors de la détermination du gain sur la limite.

Réseau de neurones

A.1 Structure du réseau de neurones

Les réseaux de neurones sont des algorithmes qui peuvent être entraînés à partir de la simulation pour identifier un signal et le séparer du bruit de fond. La figure A.1 représente la structure du réseau de neurones. En entrée, des variables appelées neurones reçoivent les informations caractéristiques du signal et du bruit de fond. Ces neurones sont connectés entre eux par des synapses. On associe une fonction d'activation (A) à chaque neurone et un poids à chaque synapse (coefficient synaptique).

Dans cette annexe, nous utilisons un réseau de neurones à couche (*Multi-Layer Perceptron*). Ces informations des neurones de la couche d'entrée sont traitées par une ou plusieurs couches de neurones cachées et sont renvoyées vers la couche de sortie, qui comprend dans notre cas d'étude un unique neurone. Le réseau de neurones détermine la valeur des coefficients synaptiques afin de pouvoir répondre si la topologie de l'évènement correspond à un jet issu de quark b (sortie du réseau de neurones égale à 1) ou non (sortie égale à 0). Pour voir ce qu'il se passe, suivons un évènement passant à travers le réseau de neurones défini dans la figure A.1. Nous choisissons i variables (x_i) de cet évènement dans la couche d'entrée. La fonction d'activation utilisée est linéaire :

$$X_i = A(x_i) = x_i. \quad (\text{A.1})$$

Les j neurones de la couche cachée vont traiter les informations de la couche d'entrée en tenant compte des coefficients synaptiques :

$$y_j = \sum_i \omega_{ij} \times X_i \quad (\text{A.2})$$

et en sachant que la fonction de probabilité de sortie (Z continue) peut être approximée par une combinaison linéaire de sigmoïdes. En utilisant une fonction d'activation de type

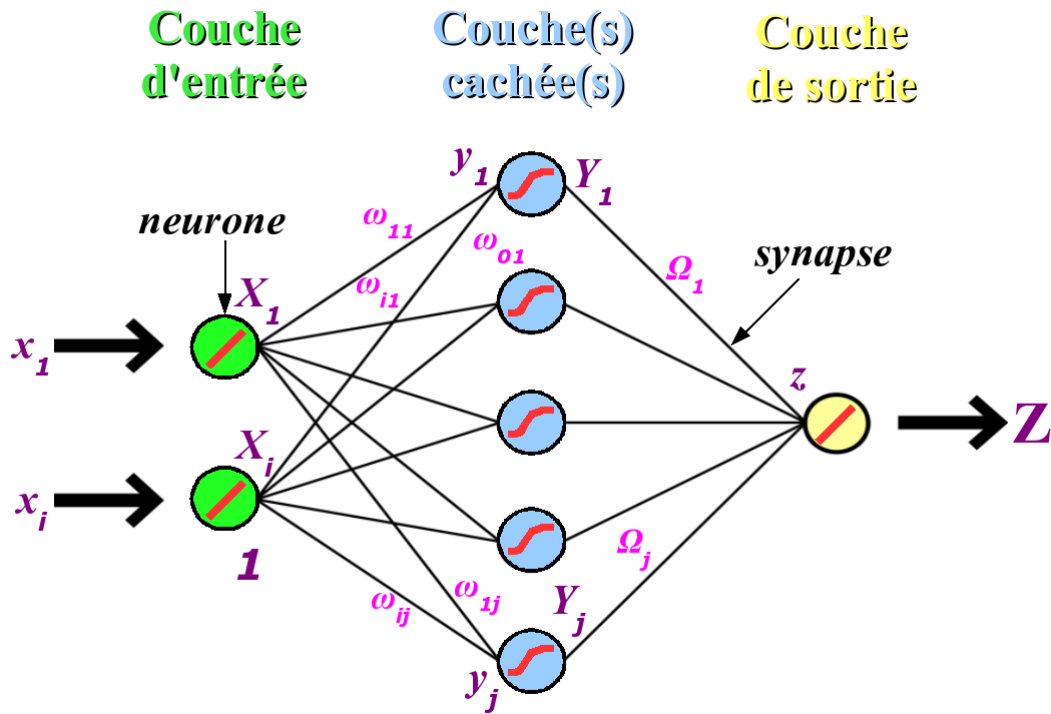


FIG. A.1: Structure d'un réseau de neurone perceptron multi-couche (*Multi-Layer Perceptron*).

sigmoïde pour les y_j , on définit :

$$Y_j = A(y_j) = \frac{1}{1 + e^{-y_j}} \quad (\text{A.3})$$

Et on procède de la même manière pour la sortie du réseau de neurones (avec une fonction d'activation linéaire) :

$$z = \sum_j \Omega_j \times Y_j \quad (\text{A.4})$$

$$Z = A(z) = z \quad (\text{A.5})$$

Il est possible de rajouter une deuxième couche possédant un unique neurone dans la couche cachée (avec une fonction d'activation de type sigmoïde). Cela a pour effet de normaliser la sortie du réseau de neurone entre 0 et 1. En effet dans ce cas, on aurait :

$$Y' = A(y') = \frac{1}{1 + e^{-y'}} \quad (\text{A.6})$$

avec

$$y' = \sum_j \Omega_j \times Y_j \quad (\text{A.7})$$

Cette étude a été effectuée lors du développement de la méthode d'étiquetage NN [52]. La figure A.2 montre la sortie du réseau de neurones pour les configurations avec ou sans l'unique neurone dans la couche cachée. On peut voir que le pouvoir discriminant n'est pas sensiblement modifié mais que la sortie est normalisée entre 0 et 1.

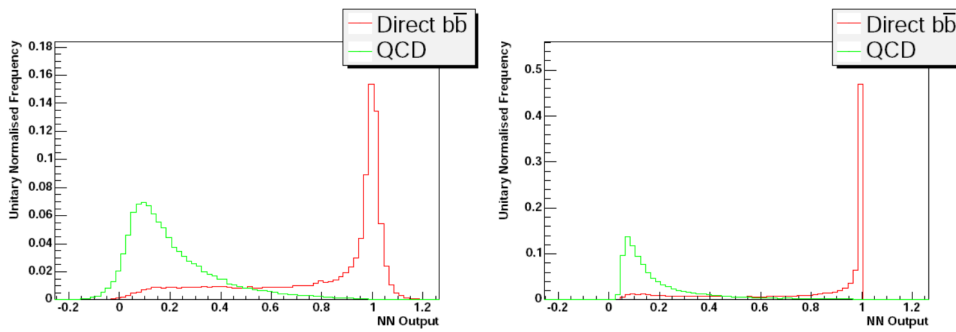


FIG. A.2: Sortie du réseau de neurones pour les configurations avec (à droite) ou sans (à gauche) l'unique neurone dans la couche cachée. Cette étude a été effectuée lors du développement de la méthode d'étiquetage NN [52].

A.2 Apprentissage du réseau de neurones

La méthode d'apprentissage utilisée est dite supervisée : on compare la sortie du réseau de neurones Z à la valeur attendue t . Pour ce faire, on utilise les exemples issus du lot d'évènements de signal (jet issu de quark b) pour lequel $t = 1$ et du lot d'évènements de bruit de fond (jet issu de quark u , d ou s ou issu de gluons) pour lequel $t = 0$. Lorsque tous les évènements de simulation ont été utilisés, on a effectué une époque. On entraîne le réseau de neurones en utilisant un grand nombre d'époque. A chaque itération les coefficients synaptiques sont ajustés.

A.2.1 Traitement de l'erreur

Dans la configuration initiale du réseau de neurones, les coefficients synaptiques ont été choisis de manière aléatoire entre -0.5 et 0.5 . L'entraînement du réseau de neurones consiste à minimiser l'erreur faite entre Z et t en ajustant les coefficients synaptiques. On définit l'erreur

E faite à chaque époque contenant p évènements de poids ω_p :

$$E = \sum_p \frac{1}{2} \omega_p (Z_p - t_p)^2 \quad (\text{A.8})$$

Ensuite, il existe plusieurs algorithmes d'entraînement qui utilise la dérivée de l'erreur :

$$\frac{dE}{d\omega_{ij}} = \sum_p \frac{de_p}{d\omega_{ij}} \quad (\text{A.9})$$

où e_p est l'erreur faite à chaque évènement. On appelle cette technique : la "rétropropagation" des erreurs.

Les techniques d'apprentissage cherchent le minimum de l'erreur. La méthode utilisée pour le développement de SLTNN est appelée BFGS [82] (Broyden, Fletcher, Goldfarb, Shanno). La recherche du minimum de l'erreur nécessite de déterminer les zéros du gradient et aussi d'évaluer les points d'inflexions de la dérivée seconde. La méthode BFGS est un algorithme qui permet d'évaluer très rapidement les points d'inflexion afin de rechercher le minimum global de l'erreur, sans faire le calcul explicite de la dérivée seconde.

A.3 Arbres de décision

Les arbres de décision (voir figure A.3) sont une méthode d'apprentissage [83] multivariée au même titre que les réseaux de neurone.

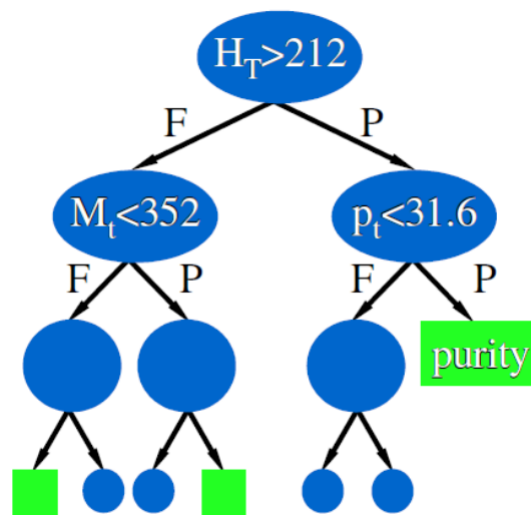


FIG. A.3: Schéma d'un arbre de décision. Les nœuds sont représentés par les bulles bleues et les feuilles par les boîtes vertes. Les flèches noires représentent les tests effectués sur un nœud.

L'arbre de décision utilise un échantillon d'entraînement (contenant du signal et du bruit de fond) et x_i variables choisies par l'utilisateur. Pour chaque variable x_i , l'arbre de décision

va déterminer quelle valeur sur la variable x_i va permettre la meilleure séparation entre le signal et le bruit de fond. La variable x_i donnant la meilleure séparation sera choisie (le nœud nommé H_T sur la figure). Le lot initial se sépare ainsi en deux : le lot qui réussit le test (P) et le lot qui ne le réussit pas (F).

L'opération de recherche de la meilleure variable discriminante est réitérée sur les lots P et F, construisant ainsi un arbre. On continue de cette manière jusqu'à ce que les nœuds ne contiennent que 100 événements ou que l'on ne puisse plus améliorer la séparation signal/bruit de fond. Dans ce cas, le nœud s'appelle une feuille et on associe à cette dernière le critère de pureté : $p = s/(s + b)$, avec s le nombre d'événements de signal et b le nombre d'événements de bruit de fond. La valeur de la feuille correspond à la variable de sortie de l'arbre de décision.

Ensuite, il est possible d'augmenter le pouvoir discriminant de cet arbre T_n en effectuant un *boost* (il existe plusieurs méthodes, mais celle-ci est celle qui a été utilisée). Cela consiste à replacer les variables x_i les plus mal classées de l'arbre T_n , en augmentant le poids associé à ces variables pour entraîner un arbre de décision T_{n+1} . Cette méthode est connue sous le nom de AdaBoost [84].

La figure A.4 montre la comparaison de l'efficacité d'étiquetage des jets issus de quark b en fonction du taux de mauvaise identification entre l'utilisation d'un réseau de neurone et un arbre de décision *boosté* dans la méthode d'étiquetage SLTNN. Le réseau de neurone a une configuration optimisée et l'arbre de décision a une configuration par défaut. On peut voir que quelque soit la qualité du muon, l'arbre de décision a amélioré les performances de la méthode SLTNN. Cette étude est préliminaire, les arbres de décisions ne seront pas encore utilisés dans le développement de SLTNN pour l'instant.

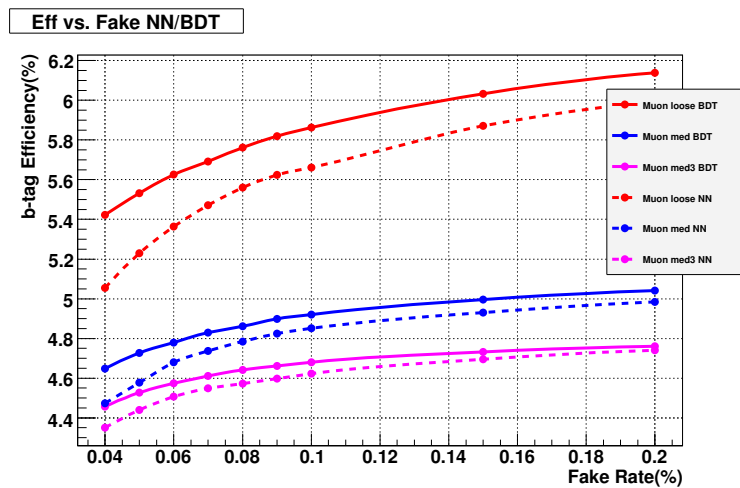


FIG. A.4: Comparaison de l'efficacité d'étiquetage des jets issus de quark b en fonction du taux de mauvaise identification entre l'utilisation d'un réseau de neurones (courbes pointillées) et un arbre de décision *boosté* (courbes en trait plein) dans la méthode d'étiquetage SLTNN. Ce résultat a été effectué avec les trois qualités de muon *loose* (en rouge), *med* (en bleu) et *med3* (en violet). Le réseau de neurone a une configuration optimisée et l'arbre de décision a une configuration par défaut.

