

Table des matières

1	Le modèle standard et le processus $t\bar{t}H$ au LHC	7
1.1	Le modèle standard	8
1.1.1	L'interaction forte	8
1.1.2	Symétrie et brisure électrofaible	9
1.1.3	Les limites du modèle standard	11
1.2	Production de paire de quark top	11
1.3	Le boson de Higgs	12
1.3.1	Production au LHC	12
1.3.2	Modes de désintégration	13
1.4	De la théorie à la génération d'événements	14
2	Le contexte expérimental : le LHC et le détecteur ATLAS	19
2.1	Le collisionneur LHC et ses performances au Run 2	20
2.2	Le détecteur ATLAS	22
2.2.1	Le Détecteur Interne	24
2.2.2	Calorimètre	27
2.2.3	Spectromètre à muon	33
2.2.4	Système de déclenchement et d'acquisition	34
3	La reconstruction des objets et la simulation	38
3.1	Trace	39
3.2	Vertex primaire	40
3.3	Candidat électron	42
3.4	Candidat muon	45
3.5	Jet	47
3.6	Grand jet	48
3.7	Simulation du détecteur	50
4	Étiquetage des jets de quark <i>beauté</i>	54
4.1	Introduction	55
4.2	Les algorithmes d'étiquetage des jets de b	57
4.2.1	L'algorithme IPxD et les paramètres d'impact des traces	57

4.2.2	L'algorithme JetFitter	62
4.2.3	L'algorithme SV1	64
4.2.4	Le classifieur $MV2$	66
4.2.5	Candidat b	70
4.3	Etude de la sélection des traces de JF à haute impulsion	70
4.4	Investigation de nouvelles variables de JetFitter	76
4.4.1	Variables angulaires de JetFitter	76
4.4.2	Masse SV de JetFitter	79
4.4.3	Amélioration apportée par les nouvelles variables	82
4.5	Etude de l'entraînement de l'arbre de décision $MV2$	86
4.5.1	Redéfinition de l'échantillon d'entraînement du BDT $MV2$	86
4.5.2	Optimisation du nouvel entraînement du BDT $MV2$	91
4.5.3	Conclusion	93
5	Recherche du processus $t\bar{t}H$ de haute impulsion dans le canal $b\bar{b}$	97
5.1	Motivation du canal $t\bar{t}H(b\bar{b})$ boosté	98
5.2	Modélisation du signal et des bruits de fond	99
5.2.1	Le signal $t\bar{t}H$	99
5.2.2	Le bruit de fond $t\bar{t}$	99
5.2.3	Les autres bruits de fond	100
5.3	Sélection d'objets d'analyse et d'événements	101
5.3.1	Les candidats électron	101
5.3.2	Les candidats muon	101
5.3.3	Les jets	101
5.3.4	Les candidats b	102
5.3.5	Les grands jets	102
5.3.6	Candidat <i>top lâche</i>	102
5.4	Etude du canal <i>top</i> boosté	103
5.4.1	Sélection <i>3211</i> et <i>3111</i>	103
5.4.2	Suppression du recouvrement candidat électron/grand jet	103
5.4.3	Etiquetage du quark <i>top</i> boosté	105
5.5	Etude du canal <i>top</i> et Higgs boostés	116
5.5.1	Stratégie et sélection d'événements	116
5.5.2	Pureté des candidats <i>top</i> et Higgs boostés	118
5.5.3	Classification des événements boostés	121
5.6	Incertitudes de l'analyse boostée	130
5.6.1	Modélisation du signal et du bruit de fond	130
5.6.2	Objets reconstruits	132
5.6.3	Luminosité	134
5.7	Sensibilité du canal <i>top</i> et Higgs boostés dans l'analyse $t\bar{t}H(b\bar{b})$	134
5.7.1	Régions de contrôle	134

5.7.2	Limite et force du signal avec 36.5/fb	134
5.7.3	Résultat incluant les incertitudes systématiques	139
5.7.4	Perspectives à haute luminosité	141
5.7.5	Conclusion	143



Introduction

La nature est gouvernée par un ensemble de quatre forces fondamentales : la gravitation, l'électromagnétisme, l'interaction faible et l'interaction forte. Ces trois dernières forces se manifestent à petite échelle comme des interactions entre particules élémentaires, dont la description est donnée par le modèle standard. En 2012, les collaborations ATLAS et CMS observent indépendamment une nouvelle particule de masse d'environ 125 GeV produite lors des collisions de protons. Les collisions sont fournies par le collisionneur LHC à une énergie dans le centre de masse de $\sqrt{s} = 8$ TeV jamais égalée auparavant. Cette observation marque une consécration pour la physique des particules, aussi bien sur le plan théorique que sur le plan expérimental. D'un point de vue théorique, parce que la découverte est interprétée comme la production d'une particule élémentaire du modèle standard appelée le boson de Higgs, dont l'existence a été postulée dans les années 1960 par les théoriciens Brout, Englert et Higgs pour expliquer l'origine de la masse des bosons de jauge. D'un point de vue expérimental, parce que cette observation témoigne du niveau d'excellence atteint par les expériences du LHC, étant donnée l'extrême rareté des collisions de protons produisant des bosons de Higgs et la grande difficulté à les détecter.

Actuellement, les activités autour du boson de Higgs se concentrent sur l'étude de ses propriétés afin de déterminer leur degré de compatibilité avec celles prédites par le modèle standard. L'étude réalisée dans cette thèse s'inscrit dans cette perspective, en se focalisant sur l'étude du couplage du boson de Higgs avec le quark *top*. Le quark *top*, fermion le plus lourd du modèle standard, interagit plus fortement avec le boson de Higgs que tout autre fermion. Cette caractéristique intrigue les physiciens et semble indiquer que le quark *top* joue un rôle prédominant à très hautes échelles d'énergie, supérieures au TeV. L'observation de l'interaction du quark *top* avec le boson de Higgs pourrait faire intervenir des phénomènes qui ne sont pas prédits par le modèle standard, potentiellement observables au LHC. Pour étudier cette interaction, les collisions de protons produisant une paire de quarks *top*/anti-*top* en association avec un boson de Higgs (dits événements $t\bar{t}H$), qui font intervenir un couplage entre boson de Higgs et quark *top*, sont analysées par les expériences ATLAS et CMS.

Au cours de cette thèse, une analyse d'événements $t\bar{t}H$ a été développée au sein de la collaboration ATLAS spécifiquement pour des topologies *boostées*, pour lesquelles quarks *top* et bosons de Higgs sont émis à haute impulsion. L'objectif de cette analyse est d'exploiter les topologies boostées pour améliorer la recherche du processus $t\bar{t}H$ tel qu'il est prédit par le modèle standard et contraindre de nouveaux modèles. Les événements analysés proviennent de données récoltées au cours des années 2015 et 2016. L'année 2015 a été marquée par la reprise des collisions *pp* à une énergie dans le centre de masse de 13 TeV. Cette étude est accompagnée de travaux réalisés sur l'identification des jets de quark *b*, qui sont le signe d'une production de quarks *b* dans les collisions de protons. Les quarks *b* jouent un rôle primordial dans cette analyse dans la mesure où chacun des quarks *top* se désintègre presque exclusivement en un boson *W* et un quark *b* et où le canal choisi pour la désintégration du boson de Higgs est une paire $b\bar{b}$.

La thèse s'organise de la manière suivante : une brève introduction au modèle standard

de la physique des particule insistant sur le mécanisme de Higgs sera donnée au Chapitre 1, avec une description de la phénoménologie des événements $t\bar{t}H$ au LHC. Le dispositif expérimental sera décrit au Chapitre 2, avec une introduction sur le collisionneur LHC et le détecteur ATLAS. Les objets physiques reconstruits dans le détecteur qui sont utiles pour l'analyse $t\bar{t}H$ boostée seront abordés au Chapitre 3. Au Chapitre 4 seront présentés l'étiquetage de jet de b tel qu'il est réalisé dans la collaboration ATLAS et les travaux effectués sur les algorithmes d'identification des jets de quark b , notamment JetFitter et $MV2$. Enfin, l'étude des topologies boostées d'événements $t\bar{t}H$ sera présentée au Chapitre 5, avec le développement de techniques d'identification de quarks top boostés, ainsi que la mise en place d'une recherche spécifique d'événements présentant à la fois un quark top et un boson de Higgs boostés. Ce chapitre se conclut avec la présentation des résultats sur la sensibilité de cette analyse boostée lorsqu'elle est combinée avec l'analyse standard $t\bar{t}H(b\bar{b})$.

Chapitre 1

Le modèle standard et le processus $t\bar{t}H$ au LHC

Rapport-Gratuit.com

1.1 Le modèle standard

Le modèle standard de la physique des particules est une théorie quantique des champs qui décrit les interactions électromagnétiques, faibles et fortes entre les particules élémentaires. Ce modèle contient des particules de spin 1/2, ou fermions, comme composants élémentaires de la matière :

- six leptons (et leurs antiparticules) organisés en trois générations avec les électrons e , muons μ et taus τ et leur neutrino associé :

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \quad (1.1)$$

- six quarks (et leurs antiparticules) organisés en trois générations :

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} s \\ c \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}. \quad (1.2)$$

Le modèle contient également les bosons de jauge, de spin 1, qui véhiculent les interactions fondamentales :

- le photon γ pour la force électromagnétique ;
- les bosons Z , W^+ et W^- pour la force faible ;
- les huit gluons g pour la force forte.

Enfin, le modèle standard contient le boson de Higgs, de spin 0, dont le champ est responsable de la génération des masses des particules.

Les interactions du modèle standard découlent d'une symétrie de groupe de la forme $SU(3) \times SU(2) \times U(1)$, où $SU(3)$ est le groupe de symétrie lié à l'interaction forte et $SU(2) \times U(1)$ le groupe de symétrie lié à l'interaction électrofaible. La symétrie $SU(2) \times U(1)$ est brisée par le champ de Higgs, générant ainsi la masse des particules élémentaires. Le formalisme de la théorie repose sur un lagrangien qui encode les interactions des champs de particules. Celui-ci peut s'exprimer en une somme de deux lagrangiens : le premier décrit l'interaction forte (cf. Section 1.1.1) et le second l'interaction électrofaible (cf. Section 1.1.2).

1.1.1 L'interaction forte

La chromodynamique quantique (QCD) est la théorie quantique des champs de l'interaction forte. Elle se manifeste entre particules colorées, à savoir les quarks et les gluons. Elle est basée sur le groupe de symétrie locale $SU(3)$ de couleur, qui est une composante du groupe de symétrie du modèle standard $SU(3) \times SU(2) \times U(1)$. Son lagrangien est donné par :

$$\mathcal{L}_{QCD} = i \sum_f \bar{q}_f \gamma^\mu D_\mu q_f - \frac{1}{4} G_{\mu\nu}^i G_i^{\mu\nu} \quad (1.3)$$

où q_f est le champ spineur associé au quark de saveur f . Les quarks sont des triplets de couleur. La dérivée covariante D_μ est définie comme

$$D_\mu = \partial_\mu - ig_s \frac{\lambda^i}{2} G_\mu^i \quad (1.4)$$

où $g_s = \sqrt{4\pi\alpha_s}$, α_s étant la constante de couplage forte, et λ_i les huit générateurs du groupe $SU(3)$. Les champs tensoriels $G_{\mu\nu}$ sont définis par :

$$G_{i,\mu\nu} = \partial_\mu G_{i,\nu} - \partial_\nu G_{i,\mu} - g_s f_{ijk} G_{j,\mu} G_{k,\nu} \quad (1.5)$$

où $G_{i,\nu}$ sont les huit champs de gluon et les f_{ijk} sont les constantes de structure du groupe $SU(3)$ qui satisfont les relations $[\lambda_i, \lambda_j] = i f_{ijk} \lambda^k$.

A basse échelle d'énergie ($Q < \Lambda_{QCD}$), ni les quarks ni les gluons ne sont observés comme des particules libres. Ils se combinent pour former des particules neutres de couleur (singlet de couleur), ou hadrons. L'intensité de la force forte diminuant avec l'échelle d'énergie Q (c'est le phénomène dit de liberté asymptotique), les quarks se comportent comme des particules quasiment libres dans les processus de haute énergie accessibles au LHC.

1.1.2 Symétrie et brisure électrofaible

L'interaction électrofaible du modèle standard décrit l'interaction électromagnétique, médiée par les photons, et l'interaction faible, médiée par les bosons Z et W (W^+ et W^-). Contrairement aux gluons et aux photons, les bosons Z et W sont massifs. Expérimentalement, leur masse est de $m_Z = 91.1876 \pm 0.0021$ GeV et $m_W = 80.387 \pm 0.016$ GeV respectivement [1]. Comme toutes les particules élémentaires massives, les bosons Z et W acquièrent une masse en interagissant avec le champ de Higgs, dont la valeur moyenne dans le vide v est non nulle ($v = 246$ GeV). Le photon et les gluons restent sans masse car ils n'interagissent pas directement avec le champ de Higgs, sachant que celui-ci n'est ni chargé électriquement ni coloré. Lorsque le champ de Higgs acquiert une valeur moyenne dans le vide, il brise la symétrie $SU(2) \times U(1)$ sur laquelle repose l'interaction électrofaible. Avant cette brisure de symétrie, le lagrangien électrofaible peut s'écrire comme une somme de quatre termes :

$$\mathcal{L}_{EW} = \mathcal{L}_g + \mathcal{L}_f + \mathcal{L}_H + \mathcal{L}_y. \quad (1.6)$$

Le terme \mathcal{L}_g décrit la dynamique des bosons de jauge W et B associés aux groupes $SU(2)$ et $U(1)$ respectivement :

$$\mathcal{L}_g = -\frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \quad (1.7)$$

avec leurs champs tensoriels associés $W_a^{\mu\nu}$ ($a = 1, 2, 3$) et $B^{\mu\nu}$. Les interactions électrofaibles traitent différemment la chiralité droite et la chiralité gauche des fermions. Les fermions de chiralité gauche, ou fermions gauche, se transforment en doublets selon $SU(2)$. Ils sont notés $L^p = (\nu_L^p, e_L^p)$ pour les leptons et $Q^p = (u_L^p, d_L^p)$ pour les quarks d'une même génération p . Les fermions droits, u_R^p et d_R^p pour les quarks, e_R^p pour les leptons chargés, sont des singlets de $SU(2)$. La dynamique électrofaible des fermions est décrite par le terme suivant :

$$\mathcal{L}_f = i[\bar{Q}_L^p \not{D} Q_L^p + \bar{u}_R^p \not{D} u_R^p + \bar{d}_R^p \not{D} d_R^p + \bar{L}_L^p \not{D} L_L^p + \bar{e}_R^p \not{D} e_R^p] \quad (1.8)$$

en sommant sur les p pour les trois générations, et avec $\not{D} \equiv \gamma^\mu D_\mu$. La dérivée covariante

$$D_\mu = \partial_\mu + \frac{i}{2} g \tau_a W_\mu^a + \frac{i}{2} g' Y B_\mu \quad (1.9)$$

fait intervenir les générateurs τ_a et Y des groupes $SU(2)$ et $U(1)$, ainsi que leur constante de couplage associée g (isospin faible) et g' (hypercharge faible) respectivement. Les τ_a sont nuls pour les composantes droites des fermions.

Le terme \mathcal{L}_H décrit la dynamique du champ de Higgs H , qui est un champ scalaire, complexe et doublet de $SU(2)$:

$$\mathcal{L}_H = |D_\mu H|^2 - \lambda \left(|H|^2 - \frac{v}{2} \right)^2 \quad (1.10)$$

où λ est constante d'auto-couplage. L'interaction du champ de Higgs avec les fermions est décrite par les termes de Yukawa :

$$\mathcal{L}_y = Y_{pq}^d \bar{Q}_L^p H d_R^q + Y_{pq}^u \bar{Q}_L^p \tilde{H} u_R^q + Y_{pq}^e \bar{L}_L^p H e_R^q + h.c \quad (1.11)$$

où les Y_{pq}^d , Y_{pq}^u et Y_{pq}^e sont les matrices de couplage de Yukawa pour les quarks de charge $-1/3$, $+2/3$ et pour les leptons chargés respectivement. Les conjugués hermitiens des termes explicitement écrits dans \mathcal{L}_y sont ajoutés avec la notation $h.c$. Le terme \mathcal{L}_y du lagrangien est responsable de la génération des masses des fermions après brisure de symétrie électrofaible. Les neutrinos ne sont pas inclus dans ces termes de Yukawa car la génération de leur masse pourrait être plus complexe, du fait de leur masse extrêmement faible par rapport aux autres fermions.

Avec la brisure de symétrie électrofaible, le champ de Higgs est transformé, par exemple, de la manière suivante :

$$H = \begin{pmatrix} \phi^+ \\ \phi_0 \end{pmatrix} \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix} \quad (1.12)$$

où h représente un champ scalaire réel de moyenne nulle dans le vide. Le lagrangien électrofaible après brisure spontanée de symétrie peut être réécrit en terme de champs état propre de masse, dont les particules associées sont observables [2]. La masse des bosons Z et W vient du terme cinétique $|D_\mu H|^2$ de \mathcal{L}_H . Le lien entre les bosons W_3 et B d'une part et le photon et le boson Z d'autre part est donné par :

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}, \quad \cos \theta_W = \frac{g}{\sqrt{g^2 + g'^2}} \quad (1.13)$$

où θ_W est l'angle de Weinberg. Les bosons W_1 et W_2 se combinent pour donner les bosons de jauge massifs

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2). \quad (1.14)$$

Les masses et couplages de Yukawa des fermions apparaissent avec la diagonalisation des matrices de Yukawa. Le lagrangien de Yukawa après brisure de symétrie devient (ici seuls les quarks de charge $+2/3$ sont considérés) :

$$\mathcal{L}_y = \frac{1}{\sqrt{2}} y_p^u \bar{u}_L^p h u_R^p + m_p^u \bar{u}_L^p u_R^p + h.c \quad (1.15)$$

où y_p^u est le couplage de Yukawa du quark de charge $+2/3$ de la $p^{\text{ème}}$ génération et $m_p^u = y_p^u v / \sqrt{2}$ est sa masse associée. Par conséquent le modèle standard prédit une proportionnalité entre la masse des fermions et leur couplage de Yukawa.

1.1.3 Les limites du modèle standard

Pendant les 30 dernières années les expériences de physique des particules ont testé le modèle standard aussi bien en collisions leptoniques qu'en collisions hadroniques. Le succès du modèle standard provient de sa capacité à prédire les observations réalisées jusqu'à présent. Il reste cependant de nombreuses questions ouvertes qui laissent à penser que le modèle standard n'est pas complet :

- le grand nombre de paramètres (18) de la théorie qui doivent être déterminés par des mesures ;
- la génération de la masse des neutrinos, dont les masses sont très faibles comparées à celles des autres fermions ;
- la théorie n'est pas en mesure d'expliquer certaines symétries comme celle qui entraîne la conservation du nombre baryonique ;
- le modèle standard ne propose aucune explication plausible concernant les observations cosmologiques, à savoir l'origine de la matière noire, de l'énergie sombre ou la prédominance de la matière sur l'antimatière ;
- la théorie n'explique pas l'existence d'exactly 3 générations pour les quarks et les leptons ;
- le modèle standard n'explique pas la gravitation.

1.2 Production de paire de quark *top*

Depuis sa découverte en 1995 par les deux expériences CDF et D0 au Tevatron, le quark *top* fait l'objet d'intenses études [3]. Il appartient avec le quark *b* à la troisième génération de quarks, et c'est la particule la plus lourde du modèle standard. Sa masse est de 173.34 ± 0.76 GeV [4]. Le quark *top* est donc le fermion qui interagit le plus fortement avec le boson de Higgs. Son couplage de Yukawa est proche de 1, bien au dessus de ceux des autres fermions. Pour cette raison, le quark *top* joue un rôle particulier dans de nombreuses théories qui proposent de résoudre les limitations du modèle standard. Les mesures précises des productions de quarks *top* à haute énergie sont utilisées pour tester le modèle standard. Le quark *top* possède un temps de vie plus court que le temps typique d'hadronisation, sa largeur $\Gamma = 1.41 \pm 0.17$ GeV [5] étant supérieur à l'échelle de l'interaction forte $\Lambda_{QCD} \sim 0.2$ GeV. Il se désintègre donc avant de former un hadron, contrairement aux autres quarks produits dans les collisions de haute énergie.

Le mode dominant de production de quark *top* au LHC est la production par paire de quark top-antitop ($t\bar{t}$) via l'interaction forte. Cette production fait intervenir deux partons provenant des protons du LHC dans l'état initial : une paire $q\bar{q}$ ou une paire gluon gg . Ce processus est un des bruits de fond principaux du signal $t\bar{t}H$. Sa section efficace théorique au LHC est de 832^{+46}_{-51} pb à une énergie de centre de masse proton-proton de $\sqrt{s} = 13$ TeV [6]. Elle est calculée au second ordre en α_s (NNLO) et comprend la resommation des termes logarithmiques dominants pour les termes de gluon de faible énergie jusqu'au second

ordre (NNLL).

Le quark *top* se désintègre presque exclusivement en un boson W^+ et un quark b . Le boson W^+ peut ensuite se désintégrer de façon hadronique avec l'émission d'une paire de quarks $q\bar{q}'$, principalement $u\bar{d}$ et $s\bar{c}$ ($\Gamma_{W \rightarrow q\bar{q}'} \sim 70\%$), ou de façon leptonique en une paire de leptons $l^+\nu_l$, où $l = e, \mu, \tau$ ($\Gamma_{W \rightarrow l\nu_l} \sim 30\%$). Le quark *top* est nommé *quark top hadronique* dans le premier cas, et *quark top leptonique* dans le second. Les paires $t\bar{t}$ sont classées selon le nombre de quark *top* leptonique : tout hadronique, semi-leptonique et tout leptonique pour 0, 1 et 2 quark *top* leptonique(s) respectivement, avec des taux de branchement de $\sim 46\%$, $\sim 45\%$ et $\sim 9\%$.

1.3 Le boson de Higgs

Les recherches du boson de Higgs dans ses principaux modes de production et de désintégration laissent à penser que le boson de Higgs découvert en 2012 est vraisemblablement compatible avec celui prédit par le modèle standard (Figure 1.3) [7, 8]. Sa masse mesurée au Run 1 du LHC (période s'étalant de 2011 à 2012) par les expériences ATLAS et CMS a pour valeur combinée : $m_H = 125.0.9 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})$ GeV. Cette particule a été observée dans les modes de désintégration en paires de photons, de bosons Z , de bosons W et de leptons τ avec une signification de plus de 5σ . Le boson de Higgs découvert est également compatible avec l'hypothèse d'une particule scalaire de parité positive, tel qu'il est prédit par le modèle standard.

1.3.1 Production au LHC

Le boson de Higgs du modèle standard peut être produit dans les collisions pp du LHC selon quatre principaux modes dont les diagrammes de Feynman dominants sont présentés Figure 1.1 : par une boucle de quark *top* engendrée par une *fusion de gluons*, par une *fusion de bosons Z/W* radiés par des quarks de l'état initial (qqH), par radiation à partir d'une paire de quark/antiquark *top* ($t\bar{t}H$) ou enfin par radiation à partir d'un boson Z/W (VH). Leur section efficace en fonction de l'énergie dans le centre de masse des protons est présentée Figure 1.2a pour un boson de Higgs de $m_H = 125$ GeV [9]. Les bosons de Higgs sont produits le plus souvent par fusion de gluon, avec une section efficace de presque $\sigma = 50$ pb à 13 TeV, suivi du processus de fusion de bosons W/Z avec presque $\sigma = 4$ pb, et des processus de troisième ordre VH et $t\bar{t}H$ avec des sections efficaces respectives de ~ 0.9 pb et ~ 0.5 pb.

Le boson de Higgs se couple au quark *top* pour les processus de fusion de gluon et $t\bar{t}H$. Les mesures des sections efficaces des deux processus permettent toutes deux d'extraire le couplage de Yukawa du quark *top*. Cependant, une déviation par rapport au modèle standard de la section efficace du processus de fusion de gluon pourrait non seulement provenir du couplage de Yukawa du quark *top* y_t mais aussi d'une particule non prédite par le modèle standard contribuant à la boucle. En revanche, le processus $t\bar{t}H$ permet de

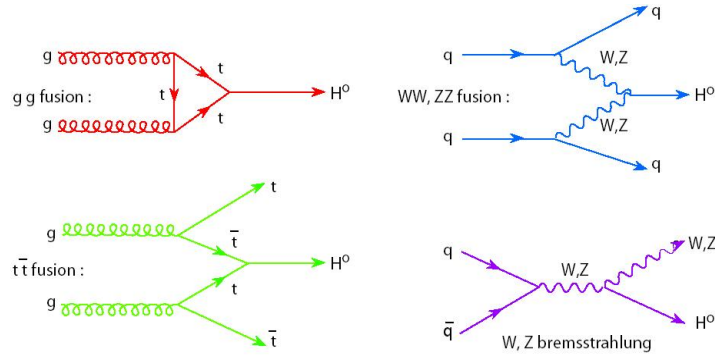


FIGURE 1.1: Diagrammes de Feynman dominants des quatre principaux modes de production du boson de Higgs au LHC.

réaliser une mesure directe du couplage y_t .

1.3.2 Modes de désintégration

Le boson de Higgs de masse $m_H = 125$ GeV présente un large éventail de désintégrations, comme indiqué Figure 1.2b, conduisant à une variété de signatures recherchées par les expériences ATLAS et CMS.

$H \rightarrow q\bar{q}$

Le boson de Higgs se désintègre en une paire $b\bar{b}$ avec un rapport de branchement $\Gamma_{b\bar{b}} \sim 57\%$. Malgré un rapport de branchement élevé, ce canal de désintégration reste difficile à observer pour les modes de productions de Higgs dominants via fusion de gluons ou de bosons vecteurs, en raison d'un bruit de fond extrêmement important produisant des jets (*multijet*). Les modes de productions VH et $t\bar{t}H$ (Figure 1.3) sont sensibles à la désintégration $H \rightarrow b\bar{b}$ malgré une section efficace bien plus faible que le processus $gg \rightarrow H$, car les désintégrations leptoniques de boson V et de quark *top* permettent de supprimer presque totalement le bruit de fond *multijet*. Le canal $c\bar{c}$ est encore plus difficile à observer, avec un faible rapport de branchement $\Gamma_{c\bar{c}} \sim 3\%$ et une difficulté à identifier les quarks c dans l'environnement hadronique du LHC. Les désintégrations en quarks plus légers sont hors d'atteinte du LHC.

$H \rightarrow VV^*$ ou $\gamma\gamma$

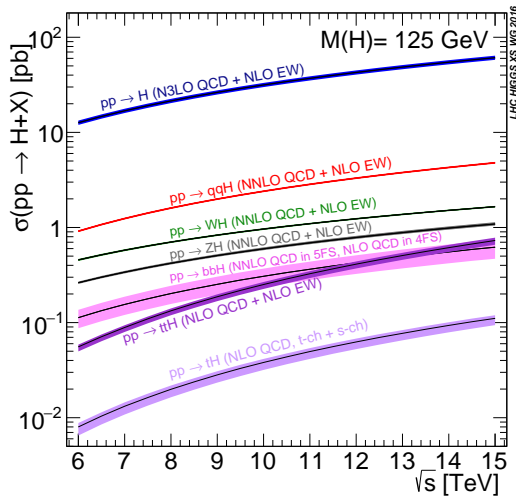
Le boson de Higgs se désintègre en deux bosons vecteur, WW^* ou ZZ^* (dont l'un des deux est émis hors couche de masse), avec des rapports de branchement de $\sim 21\%$ et $\sim 3\%$ respectivement. Les bosons vecteurs sont recherchés principalement dans leurs modes de désintégration en leptons. Bien que le mode Z soit plus rare, le canal $ZZ^* \rightarrow 4l$ avec $l = e, \mu$ possède un très faible bruit de fond et permet de reconstruire la masse du boson de

Higgs avec une très bonne résolution. Le canal WW^* est recherché dans sa désintégration en $2l\nu$ et $lvq\bar{q}'$.

Le boson de Higgs ne se couple pas directement aux photons, mais indirectement via des boucles de bosons W et de quark top . Le rapport de branchement correspondant $\Gamma_{\gamma\gamma} \sim 0.2\%$ est relativement faible. Cependant il est possible de mesurer la masse du boson Higgs dans ce canal de façon précise du fait de la bonne résolution en énergie des photons détectés dans CMS et ATLAS.

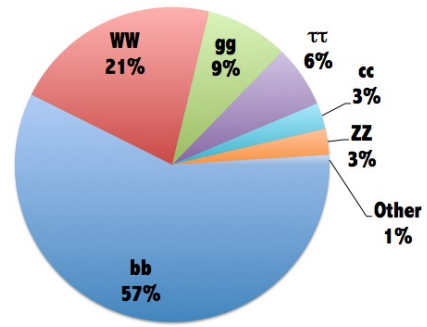
$$H \rightarrow l^+l^-$$

Les couplages du boson de Higgs aux leptons sont étudiés dans les désintégrations $H \rightarrow \tau\bar{\tau}$ ($\Gamma_{\tau\bar{\tau}} \sim 6\%$) et $H \rightarrow \mu\bar{\mu}$ ($\Gamma_{\mu\bar{\mu}} \sim 0.02\%$). Le second mode de désintégration est limité actuellement par la faible statistique lié au très faible rapport de branchement.



(a)

Higgs decays at $m_H=125\text{GeV}$



(b)

FIGURE 1.2: (a) Sections efficaces des modes de production du boson Higgs en fonction de l'énergie dans le centre de masse \sqrt{s} de la collision pp [9]. (b) Rapport de branchement des différents modes de désintégration du boson de Higgs.

Dans l'analyse d'événements $t\bar{t}H$ présentée dans cette thèse, le boson de Higgs est recherché dans sa désintégration en paire $b\bar{b}$ ($t\bar{t}H(b\bar{b})$).

1.4 De la théorie à la génération d'événements

La plupart des événements d'intérêt analysés par l'expérience ATLAS mettent en jeu des collisions pp à grand transfert d'énergie Q , *i.e* à haute échelle d'énergie. Ces collisions sont simulées à partir de modèles physiques pour tester leur validité. La simulation d'une collision d'intérêt est décomposée en plusieurs processus qui se produisent à différentes échelles d'énergie, depuis la collision dure à $Q \gg m_p (\sim 1 \text{ GeV})$ jusqu'aux phénomènes

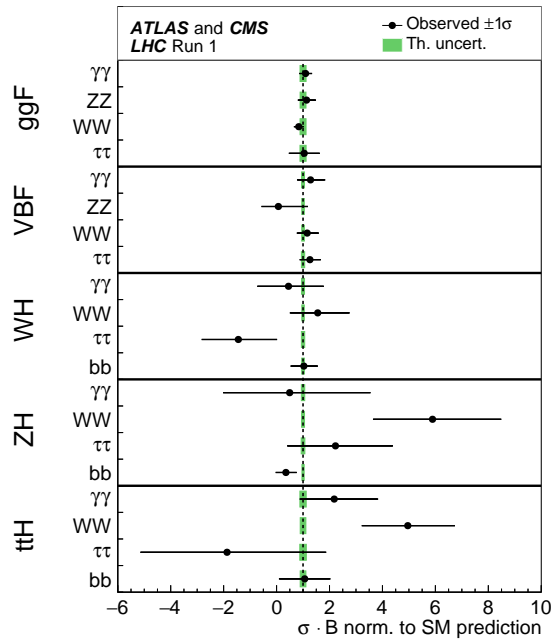


FIGURE 1.3: Mesures de $\sigma \times B$ (section efficace \times rapport de branchement) normalisées à sa valeur prédite par le modèle standard, pour chaque canal spécifique $i \rightarrow H \rightarrow f$, obtenues à partir de la combinaison des mesures ATLAS et CMS [7]. Les barres d'erreur indiquent les intervalles de 1σ . Les bandes vertes autour de 1 indiquent les incertitudes théoriques des prédictions.

d'hadronisation à $Q \sim m_p$, responsables de la formation des hadrons susceptibles d'interagir avec le détecteur. La collision dure fait intervenir dans l'état initial deux partons quasiment libres provenant des deux protons du LHC. Les deux partons a et b emportent une fraction x_a et x_b de l'impulsion du proton dont ils sont issus, ce qui implique une énergie dans le centre de masse de $x_a x_b \sqrt{s}$. Pour un processus donné $a, b \rightarrow c$ (par exemple $gg \rightarrow t\bar{t}$), la section efficace de la collision dure $d\hat{\sigma}_{a,b \rightarrow c}(\mu_F, \mu_R)$ est calculable en théorie des perturbations. Du fait de la nécessité de calculer des corrections QCD au delà du premier ordre en théorie des perturbations, la section efficace partonique acquiert une dépendance avec l'échelle de factorisation μ_F . De plus, elle acquiert une dépendance avec l'échelle de renormalisation μ_R due à la renormalisation de la constante de couplage forte α_s . Le lien entre l'impulsion d'un parton et celle de son proton d'origine est donné par les fonctions de distributions partoniques (PDF). Il existe une PDF par type de parton a , classé selon qu'il soit un gluon, un quark ou un anti-quark, et selon sa saveur. La PDF $f_a(x_a, \mu_F)$ exprime la densité de probabilité qu'un parton de type a avec une fraction x_a d'impulsion du proton d'origine. La dépendance avec l'échelle de factorisation μ_F est décrite par les équations d'évolution DGLAP [10]. Finalement, la section efficace hadronique est obtenue en convoluant les PDFs $f_a(x_a, \mu_F)$ et $f_b(x_b, \mu_F)$ avec les sections efficaces partoniques $d\hat{\sigma}_{a,b \rightarrow c}(\mu_F, \mu_R)$:

$$d\sigma(pp \rightarrow c) = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b f_a(x_a, \mu_F) f_b(x_b, \mu_F) d\hat{\sigma}_{a,b \rightarrow c}(\mu_F, \mu_R). \quad (1.16)$$

A ce stade, des événements sont générés à partir d'une méthode Monte Carlo, grâce à un logiciel appelé générateur d'événements. Chaque événement Monte Carlo généré contient les quadri-impulsions des particules de l'état final au niveau de la collision dure. La distribution des événements selon ces quadri-impulsions est donnée par la section efficace hadronique. Le calcul de la section efficace hadronique est obtenue grâce aux mesures des PDFs et au calcul des sections efficaces partoniques par le générateur d'événements. La connaissance actuelle des PDF utilisées au LHC vient principalement de mesures réalisées aux expériences du LHC, du Tevatron et de Hera.

De nombreux processus produisent des quarks et gluons dans l'état final au niveau de la collision dure, émis typiquement à haute impulsion transverse. Ils perdent de l'énergie en rayonnant de nouveaux partons par interaction forte, jusqu'à une échelle d'énergie de l'ordre de 1 GeV. Ils entrent alors dans un régime non perturbatif et s'hadronisent. Ces deux étapes sont simulées par des modèles phénoménologiques de *gerbe de partons* et d'hadronisation. Les processus de basse énergie provenant d'événements sous-jacents et de radiations de photons additionnelles sont également simulés à partir de modèles phénoménologiques. A la suite de ces simulations, les événements Monte Carlo sont finalement décrits comme un ensemble de particules dans l'état final qui sont utilisées en entrée de la simulation du détecteur, dans laquelle chaque particule est propagée à l'intérieur du détecteur ATLAS. Cette dernière étape est introduite Section 3.7.

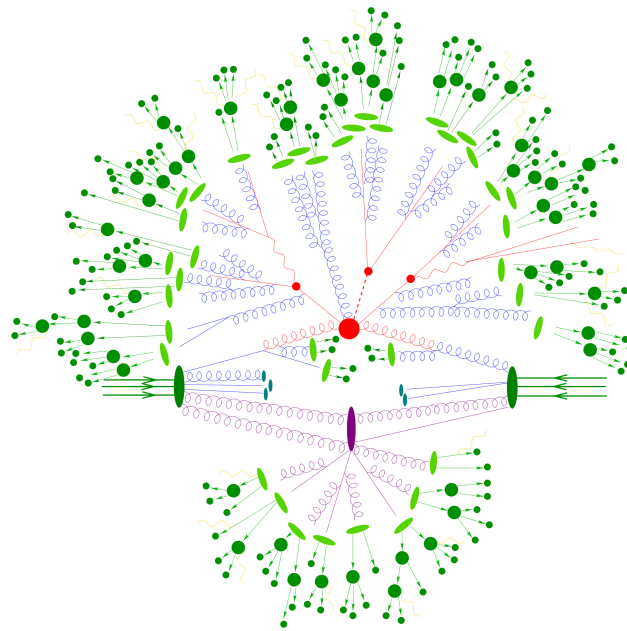


FIGURE 1.4: Schéma représentant la **collision dure** produisant un événement $t\bar{t}H(b\bar{b})$ semi-leptonique, accompagnée d'**événements sous-jacents**. Les **partons de la collision dure** radient des **partons de fragmentation** (phénomène de gerbe de parton) qui s'**hadronisent** pour former des hadrons.

Bibliographie

- [1] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38 :090001, 2014.
- [2] The Standard Model Higgs Boson. <https://www.nikhef.nl/~ivov/HiggsLectureNote.pdf>, 2013.
- [3] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74 :2632–2637, 1995.
- [4] First combination of Tevatron and LHC measurements of the top-quark mass. 2014.
- [5] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10) :100001, 2016.
- [6] Michal Czakon and Alexander Mitov. Top++ : A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, 185 :2930, 2014.
- [7] Georges Aad et al. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 08 :045, 2016.
- [8] J R Andersen et al. Handbook of LHC Higgs Cross Sections : 3. Higgs Properties. 2013.
- [9] LHC Higgs Cross Section WG Picture Gallery. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHSWGCCrossSectionsFigures>, 2016.
- [10] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126 :298–318, 1977.

Chapitre 2

Le context expérimental : le LHC et le détecteur ATLAS

2.1 Le collisionneur LHC et ses performances au Run 2

Le collisionneur LHC (Large Hadron Collider) est un accélérateur et collisionneur de hadrons installé à l'intérieur d'un tunnel circulaire de 26.7km de long situé au Cern. Le tunnel a été construit à l'origine pour accueillir le collisionneur e^+e^- LEP [11]. Deux faisceaux de hadrons sont accélérés dans des directions opposées le long du tunnel grâce à des cavités accélératrices tandis qu'un système d'aimants supraconducteurs courbent les trajectoires des faisceaux et les focalisent. Le LHC est principalement un collisionneur proton-proton, mais des prises de données (*runs*) dédiées aux collisions d'ions lourds sont effectuées notamment à partir de noyaux d'atomes de plomb [12]. Les collisions ont lieu en quatre points de l'anneau. Autour de deux d'entre eux se trouvent les détecteurs généralistes des expériences ATLAS et CMS (Figure 2.1).

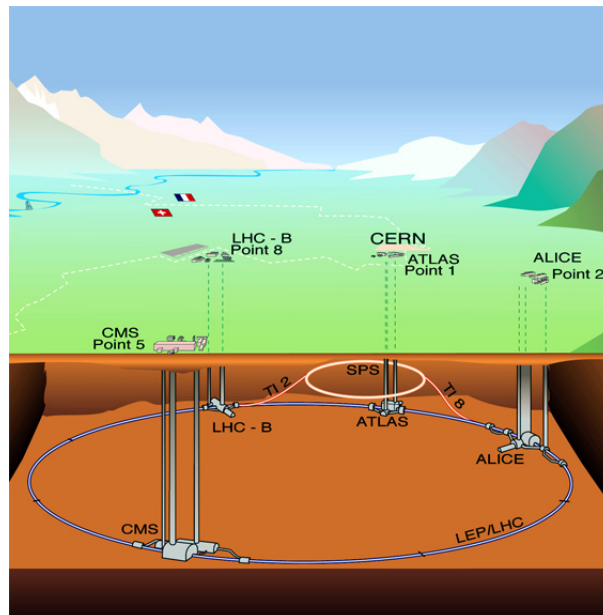


FIGURE 2.1: Schéma du tunnel du LHC avec ses quatre expériences ATLAS, ALICE, CMS et LHCb.

En amont du LHC, les faisceaux sont accélérés par une chaîne d'accélération (présentée dans la référence [13]), formée du Linac2 qui alimente le Proton Synchrotron Booster (PSB) en protons de 50 MeV. Les protons sont ensuite injectés à 1.4 GeV dans le PS (Proton Synchrotron) qui les accélère à 25 GeV avant leur passage dans le SPS (Super Proton Synchrotron) pour une montée en énergie à 450 GeV. L'énergie maximale atteinte par les protons qui circulent au LHC est limitée par le champ magnétique des aimants dipolaires nécessaires pour maintenir les particules le long de leur trajectoire circulaire : le champ magnétique nominal est de 8.3 T ce qui correspond à une énergie des protons de 7 TeV par faisceau. Au Run 2, qui a débuté en 2015 et qui s'achèvera en 2018, l'énergie des faisceaux est maintenue à 6.5 TeV, entraînant des collisions pp avec une énergie dans le centre de

masse $\sqrt{s} = 13$ TeV.

La luminosité instantanée est également un paramètre essentiel du collisionneur. La luminosité instantanée ℓ relie le taux de collisions proton-proton R pour un processus donné de section efficace σ de la manière suivante :

$$R = \ell \times \sigma. \quad (2.1)$$

Elle dépend des caractéristiques des faisceaux au niveau du point de collision. Les faisceaux de protons sont divisés en paquets de protons successifs. Au Run 2, chaque faisceau est constitué d'environ $n_b = 2000$ paquets de $n_p = 1.1 \times 10^{11}$ protons séparés de 25 ns (soit à une distance de 7.5 m). La luminosité instantanée est donnée par l'expression :

$$\ell = f \frac{n_p^2}{4\pi\sigma_r^2} \quad (2.2)$$

où f est la fréquence de collision de deux paquets, et σ_r est la taille transverse du faisceau définie comme la déviation standard de la distribution des protons dans le plan transverse à la direction du faisceau (typiquement $\sigma_r \sim 20 \mu\text{m}$). Cette expression simplifiée de la luminosité fait l'hypothèse que l'angle entre les directions des faisceaux est nul, que le profil du paquet ne dépend pas de la position selon l'axe du faisceau, et que les profils des paquets ne sont pas altérés pendant les collisions. La taille transverse du faisceau σ_r est souvent exprimée comme le produit de l'émittance transverse ϵ_r et de la fonction d'amplitude β^* . L'émittance d'un paquet est une mesure combinée de la taille transverse et de la divergence angulaire des protons dans un paquet. La fonction d'amplitude β^* est une quantité optique et est déterminée par la configuration des quadrupôles, exprimée par

$$\beta^* = \pi\sigma_r^2/\epsilon_r. \quad (2.3)$$

L'émittance transverse normalisée $\epsilon_n = \epsilon_r/\gamma$, où γ est le facteur de Lorentz des protons, est utilisée pour quantifier les performances du LHC. En prenant en compte l'angle entre les faisceaux au niveau des croisements de paquets, l'expression de la luminosité en fonction de l'émittance transverse normalisée ϵ_n et de la fonction d'amplitude β^* devient :

$$\ell = f \frac{n_p^2\gamma}{4\epsilon_n\beta^*} \times F \quad (2.4)$$

où F est le facteur de réduction géométrique défini par :

$$F = \left(1 + \left(\frac{\theta\sigma_z}{\sigma_r} \right)^2 \right)^{-1/2} \quad (2.5)$$

et σ_z est défini comme le profil longitudinal (parallèlement au faisceau) d'un paquet, et θ la moitié de l'angle entre les faisceaux de protons. Finalement, pour un processus de section efficace σ , le nombre d'événements N récolté sur une période de prise de donnée est égal à l'intégral du taux d'événements sur cette période

$$N = \sigma \times \int_t \ell dt = \sigma \times \mathcal{L} \quad (2.6)$$

où \mathcal{L} est défini comme la luminosité intégrée, ou simplement luminosité.

Les paramètres intervenant dans le calcul de la luminosité (Eq. 2.4) sont renseignés dans la Table 2.1 [14]. L'année 2016 a été une réussite pour le LHC [15]. En fournissant aux expériences ATLAS et CMS une luminosité de 40/fb, les performances du LHC ont surpassé celles attendues, avec des prédictions de $\mathcal{L} = 25/\text{fb}$. Au cours de cette même année, une luminosité de 33.2/fb a été enregistrée par le détecteur ATLAS lorsque celui-ci était pleinement opérationnel. Cette amélioration par rapport à 2015 (voir Figure 2.2) a été obtenue grâce à une réduction de la taille des faisceaux au niveau des points d'interaction, avec d'une part la production de faisceaux plus fins aux niveau de la chaîne d'injection et d'autre part la réduction de l'angle de croisement entre les faisceaux. La luminosité instantanée a atteint une valeur maximale de $1.4 \times 10^{34}/\text{cm}^2/\text{s}$, 40% au-dessus de sa valeur de conception. Le taux de collisions pp inélastiques à 13 TeV, de section efficace $\sigma_{inel} \sim 70$ mb [16], pour une luminosité instantanée typique de $10^{34}/\text{cm}^2/\text{s}$ est d'environ 1 GHz, soit un nombre moyen de collisions de 25 par croisement de paquets. Ce flux de particules émanant des points de collision expose les détecteurs ATLAS et CMS à de fortes irradiations.

	Design	2011	2012	2015	2016 (June)
Energie du faisceau (TeV)	7.0	3.5	4.0	6.5	6.5
Nombre de Protons (10^{11})/bunch n_p	1.15	1.3	1.5	1.1	1.1
Nombre de Paquets n_b	2808	1380	1380	2244	2076
Espacement de paquet (ns)	25	50	50	25	25
Emittance normalisée ϵ_n (μm)	3.75	2.4	2.4	3.5	3.4
Demi angle de croisement θ (μrad)	143	120	146	146	185
Fonction d'amplitude β^*	0.55	1.0-1.5	0.6	0.8	0.4
Luminosité \mathcal{L} (/fb)		5.5	22.8	4.2	40

TABLE 2.1: Conditions typiques des opérations du LHC lors des collisions de protons au cours des années 2011, 2012, 2015 et 2016 (jusqu'à juin) et conditions prévues à la conception du LHC [14].

2.2 Le détecteur ATLAS

Lors de la prise de données, le détecteur ATLAS est exposé à un environnement fort contraignant. Il subit d'importantes doses de radiations notamment à proximité du tube à faisceau. De plus, il doit opérer à un taux de collisions pp élevé tout en fournissant des données de qualité. Malgré ces conditions difficiles, le détecteur doit collecter efficacement les événements intéressants pour permettre à la collaboration ATLAS d'étudier avec précision les processus de hautes énergies. Ceci entraîne les exigences suivantes sur le détecteur ATLAS :

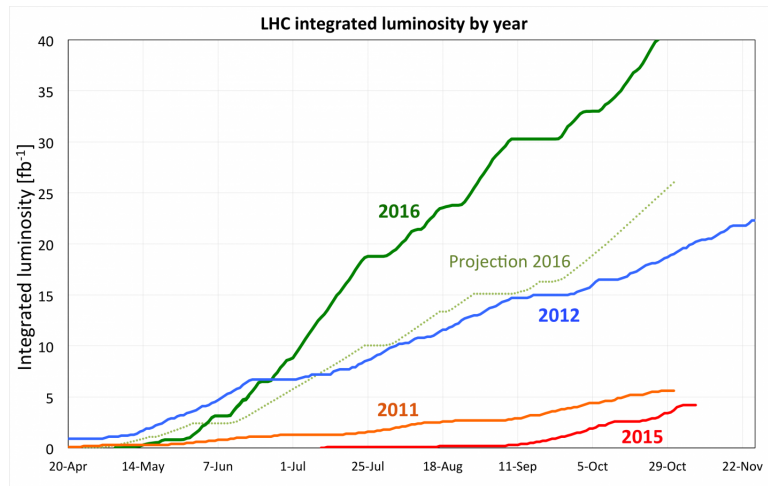


FIGURE 2.2: Luminosité intégrée fournie par le LHC en 2011, 2012, 2015 et 2016 et prévision faites début 2016 des performances pour l'année 2016.

- importante couverture angulaire grâce à une acceptation élevée en pseudo-rapacité et une couverture azimutale quasiment complète ;
- bonne résolution des trajectoires de particules chargées et efficacité de leur reconstruction dans le trajectographe, pour l'identification de leptons chargés et des jets de b ;
- très bon calorimètre électromagnétique pour l'identification des électrons et photons et leur mesure d'énergie, complété par un calorimètre hadronique hermétique pour la précision de la mesure de l'énergie des jets et de l'énergie transverse manquante ;
- identification précise des muons et de leur impulsion couvrant un large spectre en impulsion grâce au spectromètre à muon ;
- rapidité et flexibilité du système de déclenchement pour enregistrer les événements intéressants.

Le point d'interaction nominal, situé au centre du détecteur ATLAS, est défini comme l'origine du système de coordonnées. Un repère orthonormé direct est formé de l'axe z selon la direction du faisceau de proton et d'un plan $x - y$ perpendiculaire (ou transverse). Les plans longitudinaux sont parallèles à l'axe z . La direction de l'axe x pointe vers le centre de l'anneau du LHC.

Les directions sont exprimées à l'aide des coordonnées (η, ϕ) , où ϕ est l'azimut mesuré autour de l'axe du faisceau et où $\eta = -\log(\tan(\theta/2))$ est la pseudo-rapacité, avec θ l'angle polaire mesurant l'angle par rapport à l'axe z . Les positions sont exprimées avec les coordonnées cylindriques usuelles (r, ϕ, z) .

Une coupe du détecteur ATLAS est représentée Figure 2.3, laissant apparaître les sous-détecteurs internes jusqu'au point de collision. Les sous-détecteurs ont une forme cylindrique et sont encastrés les uns dans les autres à partir du point de collision. Le sous-détecteur le

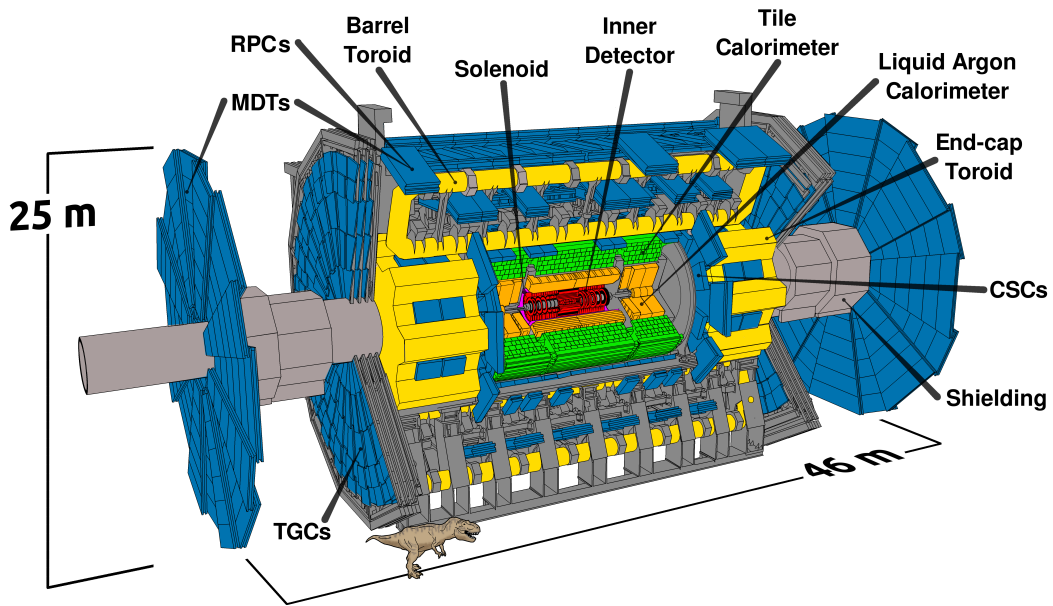


FIGURE 2.3: Vue en coupe du détecteur ATLAS.

plus proche du point de collision, appelé Détecteur Interne, est un trajectographe : il sert à mesurer les trajectoires des particules chargées. Les impulsions des particules chargées sont déterminées grâce à la courbure de leurs trajectoires sous l'effet du champ magnétique de l'aimant solénoïdal entourant le Détecteur Interne. Un calorimètre, qui entoure le Détecteur Interne et l'aimant solénoïdal, mesure l'énergie déposée par les hadrons, électrons et photons. La trajectoire des muons qui ont traversé le calorimètre est mesurée dans le spectromètre à muon, dans lequel ils sont déviés par le champ magnétique de l'aimant toroidal.

Le Détecteur Interne, le calorimètre et le spectromètre à muon seront présentés brièvement dans les sous-sections suivantes. Une description détaillée du détecteur ATLAS et de ses sous-détecteur peut être trouvée dans le Technical Design Report [17].

2.2.1 Le Détecteur Interne

En traversant le Détecteur Interne, les particules chargées déposent leur énergie dans la matière le long de leur trajectoire par interaction électromagnétique. Les pertes d'énergie sont détectées en plusieurs positions (ou *coups*) par le sous-détecteur, ce qui permet de mesurer les trajectoires empruntées par les particules chargées, ou *traces*. Les traces sont utilisées dans de nombreuses applications : pour trouver les positions des collisions pp d'intérêt ou des désintégrations de hadrons de courte durée de vie (autour de la ps). Elles participent également à l'identification des électrons, muons, taus ou encore des jets. Les traces sont reconstruites grâce à la combinaison de trois sous-détecteurs présentés dans les sections suivantes : les deux détecteurs de silicium à haute granularité avec le Détecteur à Pixel et le Trajectographe Semiconducteur (SCT), et le Trajectographe à Radiation de

Transition (TRT). Le Détecteur Interne est contenu dans un volume cylindrique d'environ $\Delta z = 6.2$ m de longueur et de $\Delta r = 2,1$ m de profondeur, dont la structure interne est présentée Figure 2.4. Il est entouré d'un aimant solénoïdal créant un champ de 2 T parallèle au faisceau à l'intérieur du volume du détecteur (Figure 2.6). Le champ courbe les particules chargées dans le plan transverse. La mesure de la courbure des traces permet de mesurer l'impulsion des particules chargées.

Le Détecteur Interne est soumis à un flux intense de particules émergeant du point de collision, avec typiquement 1000 particules émises au niveau du point de collision par croisement de paquets de protons produisant 25 collisions inélastiques. Pour permettre une reconstruction efficace des traces dans ces conditions, le Détecteur Interne possède une granularité élevée et une réponse rapide, et est résistant aux radiations. La granularité est d'autant plus importante que l'on s'approche de l'axe du faisceau afin de limiter le taux d'occupation du détecteur, et en particulier de limiter la fraction de cellules traversées par plusieurs particules chargées à la fois.

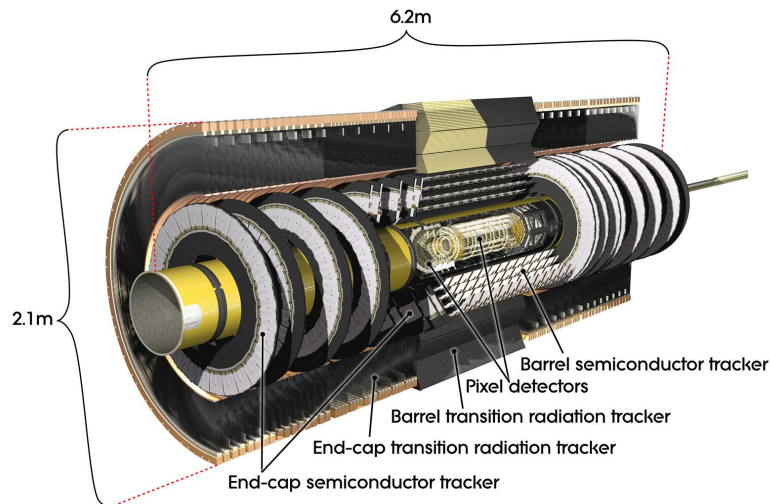


FIGURE 2.4: Vue en coupe (dans le plan longitudinal) du Détecteur Interne d'ATLAS.

