N° d'ordre NNT : 2021LYSE1118

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale N° 52
École Doctorale de Physique et d'Astrophysique

Spécialité de doctorat : Physique des particules

Soutenue publiquement le 8 juillet 2021, par :

Lucas TORTEROTOT

Recherche de bosons de Higgs supplémentaires de haute masse se désintégrant en paire de taus dans l'expérience CMS au LHC à l'aide du *machine learning*

CERN-THESIS-2021-088
08/07/2021

Devant le jury composé de :

M^me^	Anne-Catherine	LE BIHAN	Chargée de recherche	CNRS	IPHC	Rapporteuse
M	David	ROUSSEAU	Directeur de recherche	CNRS	IJCLab	Rapporteur
M^me^	Corinne	AUGIER	Professeure	UCBL	IP2I	Examinatrice
M^me^	Lucia	DI CIACCIO	Professeure	USMB	LAPP	Examinatrice
M	Steve	MUANZA	Directeur de recherche	CNRS	CPPM	Examinateur
M	Roger	WOLF	Privat-docent	KIT	ETP	Examinateur
M	Colin	BERNET	Chargé de recherche	CNRS	IP2I	Directeur de thèse

MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

ÉCOLE DOCTORALE 52

PHAST
PHYSIQUE
ET ASTROPHYSIQUE
UNIVERSITÉ DE LYON

UNIVERSITÉ DE LYON

Lyon 1

N° d'ordre NNT : 2021LYSE1118

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale N° 52
École Doctorale de Physique et d'Astrophysique

Spécialité de doctorat : Physique des particules

Soutenue publiquement le 8 juillet 2021, par :

Lucas TORTEROTOT

Recherche de bosons de Higgs supplémentaires de haute masse se désintégrant en paire de taus dans l'expérience CMS au LHC à l'aide du *machine learning*

Devant le jury composé de :

M^me	Anne-Catherine	LE BIHAN	Chargée de recherche	CNRS	IPHC	Rapporteuse
M	David	ROUSSEAU	Directeur de recherche	CNRS	IJCLab	Rapporteur
M^me	Corinne	AUGIER	Professeure	UCBL	IP2I	Examinatrice
M^me	Lucia	DI CIACCIO	Professeure	USMB	LAPP	Examinatrice
M	Steve	MUANZA	Directeur de recherche	CNRS	CPPM	Examinateur
M	Roger	WOLF	Privat-docent	KIT	ETP	Examinateur
M	Colin	BERNET	Chargé de recherche	CNRS	IP2I	Directeur de thèse

cnrs dépasser les frontières IN2P3 Les deux infinis

CERN

CMS Compact Muon Solenoid

iP2i LES 2 INFINIS LYON

Version du 25 juin 2021

Résumé

Recherche de bosons de Higgs supplémentaires de haute masse se désintégrant en paire de taus dans l'expérience CMS au LHC à l'aide du *machine learning*

Malgré plusieurs décennies de prédictions expérimentalement vérifiées, le modèle standard (SM) souffre de lacunes que des théories allant au-delà (BSM, *Beyond SM*) tentent de combler. L'une d'entre elles, l'extension supersymétrique minimale du modèle standard (MSSM), prédit l'existence de cinq bosons de Higgs dont trois neutres. Le plus léger, h, doit correspondre au boson découvert en 2012. Les deux autres, H et A, sont de haute masse et additionnels par rapport au SM. Leur phénoménologie au LHC motive l'étude des événements avec une paire de leptons τ.

Dans ce contexte, les événements récoltés par la collaboration CMS de 2016 à 2018 lors des collisions de protons du LHC avec une énergie dans le centre de masse de 13 TeV, correspondant à une luminosité intégrée de 137 fb^{-1}, sont exploités dans cette thèse.

Les jets sont des objets physiques omniprésents lors des collisions au LHC. Leur calibration à l'aide d'événements contenant un photon et un jet (γ + jet) est présentée ainsi que les résultats obtenus pour l'année 2018. Cette étude est directement utilisée dans la calibration officielle de la collaboration CMS.