Annexe B

**Distributions supplémentaires du
développement de la méthode
d'étiquetage SLTNN**

B.1 Coefficients de corrélation

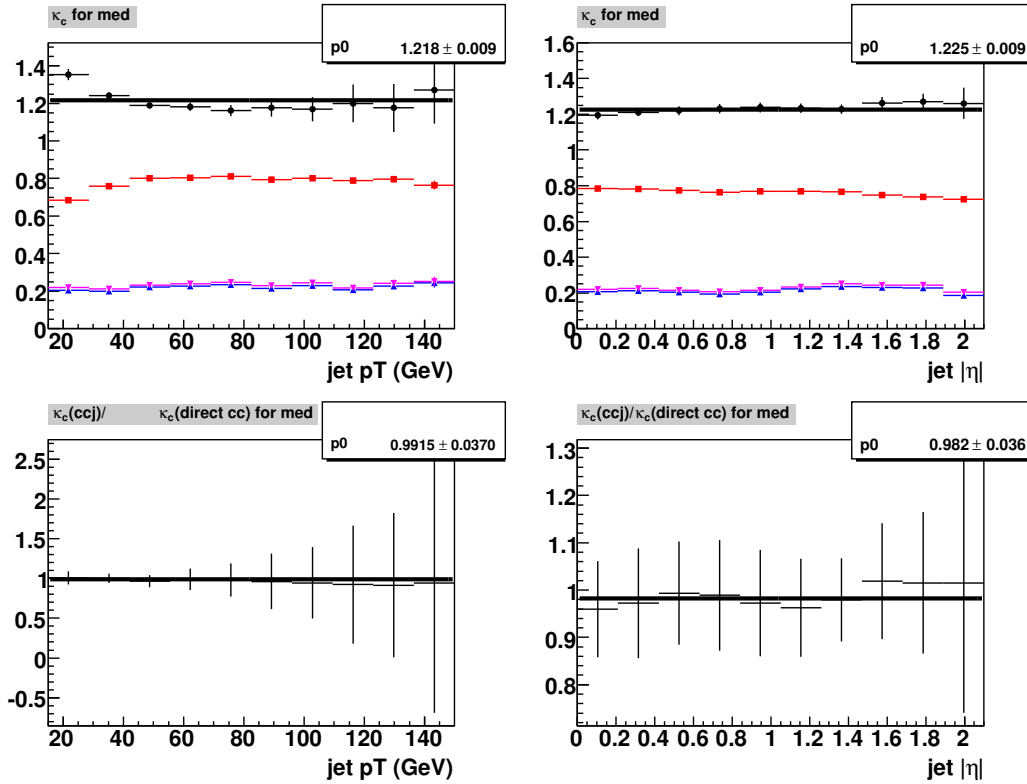


FIG. B.1: Distributions du coefficient de corrélation κ_{cl} pour un muon de qualité *med* et le point de fonctionnement de SLTNN *MEDIUM*. Le coefficient est paramétrisé en fonction du p_T (à gauche) ou de η (à droite) du jet. En bas, rapport entre κ_{cl} obtenu avec le processus simulé cc -jets et celui obtenu avec direct $c\bar{c}$. En haut, κ_{cl} est obtenu avec la combinaison des deux processus simulés (points noir) ajusté par une constante (ϵ_{cl}^{SLTNN} en rouge, $\epsilon_{cl}^{modJLIP}$ en violet et $\epsilon_{cl}^{SLTNN,modJLIP}$ en bleu). Notons que $\epsilon_{cl}^{modJLIP}$ et $\epsilon_{cl}^{SLTNN,modJLIP}$ sont pratiquement superposés. Les erreurs sont uniquement statistiques. En faisant la somme quadratique des erreurs du coefficient de corrélation de la figure du haut et du bas, on détermine l'erreur reportée dans les tableaux 3.10 et 3.11.

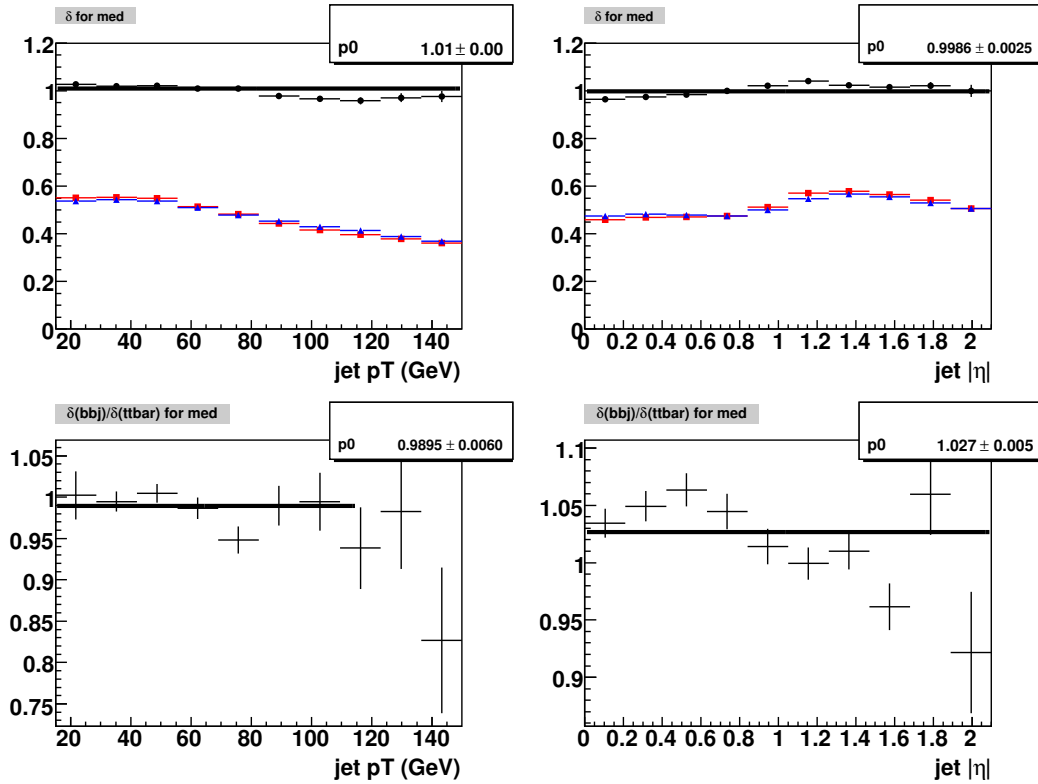


FIG. B.2: Distributions du coefficient de corrélation δ pour un muon de qualité *med* et le point de fonctionnement de SLTNN *MEDIUM*. Le coefficient est paramétrisé en fonction du p_T (à gauche) ou de η (à droite) du jet. En bas, rapport entre δ obtenu avec le processus simulé $bb+jets$ et celui obtenu avec $t\bar{t}$. En haut, δ est obtenu avec la combinaison des deux processus simulés (points noirs) ajusté par une constante ($\epsilon_b^{\text{modJLIP}}$ dans le lot enrichi en rouge ou dans le lot muonique en bleu). Notons que $\epsilon_b^{\text{modJLIP}}$ et dans le lot muonique et enrichi sont pratiquement superposés. Les erreurs sont uniquement statistiques. En faisant la somme quadratique des erreurs du coefficient de corrélation de la figure du haut et du bas, on détermine l'erreur reportée dans les tableaux 3.10 et 3.11.

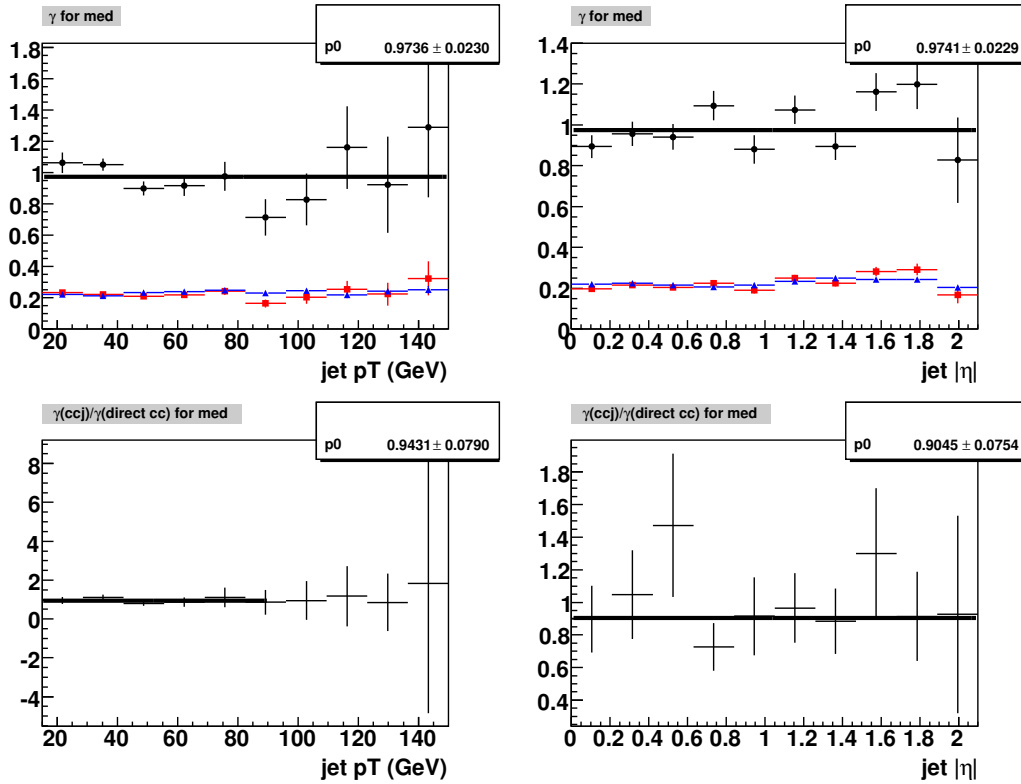


FIG. B.3: Distributions du coefficient de corrélation γ pour un muon de qualité *med* et le point de fonctionnement de SLTNN *MEDIUM*. Le coefficient est paramétrisé en fonction du p_T (à gauche) ou de η (à droite) du jet. En bas, rapport entre γ obtenu avec le processus simulé $cc+\text{jets}$ et celui obtenu avec $\text{direct } c\bar{c}$. En haut, γ est obtenu avec la combinaison des deux processus simulés (points noir) ajusté par une constante ($\epsilon_{cl}^{\text{modJLIP}}$ dans le lot enrichi en rouge ou dans le lot muonique en bleu). Notons que $\epsilon_{cl}^{\text{modJLIP}}$ et dans le lot muonique et enrichi sont pratiquement superposés. Les erreurs sont uniquement statistiques. En faisant la somme quadratique des erreurs du coefficient de corrélation de la figure du haut et du bas, on détermine l'erreur reportée dans les tableaux 3.10 et 3.11.