Détecteur à Pixel

La composante du Détecteur Interne la plus proche des faisceaux de protons est le Détecteur à Pixel. Ce sous-détecteur utilise le silicium comme matériau actif, et possède une granularité très élevée, ce qui garantit un ensemble de mesures de haute précision au plus près possible du point d'interaction. Le système fournit en moyenne quatre mesures de positions pour une particule chargée émergeant du tube à faisceau, suite à son passage dans les différentes couches sensibles du sous-détecteur sur l'acceptance $|\eta| < 2.5$. Le sous-détecteur, par sa proximité avec le point de collision et sa granularité élevée, détermine principalement la résolution des paramètres d'impact des traces, et par conséquent la capacité du Détecteur Interne à trouver des particules à longue durée (\sim ps) de vie telles que hadrons

b. Le système est composé dans le tonneau de quatre couches concentriques à des rayons r moyens de 3 cm, 5 cm, 9 cm et 12 cm, présentées Figure 2.5, et de six disques de part et d'autre du tonneau, dont les rayons r sont compris entre 9 et 15 cm. La couche la plus interne du Détecteur à Pixel, appelée IBL pour Insertable B Layer, a été installée avant la reprise des prises de donnée du Run 2 pour améliorer les performances de la reconstruction des traces et pour faire face à l'augmentation de la luminosité instantanée prévue dans le futur, qui entraînera une densité de traces plus élevée et davantage d'irradiation. Elle couvre une pseudo-rapacité de $|\eta| < 2.9$. La couche suivante est nommée B-layer.

La résolution sur la mesure de la position des traces dans les couches sensibles définit la résolution spatiale intrinsèque du détecteur. La taille minimale en $r - \phi \times z$ des pixels est de $50 \times 250 \mu\text{m}^2$ pour l'IBL et de $50 \times 400 \mu\text{m}^2$ et pour les couches du tonneau suivantes. La résolution intrinsèque obtenue dans le tonneau est de $8 \times 40 \mu\text{m}^2$ en $r - \phi$ pour l'IBL et de $10 \times 115 \mu\text{m}^2$ dans les couches suivantes du tonneau [18].

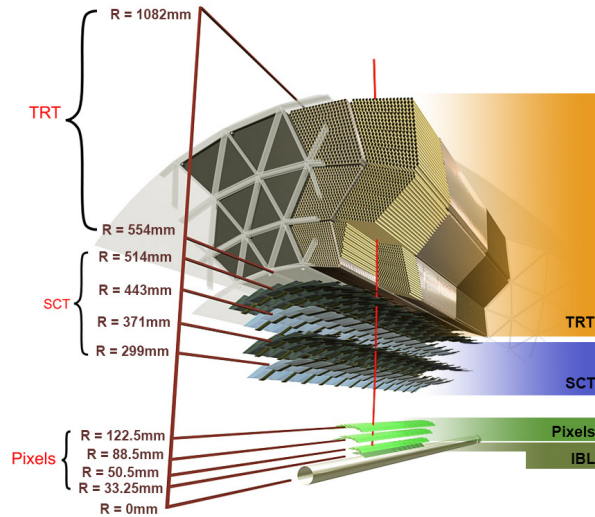


FIGURE 2.5: Vue en coupe (dans le plan transverse) des parties actives du Détecteur Interne d'ATLAS.

Trajectographe à Semiconducteur

Le Trajectographe à Semiconducteur forme la couche intermédiaire du détecteur interne. Il contribue principalement à la mesure de l'impulsion des traces, grâce à son relativement grand bras de levier (extension radiale) et sa granularité élevée. Il est formé de quatre couches concentriques dans le tonneau, visibles sur la Figure 2.5, et de neuf disques à chaque extrémité. L'ensemble du SCT occupe un cylindre de $\Delta z = 5.6$ m de long délimité par sa couche la plus externe de rayon $r = 56$ cm et sa couche la plus interne situé à $r = 27$ cm de l'axe du faisceau. Sa géométrie est conçue pour offrir en moyenne quatre mesures précises de position par trace jusqu'à des pseudo-rapidités de $|\eta| < 2.5$. Le détecteur a une surface active de 61 m^2 de détecteurs de silicium, avec 6.2 millions canaux de lecture. La précision

intrinsèque dans la région du tonneau est de $17 \mu\text{m}$ en $r - \phi$ et de $580 \mu\text{m}$ en z [18].

Trajectographe à Radiation de Transition

Le Trajectographe à Radiation de Transition (TRT) constitue la partie la plus à l'extérieur du Détecteur Interne. Environ 370,000 tubes à dérive de 4 mm de diamètre communément appelées *pailles* sont utilisés pour la reconstruction des traces des particules chargées. Les pailles du tonneau sont alignées parallèlement à l'axe du faisceau (Figure 2.5), tandis que les pailles des bouchons sont orientées radialement. Afin de réduire l'occupation dans le tonneau, les pailles sont divisées à $\eta = 0$ en deux moitiés de 144 cm de long qui sont lues séparément à chaque extrémité du tonneau. Avec cette géométrie, le TRT fournit en moyenne 36 mesures d'impact jusqu'à $|\eta| = 2.1$. La partie tonneau du TRT fournit une position selon $r - \phi$, avec une précision intrinsèque de $170 \mu\text{m}$ par paille. Même si la résolution intrinsèque du TRT ne peut rivaliser avec celle de détecteurs fondés sur une technologie à silicium, le nombre élevé de mesures et le long bras de levier dont bénéficie le TRT permet au détecteur de contribuer significativement à la résolution en impulsion des traces. Le TRT participe également à l'identification des électrons à l'aide de mesures de rayonnements de transition qui sont émis par les particules chargées traversant les feuilles de plastique agencées entre les pailles.

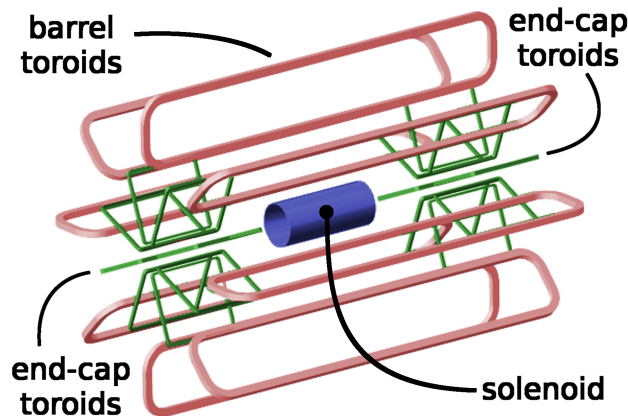


FIGURE 2.6: Système d'aimants du détecteur ATLAS. Le solénoïde permet de courber les trajectoires des particules chargées dans le détecteur interne et l'aimant toroïdal courbe les particules chargées en sortie du calorimètre.

2.2.2 Calorimètre

Le calorimètre est une composante majeure du détecteur ATLAS. Il permet de mesurer l'énergie des électrons, des photons et des hadrons produits lors des collisions pp . Ces particules déposent toute leur énergie dans le calorimètre en formant une gerbe de particules secondaires, de type électromagnétique pour les photons et électron ou hadronique pour les hadrons. Le calorimètre est divisé en une partie électromagnétique, dont l'objectif principal

est de fournir une mesure précise de l'énergie des photons et des électrons, et d'une partie hadronique, qui récupère l'énergie des hadrons n'ayant pas été absorbée en amont dans le calorimètre électromagnétique. Le calorimètre mesure également l'énergie transverse manquante, utilisée pour déduire la production de neutrinos, et participe à l'identification des électrons et des photons dans un environnement hadronique.

L'hermiticité est une caractéristique essentielle du calorimètre. D'une part, la profondeur du calorimètre, d'un minimum de 10 longueur d'interaction (λ), est suffisante pour contenir complètement l'énergie des hadrons produits au LHC. D'autre part, son extension en pseudo-rapacité jusqu'à $|\eta| = 4.9$ lui permet de couvrir un grand angle solide afin de capturer la majorité des particules provenant des collisions d'intérêt. L'hermiticité du calorimètre garantit de bonnes résolutions sur l'énergie transverse manquante et sur l'énergie des jets (cf. Section 3.5) jusqu'à des échelles d'énergie de l'ordre du TeV, et permet de protéger le spectromètre à muon (cf. Section 2.2.3) d'une contamination provenant de hadrons.

Deux technologies sont utilisées pour former le calorimètre d'ATLAS. On distingue les calorimètres à argon liquide (LAr), qui sont les plus proches du faisceau avec un rayon inférieur à $r = 2$ m (en jaune sur la Figure 2.7), et le calorimètre à tuile (Tile) qui entoure les calorimètres LAr à partir d'un rayon $r = 2.4$ m (en gris sur la Figure 2.7). Le calorimètre LAr et le calorimètre à tuile sont tout deux des calorimètres à échantillonnage. Le premier utilise de l'argon liquide comme milieu actif, choisi pour sa résistance aux radiations et sa réponse linéaire aux dépôts d'énergie par ionisation. Le second est composé de fer jouant le rôle d'absorbeur et de tuiles scintillantes comme matériau actif.

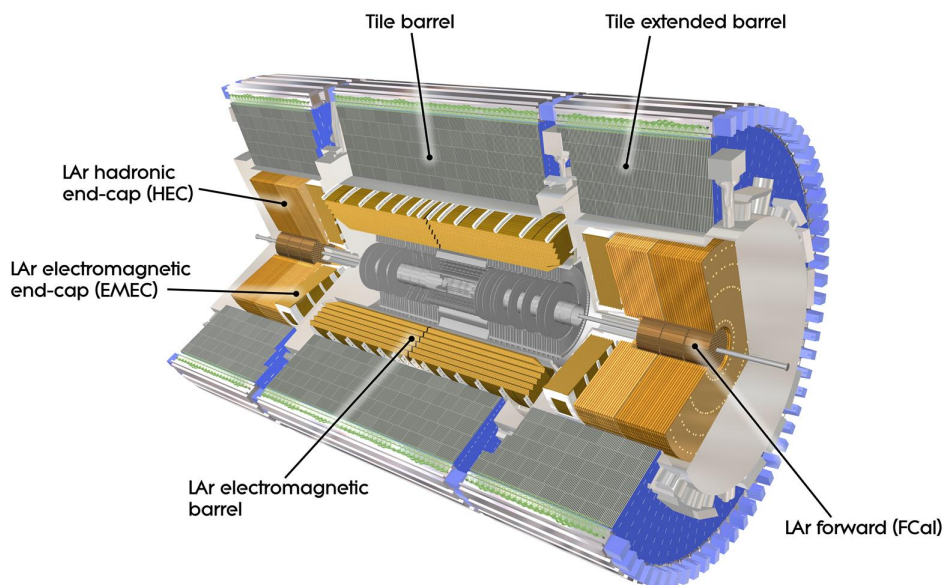


FIGURE 2.7: Vue en coupe du calorimètre d'ATLAS.

Calorimètre à argon liquide

Le calorimètre à argon liquide est utilisé aussi bien comme calorimètre électromagnétique – couvrant un intervalle en pseudo-rapacité $|\eta| < 3.2$ – que comme calorimètre hadronique entre $1.4 < |\eta| < 3.9$. La partie électromagnétique est formée d'un tonneau de $\Delta z = 6.4$ m de long (LAr electromagnetic barrel) et de rayons interne et externe d'environ $r = 1.5$ et $r = 2.0$ m respectivement. Deux roues sont situées de part et d'autre du tonneau, formant les bouchons électromagnétiques (LAr electromagnetic endcap). Les bouchons sont renforcés avec une partie hadronique constituée de deux roues coaxiales dans chaque bouchon (LAr hadronic endcap). Le tonneau et les deux bouchons sont hébergés dans trois cryostats qui maintiennent l'argon à sa température nominale de 89 K. Les rayons externes des bouchons sont alignés avec celui du tonneau. A très haute rapidité, le détecteur le plus à l'avant (Forward) permet de récupérer l'énergie déposées par les particules dans la région $3.1 < |\eta| < 4.9$; cette composante ne sera pas décrite dans cette partie.

L'absorbeur du calorimètre électromagnétique est composé d'une succession de plaques de plomb en forme d'accordéon assemblées radialement, schématisées Figure 2.8a. Le volume entre les plaques est rempli d'argon liquide. Il est divisé en deux par une couche d'électrodes de cuivre-kapton parallèle aux plaques de plomb voisines. Le champ électrique appliqué entre les électrodes et l'absorbeur entraîne une dérive des électrons ionisés dans l'argon liquide par le passage des particules chargées, induisant un courant dont l'amplitude est proportionnelle aux pertes d'énergies dans l'argon liquide. Les électrodes sont segmentées en cellules selon la profondeur en trois couches, et selon l'azimut ϕ et la pseudo-rapacité η .

Une segmentation du tonneau à $|\eta| < 1.4$ est présentée Figure 2.8b. Elle est suffisamment fine en (η, ϕ) pour mesurer le profil latéral des gerbes électromagnétiques. Le calorimètre électromagnétique est conçu de sorte à collecter la plus grande fraction de l'énergie déposée par la gerbe électromagnétique dans la deuxième couche (de profondeur $16X_0$ à $\eta = 0$), tandis que la couche arrière avec la granularité la plus grossière ne recueille que la queue de la gerbe (de profondeur $2X_0$ à $\eta = 0$). La granularité très fine en η de la couche avant permet non seulement de mesurer précisément la direction des gerbes mais aussi d'identifier les paires de photons collimées provenant de désintégrations de π^0 . La granularité angulaire des bouchons électromagnétiques décroît avec la pseudo-rapacité $|\eta|$ mais garantit des mesures de précision jusqu'à $|\eta| < 2.5$.

La calibration en énergie du calorimètre à argon liquide découle de l'étalonnage du signal électronique résultant de l'énergie déposée par des gerbes électromagnétiques mesurées à partir de tests en faisceau, connue sous le nom d'échelle électromagnétique. La résolution du calorimètre électromagnétique à argon liquide obtenue à partir de faisceaux test d'électrons est de $\sigma_E/E(\pi) = 10\%/\sqrt{E} \oplus 0.5\%$. Pour maintenir une réponse constante du calorimètre tout au long des prises de données, la réponse de l'électronique de lecture est calibrée en injectant un signal imitant celui produit par les particules chargées dans l'argon liquide.

Le calorimètre hadronique à argon liquide utilise du cuivre comme absorbeur. Le cuivre est réparti en plaques plates et perpendiculaires aux faisceaux de protons, de $\Delta z = 25$ mm

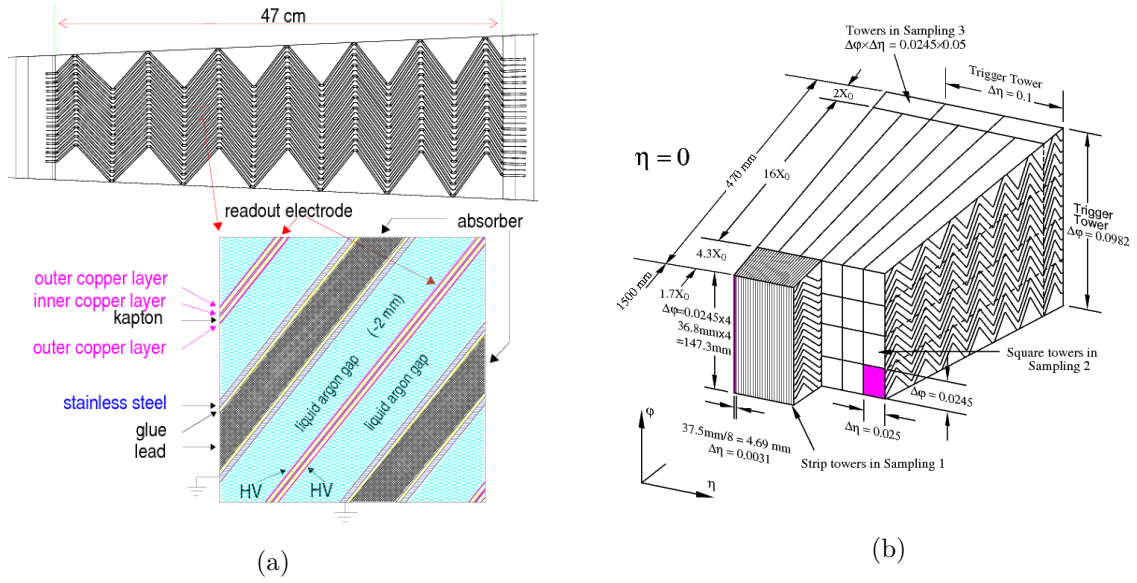


FIGURE 2.8: (a) Vue dans un plan transverse d'une section du calorimètre du tonneau. (b) Schéma d'une coupe du calorimètre électromagnétique du tonneau, permettant de visualiser la granularité en (η, ϕ) de chacune des trois couches en profondeur.

d'épaisseur dans la première roue (située à environ $4\text{m} < |z| < 5\text{m}$) de et $\Delta z = 50\text{ mm}$ d'épaisseur dans la seconde roue ($5\text{m} < |z| < 6\text{m}$). L'écart entre les plaques ($\Delta z = 8.5\text{mm}$) est rempli d'argon liquide (Figure 2.9a), et est divisé en quatre espaces de dérive par trois couches parallèles d'électrodes, dont la configuration empêche une trop importante accumulation d'ions. La segmentation du calorimètre en cellules permet d'obtenir une granularité en (η, ϕ) de 0.1×0.1 pour $1.5 < |\eta| < 2.5$ et de 0.2×0.2 pour $2.5 < |\eta| < 3.2$ (Figure 2.9b). La résolution de l'énergie des jets du calorimètre hadronique à argon liquide, déterminée à l'aide de faisceaux test et de simulations, est d'environ $\sigma_E/E(\pi) = 50\%/\sqrt{E} \oplus 3\%$ [19].

Calorimètre à tuile

Le calorimètre à tuile est un calorimètre hadronique à échantillonnage de forme cylindrique qui entoure les calorimètres LAr. Il est divisé en trois sections selon la direction des faisceaux de protons comme illustré Figure 2.7 : un tonneau central de $\Delta z = 5.6\text{ m}$ de long (Tile barrel) prolongé de part et d'autre par deux extensions de $\Delta z = 2.9\text{ m}$ de longueur chacune (Tile extended barrel), couvrant une région en pseudo-rapacité de $|\eta| < 1.7$. Chaque section est composée de 64 modules (Figure 2.10), chacun couvrant un angle azimutal $\Delta\phi = 2\pi/64 = 0.1$ [20]. Les modules sont constitués de couches alternées de plaques de fer et de scintillateurs en forme de tuile, ou *tuiles scintillantes*. L'agencement des couches permet d'obtenir une longueur d'interaction nucléaire $\lambda = 20.7\text{ cm}$ sur la profondeur du calorimètre à tuile de $\Delta r = 1.6\text{ m}$. Les tuiles scintillantes sont disposées dans le plan perpendiculaire aux faisceaux de collision et sont orientées radialement, comme illustré sur la

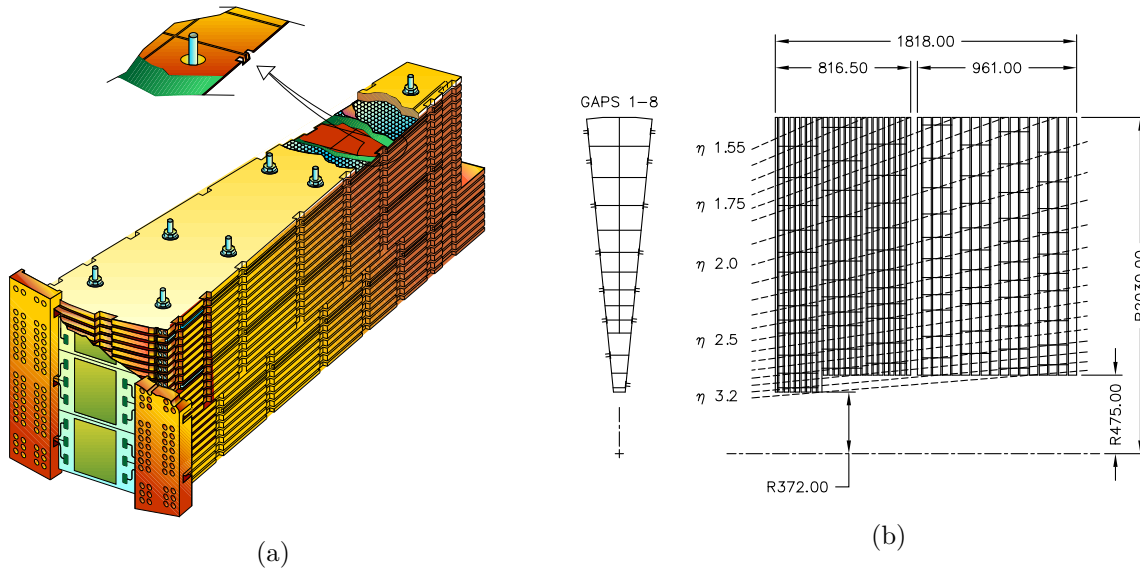


FIGURE 2.9: (a) Vue schématique d'un module du calorimètre hadronique à argon liquide. (b) Vue schématique d'un module du calorimètre selon $r - \phi$ (gauche) et de la structure du calorimètre selon $r - z$ (droite) avec sa division en cellules. Les dimensions sont en mm.

Figure 2.10. Elles sont lues par des fibres à changement de longueur d'onde (wavelength-shifting fibres) des deux côtés de chaque module. Ces fibres guident la lumière jusqu'aux photo-multiplicateurs situés dans la structure de fer au niveau du rayon externe du calorimètre, qui abrite également l'électronique frontale. Chaque paire de photo-multiplicateurs situés de part et d'autre du module rassemble plusieurs fibres pour former une cellule. Les cellules segmentent le calorimètre selon sa profondeur en trois couches, d'épaisseurs 1.5, 4.1 et 1.8 λ dans le tonneau central et 1.5, 2.6 et 3.3 λ dans ses extensions, et selon (η, ϕ) de 0.1×0.1 (0.2×0.1 pour la couche la plus profonde). La segmentation longitudinale est illustrée Figure 2.11. L'orientation innovante des tuiles du calorimètre, parallèlement aux particules entrantes à $\eta = 0$, permet aux fibres de remonter sans détour vers le rayon extérieur, ce qui entraîne une bonne herméticité calorimétrique et une association facile des tuiles avec les fibres.

Des faisceaux test d'électrons ont été utilisés pour calibrer la première couche du calorimètre à tuile, tandis que les deux autres couches ont été intercalibrées par rapport à la première à partir de muons cosmiques. Au cours des prises de données, la réponse des cellules du calorimètre varie avec l'altération des photomultiplicateurs et la dégradation des tuiles scintillantes à cause de leur exposition aux radiations. Pour maintenir la calibration du calorimètre à l'échelle électromagnétique tout au long de la vie du détecteur, trois calibrations sont réalisées à différentes étapes de la chaîne de lecture, illustrées Figure 2.12. L'électronique de lecture est calibrée en lui injectant une charge connue avec le système d'injection de charge (CIS). Ensuite, un système de laser permet d'étalonner la réponse des photo-multiplicateurs en émettant une lumière d'intensité réglable directement sur les

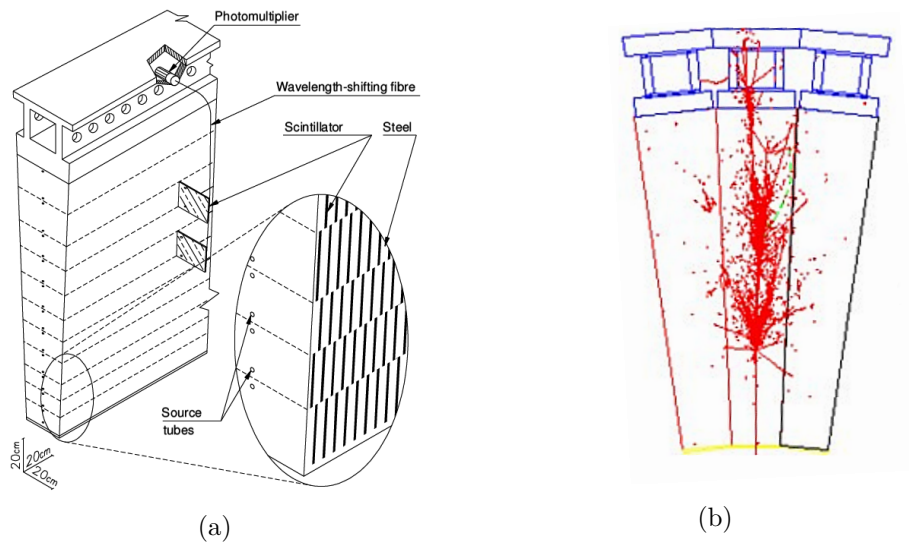


FIGURE 2.10: Schéma d'un module du calorimètre à tuile du détecteur ATLAS (a), et simulation d'une gerbe hadronique dans trois modules du calorimètre dans le plan transverse (b).

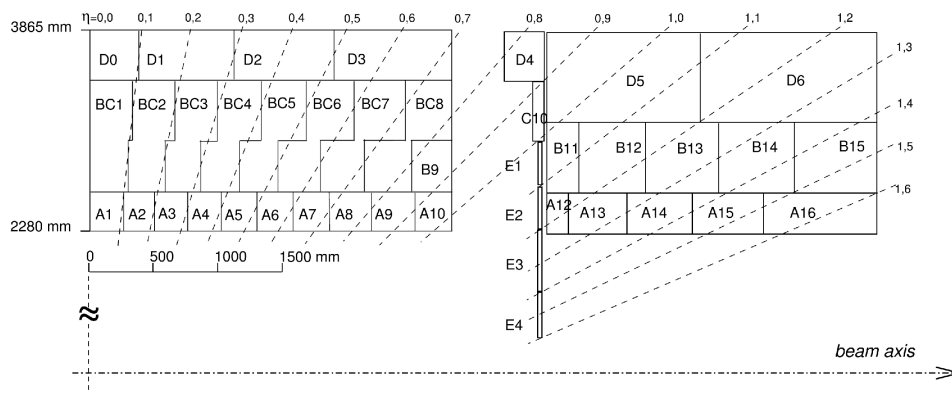


FIGURE 2.11: Segmentation du calorimètre à tuile en cellules dans le plan longitudinal, dans le tonneau central (à gauche) et son extension (à droite).

photomultiplicateurs via des fibres. Enfin, une calibration de la chaîne optique complète formée des tuiles scintillantes, des fibres et des photomultiplicateurs est effectuée à l'aide d'une source de césium ^{137}Cs qui illumine les scintillateurs en émettant des photons de 0,662 MeV [21].

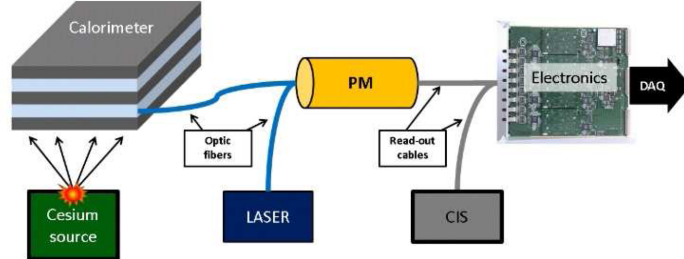


FIGURE 2.12: Chaîne de calibration du calorimètre à tuile.

La résolution sur l'énergie des hadrons uniques du calorimètre à tuile seul est obtenue avec des faisceaux de test de pions. Elle est décrite par $\sigma_E/E(\pi) = 52\%/\sqrt{E} \oplus 5.7\%$ pour un angle d'incidence $\eta = 0.2$, correspondant à une profondeur effective de 7.7λ . La résolution en énergie présente une dépendance avec l'angle d'incidence η , principalement en raison de l'augmentation de la profondeur effective et de la diminution des fuites longitudinales d'énergie lorsque η augmente. Les fuites longitudinales affectent principalement la résolution énergétique des particules d'énergie élevée et, par conséquent, reflète l'augmentation du terme constant [20].

2.2.3 Spectromètre à muon

Le spectromètre à muon est le sous-détecteur d'ATLAS le plus éloigné du faisceau, placé derrière le calorimètre hadronique. Il sert à mesurer l'impulsion des muons qui sont les seules particules avec les neutrinos susceptibles de traverser le calorimètre sans perdre toute leur énergie. Le spectromètre à muon est capable non seulement d'identifier le passage de muons énergétiques pour le système de déclenchement, mais aussi de mesurer précisément leur impulsion jusqu'à des échelles de l'ordre du TeV et des pseudo-rapidités de $|\eta| = 2.7$.

Le spectromètre à muon fournit plusieurs mesures du passage des muons à travers les chambres qui entourent le calorimètre. Les chambres forment, dans le tonneau, trois couches concentriques autour de l'axe du faisceau à des rayons de 5, 7.5 et 10 m environ, et dans chaque bouchon, quatre roues coaxiales perpendiculaires à l'axe du faisceau, à $|z| \sim 7.4, 10.5, 14$ et 21.5 m (Figure 2.13). Les trajectoires des particules chargées sont courbées en sortie du calorimètre dans le plan $r - z$ par un champ magnétique toroïdal. Ce champ est créé par trois ensembles de bobines, dont l'un est situé dans le tonneau et les deux autres dans les bouchons. Chaque ensemble contient huit bobines assemblées radialement autour de l'axe du faisceau (Figure 2.6). Le champ moyen déployé à l'intérieur des systèmes de bobines est de 0.5 T.

L'instrumentation du spectromètre est basée sur deux systèmes de sous-détecteurs : les premiers, appelés MDT et CSC pour Monitored Drift Tube et Cathode Strip Chambers, sont nécessaires pour effectuer des mesures de précisions de l'impulsion des muons. Les seconds, nommés TGC et RPC pour Thin Gap Chambers et Resistive Plate Chambers, ont été choisis pour leur rapidité au détriment d'une résolution plus grossière, et servent de système de déclenchement (Figure 2.13). Les chambres MDT sont formées de tubes à dérive en aluminium de 30 mm de diamètre, qui couvrent un domaine en pseudo-rapacité de $|\eta| < 2.7$. Les CSC sont des chambres proportionnelles multifilaires fonctionnant avec un mélange de gaz Ar/CO_2 (80%/20%). Elles sont placées sur la couche la plus interne de chaque bouchon, couvrant des pseudo-rapidités $2.0 < |\eta| < 2.7$, en raison de leur meilleure résolution en temps servant pour le système de déclenchement. Les mesures du MDT contraignent la trace du muon dans le plan de courbure, tandis que les CSC fournissent des mesures selon $r - \phi$. Ces mesures sont complétées (notamment pour déterminer la composante ϕ dans le MDT) avec des mesures des RPC ($|\eta| < 1.05$) et des TGC $1.05 < |\eta| < 2.4$.

La résolution sur l'impulsion des muons atteinte par les chambres à muon est meilleure que 10% jusqu'à des impulsions de 1 TeV.

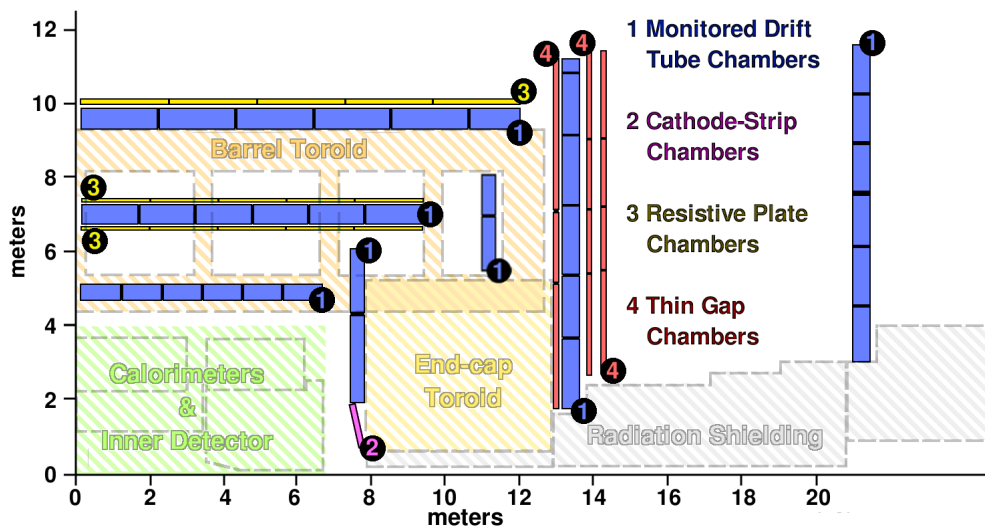


FIGURE 2.13: Coupe dans le plan longitudinal du détecteur ATLAS, présentant les différentes composantes du spectromètre à muon.

2.2.4 Système de déclenchement et d'acquisition

Le taux de données enregistrées par l'expérience ATLAS est limité par la bande passante disponible à environ 1.6 GB/s. Ce seuil ne permet pas d'enregistrer tous les événements produits au LHC. En effet, l'enregistrement des croisements de paquets de protons, avec une fréquence de 40 MHz et une taille d'environ 1 MB, nécessiterait une bande passante bien supérieure. Un système de déclenchement (Trigger) est utilisé pour ne sélectionner que les événements les plus intéressants, qui sont enregistrés pour être analysés par la collaboration.

Au Run 2, le système de déclenchement est composé de deux niveaux. Le premier niveau sélectionne les événements contenant les signatures de haute impulsion transverse provenant : des muons, électrons/photons, jets, des leptons τ se désintégrant en hadrons et de l'énergie transverse manquante, à partir des informations du spectromètre à muon (TGC et RPC) et du calorimètre (une granularité grossière du calorimètre est utilisée à ce niveau). Le temps alloué pour accepter ou non un événement est de $2.5 \mu\text{s}$. Le système de déclenchement de niveau 1 réduit le taux d'événements de 40 MHz à 100 kHz. Lorsqu'un événement passe ce premier niveau, la zone du détecteur susceptible de détecter la signature identifiée au niveau 1, dite *région d'intérêt*, est envoyée au second niveau du système de déclenchement pour guider la sélection. La sélection des événements réalisée au niveau 2 à partir de logiciels ou HLT exploite la granularité complète du détecteur au niveau de la région d'intérêt puis au niveau de l'événement entier. L'identification des signatures à ce niveau est beaucoup plus sophistiquée [22], notamment grâce à une reconstruction des traces, et se rapproche de celle effectuée hors ligne une fois les événements enregistrés, décrite au Chapitre 3. En sortie du niveau 2, le taux d'événement enregistré est de 1 kHz avec un temps alloué d'environ 200 ms.

Les signaux du détecteur sont réceptionnés et acheminés à travers le système de déclenchement jusqu'à l'enregistrement final à l'aide du système d'acquisition (DAQ). La structure du système de déclenchement (Trigger) et d'acquisition de données (DAQ) au Run 2 est illustrée Figure 2.14 [23]. Les données de l'événement sont initialement lues via une électronique spécifique appelée front-end. Ces systèmes effectuent une mise en forme initiale des signaux électroniques, une conversion analogique-numérique et une agrégation des signaux reçus par les capteurs des détecteurs. Des portions de ces données associées au calorimètre et au spectromètre à muon sont ensuite traitées par le système de déclenchement de niveau 1. Si un événement passe la sélection du niveau 1, un signal est renvoyé au front-end, ce qui permet de lire les données associées à l'événement pour toutes les composantes du détecteur. Ces données sont alors transmises grâce aux pilotes de lecture (ROD) jusqu'au système de lecture (ROS) via des fibres optiques. Une fois la décision au niveau 2 prise, l'événement est écrit pour un stockage permanent.

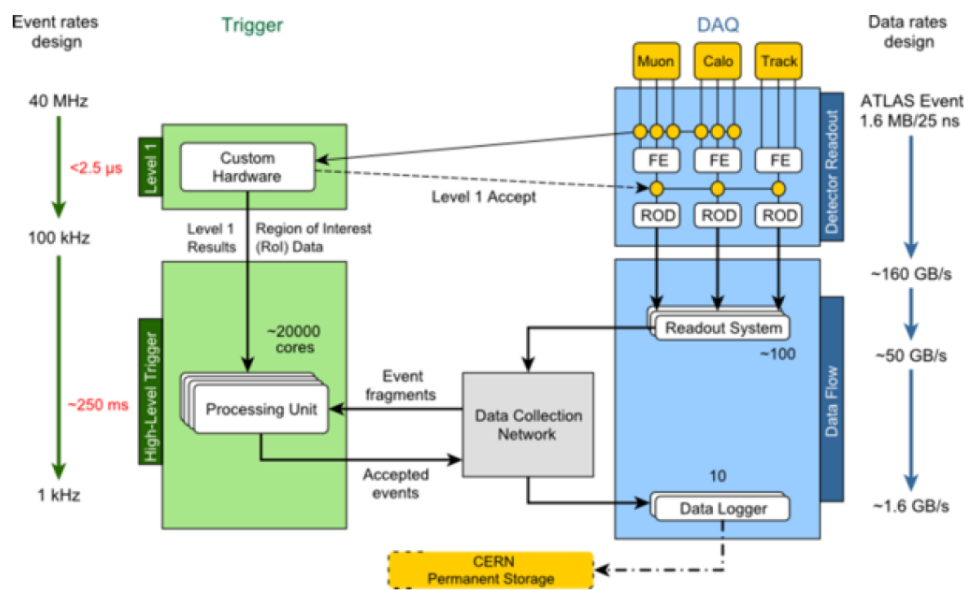


FIGURE 2.14: Diagramme du système de déclenchement et d'acquisition d'ATLAS au Run 2. [23]

Bibliographie

- [11] Grand collisionneur électron-positon (LEP). <https://cds.cern.ch/record/1997707>, Jan 2013.
- [12] Ions lourds et plasma de quarks et de gluons. <https://cds.cern.ch/record/1997843>, Mar 2013.
- [13] Le complexe d'accélérateurs. <https://cds.cern.ch/record/1997501>, Nov 2012.
- [14] Roderik Bruce, Gianluigi Arduini, Hannes Bartosik, Riccardo De Maria, Massimo Giovannozzi, Giovanni Iadarola, John Jowett, Kevin Shing Bruce Li, Mike Lamont, Anton Lechner, Elias Metral, Daniele Mirarchi, Tatiana Pieloni, Stefano Redaelli, Giovanni Rumolo, Benoit Salvant, Rogelio Tomas Garcia, and Jorg Wenninger. LHC Run 2 : Results and Challenges . Technical Report CERN-ACC-2016-0103, CERN, Geneva, Jul 2016.
- [15] by Corinne Pralavorio. 2016, une année exceptionnelle pour le LHC . <https://cds.cern.ch/record/2240062>, Dec 2016.
- [16] M. Aaboud et al. Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC. *Phys. Rev. Lett.*, 117(18) :182002, 2016.
- [17] A. Airapetian and al. *ATLAS detector and physics performance : Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1999.
- [18] ATLAS Inner Detector Alignment Performance with February 2015 Cosmic Rays Data. Technical Report ATL-PHYS-PUB-2015-009, CERN, Geneva, Apr 2015.
- [19] M Aleksa and al. Construction, assembly and tests of the ATLAS electromagnetic end-cap calorimeters. *JINST*, 3 :P06002, 2008.
- [20] Ana Maria Henriques Correia. The ATLAS Tile Calorimeter. Technical Report ATL-TILECAL-PROC-2015-002, CERN, Geneva, Mar 2015.
- [21] Alexander Solodkov. Performance of the ATLAS Tile Calorimeter. Technical Report ATL-TILECAL-PROC-2015-024, CERN, Geneva, Dec 2015.
- [22] A. Ruiz Martínez. The Run-2 ATLAS Trigger System. *J. Phys. Conf. Ser.*, 762(1) :012003, 2016.
- [23] J G Panduro Vazquez. The ATLAS Data Acquisition System : from Run 1 to Run 2. Technical Report ATL-DAQ-PROC-2014-035, CERN, Geneva, Oct 2014.

Chapitre 3

La reconstruction des objets et la simulation

Lorsqu'un événement est sélectionné par le système de déclenchement du détecteur ATLAS, il est enregistré de manière permanente pour être analysé. En partant des données brutes, constituées des signaux électriques du détecteur, l'événement est interprété en terme d'*objets physiques* dans un processus appelé *reconstruction d'événements*. Ces objets physiques permettent d'identifier les particules produites aux premiers instants de la collision dure, comme les photons, les électrons, les muons, etc, et par conséquent de déduire les processus physiques mis en jeux.

Les objets physiques de premier plan de l'analyse $t\bar{t}H$ dans le canal *lepton + jet* sont décrits dans ce chapitre. La paire $t\bar{t}$ des événements $t\bar{t}H$ est recherchée dans sa désintégration semi-leptonique, émettant un muon ou un électron, quatre quarks dont deux de saveur b . Le boson de Higgs est recherché dans sa désintégration en paire de quarks $b\bar{b}$. Les performances de l'analyse reposent sur les propriétés des candidats électron (cf. Section 3.3), muon (cf. Section 3.4), des jets (cf. Section 3.5), des grands jets (cf. Section 3.6) qui sont exposés dans les sections suivantes, et des candidats b qui sont présentés au chapitre 4 dédié aux études sur l'étiquetage de jets issus de l'hadronisation de quarks b , dits *jets de b* . Les traces (cf. Section 3.1) et les vertex primaires (cf. Section 3.2) sont également définis, car ils interviennent dans la description des objets mentionnés ci-dessus. Malgré la présence de neutrino dans le signal, l'énergie transverse n'est pas considérée directement dans l'analyse et n'est pas détaillée dans ce chapitre.

3.1 Trace

Les trajectoires des particules chargées produites dans les événements d'intérêt sont reconstruites en traces à partir de leurs dépôts d'énergie mesurés dans le Détecteur Interne.

La reconstruction des traces commence avec le rassemblement des dépôts d'énergie dans les détecteurs à silicium (Pixel et SCT) en amas, compatibles avec l'interaction d'une particule chargée avec une couche sensible des détecteurs à silicium. Les positions de ces amas, ou *coups*, servent de points de départ à la formation des traces. Ces 'début de trace' sont ensuite extrapolés jusqu'au TRT avec la méthode du filtre de Kalman [24]. Les traces qui ne proviennent pas de particules chargées, appelées *fausses traces*, sont rejetées en appliquant des critères liés à la qualité de la reconstruction des traces. Par exemple, chaque nouvel amas associé à la trace renforce sa qualité, alors que les trous la dégrade. Le χ^2 du fit de la trace est également considéré pour pénaliser les traces mal ajustées. Finalement, considérer l'impulsion transverse des traces permet de promouvoir les traces de haute impulsion et de supprimer le grand nombre de traces de basse impulsion transverse assignées de manière incorrecte à des amas.

La reconstruction des traces est conçue pour identifier les particules chargées avec une grande efficacité. L'efficacité typique de reconstruction de trace en fonction de leur impulsion transverse ou de leur pseudo-rapacité η_{trace} sont présentées Figure 3.1 [25], pour des traces de collisions inélastiques pp simulées; l'efficacité de reconstruction étant définie ici comme la fraction des particules chargées issues de collisions pp qui sont reconstruites en

traces. La reconstruction des particules chargées en traces est affectée par la quantité de matière traversée, et l'efficacité de reconstruction décroît avec la pseudo-rapidité de la trace, notamment dans la région $|\eta_{trace}| > 1$ car les particules passent à travers davantage de matière. La légère augmentation d'efficacité à $|\eta_{trace}| \sim 2$ est due aux particules traversant un nombre croissant de couches dans les bouchons. Les particules de hautes impulsions sont plus facilement reconstruites notamment parce que leur trajectoire est moins déviée par diffusion multiple lors de leur passage dans la matière (Figure 3.1).

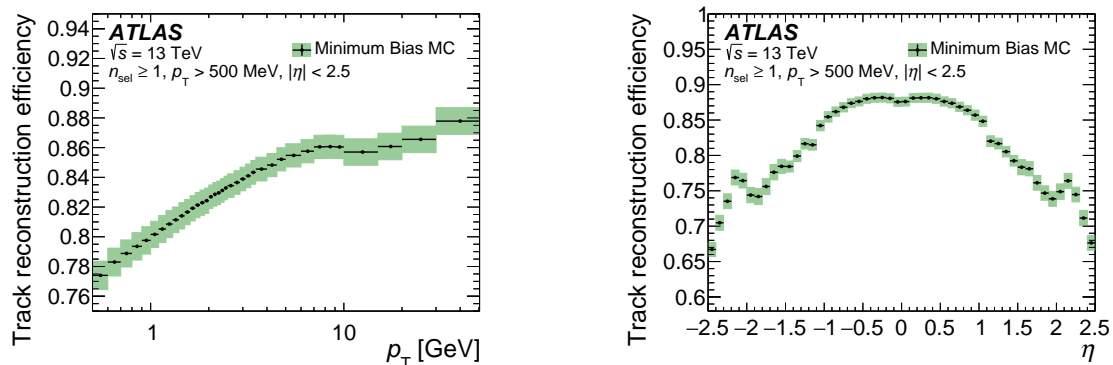


FIGURE 3.1: Efficacité de reconstruction de trace en fonction de l'impulsion transverse (a) et de la pseudo-rapidité (b) de ces traces, estimée à partir d'un échantillon de collision inélastiques, simulé avec Pythia 8 A2. Les incertitudes statistiques sont représentées par des barres verticales noires tandis que les incertitudes totales sont données par les bandes vertes [25].

L'ajout de l'IBL au Run 2 a eu un impact majeur sur la reconstruction des traces, notamment sur leur résolution spatiale, qui se traduit par une amélioration de la résolution sur les paramètres d'impact d_0 et z_0 . Cette amélioration est plus prononcée pour les traces de basse impulsion transverse (Figure 3.2) [26].

3.2 Vertex primaire

Lors d'un croisement de paquets de protons, les traces de particules produites au point de la collision pp sont accompagnées de traces de particules secondaires ou de traces originaires de collisions d'empilement. Afin d'identifier ces particules chargées qui ne proviennent pas de la collision d'intérêt, il est essentiel de connaître avec précision la position de cette collision.

Les traces produites dans une collision inélastique ont une origine commune, appelée vertex primaire. La reconstruction de ces vertex primaires est un procédé itératif, qui débute avec la recherche d'une intersection de traces formant un premier vertex primaire. Les traces dont l'origine n'est pas compatible avec la position fittée du vertex primaire sont utilisées pour la recherche d'autres vertex primaires, et ainsi de suite [27]. La recherche de vertex primaires est privilégiée dans la région de croisement de paquets de protons, ou *région*

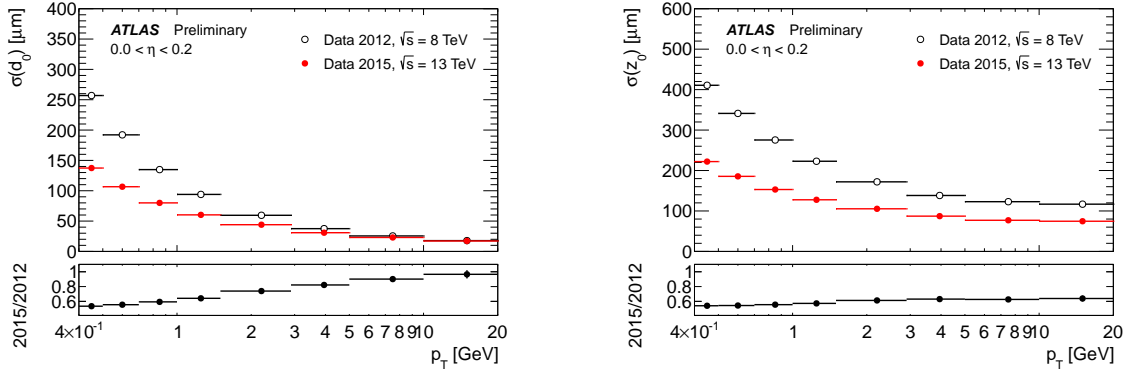


FIGURE 3.2: Résolution des paramètres d’impact transversal (gauche) et longitudinal (droite) mesurés à partir des données en 2015 à $\sqrt{s} = 13$ TeV, avec le Détecteur Interne comprenant l’IBL, en fonction de p_T , pour des valeurs de $0.0 < |\eta_{trace}| < 0.2$. Ces résolutions sont comparées à celles mesurées à partir des données en 2012, à $\sqrt{s} = 8$ TeV. Les données en 2015 sont collectées grâce à un déclenchement sur des collisions inélastiques. Les données en 2012 proviennent d’un mélange de déclenchements sur les jets, sur les τ et sur l’énergie transverse manquante [26].

lumineuse [28]. Ses dimensions dans le plan transverse¹ et longitudinal sont présentées pour des données 2015 sur la Figure 3.3 [29]. Dans le plan transverse, la résolution σ_r spatiale du vertex primaire est comparable avec celle de la région lumineuse autour de $\sim 20 \mu\text{m}$. La région lumineuse est en revanche bien plus étalée dans le plan longitudinal, avec une taille d’environ 50 mm, et ne permet pas de contraindre la résolution longitudinale σ_z du vertex primaire. Cette composante est donc plus sensible au nombre de traces appartenant au vertex primaire. Pour des événements $t\bar{t}$ qui peuvent produire plusieurs dizaines de traces, la résolution σ_z obtenue est d’environ 50 μm .

Les collisions inélastiques d’un même croisement de paquet sont séparées d’environ 2 mm en moyenne pour un empilement de 25 collisions par croisement de paquet, et la résolution longitudinale des vertex primaires permet de les reconstruire séparément. Parmi les vertex primaires candidats pour un même événement, un unique vertex primaire est défini. Dans l’analyse $t\bar{t}H$ comme pour de nombreuses analyses, le vertex primaire est choisi comme étant celui qui possède la somme en quadrature des impulsions transverses des traces associées la plus élevée. Sur la base de cette définition, la fraction de vertex primaires qui proviennent de la collision d’intérêt, ou efficacité de vertex primaire, est présentée Figure 3.4 [30] en fonction du nombre moyen de collisions d’empilement pour trois processus physiques représentatifs : $t\bar{t}$, $Z \rightarrow \mu\mu$, et $H \rightarrow \gamma\gamma$. Les meilleures efficacités sont obtenues pour le processus $t\bar{t}$, qui produit un grand nombre de particules chargées de haute impulsion transverse, contrairement au processus $H \rightarrow \gamma\gamma$. De plus, avec l’augmentation du nombre de collisions d’empilement,

1. La dispersion de la région lumineuse est en très bonne approximation isotrope dans le plan transverse et est présentée selon l’axe x .

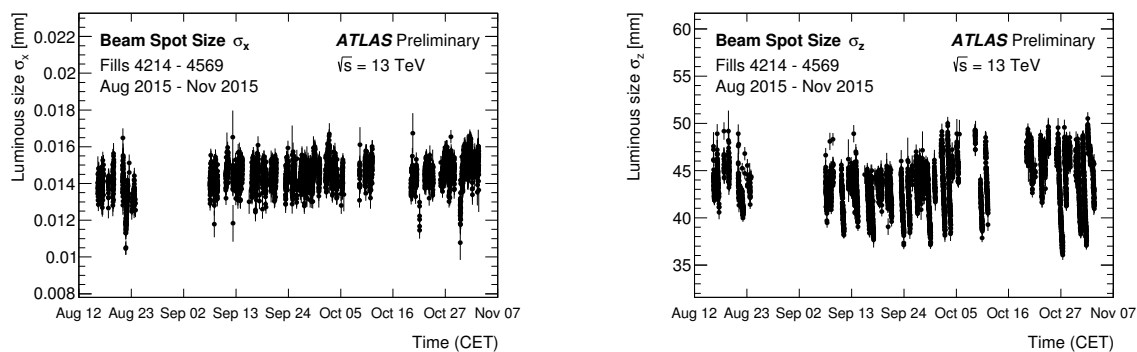


FIGURE 3.3: La taille de la région lumineuse selon l'axe x , à gauche, et l'axe z , à droite, dans ATLAS pour des collisions inélastiques pp enregistrées en 2015 à $\sqrt{s} = 13$ TeV. Les points de données sont le résultat d'un fit de maximum de vraisemblance de la distribution spatiale des vertex primaires recueillis sur cinq minutes. Les erreurs sont uniquement statistiques [29].

il est plus probable de choisir un vertex primaire provenant d'une collision d'empilement.

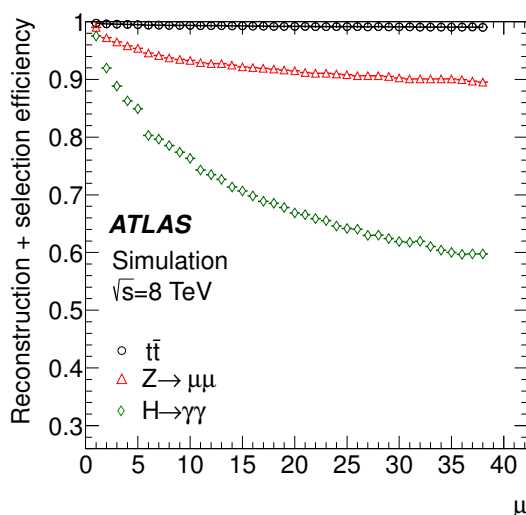


FIGURE 3.4: Efficacité de reconstruire et de sélectionner le vertex primaire de la collision d'intérêt en fonction du nombre moyen de collisions pp par croisement de paquets de protons, pour des événements reconstruits simulés $t\bar{t}$, $Z \rightarrow \mu\mu$, et $H \rightarrow \gamma\gamma$ [30].

3.3 Candidat électron

Un candidat électron est reconstruit dans le détecteur à partir d'un amas d'énergie déposée dans le calorimètre électromagnétique, compatible avec le développement d'une gerbe électromagnétique, et d'une trace joignant le vertex primaire à l'amas calorimétrique (Figure 3.5). Cette trace associée, reconstruite avec les coups du Détecteur Interne, tient

compte de possibles émissions de photons par bremsstrahlung.

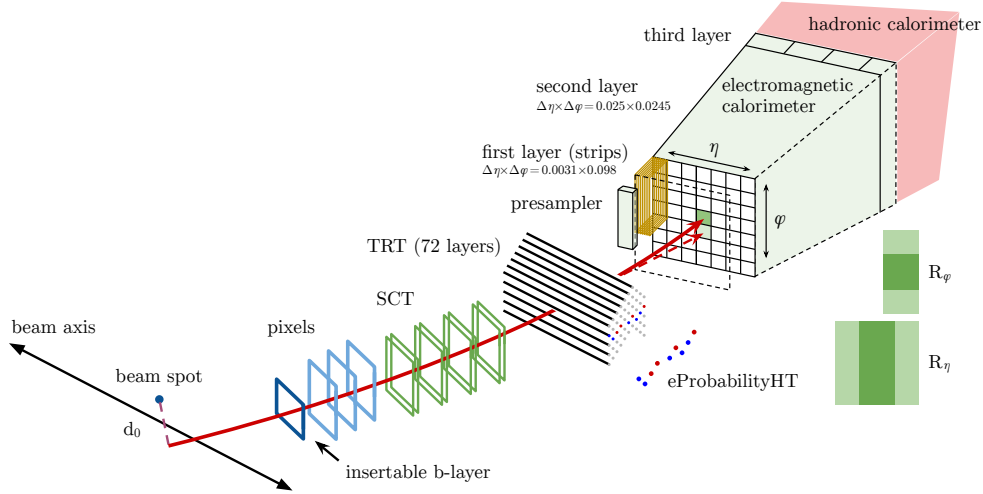


FIGURE 3.5: Vue schématique de la reconstruction d'un électron [31].

Les propriétés du candidat électron reconstruit sont ensuite exploitées pour rejeter principalement les candidats qui ne proviennent pas d'électrons, appelés *faux* candidats électron (les *vrais* candidats sont quant à eux associés à un électron) [31]. Par exemple, la forme de l'amas d'énergie du calorimètre et sa position permettent de supprimer des faux candidats électron issus de hadrons chargés, dont la gerbe hadronique est moins dense en énergie et située plus en profondeur dans le calorimètre que la gerbe formée par un électron. Le TRT fournit également une capacité d'identification d'électrons grâce à la détection de photons à rayons X. Au final, les informations du candidat électron composé de son amas calorimétrique et de sa trace, ses caractéristiques cinématiques (p_T , η , ϕ) et la qualité de l'association amas-trace sont combinées avec une méthode de vraisemblance pour distinguer au mieux les vrais des faux candidats électron. Pour des candidats électron d'impulsion transverse supérieure à une vingtaine de GeV, la méthode de vraisemblance est optimisée dans la simulation pour séparer les vrais candidats provenant des désintégrations $Z \rightarrow ee$ des faux candidats issus d'événements *dijet* (correspondant deux jets de haute impulsion émis dos à dos).