Aucun excès significatif correspondant aux bosons H et A par rapport aux bruits de fond attendus n'est observé dans l'analyse des événements avec une paire de τ. Des limites d'exclusion sur le produit de la section efficace de production de H et A avec leur rapport de branchement aux τ sont donc données en fonction de leur masse pour les modes de production par fusion de gluon ou en association avec des quarks b. Ces limites sont comprises entre 15 pb à 110 GeV et 3×10^{-4} pb à 3,2 TeV pour la fusion de gluons et entre 1,2 pb à 110 GeV et 3×10^{-4} pb à 3,2 TeV en association avec des quarks b. Dans le scénario M_h^{125}, ces limites se traduisent en une région d'exclusion dans le plan $(m_A, \tan\beta)$. Les valeurs de m_A inférieures à 600 GeV sont exclues. Cette limite passe à 2 TeV pour $\tan\beta \gtrsim 50$. Dans le scénario $M_{H_1}^{125}(\text{CPV})$, cette région est donnée dans le plan $(m_{H^\pm}, \tan\beta)$. Les valeurs de $m_{H^\pm}$ inférieures à 400 GeV sont exclues. Lorsque $\tan\beta \simeq 20$, l'exclusion s'étend jusqu'à $m_{H^\pm} \simeq 1{,}4$ TeV.

Pour les analyses dans lesquelles une particule se désintègre en $\tau^+\tau^-$, comme celle présentée dans cette thèse, la reconstruction de la masse d'une paire de τ est cruciale. Les neutrinos issus des désintégrations des τ compliquent cette tâche car ils sont invisibles dans le détecteur. Le *machine learning* apporte une solution. Le modèle obtenu dans cette thèse permet de reconstruire la masse d'une paire de τ de 50 GeV à 800 GeV avec une résolution de 20 % à 50 GeV, 26 % à 250 GeV et 22 % à 800 GeV. Ce modèle est 60 fois plus rapide et propose une meilleure description du boson Z que l'algorithme SVFIT actuellement utilisé au sein de la collaboration CMS.

Mots clés

Higgs, CMS, Jets, Calibration, Tau, Reconstruction de la masse, Réseau de neurones, Régression.

Abstract

Search for additional heavy Higgs bosons decaying to tau lepton pair in the CMS experiment at LHC with machine learning techniques

Despite decades of correct predictions, physicists are convinced that the Standard Model (SM) does not show us the whole picture. Among the various extensions going beyond the SM (BSM), the Minimal Supersymmetric extension of the SM (MSSM) predicts two charged Higgs bosons, $H^{\pm}$, and three neutrals: h corresponding to the observed one discovered in 2012 and H and A being additionnals with respect to the SM. The MSSM phenomenology motivates the focus on events containing a di-τ pair. The MSSM $H/A \to \tau\tau$ analysis is thus the core of this thesis.

In this thesis, the search for H and A is performed on the data collected with the CMS detector from 2016 to 2018 on proton collisions at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 137 fb^{-1}.

Jets are complex physics objets obtained in the proton collisions occuring at the CERN LHC. Their calibration by the study of events containing a photon and a jet ($\gamma +$ jet) and the corresponding results on 2018 data are introduced. These results are used in the official CMS jet calibration.

No significant deviation above the expected background is observed in the MSSM $H/A \to \tau\tau$ analysis. Model-independent limits are then set on the product of the cross section and branching fraction for the production via gluon-fusion or in association with b quarks. These limits range from 15 pb at 110 GeV to 3×10^{-4} pb at 3,2 TeV for gluon-fusion and from 1,2 pb at 110 GeV to 3×10^{-4} pb at 3,2 TeV for b-associated production. In the M_h^{125} scenario, these limits translate into an exclusion region in the $(m_A, \tan\beta)$ plane. Values of m_A below 600 GeV are excluded and this goes up to 2 TeV for $\tan\beta \gtrsim 50$. In the $M_{H_1}^{125}(\text{CPV})$ scenario, the region is given in the $(m_{H^{\pm}}, \tan\beta)$ plane. Values of $m_{H^{\pm}}$ below 400 GeV are excluded and this goes up to 1,4 TeV for $\tan\beta \simeq 20$.