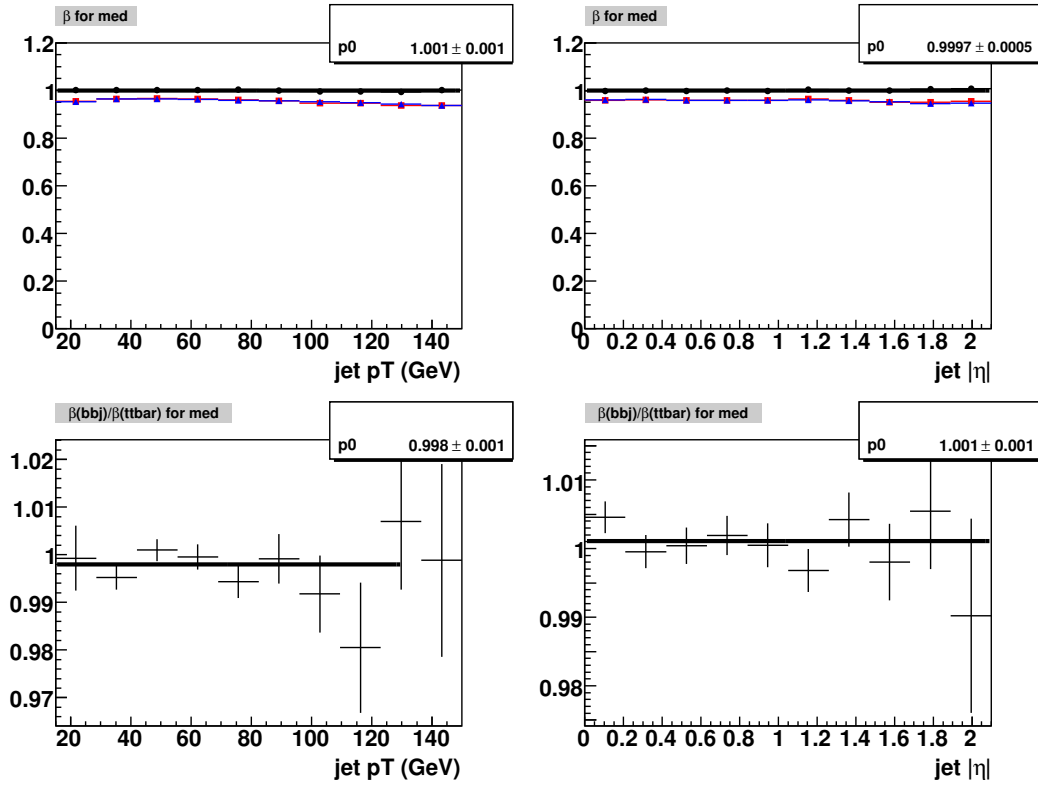


FIG. B.4: Distributions du coefficient de corrélation β pour un muon de qualité *med* et le point de fonctionnement de SLTNN *MEDIUM*. Le coefficient est paramétrisé en fonction du p_T (à gauche) ou de η (à droite) du jet. En bas, rapport entre β obtenu avec le processus simulé bb +jets et celui obtenu avec $t\bar{t}$. En haut, β est obtenu avec la combinaison des deux processus simulés (points noir) ajusté par une constante ($\epsilon_b^{\text{SLTNN}}$ dans le lot enrichi en rouge ou dans le lot muonique en bleu). Notons que $\epsilon_b^{\text{SLTNN}}$ et dans le lot muonique et enrichi sont pratiquement superposés. Les erreurs sont uniquement statistiques. En faisant la somme quadratique des erreurs du coefficient de corrélation de la figure du haut et du bas, on détermine l'erreur reportée dans les tableaux 3.10 et 3.11.

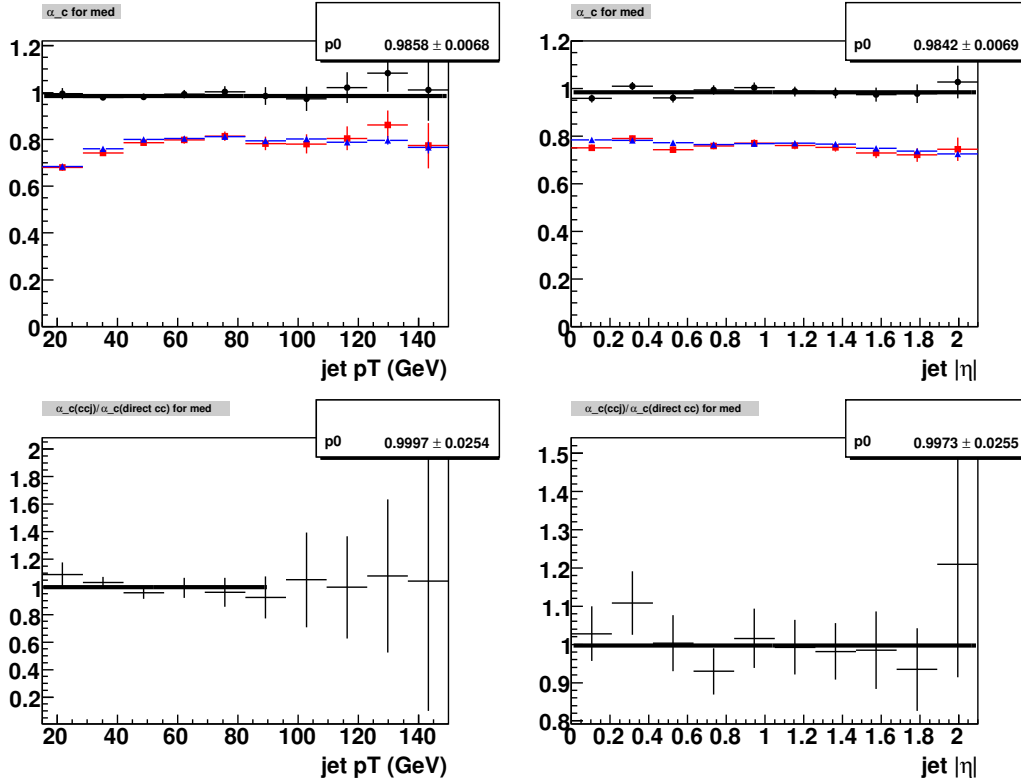


FIG. B.5: Distributions du coefficient de corrélation α pour un muon de qualité *med* et le point de fonctionnement de SLTNN *MEDIUM*. Le coefficient est paramétrisé en fonction du p_T (à gauche) ou de η (à droite) du jet. En bas, rapport entre α obtenu avec le processus simulé cc +jets et celui obtenu avec $\text{direct } c\bar{c}$. En haut, α est obtenu avec la combinaison des deux processus simulés (points noir) ajusté par une constante ($\varepsilon_{cl}^{\text{SLTNN}}$ dans le lot enrichi en rouge ou dans le lot muonique en bleu). Notons que $\varepsilon_{cl}^{\text{SLTNN}}$ et dans le lot muonique et enrichi sont pratiquement superposés. Les erreurs sont uniquement statistiques. En faisant la somme quadratique des erreurs du coefficient de corrélation de la figure du haut et du bas, on détermine l'erreur reportée dans les tableaux 3.10 et 3.11.

B.2 Variables d'entrées de SLTNN dans les lots de bruit de fond simulés

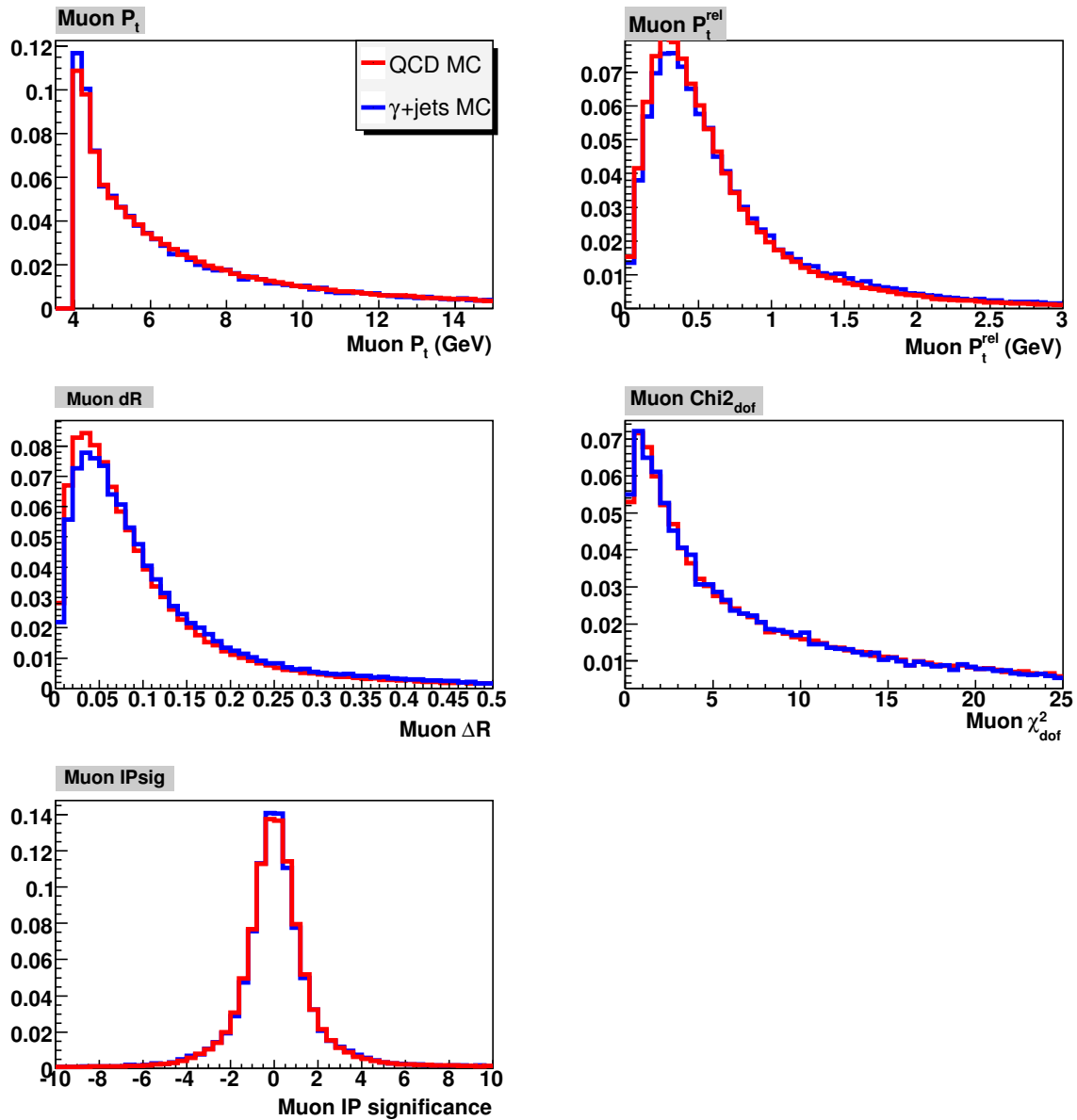


FIG. B.6: Variables d'entrée du réseau de neurones relatives au muon de qualité *med* pour les lots simulés de bruit de fond QCD $q\bar{q}$ (rouge) et γ +jets (bleu). De haut en bas et de gauche à droite : p_T du muon, p_T^{rel} du muon, $\Delta R(\text{muon}, \text{jet})$, χ^2_{ddl} du muon et IP_{sig} du muon. Ces distributions sont normalisées à l'unité.

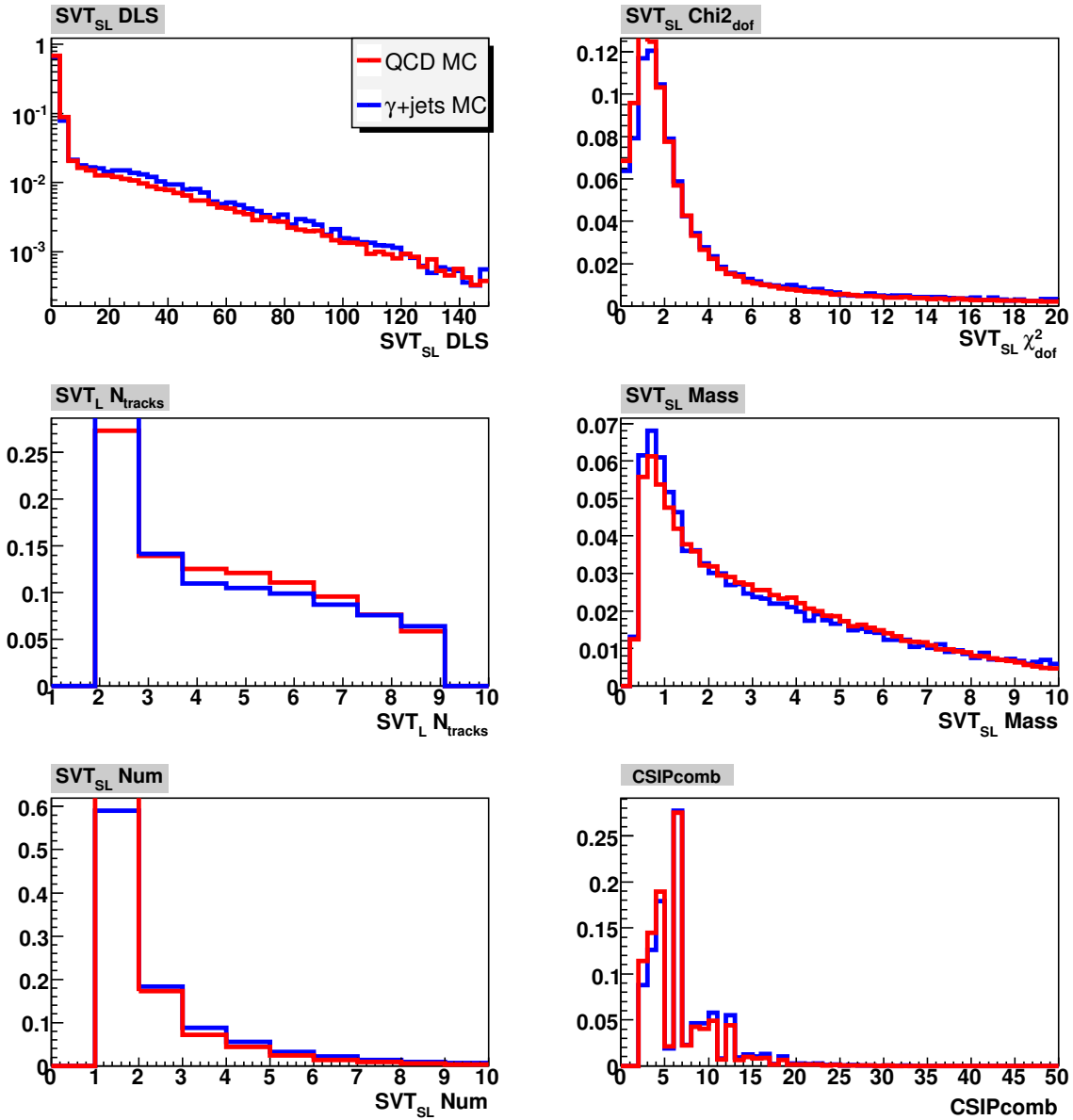


FIG. B.7: Variables d'entrée du réseau de neurones relatives au vertex secondaire pour les lots simulés de bruit de fond QCD $q\bar{q}$ (rouge) et γ +jets (bleu) pour un muon de qualité *med*. De haut en bas et de gauche à droite : $SVT_{SL}DLS$, $SVT_{SL}\chi^2_{dof}$, $SVT_{SL}N_{Tracks}$, $SVT_{SL}Mass$, $SVT_{SL}Num$ et CSIPcomb. Ces distributions sont normalisées à l'unité.