Trois niveaux d'identification sont définis, appelés *Loose*, *Medium*, *Tight*, ordonnés selon leur efficacité à rejeter le bruit de fond, au détriment d'une perte d'efficacité d'identification des vrais candidats électron. Leur efficacité d'identification, définies comme la fraction de candidats électron reconstruit qui passent un critère d'identification donné, sont présentées Figure 3.6 [31] pour les critères d'identification *Loose*, *Medium*, *Tight* dans des événements candidat $Z \rightarrow ee$. Les performances d'identification des trois critères s'améliorent en général avec l'énergie du candidat électron reconstruit, avec une augmentation de l'efficacité d'identification des vrais candidats, qui dépasse les 90% pour des candidats *Medium* de $p_T > 40$ GeV, et une diminution de l'efficacité d'identification des faux candidats, qui passe en dessous des 0.3% pour ces mêmes critères d'identification.

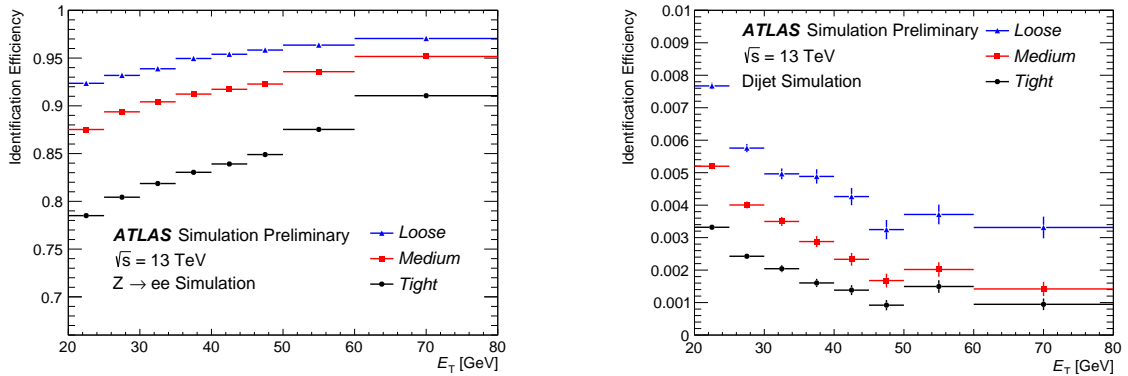


FIGURE 3.6: L’efficacité d’identification des vrais candidats électron provenant des désintégrations $Z \rightarrow ee$ (à gauche) et l’efficacité d’identification de faux candidats électron produits dans les événements *dijet* (à droite) sont présentées en fonction de l’impulsion transverse du candidat à l’aide de simulation. Les efficacités sont mesurées pour les critères d’identification *Loose*, *Medium*, *Tight* [31].

Les analyses qui recherchent des électrons provenant de désintégrations de particules lourdes telles que les bosons Z et W demandent aux candidats électron d’être isolés d’une activité hadronique. L’isolation permet de rejeter les candidats électron issus de photons convertis produits dans la désintégration faible de hadrons, d’électrons provenant de désintégrations faibles de hadrons ou bien de hadrons légers. L’isolation est exprimée en terme d’énergie déposée dans le calorimètre au voisinage de l’amas calorimétrique du candidat et comme la somme des impulsions transverses des traces voisines de la trace de l’électron. Pour l’isolation *Gradient* utilisée dans les analyses $t\bar{t}H$, ces quantités sont combinées pour isoler les candidats électron en fonction de leur énergie transverse E_T . Les efficacités d’isolation, définies comme les fractions de candidats électron identifiés qui passent le critère d’isolation *Gradient*, sont typiquement de 90 (99) % pour $E_T = 25$ (60) GeV.

Afin d’obtenir des résultats fiables, les simulations sont corrigées pour reproduire les efficacités de reconstruction, d’identification et d’isolation mesurées avec les données. Ces calibrations des efficacités sont fondées sur la méthode dite de *tag-and-probe* à l’aide des résonances Z et J/ψ , nécessitant la présence d’un électron identifié et isolé comme *tag* alors que l’autre sert à la calibration. Les mesures observées dans les données sont comparées aux simulations pour obtenir des corrections (ou facteurs d’échelle) en fonction de l’énergie transverse du candidat électron E_T et de sa pseudorapidité η .

L’énergie du candidat électron est également calibrée [32]. La calibration est effectuée en deux grandes étapes. La première calibration, estimée dans la simulation, corrige l’énergie des candidats électron afin que l’énergie des vrais candidats corresponde à celle de leur électron associé. Cette correction dépend des propriétés du candidat électron. Cette étape permet notamment de corriger des pertes d’énergie de l’électron dans les matériaux passifs situés en amont du calorimètre, non prises en compte dans la reconstruction du candi-

dat électron. Après cette calibration, la réponse en énergie des candidats électron est plus uniforme en fonction de la pseudo-rapidité du candidat. La seconde étape de calibration corrige dans la simulation l'échelle d'énergie du candidat électron et sa résolution à partir d'événements observés de candidats $Z \rightarrow ee$.

3.4 Candidat muon

Le candidat muon utilisé dans l'analyse $t\bar{t}H(b\bar{b})$ et dans de nombreuses analyses ATLAS est reconstruit en ajustant une trace reconstruite à partir de segments (de traces) dans les différentes couches du spectromètre avec une trace du Détecteur Interne. Deux candidats muon typiques, vraisemblablement produits par la désintégration d'un boson Z , sont représentés Figure 3.7 [33] au sein du détecteur ATLAS. Les propriétés des candidats muon

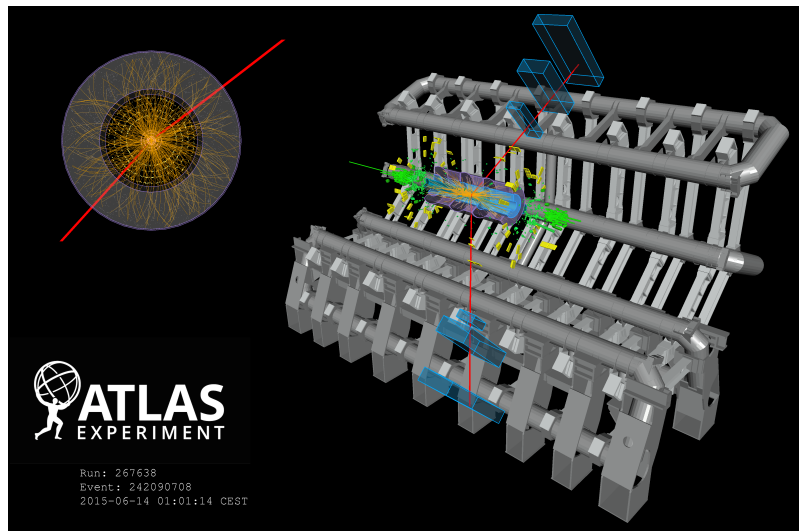


FIGURE 3.7: Représentation d'un événement candidat $Z \rightarrow \mu\mu$ ($m_{\mu\mu} = 90.2$ GeV) de collisions pp à 13 TeV enregistrées par ATLAS en 2015. Les lignes rouges montrent les chemins de deux candidats muon à travers le détecteur, avec des impulsions transverses d'environ 45 GeV. Les barres vertes et jaunes indiquent les dépôts d'énergie dans les calorimètres électromagnétique et hadronique. Les traces de particules chargées reconstruites à partir des impacts (coups) dans le Détecteur Interne sont représentées sous forme d'arcs oranges, se courbant sous l'effet du champ magnétique solénoïdal [33].

reconstruits sont ensuite utilisées pour rejeter ceux originaires de hadrons se désintégrant en vol, comme par exemple les Kaons chargés. Les traces de ces candidats muon reconstruits sont caractérisées par la présence d'une forte courbure au niveau de la désintégration du hadron. Par conséquent, la qualité du fit de leur trace et la compatibilité de leurs mesures d'impulsion dans le Détecteur Interne et le spectromètre à muon sont dégradées.

Des critères de qualité sont imposés sur les impacts utilisés pour former la trace du muon, afin de garantir une mesure robuste de son impulsion [34]. Par exemple, la trace du muon

doit contenir au moins cinq impacts du SCT, et moins de trois *trous* dans le Pixel/SCT ; un trou étant défini comme un senseur actif traversé par la trace sans qu'il n'ait pas enregistré de coup. Plusieurs niveaux d'identification de muon fondés sur les propriétés du muon reconstruit sont disponibles : *Loose*, *Medium* et *Tight*, classés par efficacité d'identification décroissante. Le candidat de l'analyse $t\bar{t}H(b\bar{b})$ est identifié avec le critère *Medium*, pour une pseudo-rapidité $|\eta| < 2.5$. Pour les candidats *Medium* provenant d'un échantillon $t\bar{t}$, d'impulsion transverse comprise entre 20 et 100 GeV, l'efficacité d'identifier un muon issu de la désintégration d'un boson W est de 96% pour une efficacité d'identifier, à la place, un hadron se désintégrant en vol de 0.17%.

Les candidats muon originaires de muons venant de la désintégration de bosons Z et W sont souvent isolés. L'activité autour de la direction de la trace du muon au niveau du vertex primaire, mesurée à l'aide des traces ou des dépôts calorimétriques, permet de quantifier le niveau d'isolation du muon, et ainsi de rejeter les candidats muon provenant de désintégrations de hadrons c ou b qui sont entourés d'une forte activité. Les candidats muon de l'analyse $t\bar{t}H(b\bar{b})$ sont sélectionnés avec le critère d'isolation *Gradient* dont l'efficacité d'isolation, calculée à partir de candidats muon provenant de la désintégration de bosons Z , varie en fonction de l'impulsion transverse du muon pour atteindre 90 (99)% à 25 (60) GeV. L'efficacité d'isolation est connue avec une précision au niveau de 0.2% pour les muons d'impulsion transverse supérieure à 20 GeV.

Les efficacités d'identification, d'isolation du muon ainsi que son échelle d'impulsion et sa résolution en impulsion prédites par la simulation sont calibrées avec des candidats provenant des processus $J/\psi \rightarrow \mu\mu$ et $Z \rightarrow \mu\mu$ observés dans les données. Les candidats muon de haute impulsion transverse (sélectionnés dans l'analyse $t\bar{t}H(b\bar{b})$) sont principalement calibrés grâce au second processus physique. Les mesures d'efficacité sont réalisées avec la méthode *tag-and-probe*. La calibration de l'impulsion du candidat muon est essentielle pour la précision des mesures réalisées dans ATLAS, et son impact sur la distribution de la masse invariante dimuon pour des événements candidats $Z \rightarrow \mu\mu$ est présenté Figure 3.8 [34]. L'impulsion corrigée du muon est obtenue en combinant les impulsions corrigées de la trace correspondante mesurées dans le Détecteur Interne et dans le spectromètre à muon. Les distributions simulées de masses invariantes dimuon calculées avec ces deux impulsions sont fittées et comparées aux données pour extraire les paramètres qui corrigent leur échelle d'impulsion et leur résolution. La calibration de l'échelle d'impulsion corrige de l'incertitude sur la description du champ magnétique et sur les dimensions du détecteur dans la simulation. L'incertitude sur la résolution dans la simulation vient de la difficulté de modéliser les différentes sources de fluctuations dans le détecteur. La résolution de l'impulsion transverse du muon peut se décomposer en une somme quadratique de trois termes :

$$\frac{\sigma(p_T)}{p_T} = r_0/p_T \oplus r_1 \oplus r_2 \cdot p_T. \quad (3.1)$$

Le premier terme tient principalement compte des fluctuations des pertes d'énergie dans la matière traversée (négligeable devant les autres termes), le second terme provient de la diffusion multiple, des inhomogénéités du champ magnétique et des déplacements locaux

des impacts dans le trajectographe. Finalement, le troisième terme est dû à la résolution spatiale de la mesure des impacts et de l’alignement du détecteur.

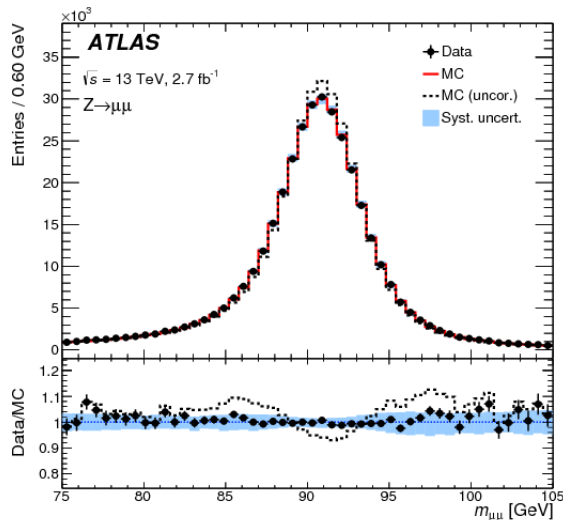


FIGURE 3.8: Distribution de la masse invariante dimuon d’événements candidat $Z \rightarrow \mu\mu$ pour des muons reconstruits avec la combinaison du Détecteur Interne et du spectromètre à muon. La figure supérieure montre la distribution de masse invariante pour les données (points) et pour la simulation avec et sans calibration de l’impulsion. La figure inférieure montre le rapport des distributions dans les données et de la simulation. La bande bleue représente l’incertitude systématique du rapport liée aux corrections d’impulsion [34].

Pour des désintégrations $Z \rightarrow \mu\mu$, l’incertitude sur l’échelle d’impulsion varie d’un minimum de 0.05% à $|\eta| < 1.0$ jusqu’à 0.3% pour $|\eta| \sim 2.5$. Pour ces candidats muon, la résolution en impulsion transverse est de 2.3% à faibles pseudo-rapidités et atteint 2.9% dans les bouchons. Après calibration de l’impulsion des muons, la résolution en impulsion transverse dans les données et les simulations sont en accord à mieux que 5%.

3.5 Jet

Un jet est un ensemble de particules émises autour d’une même direction. Dans l’expérience ATLAS, la reconstruction des jets a pour objectif de capturer les produits de l’hadronisation de partons émis dans les collisions dures avec une haute impulsion transverse, typiquement supérieure à 20 GeV.

Les jets standard sont reconstruits à partir des dépôts calorimétriques. D’abord, les cellules voisines du calorimètre, dont l’énergie est significativement supérieure au bruit provenant des collisions d’empilement ou de l’électronique, sont regroupées en amas *topologiques* compatibles avec des gerbes de particules secondaires [35]. Ensuite les amas topologiques sont rassemblés en jets à l’aide de l’algorithme anti- k_t de rayon $R = 0.4$ [36]. En sortie de

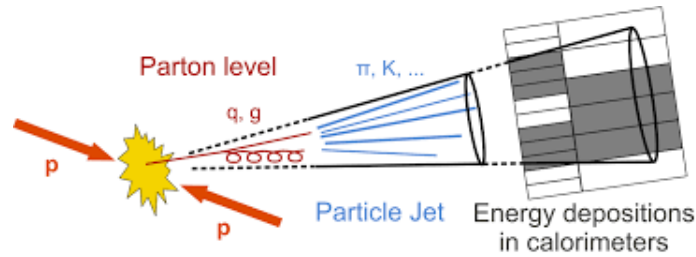


FIGURE 3.9: Schéma représentant un même jet à plusieurs instants après la collision pp dure : avant hadronisation ($\sim 10^{-22}$ s) en terme de partons, puis en terme de particules relativement stables ($\sim 10^{-11}$ s) qui n'ont pas encore atteint le calorimètre, et en enfin en terme d'énergie déposée par les particules interagissant avec le calorimètre ($\sim 10^{-8}$ s).

l'algorithme, les paires de jets ne peuvent être séparés à moins de $\Delta R(\text{jet}, \text{jet}) = 0.4$. Les jets d'impulsion transverse inférieure à 20 GeV et de pseudo-rapacité supérieure à $|\eta| = 4.5$ sont rejetés. De plus, des critères de qualité sont imposés pour supprimer les jets ne provenant pas de la collision dure mais, par exemple, de l'interaction du faisceau de proton avec le gaz du tube à faisceau, du bruit du calorimètre ou encore des collisions pp sous-jacentes [37]. Un des critères de qualité utilisé est lié au nombre de traces associées au jet qui sont originaires du vertex primaire.

L'énergie finale des jets est le résultat d'une succession de calibrations [38]. Tout d'abord, les amas topologiques sont classés selon leur aspect *électromagnétique* ou *hadronique* à l'aide de leur forme, de leur profondeur et de leur densité d'énergie. Sur la base de cette classification, des corrections d'énergie sont dérivées de simulations Monte Carlo de pions. Elles permettent de corriger des effets de non-compensation, des pertes de signal en raison des effets de seuil de bruit et de l'énergie perdue dans les régions non instrumentées. L'étalonnage final de l'énergie du jet (JES : *Jet Energy Scale*) relie la réponse calorimétrique à la véritable énergie du jet. Pour estimer cette correction, des *vrais* jets sont reconstruit *au niveau particule* en regroupant les particules simulées de temps de vie supérieur à 10 ps, sauf les muons et les neutrinos, à l'aide du même algorithme utilisé pour reconstruire les jets (anti- k_T de rayon $R = 0.4$) (Figure 3.9). L'énergie vraie d'un jet est donnée par l'énergie du *vrai* jet qui lui est associé ($\Delta R(\text{jet}, \text{vrai jet}) < 0.3$). Ces calibrations corrigent des effets de pertes d'énergie liées aux matériaux passifs, aux particules du *vrai* jet qui se retrouvent hors du jet ou aux dépôts d'énergie non reconstruits en amas topologiques. Enfin, une correction supplémentaire est appliquée pour corriger la résolution en énergie des jets mesurés dans la simulation par rapport à celle des données (JER : *Jet Energy Resolution*).

3.6 Grand jet

Les grands jets sont l'équivalent des jets standard décrits dans la section précédente, mais ils sont reconstruits pour capturer les produits de désintégrations hadroniques de particules

lourdes, comme les bosons Z , W , de Higgs ou le quark top boostés (Figure 3.10), c'est-à-dire d'impulsion transverse grande devant leur masse [39]. L'angle d'ouverture du grand jet est plus important que celui des jets standard.



FIGURE 3.10: Vue schématique de la désintégration hadronique d'un quark top $t \rightarrow bq\bar{q}'$, dans un cas non boosté où les quarks $bq\bar{q}'$ sont reconstruits dans des jets standard (à gauche), et dans un cas boosté où ces mêmes quarks sont reconstruits dans un même grand jet (à droite).

Comme pour les jets standard, les grands jets sont formés d'amas topologiques rassemblés à l'aide de l'algorithme anti- k_t , de rayon $R = 1.0$. Ils sont ensuite *élagués* [40] pour réduire la contamination provenant de l'énergie déposée par les collisions d'empilement. La procédure d'*élaguage* retire les dépôts d'énergie les moins denses à l'intérieur du grand jet. Les amas d'énergie topologiques à l'intérieur du grand jet sont dans un premier temps regroupés avec un algorithme k_t [41] de rayon $R_{sub} = 0.2$. Les sous-jets obtenus dont l'impulsion transverse est inférieure à $f_{cut} = 5\%$ de celle du grand jet parent sont ensuite supprimés. Les propriétés du grand jet utilisées par les analyses comme l'impulsion et la masse sont calculées à partir des sous-jets restants.

L'échelle d'énergie des grands jets et leur résolution en énergie sont aussi calibrées de manière analogue à ce qui est fait pour les jets standard. Etant donné l'importance de la masse des grands jets pour identifier les particules lourdes, une dernière étape est ajoutée à la procédure d'étalonnage dans laquelle la masse du jet [42] est calibrée en fonction de la simulation MC d'événements *dijet*. Un étalonnage explicite de la masse du jet est important, car cette masse est particulièrement sensible aux faibles dépôts d'énergie émis à grands angles par rapport à l'axe du grand jet.

Plusieurs variables de sous-structures sont utilisées dans l'analyse boostée pour décrire la forme des dépôts d'énergie à l'intérieur des grands jets. La variable τ_{32} joue un rôle important dans l'identification de quark top hadronique boosté. Elle est définie comme le ratio de variables 'N-subjettiness' τ_3/τ_2 [43]. Les variables τ_3 et τ_2 expriment à quel point un jet peut être décrit comme provenant de 3 et 2 partons durs respectivement. Le rapport τ_{32} permet une discrimination entre les grands jets produits par l'émission de trois partons durs de ceux produits par deux partons durs. Le ratio τ_{21} , calculé de manière équivalente à τ_{32} , permet une discrimination entre l'émission de deux et d'un unique parton dur au sein d'un grand jet. Les variables C_2 [44] et D_2 [45] sont conçues pour identifier les désintégrations

hadroniques des bosons Z , W ou de Higgs boostés dans un environnement hadronique.

3.7 Simulation du détecteur

Après la génération des événements de collisions dures par les générateurs Monte Carlo (cf. Section 1.4), l'étape de simulation des événements consiste à reproduire le passage des particules générées dans le détecteur. Cette phase est réalisée par le logiciel Geant4 [46]. Ce logiciel décrit les détails de l'interaction des particules traversant la matière du détecteur. Ces interactions font intervenir des processus électromagnétiques comme l'effet photoélectrique, la diffusion Compton, les rayonnement de bremsstrahlung ou encore la conversion de photons, des processus hadroniques impliquant des collisions de hadrons incidents sur des noyaux, mais aussi des désintégrations de particules ayant lieu dans le détecteur. Les distorsions et les effets de désalignement du détecteur sont également pris en compte. La dernière étape de la simulation, appelée *digitisation*, simule la réponse de l'électronique du détecteur suite à l'énergie déposée dans la matière des sous-détecteurs et prend en compte le bruit électronique. Au final, la réponse du détecteur aux collisions d'empilement, aux halos ou gaz de faisceaux sont simulés avec la collision d'intérêt.

La complexité de la géométrie du détecteur et des descriptions physiques employées par la simulation ATLAS de Geant4 nécessite une consommation CPU importante pour simuler les millions d'événements nécessaires à l'analyse. Pour certaines études, il est parfois nécessaire de considérer une simulation simplifiée du passage des particules dans le détecteur, appelée *simulation rapide* par opposition à la *simulation complète*, afin d'obtenir une statistique Monte Carlo plus importante. Environ 75% de la consommation de la simulation complète est dépensée à simuler les particules électromagnétiques formant des gerbes dans les calorimètres (Figure 3.11). L'approche proposée par la simulation rapide est donc d'éliminer les particules électromagnétiques de faible énergie et de les remplacer par des gerbes pré-simulées mémorisées. La version de simulation rapide appelée ATLFAST-II [47] permet de gagner jusqu'à un facteur 100 sur le temps de simulation des traces du Détecteur Interne, des muons et des processus impliquant le calorimètre.

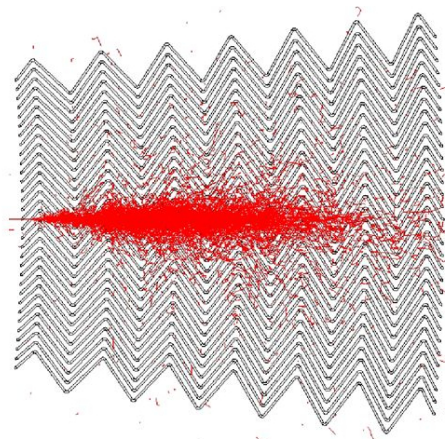


FIGURE 3.11: Simulation Geant4 d'une gerbe électromagnétique déclenchée par un électron de 50 GeV d'énergie dans le calorimère électromagnétique d'ATLAS. Les traits rouges symbolisent le passage des électrons/positrons à travers la matière.

Bibliographie

- [24] R Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Methods Phys. Res., A*, 262(HEPHY-PUB-503) :444. 19 p, Jun 1987.
- [25] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, R. Aben, M. Abolins, O. S. AbouZeid, N. L. Abraham, H. Abramowicz, and et al. Charged-particle distributions in $\sqrt{s} = 13$ TeV pp interactions measured with the ATLAS detector at the LHC. *Physics Letters B*, 758 :67–88, July 2016.
- [26] Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-018, CERN, Geneva, Jul 2015.
- [27] G Piacquadio, K Prokofiev, and A Wildauer. Primary vertex reconstruction in the atlas experiment at lhc. *Journal of Physics : Conference Series*, 119(3) :032033, 2008.
- [28] Characterization of Interaction-Point Beam Parameters Using the pp Event-Vertex Distribution Reconstructed in the ATLAS Detector at the LHC. Technical Report ATLAS-CONF-2010-027, CERN, Geneva, May 2010.
- [29] Atlas experiment - beam-spot measurements. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BeamSpotPublicResults>. Accessed : 2017-07-30.
- [30] Morad Aaboud et al. Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC. *Eur. Phys. J., C77(5)* :332, 2017.
- [31] Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data. Technical Report ATLAS-CONF-2016-024, CERN, Geneva, Jun 2016.
- [32] Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2016-015, CERN, Geneva, Aug 2016.
- [33] Atlas experiment - event displays from collision data. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun2Collisions>. Accessed : 2017-07-30.
- [34] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, R. Aben, M. Abolins, O. S. AbouZeid, N. L. Abraham, H. Abramowicz, and et al. Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV. *European Physical Journal C*, 76 :292, May 2016.

- [35] W Lampl, S Laplace, D Lelas, P Loch, H Ma, S Menke, S Rajagopalan, D Rousseau, S Snyder, and G Unal. Calorimeter Clustering Algorithms : Description and Performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008.
- [36] M. Cacciari, G. P. Salam, and G. Soyez. The anti- k_t jet clustering algorithm. *Journal of High Energy Physics*, 4 :063, April 2008.
- [37] Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data. Technical Report ATLAS-CONF-2012-020, CERN, Geneva, Mar 2012.
- [38] Georges Aad et al. Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J.*, C73(3) :2304, 2013.
- [39] Georges Aad et al. Performance of jet substructure techniques for large- R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *JHEP*, 09 :076, 2013.
- [40] D. Krohn, J. Thaler, and L.-T. Wang. Jet trimming. *Journal of High Energy Physics*, 2 :84, February 2010.
- [41] S Catani, Yu L Dokshitzer, Michael H Seymour, and Bryan R Webber. Longitudinally-invariant k_{\perp} -clustering algorithms for hadron-hadron collisions. *Nucl. Phys. B*, 406(CERN-TH-6775-93. LU-TP-93-2) :187–224. 38 p, Feb 1993.
- [42] Jet mass reconstruction with the ATLAS Detector in early Run 2 data. Technical Report ATLAS-CONF-2016-035, CERN, Geneva, Jul 2016.
- [43] J. Thaler and K. Van Tilburg. Identifying boosted objects with N-subjettiness. *Journal of High Energy Physics*, 3 :15, March 2011.
- [44] A. J. Larkoski, G. P. Salam, and J. Thaler. Energy correlation functions for jet substructure. *Journal of High Energy Physics*, 6 :108, June 2013.
- [45] A. J. Larkoski, I. Moult, and D. Neill. Power counting to better jet observables. *Journal of High Energy Physics*, 12 :9, December 2014.
- [46] S. Agostinelli et al. GEANT4 : A Simulation toolkit. *Nucl. Instrum. Meth.*, A506 :250–303, 2003.
- [47] G. Aad, B. Abbott, J. Abdallah, A. A. Abdelalim, A. Abdesselam, O. Abdinov, B. Abi, M. Abolins, H. Abramowicz, H. Abreu, and et al. The ATLAS Simulation Infrastructure. *European Physical Journal C*, 70 :823–874, December 2010.

Chapitre 4

Etiquetage des jets de quark *beauté*

4.1 Introduction

L'étiquetage des jets de b désigne l'identification des jets provenant de l'hadronisation de quarks b , appelés *jets de b* . Il est conçu pour rejeter, dans la mesure du possible, les jets de bruit de fond issus de saveurs plus légères, que l'on sépare habituellement dans les prédictions en *jets de c* (provenant de l'hadronisation de quarks c) et en *jets légers* (issus de quarks $u/d/s$). L'étiquetage des jets de b est appliqué aux jets des événements dans de nombreuses analyses pour séparer les états finals selon leur teneur en quarks b , que ce soit pour sélectionner ou bien pour supprimer des événements riches en jets de b . Identifier les jets de b permet non seulement d'étudier le quark *top*, qui se désintègre presque exclusivement en émettant un quark b ($\Gamma_{t \rightarrow Wb} \sim 100\%$), mais aussi d'étendre la recherche du boson de Higgs en profitant du taux de branchement élevé de sa désintégration en une paire de quarks b ($\Gamma_{H \rightarrow b\bar{b}} \sim 57\%$). Par ailleurs, l'identification de jet de b s'avère particulièrement utile dans un environnement hadronique, où les jets légers, issus de quarks u, d, s , sont produits en abondance par interaction forte.

L'étiquetage des jets de b repose sur les propriétés des *hadrons b* formés lors de l'hadronisation de quarks b . Un hadron b possède un quark de valence de saveur b qui lui confère ses caractéristiques propres. Son temps de vie moyen par exemple, d'environ 1.6 ps, est essentiel à l'étiquetage. Il est trop court pour permettre une détection directe du hadron b , car ce dernier se désintègre avant d'atteindre les premières couches du détecteur à Pixel. Son parcours moyen dans le plan transverse est de ~ 5.1 mm pour des jets de b de processus $t\bar{t}$ (à comparer avec un rayon de 33 mm pour l'IBL). En revanche, le temps de vie du hadron b est suffisamment long pour distinguer – dans les limites de la résolution du trajectographe – les traces issues de sa désintégration des traces issues de la collision dure. Ces dernières, appelées *traces de fragmentation*, sont dominantes dans les jets légers. Contrairement aux traces de fragmentation, les traces provenant de la désintégration du hadron b peuvent être significativement éloignées, ou *déplacées*, du vertex primaire. De plus, elles convergent en un point appelé vertex secondaire (SV), identifié comme la position de la désintégration du hadron b , et distinct du vertex primaire. Le déplacement des traces et les propriétés des vertex secondaires reconstruits au sein des jets sont utilisées pour distinguer les jets de b des jets de bruit de fond.

Outre son temps de vie caractéristique, le hadron b possède une masse d'environ ~ 5 GeV, grande devant les masses typiques de ses particules filles. Il emporte également une relativement grande fraction de l'énergie du jet de b parent, en moyenne de 70%. L'énergie restante du jet de b est portée, principalement, par des traces de fragmentation. Enfin, l'étiquetage des jets de b tire avantage de la nature des particules filles du hadron b . Dans presque toutes les désintégrations, un *hadron c* composé d'un quark c de valence est produit, car le terme de matrice CKM $|V_{cb}|^2$ est grand devant $|V_{ub}|^2$ ($1.7 \times 10^{-3} \gg 1.2 \times 10^{-5}$). Ses propriétés sont proches de celles du hadron b . Les hadrons c possèdent des temps de vie moyens compris entre ~ 0.5 et ~ 1.0 ps et une masse d'environ 2 GeV. Les traces issues de sa désintégration sont aussi déplacées, et peuvent être combinées avec celles du hadron b pour

reconstruire un unique vertex secondaire, ou utilisées pour reconstruire un nouveau vertex, appelé vertex tertiaire, dans la direction de vol du hadron b donnée par l'axe PV-SV (Figure 4.1). Le nombre de particules chargées issues des désintégrations de hadrons provenant de b et c de jets de b (Figure 4.2) sont en moyenne de ~ 2.7 et ~ 3 respectivement. Il est intéressant de noter que les hadrons b et c peuvent également se désintégrer en émettant un électron ou un muon, chacun avec un taux de branchement de $\sim 20\%$. La cascade de désintégration $b \rightarrow c$ produit donc un lepton chargé dans $\sim 40\%$ des cas.

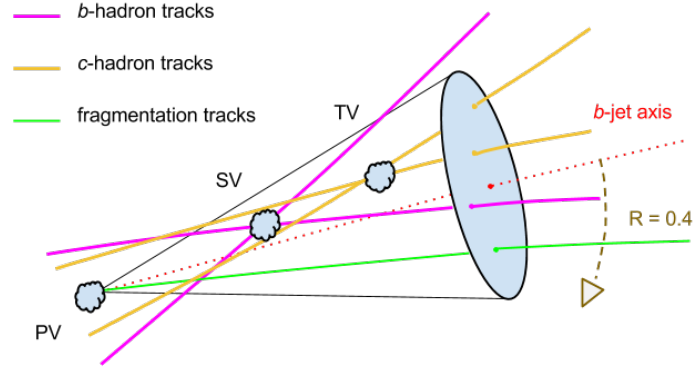


FIGURE 4.1: Schéma simplifié d'un jet de b et des vertex provenant des désintégrations du hadron b (SV) et du hadron c (TV).

Certains jets de bruit de fond présentent des traces qui imitent celles issues de la désintégration de hadrons b , et peuvent être incorrectement étiquetés b . Ces jets de bruit de fond qui ressemblent aux jets de b peuvent contenir des désintégrations faibles de hadrons dont le temps de vie est semblable à celui du hadron b , comme celles de K_S^0 ou Λ^0 , appelés V^0 , ou encore de hadrons c . Ils peuvent également contenir des collisions de particules avec la matière du détecteur, comme des *interactions hadroniques* ou des *conversions*. Les premières sont déclenchées par un hadron, et produisent un ensemble de particules secondaires. Les secondes sont amorcées par des photons, qui se convertissent en paires électron-positron. D'autre part, les traces de fragmentation peuvent être déplacées à cause des diffusions multiples dans le détecteur, de l'incertitude sur l'alignement du détecteur, ou bien à cause de la résolution du vertex primaire. Par la suite, les jets étiquetés (identifiés) comme provenant d'un jet de b seront appelés *candidats b* , qu'ils proviennent en vérité d'un jets de b ou de jets de bruit de fond. Dans la simulation, un jet est défini comme étant un jet de b s'il contient un hadron b tel que ΔR (jet de b , hadron b) < 0.3 ; et un jet est défini comme jet de c s'il n'est pas un jet de b et s'il contient un hadron c tel que ΔR (jet de c , hadron c) < 0.3 . Finalement, un jet est défini comme un jet léger s'il n'est ni un jet de b ni un jet de c .

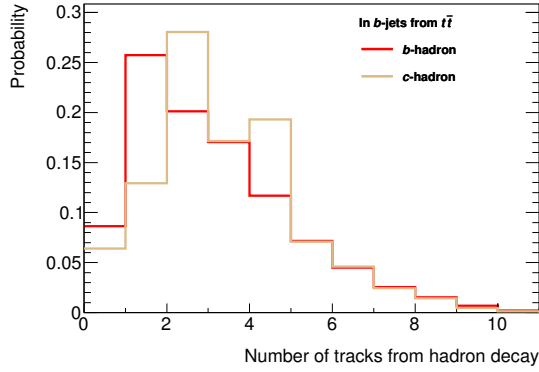


FIGURE 4.2: Distribution du nombre de traces émises lors de la désintégration de hadrons b et de hadrons c provenant de jets de b . Ces jets sont collectés à partir d'un échantillon $t\bar{t}$.

4.2 Les algorithmes d'étiquetage des jets de b

Les traces sont les éléments de base utilisés par les algorithmes d'étiquetage des jets de quark b pour rechercher les hadrons b au sein des jets, caractéristiques des jets de b . La première étape de l'étiquetage réalisé dans l'expérience ATLAS [48] consiste à associer des traces aux jets. Comme les particules issues de la désintégration du hadron b sont émises de manière privilégiée selon la direction du jet associé, un jet d'impulsion \vec{p}_{jet} se voit associer les traces d'impulsion \vec{p}_{trace} (estimée au point de moindre approche du vertex primaire) si celles-ci satisfont le critère angulaire suivant $\Delta R(\vec{p}_{jet}, \vec{p}_{trace}) < R_{cone}$. Le rayon du cône R_{cone} vaut 0.43 à $p_T = 20$ GeV, et décroît avec l'impulsion transverse du jet ; il atteint 0.26 à $p_T = 150$ GeV [49]. Les petits rayons à haute impulsion réduisent la contamination de traces provenant de désintégrations de hadrons b dans les jets de bruits de fond voisins des jets de b , et conservent en même temps la proportion des traces issues de la désintégration du hadron b – de plus en plus collimées – associées aux jets de b .

Les traces sont ensuite utilisées en entrée des trois algorithmes d'étiquetage des jets de b appelés IPxD, SV1 et JetFitter, décrits dans les sections suivantes. L'algorithme IPxD teste si les jets proviennent de l'hadronisation d'un quark b grâce aux déplacements par rapport au vertex primaire de leurs traces associées, tandis que les deux autres algorithmes reconstruisent des vertex secondaires à partir des traces associées. Les propriétés des vertex secondaires reconstruits au sein des jets par SV1 et JetFitter, ainsi que l'information obtenue par IPxD, sont combinées avec le classifieur $MV2$ à l'aide d'une méthode multivariée (Figure 4.3) afin d'identifier au mieux les jets de b .

4.2.1 L'algorithme IPxD et les paramètres d'impact des traces

Les traces originaires de la cascade de désintégration des hadrons b sont souvent significativement déplacées par rapport au vertex primaire. Cette propriété est exploitée par l'algorithme IPxD [49, 50]. Les variables à l'origine de la discrimination d'IPxD, à savoir les

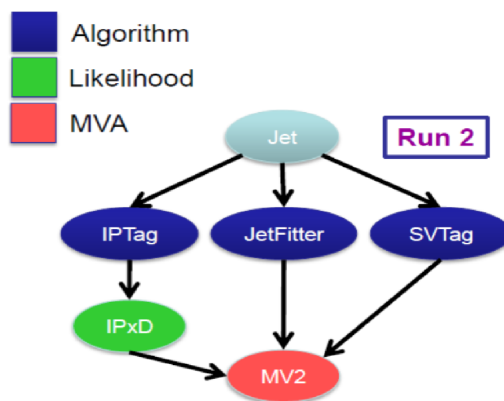


FIGURE 4.3: Diagramme des étapes principales de l'étiquetage des jets de b : les traces associées aux jets sont données en entrée des algorithmes IPxD (ou IPxTag), SV1 (ou SVTag) et JetFitter. Les informations des trois algorithmes sont ensuite combinées avec le classifieur $MV2$ qui associe au final un poids à chaque jet selon son 'degré de ressemblance' avec un jet de b .

paramètres d'impact signés des traces $IP_{r\phi}$ et IP_z , seront définis et discutés, en premier lieu, afin de comprendre la forme de leur distributions. Ensuite, la sélection des traces utilisées par l'algorithme IPxD et la combinaison des informations des paramètres d'impact signés des traces sélectionnées dans le but d'identifier les jets de b seront présentées dans cette section.

Les paramètres d'impact signés $IP_{r\phi}$ et IP_z

Dans la suite du chapitre, la position du vertex primaire est utilisée comme origine du repère $(Oxyz)$. Les paramètres d'impact sont calculés à partir des coordonnées (d_0, z_0, θ) du point P de la trace le plus proche de l'axe (Oz) . Les paramètres d_0, z_0 correspondent aux distances du point P à l'origine dans le plan transverse et longitudinal respectivement, et l'angle θ est défini comme l'angle entre l'axe du faisceau et le vecteur \vec{OP} . Les paramètres d'impact transversal et longitudinal sont donnés par

$$IP_{r\phi} = d_0, IP_z = z_0 \sin(\theta).$$

En approximant la trace à une ligne droite en P , IP_z correspond à la distance la plus proche entre la trace et le vertex primaire dans le plan tangent à la trace en P .

En moyenne, les paramètres d'impact d'une trace produite par la désintégration d'une particule ne dépend pas de l'impulsion de la particule, étant donné leur invariance par boost de la particule mère (Figure 4.4). Tant que la particule mère est relativiste, l'augmentation de la distance de vol parcourue de la particule mère est compensée par une diminution de l'angle entre la particule mère et particule secondaire à l'origine de la trace.

IPxD exploite aussi la direction du vecteur joignant le vertex primaire au point de la trace le plus proche du vertex primaire, \vec{IP}_{3D} , pour distinguer les traces de vertex secondaires

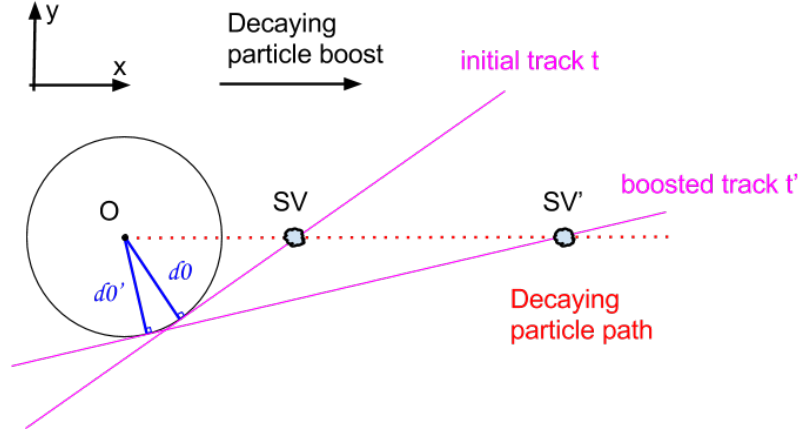


FIGURE 4.4: Illustration de l'invariance du paramètre d'impact $IP_{r\phi} = d_0$ par boost de Lorentz, effectué ici selon la direction de la particule se désintégrant. Pour simplifier, la particule mère et la trace provenant de sa désintégration sont émises dans le plan transverse.

des traces de fragmentation. Négligeant les effets de résolution des traces, le vecteur \vec{IP}_{3D} d'une trace originaire de la désintégration d'une particule est systématiquement orienté selon la direction $\hat{p}_{mère}$ de cette dernière, autrement dit le produit scalaire $\vec{IP}_{3D} \cdot \hat{p}_{mère}$ est positif. Sachant que le hadron b porte une large fraction de l'énergie du jet, il est émis préférentiellement dans l'axe du jet \hat{p}_{jet} . Le signe du paramètre d'impact 3D – défini comme le signe de $\vec{IP}_{3D} \cdot \hat{p}_{jet}$ – d'une trace issue de la désintégration de hadron b est donc positif dans la plupart des cas. En revanche, le signe des paramètres d'impact des traces de fragmentation ne présente aucune valeur privilégiée. Les traces de fragmentation doivent leur déplacement à une erreur sur la mesure de la trace due aux effets d'instrumentation (diffusion, alignement) et le signe de $\vec{IP}_{3D} \cdot \hat{p}_{jet}$ est négatif ou positif de façon équiprobable (Figure 4.5).

Deux signes sont obtenus pour chacun des paramètres d'impact $IP_{r\phi}$ et IP_z à partir du signe des produits scalaires $(\vec{IP}_{3D})_{\perp} \cdot \hat{p}_{jet}$ et $(\vec{IP}_{3D})_{\parallel} \cdot \hat{p}_{jet}$, où $(\vec{IP}_{3D})_{\perp}$ et $(\vec{IP}_{3D})_{\parallel}$ sont les projections transverses et longitudinales de \vec{IP}_{3D} .

L'algorithme IPxD

L'algorithme IPxD combine les significations des paramètres d'impact des traces associées au jet pour déterminer si celui-ci provient ou non d'un quark b . La signification statistique du paramètre d'impact $S_{IP_{r\phi}}$ (respectivement S_{IP_z}) est définie comme l'inverse de l'incertitude relative du paramètre d'impact $IP_{r\phi}$ (respectivement IP_z).

Les traces de mauvaise qualité, ou non susceptibles de provenir de la désintégration d'un hadron b sont retirées de la sélection de traces d'IPxD, et par conséquent ne participent pas à la combinaison des significations des paramètres d'impact. Une première sélection sur les paramètres d'impact ($IP_{r\phi} \leq 1$ mm et $IP_z \leq 1.5$ mm) permet d'éviter une contamination trop importante de traces provenant de collisions pp d'empilement et de celles issues des

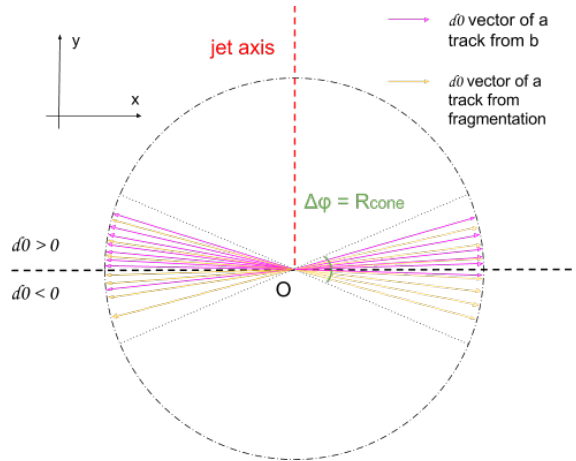


FIGURE 4.5: Schéma des vecteurs \vec{IP}_{3D} (à $|\vec{IP}_{3D}|$ fixé) pour des traces de désintégrations de hadrons b et des traces de fragmentation. Autant de traces provenant de la désintégration de b et de fragmentation sont considérées ici pour l'illustration. Elles sont émises dans le plan transverse, pour simplifier la visualisation du phénomène.

désintégrations faibles de V^0 . Le critère cinématique $p_{T,track} \geq 1$ GeV réduit la proportion de traces associées de fragmentation, qui ont en moyenne une impulsion transverse plus faible que les traces provenant de la désintégration du hadron b . Enfin, l'algorithme demande que les traces soient reconstruites à partir d'au moins deux coups dans le détecteur à pixel, dont au moins un dans sa couche la plus interne (l'IBL), pour garantir une relativement bonne précision de la mesure du paramètre d'impact.

Afin de prendre en compte la dépendance de la mesure du paramètre d'impact avec la qualité et la quantité des coups ayant servis à la reconstruction d'une trace dans les sous-détecteurs en silicium du trajectographe, quatorze catégories exclusives de traces sont définies selon leur contenu en coups (Fig. 4.6 [51]). Tout d'abord, les traces sont séparées selon qu'elles ont un coup dans la première couche (IBL) ou la seconde couche (b-layer) du Détecteur à Pixel. L'absence de coup(s) peut être attendue, si par exemple elle provient d'un module défectueux, ou inattendue dans le cas contraire. Ensuite, une trace peut contenir des coups dits *divisés* ou *partagés*. Le coup divisé est identifié comme provenant de plusieurs particules chargées. Le coup partagé (par définition différent du coup divisé) est un coup compatible avec le passage d'une unique particule chargée mais associé à plus d'une trace. La fraction relative de chaque catégorie évolue avec l'impulsion du jet. Les particules d'un jet deviennent de plus en plus collimées selon la direction de celui-ci avec son impulsion, et donc plus proches spatialement les unes des autres. Cet environnement dense conduit notamment à une plus grande proportion de traces reconstruites à partir de coups divisés ou partagés.

La densité de probabilité de la signification du paramètre d'impact $IP_{r\phi}$ signé des traces sélectionnées par IPxD et de catégorie 'Good' (numérotée 13), est présentée Fig. 4.7 [51], pour chaque saveur f de jets : légers, c et b , et est notée $PDF(S_{IP_{r\phi}}|Q = 13, f)$. Pour

#	Category	Fractional contribution [%]		
		b-jets	c-jets	light-jets
0	No hits in first two layers; expected hit in IBL and b-layer	1.9	2.0	1.9
1	No hits in first two layers; expected hit in IBL and no expected hit in b-layer	0.1	0.1	0.1
2	No hits in first two layers; no expected hit in IBL and expected hit in b-layer	0.04	0.04	0.04
3	No hits in first two layers; no expected hit in IBL and b-layer	0.03	0.03	0.03
4	No hit in IBL; expected hit in IBL	2.4	2.3	2.1
5	No hit in IBL; no expected hit in IBL	1.0	1.0	0.9
6	No hit in b-layer; expected hit in b-layer	0.5	0.5	0.5
7	No hit in b-layer; no expected hit in b-layer	2.4	2.4	2.2
8	<i>Shared</i> hit in both IBL and b-layer	0.01	0.01	0.03
9	At least one <i>shared</i> pixel hits	2.0	1.7	1.5
10	Two or more <i>shared</i> SCT hits	3.2	3.0	2.7
11	<i>Split</i> hits in both IBL and b-layer	1.0	0.87	0.6
12	<i>Split</i> pixel hit	1.8	1.4	0.9
13	<i>Good</i>	83.6	84.8	86.4

FIGURE 4.6: Description des catégories de traces considérées par IPxD avec leur fraction calculées pour un échantillon $t\bar{t}$. Les traces ne satisfaisant pas l'une des treize premières catégories sont associées à la catégorie 'Good' [51].

chaque saveur, les valeurs proches de zéro sont dominées par des traces de fragmentation, distribuées selon une gaussienne de largeur $\sigma \sim 1$. Les traces provenant de désintégrations faibles de hadrons, de conversions et d'interactions hadroniques, peuplent les queues de distribution à décroissance approximativement exponentielle, principalement aux valeurs positives. Leur fraction est beaucoup plus élevée pour les traces de jets de b que ceux de jet légers ou de c .

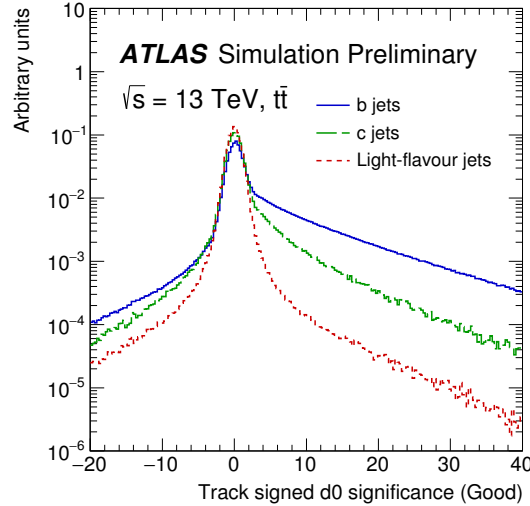


FIGURE 4.7: Densités de probabilité de la signification du paramètre d'impact $IP_{r\phi} = d_0$ des traces d'IPxD pour des jets légers, c et b originaire d'un échantillon $t\bar{t}$ [51].

La vraisemblance qu'une trace de signification du paramètre d'impact signé $S_{IP_{r\phi}}$ provienne d'un jet de saveur f est définie comme la densité de probabilité d' $S_{IP_{r\phi}}$ estimée

pour une catégorie Q et un jet de saveur f , notée $\text{PDF}(S_{IP_{r\phi}}|Q, f)$. Le produit des vraisemblances des n traces d'un jet sélectionnées par IPxD permet de définir la vraisemblance de ce jet $L_f = \prod_{i=1}^n \text{PDF}(S_{IP_{r\phi}}|Q, f)$. Finalement les discriminants finals sont donnés par les logarithmes des rapports de vraisemblance suivants :

$$\text{IP2D}_{ff'} = \log\left(\frac{L_f}{L_{f'}}\right) \quad (4.1)$$

pour $ff' = \{bu, cu, bc\}$, séparant au mieux les jets de b des jets légers, les jets de c des jets légers et des jets de b des jets de c respectivement. La méthode permettant de calculer ces trois variables $\text{IP2D}_{ff'}$ se généralise dans le cas tridimensionnel pour obtenir les variables $\text{IP3D}_{ff'}$. Ces dernières sont calculées à partir de la densité de probabilité du couple de significations du paramètre d'impact signé ($S_{IP_{r\phi}}, S_{IP_z}$) au lieu de la densité de probabilité de $S_{IP_{r\phi}}$.

4.2.2 L'algorithme JetFitter

Les traces issues de la désintégration d'un hadron b sont non seulement déplacées, mais peuvent aussi se croiser pour former un ou plusieurs vertex secondaires. Ces vertex secondaires sont reconstruits par l'algorithme JetFitter. La sélection des traces qui serviront à reconstruire les vertex secondaires est bien plus lâche que celle adoptée par IPxD. En effet, la reconstruction d'un vertex nécessite l'intersection d'au moins deux traces, et son efficacité se trouve limitée par la multiplicité de particules chargées émises dans la désintégration de hadron b . C'est pourquoi une haute efficacité de trouver les traces de vertex secondaires est cruciale pour la reconstruction de vertex secondaires.

Quantité de trace	Sélection JetFitter	Sélection SV1
Impulsion transverse	$p_T > 770 \text{ MeV}$	$p_T > 700 \text{ MeV}$
Paramètre d'impact $IP_{r\phi}$	$ IP_{r\phi} < 3.5 \text{ mm}$	$ IP_{r\phi} < 5 \text{ mm}$
Paramètre d'impact IP_z	$ IP_z < 5 \text{ mm}$	$ IP_z < 25 \text{ mm}$
Incertitude sur d_0	$\sigma(d_0) < 350 \mu\text{m}$	$\sigma(d_0) < 1 \text{ mm}$
Incertitude sur z_0	$\sigma(z_0) < 2.5 \text{ mm}$	$\sigma(z_0) < 5 \text{ mm}$
Qualité du fit de trace	$\chi^2/dof < 3.5$	N/A
Nombre minimal de coup de Pixel	$n_{pix} \geq 1$	$n_{pix} \geq 1$
Nombre minimal de coup de SCT	$n_{SCT} \geq 4$	$n_{SCT} \geq 4$
Nombre minimal de coup de Pixel+SCT	$n_{pix} + n_{SCT} \geq 7$	N/A

TABLE 4.1: Coupures de sélection appliquées aux traces utilisées en entrée des algorithmes de reconstruction de vertex JetFitter et SV1.

Au sein d'un jet de b , le hadron c de la cascade de désintégration $b \rightarrow c$ parcourt une distance suffisamment élevée pour permettre la reconstruction des désintégrations de hadrons b et c en deux vertex secondaires distincts. Typiquement, pour des jets de b du processus

$t\bar{t}$ d'impulsion transverse moyenne ~ 70 GeV, la distance de vol transverse du hadron c est d'environ 1.6 mm en moyenne, grande devant la résolution spatiale transverse des vertex reconstruits, *i.e* ~ 40 μm . La reconstruction des vertex de la cascade de désintégration $b \rightarrow c$ est effectuée par l'algorithme JetFitter (JF) [52]. Comme l'angle entre les impulsions des hadrons b et c est en général très faible, les vertex sont recherchés de manière privilégiée le long d'un même axe, appelé *axe JetFitter*, qui estime la direction de vol du hadron b .

Les traces sélectionnées pour reconstruire les vertex présentent des coupures relâchées sur les paramètres d'impact (Table 4.1), afin de conserver une grande proportion de traces issues des désintégrations de hadrons b et c , au détriment d'une contamination de traces de vertex originaire de désintégrations de hadrons V^0 produits au vertex primaire, de conversions de photon ou d'interactions hadroniques.

Une reconstruction de paires de traces formant des vertex secondaires est réalisée par un filtre de Kalman, à partir de traces qui ne proviennent pas du vertex primaire. Les vertex à deux traces ayant vraisemblablement une origine commune avec le vertex primaire sont en premier lieu rejetés. Les vertex situés au niveau des premières couches du détecteur à Pixel sont identifiés comme provenant d'interactions hadroniques. Ensuite, les conversions de photons sont identifiées à partir de vertex de faibles masses et dont les deux traces sont de charges opposées. Les vertex de mésons K_s^0 et de baryons Λ sont identifiés notamment grâce à leur masse, sélectionnée autour de 498 MeV et de 1116 MeV respectivement¹. Pour chaque vertex identifié comme provenant d'un hadron V^0 , une *trace neutre* est définie. Cette trace neutre passe par la position du vertex associé et son impulsion est donnée par l'impulsion de ce même vertex. Les paramètres d'impact obtenus permettent de classer ces candidats V^0 comme particules primaires, provenant du vertex primaire, ou comme particules secondaires dans le cas contraire. Finalement, les traces des candidats V^0 primaires sont rejetées avec celles des candidats $\gamma \rightarrow ee$ et des candidats de collisions hadroniques. Les traces neutres des candidats V^0 secondaires seront utilisées pour le fit de la cascade de désintégration $b \rightarrow c$.

Un premier fit de la cascade de désintégration $b \rightarrow c$ est effectué à l'aide du filtre de Kalman à partir des traces restantes associées à des vertex à deux traces. Les positions des vertex et l'orientation de l'axe de JetFitter sont ajustées pendant cette procédure. Le fit commence en supposant que l'axe JetFitter passe par le vertex primaire et est orienté selon l'axe du jet. Un second fit est réalisé, en utilisant le résultat du fit précédent en entrée, et en rajoutant les traces qui ne sont pas associées ni aux vertex à deux traces ni au vertex primaire. L'intersection de l'axe JetFitter et d'une trace peut donner lieu à la reconstruction d'un vertex secondaire, appelé *vertex secondaire à une trace*. Enfin, les vertex compatibles entre eux sont rassemblés et les vertex peu probables sont rejetés.