To test any theory involving Higgs or Z bosons decaying to $\tau^+\tau^-$, the reconstruction of di-τ mass in a faster and more accurate way than the existing methods is crucial. However, it is an arduous task due to existence of neutrinos as decay product of each τ lepton which are invisible to detectors at LHC. Machine learning techniques bring a solution for this task. The reconstruction of the di-τ mass by a deep neural network (DNN) is achieved in this thesis from 50 GeV to 800 GeV with a 20 % resolution at 50 GeV, 26 % at 250 GeV and 22 % at 800 GeV. This DNN is 60 times faster and better at describing the Z boson than the SVFIT algorithm currently used in CMS.

Keywords

Higgs, CMS, Jets, Calibration, Tau, Mass reconstruction, Neural Network, Regression.

Remerciements

Cette thèse n'aurait pas pu être ce qu'elle est ni menée à son terme sans la présence, l'accompagnement et le soutien de mes collègues, amis et membres de ma famille. Je souhaite ici leur exprimer ma reconnaissance, tant professionnelle que personnelle.

Merci à Colin, mon « papa de thèse », d'avoir accepté de m'encadrer sur ce sujet. Tu m'as amené à être encore plus rigoureux et tu as su m'orienter sur des travaux réalisables malgré les aléas et contraintes de cette deuxième moitié de thèse. Tu m'as dans le même temps toujours laissé le choix et je t'en suis reconnaissant. Je te souhaite le meilleur dans ta nouvelle vie professionnelle !

Merci à Gaël, mon « grand frère de thèse », de m'avoir fait découvrir tous les outils que j'ai utilisés dans ma thèse. Que ce soit le terminal et *bash*, Git ou même HEPPY, tu m'as donné les clefs de leur maîtrise ! Et ce, toujours dans la bonne humeur même quand tu étais sous pression. Je ne connais que peu de personnes avec une telle patience.

Merci à Ece, ma « maman de thèse », pour ton soutien sur les événements $\gamma +$ jets, pour ton aide et tes conseils précieux sur le projet de *machine learning* et tes retours sur mes travaux. Nous avons formé une très bonne équipe et je ne serai pas allé si loin sans toi. Je te souhaite bonne chance pour ton prochain poste, et surtout qu'il soit moins mouvementé !

Merci à Hugues, mon « cousin de thèse », d'avoir pris le temps nécessaire à la transmission de ta responsabilité sur l'analyse des événements $\gamma +$ jets. Le C++ m'était inconnu, mais tu m'as donné les informations nécessaires pour comprendre le code associé et fournir les résultats à la collaboration.

Merci à Yi, Giuseppe, Juska et Mikko pour leur écoute et leurs instructions claires quant aux suites attendues à l'analyse des événements $\gamma +$ jets. Mention spéciale à Juska pour tes réponses à mes questions pas forcément évidentes, et pour ton invitation à venir avec mes futurs élèves au point 5. Ce ne sera pas pour tout de suite visiblement, mais ce n'est que partie remise !

Vielen Dank aux membres de l'*Institut für Experimentelle Teilchenphysik* du *Karlsruher Institut für Technologie* de m'avoir accueilli dans votre groupe, même à distance, afin de mener ensemble l'analyse des événements $H \to \tau\tau$ du Run II. Merci à Günter pour ta bonne humeur et l'attention portée à ma bonne santé en ces temps difficiles. Merci à Roger d'avoir accepté cette collaboration. Merci à Artur pour ton aide et tes nombreux conseils, lors de l'implémentation du scénario CPV mais aussi tout au long de l'analyse. Merci à Sebastian B. et Maximilian pour vos indications sur le fonctionnement des ressources informatiques au KIT, sans lesquelles je n'aurais pas pu les prendre en main aussi rapidement. Merci à Sebastian W. pour notre discussion fructueuse autour de la fonction de coût utilisée pour obtenir notre estimateur de la masse d'une paire de τ. Merci à Felix pour l'intérêt que tu portes à cet estimateur et tes retours sur son utilisation dans une nouvelle analyse.

Merci aux membres du groupe CMS MSSM HTT pour notre collaboration autour de cette analyse. En plus des membres de KIT, je pense à Aleksei, Christopher, Daniel, David, Georges, Janek, Janik, Mareike, Oleg, Suman, et tous ceux que je n'ai pas mentionné ici. Vos retours sur mes contributions à la note d'analyse m'ont aussi aidé pour ce manuscrit. Leurs qualités n'en sont que meilleures.