B.3 Variables d'entrées de SLTNN dans les lots de données et de signal simulé

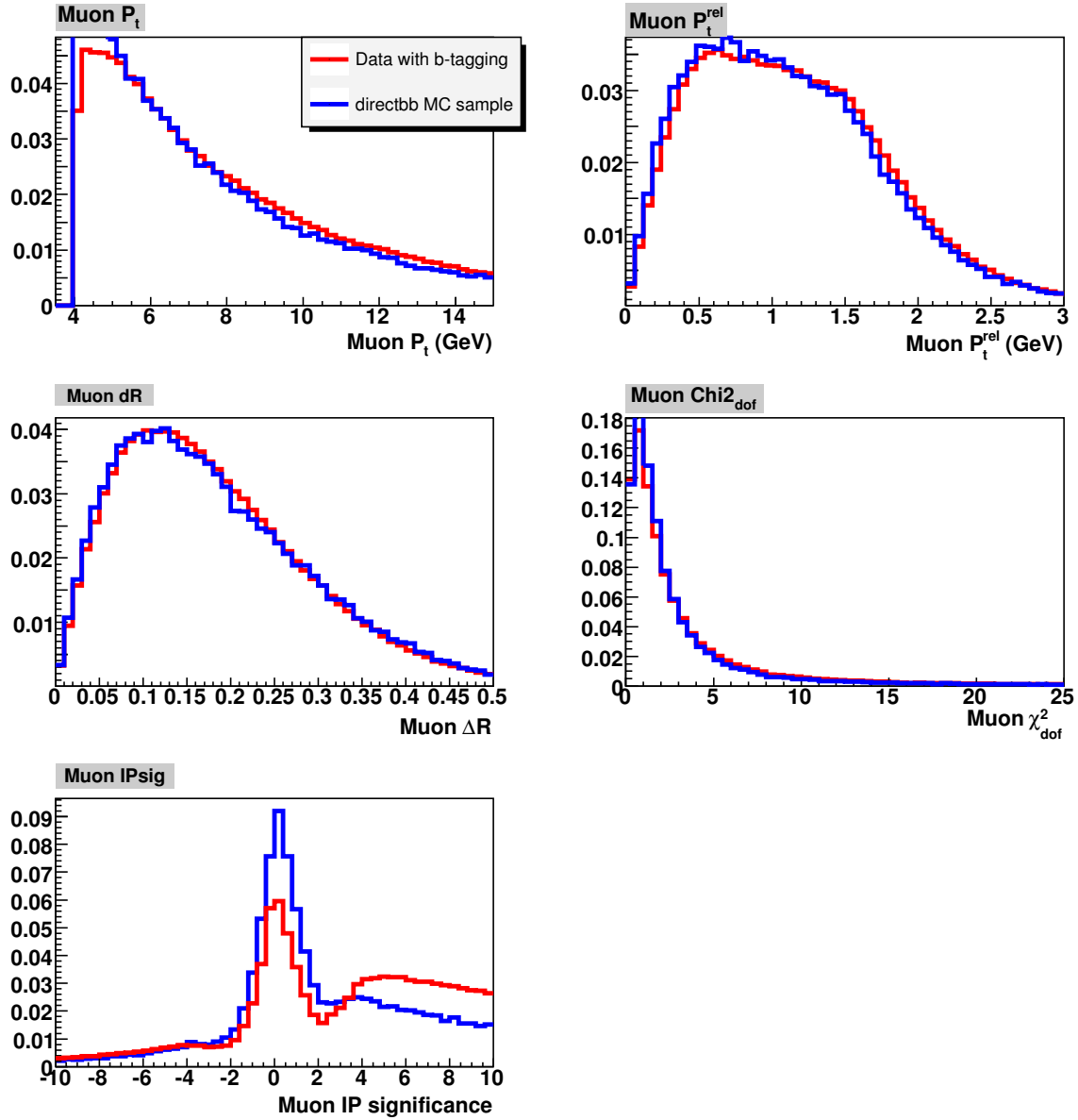


FIG. B.8: Variables d'entrée du réseau de neurones relatives au muon de qualité *med* pour les lots de données $\text{MU}_{\text{inclusive}}$ étiqueté par le point de fonctionnement *Tight* de NN (rouge) et le lot simulé $\text{direct } b\bar{b}$ (bleu). De haut en bas et de gauche à droite : p_T du muon, p_T^{rel} du muon, $\Delta R(\text{muon}, \text{jet})$, χ_{dof}^2 du muon et IP_{sig} du muon. Ces distributions sont normalisées à l'unité.

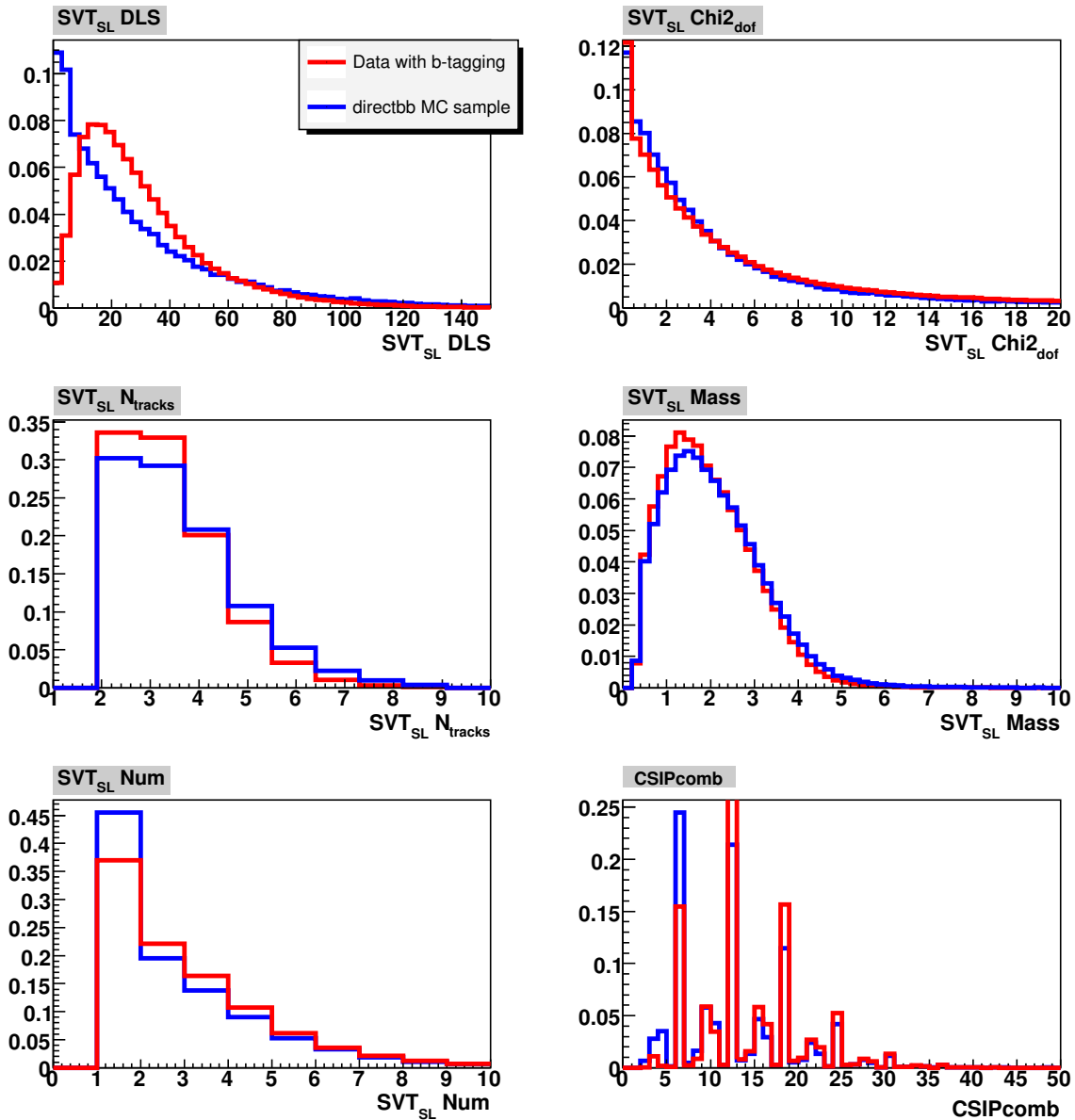


FIG. B.9: Variables d'entrée du réseau de neurones relatives au vertex secondaire pour les lots de données MUinclusive étiqueté par le point de fonctionnement Tight de NN (rouge) et le lot simulé direct $b\bar{b}$ (bleu) pour un muon de qualité *med*. De haut en bas et de gauche à droite : $SVT_{SL}DLS$, $SVT_{SL}\chi^2_{dof}$, $SVT_{SL}N_{Tracks}$, $SVT_{SL}Mass$, $SVT_{SL}Num$ et $CSIPcomb$. Ces distributions sont normalisées à l'unité.

B.4 Variables d'entrées de SLTNN dans les lots de données et de signal plus bruit de fond simulé

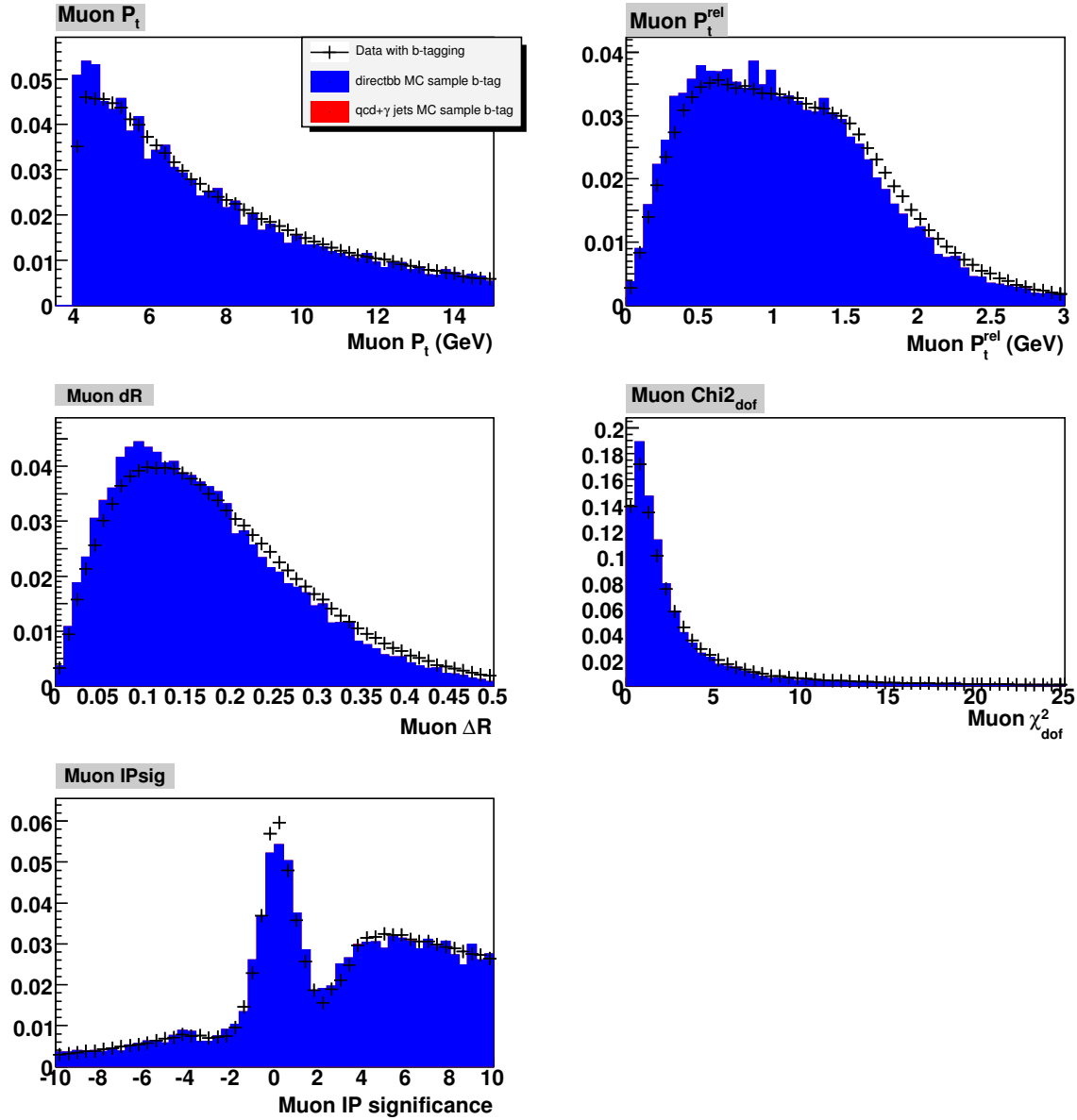


FIG. B.10: Variables d'entrée du réseau de neurones relatives au muon de qualité *med* pour les lots de données MUinclusive (points noirs), le lot simulé direct $b\bar{b}$ (bleu) et les lots simulés combinés QCD $q\bar{q}$ et γ +jets (rouge). Les événements sont étiquetés par le point de fonctionnement Tight de NN. De haut en bas et de gauche à droite : p_T du muon, p_T^{rel} du muon, $\Delta R(\text{muon}, \text{jet})$, χ^2_{dof} du muon et IP_{sig} du muon. Ces distributions sont normalisées à l'unité.