Huit variables sont calculées à partir des propriétés des vertex reconstruits résultant du fit pour être utilisées lors de la combinaison des algorithmes (Table 4.2). Ces variables dépendent des caractéristiques de la particule mère à l'origine du/des vertex secondaire(s),

1. Les masses de mésons K_s^0 et de baryons Λ sont calculées à partir de l'hypothèse que les traces viennent de pions chargés et d'un système proton pion chargé respectivement.

comme sa masse, sa fraction d'énergie du jet associé, sa distance parcourue par rapport au vertex primaire, et par conséquent présentent des distributions différentes pour les jets de b et pour les jets de bruit de fond (Figure 4.8).

L'énergie et l'impulsion du/des vertex secondaires, qui interviennent dans la définition des variables, sont définies comme la somme des énergies et impulsions respectivement des traces associées au(x) vertex secondaire(s). Pour le calcul de l'énergie des traces, on fait l'hypothèse que les traces viennent de pions chargés et une masse de 139 MeV leur est attribuée. La direction du jet est donnée par la direction de son impulsion. Les distances des vertex secondaires sont calculées par rapport au vertex primaire. La *significance de la distance 3D* (3D length significance) utilisée comme en entrée de MV2 est définie comme le barycentre des distances de vertex secondaires L_i pondérées par leur incertitude σ_i :

$$\frac{\sum_i{}_{SV} L_i/\sigma_i}{\sum_i{}_{SV} 1/\sigma_i}. \quad (4.2)$$

Le calcul de la masse *standard* utilisée en entrée de MV2 fait intervenir une correction pour estimer la contribution des particules neutres sur la masse du vertex. Cette correction est définie comme la somme scalaire des impulsions transverses à l'axe JetFitter des traces associées au(x) vertex secondaire(s), notée $\sum_{tracks \in SV} p_{T,track}^{JF\ axis}$. Elle est ajoutée à la masse calculée à partir des impulsions des traces associées au(x) vertex secondaire(s) notée m_{SV} , définissant la masse $m_{SV,corr}$:

$$m_{SV,corr} = m_{SV} + \sum_{tracks \in SV} p_{T,track}^{JF\ axis}. \quad (4.3)$$

De plus, afin de faciliter l'utilisation de la variable dans l'algorithme multivarié, la masse est bornée à 10 GeV en appliquant la transformation suivante :

$$m'(GeV) = 5 + 5 \times \frac{2}{\pi} \arctan\left(\frac{\pi}{2} \times \frac{m - 5}{5}\right). \quad (4.4)$$

La masse JetFitter fait référence à cette masse corrigée dans la suite de ce manuscrit. Une correction similaire est appliquée à la signficance 3D de la distance afin de maintenir ses valeurs dans un intervalle donné.

4.2.3 L'algorithme SV1

La stratégie de l'algorithme SV1 [53] est de reconstruire la désintégration du hadron b en un unique vertex secondaire. La conception de SV1 le rend très efficace pour la réjection des jets de c , mais inadapté aux jets de b de très haute impulsion transverse, lorsque les désintégrations de hadrons b et c sont significativement séparées.

La sélection des traces de SV1 (Table 4.1) est équivalente à celle de JetFitter, permettant de sélectionner un grand nombre de traces non issues du vertex primaires pour la recherche des vertex secondaires. Les vertex ne provenant pas de désintégrations de hadron b/c sont rejetés de manière analogue à ce qui est fait par l'algorithme JetFitter. Des critères topologiques sur les vertex permettent de rejeter une partie des faux vertex secondaires, comme

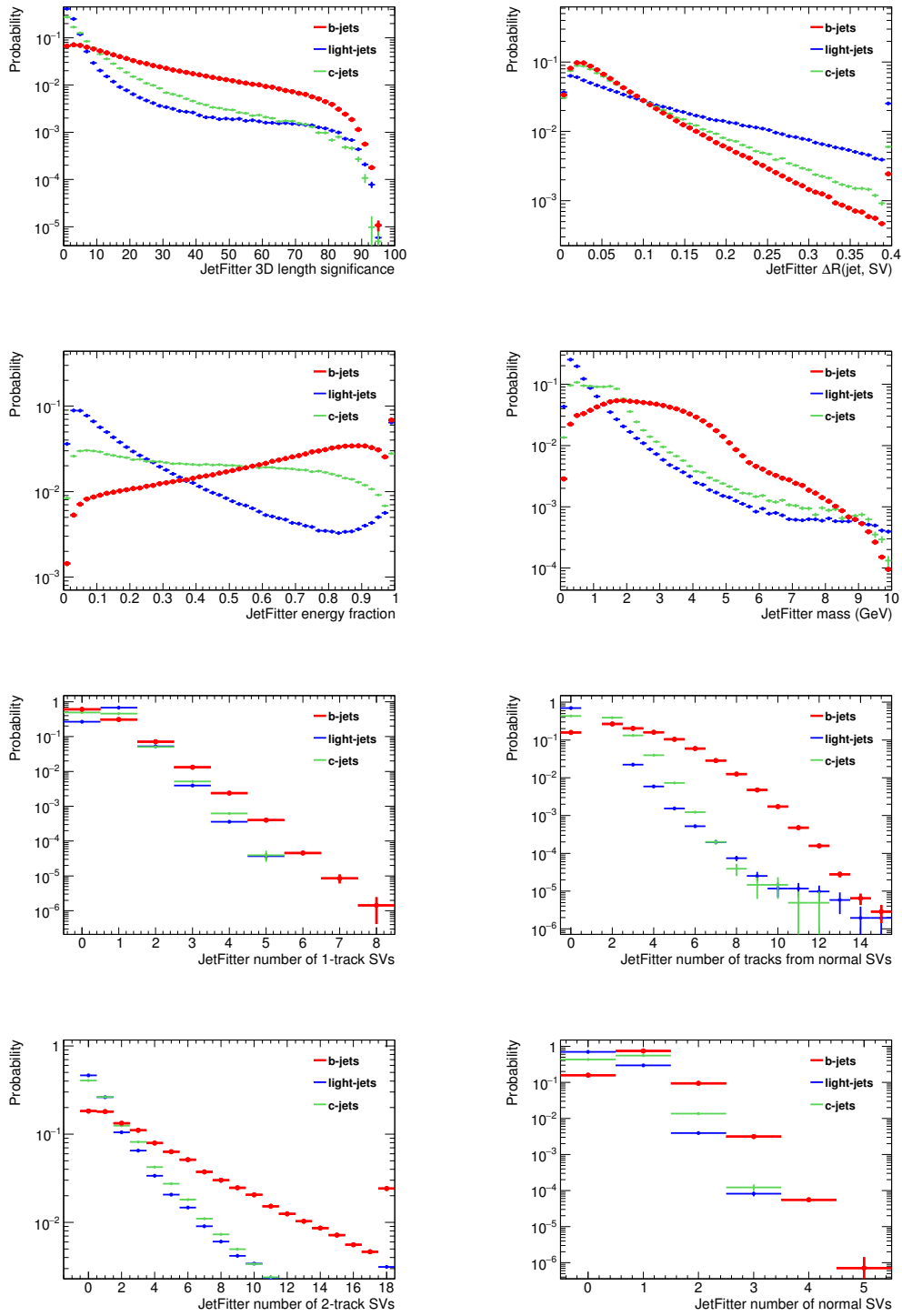


FIGURE 4.8: Distributions des huit variables JetFitter utilisées comme variables d'entrée dans la combinaison $MV2$. Ces variables sont calculées à partir des vertex secondaires reconstruits par l'algorithme JetFitter pour chaque jet. Leurs distributions sont représentées pour des jets légers (bleu), des jets de b (rouge) et de c (vert). Les jets sont issus d'un échantillon $t\bar{t}$.

Variable de JetFitter	Définition
Significace de la distance 3D	Distance moyenne des vertex secondaires selon l'axe JetFitter divisé par son incertitude
ΔR (jet,SV)	Angle ΔR entre le jet et l'impulsion des vertex secondaires
Fraction d'énergie	Energie des vertex secondaires divisée par l'énergie des traces associées au jet
Masse	Masse des vertex secondaires
Nombre de vertex à 1 trace	Nombre de vertex secondaires qui ne possèdent qu'une seule trace
Nombre de <i>vertex normaux</i>	Nombre de vertex secondaires à au moins deux traces
Nombre de traces associées aux <i>vertex normaux</i>	Nombre de traces dans les vertex à au moins deux traces
Nombre de vertex à 2 traces	Nombre de vertex à deux traces reconstruits avant le fit de la cascade de désintégration $b \rightarrow c$

TABLE 4.2: Définition des variables calculées à partir des vertex secondaires reconstruits par l'algorithme JetFitter et utilisées en entrée de l'algorithme *MV2*.

par exemple ceux situés après l'IBL mais dont une trace associé au vertex possède un coup dans cette couche du Détecteur à Pixel. Un vertex est fitté à l'aide de la méthode de Billoir à partir des traces restantes [54]. Les traces qui dégradent davantage la qualité du fit (qui font augmenter le χ^2 du fit) sont retirées une à une. Finalement, les propriétés des vertex secondaires reconstruits sont explorées pour l'identification de jets de b [55]. Au total, huit variables de l'algorithme SV1 sont utilisées en entrée du classifieur *MV2*.

4.2.4 Le classifieur *MV2*

La dernière étape de l'étiquetage des jets de b consiste à combiner l'information des trois algorithmes IPxD, SV1 et JF afin de séparer les jets de b des jets de bruit de fond, composé d'un mélange de jets légers et de jets de c . Cette combinaison est réalisée à l'aide d'un arbre de décision boosté (BDT). Dans cette section nous décrivons les variables d'entrée du BDT, les jets qui composent l'échantillon d'entraînement et celui d'évaluation, et enfin les réglages de l'algorithme de BDT.

Echantillons et variables d'entrée du BDT

La distinction entre jets de b et jets de bruit de fond se fonde sur un ensemble de 22 variables, choisies pour leur caractère discriminant. Il s'agit des rapports de vraisemblance d'IPxD, des huit variables liées au vertex reconstruit par SV1 ainsi qu'aux huit variables calculées à partir de ceux reconstruits par JetFitter. Pour l'entraînement comme pour la

phase d'évaluation, les variables d'entrée sont calculées à partir de jets simulés provenant de la production de paire $t\bar{t}$ semi-leptonique (où un lepton chargé e ou μ est produit lors de la désintégration de boson W). Le processus $t\bar{t}$ est choisi comme échantillon de référence, car riche en jets de b originaires des désintégrations de quarks top et en jets légers et de c provenant des désintégrations de bosons W hadroniques.

Cependant, ces jets ont des caractéristiques propres à la production de paire $t\bar{t}$, et quelques corrections sont nécessaires pour rendre l'étiquetage des jets de b moins dépendant du processus de production. Tout d'abord, pour que l'étiquetage des jets de b n'exploite pas les différences cinématiques entre les jets de b et les jets de bruit de fond, le spectre en impulsion transverse p_T et en pseudo-rapacité η doit être identique pour les jets de b et ceux de bruit de fond. La pondération choisie applique un poids aux jets de b et de c pour que leur distribution respective (p_T, η) corresponde à celle des jets légers. De plus, les variables p_T et η des jets sont ajoutées à la liste des 22 variables d'entrée, si bien que l'optimisation est réalisée par intervalle $dp_T, d\eta$ et devient moins dépendante du choix du spectre initial en (p_T, η) . Enfin, la fraction de jets de bruit de fond de saveur c utilisée dans l'échantillon d'entraînement est artificiellement modifiée. Trois optimisations de l'étiquetage des jets de b sont effectuées pour trois valeurs de fractions de jets de c : 0, 10 et 20 %, obtenues en pondérant les jets de c .

Une séparation optimale entre jets de b et jets de bruit de fond fait appel au calcul du rapport de vraisemblance, calculée à partir des densités de probabilité des 24 variables d'entrée x , pour l'hypothèse de signal ($\text{PDF}(x|b)$) et celle de bruit de fond ($\text{PDF}(x|cl)$)

$$\text{LLR} = \log \left(\frac{\text{PDF}(x|b)}{\text{PDF}(x|cl)} \right).$$

Dans la pratique, déterminer la densité de probabilité d'un espace à 24 variables à partir d'événements Monte Carlo est irréalisable, faute d'un nombre suffisant d'événements. C'est pourquoi un algorithme multivarié – en l'occurrence un BDT – est utilisé. Sa réponse optimale tient compte de la statistique limitée. Le BDT renvoie une unique variable de sortie en fonction des valeurs des 24 variables d'entrée. Cette variable est appelée $MV2cX$, où X indique la fraction de c , en pourcent, du bruit de fond utilisé pour l'entraînement (par exemple $MV2c20$ pour 20%). Pour les études qui ne dépendent pas de cette fraction, la variable $MV2cX$ sera noté simplement $MV2$ par la suite. L'application f

$$f : x \rightarrow MV2$$

est déterminée à la suite d'un entraînement du BDT, qui sera décrit dans la section suivante. Le poids $MV2$ d'un jet prend des valeurs comprises entre -1 et 1 : plus sa valeur est proche de 1, plus le jet a de chance d'être un jet de b , et plus sa valeur tend vers -1, plus il a de chance d'être un jet de bruit de fond.

Pour gagner du temps de calcul sur l'entraînement et son évaluation, l'entraînement se focalise sur un ensemble de *bons jets*, pour lesquels les algorithmes IPxD, SV1 et JetFitter sont définis : un bon jet contient au moins une trace sélectionnée par IPxD, un vertex secondaire reconstruit par SV1 et au moins un vertex secondaire reconstruit par JetFitter.

L'échantillon d'entraînement est constitué d'un ensemble de bons jets et de jets qui ne sont pas de bons jets, appelés *mauvais jets*. Pour un mauvais jet, au moins un des trois algorithmes IPxD, SV1 et JetFitter n'est pas défini, *i.e.*, aucune trace n'est sélectionnée par IPxD, ou aucun vertex n'a été reconstruit par SV1, ou aucun vertex n'a été reconstruit par JetFitter. Les échantillons de bons jets et de mauvais jets d'entraînement contiennent chacun 10^5 jets de b , autant de jets légers, et 5×10^4 jets de c . Les mauvais jets sont d'abord *dévalués* avec l'application d'un poids de 10^{-6} . Puis la pondération pour égaliser le spectre en impulsion transverse p_T et en pseudo-rapacité η des différentes saveurs de jets est appliquée indépendamment pour l'ensemble des bons jets et pour celui des mauvais jets (Fig. 4.9)). Un échantillon de test qui sert à la mesure du surentraînement est construit de la même manière que l'échantillon d'entraînement, mais à partir d'un échantillon statistiquement indépendant. Un dernier échantillon statistiquement indépendant des deux premiers, dit d'évaluation, sert à quantifier les performances de l'entraînement. C'est un échantillon inclusif formé des jets de l'échantillon $t\bar{t}$, obtenu sans appliquer de poids aux jets et sans égaliser artificiellement le nombre de bons et de mauvais jets comme pour l'échantillon d'entraînement.

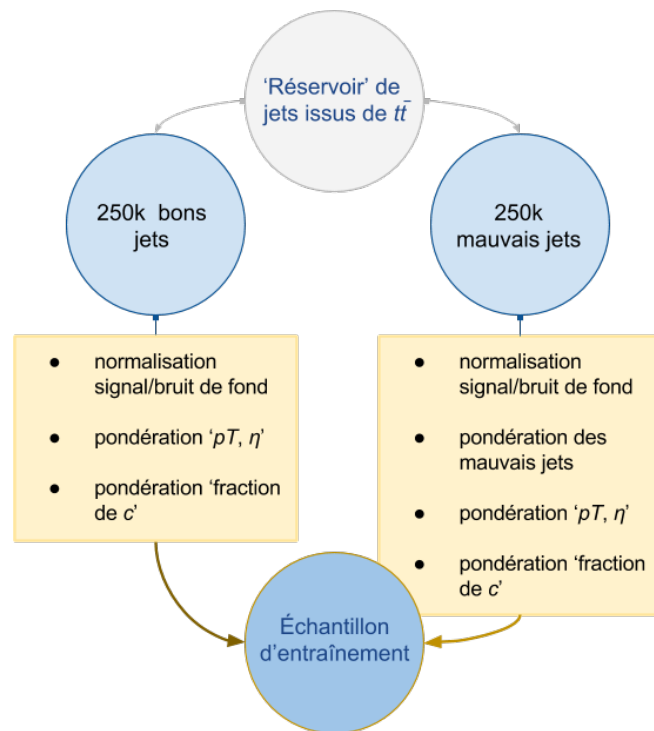


FIGURE 4.9: Schéma résumant la construction de l'échantillon d'entraînement, formé de deux ensembles de bons jets et de mauvais jets.

Gradient boosting et arbre de décision

Le BDT utilisé pour obtenir *MV2* fait intervenir deux concepts : celui dit de *gradient boosting*, et celui d'*arbre de décision*. Le gradient boosting est une méthode générale permettant d'améliorer un classifieur $h(x)$ de manière itérative, en se focalisant à chaque nouvelle itération sur les mauvaises classifications précédemment commises. Formellement, à chaque nouvelle itération m , $h(x)$ est actualisé afin de faire décroître une fonction de perte $L(y, h(x))$, où y (qui prend la valeur $+1$ ou -1 pour le signal ou le bruit de fond) est la variable cible, de la manière suivante :

$$h_m = h_{m-1} - \gamma b_m(x), \gamma > 0 \quad (4.5)$$

où γ est le taux d'apprentissage, et $b_m(x)$ est le gradient de la fonction de perte défini par

$$b_m(x) = \left(\frac{\partial L(y, h(x))}{\partial h(x)} \right)_{h(x)=h_{m-1}(x)} \quad (4.6)$$

estimé à l'aide d'une régression. Le gradient boosting utilisé pour *MV2* est implémenté à l'aide de la bibliothèque TMVA [56] de ROOT, qui se sert de la déviance binomiale $L(y, h(x)) = \log(1 + \exp(-2h(x)y))$ comme fonction de perte, et des arbres de décision comme régressions pour estimer les gradients $b_m(x)$. La première étape de l'algorithme initialise le classifieur à une valeur nulle, puis procède à l'actualisation de $h(x)$ comme indiqué en (4.5) jusqu'à ce que le nombre d'itération ait atteint le nombre maximal d'arbres de décision défini. Ce nombre est fixé à 400 par défaut pour l'entraînement *MV2*. Une normalisation finale garantit que les valeurs du classifieur sont comprises entre -1 et $+1$. Enfin, une coupure sur *MV2* permet de classifier les événements.

Un arbre de décision, à la $m^{\text{ème}}$ itération, fragmente l'espace des variables d'entrée en un ensemble de régions, ou feuilles, obtenues par successions de coupures de type ' $x > a$ ', où x et a représentent une variable d'entrée et sa valeur de coupure associée, puis assigne à chaque feuille un estimateur $\hat{b}_m(x)$ de la variable cible $b_m(x)$. Cet estimateur est égal à la valeur moyenne de $b_m(x)$ sur les jets de la feuille.

L'arbre de décision croît à partir d'une première coupure sur l'échantillon d'entrée, qui sépare ce premier noeud en deux nouveaux noeuds. Ces derniers sont à leur tour séparés en deux, et ainsi de suite. Comme pour les feuilles, la valeur moyenne de $b_m(x)$ est calculée pour chaque noeud, ainsi que son erreur quadratique moyenne. En un noeud, la coupure choisie minimise l'erreur quadratique moyenne des deux noeuds de sortie. Certains critères permettent de limiter l'expansion de l'arbre en arrêtant la division de certains noeuds, alors appelés feuilles. En particulier, si les événements d'un noeud ont passé un certain nombre n_{depth} coupures, le noeud devient feuille. Une profondeur maximale $n_{depth} = 10$ est utilisée pour l'entraînement *MV2*. L'expansion d'un noeud s'arrête également si celui-ci contient un nombre d'événement inférieur à une fraction, appelée MNS (Minimum Node Size), du nombre d'événements de l'entraînement. Le MNS est fixé à 0.5% pour l'entraînement *MV2*.

4.2.5 Candidat b

Les candidats b sont des jets qui possèdent un poids de sortie du classifieur $MV2$ minimal, ou coupure inférieure sur le poids $MV2c20$. Les performances d'identification des jets de b sont quantifiées en terme d'efficacité de jet léger, de b et de c , correspondant respectivement à la fraction de jets légers, de b et de c d'événements $t\bar{t}$ qui passent une telle coupure. Quatre coupures, ou *points de fonctionnement*, sont mises à disposition pour les analyses d'ATLAS, correspondant à des efficacités de jet de b de 60%, 70%, 77%, et 85% respectivement. Pour chaque point de fonctionnement, les efficacités estimées de jet léger, de b et de c sont calibrées dans les données en fonction de l'impulsion transverse et de la pseudo-rapacité des jets, à partir d'événements candidats $t\bar{t}$ [57] et $W + c$ [58]. Les *réjections* (inverse des efficacités) de jet léger et de jet de c sont présentées dans la Table 4.3 pour ces points de fonctionnement. Les candidats b utilisés depuis 2016 sont étiquetés avec une coupure sur $MV2c10$, après une modification de $MV2$ qui sera décrite dans cette étude.

ϵ_b	$1/\epsilon_u$	$1/\epsilon_c$
60%	1163	37
70%	279	13
77%	100	7
85%	27	3

TABLE 4.3: Efficacités d'identification de jet de b pour les quatre points de fonctionnement fournis par le groupe d'étiquetage des jets de b et leur réjection de jet léger et de jet de c associées.

4.3 Etude de la sélection des traces de JF à haute impulsion

L'identification des jets de b réalisée dans l'expérience ATLAS est fortement dépendante de l'impulsion du jet. Elle atteint son maximum en terme de performances pour des jets d'impulsion transverse autour de 100 GeV, comme c'est le cas pour JetFitter (Figure 4.10). Les deux principales causes à l'origine de la dégradation des performances de l'identification sont d'une part l'interaction particule-matière, qui prédomine à basse impulsion de jet ($p_{T,jet} < 100$ GeV), et d'autre part la densité élevée de particules produites au sein des jets de haute impulsion ($p_{T,jet} > 100$ GeV). A basse impulsion de jet, les particules du jet considéré sont émises avec une plus faible impulsion, et sont plus susceptibles d'être déviées de leur trajectoire initiale (par diffusion élastique) ou de produire des particules secondaires (par collision inélastique) en interagissant avec le détecteur ou le tube faisceau. Les diffusions élastiques altèrent notamment la résolution des paramètres d'impact des traces, tandis que les collisions inélastiques produisent de potentiels vertex secondaires, ce qui conduit à une détérioration de l'identification des jets de b . D'autre part, par effet cinématique, plus l'impulsion d'un jet augmente, plus le nombre de traces produites augmente, et plus les

traces sont émises dans une même direction. Il en résulte un rapprochement des traces les unes par rapport aux autres, et une augmentation de la densité de traces au coeur du jet. Par conséquent, la reconstruction des traces et des vertex secondaires est dégradée, avec par exemple la présence de coups partagés ou divisés au sein des traces et de vertex secondaires reconstruits avec des traces de fragmentation. Par ailleurs, dans de nombreux processus physiques, les jets produits à haute impulsion ne sont pas isolés, mais entourés d'autres jets, à l'instar des productions de quark *top*. Lorsqu'un quark *top* possède une impulsion transverse grande devant sa masse, les trois jets issus de sa désintégration hadronique peuvent se chevaucher, ce qui ajoute encore de la complexité dans l'identification des jets de *b*.

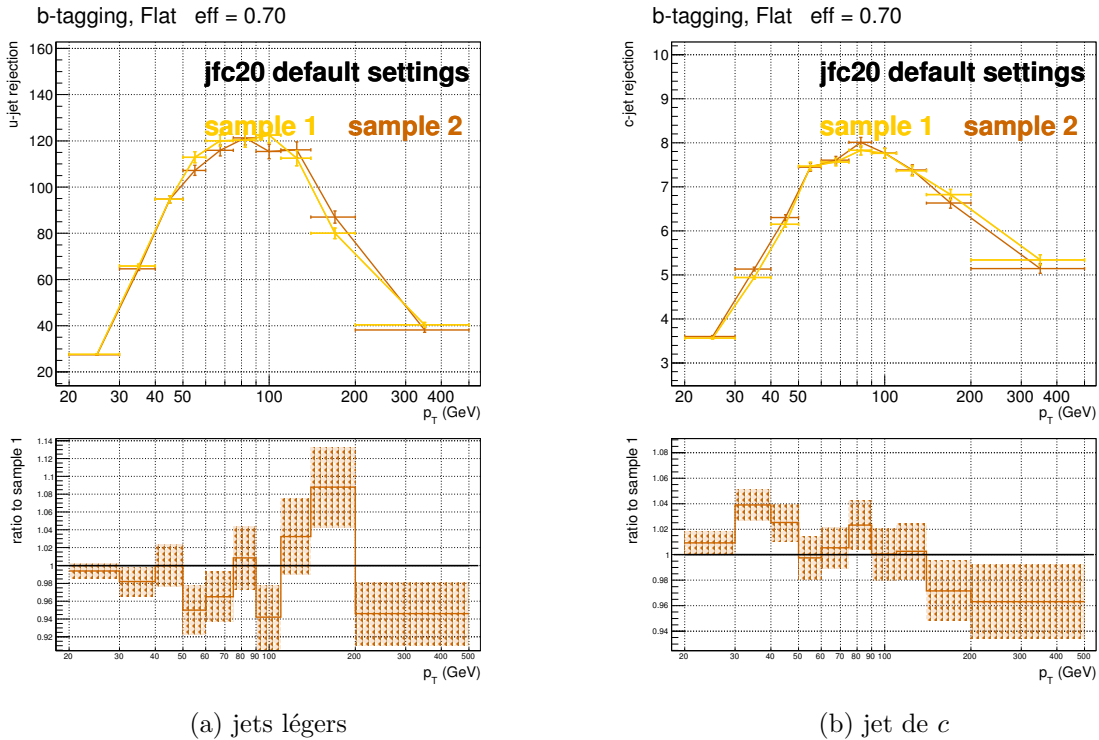


FIGURE 4.10: Réjection de jet léger (a) et jet de *c* (b) en fonction du p_T du jet. Dans chaque bin, l'efficacité de jet de *b* est fixée à 70%. Deux performances sont obtenues avec deux BDT entraînés sur des échantillons statistiquement indépendants, avec comme variables d'entrée les variables de l'algorithme JetFitter. Les jets sont collectés à partir d'un échantillon de $t\bar{t}$.

La dégradation de la reconstruction des vertex secondaires par JetFitter à mesure que l'impulsion du jet augmente est visible sur la Figure 4.11, avec une augmentation linéaire de la fraction de jets légers qui possèdent un ou plusieurs vertex secondaires (Figure 4.11b), tandis que cette même fraction reste constante pour les jets de *b* (Figure 4.11a), et ce à partir de 100 GeV environ. De plus, les vertex reconstruits au sein des jets légers de haute impulsion ressemblent de plus en plus à ceux des jets de *b*, car la fraction de vertex secondaires formés d'au moins deux traces augmente. Afin d'améliorer les performances de JetFitter à haute impulsion, il faut chercher à réduire la fraction de jets légers qui présentent un ou plusieurs

vertex secondaires en sortie de l'algorithme de JetFitter, tout en maintenant une bonne efficacité de reconstruire les vertex secondaires des jets de b , en particulier ceux provenant de la désintégration de hadrons b .

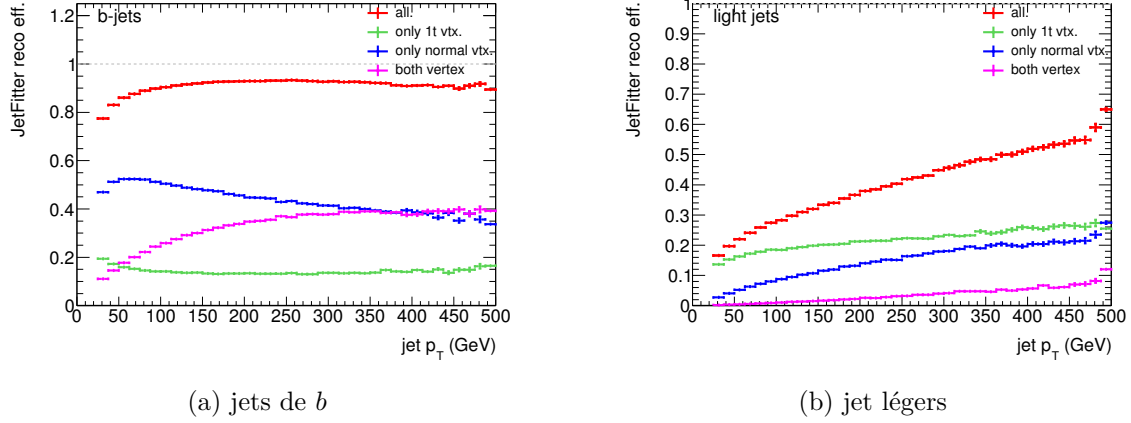


FIGURE 4.11: Efficacité de reconstruire un ou plusieurs vertex avec l'algorithme JetFitter en fonction du p_T du jet, pour les jets de b (a) et les jets légers (b). Les jets sont séparés selon le type de vertex secondaires : les jets possédant **seulement des vertex à une trace** (vert), **seulement des vertex à au moins deux traces** (bleu), les jets possédant **à la fois des vertex à une et au moins deux traces** (violet), et les jets contenant **tout type de vertex** (rouge). Les jets sont collectés à partir d'un échantillon $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1$ TeV).

Pour atteindre cet objectif, l'étude présentée dans cette section explore une sélection alternative des traces utilisées pour reconstruire les vertex secondaires, avec l'implémentation d'une coupure, ou seuil, ne retenant que les traces au-dessus d'une certaine impulsion transverse proportionnelle à l'impulsion du jet considéré. En effet, les traces de hadrons b ont une impulsion moyenne supérieure aux traces de fragmentation, et cette impulsion augmente avec l'impulsion du jet (Figure 4.12). Il semble alors approprié d'accroître linéairement la coupure sur l'impulsion transverse des traces avec l'impulsion du jet. La dépendance des impulsions transverses moyennes des traces avec l'impulsion transverse du jet associé découle de celle des impulsions transverses moyennes de hadron b avec l'impulsion transverse du jet associé (Figure 4.14a). Les prédictions sont estimées à l'aide de jets collectés à partir d'échantillons $Z \rightarrow t\bar{t}$. Ce processus est adéquat pour l'étude de jets de très haute impulsion ($p_T > 300$ GeV), car les jets sont issus de désintégrations de bosons Z' leptophobiques très massifs (de masses $m_{Z'} = O(\text{TeV})$) [59]. La statistique disponible en jets de haute impulsion d'échantillons $Z \rightarrow t\bar{t}$ est donc bien supérieure à celle d'échantillons $t\bar{t}$.

En pratique, effectuer cette coupure revient à appliquer une coupure constante sur le rapport entre l'impulsion transverse de la trace et l'impulsion transverse du jet, noté $f_{p_T}^{trk,j}$, dont la distribution pour des jets d'impulsion transverse supérieure à 300 GeV est présentée Figure 4.13. Les traces de fragmentation, qu'elles proviennent des jets de b ou des jets légers, ont la propriété d'être plus souvent émises à faible $f_{p_T}^{trk,j}$ ($< 1\%$) en comparaison avec les

traces des désintégrations de hadrons b .

Par la suite, l'effet sur JetFitter de coupures sur les rapports des impulsions transverses des traces divisées par la somme scalaire des impulsions transverses des traces associées au jet, noté $f_{p_T}^{trk}$, sera considéré au lieu de $f_{p_T}^{trk,j}$. La nouvelle variable possède l'avantage d'être indépendante de la calibration en énergie des jets. En revanche, la somme scalaire des impulsions transverses des traces associées au jet, bien que proportionnelle à l'impulsion du jet (Figure 4.14b) est moins corrélée avec l'impulsion transverse d'une trace de hadron b associée pour un jet de b .

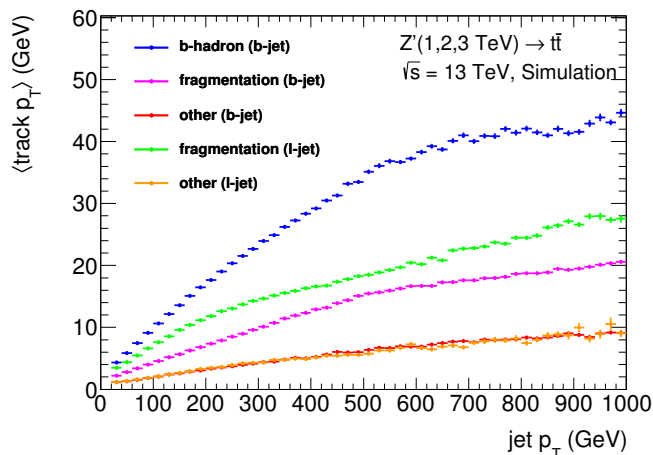


FIGURE 4.12: p_T moyen d'une trace en fonction du p_T du jet associé. Les traces considérées sont séparées selon leur jet d'origine, et sont collectées à partir d'un échantillon $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1 \text{ TeV}$).

Trois coupures sur le rapport $f_{p_T}^{trk}$ sont testées : une valeur centrale à 1.5%, une sélection plus stricte à 3.5%, et une dernière plus lâche à 0.5%. Lorsque le jet possède une somme scalaire des impulsions transverses des traces le composant supérieure à 600 GeV, l'impulsion minimale d'une trace sélectionnée par JetFitter est fixée à $f_{p_T}^{trk} \times 600 \text{ GeV}$. Pour chaque coupure, la reconstruction des vertex secondaires par JetFitter et l'estimation des performances sont effectuées. Cette estimation est réalisée à l'aide d'un BDT, entraîné avec les mêmes huit variables de JetFitter utilisées en entrée de l'entraînement $MV2$.

Figure 4.15 montre les performances de ces trois configurations de JetFitter, comparées avec l'algorithme par défaut, qui n'applique pas de coupure sur le rapport $f_{p_T}^{trk}$. Tout d'abord, on remarque que plus la coupure sur le rapport $f_{p_T}^{trk}$ est élevée, plus l'efficacité de reconstruire au moins un vertex secondaire est faible, conduisant à de moins bonnes performances à haute efficacité de jet de b . Cet effet est donc plus marqué pour la coupure à $f_{p_T}^{trk} = 3.5\%$. De même, plus la coupure sur $f_{p_T}^{trk}$ augmente, plus les performances de JetFitter sont bonnes à basse efficacité de jet de b . Pour une efficacité de jet de b de 60%, une hiérarchie des performances se profile, avec de meilleures rejections de jet léger pour la coupure à 1.5%, suivie de celle à 0.5% suivie de l'algorithme par défaut et de la coupure à 3.5%. En outre,

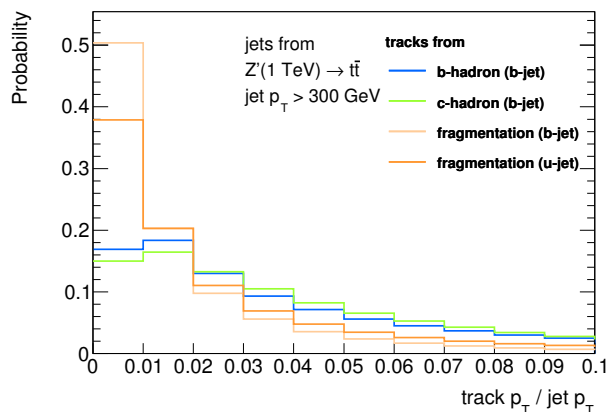
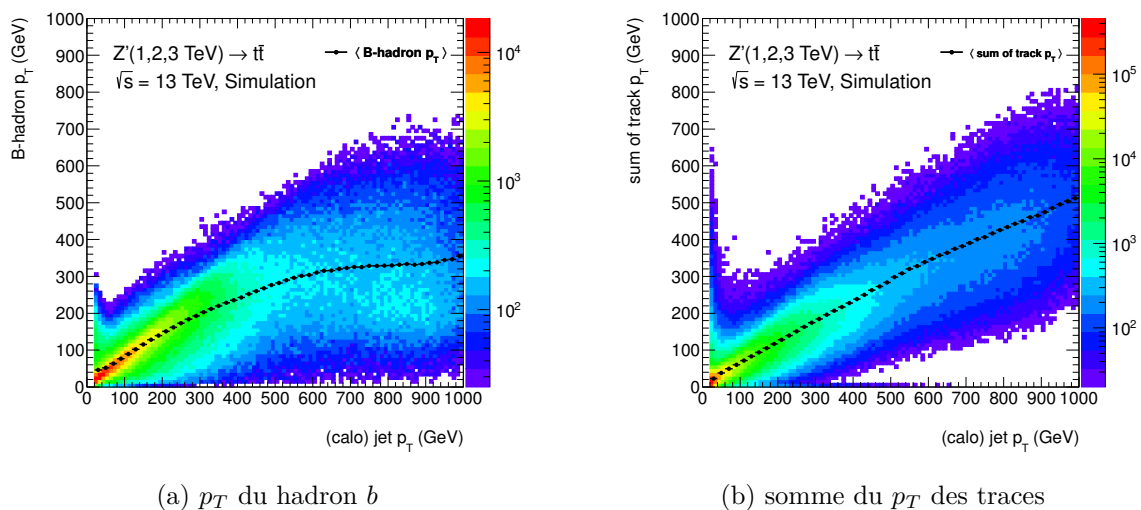


FIGURE 4.13: Distributions du p_T des traces divisés par le p_T du jet associé ($f_{p_T}^{trk,j}$), pour des traces de fragmentation provenant de jet légers et de jets de b , et des traces issues des désintégrations de hadrons b et c au sein de jets de b . Les traces sont collectées à partir de jets de $p_T > 300$ GeV d'un échantillon $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1$ TeV).



(a) p_T du hadron b

(b) somme du p_T des traces

FIGURE 4.14: Distributions 2D du p_T des hadrons b (a) et de la somme des p_T des traces (b) en fonction du p_T du jet associé. Les points noirs représentent la valeur moyenne du p_T par bin de jet p_T . Les jets considérés sont obtenus à partir d'un échantillon de $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1, 2, 3$ TeV).

les modifications des performances fluctuent avec l'impulsion du jet et ne montrent pas de tendance particulière.

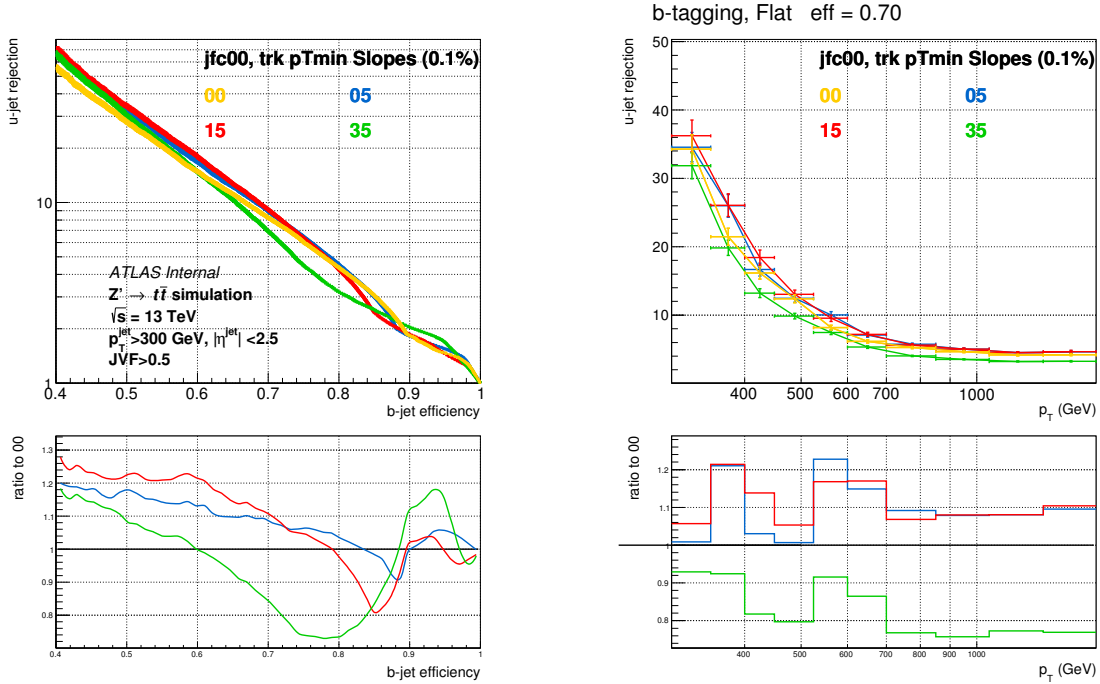


FIGURE 4.15: Réjection de jet léger en fonction de l'efficacité de jet de b (gauche) et réjection de jet léger en fonction du p_T du jet pour une efficacité de jet de b de 70% à chaque bin de jet p_T (droite). Les performances sont obtenues avec un BDT dont les variables d'entrée sont les variables de l'algorithme JetFitter, et pour différentes coupures sur le rapport $f_{p_T}^{trk}$: 0.5, 1.5 et 3.5 %, comparées à l'algorithme par défaut sans coupure sur $f_{p_T}^{trk}$. Les jets de $p_T > 300$ GeV sont collectés à partir d'un échantillon de $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 0.4, 1, 3, 4$ TeV).

Les résultats obtenus sont encourageants, car ils montrent qu'il est possible d'améliorer JetFitter pour des jets de haute impulsion en étant plus sélectif sur l'impulsion transverse des traces utilisées par l'algorithme pour reconstruire les vertex secondaires. Cependant, les coupures proposées ont un impact modéré sur l'amélioration des performances de JetFitter, en deçà des attentes. Etant donné l'environnement complexe des événements $Z' \rightarrow t\bar{t}$ utilisés, il serait utile de poursuivre ces études en se plaçant dans un cadre simplifié, dans lequel les jets seraient isolés. Ces jets peuvent être obtenus à partir d'échantillons *di-jet* pour le bruit de fond et de $Z' \rightarrow b\bar{b}$ pour le signal. En effet, l'échantillon $Z' \rightarrow t\bar{t}$ donne de nombreux jets de b reconstruits à partir de l'hadronisation de quarks issus de la désintégration du même quark *top* ($t \rightarrow bW \rightarrow bq\bar{q}'$), et ce d'autant plus que l'impulsion des jets de b considérés augmente. Cela a pour effet de diminuer la fraction de l'impulsion du jet portée par le hadron b , et par conséquent celle des particules de sa désintégration, qui devient plus proche de celle des particules issues du vertex primaire. Une coupure sur l'impulsion des traces proportionnelle

à l'impulsion du jet devient alors moins pertinente.

Une autre amélioration possible serait de modifier la coupure sur l'impulsion transverse des traces afin qu'elle ne soit non pas proportionnelle à la somme scalaire de l'impulsion transverse des traces associées au jet, mais proportionnelle à l'impulsion transverse du jet lui-même. Il faut noter enfin que l'idée originelle de JetFitter qui consiste, lors de la reconstruction de vertex secondaire, à faire l'hypothèse que le hadron b se désintègre préférentiellement le long de l'axe du jet de b associé, trouve ses limites dans cette topologie complexe, où le jet de b voit son impulsion désaxée par rapport à celle du hadron b parce qu'il n'est pas isolé.

4.4 Investigation de nouvelles variables de JetFitter

Une fois les vertex secondaires reconstruits par l'algorithme JetFitter, des variables de sortie sont calculées à partir des positions et énergie-impulsions de ces vertex secondaires. Elles sont choisies pour leur fort pouvoir discriminant, c'est-à-dire leur faculté à séparer les jets de b des jets de bruit de fond. Par exemple, un vertex secondaire de haute masse calculée par JetFitter a plus de chance de provenir de l'hadronisation d'un quark b qu'un jet de basse masse. C'est pourquoi la masse calculée par JetFitter fait partie des variables discriminantes utilisées en entrée de $MV2$.

En général, le pouvoir discriminant des variables calculées par JetFitter diminue à haute impulsion en raison de la dégradation de la reconstruction des vertex secondaires. Les vertex secondaires de jets de haute impulsion sont en effet plus souvent formés de traces de fragmentation ou/et de fausses traces. Parmi l'ensemble des huit variables de sortie de JetFitter utilisées en entrée de $MV2$, deux d'entre elles ont été étudiées dans cette partie car leur discrimination est particulièrement affectée à haute impulsion de jet : d'une part, la distance angulaire entre l'impulsion du jet et celle des vertex secondaires, notée $\Delta R_{jet,SV}$, et d'autre part, la masse des vertex secondaires. Les jets étudiés proviennent d'événements simulant la production d'hypothétiques nouveaux bosons de jauge lourds Z' se désintégrant en une paire $t\bar{t}$. Trois échantillons avec une masse pour le boson Z' de 1, 2 et 3 TeV sont considérés pour couvrir un large spectre en impulsion transverse de jet. Pour l'étude des variables angulaires de JetFitter, seuls les jets ayant au moins un vertex secondaires à plusieurs traces seront considérés. En effet, les variables angulaires ne sont pas discriminantes pour les jets ne possédant que des vertex secondaires à une trace.

4.4.1 Variables angulaires de JetFitter

La variable angulaire $\Delta R_{jet,SV}$ (définition Table 4.2) exploite essentiellement la propriété qu'a le hadron b d'emporter une grande fraction de l'énergie du jet associé (70 % en moyenne), et par conséquent d'être émis dans la direction de ce dernier. Il en résulte que l'angle $\Delta R_{jet,SV}$ est plus petit pour des jets de b que pour les jets légers. Pourtant il arrive que l'impulsion d'un hadron ne soit que partiellement reconstruite, car les particules

issues de la désintégration de ce hadron ne sont pas toutes associées à un vertex secondaire (particules neutres, particules chargées dont la trace ne satisfait pas les critères de sélection de JetFitter). Cela entraîne une dispersion de la distribution en $\Delta R_{jet,SV}$ vers les grands angles. Cette dispersion est plus prononcée pour les jets de b que les jets de c , en partie parce que le hadron c se désintègre en moins de particules que le hadron b . Pour cette raison la distribution de l'angle $\Delta R_{jet,SV}$ apparaît presque identique pour les jets de b et de c , malgré le fait que la direction du hadron c soit moins alignée avec le jet associé que ne l'est celle du hadron b (Figures 4.16a et 4.16b).

Les distributions de l'angle $\Delta R_{jet,SV}$ des jets de b , c et légers sont très sensibles à l'environnement du jet. On peut identifier deux régimes. D'une part, le régime résolu pour lequel les jets sont relativement isolés (à basse impulsion de jet $p_{T,jet} < 200$ GeV), et d'autre part, le régime boosté avec des jets très proches les uns des autres ($p_{T,jet} > 400$ GeV). Tant que les jets sont dans un régime résolu, la valeur de l'angle $\Delta R_{jet,SV}$ pour les jets de b et pour les jets de c est distribuée de manière identique, et les jets légers dominent à grand $\Delta R_{jet,SV}$ (Figures 4.16a et 4.16b). Plus l'impulsion transverse du jet augmente, plus l'angle $\Delta R_{jet,SV}$ diminue par effet de boost. Lorsque les jets atteignent le régime boosté, la hiérarchie des angles $\Delta R_{jet,SV}$ pour les jets de b/c et les jets légers est modifiée. Les premiers signes d'une fusion des jets issus du boson W en un unique jet deviennent visibles avec l'augmentation de l'angle $\Delta R_{jet,SV}$ pour les jets de c autour de 400 GeV (Figure 4.16c). La direction du jet de c s'identifie alors avec celle du boson W et non plus avec celle du quark c provenant du même boson W . Par conséquent l'impulsion des vertex secondaires provenant de la désintégration de hadron c est alors plus éloignée de l'axe du jet associé. Pour des jets d'impulsion transverse encore plus élevée (Figure 4.16d), un phénomène de fusion semblable se produit affectant cette fois-ci les jets de b . Les jets issus de la désintégration du boson W se recouvrent avec le jet de b issu de la désintégration du quark top associé pour n'en former qu'un. A son tour, le jet de b voit son angle $\Delta R_{jet,SV}$ augmenter, tandis que celui du jet de c diminue, par effet de boost.

Dans l'étude présentée ici, une variable supplémentaire a été introduite pour apporter une information supplémentaire sur les vertex secondaires reconstruits par JetFitter tout en étant moins sensible à la direction du jet. Il s'agit de l'angle entre l'axe JetFitter et l'impulsion des vertex secondaires, notée $\Delta R_{JF,SV}$. L'axe JetFitter est ajusté au cours de la recherche de vertex secondaires par l'algorithme afin de passer par les positions des vertex secondaires et du vertex primaire. Par exemple, lorsque l'algorithme JetFitter reconstruit un unique vertex secondaire au sein d'un jet, l'axe JetFitter est la droite qui passe par le vertex primaire et le vertex secondaire, représentée Figure 4.17. L'angle $\Delta R_{JF,SV}$ est particulièrement intéressant pour rejeter les jets de c , comme il sera montré par la suite.

Une part importante des vertex secondaires à deux traces dans les jets légers est originaire de désintégrations de V^0 et de conversions de photons produits au vertex primaire. L'impulsion de ces vertex secondaires est égale à l'impulsion des V^0/γ d'origine, et par conséquent sa direction est égale à l'axe JetFitter, modulo les variations dues à la résolution sur l'impulsion des traces et la position des vertex. Ainsi les jets légers de faible

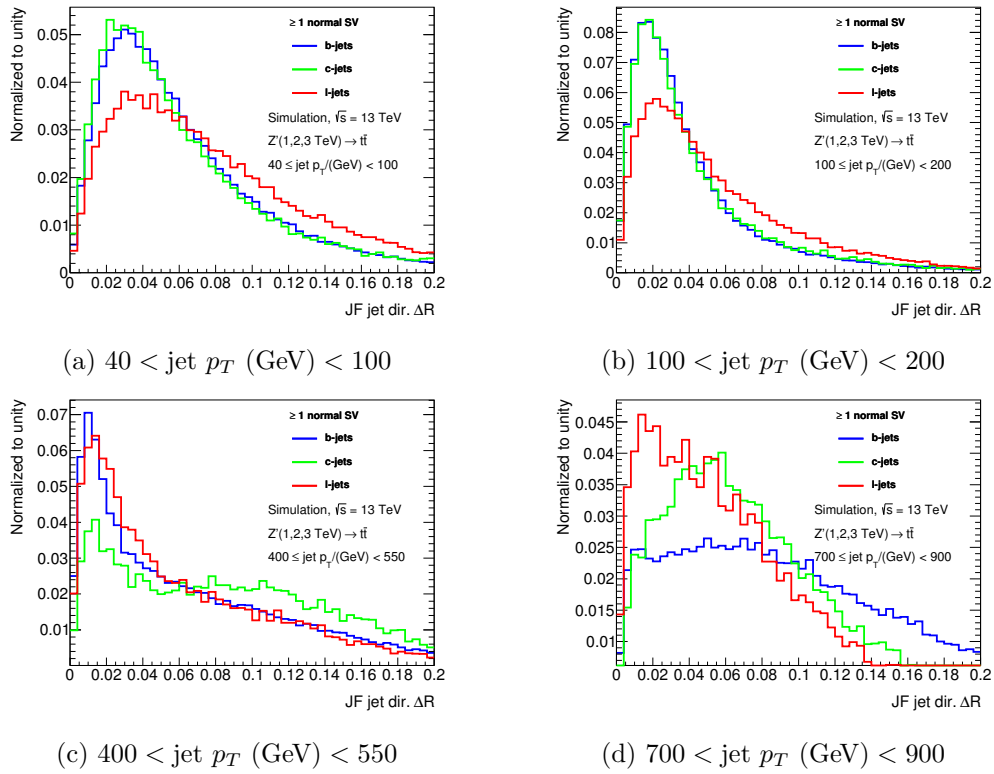


FIGURE 4.16: Distributions de la variable $\Delta R_{jet,SV}$ entre l'impulsion du jet et celle du/des vertex reconstruit(s) par JetFitter, pour des jets légers (trait rouge), des jets de c (trait vert) et des jets de b (trait bleu). Chacune des quatre figures correspond à un intervalle en impulsion transverse de jet. Les jets sont issus d'un échantillon $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1, 2, 3 \text{ TeV}$).

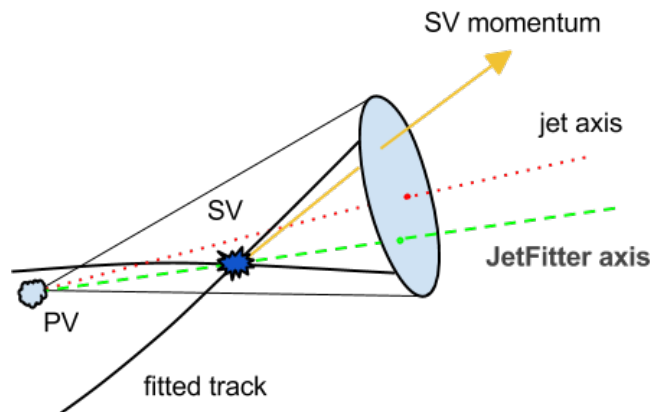


FIGURE 4.17: Schéma d'un vertex secondaire reconstruit au sein d'un jet à partir de deux traces, avec son vecteur impulsion, l'axe JetFitter et la direction du jet associé.

impulsion peuplent fréquemment les faibles valeurs de $\Delta R_{JF,SV}$ (autour de $\sim 2 \times 10^{-3}$), tandis que les jets de b sont plus fréquents à plus grand angle ($\Delta R_{JF,SV} \sim 1.5 \times 10^{-2}$), notamment à cause des particules neutres émises lors de la désintégration des hadrons b (Figure 4.18a). Les angles $\Delta R_{JF,SV}$ des jets de c se situent en moyenne entre ceux des jets légers et ceux des jets de b . Les $\Delta R_{JF,SV}$ de chacune des saveurs de jet tendent à se rapprocher de 0 avec l'augmentation de l'impulsion transverse du jet, sans que les formes des distributions ne soient drastiquement modifiées à très haute impulsion de jet, *i.e* lorsque les jets d'un même quark *top* commencent à fusionner (Figures 4.18c et 4.18d). L'identification des jets de b fondée sur cette variable est donc moins sensible à l'environnement du jet que dans le cas de l'angle $\Delta R_{jet,SV}$, et donc moins dépendante de la topologie de l'événement considéré.

Pour mesurer l'effet de l'angle $\Delta R_{JF,SV}$ sur les performances de JetFitter, un BDT est entraîné avec et sans cette variable en entrée, en plus des huit variables JetFitter qui sont utilisées en entrées de *MV2* (cf. Section 4.4.3). On peut s'attendre à ce que la corrélation entre l'angle $\Delta R_{JF,SV}$ et la masse JetFitter soit exploitée par l'algorithme multivarié. Par exemple, lorsque le $\Delta R_{JF,SV}$ est très faible pour des jets de c ou de b , de l'ordre de 10^{-3} , cela implique que le hadron c ou b a plus de chance de se désintégrer exclusivement en particules chargées. La résolution sur la masse de ces hadrons devient meilleure et la séparation des jets de b et de c est par conséquent améliorée (Figure 4.19). Ces améliorations seront quantifiées par la suite.

4.4.2 Masse SV de JetFitter

La masse des vertex secondaires reconstruits par JetFitter (définition Table 4.2) est naturellement plus grande en moyenne pour les jets de b ($m_b \sim 5$ GeV) que pour les jets de c ($m_c \sim 2$ GeV) et plus encore pour les jets légers. La résolution sur la masse des hadrons b et c est assez mauvaise en raison des particules neutres émises dans leur désintégration. A basse et moyenne impulsion, un jet de grande masse calculée par JetFitter a plus de chance d'être un jet de b qu'un jet léger (Figures 4.20a et 4.20b). Néanmoins, c'est moins le cas lorsque le jet gagne en impulsion, et que les masses calculées par JetFitter deviennent de plus en plus élevées. Les jets légers, dont une grande part possède de faux vertex secondaires, ont une masse qui croît plus vite avec l'impulsion transverse du jet. La queue dans la distribution de la masse calculée par JetFitter des jets légers dégrade alors la discrimination entre ces derniers et les jets de b (Figures 4.20c et 4.20d). A une impulsion transverse de jet d'environ 500 GeV, des problèmes liés à la topologie de l'événement apparaissent, avec par exemple des jets de c dont la masse est plus élevée que celle de jets de b , car les jets de c contiennent l'hadronisation des deux quarks provenant de la désintégration d'un boson W .