Merci à Davide pour l'interface de notre réseau de neurones en C++. Pour sûr, cela ouvre la porte à une utilisation sur de nouvelles analyses à CMS. Merci aussi pour tes retours sur son application aux événements $hh \to b\bar{b}\tau\tau$, dont la topologie différente de ceux que j'ai étudiés nous montre son comportement dans cette situation inédite.

Merci aux doctorants de l'IP2I pour toutes les discussions, scientifiques, techniques ou plus personnelles, les pauses café, les parties de billard, de ping-pong, de pétanque, les soirées et autres sorties. Je pense en particulier à Antoine L., Grégoire, Jean-François et surtout Aurélien, dont je souhaite saluer l'honnêteté et la franchise. À Corentin et Martin aussi, le voyage des nouveaux doctorants au

CERN n'aurait pas eu lieu sans votre aide. À tous, je vous souhaite bon courage pour la fin de vos thèses et pour la suite.

Merci à Antoine C., pour ton parrainage, les discussions que nous avons eues et les opportunités de diffusion scientifique auprès du jeune public.

Merci aux membres de mon jury, Corinne AUGIER, Lucia DI CIACCIO, Anne-Catherine LE BIHAN, Steve MUANZA, David ROUSSEAU et Roger WOLF, d'avoir accepté d'en faire partie et d'évaluer rigoureusement mes travaux.

Merci aux membres de la direction de l'IP2I, ou IPNL initialement, pour le cadre dans lequel vous m'avez permis de réaliser ma thèse. Vous avez toujours été à l'écoute et présents, même lors des démarches administratives les plus complexes.

Merci aux membres du groupe CMS de l'IP2I et du CERN pour votre accueil, vos conseils et la découverte du monde de la recherche. Je le quitte aujourd'hui pour celui de l'enseignement, mais par préférence et non par rejet. Vous avez été de formidables collègues.

Merci aux guides de la collaboration CMS pour votre accueil et en particulier à Jacob pour ton énergie et ton enthousiasme à organiser les visites au point 5. J'espère que cette activité pourra reprendre bientôt pour faire découvrir ce fantastique lieu de sciences au public.

Merci à vous deux, Papa et Maman, d'avoir été des parents qui m'ont permis d'en arriver là. Vous m'avez laissé le choix de suivre cette voie sans fermer les autres et soutenu à chaque étape. Vous m'avez transmis bien plus que la myopie familiale et ces quelques mots sont loin d'exprimer toute ma reconnaissance.

Merci à Sarah, pour tout ce que nous vivons à deux. Merci pour le rythme de vie plus sain que celui que je me serai fixé sans toi. Merci pour ces moments de décompression tout au long de nos thèses. Merci pour le temps que tu as passé à relire ce manuscrit pour y relever mes maladresses.

Et à tous ceux que j'aurais oublié ici, merci.

Table des matières

Introduction **1**

1 Particules, interactions et phénoménologie **3**
1 Particules du modèle standard . . . 4
1.1 Fermions . . . 4
1.2 Bosons . . . 5
2 Formalisme théorique et interactions . . . 6
2.1 Introduction au formalisme du lagrangien . . . 6
2.2 Interaction électromagnétique . . . 9
2.3 Interaction électrofaible . . . 10
2.4 Mécanisme de Higgs et brisure spontanée de symétrie . . . 15
2.5 Interaction forte . . . 18
2.6 Formation des jets . . . 20
3 Succès et limites du modèle standard . . . 24
3.1 Succès . . . 24
3.2 Limites . . . 25
4 Au-delà du modèle standard . . . 27
4.1 Supersymétrie . . . 28
4.2 Modèles à deux doublets de Higgs . . . 28
4.3 Extension supersymétrique minimale du modèle standard . . . 30
4.4 Violation de CP avec les bosons de Higgs . . . 32
5 Phénoménologie des bosons de Higgs du MSSM . . . 32
5.1 Production de bosons de Higgs . . . 33
5.2 Désintégrations des bosons de Higgs . . . 36
5.3 Désintégrations des leptons τ . . . 38
6 Conclusion . . . 39