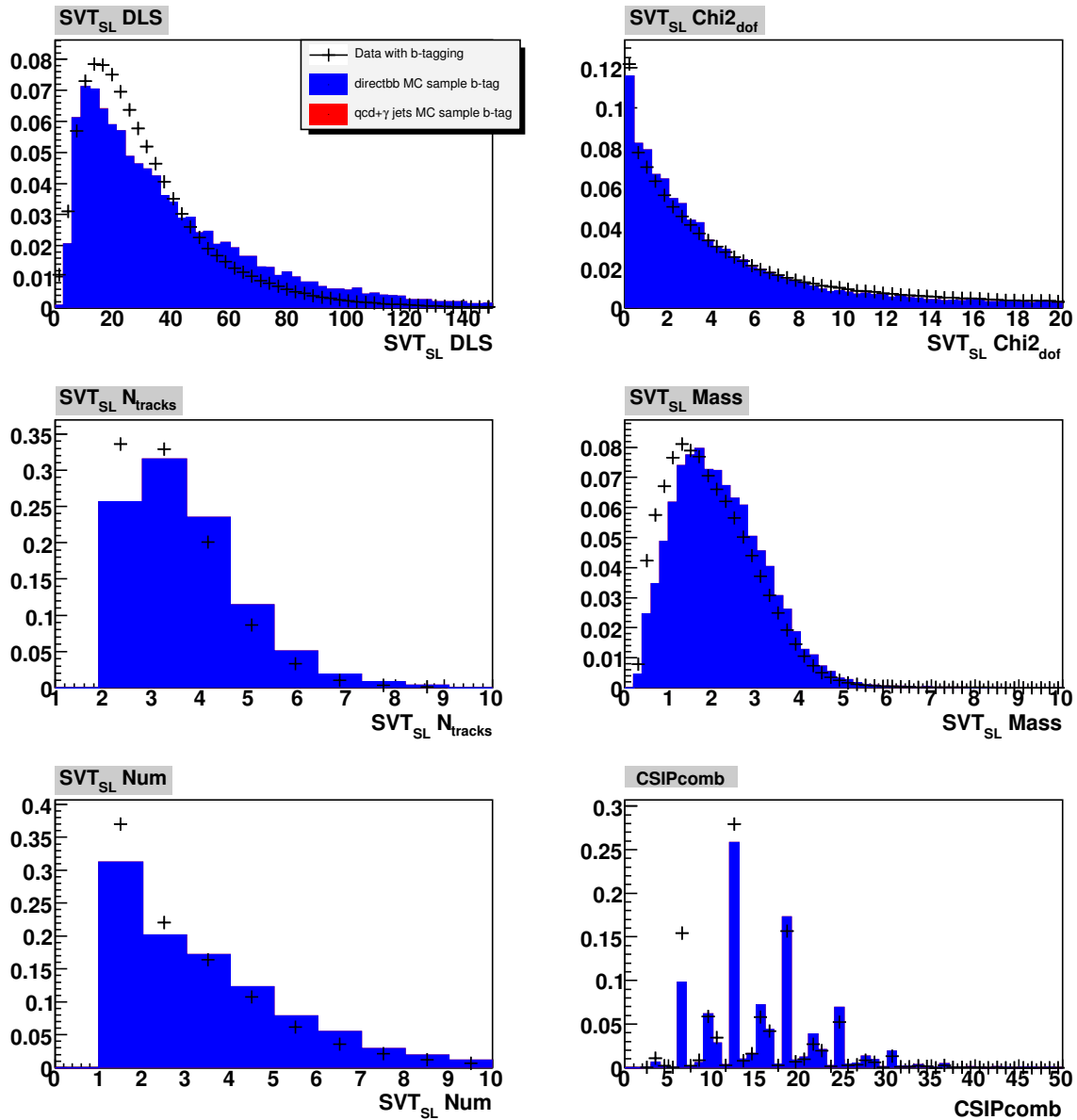


FIG. B.11: Variables d'entrée du réseau de neurones relatives au vertex secondaire pour les lots de données MU inclusive (points noirs), le lot simulé direct $b\bar{b}$ (bleu) et les lots simulés combinés QCD $q\bar{q}$ et γ +jets (rouge) pour un muon de qualité *med*. Les événements sont étiquetés par le point de fonctionnement Tight de NN. De haut en bas et de gauche à droite : $SVT_{SL}DLS$, $SVT_{SL}X_{dd}^2$, $SVT_{SL}N_{Tracks}$, $SVT_{SL}Mass$, $SVT_{SL}Num$ et $CSIPcomb$. Ces distributions sont normalisées à l'unité.

Annexe C

Autres combinaisons d'étiquetage dans l'analyse $ZH \rightarrow \nu\bar{\nu}b\bar{b}$

Les distributions suivantes montrent les résultats obtenus avec la méthode d'étiquetage SLTNN dans l'analyse $ZH \rightarrow \nu\bar{\nu}b\bar{b}$. Cet algorithme utilise trois qualités de muon différentes : *loose*, *med* et *med3*, ainsi que trois points de fonctionnements : *LOOSE*, *MEDIUM* et *TIGHT*. Il existe donc en tout neuf combinaisons possible pour étiqueter un jet. Le tableau suivant résume ces possibilités et leur nomenclature :

qualité de μ	points de fonctionnement		
	LOOSE	MEDIUM	TIGHT
loose	lL	lM	lT
medium	mL	mM	mT
medium3	m3L	m3M	m3T

TAB. C.1: Les neuf possibilités d'étiquetage avec SLTNN.

Cette annexe propose de voir les distributions correspondantes à des événements où l'un des deux jets est étiqueté selon l'un de neuf critères et l'autre ne doit pas être étiqueté lL. Les distributions relatives aux cas lL, lM et lT ont déjà été montrées.

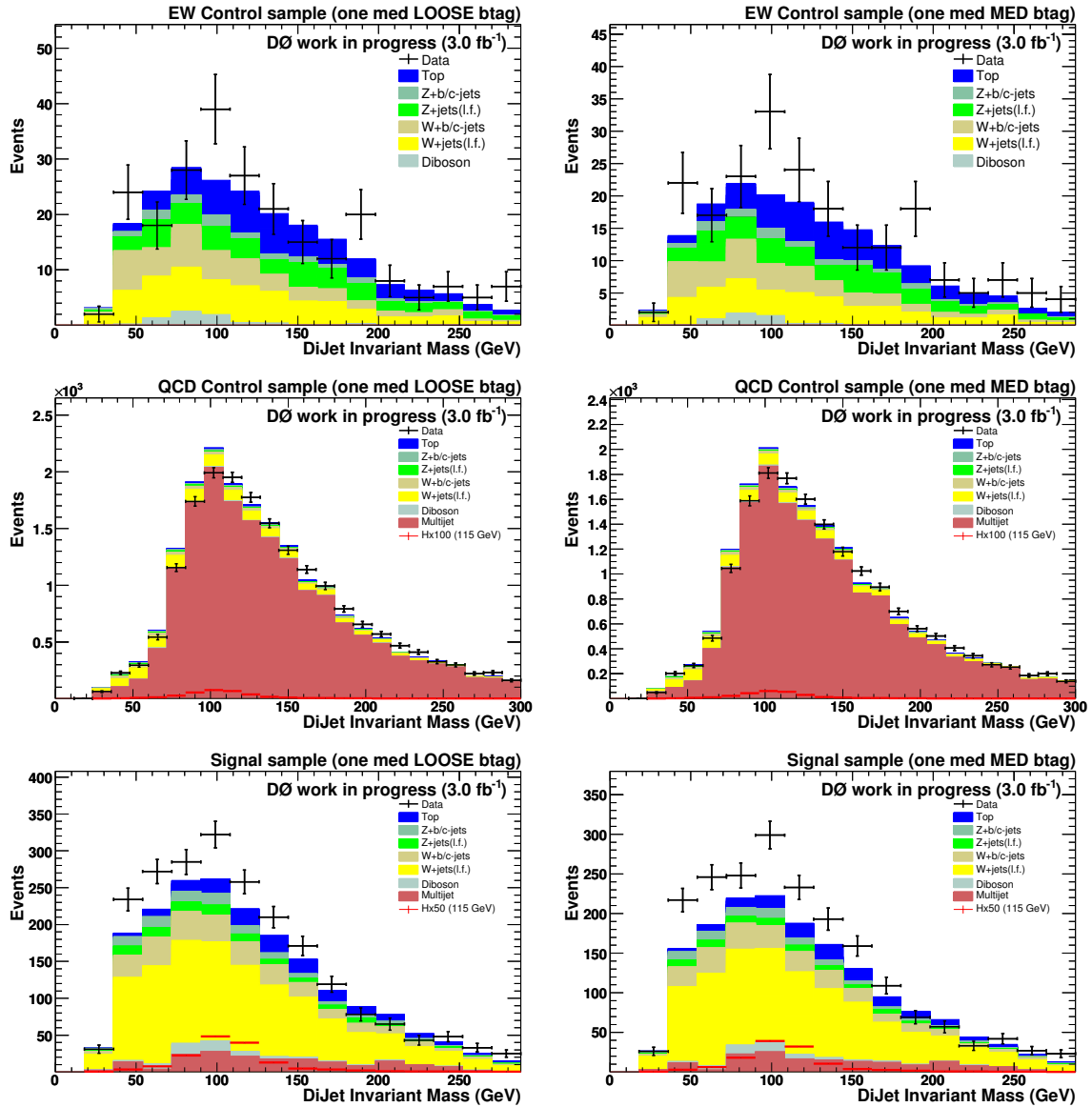


FIG. C.1: Masse invariante des deux jets principaux dans l'échantillon de contrôle électrofaible (haut), dans l'échantillon de contrôle QCD (milieu) et dans l'échantillon de signal (bas) après étiquetage avec SLTNN dans le cas mL (à gauche) et mM (à droite).

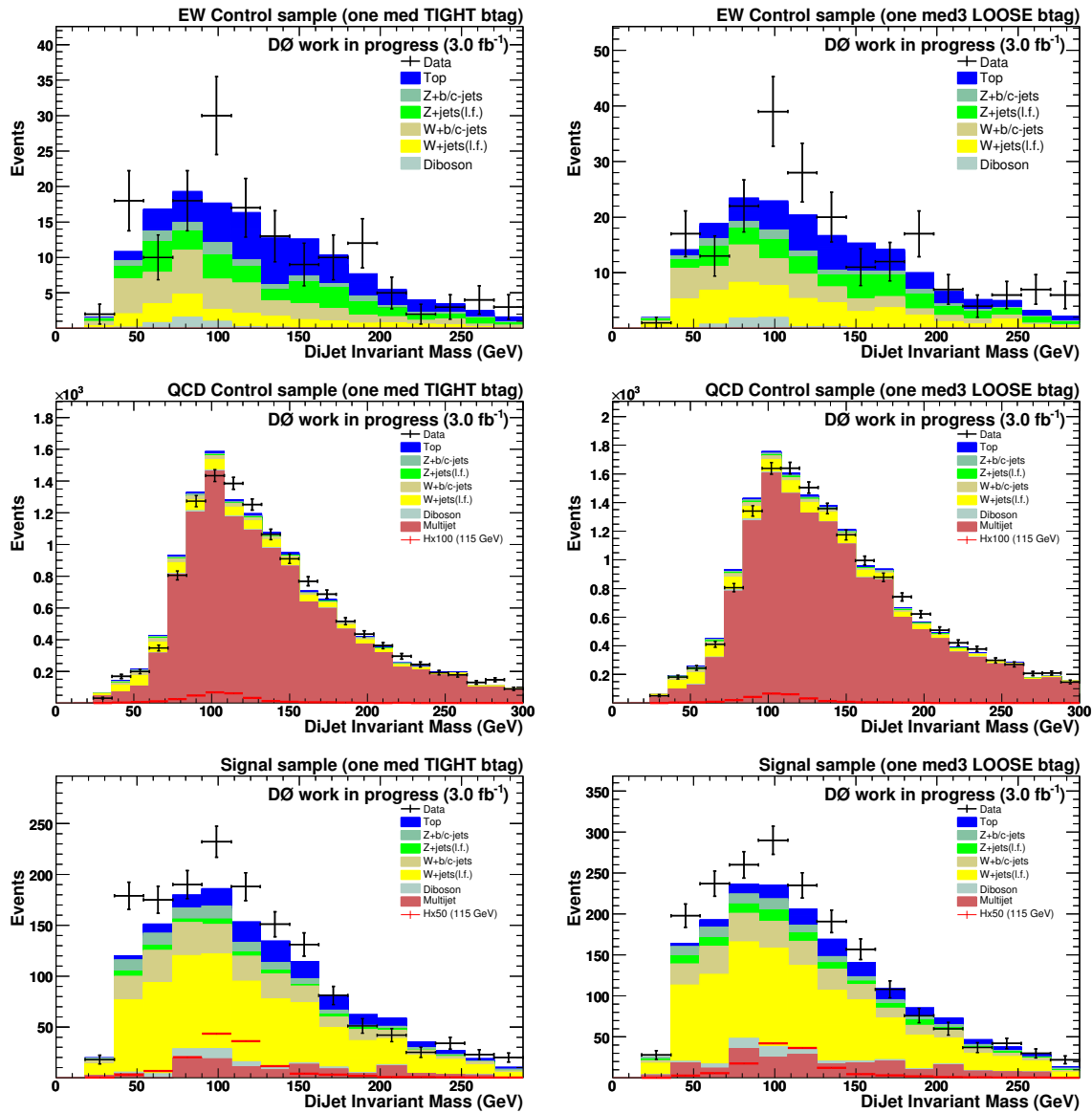


FIG. C.2: Masse invariante des deux jets principaux dans l'échantillon de contrôle électrofaible (haut), dans l'échantillon de contrôle QCD (milieu) et dans l'échantillon de signal (bas) après étiquetage avec SLTNN dans le cas m_T (à gauche) et m_{3L} (à droite).