La masse non corrigée, considérée ici comme une alternative à la masse calculée par JetFitter, est potentiellement intéressante car elle n'introduit pas le terme $2 \sum_{tracks \in SV} p_{T,track}^{JF \text{ axis}}$ présent dans le calcul de la masse standard (définie à la sous-section 4.2.2), notamment res-

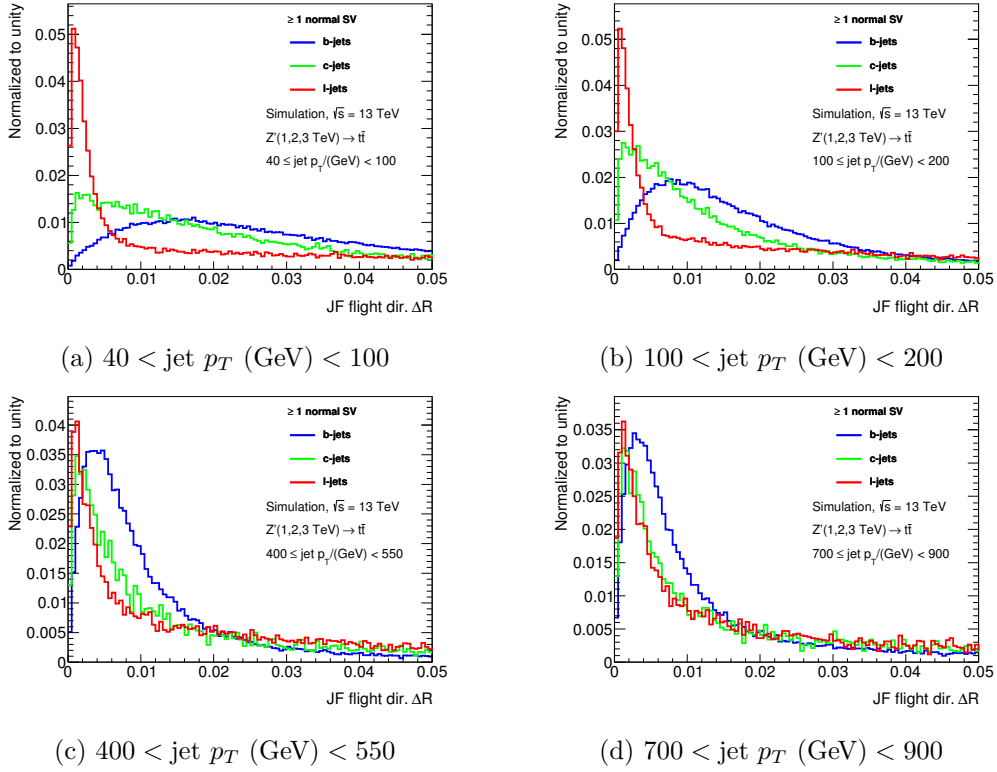


FIGURE 4.18: Distributions du $\Delta R_{JF,SV}$ entre l'axe JetFitter et l'impulsion du/des vertex reconstruit(s) par JetFitter, pour des jets légers (trait rouge), des jets de c (trait vert) et des jets de b (trait bleu). Chacune des quatre figures correspond à un intervalle en impulsion transverse de jet Δp_T . Les jets sont issus d'un échantillon $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1, 2, 3$ TeV).

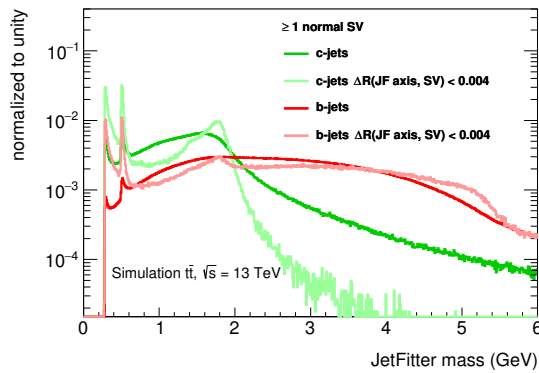
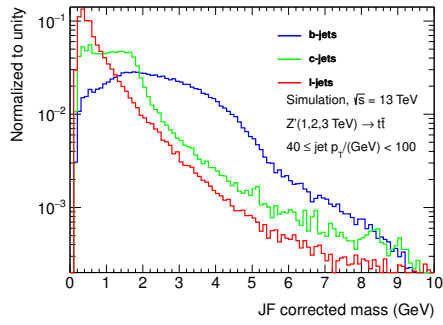
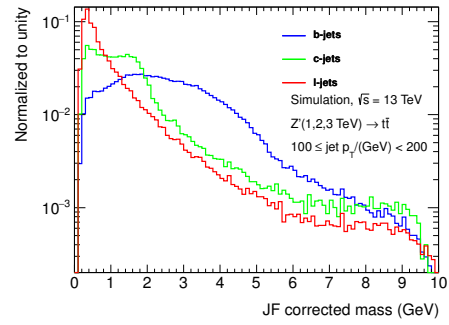


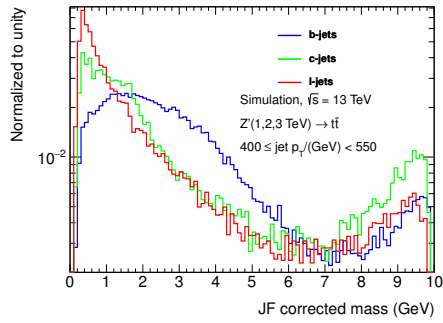
FIGURE 4.19: Distribution de la masse calculée par JetFitter pour des jets de b et de c qui possèdent au moins un vertex secondaire à au moins deux traces, avec et sans la coupure $\Delta R_{JF,SV} < 0.004$.



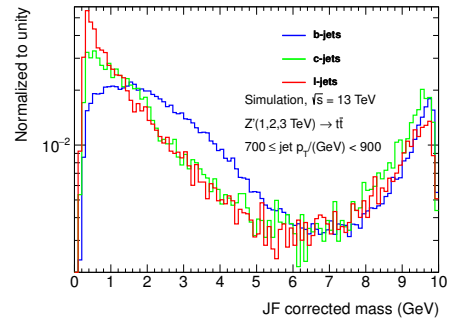
(a) $40 < \text{jet } p_T$ (GeV) < 100



(b) $100 < \text{jet } p_T$ (GeV) < 200



(c) $400 < \text{jet } p_T$ (GeV) < 550



(d) $700 < \text{jet } p_T$ (GeV) < 900

FIGURE 4.20: Distributions de la masse calculée par JetFitter, pour des jets légers (trait rouge), des jets de c (trait vert) et des jets de b (trait bleu). Chacune des quatre figures correspond à un intervalle en impulsion transverse de jet. Les jets sont issus d'un échantillon $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1, 2, 3$ TeV).

possible des larges queues de distribution à grande masse (Figure 4.21). Les queues de distributions sont toujours présentes à haute impulsion transverse de jet mais sont moins prononcées que pour la masse standard. La masse non corrigée, prise individuellement, possède une meilleure discrimination que la masse standard.

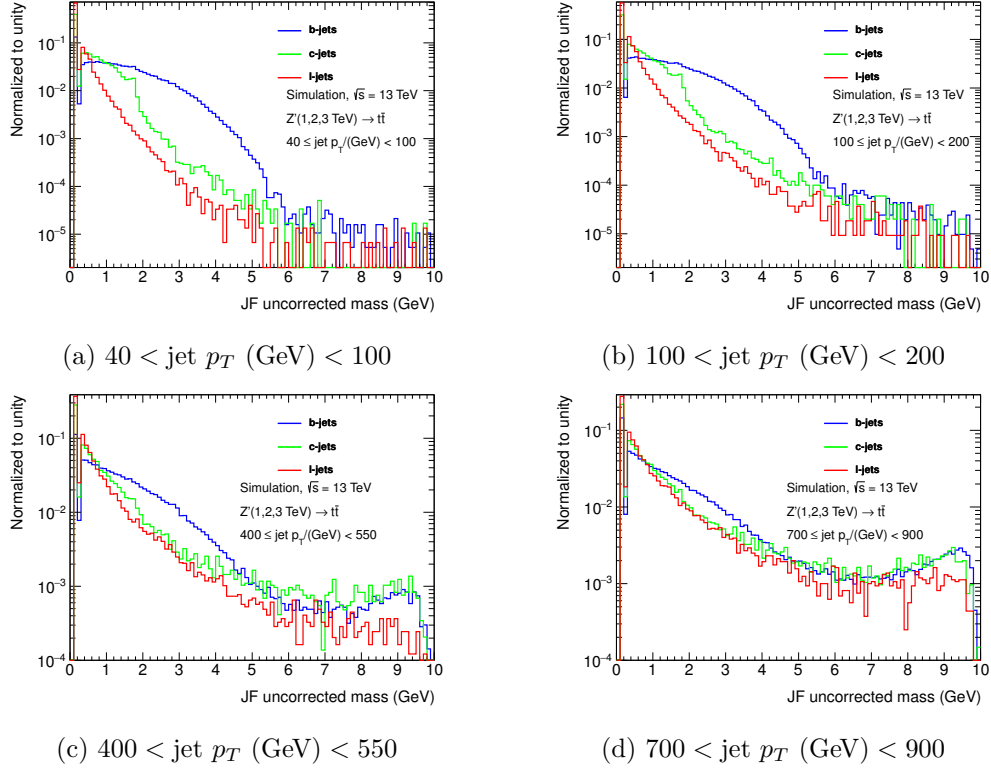


FIGURE 4.21: Distributions de la masse non corrigée de JetFitter, pour des jets légers (trait rouge), des jets de c (trait vert) et des jets de b (trait bleu). Chacune des quatre figures correspond à un intervalle en impulsion transverse de jet. Les jets sont issus d'un échantillon $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1, 2, 3 \text{ TeV}$).

4.4.3 Amélioration apportée par les nouvelles variables

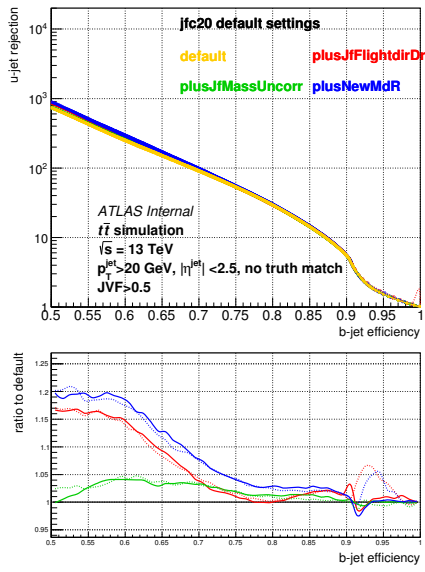
Les deux variables de JetFitter proposées dans cette étude, à savoir l'angle $\Delta R_{JF,SV}$ et la masse non corrigée, sont toutes deux fortement corrélées respectivement à l'angle $\Delta R_{jet,SV}$ et à la masse standard. Elles apportent toutefois une information supplémentaire, qui se traduit par une amélioration significative des performances de JetFitter. Pour évaluer cette amélioration, quatre BDT sont entraînés en présence ou en l'absence des nouvelles variables. Le premier BDT est entraîné avec les huit variables de JetFitter présentes dans la liste des variables d'entrée du BDT $MV2$ par défaut, et ses performances servent de référence. Les trois autres BDT sont entraînés non seulement avec ces mêmes huit variables,

2. Le terme est proportionnel en moyenne à la somme scalaire des impulsions des traces des vertex secondaires et donc, grossièrement, à l'impulsion du jet.

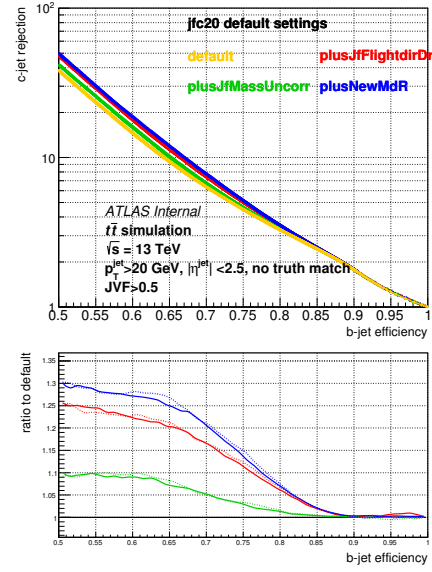
mais avec soit l'angle $\Delta R_{JF,SV}$, soit la masse non corrigée, soit les deux. Les réglages internes des quatre BDT sont identiques à ceux du BDT $MV2$. Les procédures de pondération de jets sont également inchangées, sauf pour la pondération des mauvais jets, qui n'est pas appliquée ici.

Tout d'abord, ajouter l'angle $\Delta R_{JF,SV}$ devient intéressant pour rejeter les jets légers à des efficacités de jet de b en dessous des 65%. Dans cette région l'amélioration sur la réjection des jets légers dépasse les 10% pour atteindre un gain de 15% au point de fonctionnement de 60% (Figure 4.22a). Cette tendance est accompagnée d'une amélioration de la réjection de jet de c , avec un gain d'environ 20% à une efficacité de jet de b de 60% (Figure 4.22b). Ce gain est d'autant plus remarquable que l'amélioration de la réjection des jets de c est souvent difficile, car les désintégrations de hadrons b et c sont très similaires. Les bénéfices liés à l'addition de cette variable sont présents à haute impulsion transverse de jet, et principalement autour de 100 GeV, quelque soit la saveur du jet de bruit de fond (Figure 4.23). L'amélioration apportée par la masse non corrigée est plus faible avec un gain sur la réjection de jet léger d'environ 2% et sur celle des jets de c de 10% pour une efficacité de jet de b de 60% (Figure 4.22). Même si la masse non corrigée est plus discriminante que la masse standard, l'amélioration au sein du BDT est relativement faible, notamment en ce qui concerne la réjection de jet léger. En fait, la masse non corrigée sépare deux composantes : les jets qui ne contiennent qu'un seul vertex à une trace des autres jets, dont la majorité contient au moins un vertex à au moins deux traces et pour lesquels la discrimination entre jets de b et jets légers est meilleure. En incluant la variable du nombre de vertex secondaire à au moins deux traces dans le BDT, ce dernier peut discriminer la masse standard selon que le jet ait ou non un vertex secondaire à au moins deux traces, et donc la discrimination du couple de variables (masse standard, nombre de vertex secondaires à au moins deux traces) est presque équivalente à celle du couple (masse non corrigée, nombre de vertex secondaires à au moins deux traces).

En conclusion l'angle $\Delta R_{JF,SV}$ et la masse non corrigée ont été étudiés pour apporter une solution à la dégradation de l'identification des jets de b pour des jets de haute impulsion, là où les variables de JetFitter : l'angle $\Delta R_{jet,SV}$ et la masse standard, montrent une perte de discrimination. Les variables proposées contribuent à améliorer les performances de JetFitter à haute impulsion, mais aussi à des impulsions transverses de jet autour de 100 GeV. Leur impact sur les performances de $MV2$ serait bien entendu plus faible que sur celles de JetFitter. L'angle $\Delta R_{JF,SV}$ semble être une variable prometteuse, et est dorénavant utilisée dans le développement d'un étiqueteur de jets de c , pour sa capacité à améliorer la séparation des jets de b par rapport aux jets de c . En outre, il serait intéressant de comprendre si, avec l'ajout de l'angle $\Delta R_{JF,SV}$, les performances de JetFitter sont moins dépendantes de la topologie de l'événement. Enfin, la masse non corrigée, lorsqu'elle est combinée avec les autres variables de JetFitter, apporte une amélioration modérée aux performances de JetFitter.

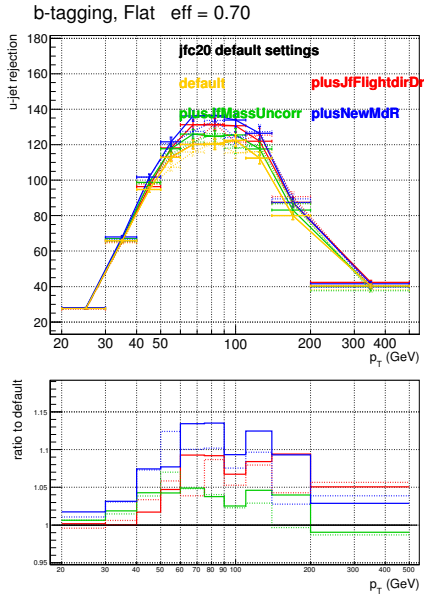


(a) jets légers

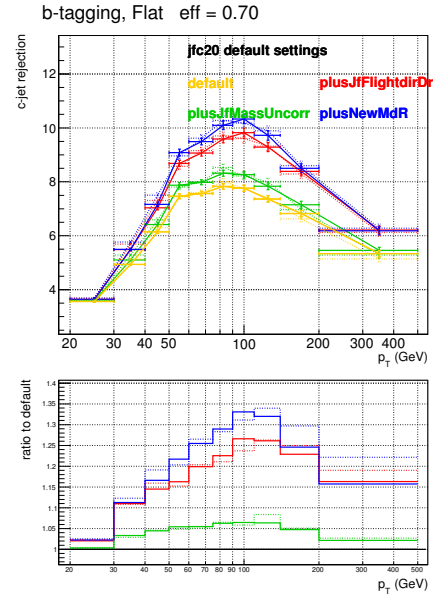


(b) jets de c

FIGURE 4.22: Réjections de jet léger (gauche) et de jet de c (droite) en fonction de l'efficacité de jet de b . Les quatre courbes correspondent à l'évaluation de quatre entraînements indépendants, qui se distinguent par leurs variables d'entrée. L'entraînement par défaut (jaune) est comparé à un entraînement avec l'ajout de la masse non corrigée de JetFitter (vert), ou l'ajout du ΔR (axe JF, vertex) (rouge), ou bien l'ajout de ces deux variables (bleu). Les jets utilisés pour l'entraînement et l'évaluation sont issus d'un échantillon $t\bar{t}$.



(a) jets légers



(b) jets de c

FIGURE 4.23: Réjections de jet léger (gauche) et de jet de c (droite) par intervalle d'impulsion de jet. Dans chaque intervalle, l'efficacité de jet de b est de 70%. Les quatre courbes correspondent à l'évaluation de quatre entraînements indépendants, qui se distinguent par leurs variables d'entrée. L'entraînement par défaut (jaune) est comparé à un entraînement avec l'ajout de la masse de JetFitter non corrigée (vert), ou l'ajout du ΔR (axe JF, vertex) (rouge), ou bien l'ajout de ces deux variables (bleu). Les jets utilisés pour l'entraînement et l'évaluation sont issus d'un échantillon $t\bar{t}$.

4.5 Etude de l'entraînement de l'arbre de décision *MV2*

L'entraînement du BDT *MV2* est une étape déterminante de l'identification des jets de b . L'identification des jets de b dépend non seulement du choix des variables d'entrée et des réglages internes du BDT, mais aussi du choix des jets utilisés pendant l'entraînement, appelés *jets d'entraînement*. Dans cette partie, une nouvelle configuration du BDT *MV2* est étudiée. Avec cette configuration, les mauvais jets ne sont plus négligés pendant la phase d'entraînement. Les motivations et les implications de ce nouvel entraînement sont discutées ci-dessous.

4.5.1 Redéfinition de l'échantillon d'entraînement du BDT *MV2*

Le BDT *MV2* doit être entraîné chaque fois qu'une modification du détecteur ou du logiciel de reconstruction d'ATLAS affecte ses variables d'entrée. Le nouvel entraînement, qui peut nécessiter quelques réglages internes du BDT, est par conséquent plus optimal dans ces nouvelles conditions. L'entraînement du BDT *MV2* présente une instabilité face à de petits changements des conditions initiales de l'entraînement (variables d'entrée, réglages internes, échantillons d'entraînements). Cela se traduit par d'importantes variations des performances de *MV2*, sans aucun signe significatif de surentraînement.

Pour quantifier cette instabilité sur les performances d'identification des jets de b , deux BDT *MV2*, entraînés dans les mêmes conditions mais avec des échantillons d'entraînement statistiquement indépendants, sont comparés. Un échantillon d'entraînement du BDT *MV2* est composé de 2.5×10^5 bons jets et d'une même quantité de mauvais jets. Les bons jets, contrairement aux mauvais jets, possèdent un vertex secondaire par l'algorithme SV1, au moins un vertex secondaire reconstruit par l'algorithme JetFitter et possède au moins une trace associée qui passe la sélection de IPxD. Les sous-échantillons de mauvais jets et de bons jets contiennent chacun 10^5 jets de b , 10^5 jets légers et $5 \cdot 10^4$ jets de c . Cette séparation en bons et mauvais jets permet de gagner du temps CPU pendant l'entraînement du BDT pour l'identification des bons jets. Au cours de l'entraînement du BDT *MV2*, les mauvais jets sont dévalués en leur attribuant un poids presque nul de 10^{-6} . Le BDT se focalise donc essentiellement sur les bons jets. Il est performant pour rejeter les jets légers dont les vertex secondaires ont une topologie proche de celle des jets de b . En revanche, il n'apprend pas à séparer spécifiquement les mauvais jets. Ces derniers constituent pourtant une part non négligeable de l'ensemble des jets produits dans les événements $t\bar{t}$. Par exemple, $\sim 25\%$ des jets de b provenant d'événements $t\bar{t}$ sont de mauvais jets (Table 4.4). Les performances de ce BDT *MV2*, évaluées sur un échantillon $t\bar{t}$, sont donc sous-optimales comme il sera démontré par la suite. Par ailleurs, le BDT *MV2* n'est pas surentraîné, car la distribution du poids de sortie du BDT est identique en utilisant les jets d'entraînements et en utilisant les jets de test³(Figure 4.24).

Les deux BDT *MV2* entraînés avec deux échantillons statistiquement indépendants sont ensuite évalués sur deux échantillons $t\bar{t}$ eux aussi statistiquement indépendants. Les per-

3. Echantillon de test statistiquement indépendant de l'échantillon d'entraînement.

Sélection	bon jet	IPxD valide
jets de b	75.7%	99.8%
jets de c	30.3%	99.8%
jets légers	2.4%	99.4%

TABLE 4.4: Fraction de jets qui passent le critère *bon jet* et qui passent la sélection de trace de IPxD (*IPxD valide*) par saveur de jet b , c et légers, estimée à partir de jets d'échantillons $t\bar{t}$.

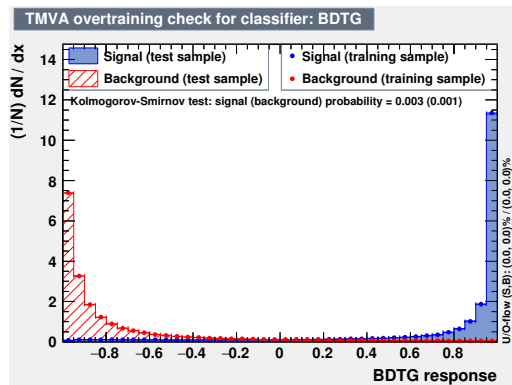


FIGURE 4.24: Test de surentraînement du BDT $MV2$: distribution du poids de sortie du BDT $MV2c10$ pour les jets de signal (bleu) et les jets de bruit de fond (rouge) pour des jets d'entraînement (points) et de test (lignes). L'échantillon d'entraînement considéré ici est celui d'origine, pour lequel les mauvais jets sont dévalués.

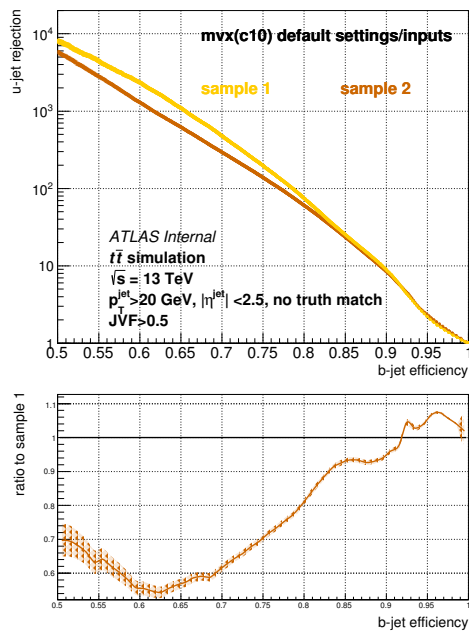
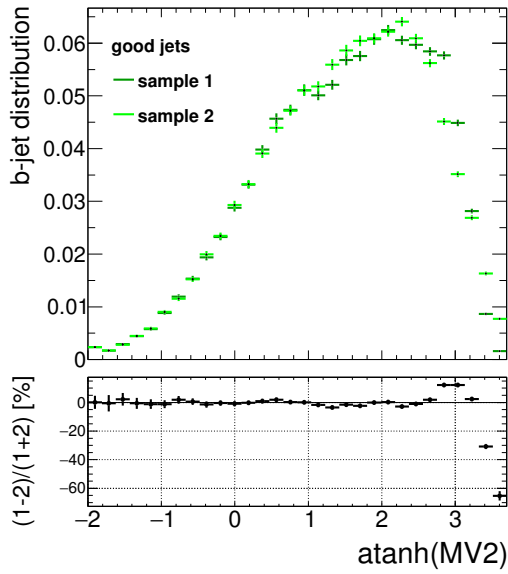


FIGURE 4.25: Réjection des jets légers en fonction de l'efficacité de jet de b évaluée pour deux BDT $MV2c10$ entraînés à partir de deux échantillons statistiquement indépendants, notés "sample 1" et "sample 2" (figure du haut), et leur rapport (figure du bas). L'échantillon d'entraînement considéré ici est celui d'origine, pour lequel les mauvais jets sont dévalués. La réjection des deux BDT est évaluée sur deux échantillons $t\bar{t}$ statistiquement indépendants. La bande en pointillé représente l'incertitude statistique sur le rapport des réjections.

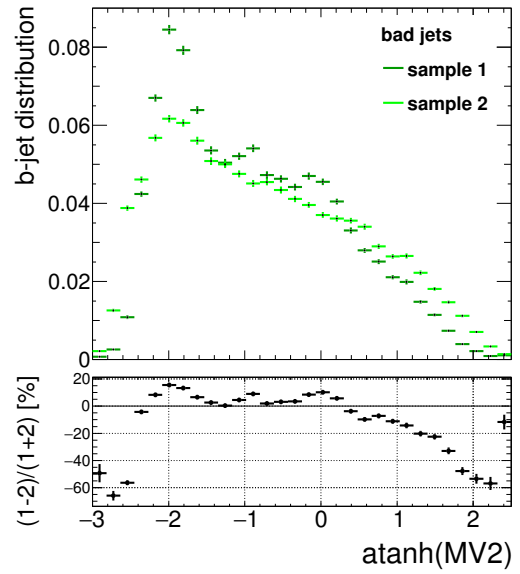
performances obtenues sont significativement différentes : la réjection de jet léger pour les deux BDT est incompatible sur une large plage d'efficacité d'identification de jet de b , en particulier sur la zone intéressante comprise entre 50 et 85% qui comprend les points de fonctionnement (Figure 4.25). A une efficacité d'identification de jet de b de 70%, le désaccord sur les réjections atteint 60%. En revanche, la classification des deux BDT $MV2$ est en relativement bon accord lorsque seuls les bons jets sont considérés. En effet, les poids de sortie des deux entraînements sont distribués de la même façon pour les bons jets (Figures 4.26a et 4.26c), sauf pour des efficacités de jet de b supérieures à $\sim 90\%$. Le désaccord entre les deux entraînements provient principalement des mauvais jets, aussi bien pour les jets de b que pour les jets de bruit de fond (Figures 4.26b et 4.26d). Les jets légers forment 90% du bruit de fond et contribuent à 98% de l'ensemble des jets légers d'un échantillon inclusif $t\bar{t}$. Une différence de performances pour ces mauvais jets se répercute donc sur l'ensemble des jets issus du processus $t\bar{t}$.

La distribution des poids des deux BDT $MV2$ est très différente pour les mauvais jets. Cette différence est due à la dévaluation des mauvais jets lors de l'entraînement, qui a pour conséquence que l'entraînement ne s'effectue de fait que sur les bons jets. En effet, le nombre total de mauvais jets d'entraînement est quasiment nul du fait de leur dévaluation par un poids très faible, bien que leur statistique soit égale à celle des bons jets d'entraînement. Plus le BDT s'entraîne, plus il se focalise sur les jets difficiles à classifier et se concentre sur des régions à statistique de plus en plus faible. Or, il se peut que la statistique de bons jets au niveau d'un noeud soit très faible tandis que celle des mauvais jets soit suffisamment élevée pour permettre une coupure du noeud. Pour ce noeud, la coupure optimale dépend principalement des bons jets, et présente une incertitude liée à la statistique limitée de ces derniers. Cette incertitude n'a cependant pas d'impact significatif sur la classification globale des bons jets, contrairement à celle des mauvais jets. L'accumulation de ce type de noeuds peut aboutir à des variations significatives de classification sur l'ensemble des mauvais jets. En somme, le traitement réservé aux mauvais jets pendant l'entraînement du BDT $MV2$ est responsable de l'instabilité des performances de ce dernier : la dévaluation de ces jets malgré leur importante contribution à l'échantillon d'évaluation, et leur statistique élevée.

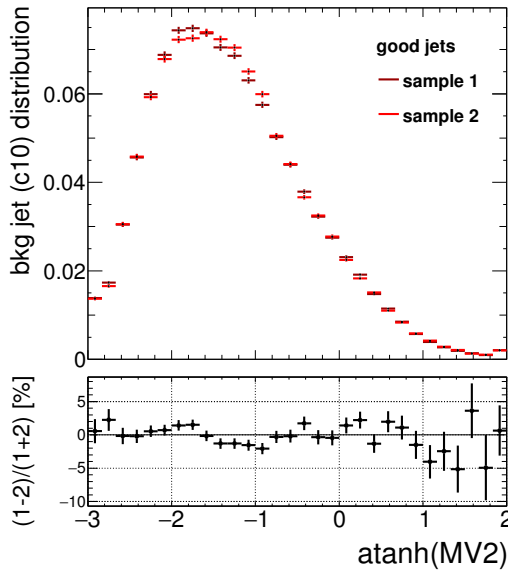
Afin de valider cette hypothèse, un nouveau BDT $MV2$ est considéré, entraîné avec des jets qui passent la sélection d'IPxD. Contrairement à l'entraînement du BDT $MV2$ précédent, le nombre d'événements Monte Carlo des bons jets d'une part et des mauvais jets d'autre part ne sont pas artificiellement égalés, et aucune distinction n'est faite entre les bons et les mauvais jets qui passent la sélection d'IPxD, *i.e* les mauvais jets ne sont pas dévalués. Pour obtenir avec le nouveau BDT $MV2$ des performances similaires à celles d'origine pour les bons jets, la statistique totale des échantillons de jets d'entraînement de b , de c et légers est augmentée afin que le nombres de bons jets de b , c et légers soient approximativement égal à celui d'origine, c'est-à-dire à 10^5 , 10^5 et 5×10^4 respectivement. Comme presque la totalité des jets du processus $t\bar{t}$ passent cette sélection (Table 4.4), l'échantillon d'entraînement est semblable à l'échantillon d'évaluation. Deux BDT $MV2$ sont entraînés



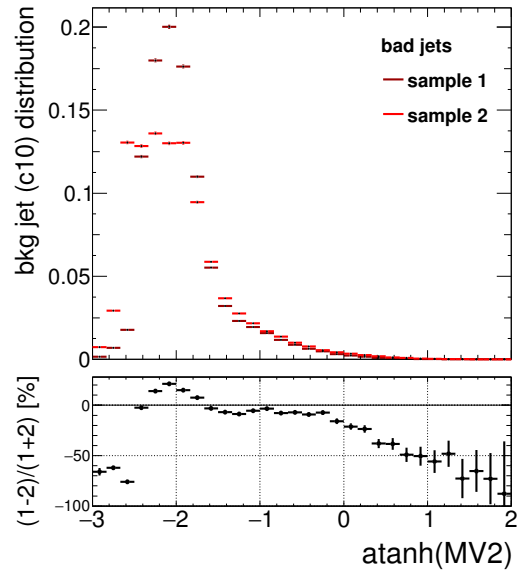
(a) bons jets, signal



(b) mauvais jets, signal



(c) bons jets, bruit de fond



(d) mauvais jets, bruit de fond

FIGURE 4.26: Comparaison des distributions de l'arctangente hyperbolique des poids de sortie de deux BDT $MV2c10$ entraînés à partir de deux échantillons statistiquement indépendants, notées "sample 1" et "sample 2". L'échantillon d'entraînement considéré ici est celui d'origine, pour lequel les mauvais jets sont dévalués. Les distributions sont calculées pour des bons jets de signal (a) et de bruits de fond (c), et des mauvais jets de signal (b) et de bruits de fond (droite).

dans cette nouvelle configuration à partir d'échantillons statistiquement indépendants. Ensuite, leurs performances sont évaluées sur deux échantillons $t\bar{t}$, de même statistiquement indépendants. Leurs différences en terme de réjection de jet légers sont compatibles avec les erreurs statistiques (Figure 4.27). Ce test démontre que le nouvel entraînement sans traitement particulier des mauvais jets ne présente pas les instabilités du BDT d'origine.

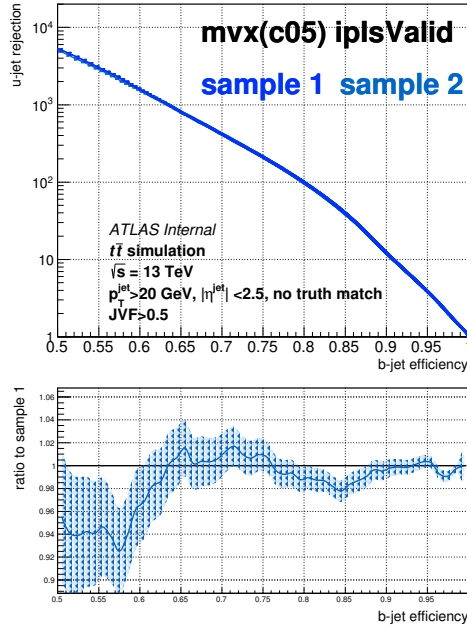


FIGURE 4.27: Réjection de jet léger en fonction de l'efficacité de jet de b évaluée pour deux BDT $MV2c05$ entraînés à partir de deux échantillons statistiquement indépendants, notés "sample 1" et "sample 2" (figure du haut), et leur rapport (figure du bas). Les deux BDT sont entraînés sur des jets qui passent la sélection IPxD. Les performances des deux BDT sont évaluées sur deux échantillons $t\bar{t}$ statistiquement indépendants. La bande en pointillé représente l'incertitude statistique sur le rapport des réjections.

4.5.2 Optimisation du nouvel entraînement du BDT $MV2$

A ce stade, les performances du BDT $MV2c20$ d'origine ne sont pas comparables avec celles du nouveau BDT $MV2c20$ (entraîné sans dévaluation des mauvais jets), en raison d'une différence de composition en jets de c dans l'échantillon de bruit de fond utilisé pour l'entraînement. Le nouveau BDT $MV2c20$ est optimal pour rejeter un bruit de fond composé de 20% de jets de c , alors que le BDT d'origine est optimal pour rejeter un bruit de fond de bons jets contenant 20% de jets de c . Le nouveau BDT est par construction plus performant que le BDT d'origine pour rejeter les jets de c , au détriment des jets légers. En effet, comme les jets de c ont plus de chance d'être de bons jets que les jets légers ($\sim 30\%$ contre $\sim 2\%$), un échantillon $t\bar{t}$ avec seulement 3% de jets de bruit de fond provenant de jets de c contient

un sous-ensemble de bons jets de bruit de fond dont 20% sont des jets de c .

Pour une comparaison équitable des performances du nouveau BDT avec l'entraînement d'origine, les jets de c sont retirés des échantillons d'entraînement. De cette manière, on s'acquitte des différences de composition du bruit de fond d'entraînement en jets de c . Le nouveau BDT $MV2c00$ gagne principalement là où SV1 ne reconstruit pas de vertex secondaire, c'est-à-dire à haute efficacité de jet de b (Figure 4.28). A basse efficacité de jet de b , lorsque les jets de b ont plus de chance d'être de bons jets, le BDT $MV2c00$ d'origine obtient de meilleures performances. Il est donc nécessaire de modifier les réglages internes du nouveau BDT $MV2$ pour que sa classification des bons jets soit aussi bonne que celle du BDT $MV2$ d'origine.

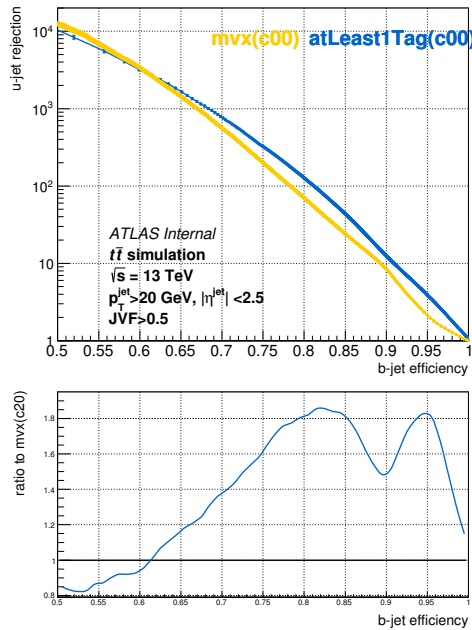


FIGURE 4.28: Réjection de jet léger en fonction de l'efficacité de jet de b évaluée pour le nouveau BDT $MV2c00$ noté 'atLeast1Tag(c00)' et celui d'origine, noté 'mvx(c00)' (figure du haut), et leur rapport (figure du bas). Les réjections des deux BDT sont évaluées sur un même échantillon $t\bar{t}$.

Comme la statistique de l'échantillon d'entraînement du nouveau BDT $MV2$ est supérieure à celle d'origine, les réglages internes du BDT doivent être réoptimisés, notamment avec l'augmentation du nombre d'arbres de décision, de leur profondeur, et avec l'abaissement du MNS. La statistique maximale dans une feuille d'arbre de décision est égale au MNS multiplié par la statistique totale de l'échantillon d'entraînement. Pour optimiser le nouveau BDT, le MNS d'origine peut par exemple être diminué pour compenser la statistique plus élevée du nouvel échantillon d'entraînement. Cette hypothèse est validée avec l'observation d'un gain des performances, atteignant $\sim 15\%$ pour la réjection de jet léger à 60% d'efficacité de jet de b , obtenu grâce à la réduction du MNS, passant de 0.5% (valeur

d'origine) à 0.1% (Figure 4.29).

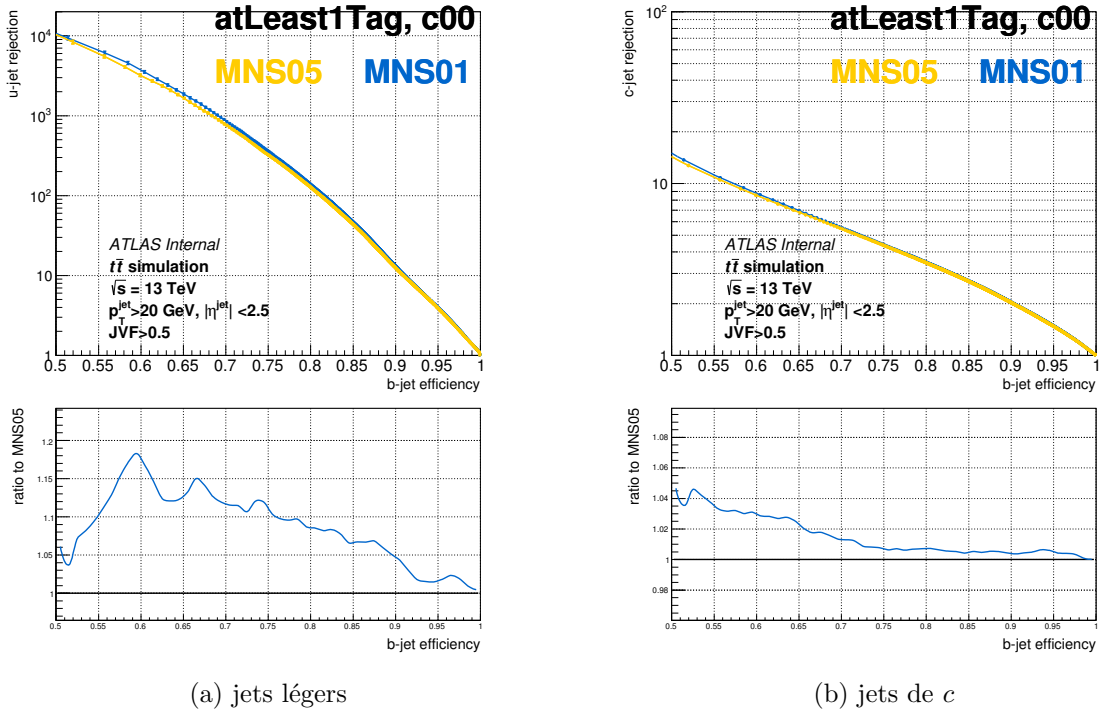


FIGURE 4.29: Réjections de jet léger (a) et jet de c (b) en fonction de l'efficacité de jet de b évaluée pour deux BDT $MV2c00$ entraînés avec des MNS de 0.5% et 0.1% notés 'MNS05' et 'MNS01', (figure du haut), et leur rapport (figure du bas). Les deux BDT sont entraînés à partir d'un même échantillon de jets pour lesquels au moins un des trois algorithmes IPxD, SV1 et JetFitter est défini. Les réjections des deux BDT sont évaluées sur un même échantillon $t\bar{t}$.

4.5.3 Conclusion

À la fin de l'année 2015, le BDT $MV2$ a été réentraîné pour prendre en compte des améliorations dans les algorithmes d'identification des jets de b et les nouvelles conditions de la prise de données de 2016, en particulier le changement dans le nombre de collisions d'empilement [51]. Ce réentraînement a bénéficié de l'optimisation des traces et vertex secondaires sélectionnés par les algorithmes IPxD et SV1. Le changement qui a principalement contribué à l'amélioration des performances en 2016 a été la prise en compte du nouvel échantillon d'entraînement du BDT $MV2$ étudié dans cette partie, c'est-à-dire sans évaluer artificiellement la statistique des bons jets avec celles des mauvais jets ni dévaluer les mauvais jets. L'optimisation débutée dans la section précédente a été finalisée au sein de la collaboration, dont les réglages sont décrits ci-après.

L'optimisation des réglages internes du nouveau BDT $MV2$ sont les suivants : le MNS est réduit de 0.5% à 0.05%, le nombre d'arbres de décision est augmenté de 400 à 1000 et leur

profondeur augmentée de 10 à 30. Le BDT est entraîné avec un total de 2.5 millions de jets issus d'événements $t\bar{t}$, comprenant 1 million de jets de b et autant de jets légers, ainsi que 0.5 million de jets de c . Les distributions en (p_t, η) des jets de b et de c sont pondérées pour égaler celle des jets légers. Trois entraînements sont effectués, appelés $MV2c00$, $MV2c10$ et $MV2c20$, avec trois fractions de jets de c pour le bruit de fond de 0%, 7% et 15% respectivement. Le nouveau BDT $MV2c10$ a été choisi pour remplacer le $MV2c20$ en 2016. Les deux configurations obtiennent des réjections de jet léger similaires pour des efficacités de jet de b identiques, avec un léger avantage pour le nouveau BDT $MV2c10$ à haute efficacité de jet de b (approximativement 4% meilleur à 77% d'efficacité de jet de b) (Figure 4.30a). Le nouveau BDT est plus performant que l'ancien à bas (< 50 GeV) et haut (> 200 GeV) p_T de jet (Figure 4.31a). Sa réponse est donc plus uniforme avec l'impulsion transverse du jet, bien que les performances soient toujours meilleures autour de 100 GeV. De plus, le gain est d'autant plus grand que la pseudo-rapidité du jet augmente (Figure 4.31b). Le nouveau BDT $MV2c10$ se distingue également par sa capacité à rejeter plus de jets de c que le BDT $MV2c20$ d'origine, avec un gain de 40% à 77% d'efficacité de jet de b (Figure 4.30b). Le gain apporté à la réjection de jet de c concerne en particulier les impulsions vers 120 GeV et les pseudo-rapidités autour de 1 (Figures 4.31c et 4.31d) [51].

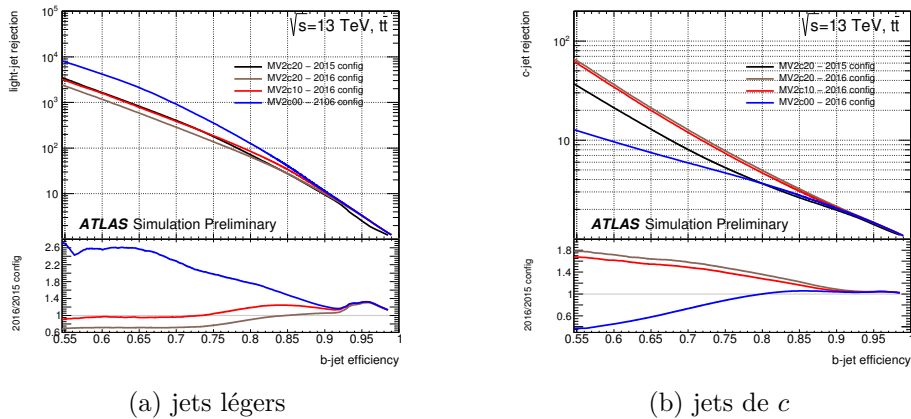
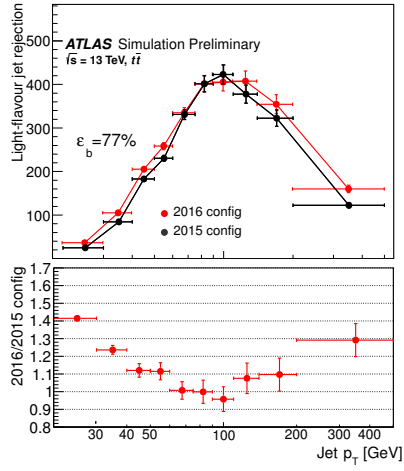
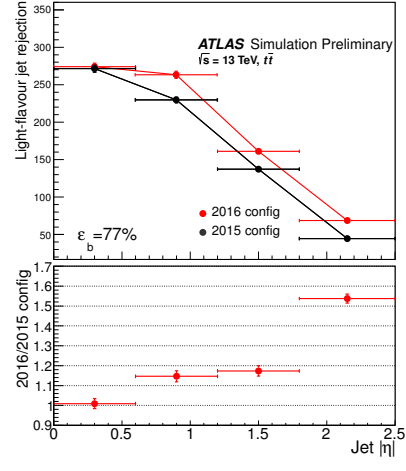


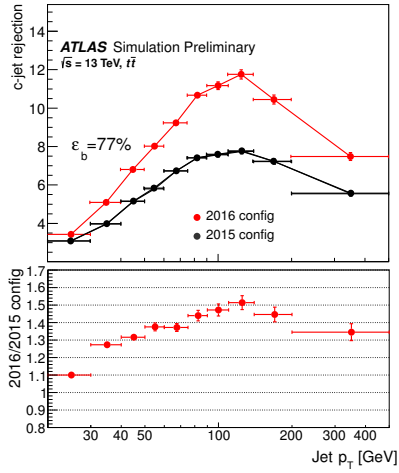
FIGURE 4.30: Réjections de jet léger (gauche) et de jet de c (droite) en fonction de l'efficacité de jet de b . Les quatre courbes correspondent à l'évaluation de quatre entraînements indépendants. Le BDT $MV2$ d'origine, ou configuration 2015, (courbe noire) est comparé aux nouveaux BDT $MV2$, ou configuration 2016, $MV2c00$ (courbe bleue), $MV2c10$ (courbe rouge) et $MV2c20$ (courbe marron) entraînés avec 0, 7 et 15 % de jets de c dans le bruit de fond. Les jets d'évaluation sont issus d'un échantillon $t\bar{t}$ [51].



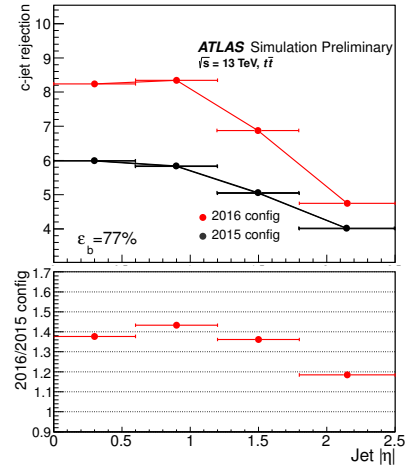
(a) jets légers



(b) jets légers



(c) jets de c



(d) jets de c

FIGURE 4.31: Réjections de jet léger et de jet de c en fonction du p_T du jet (a et c) et de sa pseudo-rapidité (b et d), pour le nouveau BDT $MV2c10$ (2016 config) et le BDT $MV2c20$ d'origine (2015 config). L'efficacité de jet de b est choisie égale à 77% pour ces distributions. Les jets d'entraînement et d'évaluation sont issus d'un échantillon $t\bar{t}$ [51].

Bibliographie

- [48] Georges Aad et al. Performance of b -Jet Identification in the ATLAS Experiment. *JINST*, 11(04) :P04008, 2016.
- [49] Performance of Impact Parameter-Based b -tagging Algorithms with the ATLAS Detector using Proton-Proton Collisions at $\sqrt{s} = 7$ TeV. Technical Report ATLAS-CONF-2010-091, CERN, Geneva, Oct 2010.
- [50] Impact parameter-based b -tagging algorithms in the 7 TeV collision data with the ATLAS detector : the TrackCounting and JetProb algorithms. Technical Report ATLAS-CONF-2010-041, CERN, Geneva, Jul 2010.
- [51] Optimisation of the ATLAS b -tagging performance for the 2016 LHC Run. Technical Report ATL-PHYS-PUB-2016-012, CERN, Geneva, Jun 2016.
- [52] Giacinto Piacquadio and Karl Jakobs. Identification of b -jets and investigation of the discovery potential of a Higgs boson in the $WH \rightarrow l\nu b\bar{b}$ channel with the ATLAS experiment, 2010. Presented on 11 Jan 2010.
- [53] b -tagging in dense environments. Technical Report ATL-PHYS-PUB-2014-014, CERN, Geneva, Aug 2014.
- [54] V Kostyukhin. VKalVrt - package for vertex reconstruction in ATLAS. Technical Report ATL-PHYS-2003-031, CERN, Geneva, Aug 2003. revised version number 1 submitted on 2003-09-24 11 :10 :53.
- [55] Secondary vertex finding for jet flavour identification with the ATLAS detector. Technical Report ATL-PHYS-PUB-2017-011, CERN, Geneva, Jun 2017.
- [56] P Speckmayer, A Hocker, J Stelzer, and H Voss. The toolkit for multivariate data analysis : TMVA 4. *J. Phys. : Conf. Ser.*, 219 :032057, 2010.
- [57] Calibration of b -tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment. Technical Report ATLAS-CONF-2014-004, CERN, Geneva, Feb 2014.
- [58] Calibration of the performance of b -tagging for c and light-flavour jets in the 2012 ATLAS data. Technical Report ATLAS-CONF-2014-046, CERN, Geneva, Jul 2014.
- [59] Georges Aad et al. A search for $t\bar{t}$ resonances using lepton-plus-jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 08 :148, 2015.

Chapitre 5

Recherche du processus $t\bar{t}H$ de haute impulsion dans le canal $b\bar{b}$

5.1 Motivation du canal $t\bar{t}H(b\bar{b})$ boosté

L'analyse présentée dans cette thèse se focalise sur la recherche d'événements $t\bar{t}H(b\bar{b})$ produits avec des quarks *top* et bosons de Higgs de haute impulsion transverse, qu'on appelle le canal boosté. Dans cette topologie, les jets produits lors de la désintégration du quark *top* ou du boson de Higgs peuvent être très proches. Elle est développée dans le but d'être combinée avec l'analyse résolue $t\bar{t}H(b\bar{b})$ dans le canal *lepton + jets*, c'est-à-dire l'analyse où les jets provenant de la désintégration du quark *top* ou du boson de Higgs sont suffisamment séparés pour être reconstruits indépendamment. D'un point de vue expérimental, l'analyse boostée utilise des grands jets de rayon $R = 1.0$, en plus des jets de rayon $R = 0.4$ sélectionnés par l'analyse résolue. Leur taille importante rend possible l'étude des caractéristiques de leurs dépôts d'énergie internes, appelées sous-structures. Pour des topologies boostées, l'analyse de sous-structures permet de révéler des caractéristiques qui ne pourraient être résolues avec des jets standards, notamment lorsque les produits de désintégration d'un quark *top* hadronique, ou d'un boson de Higgs, commencent à se chevaucher dans un même jet de rayon $R = 0.4$. L'analyse résolue est confrontée à la difficulté d'identifier les jets issus de la désintégrations des événements $t\bar{t}H$ à cause de la grande multiplicité de jets au sein de ces événements. En revanche, l'analyse boostée, qui identifie des grands jets provenant de quarks *top* hadroniques et/ou bosons de Higgs, est limitée par la faible fraction des événements présentant un quark *top* hadronique et/ou boson de Higgs boostés (Figure 5.1 et Table 5.1).

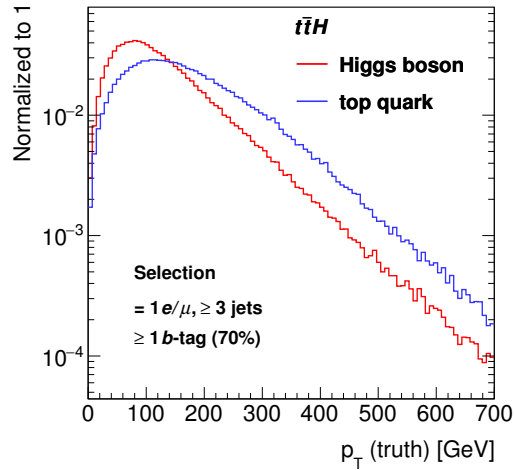


FIGURE 5.1: Distributions normalisées de l'impulsion transverse du boson de Higgs et du quark *top* hadronique pour des événements $t\bar{t}H$ simulés qui passent la sélection demandant un candidat électron/muon, ≥ 3 jets et ≥ 1 candidat *b* (point de fonctionnement 70%) dans l'événement.

D'un point de vue théorique, le canal boosté peut non seulement améliorer la recherche d'un signal du modèle standard, mais aussi être sensible à des modèles de nouvelle physique

truth p_T cut (GeV)	efficiency
Higgs $p_T \geq 200$	21%
t -quark $p_T \geq 250$	25%
both	8%

TABLE 5.1: Fraction des événements $t\bar{t}H$ qui passent la sélection demandant un candidat électron/muon, ≥ 3 jets et ≥ 1 candidat b (point de fonctionnement 70%) avec un boson de Higgs et d'un quark top hadronique d'impulsion transverse supérieure à 200 GeV et 250 GeV.

(en particulier dans le cadre de modèles effectifs) qui peuvent induire des topologies $t\bar{t}H$ plus boostées.

5.2 Modélisation du signal et des bruits de fond

L'ensemble des échantillons Monte Carlo utilisés dans l'analyse sont décrits dans cette partie. Les échantillons de bruit de fond sont générés avec une masse du quark top de 172.5 GeV. La désintégration des hadrons c et b est simulée avec EvtGen v1.2.0 [60].