2 Dispositif expérimental **41**
1 Introduction . . . 41
2 CERN : Conseil Européen pour la Recherche Nucléaire . . . 42
2.1 Origines du CERN . . . 42
2.2 Réalisations du CERN . . . 42
2.3 Accélérateurs du CERN . . . 43
3 LHC : *Large Hadron Collider* . . . 43
3.1 Exploitation du LHC . . . 43
3.2 Accélération de protons . . . 45
3.3 Collisions de protons . . . 45
3.4 Luminosité, nombre d'événements et statistiques . . . 47
3.5 Empilement . . . 49
3.6 Expériences du LHC . . . 49
4 CMS : *Compact Muon Solenoïd* . . . 50
4.1 Vue d'ensemble et système de coordonnées . . . 50
4.2 Solénoïde . . . 51
4.3 Trajectographe . . . 53

4.4 Calorimètre électromagnétique ou ECAL . . . 54
4.5 Calorimètre hadronique ou HCAL . . . 57
4.6 Chambres à muons . . . 58
4.7 Prise de données à CMS . . . 60
5 Reconstruction des événements . . . 61
5.1 Éléments d'identification du *Particle Flow* . . . 62
5.2 Identification et reconstruction des particules . . . 64
5.3 Reconstruction des jets . . . 69
5.4 Énergie transverse manquante . . . 71
5.5 Taus hadroniques . . . 74
6 Simulation d'événements . . . 76
6.1 Génération d'événements . . . 76
6.2 Simulation du détecteur . . . 76
6.3 Corrections apportées aux simulations . . . 77
7 Conclusion . . . 81

3 Calibration en énergie des jets **83**
1 Introduction . . . 83
2 Étapes de la calibration . . . 84
2.1 Correction de l'empilement . . . 85
2.2 Correction de la réponse du détecteur en p_T et en η . . . 87
2.3 Corrections résiduelles . . . 87
2.4 Incertitudes sur la correction en énergie des jets . . . 90
2.5 Correction de la résolution en énergie . . . 92
3 Phénoménologie des événements γ + jets . . . 92
3.1 Principe des événements γ + jets et réponse équilibrée . . . 92
3.2 Activité additionnelle des jets et réponse MPF . . . 94
4 Correction résiduelle absolue en p_T avec les événements γ + jets . . . 95
4.1 Événements utilisés . . . 95
4.2 Analyse . . . 98
4.3 Résultats . . . 103
5 Correction de la résolution en énergie avec les événements γ + jets . . . 110
5.1 Définition de la résolution en énergie des jets . . . 110
5.2 Analyse . . . 110
5.3 Résultats . . . 111
6 Conclusion . . . 113

4 Recherche de bosons de Higgs supplémentaires de haute masse **115**
1 Introduction . . . 116
2 Corrections . . . 117
2.1 Efficacité des chemins de déclenchement . . . 117
2.2 Impulsions des particules générées . . . 118
3 Sélection des événements . . . 119
3.1 Sélection pour l'enregistrement des données . . . 120
3.2 Obtention du dilepton et sélection finale . . . 120
3.3 Sélection des jets . . . 124
4 Modélisation du bruit de fond . . . 125
4.1 Méthode des données encapsulées ou *embedding* . . . 127
4.2 Estimation du bruit de fond QCD dans le canal $e\mu$ . . . 129
4.3 Méthode des facteurs de faux ou *fake factors* . . . 130
5 Catégorisation des événements et variables discriminantes . . . 135
5.1 Catégories « BSM » . . . 135
5.2 Catégories « SM » . . . 138
5.3 Combinaison des catégories SM et BSM . . . 141

6 Incertitudes systématiques . . . 142
6.1 Incertitudes de normalisation . . . 142
6.2 Incertitudes de forme . . . 143
7 Extraction du signal . . . 146
7.1 Modèle de vraisemblance . . . 146
7.2 Méthode CL_s . . . 149
7.3 Limites indépendantes du modèle . . . 150
7.4 Interprétation dans les scénarios du MSSM . . . 150
8 Résultats et interprétations . . . 152
8.1 Limites indépendantes du modèle . . . 152
8.2 Limites du scénario M_h^{125} . . . 153
8.3 Limites du scénario $M_{H_1}^{125}$ avec violation de CP . . . 154
9 Conclusion . . . 154