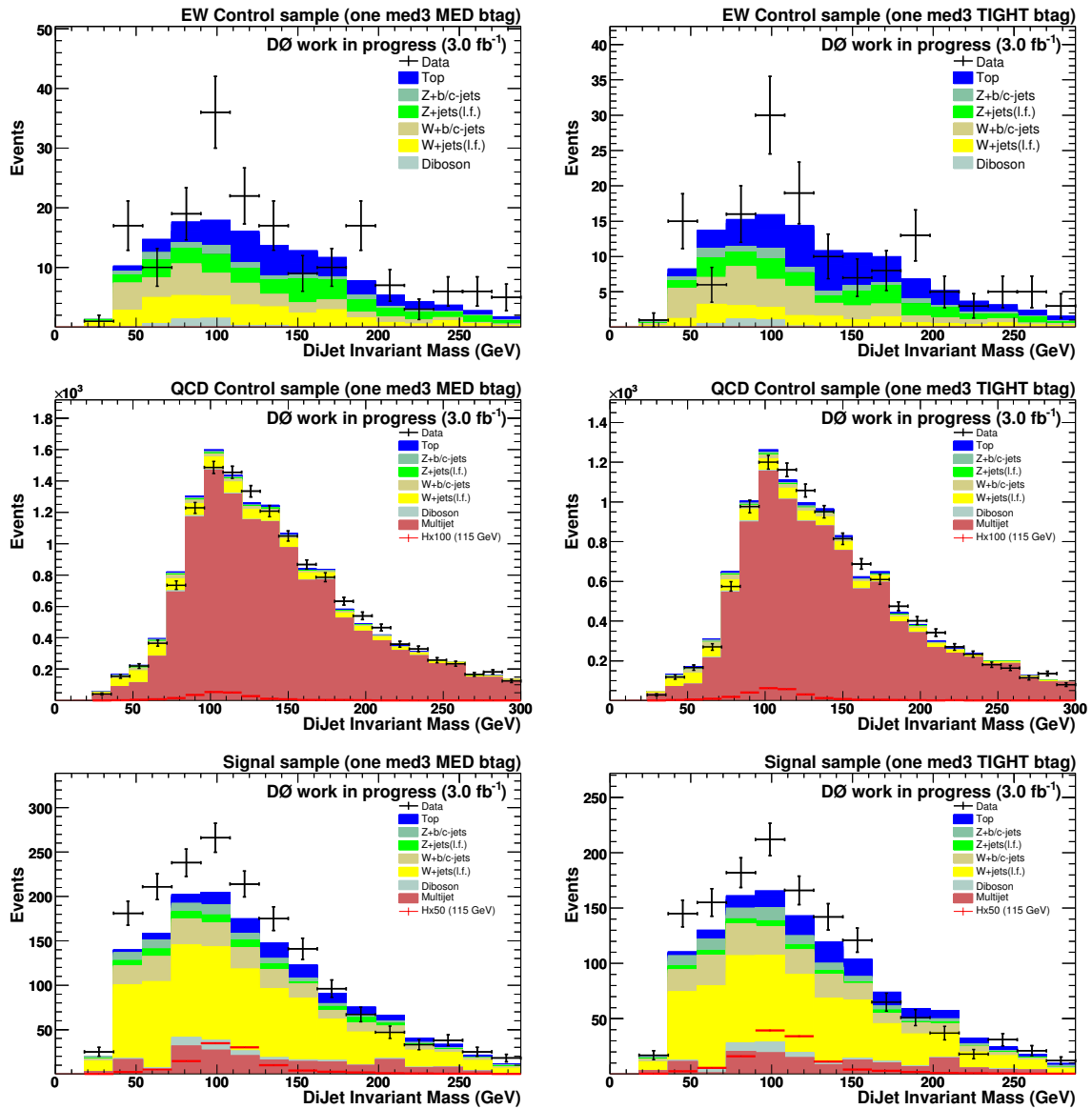


FIG. C.3: Masse invariante des deux jets principaux dans l'échantillon de contrôle électrofaible (haut), dans l'échantillon de contrôle QCD (milieu) et dans l'échantillon de signal (bas) après étiquetage avec SLTNN dans le cas m3M (à gauche) et m3T (à droite).

Distributions des variables d'entrées du SLTNN dans l'analyse $ZH \rightarrow \nu\bar{\nu}b\bar{b}$

D.1 Variables d'entrées de SLTNN dans l'analyse avant *b-tagging*

Cette section présente les distributions des 11 variables d'entrée utilisées pour entraîner SLTNN, pour le jet principal (jet de plus haut p_T) ou secondaire avant *b-tagging*. De haut en bas et de gauche à droite, on trouvera :

- P_T du muon
- P_T relatif du muon
- $\Delta R(\text{muon}, \text{jet})$
- χ^2 du muon
- paramètre d'impact en signficance du muon
- distance d'approche minimale en signficance du vertex secondaire
- χ^2 par degré de liberté du vertex secondaire
- nombre de traces associées au vertex secondaire
- masse du vertex secondaire
- nombre de vertes secondaires
- CSIP comb.

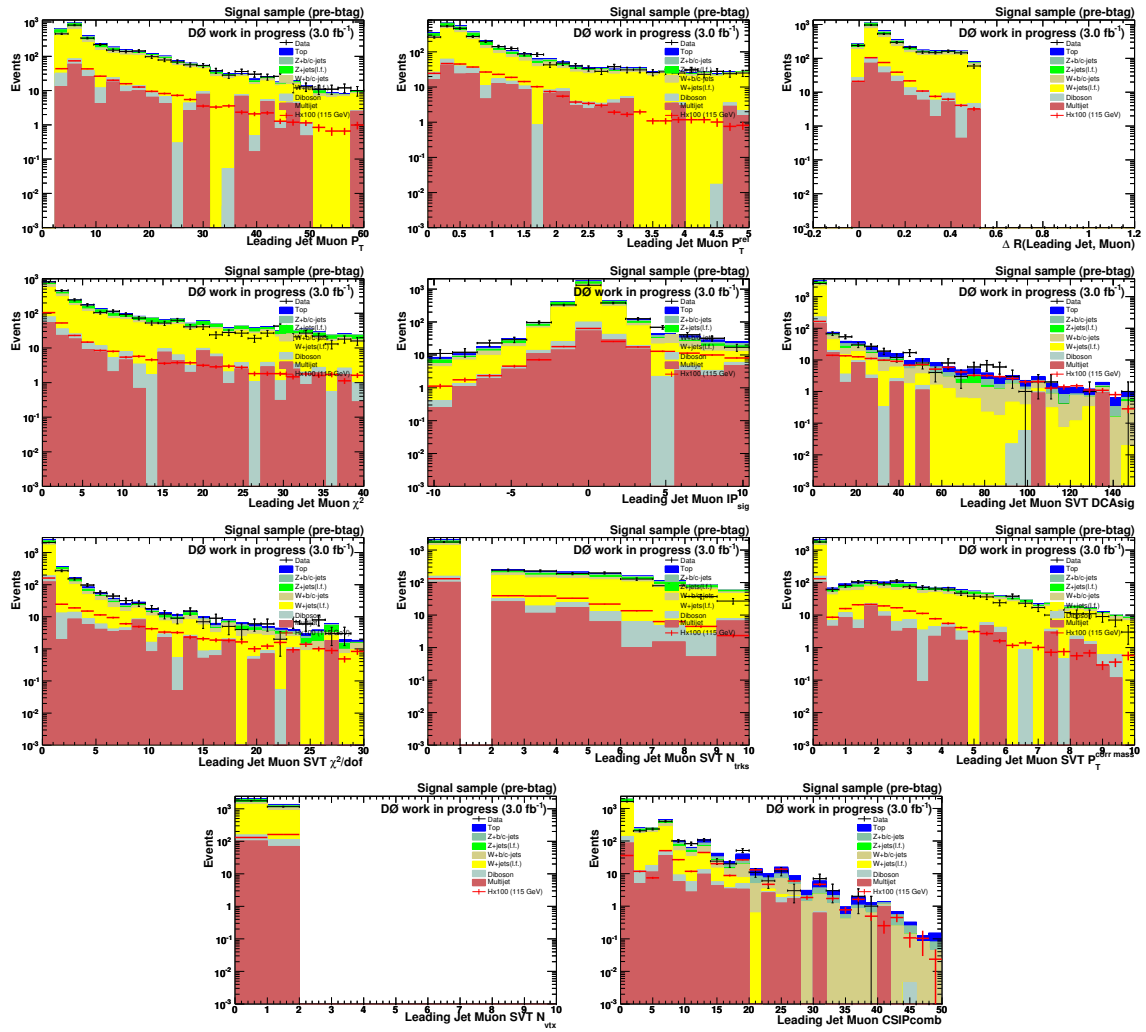


FIG. D.1: Distributions des variables d'entrée du SLTNN avant étiquetage du jet principal dans l'échantillon de signal.

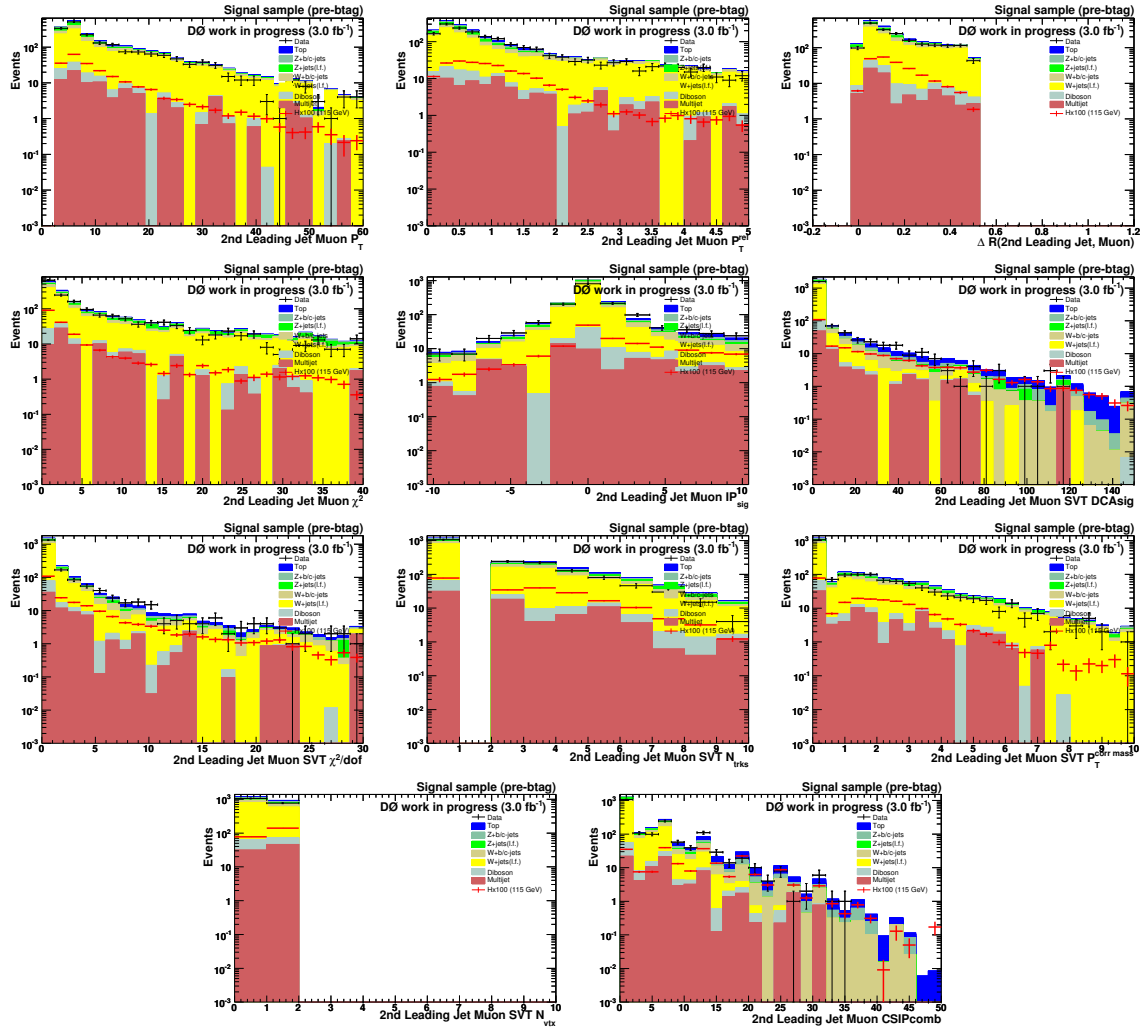


FIG. D.2: Distributions des variables d'entrée du SLTNN avant étiquetage du jet secondaire dans l'échantillon de signal.