5.2.1 Le signal $t\bar{t}H$

Le signal $t\bar{t}H$ est simulé à l'aide de MadGraph5_aMC@NLO [61] (mentionné dans la suite simplement comme aMC@NLO) version 2.3.2 pour le calcul de l'élément de matrice, interfacé avec Pythia 8.210 [62] pour l'hadronisation utilisé avec la configuration A14 [63] pour les paramètres réglables de la modélisation des événements sous-jacents. Les fonctions de distribution partonique (PDF) NNPDF3.0NLO [64] sont utilisées, et les échelles de factorisation et de renormalisation sont fixées à $\mu_F = \mu_R = H_T/2$, où H_T est défini comme la somme scalaire des masses transverses $\sqrt{p_T^2 + m^2}$ de toutes les particules de l'état final. Les quarks top sont désintégrés en utilisant MadSpin [65], qui préserve les corrélations de spin. La masse du boson de Higgs est fixée à 125 GeV et tous ses modes de désintégration sont générés. Les rapports de branchement sont calculés avec HDECAY [66]. La section efficace $t\bar{t}H$ du Monte Carlo est normalisée à la section efficace extraite des calculs à l'ordre NLO [67–70] en $\alpha_s : 0.5608\text{pb}^{+5.8\%}_{-9.3\%}$ (échelles QCD) $\pm 8.9\%$ (PDF et α_s)

5.2.2 Le bruit de fond $t\bar{t}$

L'échantillon utilisé par défaut pour modéliser le bruit de fond $t\bar{t}$ est simulé à l'aide du générateur Powheg-Box v2 [71] avec la PDF CT10. Le paramètre $hdamp$, qui contrôle l'impulsion transverse de la première émission additionnelle au-delà de la configuration de Born, est fixé à la masse du quark top . Les échelles de factorisation et de renormalisation sont fixées à $\sqrt{p_{T,t}^2 + m_t^2}$, avec m_t et $p_{T,t}$ étant la masse et l'impulsion transverse du quark

top. Les radiations partoniques et l'hadronisation sont modélisées avec Pythia 6.428 [72] et le jeu de PDF CTEQ6L1 [73] et utilisant le réglage Perugia2012 [74]. L'échantillon est normalisé à la section efficace théorique de 832_{-51}^{+46} pb [75], calculée à l'ordre NNLO en QCD, et comprend la resommation des termes logarithmiques dominants pour les termes de gluon doux au second ordre (NNLL) [76]. Les échantillons $t\bar{t}$ alternatifs utilisés pour dériver des incertitudes systématiques seront décrits dans la section 5.6.

Le bruit de fond $t\bar{t}$ est classé selon la saveur des *vrais* jets supplémentaires – ne provenant pas des désintégrations de $t\bar{t}$ – présents dans l'événement. Les *vrais* jets sont reconstruits à partir des particules générées (à l'exception des muons et des neutrinos) avec le même algorithme que celui utilisé pour reconstruire les jets (anti- k_t de rayon $R = 0.4$). Une impulsion transverse supérieure à 15 GeV et une pseudo-rapacité $|\eta|$ inférieure à 2.5 sont requises. La saveur de ces jets est déterminée en leur associant les hadrons b ou c produits dans l'événement avec un $\Delta R(\text{vrai jet, hadron})$ inférieur à 0.4. Les *vrais* jets associés avec un hadron b , d'impulsion transverse supérieure à 5 GeV, sont étiquetés jets de b . L'association d'un *vrai* jet (non jet de b) et d'un hadron c permet de définir un jet de c . Les événements qui possèdent au moins un jet de b (en plus de ceux provenant de la désintégration des quarks *top*) sont étiquetés $t\bar{t} + b$, ceux sans jet de b supplémentaire mais avec au moins un jet de c (à l'exclusion des jets de c provenant de la désintégration de bosons W venant des quarks *top*) sont étiquetés $t\bar{t} + c$. Les événements $t\bar{t}$ restants sont classés comme $t\bar{t} + \text{léger}$.

5.2.3 Les autres bruits de fond

Les échantillons d'événements Wt et *top unique* en canal-s sont générés avec Powheg-Box 2.0 en utilisant le jeu de PDF CT10 [77, 78]. Le recouvrement entre les états finals $t\bar{t}$ et Wt est géré avec le procédé dit de "soustraction de diagrammes" [79]. Les événements *top unique* en canal-t sont générés à l'aide du générateur Powheg Box v1 avec le jeu de PDF à quatre saveurs CT10f4. Tous les échantillons de quarks *top unique* sont interfacés avec Pythia 6.428 avec le réglage Perugia2012 pour les événements sous-jacents. Les sections efficaces des échantillons de quark *top unique* en canal-t et -s sont normalisées à de la section efficace théorique approximée NNLO [80–82].

Les échantillons d'événements $W/Z + jets$ et diboson en association avec des jets sont générés à l'aide de Sherpa 2.1. Pour les échantillons $W/Z + jets$, les éléments de matrice sont calculés jusqu'à deux partons à NLO et jusqu'à quatre partons à l'ordre dominant (LO) grâce aux générateurs d'éléments de matrice Comix [83] et OpenLoops. Ces derniers sont fusionnés avec le modèle d'hadronisation de Sherpa [84], avec la prescription ME+PS@NLO [85]. Le jeu de PDF CT10 est utilisé. Les événements $W/Z + jets$ sont normalisés à la section efficace NNLO [86]. Les échantillons diboson + jets sont générés suivant la même approche mais jusqu'à un parton supplémentaire au NLO et jusqu'à trois partons supplémentaires au LO. Ils sont normalisés à leur section efficace NLO respective. Les échantillons d'événements $t\bar{t} + V$ sont générés avec un élément de matrice NLO en utilisant MG5_aMC interfacé avec Pythia 8.210 et la PDF NNPDF3.0NLO. Enfin, le réglage A14 est utilisé pour la

modélisation des événements sous-jacents.

5.3 Sélection d'objets d'analyse et d'événements

Les événements analysés sont choisis parmi les collisions à 13 TeV enregistrées par le détecteur ATLAS en 2015 et 2016, correspondant à une luminosité de 36.5/fb. Seuls les événements enregistrés lorsque tous les sous-systèmes du détecteur utiles pour la reconstruction des objets étaient opérationnels sont pris en compte. Au moins un vertex primaire avec deux ou plusieurs traces d'impulsion transverse au-dessus de 0.4 GeV est demandé dans l'événement. L'empilement moyen au cours des prises de donnée est typiquement compris entre 10 et 30. Les événements simulés sont pondérés de sorte que la distribution de l'empilement moyen corresponde à celle des données observées. Le vertex avec la plus grande somme des carrés des impulsions transverses des traces associées est considéré comme le vertex primaire.

La sélection des objets pour l'analyse *lepton + jets* est décrite ci-dessous. La sélection des événements au niveau du système de déclenchement [87] est détaillée dans la note décrivant l'analyse *lepton + jet* de 2016 [88].

5.3.1 Les candidats électron

Les candidats électron doivent passer le critère d'identification de vraisemblance *Tight* (cf. 3.3) ainsi que les sélections sur les paramètres d'impact transversal et longitudinal : $|z_0 \sin(\theta)| < 0.5$ mm et $|d_0/\sigma(d_0)| < 5$. L'impulsion transverse minimale des candidats électron est fixée à 27 GeV, et leur pseudo-rapacité doit être comprise dans l'intervalle $|\eta| < 2.5$. Pour réduire davantage le bruit de fond venant d'électrons non-prompts, de conversions et de hadrons, les candidats électron doivent également être isolés avec le critère d'isolation "Gradient". Cette isolation, basée sur les traces et les dépôts calorimétriques, dépend de l'impulsion transverse du candidat électron.

5.3.2 Les candidats muon

Les candidats muon de l'analyse doivent satisfaire le critère d'identification de qualité moyenne (cf. 3.4). La valeur absolue de la signification du d_0 du candidat muon doit être inférieure à 3, et la valeur de $|z_0 \sin(\theta)|$ doit être inférieure à 0.5 mm. De plus, leur impulsion transverse minimale doit être supérieure à 27 GeV, et leur pseudo-rapacité inférieure à $|\eta| = 2.5$. Les candidats muon doivent également être isolés en satisfaisant le point de fonctionnement d'isolation "Gradient".

5.3.3 Les jets

Les jets sont reconstruits à partir de dépôts calorimétriques avec l'algorithme anti- k_t , de rayon $R = 0.4$ (cf. 3.5). Après leur calibration en énergie, les jets doivent satisfaire $p_T > 25$ GeV et $|\eta| < 2.5$. Des critères de qualité sont imposés pour identifier les jets ne

provenant pas de collisions pp ou provenant de bruit dans le détecteur. Tout événement contenant un tel jet est retiré. Pour réduire la contribution des jets provenant des collisions d'empilement, un critère supplémentaire est exigé pour les jets d'impulsion transverse $p_T < 60$ GeV et $|\eta| < 2.4$. Il utilise le Jet Vertex Tagger (JVT) qui est basé sur l'information des traces associées au jet, et permet d'identifier les jets provenant d'un vertex créé par les collisions d'empilement.

Pour éviter de considérer une même réponse du calorimètre comme appartenant à la fois à un jet et à un candidat électron, le jet le plus proche d'un candidat électron dans le plan (η, ϕ) est retiré si $\Delta R_{e,jet} < 0.2$. Si le jet le plus proche survit à la sélection décrite ci-dessus mais se situe à $R_{e,jet} < 0.4$ du candidat électron, le candidat électron est supprimé afin de réduire la contamination d'électrons issus des désintégrations de hadrons à saveur lourdes. De même, les candidats muon sont retirés s'ils sont séparés du jet le plus proche à moins de $R = 0.4$. Toutefois, si ce jet est associé à moins de trois traces, le candidat muon est conservé et le jet est enlevé à la place. Cela évite une inefficacité pour les muons de haute impulsion transverse subissant une perte d'énergie significative dans le calorimètre.

5.3.4 Les candidats b

Les jets sont étiquetés candidats b avec l'algorithme MV2c10. Plusieurs points de fonctionnement sont considérés par la suite, correspondants à des efficacités d'identification de jet de b de 70%, 77% et 85%, obtenus avec différentes coupures sur la variable discriminante MV2c10. Seuls cinq intervalles en MV2c10 sont calibrés, correspondant à des efficacités de jet de b de [85%, 100%], [77%, 85%], [70%, 77%], [60%, 70%] et [0%, 60%]. Dans cette étude, chacun de ces intervalles en MV2c10 est associé à un entier, soit 1, 2, 3, 4 et 5 respectivement.

5.3.5 Les grands jets

Des grands jets de rayon $R = 1.0$ reconstruits avec l'algorithme anti- k_t décrits Section 3.6 sont utilisés dans l'analyse boostée. Ils possèdent la coupure en pseudo-rapacité $|\eta| < 2.0$. La sélection sur les grands jets, notamment sur l'impulsion transverse, sera étudiée dans la section suivante.

5.3.6 Candidat top *lâche*

L'étiqueteur top *lâche* permet d'identifier les grands jets d'impulsion transverse supérieure à 250 GeV provenant de désintégrations hadroniques de quarks top [89]. Il est utilisé par la sélection boostée de base qui sera détaillée par la suite. Il est fondé sur la masse calibrée du grand jet et son τ_{32} (cf Section 3.6). Une coupure sur ce couple de variables est effectuée en fonction de l'impulsion transverse du grand jet. Cette coupure maximise la réjection de grands jets provenant du bruit de fond *multijet* tout en maintenant une efficacité d'identifier un quark top hadronique issus du processus $Z' \rightarrow t\bar{t}$ de 80%. Il existe également un critère d'identification similaire appelée top *dur* pour lequel l'efficacité d'étiqueter un

quark *top* hadronique est de 50%. Les grands jets étiquetés avec les critères *top lâche* ou *top dur* sont appelés candidats *top lâche/dur* ci-après.

5.4 Etude du canal *top* boosté

Les événements sélectionnés par l'analyse $t\bar{t}H$ boostée sont caractérisés par la présence de grands jets potentiellement issus de désintégrations de quarks *top* et/ou de bosons de Higgs. L'étude présentée dans ce chapitre comprend les travaux réalisés au cours de cette thèse dans le but d'améliorer la sélection de base de l'analyse boostée appelé *3211* avec, dans une première partie, une étude sur l'identification de quark *top* hadronique et dans une seconde partie, l'étude d'un nouveau canal $t\bar{t}H(b\bar{b})$ dans lequel quarks *top* hadroniques et bosons de Higgs sont tous deux boostés.

5.4.1 Sélection *3211* et *3111*

Pour sélectionner les événements boostés $t\bar{t}H$, la sélection dite *3211* requiert :

- = 1 candidat électron ou muon ;
- ≥ 1 candidat *top lâche* d'impulsion transverse $p_T > 250$ GeV ;
- ≥ 3 jets à l'extérieur du/des candidats *top lâche*, *i.e* ΔR (jet, candidat *top lâche*) > 1.0 ;
- ≥ 2 candidats *b* (point de fonctionnement 77%) à l'extérieur du/des candidats *top lâche*.

La stratégie de cette sélection d'événements est d'identifier en premier lieu un quark *top* hadronique boosté reconstruit dans un grand jet (candidat *top*), puis de trouver les jets de *b* provenant de la désintégration du boson de Higgs parmi les jets en dehors du candidat *top*. Une préselection demandant seulement ≥ 1 candidat *b* additionnel au lieu de 2, appelée *3111*, moins sensible que la sélection *3211* sera étudiée par la suite avec d'autres définitions de candidat *top* que celle de base, *i.e* *top lâche* et $p_T > 250$ GeV.

5.4.2 Suppression du recouvrement candidat électron/grand jet

Les grands jets sont reconstruits à partir des dépôts calorimétriques, incluant ceux des candidats électron. Par conséquent, l'énergie déposée par un électron dans le calorimètre peut être considérée comme une sous-structure à l'intérieur du grand jet dans lequel il est reconstruit. Il est alors plus probable d'identifier plusieurs sous-structures au sein du grand jet en question, et de le considérer comme un candidat *top* bien qu'il ne provienne pas de la désintégration du quark *top* hadronique (faux candidat *top*).

Les événements présélectionnés (passant *3111*) proviennent principalement du processus $t\bar{t}$. Les deux quarks *top* du processus sont majoritairement émis *dos à dos*, on s'attend donc à ce que leurs produits de désintégration le soient également. La distribution de l'angle $\Delta R(\text{grand jet, lepton})$ des événements présélectionnés présente en effet un maximum autour de π , visible Figure 5.2 aussi bien pour le canal électron que pour le canal muon. Ces

événements à $\Delta R(\text{grand jet, lepton}) \sim \pi$ sont susceptibles de contenir un vrai candidat *top* (grand jet contenant les produits de désintégration du quark *top* hadronique) et un vrai candidat lepton (lepton provenant de la désintégration du quark *top* leptonique).

Cependant dans le canal électron, une grande fraction d'événements sélectionnés ont un $\Delta R(\text{grand jet, lepton}) < 1.0$ (Figure 5.2b), contrairement au canal muon, car l'énergie déposée par le muon dans le calorimètre est trop faible pour influencer l'identification de *top*.

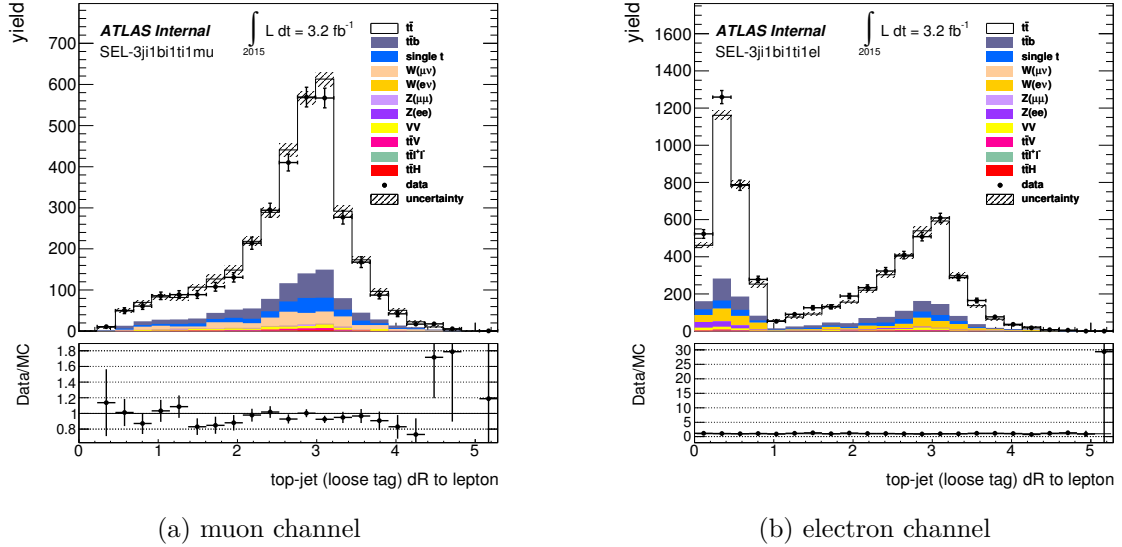


FIGURE 5.2: Distributions de l'angle ΔR entre un grand jet étiqueté 'top loose' et le lepton dans le canal muon (a) et le canal électron (b) pour les événements qui passent la sélection 3111. Les échantillons MC (histogrammes colorés) sont normalisés au données 2015 (points noirs).

Lorsque l'électron et le candidat *top* se recouvrent, presque 80% des candidats *top* sont émis à un angle ΔR du quark *top* leptonique inférieur à 0.4 pour le signal comme pour le bruit de fond $t\bar{t}$, comme indiqué Figure 5.3. Dans ce cas le candidat *top* contient le jet de b du quark *top* leptonique, et donc provient essentiellement de la désintégration du quark *top* leptonique. Par conséquent, le quark *top* hadronique de l'événement qui passe la sélection 3111 dans lequel l'électron et le candidat *top* se recouvrent n'est pas nécessairement boosté et n'est pas identifié comme candidat *top*. De plus, pour le signal, il est alors plus difficile de retrouver les jets de b du boson de Higgs, à cause des jets issus de la désintégration du quark *top* hadronique. Ce problème pourrait être résolu en retirant le dépôt d'énergie de l'électron reconstruit dans l'événement avant de procéder à la reconstruction des grands jets. Cependant, cette méthode est complexe car elle dépend de la définition de l'identification du candidat électron. Par souci de simplicité, il a été décidé de rejeter les grands jets contenant un candidat électron.

Avec la suppression des grands jets contenant un candidat électron, le nombre d'événe-

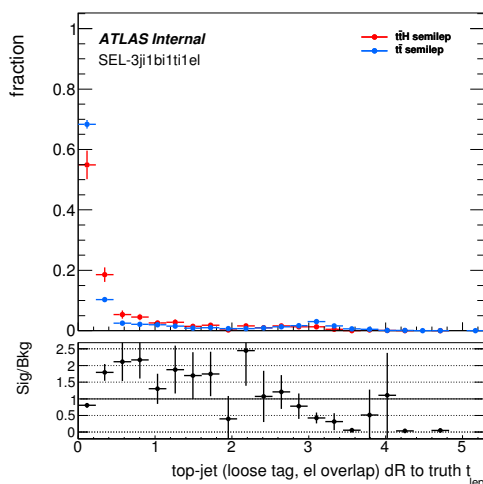


FIGURE 5.3: Distributions de l'angle ΔR entre un grand jet étiqueté 'top loose' qui contient un candidat électron et le quark top leptonique au niveau générateur, pour des événements qui passent la sélection 3111 . Le signal $t\bar{t}H$ semi-leptonique (en rouge) et le bruit de fond $t\bar{t}$ semi-leptonique (en bleu) sont normalisés à 1.

ments de signal et de bruit de fond diminue dans le canal électron (Figure 5.4). Le bruit de fond est plus affecté car il y a moins de chance de reconstruire un candidat top dans un événement de bruit de fond que dans un événement de signal. Avec la suppression de ce recouvrement le canal électron voit son rapport signal/bruit de fond augmenter de $\sim 20\%$ pour les événements qui passent la pré-sélection 3111 .

5.4.3 Etiquetage du quark top boosté

Le candidat top boosté de la sélection de base 3211 est un grand jet d'impulsion transverse supérieure à 250 GeV, identifié avec l'étiqueteur top lâche. Parce que l'étiqueteur est entraîné pour rejeter les grands jets produits dans les événements $multijet$, il n'est pas optimal pour rejeter les grands jets issus des événements $t\bar{t}$ et $t\bar{t}H$ qui ne proviennent pas de désintégrations hadroniques de quark top , comme par exemple les grands jets issus des produits de désintégration du boson W ou du boson de Higgs. Les critères d'étiquetage top lâche sont peu contraignants sur la masse du grand jet (cette dernière ne doit pas dépasser les 80 GeV pour des grands jets d'impulsion transverse inférieure à 800 GeV) afin de maintenir l'efficacité d'identifier un quark top hadronique à 80%. Ce critère ne suffit pas pour garantir que les quarks de la désintégration du top associé $t \rightarrow bq\bar{q}'$ soient contenu dans le candidat top lâche, *i.e* l'angle ΔR entre un b , q ou \bar{q}' et le vrai candidat peut être supérieur au rayon du grand jet $R = 1.0$ et ne pas faire partie du grand jet étiqueté. Une étude détaillée de la composition du candidat top a été réalisée. Les grands jets sont classés selon leur composition en terme de partons issus des désintégrations de quarks top ou de bosons de Higgs, ou *composition partonique*, définies Table 5.2. Les Figures 5.5b et 5.5a

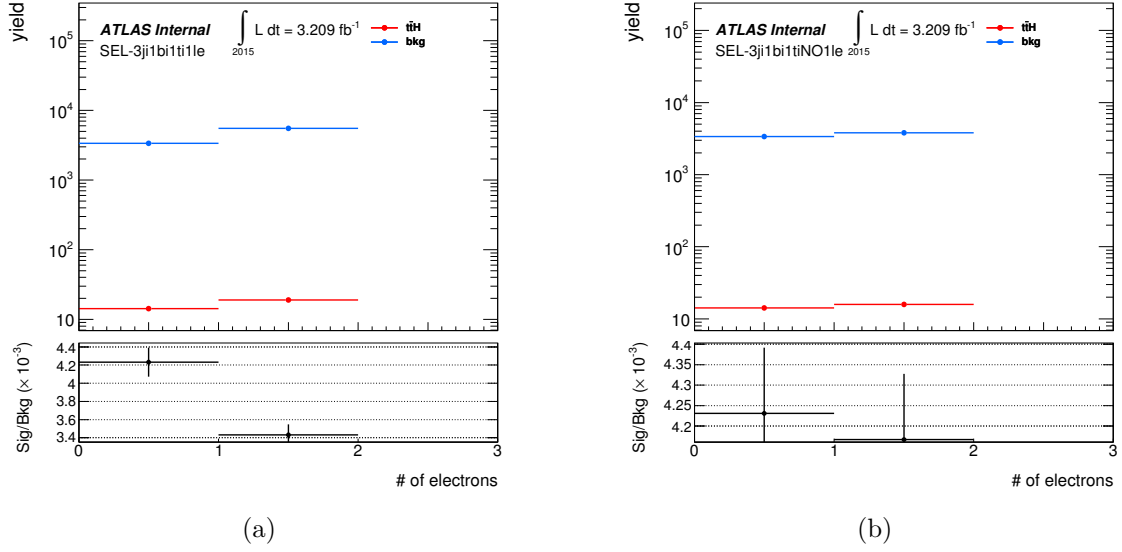


FIGURE 5.4: Distributions attendues en 2015, à $\mathcal{L} = 3.2/\text{fb}$, du nombre de candidats électron pour les événements passant la sélection 3111 sans (a) et avec (b) suppression du recouvrement grand jet/candidat electron, pour le signal $t\bar{t}H$ semi-leptonique (rouge) et le bruit de fond $t\bar{t}$ semi-leptonique (bleu).

Notation	Partons contenus dans le grand jet
$(b\bar{b})_H$	paire $b\bar{b}$ de la désintégration du boson de Higgs + X
b_{H^i}	un unique quark b de la désintégration du boson de Higgs + X
$(bq\bar{q}')_{thad}$	trois partons issus de la désintégration quark top hadronique
b_{thad^i}	quark b issu de la désintégration quark top hadronique + X
$(q\bar{q}')_W$	paire $q\bar{q}'$ de la désintégration du boson W
b_{tlep}	quark b de la désintégration du quark top leptonique
q/g_{rad}	ne contient aucun parton issu de la désintégration de la paire $t\bar{t}$
othermix	combinaison de partons n'appartenant pas aux configurations précédentes

TABLE 5.2: Définition de l'origine possible des grands jets en terme de partons de la désintégration de quark top et de boson de Higgs. Ces huit configurations sont exclusives entre elles.

montrent la composition des grands jets dans des événements $t\bar{t}$ et $t\bar{t}H$ passant la sélection 3111 avant et après application de l'étiquetage *top lâche*. Environ 20% des candidats *top* boostés contiennent les jets de la désintégration hadronique du boson W (configuration $(qq')_W$) et $\sim 26\%$ contiennent au moins le jet de b du quark *top* hadronique (configuration $b_{thad}i$) sans contenir tous les produits de la désintégration de ce dernier (configuration qui serait notée dans ce cas $(bqq')_{thad}$) (Figure 5.5b). Il est difficile d'identifier seulement les grands jets qui contiennent exclusivement tous les produits de la désintégration du quark *top* hadronique dans le cas où les grands jets sont peu boostés ($p_T^J > 250$ GeV). Cependant il est important pour l'analyse d'événements $t\bar{t}H(b\bar{b})$ boostés de capturer au moins le jet de b issu de la désintégration du quark *top* hadronique. Si ce jet de b est à l'extérieur du candidat *top* boosté, comme c'est le cas dans la configuration $(qq')_W$, il peut être considéré comme un candidat b originaire du boson de Higgs, détériorant ainsi la sensibilité de l'analyse. Environ 12% des candidats *top* boostés contiennent les deux jets de b de la désintégration du boson de Higgs et environ 16% en possèdent un des deux (Figure 5.5a). Dans ces deux configurations, l'identification du boson de Higgs en paire de candidat b à l'extérieur du candidat *top* est impossible ou sinon partielle. Ceci entraîne une perte de discrimination de la masse du candidat Higgs et par conséquent de la séparation du signal vis à vis du bruit de fond.

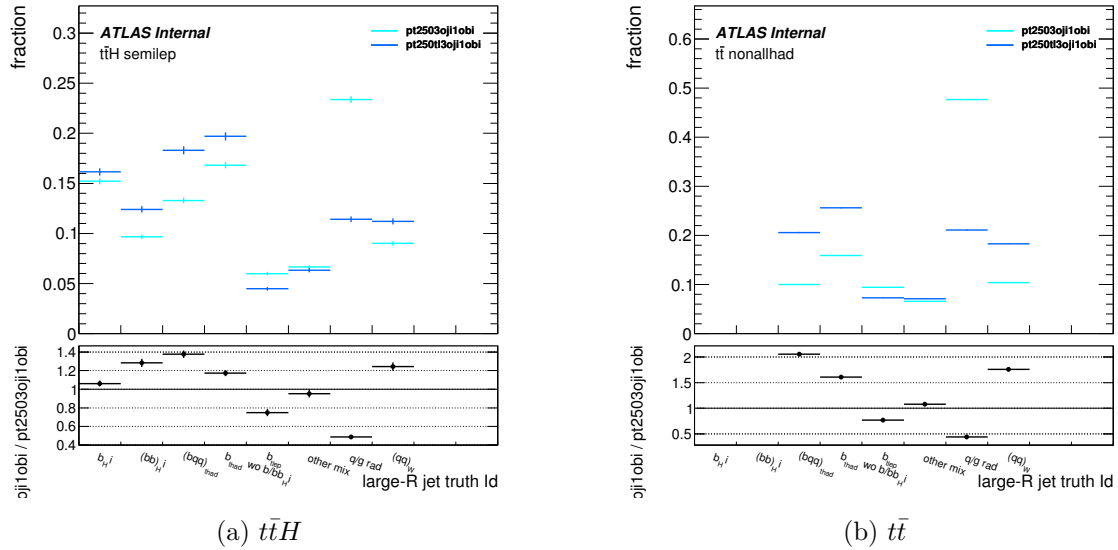


FIGURE 5.5: Composition partonique d'un **grand jet sans critère d'identification** (bleu clair) et d'un **grand jet étiqueté *top lâche*** (bleu foncé) pour des événements qui passent la sélection 3111, pour le signal $t\bar{t}H(b\bar{b})$ semi-leptonique (a) et le bruit de fond $t\bar{t}$ (b). Les deux grands jets ont une impulsion transverse $p_T > 250$ GeV.

Pour optimiser l'identification du quark *top* hadronique de l'analyse $t\bar{t}H(b\bar{b})$ boostée, l'identification peut s'appuyer sur la cinématique du grand jet, sur ses variables de sous-structures, ou bien sur l'information liée à l'identification du jet de b à l'intérieur du grand

jet. Le nombre de candidats b à l'intérieur du candidats *top lâche* (ΔR (candidat *top*, candidat b) < 1.0), ainsi que le nombre de jets qui ne sont pas candidat b , appelés *non candidats b* , à l'intérieur du candidats *top lâche*, sont considérés dans cette sous-section pour l'identification de quark *top* hadronique. Les distributions de ces deux quantités sont présentées en fonction de la composition partonique du candidat *top lâche* sur la Figure 5.6, pour le signal $t\bar{t}H$ comme pour le bruit de fond $t\bar{t}$. Pour cette étude, le point de fonctionnement 77% est utilisé pour l'identification des jets de b , et les candidats *top lâche* ont une impulsion transverse $p_T > 250$ GeV et sont entourés d'au moins 3 jets et d'au moins 1 candidat b (sélection 3111).

Tout d'abord, les candidats *top lâche* qui ne contiennent qu'un seul quark b de la désintégration d'un quark *top*, à savoir les composantes $(bq\bar{q}')_{thad}$, $b_{thad}i$ ou b_{tlep} , ont une efficacité d'obtenir un candidat b qui correspond bien à l'efficacité du point de fonctionnement fixé à 77% (Figures 5.6a et 5.6b). Dans les événements $t\bar{t}H$, les candidats *top lâche* de composition $b_H i$, qui contiennent un des deux jets de b du boson de Higgs, possèdent plus de candidats b que les composantes $(bq\bar{q}')_{thad}$, $b_{thad}i$ ou b_{tlep} (Figure 5.6a) car les candidats *top lâche* de composition $b_H i$ peuvent aussi contenir par définition tout autre parton issu de la désintégration de la paire $t\bar{t}$. Pour les événements $t\bar{t}H$, les partons des désintégrations de paire $t\bar{t}$ peuvent aussi contaminer les candidats *top lâche* qui capturent les deux jets de b originaires de la désintégration du boson de Higgs, notés $(b\bar{b})_H$. Néanmoins, cette configuration contient principalement exactement deux jets de b (Figure 5.6a). Enfin, les candidats *top lâche* ne contenant aucun parton issu de désintégration de quark *top* ou de boson de Higgs (q/g rad), ainsi que ceux contenant les produits de désintégration de boson W hadronique $((q\bar{q}')_W)$, ne produisent aucun candidat b dans environ 85% des cas (Figures 5.6a et 5.6b). En effet, les radiations de partons à l'origine des candidats *top lâche* de q/g rad n'émettent que peu de quarks b . Pour la composante $(q\bar{q}')_W$, les 10% de candidats *top lâche* qui possèdent au moins un candidat b sont enrichis en désintégrations $W \rightarrow s\bar{c}$.

La variable du nombre de non candidat b est également intéressante, car la composante $(b\bar{b})_H$ (qui correspond au cas où le candidat *top lâche* est en fait constitué des deux jets de b venant du boson de Higgs) a la particularité de contenir zéro non candidat b avec une fraction importante de 42%, comme indiqué Figure 5.6c. À l'inverse, les composantes $b_{thad}i$ possèdent très fréquemment ($\sim 95\%$) au moins un non candidat b . En effet, le critère d'identification *top lâche*, avec sa coupure sur la masse du grand jet, favorise les grands jets composés d'au moins deux jets, et le deuxième jet est généralement léger.

Afin d'améliorer l'identification de quark *top* hadronique pour le canal *top* boosté, de nouvelles conditions sont ajoutées au critère *top lâche*. La préselection de départ est la sélection 3111 définie précédemment. Si la coupure sur l'impulsion transverse du grand jet n'est pas mentionnée, c'est qu'elle est supérieure à 250 GeV (valeur par défaut). Pour quantifier les performances des critères d'identification de quark *top* hadronique, le nombre d'événements sélectionnés de signal et de bruit de fond $t\bar{t}$ sont comparés, pour une luminosité de 20/fb. De plus pour chaque critère, la pureté du candidat *top*, définie comme la fraction de candidats *top* de composition partonique $b_{thad}i$ ou $(bq\bar{q}')_{thad}$, c'est-à-dire contenant au

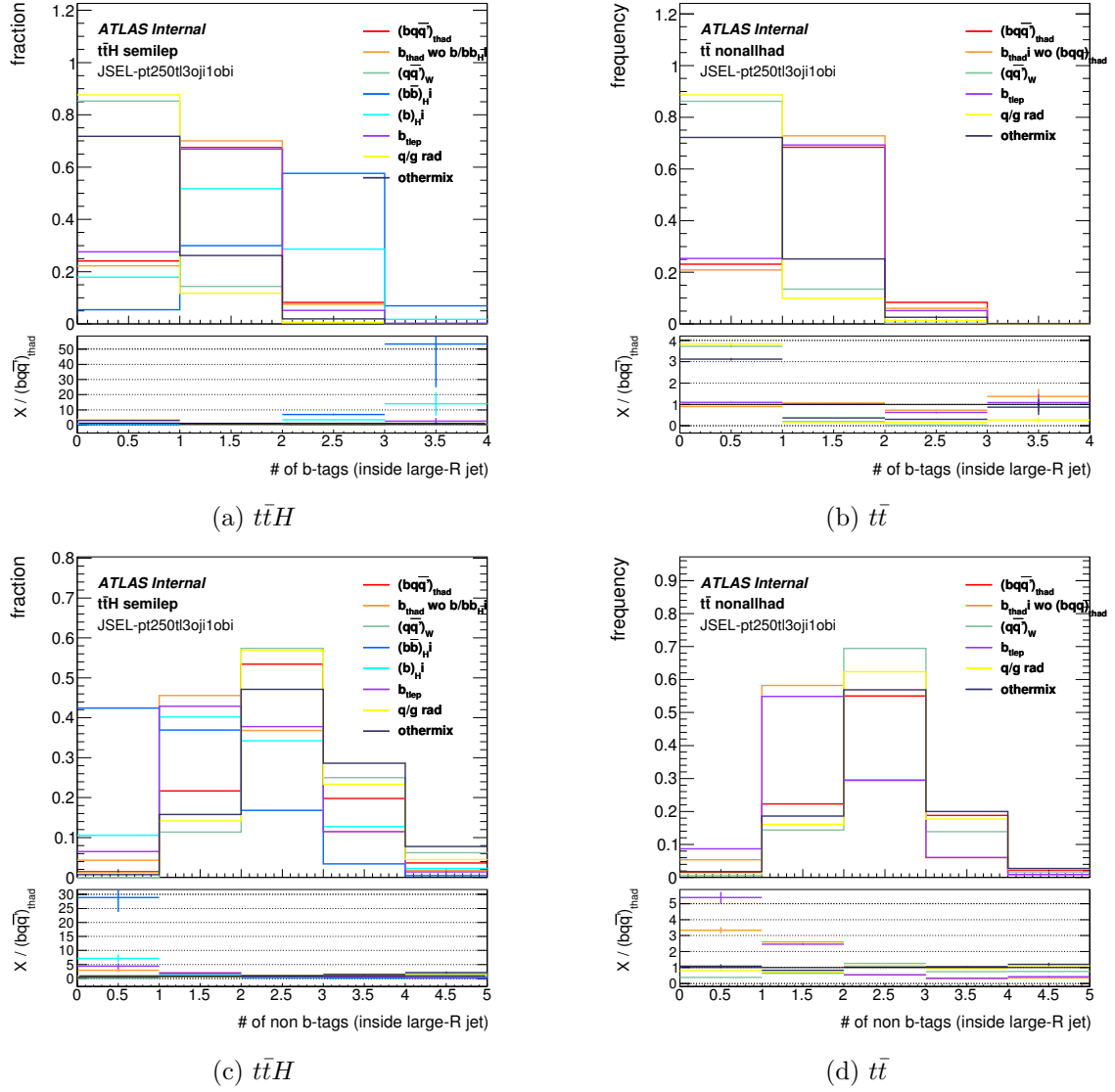


FIGURE 5.6: Distributions du nombre de candidats b pour le signal $t\bar{t}H$ (a) et le bruit de fond (b) et distributions du nombre de jets qui ne sont pas étiquetés b pour le signal $t\bar{t}H$ (c) et le bruit de fond $t\bar{t}$ (d) au sein d'un candidat *top lâche* passant la sélection 3111. Les grands jets sont séparés selon leur origine partonique (voir légende).

moins le quark b provenant de la désintégration hadronique du quark top , est calculée pour le signal comme pour le bruit de fond, et sert également d'indicateur guidant la qualité des critères d'identification du quark top hadronique. A ce stade de la sélection, c'est-à-dire avant de tenter une reconstruction du candidat Higgs à partir des jets à l'extérieur du candidat top , la pureté du candidat top des événements de signal est particulièrement importante.

La condition ' ≥ 1 candidat b à l'intérieur du grand jet' permet d'obtenir un candidat top plus pure que le critère top *dur*, aussi bien pour les événements $t\bar{t}$, avec 74.5% contre 60.3%, que pour ceux de signal, avec 46.9% contre 43.5%, et ceci sans dégrader la sensibilité de la sélection, i.e. le rapport $signal/\sqrt{bruit\ de\ fond}$ (Table 5.3). L'amélioration de la pureté du candidat top *lâche* liée à l'ajout du critère ' ≥ 1 candidat b à l'intérieur du grand jet' est plus importante pour les événements $t\bar{t}$ (74.5% contre 46.2%) que pour les événements $t\bar{t}H$ (46.9% contre 38%), car pour les événements $t\bar{t}H$ le candidat top est contaminé par des grands jets à au moins 1 candidat b provenant des désintégration $H \rightarrow b\bar{b}$. Avec l'ajout de la condition ' ≥ 1 non candidat b à l'intérieur du grand jet' en plus de la condition ' ≥ 1 candidat b à l'intérieur du grand jet', l'amélioration de la pureté du candidat top est plus importante pour le signal, car la nouvelle condition supprime 42% de la composante $(b\bar{b})_H$, sans pour autant perdre en sensibilité ($signal/\sqrt{bruit\ de\ fond} = 0.45$ contre 0.46). Au moins un second non candidat b à l'intérieur du grand jet peut être demandé pour encore augmenter la pureté du candidat top au prix d'une baisse de la sensibilité de la sélection. Cette définition de candidat top permet de mieux isoler les produits de désintégration du quark top hadronique car dans ce cas, on augmente la proportion de la configuration $(bq\bar{q}')_{thad}$.

$p_T > 250$ GeV & $ \eta < 2.0$ & ≥ 3 outside jets & ≥ 1 outside b -tag &	$t\bar{t}$		$t\bar{t}H$ semi-leptonic		S/\sqrt{B}
	purity	yield (B)	purity	yield (S)	
no criterium	25.9%	24808	30.1%	87.3	0.55
top loose	46.2%	20613	38.0%	75.5	0.53
top tight	60.3%	14003	43.5%	53.4	0.45
top loose & ≥ 1 b -tag	74.5%	16089	46.9%	58.0	0.46
top tight & ≥ 1 b -tag	81.0%	11064	50.5%	42.3	0.40
top loose & ≥ 1 b -tag & ≥ 1 non b -tag	75.4%	15326	52.5%	55.8	0.45
top tight & ≥ 1 b -tag & ≥ 1 non b -tag	81.4%	10723	54.8%	41.4	0.40
top loose & ≥ 1 b -tag & ≥ 2 non b -tag	82.7%	6811	59.5%	30.9	0.37
top tight & ≥ 1 b -tag & ≥ 2 non b -tag	84.7%	6043	60.7%	26.7	0.34
BDT	70.1%	17886	55.5%	64.9	0.49

TABLE 5.3: Nombre d'événements $t\bar{t}H$ et $t\bar{t}$ attendus passant la sélection 3111 à $\mathcal{L} = 20/\text{fb}$, pour différents critères d'identification du candidat top , leur pureté de top et sensibilité $signal/\sqrt{bruit\ de\ fond}$ (S/\sqrt{B}) associées. Les candidats top ont une impulsion transverse supérieure à 250 GeV.

L'influence de la coupure sur l'impulsion transverse du candidat top sur la pureté du candidat top a également été étudié. Comme prévu, plus la coupure sur l'impulsion transverse du candidat top augmente, plus ce dernier contient des quarks top hadroniques boostés et plus sa pureté augmente. Ce type de sélection cible les événements plus boostés et donc se focalise sur un espace des phases où les techniques de reconstruction de l'analyse résolue perdent en efficacité. Cependant les sélections dont le candidat top possède une coupure $p_T^{candidat\ top} > 300$ GeV (Table 5.4) ou $p_T^{candidat\ top} > 350$ GeV (Table 5.5) souffrent d'une plus faible sensibilité, du moins pour une luminosité de 20/fb.

$p_T > 300$ GeV & $ \eta < 2.0$ & ≥ 3 outside jets & ≥ 1 outside b -tag &	$t\bar{t}$		$t\bar{t}H$ semi-leptonic		S/\sqrt{B}
	purity	yield	purity	yield	
no criterium	26.4%	14750	32.1%	58.2	0.48
top loose	49.5%	12488	40.9%	50.6	0.45
top tight	66.3%	8413	47.9%	34.5	0.38
top loose & ≥ 1 b -tag	77.8%	9681	49.1%	38.5	0.39
top tight & ≥ 1 b -tag	85.0%	6650	53.8%	27.4	0.34
top loose & ≥ 1 b -tag & ≥ 1 non b -tag	78.8%	9226	54.7%	36.9	0.38
top tight & ≥ 1 b -tag & ≥ 1 non b -tag	85.3%	6471	58.2%	26.8	0.33
top loose & ≥ 1 b -tag & ≥ 2 non b -tag	86.1%	4675	62.6%	21.9	0.32
top tight & ≥ 1 b -tag & ≥ 2 non b -tag	88.2%	4077	64.4%	18.4	0.29

TABLE 5.4: Nombre d'événements $t\bar{t}H$ et $t\bar{t}$ attendus passant la sélection 3111 à $\mathcal{L} = 20$ /fb, pour différents critères d'identification du candidat top , leur pureté de top et sensibilité $signal/\sqrt{bruit\ de\ fond}$ (S/\sqrt{B}) associées. Les candidats top ont une impulsion transverse supérieure à 300 GeV.

Le nombre de candidats b et le nombre de non candidats b sont des variables très utiles pour purifier le candidat top . Il est possible de réoptimiser les coupures sur les variables cinématiques ou de sous-structure pour mieux identifier le quark top hadronique dans cette topologie. Pour ce faire, un BDT de type *gradient boost* [90] est entraîné avec des grands jets d'un échantillon $t\bar{t}H(b\bar{b})$ semi-leptonique afin de séparer les grands jets provenant de $b_{thad}i$ ou $(bqq')_{thad}$ des autres. Six variables sont utilisées en entrée du BDT : la masse et le τ_{32} (déjà utilisées par l'algorithme *top lâche*), deux variables de sous-structures D_2 et τ_{21} qui aident à séparer les grands jets provenant de l'hadronisation de deux partons durs de ceux provenant de l'hadronisation d'un unique parton dur, et enfin le nombre de candidats b et le nombre de non candidats b . Les distributions des variables d'entrées sont présentées Figure 5.7. Au cours de l'entraînement, le BDT peut ainsi exploiter les corrélations non triviales entre ces variables d'entrée, comme celles entre le nombre de candidats b et la masse. Les autres variables de grand jet disponibles, comme l'impulsion transverse ou C_2 (cf. Section 3.6) n'ont pas été rajoutées car elles n'apportaient pas de gain significatif sur les performances, du fait de leur corrélation avec les six variables d'entrée déjà présentes. Les

$p_T > 350 \text{ GeV} \ \& \ \eta < 2.0 \ \& \geq 3 \text{ outside jets} \ \& \geq 1 \text{ outside } b\text{-tag} \ \&$	$t\bar{t}$		$t\bar{t}H$ semi-leptonic		S/\sqrt{B}
	purity	yield	purity	yield	
no criterium	25.7%	8766	33.8%	38.7	0.41
top loose	51.0%	7388	44.0%	33.2	0.39
top tight	69.0%	4982	52.1%	21.9	0.31
top loose & ≥ 1 b -tag	79.7%	5638	52.0%	24.9	0.33
top tight & ≥ 1 b -tag	87.2%	3904	57.5%	17.2	0.28
top loose & ≥ 1 b -tag & ≥ 1 <i>non</i> b -tag	80.8%	5345	57.1%	23.7	0.32
top tight & ≥ 1 b -tag & ≥ 1 <i>non</i> b -tag	87.3%	3791	61.6%	16.8	0.27
top loose & ≥ 1 b -tag & ≥ 2 <i>non</i> b -tag	87.4%	2856	64.9%	14.3	0.27
top tight & ≥ 1 b -tag & ≥ 2 <i>non</i> b -tag	89.7%	2414	68.0%	11.6	0.24

TABLE 5.5: Nombre d'événements $t\bar{t}H$ et $t\bar{t}$ attendus passant la sélection 3111 à $\mathcal{L} = 20/\text{fb}$, pour différents critères d'identification du candidat top , leur pureté de top et sensibilité $signal/\sqrt{\text{bruit de fond}}$ (S/\sqrt{B}) associées. Les candidats top ont une impulsion transverse supérieure à 350 GeV.

réglages internes du BDT proposés par défaut par la bibliothèque TMVA ont été préservés car leur variation n'améliorait pas les performances du BDT (Table 5.6). Ces réglages sont particulièrement appropriés dans cette situation, *i.e* lorsqu'il s'agit de séparer un signal très semblable, par ses distributions de variables d'entrée, au bruit de fond, et lorsque la statistique de l'échantillon d'entraînement est faible.

Paramètre interne BDT	Valeur
nombre d'arbres de décision	1000
MNS	2.5%
profondeur des arbres	2

TABLE 5.6: Principaux réglages internes du BDT pour l'identification de quark top hadronique pour une sélection 3111 d'événements $t\bar{t}H(b\bar{b})$ semi-leptonique

Les distributions attendues du poids du BDT dans des événements $t\bar{t}H$ satisfaisant la sélection 3111 en abandonnant le critère top lâche du grand jet de l'événement est présentée Figure 5.8 pour chacune des origines partoniques de grand jet, à $\mathcal{L} = 20/\text{fb}$. Les plus grands poids sont attribués aux grands jets provenant de la configuration $(bq\bar{q}')_{thad}$ du fait de leur grande masse (autour de 172 GeV) et de leurs sous-structures identifiables car compatibles avec l'hadronisation de 3 partons durs. Les grands jets issus de radiations de partons, facilement distinguables avec leur faible masse et leur absence de candidat b ont quant à eux de faibles poids. La composante $b_H i$ est la plus difficile à séparer, car en plus d'un jet de b du boson de Higgs, elle peut contenir tout autre quark issu de la désintégration d'un quark

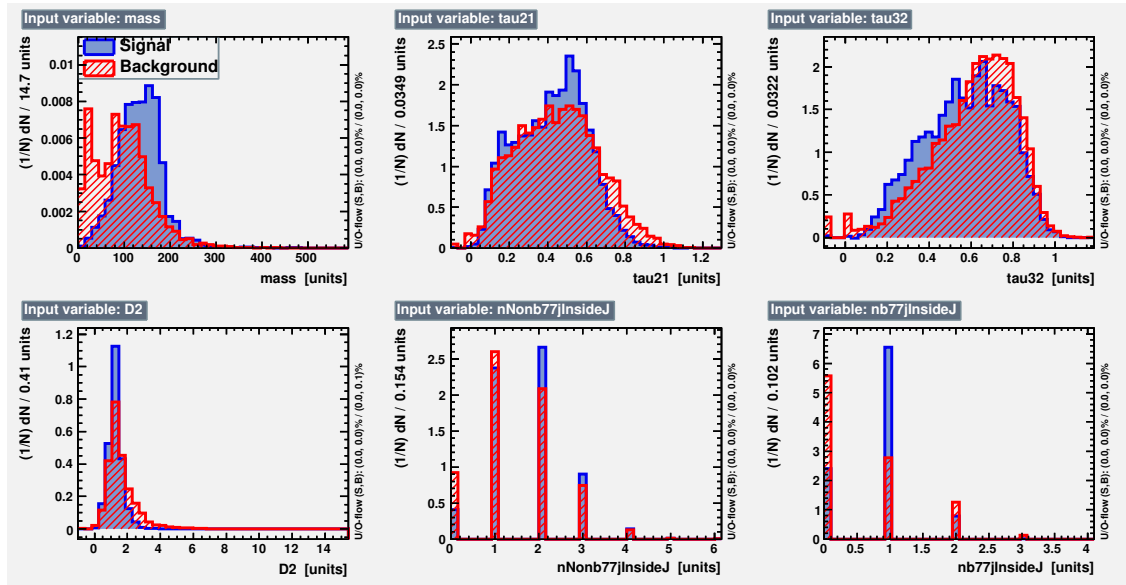


FIGURE 5.7: Distributions des variables d'entrée du *BDT* utilisées pour identifier le quark *top* hadronique pour les grands jets utilisés comme signal (bleu) et comme de bruit de fond (rouge) dans l'entraînement du *BDT*. Les grands jets d'entraînement sont issus d'événements $t\bar{t}H$ semi-leptonique avec une impulsion transverse $p_T > 250$ GeV et sont entourés d'au moins 3 jets et d'au moins 1 candidat *b*. Ici, le signal, constitué de **grands jets de composition b_{thad}^i ou $(bq\bar{q}')_{thad}$** , et le bruit de fond, constitué des **grands jets de toute autre composition**, sont normalisés à 1.

top de $t\bar{t}H$. Cet environnement dense et la présence d'un candidat b la rend similaire aux composantes $b_{thad}i$ et $(bq\bar{q}')_{thad}$. Les composantes $(qq')_W$ et $(b\bar{b})_H$, contenant principalement 0 et 2 candidats b respectivement, sont quant à elles plus facilement séparables des grands jets provenant du quark top .

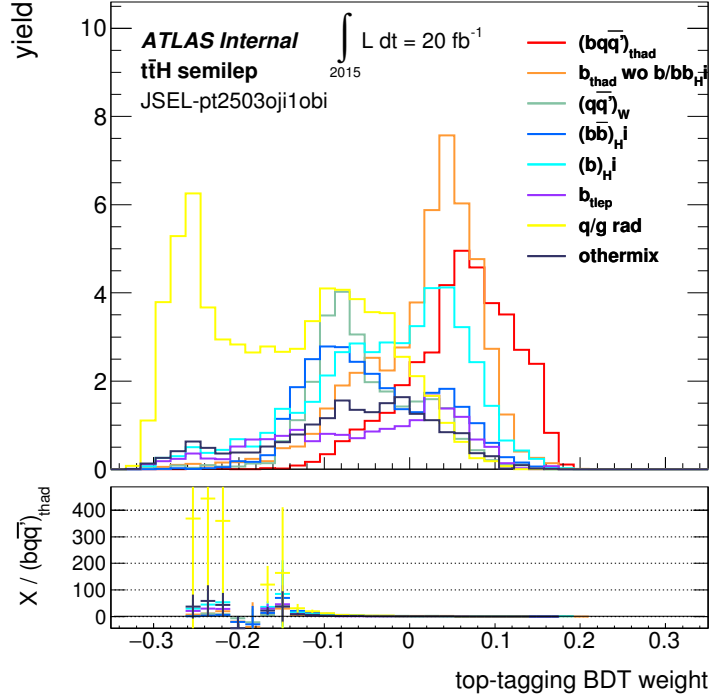


FIGURE 5.8: Distributions du poids de sortie du BDT utilisé pour identifier le quark top hadronique pour des grands jets de $p_T > 250$ GeV entouré d'au moins 3 jets et d'au moins un candidat b dans des événements $t\bar{t}H$. Les grands jets sont séparés selon leur origine partonique (voir légende). Les distributions sont normalisées à une luminosité de 20/fb.

Le candidat top provenant du BDT , ou candidat top BDT , est défini comme un grand jet dont le poids de BDT est supérieur à 0.0. Cette coupure correspond à une pureté de top de 56% pour les candidats top BDT du processus $t\bar{t}H$. La Figure 5.9 compare la composition partonique des candidats top BDT à celle des candidats top lâche. Le nombre d'événements de $t\bar{t}H$ pour ce critère d'identification de quark top hadronique est de 64, contre 58 avec le critère ' top lâche et ≥ 1 candidat b à l'intérieur du grand jet' (Table 5.3). Le critère ' $BDT > 0.0$ ' obtient une très bonne pureté de top pour les événements de signal $t\bar{t}H(b\bar{b})$ semi-leptonique, avec $\sim 56\%$ contre $\sim 47\%$ pour le critère ' top lâche et ≥ 1 candidat b à l'intérieur du grand jet', mais de moins bonnes pour les événements de bruit de fond $t\bar{t}$, avec 70% contre 75% pour ' top lâche et ≥ 1 candidat b à l'intérieur du grand jet'.

Le candidat top BDT est prometteur car il permet notamment de rejeter une grande fraction des grands jets de $(b\bar{b})_H$ par rapport au critère ' top lâche et ≥ 1 candidat b à l'intérieur du grand jet', tout en maintenant une efficacité de sélection similaire. A la suite

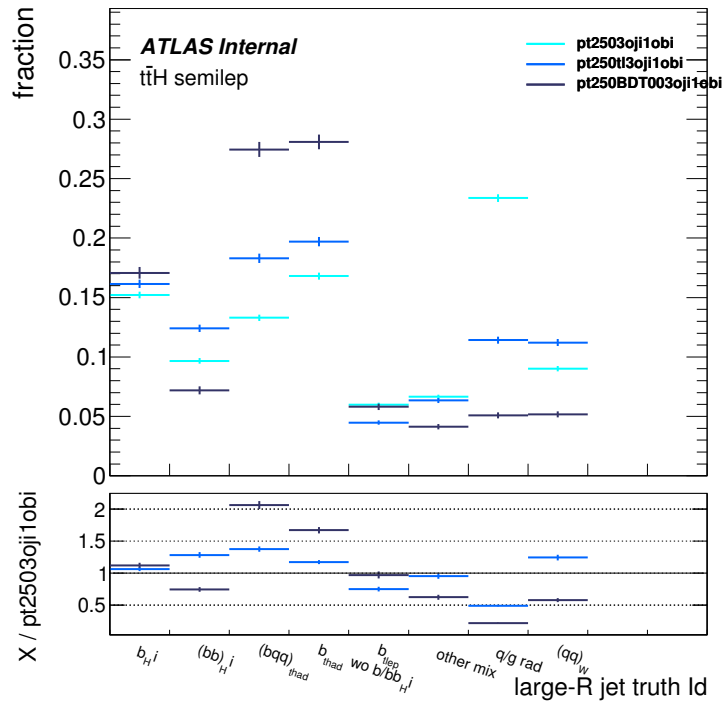


FIGURE 5.9: Composition partonique d'un **grand jet sans critère d'identification** (bleu clair), d'un **grand jet étiqueté *top lâche*** (bleu foncé) et **candidat *top tag BDT*** (noir), dans des événements $t\bar{t}H(b\bar{b})$ semi-leptonique satisfaisant la sélection 3111. Les grands jets ont un impulsion transverse $p_T > 250$ GeV.

de cette étude, le groupe d'analyse $t\bar{t}H$ boosté d'ATLAS a choisi la sélection 3211 avec un candidat top lâche & ≥ 1 candidat b & ≥ 1 non candidat b à l'intérieur du grand jet'. Ce choix est justifié par la simplicité d'implémentation du critère et sa proximité avec celui d'origine (top lâche) qui a fait l'objet de nombreuses études au sein de la collaboration. La coupure sur le nombre de candidat b et/ou sur le nombre de non candidat b rend le contrôle du candidat top dans les données plus simple qu'avec une sélection d'un candidat top provenant du BDT. Enfin, la coupure sur l'impulsion transverse du candidat top est maintenue à 250 GeV, afin de conserver une bonne efficacité de sélection des événements $t\bar{t}H$.

5.5 Etude du canal top et Higgs boostés

5.5.1 Stratégie et sélection d'événements

Une sélection d'événements comportant au moins deux grands jets est étudiée dans cette section afin d'analyser les événements $t\bar{t}H(b\bar{b})$ pour lesquels le quark top hadronique et le boson de Higgs sont tous deux boostés. La stratégie choisie ici est de définir une région boostée très pure. Tout comme la sélection 3211 , cette nouvelle sélection Higgs- top requiert un candidat électron ou muon et un candidat top boosté dans l'événement. De plus, un grand jet additionnel, appelé *candidat Higgs*, est demandé afin de capturer les produits de désintégration du boson de Higgs boosté.