5 Reconstruction de la masse d'une résonance grâce au *machine learning* 157
1 Introduction . . . 158
2 Événements utilisés ou échantillons . . . 159
2.1 Génération avec FastSim . . . 159
2.2 Sélection des événements . . . 160
2.3 Événements obtenus et pondération . . . 161
2.4 Cible et variables d'entrée des modèles . . . 162
3 Concepts communs aux modèles . . . 163
3.1 Fonction de coût . . . 163
3.2 Entraînement, descente de gradient et mini-lots . . . 163
3.3 Sous-entraînement et surentraînement . . . 164
4 Arbres de décision améliorés . . . 165
4.1 Arbres de décision . . . 165
4.2 *Gradient Boosting* et descente de gradient . . . 166
5 Réseaux de neurones profonds . . . 167
5.1 Neurones . . . 167
5.2 Réseaux de neurones . . . 169
5.3 Entraînement . . . 169
6 Optimisation des hyper-paramètres et choix d'un modèle . . . 172
6.1 Variables d'entrée . . . 173
6.2 Type de modèle . . . 175
6.3 Fonction de coût . . . 176
6.4 Algorithme d'optimisation . . . 177
6.5 Autres hyper-paramètres . . . 178
7 Discussions . . . 184
7.1 Effet de l'empilement . . . 184
7.2 Effet de la reconstruction des particules . . . 186
7.3 Effet des faux taus hadroniques . . . 187
7.4 Effet de la séparation des canaux . . . 188
7.5 Effet de la définition de $E_{\mathrm{T}}^{\mathrm{miss}}$ . . . 194
7.6 Effet de l'intervalle de masse . . . 196
7.7 Modèle final . . . 201
8 Utilisation du modèle dans les analyses CMS . . . 201
8.1 Utilisation de m_{ML} comme variable discriminante . . . 201
8.2 Comparaison de m_{ML} à m_{SVFIT} . . . 205
9 Conclusion . . . 210

Conclusion 213

A Systèmes d'unités 215

B Notions mathématiques **217**

C Diagrammes de Feynman **219**

D Jeux de données – γ + jets **221**
1 Jeux de données pour l'analyse 2018 des événements γ + jets 221
2 Jeux de données pour l'analyse 2017UL des événements γ + jets 221

E Jeux de données – $H \to \tau\tau$ **223**

F Chemins de déclenchement – $H \to \tau\tau$ **235**

G Distributions de contrôle – $H \to \tau\tau$ **241**

H Distributions des variables discriminantes – $H \to \tau\tau$ **315**

I Ajustement des paramètres de nuisance – $H \to \tau\tau$ **359**

Bibliographie **405**

Table des figures

1.1 Particules élémentaires du modèle standard. . . . 4
1.2 Trajectoires d'une particule. . . . 6
1.3 Diagrammes de Feynman possibles à partir du terme $\bar{\psi}\gamma^{\mu}eQA_{\mu}\psi$ du lagrangien $\mathcal{L}_{QED}$. 10
1.4 Diagrammes de Feynman de production de paire $W^{+}W^{-}$ à l'arbre. . . . 10
1.5 Diagramme de Feynman correspondant à l'interaction entre trois bosons. . . . 13
1.6 Forme du potentiel du Higgs V selon le signe de μ^2. . . . 16
1.7 Combinaisons des couleurs des quarks dans les hadrons. . . . 18
1.8 Interaction entre un quark et un gluon. . . . 19
1.9 Diagrammes de Feynman correspondant à l'interaction entre trois et quatre gluons. . 19
1.10 Mesure de α_s en fonction de l'échelle d'énergie. . . . 20
1.11 Spectre de masse des hadrons légers. . . . 20
1.12 Un boson Z se désintègre en paire quark-antiquark. . . . 21
1.13 Formation de deux gerbes partoniques. . . . 22
1.14 Formation de jets dans le cadre du modèle d'agglomération hadronique. . . . 22
1.15 Formation de jets dans le cadre du modèle des cordes de Lund. . . . 23
1.16 Trois jets, dont un de saveur lourde. . . . 24
1.17 Diagrammes de Feynman des corrections à la masse du boson de Higgs. . . . 26
1.18 Masses des bosons de Higgs du MSSM. . . . 31
1.19 Sections efficaces des modes de production du boson de Higgs du modèle standard. . 34
1.20 Production de boson de Higgs par fusion de gluons et de bosons vecteurs. . . . 34
1.21 Production de boson de Higgs en association avec un boson. . . . 34
1.22 Production de boson de Higgs en association avec un quark top. . . . 34
1.23 Production de boson de Higgs du MSSM par fusion de gluons et de bosons vecteurs. . 35
1.24 Production de boson de Higgs du MSSM en association avec un boson. . . . 35
1.25 Production de boson de Higgs du MSSM en association avec un quark b. . . . 36
1.26 Rapports de branchement des bosons de Higgs du MSSM. . . . 37
1.27 Désintégration $h, H, A \to \tau^{+}\tau^{-}$. . . . 38
1.28 Diagrammes de Feynman de désintégration d'un τ^{-}. . . . 38