D.2 Variables d'entrées de SLTNN dans l'analyse après *b-tagging*

Cette section présente les distributions des 11 variables d'entrée utilisées pour entraîner SLTNN ainsi que certaines variables cinématiques relatives au muon. De haut en bas et de gauche à droite, on trouvera :

- η détecteur du muon
- η physique du muon
- χ^2 du muon (*entrée de SLTNN*)
- CSIP comb. (*entrée de SLTNN*)
- $\Delta R(\text{muon, jet})$ (*entrée de SLTNN*)
- paramètre d'impact en signification du muon (*entrée de SLTNN*)
- P_T du muon (*entrée de SLTNN*)
- P_T relatif du muon (*entrée de SLTNN*)
- χ^2 par degré de liberté du vertex secondaire (*entrée de SLTNN*)
- distance d'approche minimale en signification du vertex secondaire (*entrée de SLTNN*)
- nombre de traces associées au vertex secondaire (*entrée de SLTNN*)
- nombre de vertes secondaires (*entrée de SLTNN*)
- masse du vertex secondaire (*entrée de SLTNN*)

pour le jet principal (jet de plus haut p_T) ou secondaire étiquetés par SLTNN pour un muon de type *med* et un point de fonctionnement *MEDIUM* (i.e. mM).

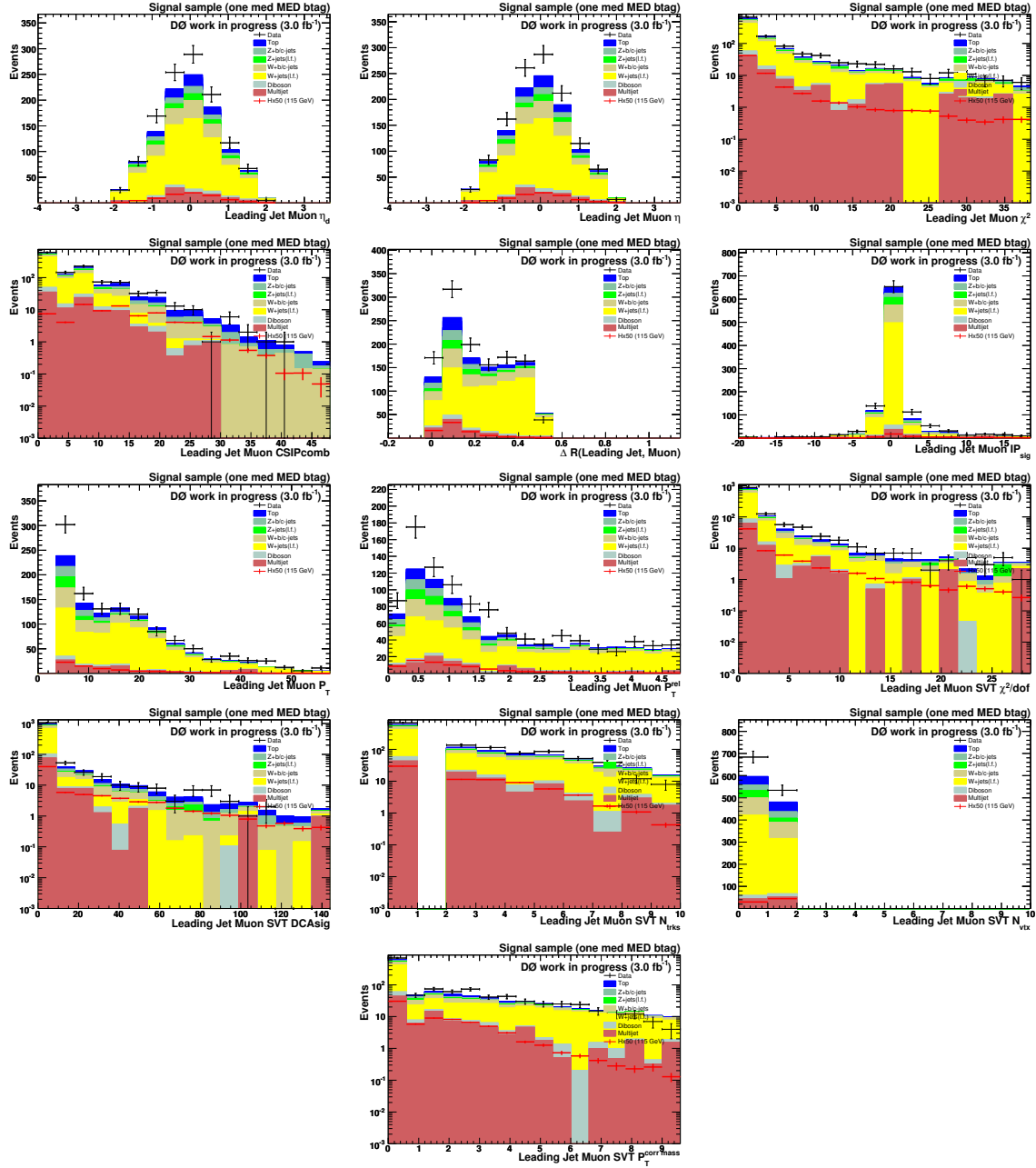


FIG. D.3: Distributions des variables d'entrée du SLTNN dans le cas de l'étiquetage mM du jet principal dans l'échantillon de signal.

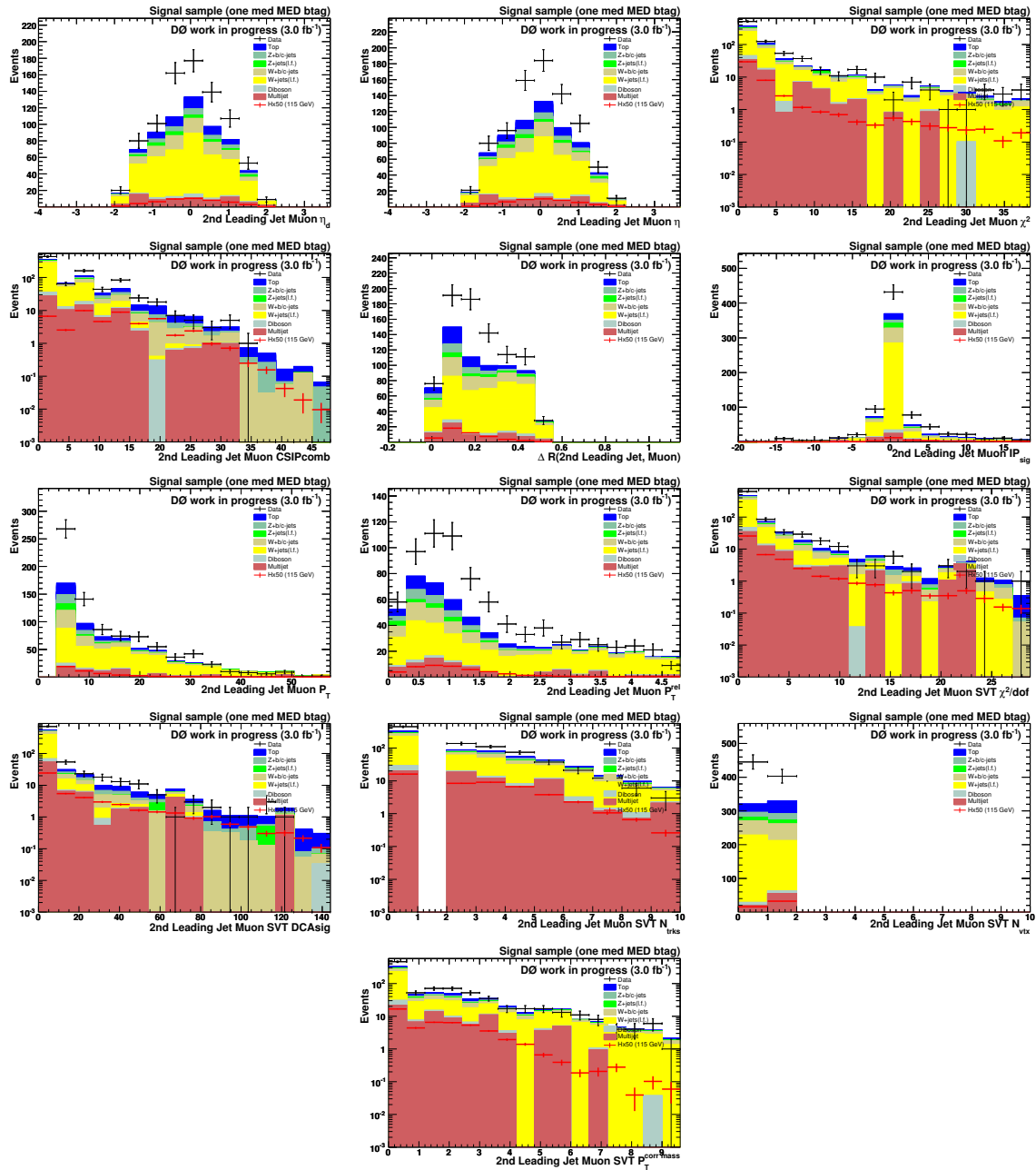


FIG. D.4: Distributions des variables d'entrée du SLTNN dans le cas de l'étiquetage mM du jet secondaire dans l'échantillon de signal.

Bibliographie

- [1] [arXiv :1007.3178v1], "*Combination of CDF and DØ results on the mass of the top quark using up to 5.6 fb^{-1} of data*",
-
- [2] S. Abachi *et al.* [DØ Collaboration], "*Observation of the Top Quark*", Phys. Rev. Lett. 74, 2632 (1995).
- [3] F. Abe *et al.* [CDF Collaboration], "*Observation of the Top Quark Production in $p - \bar{p}$ Collisions with Collider Detector at Fermilab*", Phys. Rev. Lett. 74, 2626 (1995).
- [4] S. Dawson, J.F. Gunion, H.E. Haber, G. Kane, "*The Higgs Hunter's Guide*, Frontiers in Physics, Volume 80.
- [5] P. Higgs, Phys. Rev. Lett. 12, 132 (1964).
- [6] P. Higgs, Phys. Rev. 145, 1156 (1966).
- [7] R. Brout, F. Englert, Phys. Rev. Lett. 13, 321 (1964).
- [8] Guralnik, Hagen and Kibble, Phys. Rev. Lett. 13, 585-587 (1964).
- [9] A. Djouadi, "*The Anatomy of Electro-Weak Symmetry Breaking*", hep-ph/0503172.
- [10] "*Combined CDF and DØ Upper Limits on Standard-Model Higgs-Boson Production*", arXiv :1007.4587 [hep-ex], DØ Conference Note 6096, CDF Conference Note 10241, FERMILAB-CONF-10-257-E, http://tevnpnphwg.fnal.gov/results/SM_Higgs_Summer_10/
- [11] LEP/TEV EW WG Plots for Summer 2010 web page, <http://lepewwg.web.cern.ch/LEPEWWG/plots/summer2010/>
- [12] DØ Collaboration, "*Search for the Standard Model Higgs boson in the $HZ \rightarrow b\bar{b}\nu\bar{\nu}$ channel in 5.2 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$* ", Published 2/18/10 : Phys. Rev. Lett. **104**, 071801 (2010), arXiv.org :0912.5285

- [13] DØ Collaboration, "Search for the Standard Model Higgs boson in the $HZ \rightarrow b\bar{b}\nu\bar{\nu}$ channel in 6.4 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$ ", DØ Conference Note 6087 (2010).
- [14] DØ Collaboration, "A Search for $ZH \rightarrow l^+l^-b\bar{b}$ Production in 6.2 fb^{-1} of data with the DØ detector in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96 \text{ TeV}$ ", DØ Conference Note 6089 (2010).
- [15] DØ Collaboration, "Search for WH associated production with 5.3 fb^{-1} of Tevatron data", DØ Conference Note 6092 (2010).
- [16] DØ Collaboration, "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)
- [17] CDF Collaboration, "Search for the Standard Model Higgs boson in the \cancel{E}_T plus jets sample", CDF Note 10212 (2010).
- [18] CDF Collaboration, "Combined Upper Limit on Standard Model Higgs Boson Production for ICHEP 2010", CDF Conference Note 10241 (2010).
- [19] DØ Collaboration, "Combined Upper Limits on Standard Model Higgs Boson Production from the D0 Experiment in up to 6.7 fb^{-1} of data", DØ Conference Note 6094 (2010).
- [20] Expected Performance of the ATLAS Experiment Detector, Trigger and Physics, <http://arxiv.org/abs/0901.0512>
- [21] ATLAS Sensitivity Prospects for Higgs Boson Production at the LHC Running at 7 TeV, ATL-PHYS-PUB-2010-009, <https://twiki.cern.ch/twiki/bin/view/Atlas/PublicHGPlotsAt1PhysPub2010009>
-
- [22] <http://www-bdnew.fnal.gov/tevatron/>
- [23] John D. Cockroft and Ernest T.S. Walton, Proc. Roy. Soc. (London), A129, 477 (1930) et Ernest T.S. Walton, "The Artificial Production of fast Particles", Nobel Lectures, Physics 1942-1942, Elsevier Publishing Company, Amsterdam (1964).
- [24] <http://linac.fnal.gov/>
- [25] <http://www-ad.fnal.gov/proton/booster/booster.html>
- [26] <http://www-fmi.fnal.gov/>
- [27] <http://www-bd.fnal.gov/public/maininj.html>
- [28] <http://www-bd.fnal.gov/public/antiproton.html>
- [29] V. Abazov *et al.*, DØ Collaboration, "The Upgraded DØ Detector", hep/0507191, accepté par Nucl. Instrum. Methods, 2005.