Les premières études de cette topologie sont réalisées en considérant un candidat Higgs simplement défini avec une impulsion transverse supérieure à 200 GeV et la présence d'au moins deux candidats b au sein du grand jet. Cette coupure en impulsion transverse relativement basse permet de cibler des bosons de Higgs moyennement boostés et d'augmenter l'acceptance vu la statistique limitée avec la luminosité actuelle. Plusieurs définitions pour le candidat top proches de celles étudiées dans la partie précédente pour la sélection 3211 sont examinées pour ce canal top et Higgs boostés. Pour chaque critère d'identification de quark top hadronique considéré, une sélection d'événement est réalisée, demandant un candidat top identifié avec le critère en question et un candidat Higgs (précédemment défini).

Les nombres d'événements attendus de signal $t\bar{t}H$ et de bruit de fond $t\bar{t}$ pour chaque critère d'identification de quark top hadronique sont présentés Table 5.7. Les cinq candidats top , d'impulsion transverse supérieure à 250 GeV, sont identifiés à partir du nombre de candidats b , du nombre de non candidats b et de la condition top lâche. Le point de fonctionnement d'identification de jet de b utilisé est 77%. Parmi ces candidats top , les candidats '= 1 candidat b & ≥ 1 non candidat b ', ' top lâche & = 1 candidat b ' et ' top lâche & = 1 candidat b & ≥ 1 non candidat b ' présentent les meilleures performances en terme de rapport signal S sur bruit de fond B , compris entre $S/B = 3\%$ et $S/B = 3.5\%$. Les candidats '= 1 candidat b ' et top lâche atteignent quant à eux de moins bons rapport signal sur bruit de fond, avec $S/B = 1.8\%$, car ils contiennent potentiellement des grands jets de basse masse (par exemple provenant de la configuration b_{tlep} pour les candidats '= 1

candidat b') ou des grands jets provenant de la composante $(b\bar{b})_H$ pour les candidats *top lâche*. Finalement, la sélection '*top lâche* & = 1 candidat b' ' est choisie d'une part pour sa simplicité, et d'autre part parce qu'elle combine les trois informations essentielles sur la masse du grand jet, les sous-structures et l'identification des jets de b .

top-tagging selection ($p_T > 250$ GeV)	$t\bar{t}$ non all-hadronic	$t\bar{t}H$ semi-leptonic	S/B
= 1 b -tag	606	11.0	1.8%
top loose	884	16.2	1.8%
= 1 b -tag & ≥ 1 <i>non</i> b -tag	308	9.2	3.0%
top loose & = 1 b -tag	253	8.3	3.3%
top loose & = 1 b -tag & ≥ 1 <i>non</i> b -tag	238	8.0	3.4%

TABLE 5.7: Nombre d'événements $t\bar{t}H$ et $t\bar{t}$ attendus à $\mathcal{L} = 20/\text{fb}$ pour chaque sélection d'événement. Chaque sélection d'événement contient un candidat *top*, défini comme un grand jet de $p_T > 250$ GeV passant le critère *top lâche* et la condition de la première colonne, et d'un candidat Higgs, défini comme un grand jet de $p_T > 200$ GeV contenant ≥ 2 candidats b .

Pour résumer, la sélection Higgs-*top* choisie est définie par les critères suivants :

bullet [1 candidat électron/muon], de $p_T > 25$ GeV ;

bullet [1 candidat Higgs] grand jet de $p_T > 200$ GeV, ≥ 2 candidats b tels que ΔR (grand jet, candidat b) < 1.0 ;

bullet [1 candidat *top*] grand jet de $p_T > 250$ GeV, *top lâche* et = 1 candidat b tel que ΔR (grand jet, candidat b) < 1.0 .

Le signal sur bruit de la sélection Higgs-*top* peut être encore amélioré en tenant compte de la présence d'un candidat b situé à l'extérieur des candidats *top* et Higgs. Ce candidat b , dit additionnel, a de fortes chances de provenir du jet de b issu de la désintégration du quark *top* leptonique pour les événements de signal, car les jets de b des désintégrations du quark *top* hadronique et du boson de Higgs sont principalement contenus à l'intérieur des candidats *top* et Higgs, comme il sera quantifié par la suite. En revanche, le candidat b additionnel est plus rare pour les événements de bruit de fond, notamment les événements $t\bar{t} + \text{léger}$, pour lesquels les jets de b provenant de la désintégration de la paire $t\bar{t}$ sont tous deux contenus à l'intérieur des candidats *top* et Higgs. Lorsqu'il existe au moins un candidat b additionnel, le rapport signal sur bruit S/B de la sélection Higgs-*top* présentés Table 5.8 est presque 3 fois plus élevé que lorsqu'il n'y a pas de candidat b additionnel.

La même sélection Higgs-*top* est appliquée avec un candidat Higgs d'impulsion transverse supérieure à 250 GeV, afin de retenir des bosons de Higgs plus boostés. Toutefois, cette coupure n'améliore pas significativement le rapport S/B attendue (Table 5.9). C'est pourquoi la coupure sur l'impulsion transverse du boson de Higgs est maintenue par la suite à 200 GeV.

n b -tags outside H and t -tags	$t\bar{t}$	$t\bar{t}H$ semi-leptonic	S/B	S/\sqrt{B}
$n = 0$	191 ± 9	4.8 ± 0.3	2.5%	0.35
$n \geq 1$	51 ± 5	3.4 ± 0.2	6.7%	0.48

TABLE 5.8: Nombre d'événements attendus à $\mathcal{L} = 20/\text{fb}$, avec la sélection Higgs- top et n jets étiquetés b à l'extérieur du candidat top and du candidat Higgs (d'impulsion $p_T > 200$ GeV)

n b -tags outside H and t -tags	$t\bar{t}$	$t\bar{t}H$ semi-leptonic	S/B	S/\sqrt{B}
$n = 0$	146 ± 8	4.0 ± 0.2	2.7%	0.33
$n \geq 1$	41 ± 5	2.4 ± 0.2	5.9%	0.37

TABLE 5.9: Nombre d'événements attendus pour $20/\text{fb}$, avec la sélection Higgs- top et n jets étiquetés b à l'extérieur du candidat top and du candidat Higgs (d'impulsion $p_T > 250$ GeV)

5.5.2 Pureté des candidats top et Higgs boostés

A ce stade, le rapport signal/bruit de fond de la sélection est encore trop faible pour espérer améliorer la sensibilité de l'analyse résolue. Pour aller plus loin, les propriétés des candidats top et Higgs peuvent être exploitées, parce qu'elles sont différentes pour les événements de signal et ceux de bruit de fond. Pour cela, il faut d'abord étudier la composition partonique des candidats top et Higgs.

Pour les événements de signal, la plupart des candidats top et Higgs sont correctement identifiés. En effet, presque 60% des candidats Higgs sont issus de la configuration $(b\bar{b})_H$ (Table 5.10), tandis que $\sim 70\%$ des candidats top proviennent soit de la configuration $(bq\bar{q}')_{thad}$, soit de la configuration $b_{thad}i$ (Table 5.11). Ces événements pour lesquels le quark top hadronique et le boson de Higgs sont tous deux correctement identifiés seront potentiellement les plus faciles à séparer du bruit de fond par la suite. L'identification des quarks top hadroniques et bosons de Higgs peut toutefois être incomplète, comme le montrent les Figures 5.10a et 5.11a. Par exemple, pour le candidat Higgs, la configuration $b_H i$ représente $\sim 20\%$ (32%) pour aucun jet étiqueté b à l'extérieur des candidats top et Higgs (respectivement ≥ 1 jets étiquetés b à l'extérieur).

En ce qui concerne le bruit de fond $t\bar{t}$, la représentation des événements est un peu plus complexe. Les deux candidats top et Higgs ont des compositions partoniques similaires. Dans le cas où aucun jet n'est étiqueté b à l'extérieur des candidats top et Higgs, ils contiennent majoritairement un jet de b issu de la désintégration de la paire $t\bar{t}$: i.e environ 80% proviennent des configurations $b_{thad}i$, $(bq\bar{q}')_{thad}$ ou b_{tlep} (Figures 5.10b et 5.11b). Avec la présence d'un candidat b additionnel, le bruit de fond se trouve enrichi en $t\bar{t} + b$, et une proportion importante ($\sim 35\%$) de candidats Higgs et de candidats top proviennent de la configuration q/g rad, vraisemblablement résultant de la fragmentation d'un gluon en une

paire $b\bar{b}$ ($g \rightarrow b\bar{b}$).

n b -tags outside H and t -tags	$t\bar{t}H$ semi-leptonic
$n = 0$	$(59 \pm 4)\%$
$n \geq 1$	$(57 \pm 5)\%$

TABLE 5.10: Fraction de candidat Higgs dans des événements $t\bar{t}H(b\bar{b})$ semi-leptonique pour la sélection Higgs- top associés à la désintégration du boson de Higgs en paire $b\bar{b}$ (configuration $(b\bar{b})_H$), pour 0 et ≥ 1 candidat b additionnel.

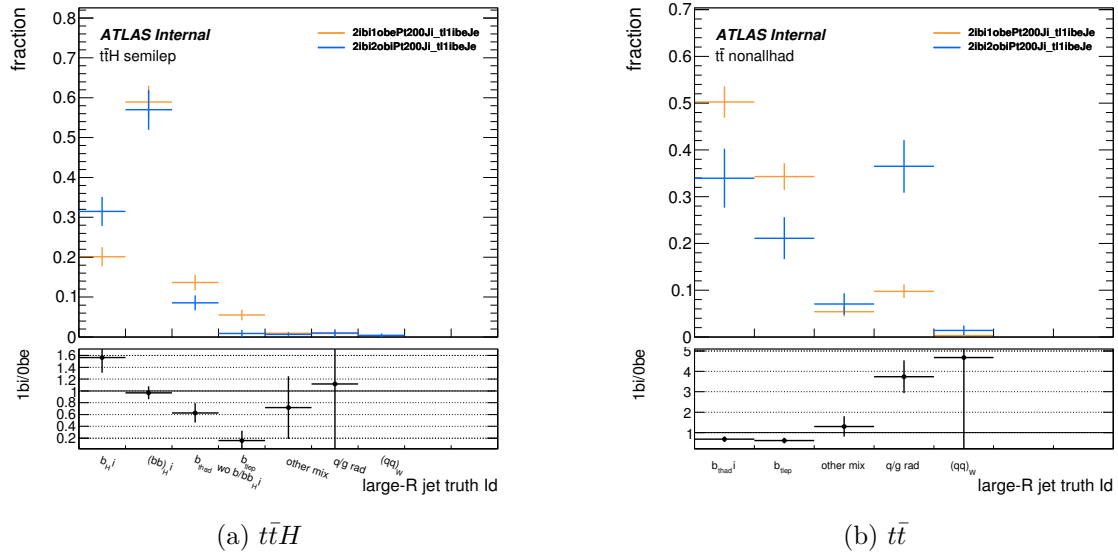


FIGURE 5.10: Composition partonique d'un candidat Higgs pour une sélection Higgs- top avec 0 candidat b additionnel (orange) et ≥ 1 candidat b additionnel (bleu). La configuration $(bq\bar{q}')_{thad}$ est incluse dans $b_{thad}i$.

Une autre sélection contenant deux grands jets a été étudiée, avec pour objectif de sélectionner des événements $t\bar{t}H(b\bar{b})$ semi-leptonique qui produisent un boson de Higgs et un boson W hadronique boostés. Ce canal n'a pas été poursuivi à cause d'une trop faible proportion de signal possédant ces deux particules lourdes boostées par rapport à la topologie top et Higgs boostés, et parce que le boson W hadronique est difficilement identifiable par rapport à des grands jets issus de partons radiés par la paire $t\bar{t}$.

La partie suivante développe la sélection d'événements Higgs- top avec l'étude des topologies d'événement pour séparer au mieux le signal $t\bar{t}H$ du bruit de fond.

n b -tags outside H and t -tags	quark matching	$t\bar{t}$ non all-hadronic	$t\bar{t}H$ semi-leptonic
$n = 0$	$(bq\bar{q}')_{thad}$ & $b_{thad}i$	$(38 \pm 3)\%$	$(70 \pm 5)\%$
	$(bq\bar{q}')_{thad}$	$(19 \pm 2)\%$	$(38 \pm 3)\%$
$n \geq 1$	$(bq\bar{q}')_{thad}$ & $b_{thad}i$	$(41 \pm 6)\%$	$67 \pm 6\%$
	$(bq\bar{q}')_{thad}$	$(24 \pm 5)\%$	$43 \pm 4\%$

TABLE 5.11: Fractions de candidat top d'événements passant la sélection Higgs- top associés aux configurations $(bq\bar{q}')_{thad}$ et $b_{thad}i$ ou associés uniquement à $(bq\bar{q}')_{thad}$, calculées pour les processus $t\bar{t}$ non tout-hadronique et $t\bar{t}H(b\bar{b})$ semi-leptonique.

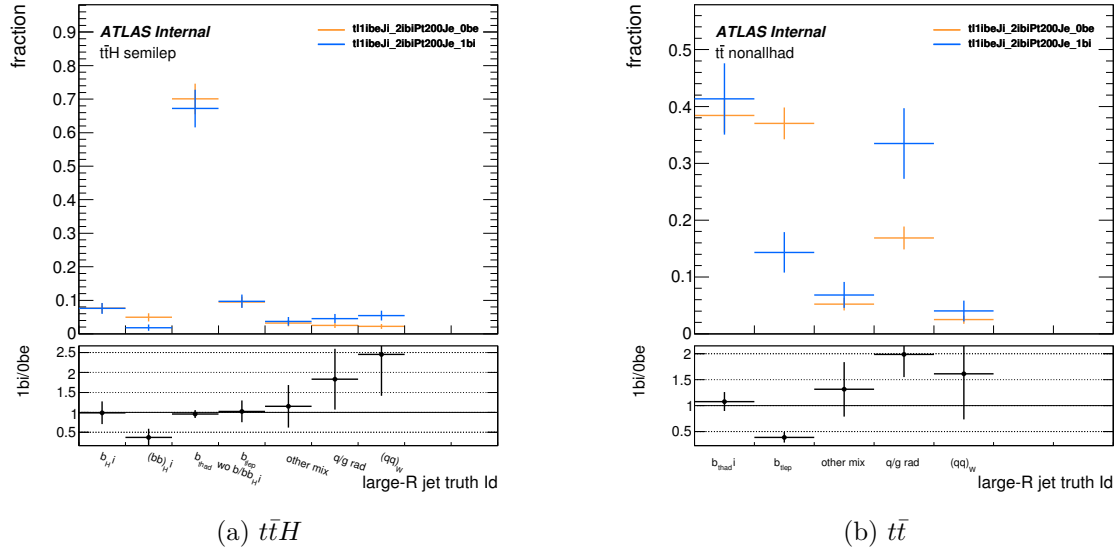


FIGURE 5.11: Composition partonique d'un candidat top pour une sélection Higgs & top tags avec 0 candidat b additionnel (orange) et ≥ 1 candidat b additionnel (bleu) pour le signal $t\bar{t}H(b\bar{b})$ semi-leptonique (a) et le bruit de fond $t\bar{t}$ (b). La configuration $(bq\bar{q}')_{thad}$ est incluse dans $b_{thad}i$.

5.5.3 Classification des événements boostés

Les différences de topologie entre le signal et le bruit de fond peuvent être exploitées par un algorithme multivarié afin d'isoler le signal. Pour ce faire, la sélection des événements boostés est légèrement modifiée par rapport à celle décrite dans les sections précédentes. Afin d'accepter un maximum d'événements de $t\bar{t}H(b\bar{b})$ boostés, le point de fonctionnement pour l'identification de jet de quark b passe de 77% à 85% d'efficacité, et la variable d'identification des jets de quark b MV2c10, dont les coupures définissent les différents points de fonctionnement, est directement utilisée en entrée de l'algorithme multivarié.

Le passage d'un candidat b 77% à un candidat b 85% implique également un léger changement de définition du candidat top . Demander exactement un candidat b 85% retire une partie des candidats top précédemment sélectionnés qui contenaient plus d'un candidat b à 85% d'efficacité. La nouvelle condition sur le nombre de candidat b 85% à l'intérieur du candidat top est relâchée à ≥ 1 candidat b 85%. Pour le candidat Higgs, la condition ≥ 2 candidats b 77% passe simplement à ≥ 2 candidats b 85%. Il existe alors un recouvrement des définitions entre candidat Higgs et du candidat top , lorsque ce dernier possède deux candidats b . Cette ambiguïté sera résolue par la suite. En plus du changement de point de fonctionnement pour l'identification de jet de b , une coupure inférieure sur la masse des grands jets est appliquée à 50 GeV. Celle-ci n'a aucune incidence sur le candidat top , dont la masse est déjà supérieure à 80 GeV du fait du critère top lâche. Elle a une faible influence sur le candidat Higgs, en raison de la corrélation entre la masse et le nombre de jets. Demander deux jets à l'intérieur d'un grand jet favorise les grands jets lourds, de masse de l'ordre de $p_{T,min}^J \cdot R_J/2 = 200 \times 1.0/2 = 100$ GeV. Enfin, la sélection Higgs- top étudiée dans la partie précédente se limite au cas où il n'existe qu'un seul candidat Higgs et un seul candidat top . Les événements qui contiennent plusieurs de ces deux candidats sont très rares ($< 1\%$). Un unique candidat Higgs et un unique candidat top sont choisis, dont les caractéristiques serviront à séparer le signal du bruit de fond. Le candidat Higgs est choisi comme celui dont la somme des valeurs MV2c10 des candidats b 85% est la plus élevée. Le candidat top est défini comme le candidat top le plus lourd. Dans toute la suite, tout candidat b de l'analyse boostée sera implicitement un candidat b du point de fonctionnement 85%. La définition de la région de signal se résume donc ainsi :

= 1 **lepton** candidat électron ou muon $p_T > 25$ GeV ;

≥ 1 **candidat Higgs** grand jet $p_T > 200$ GeV, $m > 50$ GeV & ≥ 2 candidats b . Le candidat Higgs dont la somme des MV2c10 des candidats b (WP 85%) est la plus élevée est choisi ;

≥ 1 **candidat top** grand jet $p_T > 250$ GeV, top lâche & ≥ 1 candidat b . Le candidat top dont la masse est la plus élevée, différent du candidat Higgs, est choisi.

L'analyse boostée a vocation à être combinée avec l'analyse résolue. Pour que ces analyses soient indépendantes, les événements de la région boostée sont retirés de l'analyse résolue. Les événements de l'analyse résolue sont séparés selon leur nombre de jets et de candidat b (point de fonctionnement 70%) avec au total neuf régions définies en fonction

du nombre de jets et du nombre de candidats b produits dans l'événement : $\{3, 5, \geq 6\}$ jets $\otimes \{2, 3, \geq 4\}$ candidats b [88]. Les régions avec davantage de jets et de candidats b sont enrichies en signal, à l'image des régions dites de signal ' $5j, \geq 4b$ ' ' $\geq 6j, 3b$ ' et la région la plus sensible ' $\geq 6j, \geq 4b$ '. Les six autres régions, dites de contrôle, sont enrichies en bruit de fond, notamment $t\bar{t} + \text{léger}$ et $t\bar{t} + c$, et servent principalement à ajuster les prédictions MC aux données récoltées. La région boostée peut être vue comme une dixième région de l'analyse résolue.

Environ 1090 événements sont attendus dans la région boostée dans les données 2015 et 2016 (correspondant à 36.5/fb), dont ~ 27 événements $t\bar{t}H$. Le bruit de fond est majoritairement composé de $t\bar{t} + \text{léger}$ (384) avec une part importante de $t\bar{t} + c$ (217) et de $t\bar{t} + b$ (271). Par ordre d'importance viennent ensuite les processus *top unique* (74) et $W + \text{jets}$ (58), puis les processus du second ordre comme $t\bar{t} + Z$, $t\bar{t} + W$, diboson, $Z + \text{jets}$ et autres $t\bar{t} + V$ ($O(10)$) (Table 5.12). La région résolue ' $\geq 6j, \geq 4b$ ' est quant à elle dominée par le processus $t\bar{t} + b$, avec 1710 événements $t\bar{t} + b$ parmi les 2240 événements sélectionnés. En fait, une fraction importante des événements de la région boostée n'est pas incluse originairement dans la région ' $\geq 6j, \geq 4b$ ', avec 21 événements $t\bar{t}H$ hors de la région ' $\geq 6j, \geq 4b$ ' sur les 27 événements $t\bar{t}H$ sélectionnés par l'analyse boostée (Table 5.13). Par exemple 9 d'entre eux sont présents dans la région moins sensible ' $\geq 6j, 3b$ ' parmi les 192 événements sélectionnés au total par l'analyse boostée. La sensibilité de la sélection Higgs-*top* reposerait donc moins sur l'identification des jets de quark b pour extraire le signal du bruit de fond.

	4j, 2b	4j, 3b	4j, $\geq 4b$	5j, 2b	5j, 3b	$\geq 6j, 2b$	5j, $\geq 4b$	$\geq 6j, 3b$	$\geq 6j, \geq 4b$	Boosted
$t\bar{t}H$	160 \pm 3.3	60.0 \pm 1.3	7.6 \pm 0.1	256 \pm 6.4	126 \pm 3.1	550 \pm 15.6	29.3 \pm 0.7	312 \pm 8.3	112 \pm 3.1	26.7 \pm 2.3
$t\bar{t}$ +light	422 000 \pm 5410	13 000 \pm 191	45.5 \pm 5.6	241 000 \pm 4230	9080 \pm 176	146 000 \pm 3490	61.6 \pm 6.1	6650 \pm 154	101 \pm 10.9	384 \pm 45.6
$t\bar{t} + \geq 1c$	28 200 \pm 423	2160 \pm 47.7	32.1 \pm 4.0	29 100 \pm 557	2850 \pm 64.9	31 800 \pm 804	69.8 \pm 5.2	3900 \pm 105	158 \pm 9.9	217 \pm 25.0
$t\bar{t} + \geq 1b$	11 900 \pm 188	3940 \pm 63.9	175 \pm 4.9	11 800 \pm 236	5740 \pm 114	13 300 \pm 340	544 \pm 12.4	8510 \pm 220	1710 \pm 45.7	271 \pm 30.2
$t\bar{t} + W$	255 \pm 4.9	9.9 \pm 0.3	0.1 \pm 0.03	352 \pm 8.1	20.9 \pm 0.6	562 \pm 13.6	0.4 \pm 0.06	52.7 \pm 1.4	3.1 \pm 0.2	14.2 \pm 1.1
$t\bar{t} + Z$	291 \pm 6.5	35.8 \pm 1.2	3.4 \pm 0.3	396 \pm 10.3	70.8 \pm 2.1	753 \pm 20.4	11.6 \pm 0.6	169 \pm 4.9	44.6 \pm 1.8	20.9 \pm 2.2
Single Top	26 900 \pm 413	969 \pm 24.2	23.0 \pm 3.7	13 100 \pm 278	761 \pm 22.0	7400 \pm 188	35.4 \pm 3.7	717 \pm 25.6	70.8 \pm 5.3	74.1 \pm 8.2
Diboson	728 \pm 20.8	33.3 \pm 3.3	2.2 \pm 0.8	401 \pm 16.3	35.2 \pm 4.2	335 \pm 15.9	3.1 \pm 1.7	38.6 \pm 4.4	3.7 \pm 1.0	8.4 \pm 2.5
$W + \text{jets}$	16 300 \pm 303	450 \pm 79.2	8.6 \pm 1.9	7430 \pm 207	364 \pm 16.8	4740 \pm 119	14.9 \pm 2.8	315 \pm 71.2	25.8 \pm 2.0	58.3 \pm 6.3
$Z + \text{jets}$	3140 \pm 53.6	94.2 \pm 13.2	2.0 \pm 0.4	1360 \pm 37.4	68.2 \pm 4.2	828 \pm 19.9	2.6 \pm 0.4	75.0 \pm 3.3	6.1 \pm 1.1	7.4 \pm 1.0
Other $t\bar{t} + V$	90.4 \pm 1.8	20.0 \pm 0.5	1.5 \pm 0.1	65.0 \pm 1.6	11.7 \pm 0.4	94.7 \pm 2.2	2.3 \pm 0.2	32.6 \pm 0.8	10.6 \pm 0.3	6.2 \pm 0.4
Total	510 000 \pm 6680	20 800 \pm 320	301 \pm 10.5	305 000 \pm 5480	19 100 \pm 366	206 000 \pm 5000	775 \pm 17.9	20 800 \pm 514	2240 \pm 60.9	1090 \pm 117

TABLE 5.12: Nombre d'événements attendus de l'analysis $t\bar{t}H(b\bar{b})$ avec 36.5/fb, dans chaque région de l'analyse résolue, et dans la région boostée. Seules les incertitudes liées à la statistique des événements MC sont considérées ici.

Un BDT de type *Adaboost* [56] est entraîné spécifiquement avec des événements de la région boostée afin de séparer le signal $t\bar{t}H$ des processus de bruit de fond. Les Figures 5.12-5.16, montrent les distributions de plusieurs variables liées au candidat Higgs (5.12, 5.13), au candidat *top* (5.14, 5.15), ou à la topologie de l'événement à l'extérieur ou indépendamment des candidats *top* et Higgs (5.16). La discrimination est principalement due à l'identification des jets de quark b , ainsi que des propriétés (masse, variables de sous-structures) des candidats candidats *top* et Higgs.

La liste des 11 variables retenues pour le BDT est la suivante :

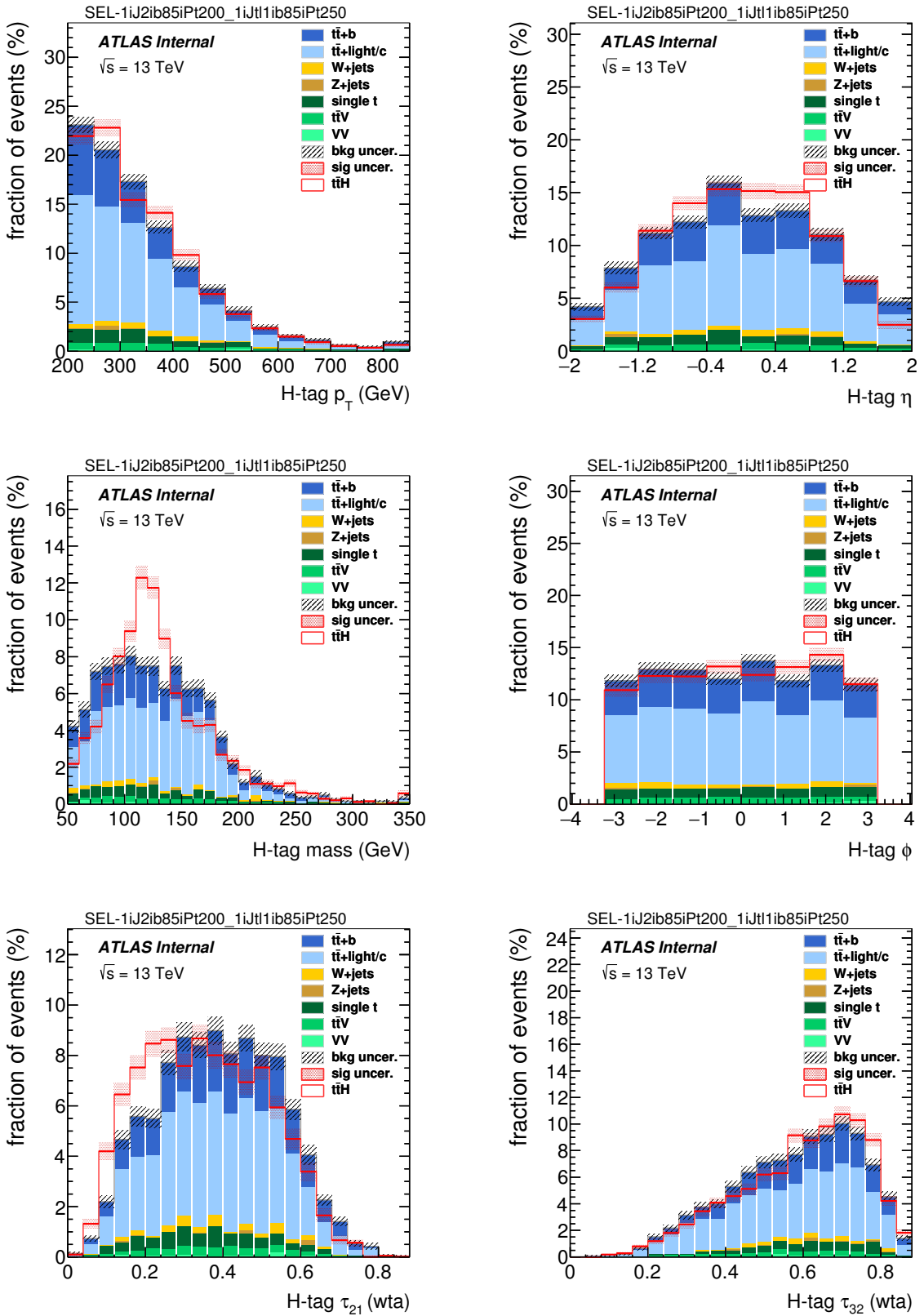


FIGURE 5.12: Distributions des variables liées au candidat Higgs pour une sélection Higgs-*top*. Les distributions du signal et du bruit de fond sont normalisées à 1.

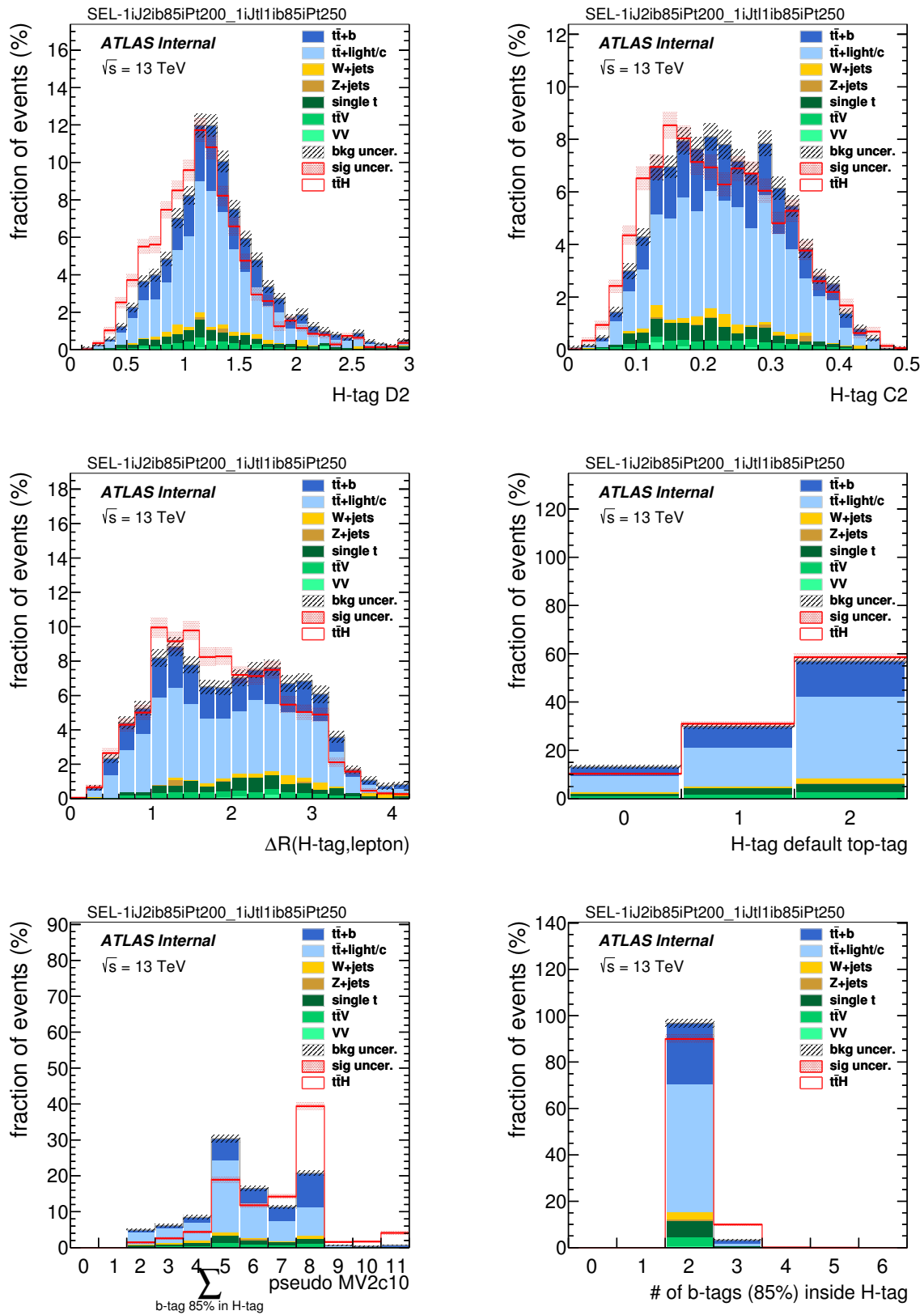


FIGURE 5.13: Distributions des variables liées au candidat Higgs pour une sélection Higgs-*top*. Les distributions du signal et du bruit de fond sont normalisées à 1.

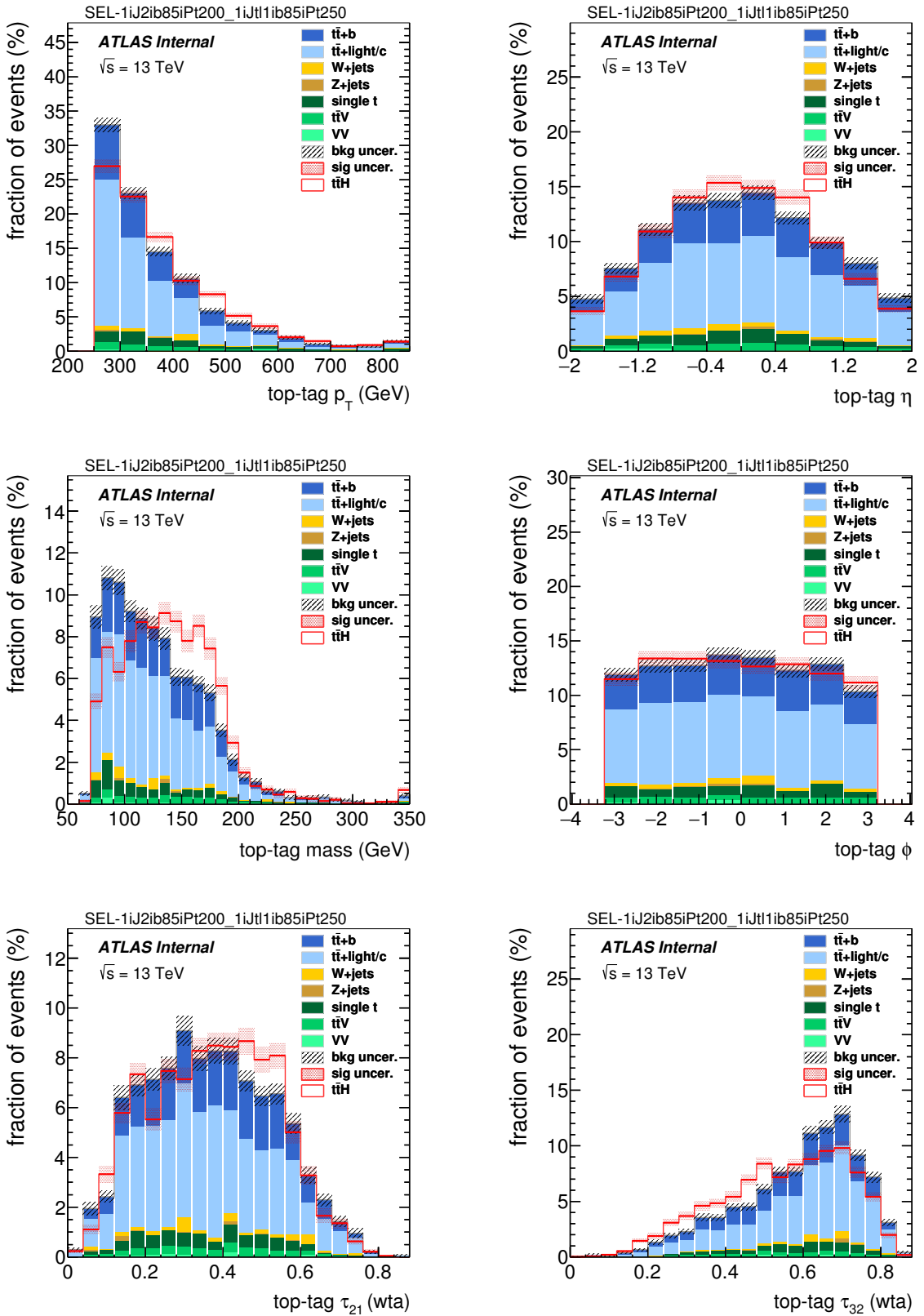


FIGURE 5.14: Distributions des variables liées au candidat top pour une sélection Higgs- top . Les distributions du signal et du bruit de fond sont normalisées à 1.

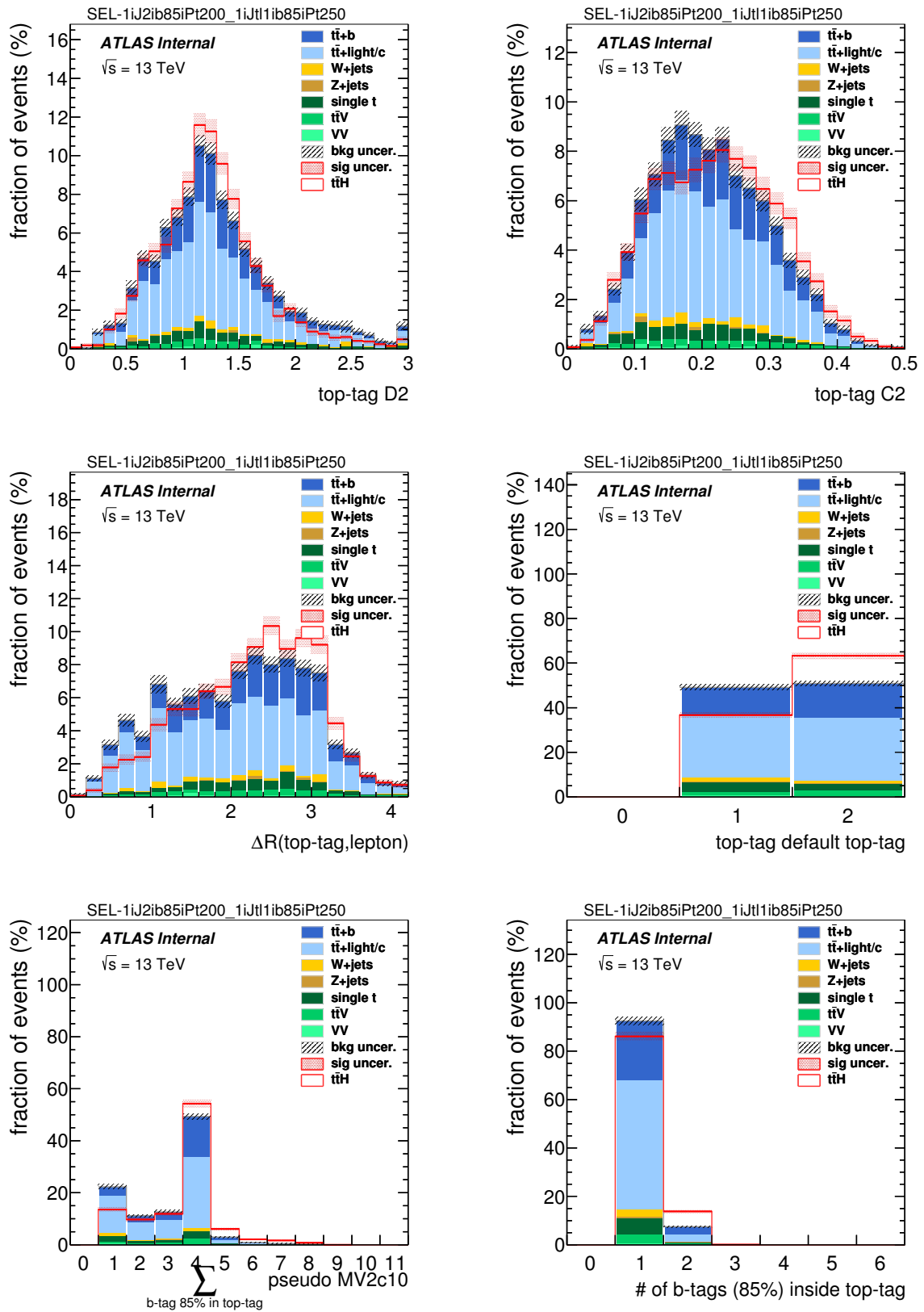


FIGURE 5.15: Distributions des variables liées au candidat top pour une sélection Higgs- top . Les distributions du signal et du bruit de fond sont normalisées à 1.

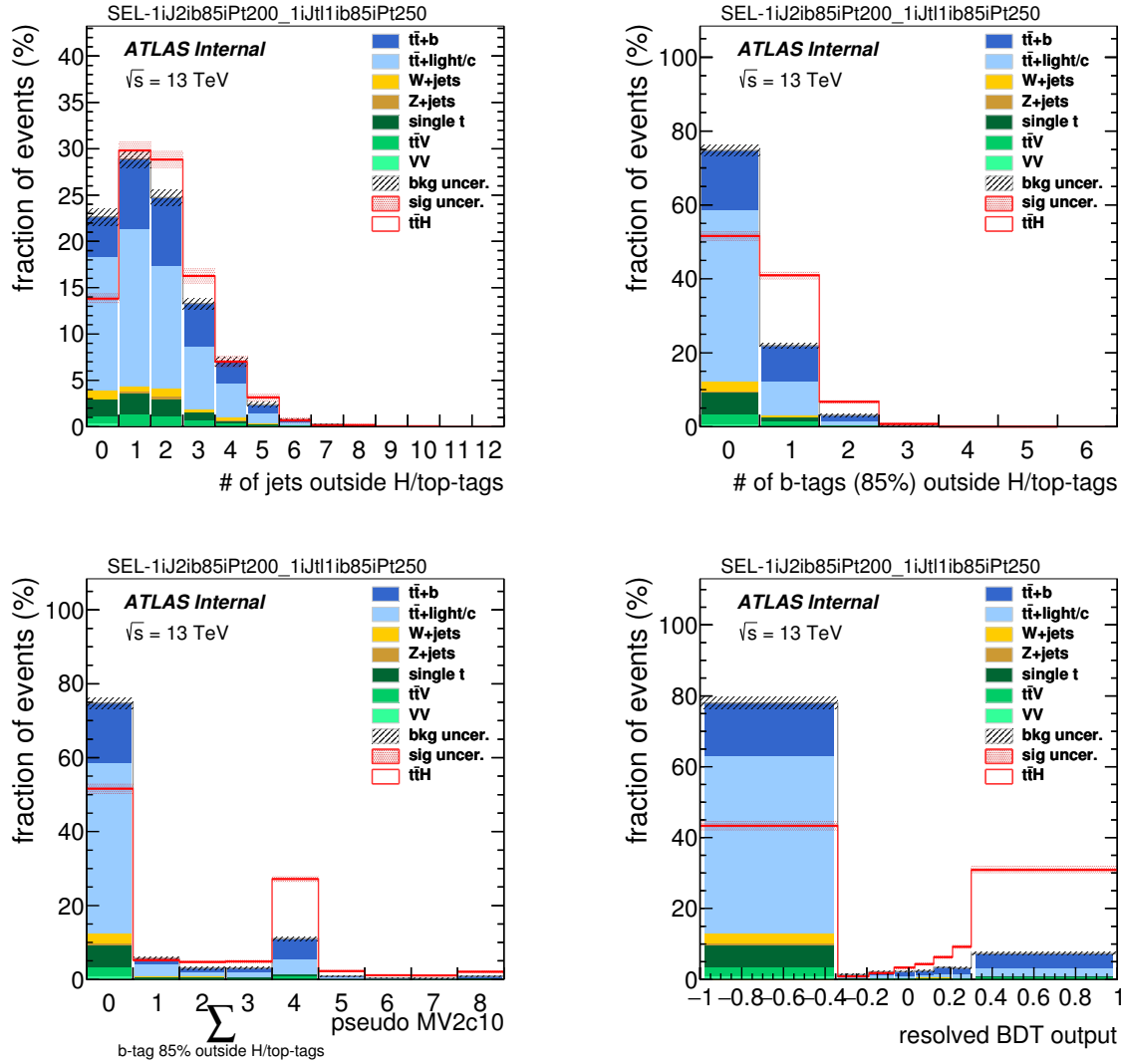


FIGURE 5.16: Distributions des variables non associées au candidat top/Higgs pour une sélection Higgs-top. Les distributions du signal et du bruit de fond sont normalisées à 1.

	4j, 2b	4j, 3b	4j, $\geq 4b$	5j, 2b	5j, 3b	$\geq 6j, 2b$	5j, $\geq 4b$	$\geq 6j, 3b$	$\geq 6j, \geq 4b$
$t\bar{t}H$	0.3 ± 0	0.4 ± 0	0.04 ± 0	1 ± 0	2 ± 0	6 ± 0	0.6 ± 0	9.0 ± 0	5.5 ± 0
$t\bar{t}+\text{light}$	22.1 ± 2.9	4.3 ± 1.05	0 ± 0	54.6 ± 4.7	13.5 ± 2.4	137 ± 7.3	0.5 ± 0.5	37.9 ± 4.7	0.4 ± 0.3
$t\bar{t}+\geq 1c$	4.4 ± 1.0	2.5 ± 1.0	0 ± 0	16.2 ± 2.3	9.3 ± 1.9	86.9 ± 6.7	0 ± 0	31.8 ± 3.6	2.5 ± 0.8
$t\bar{t}+\geq 1b$	4.9 ± 0.7	5.1 ± 0.8	0.2 ± 0.1	14.2 ± 1.3	19.7 ± 1.6	62.5 ± 2.6	4.3 ± 0.8	92.3 ± 3.0	45.6 ± 2.3
$t\bar{t}+W$	0.5 ± 0.06	0.2 ± 0.03	0 ± 0	1.4 ± 0.1	0.5 ± 0.08	5.3 ± 0.2	0.02 ± 0.008	2.0 ± 0.1	0.3 ± 0.05
$t\bar{t}+Z$	0.3 ± 0.07	0.4 ± 0.08	0 ± 0	1.4 ± 0.2	1.0 ± 0.2	6.0 ± 0.5	0.4 ± 0.1	4.6 ± 0.5	3.1 ± 0.4
Single Top	7.0 ± 1.8	1.6 ± 0.8	0 ± 0	8.4 ± 1.5	3.5 ± 0.9	20.3 ± 2.5	0.2 ± 0.2	7.7 ± 1.5	1.5 ± 0.7
Diboson	1.0 ± 0.4	0 ± 0	0 ± 0	0.6 ± 0.3	0 ± 0	1.6 ± 0.8	0.2 ± 0.2	0.4 ± 0.3	0 ± 0
$W+\text{jets}$	2.4 ± 0.6	0.8 ± 0.4	0.02 ± 0.02	5.5 ± 0.8	1.5 ± 0.4	12.9 ± 1.3	0.02 ± 0.02	3.9 ± 0.5	1.1 ± 0.3
$Z+\text{jets}$	0.2 ± 0.06	0.1 ± 0.09	0 ± 0	0.7 ± 0.1	0.3 ± 0.1	1.9 ± 0.5	0 ± 0	0.6 ± 0.1	0.06 ± 0.03
Other $t\bar{t} + V$	0.03 ± 0.01	0.03 ± 0.01	0 ± 0	0.2 ± 0.1	0.05 ± 0.02	1.7 ± 0.09	0.006 ± 0.006	2.1 ± 0.08	1.4 ± 0.05
Total	43.2 ± 3.7	15.4 ± 1.9	0.3 ± 0.1	104 ± 5.6	50.9 ± 3.6	342 ± 11	6.3 ± 1.0	192 ± 6.8	61.4 ± 2.6

TABLE 5.13: Nombre d'événements qui passent la sélection boostée attendus dans chaque région de l'analyse résolue, avec 36.5/fb. Seules les incertitudes liées à la statistique des événements MC sont considérées ici.

candidat Higgs : masse, D_2 , $\sum_{b\text{-tag}\in H} \text{MV2c10}$;

candidat top : p_T , masse, τ_{32} , $\sum_{b\text{-tag}\in top} \text{MV2c10}$;

autre : $\sum_{b\text{-tag}\notin H/top} \text{MV2c10}$, $\Delta R(\text{candidat Higgs}, \text{candidat top})$, nb de jets hors Higgs/top, BDT résolu.

Examinons les différences entre signal et bruit de fond pour ces variables. Comme nous l'avons vu dans la partie précédente, le candidat Higgs vient principalement de la désintégration du boson de Higgs pour le signal, alors que pour le bruit de fond c'est un mélange de quark top hadronique, de quark top leptonique ou de radiation dans le cas $t\bar{t} + b$ (c.f. distributions des candidat Higgs Figure 5.12 et 5.13). Par conséquent, la masse des candidats Higgs montre un pic autour de la masse du boson de Higgs à 125 GeV pour le signal, tandis que pour le bruit de fond la distribution reste plutôt plate et décroît rapidement pour des masses au dessus de celles du quark top (172 GeV). De plus, pour le signal le candidat Higgs provient davantage de l'hadronisation de deux partons durs que pour le bruit de fond. Ainsi la valeur de D_2 pour le candidat Higgs du signal est en moyenne plus petite que pour le bruit de fond. On remarque également que la somme des MV2c10 des candidats b dans le candidat Higgs est plus élevée pour le signal que pour le bruit de fond qui ne possède souvent qu'un seul vrai jet de quark b .

La composition partonique du candidat top est également bien différente selon l'origine de l'événement. La plupart des quarks top hadroniques du signal sont reconstruits – au moins partiellement – au sein de candidats top , ce qui n'est pas le cas pour le bruit de fond (Figure 5.14 et 5.15). L'impulsion transverse du candidat top est utilisée comme variable d'entrée du BDT car son spectre est plus dur pour le signal. Cette variable est non seulement discriminante, mais aussi corrélée avec la pureté du candidat top , qui est d'autant plus grande que l'impulsion transverse du candidat top est élevée. Cette corrélation est exploitée par le BDT pour améliorer la séparation des événements fortement boostés. La masse du candidat top est également intéressante. En effet le candidat top est en moyenne plus lourd dans les événements de signal que pour le bruit de fond, grâce à une proportion plus élevée

de vrais candidats top dont la masse est proche de celle du quark top . La variable de sous-structure τ_{32} est légèrement discriminante, avec de plus faibles valeurs pour le candidat top du signal, dont la topologie est plus proche de celle de l’hadronisation de trois partons. La variable d’identification des jets de quark b joue un rôle moins important que celle associée au candidat Higgs.

Les quatre dernières variables utilisées en entrée du BDT ne sont pas associées à la composition des candidats top et Higgs (Figure 5.16). La somme des MV2c10 des candidats b à l’extérieur des candidats top et Higgs est une variable qui présente une séparation particulièrement bonne. En effet, elle est plus élevée pour le signal car le bruit de fond possède moins de jets de b . Les candidats top et Higgs sont un peu plus proches l’un de l’autre dans les événements $t\bar{t}H(b\bar{b})$ que dans le bruit de fond, ce qui est à l’origine du léger pouvoir discriminant de la variable ΔR (candidat Higgs, candidat top). Le nombre de jets en dehors des candidats top et Higgs est aussi utilisé car, en moyenne, le signal en possède plus. Enfin, le poids du BDT de l’analyse résolue est aussi inclu dans la liste des variables d’entrée, et l’importance de sa présence sera discutée dans la suite avec la présentation des résultats du fit.

La plupart des variables utilisées dans le BDT sont légèrement corrélées entre elles aussi bien pour le signal que pour le bruit de fond, avec un facteur de corrélation inférieur à 10% en valeur absolue (Figure 5.17). Le poids du BDT résolu, qui est utilisé pour séparer le signal du bruit de fond pour les régions de signal de l’analyse résolue [88], est particulièrement corrélé avec les variables d’identification de b avec par exemple une corrélation de 48% et 43% avec la variable $\sum_{b-tag \in H} MV2c10$ pour le signal et le bruit de fond respectivement. Les variables cinématiques et de sous-structures du candidat top sont également corrélées entre elles.

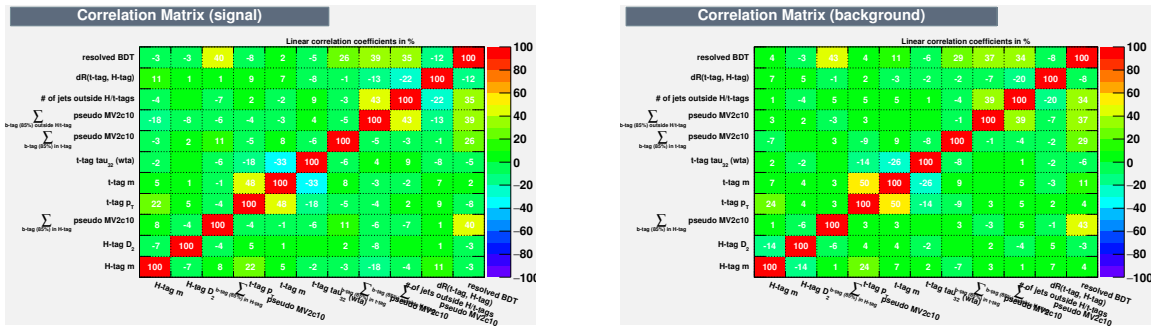


FIGURE 5.17: Corrélations des variables d’entrée du BDT du canal boosté pour le signal (a) et le bruit de fond (b) pour la sélection Higgs- top .

Les paramètres du BDT sont optimisés pour maximiser le rapport signal sur bruit tout en maintenant un surentraînement acceptable, de l’ordre du pourcent.

La distribution du poids du BDT des événements boostés est présentée Figure 5.18. Les plus grandes valeurs du BDT sont les plus enrichies en signal et le processus $t\bar{t} + b$ y est le

bruit de fond dominant.

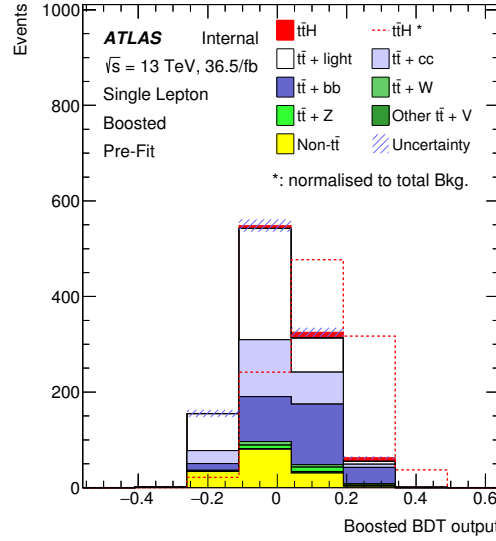


FIGURE 5.18: Distribution attendue du poids du BDT pour 36.5/fb. Les événements considérés passent la sélection Higgs-*top*.

5.6 Incertitudes de l'analyse boostée

Jusqu'à présent, seules les incertitudes provenant de la statistique des échantillons Monte Carlo ont été considérées. De nombreuses autres sources d'incertitude systématiques affectent la justesse des prédictions et doivent être prises en compte dans l'analyse avant de conclure sur la présence de signal dans les données observées. Ces incertitudes sont décrites brièvement dans cette section. Les incertitudes systématiques résultent de l'incertitude de la description des événements depuis les collisions pp jusqu'à la réponse du détecteur. Comme nous allons le voir, les incertitudes qui affectent significativement la mesure dans le canal $lepton + jets$ sont liées à la description du bruit de fond dominant $t\bar{t} + b$ et du signal $t\bar{t}H$, et à la calibration des jets et de l'étiquetage des jets de b .

5.6.1 Modélisation du signal et du bruit de fond

La modélisation du bruit de fond dans les régions de signal est un des enjeux majeurs de l'analyse. En effet, les simulations actuelles des bruits de fond principaux proviennent de calculs d'ordre trop bas en α_s (NLO pour $t\bar{t}$), et ne parviennent pas à décrire avec une précision suffisante les propriétés des événements dans les régions de signal, à haute multiplicité de jets, et plus particulièrement à haute multiplicité de jets de b .