- [30] R. Angstadt *et al.*, DØ Collaboration, "*The Layer 0 Inner Silicon Detector of the DØ Experiment*", Instrumentation and Detectors, soumis Nov 2009, arXiv :0911.2522v1 [physics.ins-det].
- [31] "*DØ Silicon Tracker Technical Design Report*", http://d0server1.fnal.gov/project/Silicon/www/tdr_final.ps
- [32] "*The DØ Upgrade Central Fiber Tracker*", Technical Design Report, http://d0server1.fnal.gov/users/stefan/www/CFT_TDR/CFT_TDR-.ps
- [33] M. Adams *et al.*, "*Technical Design Report for the Central Preshower*", DØ Note 3014.
- [34] A. Gordeev *et al.*, "*Technical Design Report for the Forward Preshower*", DØ Note 3445.
- [35] S. Abachi *et al.*, *The DØ Detector*, Nucl. Instr. Methods Phys. Res. A 338 (1994) 185.
- [36] B. Baldin *et al.*, "*Technical Design Report for the Central Muon System*", DØ Note 3365.
- [37] G. Alexeev *et al.*, "*Technical Design Report for the Forward Muon System*", DØ Note 3366.
- [38] C. Miao, "*The DØ RunII Luminosity Monitor*", Fermilab-Conf-98/395-E.
- [39] A. Schwartzman and M. Narain, "*Probabilistic Primary Vertex Selection*", DØ Note 4042, November 2002.
- [40] A. Pierce, "*A Comparison of Cone and KT Algorithms for the Reconstruction of a Higgs Boson at 110 GeV*", DØ Note 3486.
- [41] E. Busato et B. Andrieu, "*Jet Algorithms in the DØ Run II Software : Description and User's Guide*", DØ Note 4457.
- [42] G. Bernardi *et al.*, "*Jets in RunII*", 2002, The JETMET group, DØ Note 3287.
- [43] B. Abbott *et al.*, "*Jet Energy Scale at DØ*", DØ Note 3287.
- [44] <http://www-d0.fnal.gov/~daniel/jesgroup/jesgroup.html>
- [45] J. Lellouch, "*Corrections for muonic b-jets in P17*", DØ Note 5486.
- [46] Muon ID Group P20 Preliminary Certification :
http://www-d0.fnal.gov/computing/algorithms/muon/p20/p20_prelim_certif.html.
- [47] S. Calvet *et al.*, "*Towards MissingET Certification and Unclustered Energy Studies*", DØ Note 4927
- [48] M.L. Mangano *et al.*, "*ALPGEN, a generator for hard mutliparton processes in hadronic collisions*", JHEP, hep-ph/0206293, 2003.
- [49] T. Sjöstrand, L. Lonnbladn, S. Mrenna et P. Skands, "*PYTHIA 6.3 : Physics and manual*", hep-ph/0308153, 2003.

- [50] Page web du générateur COMPHEP, <http://comphep.sinp.msu.ru>
- [51] R. Brun and F. Carminati, CERN Program library Long Writeup Report W5013,1993 (unpublished).
-
- [52] T. Scanlon, "A Neural Network b -tagging Tool", DØ Note 4889.
- [53] B. Clément, D. Bloch, D. Gelé, S. Greder, I. Ripp-Baudot, "Performance of the JLIP b -tagger in $p14$ ", DØ Note 4348, March 2004.
- [54] R. Demina, A. Khanov, F. Rizatdinova, " b -tagging with Counting Signed Impact Parameter method", DØ Note 4049.
- [55] J. Clutter, "Performance of the SVT b -tagging Algorithm on P17 data", DØ Note 5385.
- [56] M. Anastasoie, S. Robinson, T. Scanlon, "Performance of the NN b -tagging Tool on $p17$ Data", DØ Note 5213.
- [57] T. Gadfort, A. Haas, D. Johnston, D. Lincoln, T. Scanlon, S. Schlobohm, "Performance of the Dzero NN b -tagging Tool on $p20$ Data", DØ Note 5554.
- [58] PDG group, "Review of Particle Physics", Journal of Physics G, Vol 33 (2006).
- [59] B. Clement, D. Bloch, D. Gele, S. Greder, A.C. Le Bilhan, I. Ripp-Baudot, "SystemD or how to get signal, backgrounds and their efficiencies with real data", DØ Note 4159.
- [60] B. Clement "Production électrofaible du quark top au Run II", Thèse de doctorat, 2006.
- [61] TMinuit,
<http://root.cern.ch/root/html/TMinuit.html>.
- [62] A. Duperrin, D. Jamin and H. Yoo, "A Soft Muon Tagger with Neural Network on $p20$ Data", DØ Note 5647.
-
- [63] Th. Millet, A. Duperrin, P. Verdier, S. Muanza, "Jets and \cancel{E}_T triggers for the New Phenomena group in the $v14$ and $v15$ trigger lists", DØ Note 5120.
- [64] F. Lacroix, C. Ochando, A. Duperrin, P. Verdier, "Higgs and New Phenomena Jets+ \cancel{E}_T triggers : L3 design and commissioning in $v15$ Run IIb trigger lists", DØ Note 5282.
- [65] J.M. Campbell and R.K. Ellis, Phys. Rev. D **60**, 113006 (1999).
- [66] M. Cacciari *et al.*, JHEP **404**, 068 (2004) ; N. Kidonakis and R. Vogt, Phys. Rev. D **68**, 114014 (2003) ; N. Kidonakis, Phys. Rev. D **74**, 114012 (2006).
- [67] N. Makovec and J-F. Grivaz, "The Relative Data - Monte Carlo Jet Energy Scale", DØ Note 4807.

- [68] N. Makovec and J-F. Grivaz, "*Shifting, Smearing and Removing Simulated Jets*", DØ Note 4914.
- [69] <https://plone4.fnal.gov/P1/DØWiki/comp/caf/caffa/LumiReWeigh>
- [70] M. Shamim and T. Bolton, "*Generator Level Reweighting of p_T of Z Boson*", DØ Note 5565.
- [71] M. Rangel, "*Vertex Confirmation SF for Jets*", CALGO Meeting, May 20, 2009, <http://www-d0.hef.kun.nl/fullAgenda.php?ida=a09823>.
- [72] S. Catani *et al.*, JHEP **0307**, 028 (2003).
- [73] A. Schwartzman "*Missing Et Significance Algorithm in RunII data*", DØ Note 4254.
- [74] S. Calvet, T. Christoudias, T. Dorland, A. Dubey, G. Facini, J-F. Grivaz, G. Hesketh, R.K. Mommsen, C. Ochando, K. Peters, "*Search for the Standard Model Higgs boson in the $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ channel in 4.0 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$* ", DØ Note 5857, http://www-d0.fnal.gov/Run2Physics/d0_private/eb/Run2EB_029/zhvbb_m09_v03.pdf.
- [75] L. Breiman *et al.*, "*Classification and Regression Trees*", Wadsworth (1984).
- [76] Y. Coadou [CDF and DØ Collaborations], "*Uses of multivariate analysis methods*", 847 **PoS TOP2006**, 016 (2006)
- [77] T. Junk, Nucl. Instrum. Methods in Phys. Res. A **434**, 435 (1999) ; A. Read, in "*1st Workshop on Confidence Limits*", CERN Report No. CERN-2000-005, 2000.
- [78] W. Fisher, "*Systematics and limit calculations*", FERMILAB-TM-2386-E, 2006, DØ Note 5309.
- [79] W. Fisher, "*Calculating limits for combined analysis*", 2006, DØ Note 4975.
- [80] W. Fisher, "*COLLIE : a confidence level limit evaluator*", 2008, DØ Note 5395.
- [81] A. Duperrin and D. Jamin, "*b-Tagging with muons (SLTNN) in $ZH \rightarrow \nu\bar{\nu}b\bar{b}$* ", DØ Note 6044.
-
- [82] R. Fletcher, "*Practical Methods of Optimization*", second edition, Wiley (1987).
- [83] L. Breiman *et al.*, "*Classification and Regression Trees*," Wadsworth (1984).
- [84] Y. Freund and R.E. Schapire, "*Experiments with a New Boosting Algorithm*," in Machine Learning : Proceedings of the Thirteenth International Conference, pp. 148-156 (1996).

Résumé

Dans le Modèle Standard de la physique des particules, le boson de Higgs permet de générer la masse des particules élémentaires. Les contraintes théoriques et expérimentales actuelles imposent au boson de Higgs d'avoir une masse comprise entre 114.4 et 158 GeV à 95% de niveau de confiance. De plus, le Tevatron a récemment exclu la zone de masse entre 100 et 109 GeV, 158 et 175 GeV à 95% de niveau de confiance. Ces résultats confortent la recherche du boson de Higgs à basse masse dans la fenêtre encore ouverte.

Le détecteur $D\bar{O}$ est situé près de Chicago, au Tevatron, collisionneur protons-antiprotons avec une énergie dans le centre de masse de 1.96 TeV. Le sujet de cette thèse est la recherche du boson de Higgs produit en association avec un boson Z. C'est un canal sensible au boson de Higgs de basse masse (<135 GeV) qui a un rapport de branchement $H \rightarrow b\bar{b}$ de l'ordre de varie entre 50% et 90% dans cette région en masse. Le canal d'étude $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ a un état final composé de 2 jets de saveurs lourdes et de l'énergie transverse manquante emportée par les neutrinos.

L'identification des jets de saveur lourde ("b-tagging") est réalisée à l'aide d'un nouvel algorithme que l'on a développé (SLTNN) : la méthode est basée sur la désintégration semi-leptonique des quarks b. L'analyse de recherche du boson de Higgs a été menée avec 3 fb^{-1} de données. L'utilisation de SLTNN a permis d'améliorer de 10% l'efficacité d'identification de boson de Higgs. En revanche, la sensibilité globale de l'analyse, une fois les bruits de fond et erreurs systématiques prises en compte, est très peu améliorée ($<1\%$).

Mots clés : $D\bar{O}$ / Tevatron / Modèle Standard / Boson de Higgs / Etiquetage des jets de saveur lourde / Muons de basse impulsion transverse

Abstract

In the Standard Model of particle physics, the Higgs boson generates elementary particle masses. Current theoretical and experimental constraints lead to a Higgs boson mass between 114.4 and 158 GeV with 95% confidence level. Moreover, Tevatron has recently excluded the mass ranges between 100 and 109 GeV, 158 and 175 GeV with 95% confidence level. These results give a clear indication to search for a Higgs boson at low mass.

The $D\bar{O}$ detector is located near Chicago, at the Tevatron, a proton-antiproton collider with an energy in the center of mass of 1.96 TeV. The topic of this thesis is the search for a Higgs boson in association with a Z boson. This channel is sensitive to low mass Higgs boson (<135 GeV) which has a branching ratio $H \rightarrow b\bar{b}$ varies between 50% and 90% in this mass range. The decay channel $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ studied has in the final state 2 heavy-flavor jets and some missing transverse energy due to escaping neutrinos. The heavy-flavor jets identification ("b-tagging") is done with a new algorithm (SLTNN) developed specifically for semi-leptonic decay of b quarks.

The Higgs boson search analysis was performed with 3 fb^{-1} of data. The use of SLTNN increases by 10% the Higgs boson signal efficiency. The global analysis sensitivity improvement, however, is rather low ($<1\%$) after taking into account the backgrounds and systematic uncertainties.

Keywords : $D\bar{O}$ / Tevatron / Standard Model / Higgs boson / b-tagging / Soft muons