Incertitudes sur le bruit de fond $t\bar{t} + jets$

Le processus $t\bar{t}$ est le bruit de fond dominant de l'analyse. Par conséquent, l'incertitude sur sa modélisation affecte grandement la force du signal, définie comme la mesure de la

section efficace du processus de production $t\bar{t}H$ normalisée à la section efficace prédite par le modèle standard. De nombreuses incertitudes sont prises en compte.

Une incertitude de $\pm 6\%$ est attribuée à l'erreur sur le calcul NNLO+NNLL de la section efficace de production $pp \rightarrow t\bar{t}$. Celle-ci découle des incertitudes sur les échelles de normalisation et de factorisation, les PDF, la constante de couplage α_s et la masse du quark *top*. Plusieurs sources d'incertitude viennent de la modélisation des événements $t\bar{t}$ dans le MC. Pour une source donnée, l'incertitude $\pm 1\sigma$ est donnée par la différence entre deux implémentations MC différentes, simulées avec la version simplifiée ATLFast-II du détecteur ATLAS (cf. Section 3.7). Dans le fit qui sera décrit dans la section suivante, les variations pour chacune de ces sources sur les composantes $t\bar{t} + \text{léger}$, $t\bar{t} + c$ et $t\bar{t} + b$ sont considérées indépendantes. Une première incertitude est associée au choix de l'élément de matrice du générateur. Les prédictions des générateurs NLO Powheg et aMC@NLO, tous deux interfacés avec le générateur d'hadronisation Herwig++, sont comparés. Une seconde source d'incertitude liée aux modèles de fragmentation et hadronisation est dérivée de la différence entre les modèles d'hadronisation Pythia6 et Herwig++ associés avec le même générateur pour l'élément de matrice. Enfin, une incertitude sur le niveau de radiation de parton est évaluée en comparant deux configurations de Powheg+Pythia6 : un échantillon avec davantage de radiations qui possède des échelles de renormalisation et de factorisation divisées par deux, le paramètre *hdamp* doublé et le paramètre Var3c du tune A14 augmenté est comparé avec un échantillon qui a des radiations plus faibles : des échelles doublées et un paramètre Var3c diminué. Les Figures 5.19 et 5.20 montrent l'effet des incertitudes systématiques de générateur, d'hadronisation, et de radiation sur la distribution du poids du BDT boosté respectivement pour les processus $t\bar{t} + \geq 1b$ et $t\bar{t} + \geq 1c$.

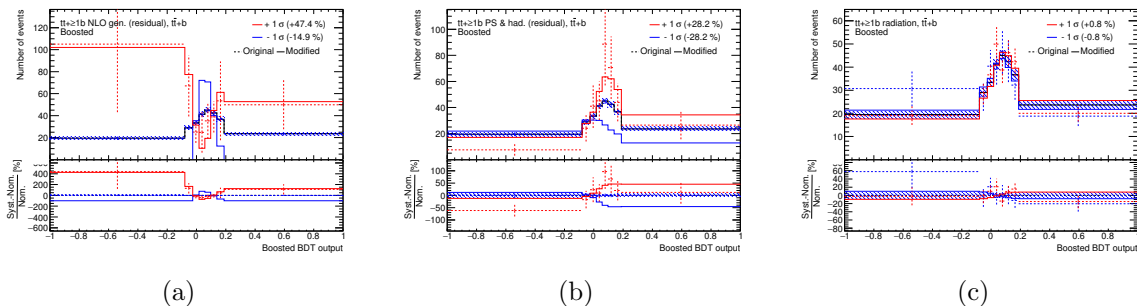


FIGURE 5.19: Distributions du poids du BDT boosté pour le processus $t\bar{t} + \geq 1b$, avec sa variation $+1\sigma$, -1σ due à son incertitude sur le générateur (a), sur l'hadronisation (b) et sur le niveau de radiations (c). Les distributions sont normalisées à 36.5/fb.

Les incertitudes actuelles sur la modélisation du bruit de fond $t\bar{t}$ ne tiennent pas suffisamment compte de notre méconnaissance des fractions relatives de $t\bar{t} + \text{léger}$, $t\bar{t} + c$ et $t\bar{t} + b$. Celle-ci affecte notamment la section efficace $t\bar{t} + b$, dominante dans les régions les plus sensibles de l'analyse, et par conséquent affecte potentiellement fortement la précision de la mesure de la section efficace $t\bar{t}H$. Les prédictions MC, actuellement au NLO, sont

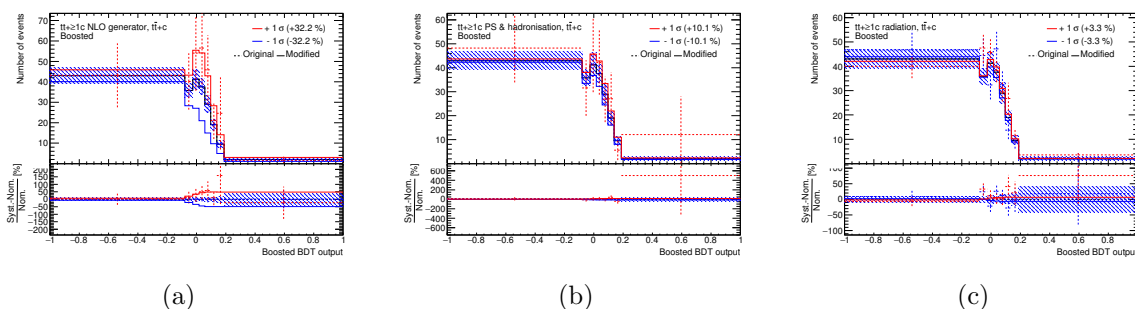


FIGURE 5.20: Distributions du poids du BDT boosté pour le processus $t\bar{t} + \geq 1c$, avec sa variation $+1\sigma, -1\sigma$ due à son incertitude sur le générateur (a), sur l’hadronisation (b) et sur le niveau de radiations (c). Les distributions sont normalisées à 36.5/fb.

insuffisantes pour décrire les événements $t\bar{t}$ quiradient des paires $c\bar{c}, b\bar{b}$ de masse comparable avec celle du boson de Higgs. Pour palier à ce problème, la normalisation des bruits de fond $t\bar{t} + c$ et $t\bar{t} + b$ est ajustée dans les données en même temps que le signal $t\bar{t}H$. Par conséquent, lorsque différents échantillons MC sont comparés pour estimer les incertitudes de modélisation, les fractions de $t\bar{t} + c$ et $t\bar{t} + b$ de tous les échantillons sont pondérées à une même valeur de référence, celle de l’échantillon nominal Powheg+Pythia6.

Incertitudes sur le bruit de fond $W/Z + jets$

Une incertitude de $\pm 30\%$ est appliquée sur la section efficace de production $W/Z + jets$, provenant de la variation des échelles de normalisation et de factorisation et des paramètres d’imbrication entre l’élément de matrice et l’hadronisation du générateur Sherpa. L’incertitude dépend de la saveur des jets radiés. Elle est plus élevée pour les radiations de quarks c et b .

Incertitudes sur le bruit de fond Diboson

Une incertitude de $\pm 50\%$ est associée à la normalisation du processus *diboson*, lié à la production de jets additionnels.

Incertitudes sur le bruit de fond top unique

Une incertitude de $\pm 30\%$ est appliquée sur la normalisation des processus de *top unique*.

Incertitudes sur le bruit de fond $t\bar{t} + V$

Une incertitude de $\pm 10\%$ liée au calcul NLO de la section efficace est appliquée.

Incertitudes sur le signal $t\bar{t}H$

Une incertitude de $\pm 10\%$ est assignée sur la normalisation du signal liée au calcul NLO de la section efficace de production $t\bar{t}H$. Comme pour le processus $t\bar{t}$, une incertitude sur le modèle d’hadronisation est considérée. Elle est évaluée en comparant l’échantillon nominal aMC@NLO+Pythia8 avec l’échantillon aMC@NLO+Herwig++ (Figure 5.21).

5.6.2 Objets reconstruits

L’incertitude sur la mesure de la section efficace $t\bar{t}H$ provient non seulement de la modélisation des collisions pp , mais aussi de la description de l’interaction des particules

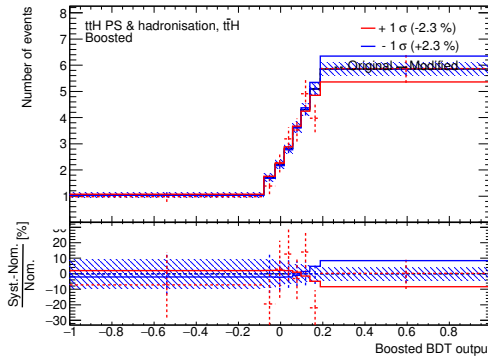


FIGURE 5.21: Distributions du poids du BDT boosté pour le processus $t\bar{t}H$, avec sa variation $+1\sigma, -1\sigma$. due à l'incertitude sur le modèle d'hadronisation. Les distributions sont normalisées à 36.5/fb.

de l'état final avec le détecteur ATLAS, ce qui se traduit par une incertitude sur les objets reconstruits qui sont considérés dans l'analyse. L'analyse est particulièrement altérée par les incertitudes liées aux jets et aux candidats b , car les événements des régions de signal possèdent ces deux objets en grand nombre. Les incertitudes associées aux candidats électron ou aux candidats muon ont un effet négligeable. C'est également le cas des incertitudes liées aux grands jets.

Incertainces liées aux jets

L'énergie des jets ainsi que leur efficacité de reconstruction sont des sources d'incertitude pour l'analyse $t\bar{t}H(b\bar{b})$. L'incertitude sur la calibration et sur la résolution en énergie des jets (JES et JER respectivement) est prise en compte. L'incertitude sur la JES est dérivée de l'information combinée de tests en faisceau, de collisions pp du LHC et de simulations. Elle s'élève à $\pm 5.5\%$ pour des jets de 25 GeV d'impulsion transverse et décroît rapidement. Elle est inférieure à $\pm 1.5\%$ pour des jets centraux d'impulsion transverse comprise entre 100 GeV et 1.5 TeV. La JER a été mesurée dans les données du Run 1 avec des événements *di-jet*. Elle est paramétrée en fonction de l'impulsion transverse et de la pseudo-rapacité des jets. Données et simulations sont en accord à mieux que 10%. Une source d'incertitude supplémentaire liée au passage des conditions du Run 1 à celles du Run 2 dans les simulations est ajoutée à celle mesurée au Run 1.

Incertainces liées aux candidats b

Les efficacités d'identifier un jet comme candidat b sont calibrées dans les données [91] pour chacune des saveurs, léger, c et b en fonction de l'impulsion transverse et la pseudo-rapacité du jet. Les quatre points de fonctionnement (85%, 77%, 70%, 60%) sont calibrés, et les corrélations entre eux sont prises en compte de manière à pouvoir utiliser plusieurs points de fonctionnement dans l'analyse.

Incertainces liées aux leptons

Les incertitudes associées aux leptons sont, comme pour les jets, liées à l'efficacité d'identification, à leur calibration en énergie et leur résolution. Elles sont négligeables par rapport

aux autres sources d'incertitude et sont donc négligées dans l'ajustement final.

Incertitudes liées aux grands jets

Les incertitudes associées aux grands jets proviennent de la calibration de la masse des grands jets (JMS : Jet Mass Scale), de l'énergie des grands jets (JES) et de la variable τ_{32} dans des données enrichies en événements *di-jet*. Ces incertitudes ont très peu d'effet sur les résultats de l'analyse. C'est pourquoi les incertitudes de ces variables, et des autres variables de sous-structures utilisées en entrée du BDT sont également négligées dans l'ajustement final.

5.6.3 Luminosité

L'incertitude sur la luminosité pour les données 2015 et 2016 est fixée à 3.4%.

5.7 Sensibilité du canal top et Higgs boostés dans l'analyse $t\bar{t}H(b\bar{b})$

5.7.1 Régions de controle

Les prédictions Monte Carlo, en particulier pour les variables de sous-structures des grands jets, sont comparées aux données dans une région de contrôle. Cette région est proche de la sélection Higgs- top , car elle contient également un candidat électron ou muon, un candidat top sélectionné comme dans la région de signal, mais un candidat Higgs plus lâche avec moins de 2 candidats b à l'intérieur. La Figure 5.22 montre les variables liées au candidat Higgs, et la Figure 5.23 celle liées au candidat Higgs top .

On remarque tout d'abord un excès de l'ordre de 10% dans les données. Il faut toutefois noter que les prédictions MC montrées ici sont celles avant ajustement, donc pour les fractions de $t\bar{t} + b$ et $t\bar{t} + c$ prédites, et non pas mesurées. Cet excès indique donc que les fractions ajustées seront légèrement supérieures à celles prédites.

La forme des distributions dans les données est raisonnablement reproduite. On note un déficit significatif à haute impulsion transverse du candidat Higgs, mais seulement au-delà de 500 GeV, nettement au-delà de la région accessible actuellement. On note également un certain désaccord dans les variables utilisant MV2c10. La calibration de l'étiquetage des jets de b qui a été utilisée ici a été améliorée depuis, en particulier la calibration des jets de c , qui est maintenant effectuée à partir des désintégrations $W \rightarrow cs$ dans les événements $t\bar{t}$. On peut penser que cette nouvelle calibration améliorerait nettement l'accord data-MC de ces variables.

5.7.2 Limite et force du signal avec 36.5/fb

L'analyse $t\bar{t}H$ a pour but de mesurer la section efficace du processus de production $t\bar{t}H$ normalisée à la section efficace prédite par le modèle standard, appelée force du signal $\mu_{t\bar{t}H}$. Les prédictions de l'analyse dépendent de paramètres a priori inconnus, avec d'une part le

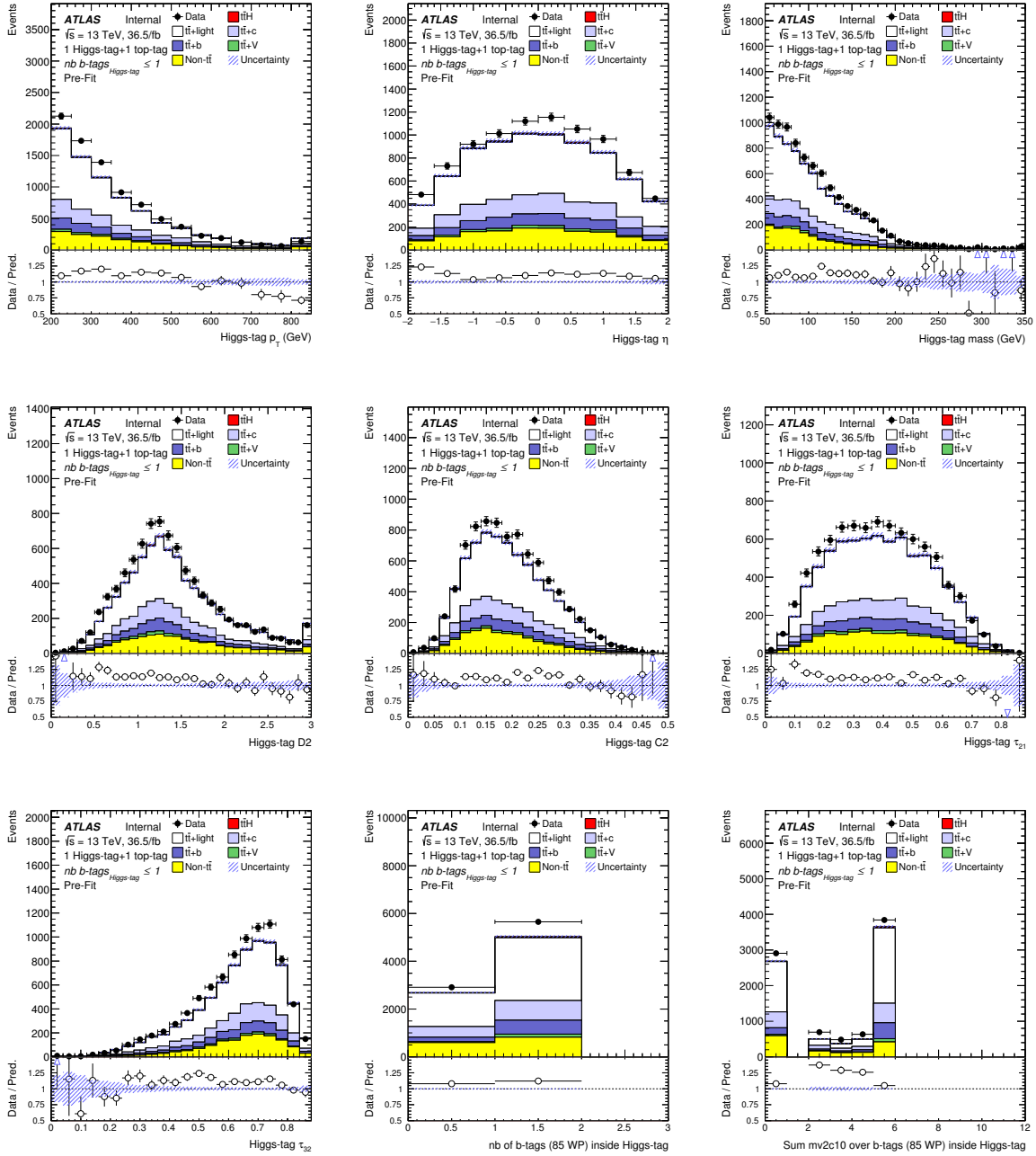


FIGURE 5.22: Distributions des variables liées au candidat Higgs pour la région de contrôle de la sélection Higgs-*top*. Le candidat Higgs contient moins de 2 candidats *b*. Les incertitudes des prédictions viennent uniquement de la statistique des événements Monte Carlo. Les distributions sont normalisées à 36.5/fb.

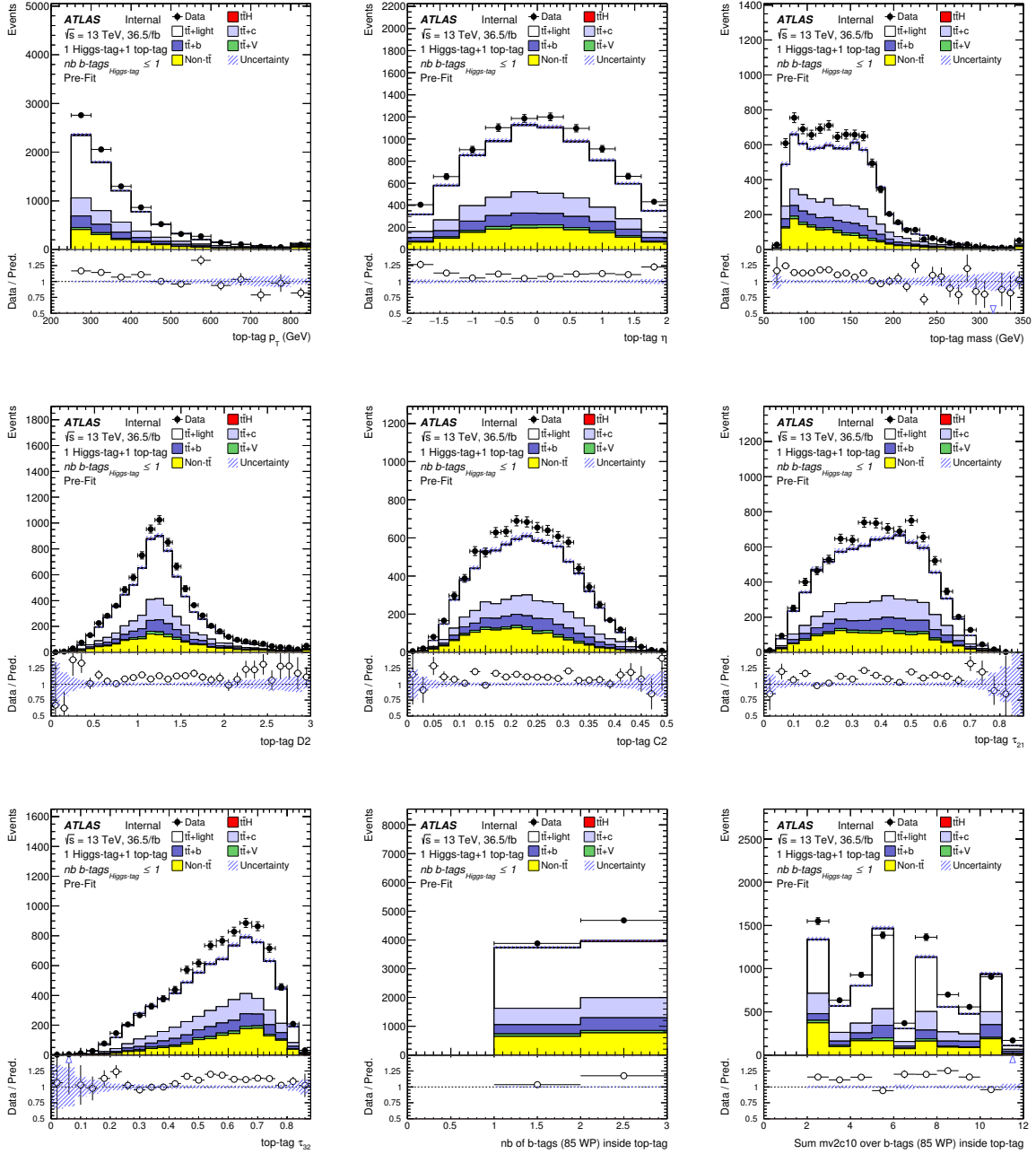


FIGURE 5.23: Distributions des variables liées au candidat top pour la région de contrôle de la sélection Higgs-*top*. Les incertitudes des prédictions viennent uniquement de la statistique des événements Monte Carlo. Le candidat Higgs contient moins de 2 candidats *b*. Les distributions sont normalisées à 36.5/fb.

paramètre d'intérêt, ici la force du signal $\mu_{t\bar{t}H}$, et d'autre part les paramètres de nuisance θ , comme les incertitudes systématiques, qui détériorent la précision de la mesure de $\mu_{t\bar{t}H}$. Ces paramètres sont mesurés dans une procédure d'ajustement (fit) des prédictions aux observations.

Les distributions ajustées dans le canal *lepton + jets* de l'analyse $t\bar{t}H(b\bar{b})$ sont le poids de BDT de classification dans les régions de signal, et H_T^{had} dans les régions de contrôle. Pour chaque région de signal, un BDT est entraîné spécifiquement dans cette même région pour séparer le signal du bruit de fond. Les intervalles de ces distributions les plus riches en signal influencent davantage la mesure de $\mu_{t\bar{t}H}$. Pour chaque région de contrôle, c'est la variable H_T^{had} qui est utilisé plutôt que le BDT. Bien que cette variable ne soit pas très discriminante, la forme de sa distribution est utilisée dans le fit pour contraindre les paramètres de nuisance.

La vraisemblance de la distribution formée des distributions de H_T^{had} dans les régions de contrôle et du poids BDT de classification dans les régions de signal, est utilisée dans la procédure de fit. La vraisemblance s'exprime comme le produit sur les intervalles de la distributions des termes de Poisson

$$L(\mu, \theta) = \prod_{i \text{ bins}} P(N_{obs}^i | N^i(\mu, \theta)) \quad (5.1)$$

où N_{obs}^i , respectivement $N^i(\mu, \theta)$, est le nombre d'événements dans un bin i des distributions observées, respectivement attendues. Le nombre d'événements observés peut provenir d'une simulation, générée avec une valeur particulière de force du signal μ' . Enfin, pour tester la compatibilité d'une observation avec l'hypothèse d'une force de signal $\mu_{t\bar{t}H}$, le rapport de vraisemblance est considéré

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})}. \quad (5.2)$$

Ici $\hat{\theta}$ au numérateur indique la valeur de θ qui maximise la vraisemblance $L(\mu, \theta)$ pour une valeur spécifique de $\mu_{t\bar{t}H}$. Le dénominateur est quant à lui la vraisemblance maximisée, *i.e* les $\hat{\mu}$ et $\hat{\theta}$ sont les estimateurs de maximum de vraisemblance.

Limite sur $\mu_{t\bar{t}H}$

Les résultats de l'analyse sont quantifiés par la limite supérieure sur la force du signal notée μ_L [92]. Le calcul de limite permet d'exclure l'hypothèse d'un signal de force supérieure à μ_L dans un échantillon testé avec un niveau de confiance de 95%. L'échantillon testé peut provenir d'une simulation contenant une force quelconque de signal ou des données observées. La limite μ_L qui servira de point de comparaison pour les études suivantes est calculée pour un échantillon de test simulé sans signal ($\mu = 0$).

Pour établir la limite supérieure, le test statistique suivant est défini

$$q_\mu = \begin{cases} -2 \log \lambda(\mu) & \text{if } \mu \geq \hat{\mu} \\ 0 & \text{if } \mu < \hat{\mu} \end{cases} \quad (5.3)$$

où $\lambda(\mu)$ est le rapport de vraisemblance profilé défini dans l'Eq. 5.2. Le niveau d'accord entre l'hypothèse d'une force de signal $\mu_{t\bar{t}H}$ et l'échantillon testé de force du signal μ' , de tests q_μ et $q_{test,\mu'}$ respectivement est donnée par la valeur- p

$$p_\mu = \int_{q_{test,\mu'}}^{\infty} f(q_\mu|\mu) dq_\mu. \quad (5.4)$$

Ici $f(q_\mu|\mu)$ est la densité de probabilité de q_μ en supposant l'hypothèse $\mu_{t\bar{t}H}$. Pour le calcul de la limite, le $q_{test,\mu'}$ est obtenu en prenant la médiane du test q_μ en supposant un signal nul $\mu = 0$. Finalement la limite μ_L est la valeur à partir de laquelle la valeur- p p_{μ_L} devient inférieure à 5%.

Résultats avec seulement l'incertitude statistique

L'analyse boostée est combinée avec l'analyse résolue en ajoutant la région boostée aux régions de l'analyse résolue. La sensibilité de cette analyse, appelée combinée, est comparée avec l'analyse résolue en comparant les limites μ_L (Table 5.14). Elle est obtenue dans le cas simplifié où seules la force du signal $t\bar{t}H$ $\mu_{t\bar{t}H}$ et les fractions $t\bar{t} + c$ et $t\bar{t} + b$, notées $k_{t\bar{t}+c}$ et $k_{t\bar{t}+b}$, sont ajustées. Les incertitudes systématiques sont fixées à leur valeur nominale, sauf pour l'incertitude statistique des Monte Carlo.

Deux analyses combinées, qui diffèrent par leur distribution de BDT, sont considérées. Les deux distributions de BDT sont obtenues sans et avec le poids du BDT de classification de l'analyse résolue en entrée du BDT de classification de la région boostée, toutes variables étant par ailleurs égales. La première configuration, avec une limite de 0.6165, est à peine plus sensible que l'analyse résolue, qui exclue une force de signal $\mu_{t\bar{t}H}$ supérieure à 0.6204, avec un gain de 0.6%. Pour la seconde configuration, le BDT de la région boostée utilise l'information de l'analyse résolue pour séparer le signal du bruit de fond en plus des variables boostées. L'analyse combinée apporte une amélioration d'environ 3% par rapport à l'analyse résolue, avec une limite à 0.6026. Cette amélioration montre qu'inclure l'analyse boostée dans l'analyse résolue apporte de l'information pour discriminer le signal du bruit de fond, ce qui se traduit par une amélioration de la sensibilité.

fit regions	-1σ	median	$+1\sigma$	gain / R
R only	0.4470	0.6204	0.8635	0
R + B	0.4442	0.6165	0.8583	0.6%
R + B (with resolved BDT)	0.4342	0.6026	0.8399	2.9%

TABLE 5.14: Limite supérieure sur la force du signal $\mu_{t\bar{t}H}$ (95% CL), avec une luminosité de 36.5/fb, pour l'analyse résolue (R only), et pour l'analyse combinée (R + B). Le BDT de la région boostée de la seconde analyse combinée est entraîné avec le poids du BDT de classification de l'analyse résolue en plus des variables boostées utilisées dans la première analyse combinée. Seules les incertitudes statistiques sont considérées dans le fit

5.7.3 Résultat incluant les incertitudes systématiques

Avec l’ajustement des paramètres de nuisance, certains paramètres sont contraints par les données, c’est-à-dire que leur incertitude systématique après ajustement est inférieure à celle avant ajustement. La Figure 5.24 montre la valeur et l’incertitude (normalisés à 1) après ajustement de tous les paramètres de nuisance ainsi que des trois paramètres d’intérêt $\mu_{t\bar{t}H}$, $k_{t\bar{t}+c}$ et $k_{t\bar{t}+b}$. Ici, les données observées sont remplacées par des données qui correspondent en tout point à la valeur attendue du bruit de fond et du signal (avec $\mu_{t\bar{t}H} = 1$). Un ajustement sur de telles données se nomme “ajustement Asimov”. Par définition, les valeurs ajustées sont donc exactement égales aux valeurs attendues. Par contre, les incertitudes sont affectées par l’ajustement. Les niveaux de contraintes pour les incertitudes présentes dans le fit de l’analyse résolue et de l’analyse combinée sont similaires. Parmi les sources d’incertitude systématique, les incertitudes sur la modélisation du bruit de fond $t\bar{t}$ sont les plus contraintes, en particulier celles de la composante $t\bar{t} + \text{léger}$. Leur niveau de contrainte est au dessus de 80%, notamment grâce aux régions de contrôle qui contiennent un nombre important d’événements $t\bar{t}$. Des contraintes non nulles mais en dessous des 50% sont obtenues pour des incertitudes systématiques liées à l’échelle d’énergie des jets ou à la résolution des jets, aux efficacités d’identification des jets de quark b , aux sections efficaces comme celles de $t\bar{t}$, *top unique* ou encore $W + \text{jets}$, et enfin à l’impulsion transverse des grands jets pour l’analyse combinée.

Les quinze premières sources d’incertitude, classées selon leur impact sur $\mu_{t\bar{t}H}$ après fit sont présentées Figure 5.25. Les incertitudes systématiques liées à la modélisation de l’hadronisation du processus $t\bar{t}H$ dominant, suivies d’une incertitude sur l’efficacité de ‘mauvaise’ identification des jets légers comme jets de quark b , puis de la section efficace de production $t\bar{t}H$. Les incertitudes sur la modélisation de $t\bar{t} + b$ sont également présentes dans ce classement malgré leur important niveau de contrainte, avec un impact sur $\mu_{t\bar{t}H}$ très différent avant et après le fit, probablement en partie due au manque de statistique Monte Carlo disponible pour estimer ces incertitudes.

Les incertitudes systématiques détériorent la sensibilité sur le paramètre d’intérêt $\mu_{t\bar{t}H}$. Elles conduisent à une augmentation de la limite supérieure sur $\mu_{t\bar{t}H}$ de 40 % pour l’analyse résolue, passant de 0.607 à 1.003 (Table 5.15). Ajouter la région boostée¹ à l’analyse résolue améliore la sensibilité (ici les variables d’entrée du BDT boosté ne contiennent pas le poids du BDT résolu). La limite supérieure μ_L de l’analyse combinée, est 0.865, soit 16% plus faible que celle de l’analyse résolue, ce qui représente une amélioration nettement plus importante que celle obtenue sans ajouter les incertitudes systématiques (< 1%). Un second test de l’impact des événements boostés consiste à comparer la sensibilité de l’analyse ‘R – B’, obtenue en retirant les événements boostés de l’analyse résolue, avec la sensibilité de l’analyse résolue. La sensibilité de l’analyse ‘R – B’ est détériorée en ne considérant que l’incertitude systématique statistique (0.633 contre 0.607) mais est améliorée en considérant

1. La limite μ_L combinée R+B en ne considérant que l’incertitude systématique liée à la statistique des événements Monte Carlo présentée Table 5.15 est légèrement différente de celle de la Table 5.14 à cause d’une réoptimisation du binning des distributions fittées.

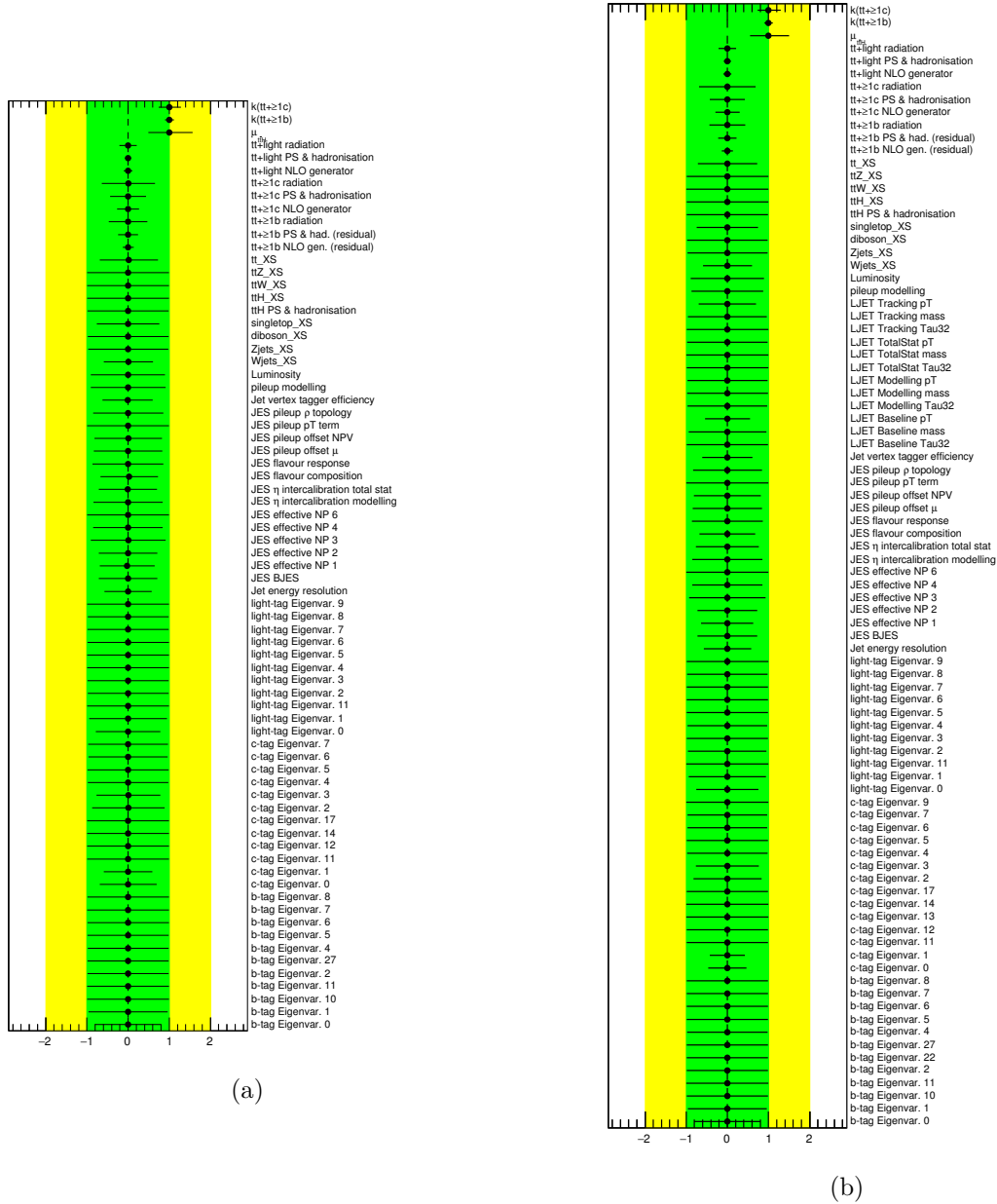


FIGURE 5.24: Contraintes sur les paramètres ajustés dans les données Asimov pour l’analyse résolue (gauche) et l’analyse combinée (droite). L’incertitude d’un paramètre ajusté est représenté par un trait noir. Pour les sources d’incertitude systématique, cette incertitude après fit est comparée à celle avant fit de ± 1 (bande verte).

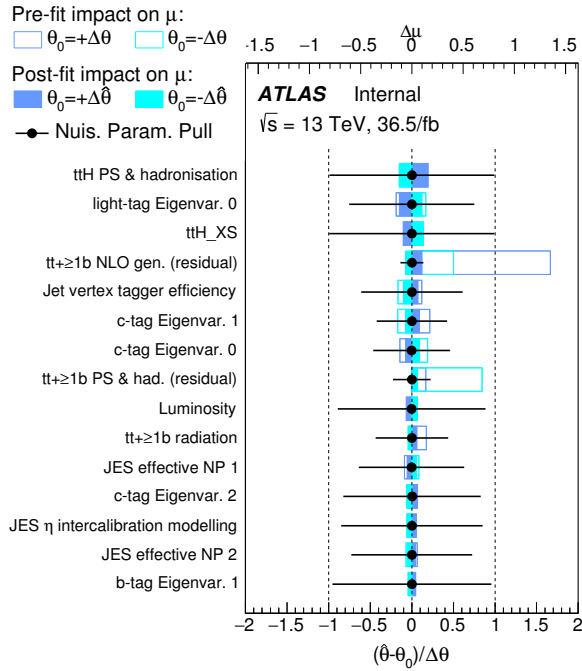


FIGURE 5.25: Les 15 systématiques qui ont le plus d’impact sur la force du signal pour 36.5/fb.

l’ensemble des incertitudes systématiques (0.920 contre 1.003). L’analyse résolue est donc très affectée par les incertitudes systématiques pour les événements passant la sélection boostées, et les enlever de la sélection de l’analyse résolue améliore sa sensibilité. Une grande partie de l’amélioration de l’analyse combinée vient du traitement particulier des événements boostés au sein d’une région spécifique, c’est-à-dire indépendamment des autres événements de l’analyse résolue. Les incertitudes systématiques dépendent en effet du caractère boosté des événements. Par exemple, les incertitudes de modélisation du bruit de fond $t\bar{t}$ augmente avec l’impulsion transverse du quark top , et les incertitudes d’identification des jets de quark b deviennent plus importantes à haute impulsion de jet. C’est d’ailleurs en incluant ces deux types d’incertitude systématique que l’écart de sensibilité se creuse entre l’analyse combinée et l’analyse résolue (Table 5.16). Toutefois, on observe probablement ici un effet du manque de statistique dans les échantillons utilisés, en particulier l’échantillon aMC@NLO utilisé pour estimer l’incertitude sur le générateur.

5.7.4 Perspectives à haute luminosité

Avec la luminosité actuelle du Run 2, l’analyse boostée est limitée par la statistique des événements sélectionnés. Le gain en sensibilité lié à l’ajout de l’analyse boostée dans l’analyse résolue devrait augmenter avec la luminosité.

Pour tester cette hypothèse, la limite sur $\mu_{t\bar{t}H}$ de l’analyse résolue est calculée avec une plus grande luminosité. Puis elle est comparée à celle obtenue quand elle est combinée avec

fit regions	stat. only	with syst.	wo LJET syst.
R-B	0.633	0.920	0.919
R+B	0.606	0.865	0.863
R only	0.607	1.003	N/A

TABLE 5.15: Limite supérieure μ_L sur la force du signal (niveau de confiance de 95%) à $\mathcal{L} = 36.5/\text{fb}$.

systematics	R only	R+B	R/R+B - 1 (%)
MC statistics	0.61	0.61	0.3
incl. cross sections	0.63	0.63	0.2
incl. pile-up & luminosity	0.63	0.63	0.1
incl. JVT	0.63	0.63	0.2
incl. $t\bar{t}$ Generation	0.74	0.72	3.7
incl. $t\bar{t}$ Parton Shower	0.79	0.75	5.5
incl. $t\bar{t}$ Radiations	0.81	0.76	7.5
incl. $t\bar{t}H$ Parton Shower	0.84	0.79	7.1
incl. b -tagging	0.98	0.84	16.1
incl. small-R jet (JES,JER)	1.00	0.86	16.2
incl. large-R jet (JES,JMS, τ_{32})	1.00	0.87	16.0

TABLE 5.16: Evolution de la limite résolue et combinée à $\mathcal{L} = 36.5/\text{fb}$ obtenue en ajoutant des sources d'incertitude systématiques les unes après les autres, regroupées en un ensemble des onze catégories.

l'analyse boostée (Table 5.17). Passer de 36.5/fb à 300/fb permet de diminuer la limite sur $\mu_{t\bar{t}H}$ de presque deux pour l'analyse résolue. L'amélioration apportée par l'analyse boostée décroît légèrement avec la luminosité, au lieu de montrer une augmentation comme attendu.

Plusieurs causes peuvent être à l'origine de ce comportement. La première est que l'analyse boostée actuelle n'est pas en mesure d'exploiter les topologies boostées pour lesquelles les partons de désintégrations du quark *top* hadronique ou du boson de Higgs sont séparés d'une distance $\Delta R < 0.4$. L'utilisation d'un étiquetage des jets de *b* standard, appliqué à des jets standards de rayon 0.4, devrait certainement être améliorée en particulier pour étiqueter le candidat Higgs. D'autre part dans l'analyse actuelle aucun effort n'a été fait pour optimiser la sensibilité à haute impulsion : coupures en impulsion transverse, entraînement du BDT, etc...

Luminosity	R only	R+B	R/(R+B) - 1 (%)
36.5	1.00	0.87	16.0
100	0.72	0.62	15.4
300	0.56	0.48	15.9

TABLE 5.17: Limite sur $\mu_{t\bar{t}H}$ pour l'analyse résolue et combinée pour différentes luminosités intégrées, et amélioration apportée par l'analyse boostée.

5.7.5 Conclusion

L'étude des topologies $t\bar{t}H(b\bar{b})$ boostées présentée dans ce chapitre a permis d'améliorer l'étiquetage des quarks *top* hadroniques boostés dans les événements $t\bar{t}H$. Elle a aussi permis de mettre en place une analyse spécifique d'événements contenant deux grands jets, utilisée pour capturer les produits des désintégrations des quarks *top* hadroniques et des bosons de Higgs modérément boostés du signal $t\bar{t}H$. L'analyse boostée à deux grands jets est combinée avec l'analyse résolue afin d'améliorer la sensibilité globale de la recherche du processus $t\bar{t}H$ prédit par le modèle standard. Les estimations pour une luminosité de 36.5/fb montrent que l'analyse boostée n'apporte qu'une faible amélioration de la sensibilité à l'analyse résolue avec un gain de 2% sur la limite sur $\mu_{t\bar{t}H}$, en considérant uniquement les incertitudes statistiques. En ajoutant les incertitudes systématiques, la combinaison des analyses résolue et boostée semble apporter un gain notable sur l'analyse résolue seule, d'environ 15% sur la limite sur $\mu_{t\bar{t}H}$. Cependant, ce résultat doit être interprété avec précaution, car la majeure partie de l'amélioration ne vient pas de l'analyse boostée elle-même, mais du fait de traiter les événements boostés indépendamment des autres événements de l'analyse résolue. En effet, retirer les événements boostés de l'analyse résolue semble réduire l'impact des incertitudes systématiques, en particulier celles sur la modélisation des événements $t\bar{t} + b$, sur l'analyse résolue. Ceci laisse à penser qu'une partie de cette amélioration est essentiellement due à des fluctuations statistiques dans les échantillons utilisés pour estimer les incertitudes sur le processus $t\bar{t} + b$. De plus, les investigations à haute luminosité,

jusqu'à $\mathcal{L} = 300/\text{fb}$, montrent que l'analyse boostée, dans sa forme actuelle, n'améliore pas plus à l'analyse résolue qu'à la luminosité actuelle.

L'une des difficultés de l'analyse boostée $t\bar{t}H$ réside dans le fait qu'une proportion relativement faible des événements de signal produisent des quarks top et des bosons de Higgs boostés, en comparaison avec des analyses boostées recherchant des résonances de masses de l'ordre du TeV par exemple. Initialement, les coupures sur les impulsions transverses des grands jets ont été fixées à des valeurs suffisamment basses (200-250 GeV) pour permettre à l'analyse boostée de sélectionner un nombre non négligeable d'événements (une dizaine) avec la luminosité actuelle. En contrepartie, lorsque les quarks top hadroniques et bosons de Higgs sont reconstruits au sein de ces grands jets de rayon $R = 1.0$, la séparation angulaire entre les partons de leur désintégration est suffisamment élevée pour qu'une majorité d'entre eux soient reconstruits séparément à l'aide de jets standards de rayon $R = 0.4$. Par exemple, seulement 7% des bosons de Higgs associés aux grands jets d'impulsion transverse supérieure à 250 GeV émettent dans leur désintégration une paire de quarks b séparés de $\Delta R < 0.4$, là où une analyse boostée basée sur des sous-structures pourrait sans ambiguïté apporter de l'information à l'analyse résolue (Figure 5.26). Davantage de luminosité permettra d'augmenter la coupure sur l'impulsion transverse pour se concentrer sur la région réellement boostée.

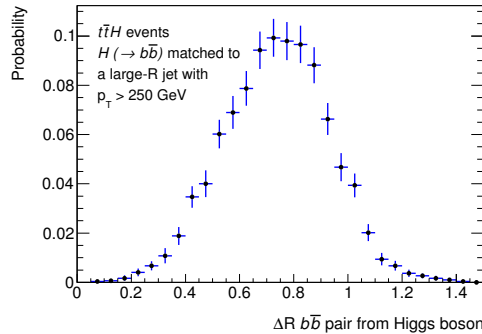


FIGURE 5.26: Distribution normalisée de l'angle ΔR entre les quarks b de la désintégration du boson de Higgs, pour des événements $t\bar{t}H(b\bar{b})$ contenant un grand jet provenant d'un boson de Higgs.

De plus, l'utilisation de sous-jets de très faible rayon, inférieur au rayon standard de 0.4, pour étiqueter les jets de quark b à l'intérieur des grands jets, avec par exemple l'utilisation de jets de traces de rayon $R = 0.2$, permettrait de résoudre les topologies non accessibles par l'analyse résolue [93]. Il faudrait pour cela porter une attention particulière à l'optimisation de l'étiquetage des jets de quark b sur ces sous-jets, pour obtenir le même niveau de performance que celle atteinte sur les jets standards. Enfin, on pourrait imaginer utiliser dans l'analyse boostée l'information obtenue par l'analyse résolue pour séparer les événements de signal et de bruit de fond. Cette méthode a été considérée dans cette étude et pourrait être testée avec une analyse boostée utilisant des sous-jets de très faible rayon pour étiqueter les

jets de quark b . Cela permettrait de combiner de manière optimale l'information boostée lorsque les produits de désintégration du quark top hadronique ou du boson de Higgs se chevauchent, et l'information résolue dans le cas contraire.

Bibliographie

- [60] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462 :152–155, 2001.
- [61] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 7 :79, July 2014.
- [62] T. Sjöstrand, S. Mrenna, and P. Skands. A brief introduction to PYTHIA 8.1. *Computer Physics Communications*, 178 :852–867, June 2008.
- [63] ATLAS Run 1 Pythia8 tunes. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014.
- [64] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali. Parton distributions for the LHC run II. *Journal of High Energy Physics*, 4 :40, April 2015.
- [65] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *Journal of High Energy Physics*, 3 :15, March 2013.
- [66] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY : a program for Higgs boson decays in the Standard Model and its supersymmetric extension. *Computer Physics Communications*, 108 :56–74, January 1998.
- [67] W. Beenakker, S. Dittmaier, M. Krämer, B. Plümper, M. Spira, and P. M. Zerwas. NLO QCD corrections to $t\bar{t}H$ production in hadron collisions. *Nuclear Physics B*, 653 :151–203, March 2003.
- [68] S. Dawson, C. Jackson, L. H. Orr, L. Reina, and D. Wackerth. Associated Higgs boson production with top quarks at the CERN Large Hadron Collider : NLO QCD corrections. *Physical Review D*, 68(3) :034022, August 2003.
- [69] Z. Yu, M. Wen-Gan, Z. Ren-You, C. Chong, and G. Lei. QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider. *ArXiv e-prints*, July 2014.
- [70] S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro. Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons. *Journal of High Energy Physics*, 6 :184, June 2015.

- [71] S. Alioli, P. Nason, C. Oleari, and E. Re. A general framework for implementing NLO calculations in shower Monte Carlo programs : the POWHEG BOX. *Journal of High Energy Physics*, 6 :43, June 2010.
- [72] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05 :026, 2006.
- [73] Pavel M. Nadolsky, Hung-Liang Lai, Qing-Hong Cao, Joey Huston, Jon Pumplin, Daniel Stump, Wu-Ki Tung, and C. P. Yuan. Implications of CTEQ global analysis for collider observables. *Phys. Rev.*, D78 :013004, 2008.
- [74] Peter Zeiler Skands. Tuning Monte Carlo Generators : The Perugia Tunes. *Phys. Rev.*, D82 :074018, 2010.
- [75] Michal Czakon and Alexander Mitov. Top++ : A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, 185 :2930, 2014.
- [76] Matteo Cacciari, Michal Czakon, Michelangelo Mangano, Alexander Mitov, and Paolo Nason. Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation. *Phys. Lett.*, B710 :612–622, 2012.
- [77] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J.*, C71 :1547, 2011.
- [78] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO single-top production matched with shower in POWHEG : s- and t-channel contributions. *JHEP*, 09 :111, 2009. [Erratum : JHEP02,011(2010)].
- [79] Stefano Frixione, Eric Laenen, Patrick Motylinski, Bryan R. Webber, and Chris D. White. Single-top hadroproduction in association with a W boson. *JHEP*, 07 :029, 2008.
- [80] Nikolaos Kidonakis. Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-. *Phys. Rev.*, D82 :054018, 2010.
- [81] Nikolaos Kidonakis. NNLL resummation for s-channel single top quark production. *Phys. Rev.*, D81 :054028, 2010.
- [82] Nikolaos Kidonakis. Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production. *Phys. Rev.*, D83 :091503, 2011.
- [83] Tanju Gleisberg and Stefan Hoeche. Comix, a new matrix element generator. *JHEP*, 12 :039, 2008.
- [84] Steffen Schumann and Frank Krauss. A Parton shower algorithm based on Catani-Seymour dipole factorisation. *JHEP*, 03 :038, 2008.
- [85] Stefan Hoeche, Frank Krauss, Marek Schonherr, and Frank Siegert. QCD matrix elements + parton showers : The NLO case. *JHEP*, 04 :027, 2013.
- [86] Measurement of W and Z Boson Production Cross Sections in pp Collisions at $\sqrt{s} = 13$ TeV in the ATLAS Detector. Technical Report ATLAS-CONF-2015-039, CERN, Geneva, Aug 2015.

- [87] S Artz and al. Upgrade of the ATLAS Central Trigger for LHC Run-2. Technical Report ATL-DAQ-PROC-2014-042. 02, CERN, Geneva, Nov 2014.
- [88] Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2016-080, CERN, Geneva, Aug 2016.
- [89] Boosted hadronic top identification at ATLAS for early 13 TeV data. Technical Report ATL-PHYS-PUB-2015-053, CERN, Geneva, Dec 2015.
- [90] P Speckmayer, A Hocker, J Stelzer, and H Voss. The toolkit for multivariate data analysis : TMVA 4. *J. Phys. : Conf. Ser.*, 219 :032057, 2010.
- [91] Calibration of ATLAS b -tagging algorithms in dense jet environments. Technical Report ATLAS-CONF-2016-001, CERN, Geneva, Feb 2016.
- [92] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *European Physical Journal C*, 71 :1554, February 2011.
- [93] Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector. Technical Report ATL-PHYS-PUB-2014-013, CERN, Geneva, Aug 2014.

Conclusion

La découverte du boson de Higgs au LHC a ouvert de nouvelles voies d'exploration en physique des hautes énergies. Pour les expériences ATLAS et CMS, l'observation de l'interaction entre le boson de Higgs et le quark *top* est actuellement un enjeu majeur. Cette interaction intervient dans les processus $t\bar{t}H$ produits dans les collisions de protons fournies par le LHC. Le canal $t\bar{t}H(b\bar{b})$ *lepton + jets*, provenant d'une paire $t\bar{t}$ se désintégrant de façon semi-leptonique et d'un boson de Higgs se désintégrant en une paire $b\bar{b}$, est l'un des canaux les plus prometteurs pour observer le processus $t\bar{t}H$. Cependant, le signal $t\bar{t}H(b\bar{b})$ tel qu'il est prédit par le modèle standard reste difficile à observer, car il est faible par rapport aux processus de bruit de fond $t\bar{t}$, en particulier par rapport au processus $t\bar{t} + b\bar{b}$, dont la modélisation est une source d'incertitude majeure. La combinaison des analyses $t\bar{t}H(b\bar{b})$, *multi-leptons* et $t\bar{t}H(\gamma\gamma)$ devrait permettre, à la fin du Run 2 avec une luminosité attendue de $\mathcal{L} = 150/\text{fb}$, aux collaborations ATLAS et CMS d'observer indépendamment le signal $t\bar{t}H$ avec une signification statistique proche de 5σ .

Au cours de cette thèse, des méthodes d'identification de quark *top* boosté ainsi qu'un canal boosté $t\bar{t}H(b\bar{b})$ ont été étudiés afin d'améliorer la sensibilité de l'analyse $t\bar{t}H(b\bar{b})$ dans le canal *lepton + jets*. Ce canal boosté se caractérise par la présence à la fois d'un quark *top* et d'un boson de Higgs boostés. Ces études ont été développées à l'aide d'outils dédiés aux topologies boostées, comme les jets de grand rayon et les variables de sous-structures associées. Une analyse multivariée a été réalisée à l'aide des variables boostées, dont les distributions simulées semblent reproduire les données avec justesse. Les estimations montrent que le canal boosté n'améliore que très légèrement l'analyse *lepton + jets* standard si on ne considère que les incertitudes statistiques. Ce premier résultat nous apprend que l'analyse boostée ne fait pas beaucoup mieux que l'analyse résolue pour séparer le signal du bruit de fond pour les événements qu'elle sélectionne, malgré l'utilisation de techniques à la fois boostées et résolues. En prenant en compte les principales sources d'incertitudes systématiques, un gain notable en sensibilité est observé mais le manque de statistique dans la simulation des bruits de fond ne permet pas de conclure de manière définitive. Enfin, les estimations montrent que le gain en sensibilité lié à l'ajout de l'analyse boostée actuelle à l'analyse *lepton + jets* standard n'augmente pas avec la luminosité. Ces conclusions indiquent que l'analyse boostée devrait se focaliser à l'avenir sur des événements bien plus boostés que ceux étudiés dans cette thèse, afin d'utiliser optimalement des événements auxquels l'analyse résolue n'est pas sensible.

Ces travaux ont été accompagnés d'études sur l'étiquetage de jets de quarks *b* avec une étude sur la sélection des traces utilisées par JetFitter pour reconstruire les vertex secondaires à haute impulsion de jet. Une sélection basée sur l'impulsion des traces afin de rejeter les traces ne provenant pas des désintégrations de hadron *b*, montre une amélioration limitée des performances d'identification de jet de *b* et indique la nécessité de considérer des techniques plus sophistiquées. Les variables calculées à partir des propriétés des vertex secondaires reconstruits par JetFitter ont également été étudiées, permettant notamment d'améliorer l'identification des jets de quarks *c*. Enfin, le développement de l'algorithme multivarié *MV2*, qui combine l'information des algorithmes de reconstruction de vertex

secondaires (JetFitter et SV1) et utilisant les paramètres d'impact des traces (IPxD), a aboutit à l'amélioration notable des performances d'identification de jets de quarks b .

Remerciements

J'aimerais remercier trois personnes qui ont beaucoup compté pour moi durant cette belle aventure de la thèse : mes superviseurs Henri Bachacou et Frédéric Déliot, et ma copine Laura Pacheco.

Tout d'abord, Frédéric, pour m'avoir fait confiance dès le début, pour m'avoir accompagné et fait voyager (au sens propre comme au figuré) dans le monde de la physique des particules. Ensuite, Henri, que j'ai vraiment découvert lors de mes séjours au Cern à Genève. Ce fut l'une des plus belles collaborations de ma vie. Henri m'a révélé les concepts simples derrière la complexité des recherches expérimentales de physique des particules, et m'a aidé (malgré mes réticences) à me focaliser sur ce qui est important dans un travail de recherche. Les conseils justes de mes superviseurs m'ont été d'une grande utilité pendant cette thèse et m'accompagneront j'en suis sûr tout au long de ma vie. Enfin, j'ai rencontré une personne magnifique au CEA, Laura, avec qui j'ai partagé des moments passionnants de physique des particules, et qui m'a toujours soutenu.

Merci à vous trois.