2.1 Emplacements des grandes expériences du LEP et du LHC. . . . 44
2.2 Complexe des accélérateurs du CERN. . . . 44
2.3 Diagramme de Feynman de la propagation d'un proton avec apparition d'une paire $q\bar{q}$. 45
2.4 Fonctions de distribution partoniques. . . . 46
2.5 Luminosité totale et empilement en 2016. . . . 48
2.6 Luminosité totale et empilement en 2017. . . . 48
2.7 Luminosité totale et empilement en 2018. . . . 48
2.8 Vue ouverte du détecteur CMS. . . . 50
2.9 Système de coordonnées du détecteur CMS. . . . 51
2.10 Vue longitudinale d'un quadrant du détecteur CMS. . . . 52
2.11 Champ magnétique dans le détecteur CMS. . . . 53
2.12 Schéma détaillé du trajectographe du détecteur CMS. . . . 54
2.13 Comparaison des trajectographes à pixels utilisés en 2016 et à partir de 2017. . . . 54
2.14 Résolution en p_{T} du trajectographe. . . . 55
2.15 Schéma du calorimètre électronique de CMS. . . . 55

2.16 Évolution temporelle de la réponse du ECAL. . . . 56
2.17 Résolution relative de l'énergie des électrons dans le ECAL lors du Run II. . . . 57
2.18 Schéma du calorimètre hadronique de CMS. . . . 57
2.19 Réponse relative du calorimètre hadronique de CMS. . . . 58
2.20 Schéma des chambres à muons de CMS. . . . 59
2.21 Distribution de la masse invariante des dimuons. . . . 59
2.22 Architecture du système de déclenchement L1. . . . 60
2.23 Coupe transverse schématique du détecteur CMS. . . . 61
2.24 Points d'interaction entre particules des événements et matière du détecteur. . . . 63
2.25 Compositions des jets reconstruits lors du Run II. . . . 70
2.26 Performances des algorithmes d'identification de la saveur des jets. . . . 73
2.27 Diagramme de Feynman de désintégration hadronique d'un τ^-. . . . 74
2.28 Représentation d'une collision de protons. . . . 77
2.29 Effet de la correction de recul de ${E_T}^{miss}$. . . . 81

3.1 Étapes successives de la JEC. . . . 84
3.2 Valeur moyenne de la réponse de jets d'événements QCD simulés. . . . 85
3.3 Contribution additionnelle de l'empilement au niveau particule. . . . 86
3.4 Réponse des jets reconstruits en fonction de p_T et η lors du Run II. . . . 87
3.5 Correction résiduelle relative en η lors du Run II. . . . 89
3.6 Correction résiduelle absolue en p_T pour $|\eta| < 1{,}3$ lors du Run II. . . . 89
3.7 Correction résiduelle de saveur en fonction de l'impulsion du jet. . . . 90
3.8 Incertitudes relatives sur la JEC en fonction de p_T et η lors du Run II. . . . 91
3.9 Résolution en énergie des jets. . . . 93
3.10 Résolution en énergie des photons. . . . 93
3.11 Diagrammes de Feynman donnant un photon et un jet dans l'état final. . . . 94
3.12 Topologies typiques des événements $\gamma +$ jets. . . . 94
3.13 Diagrammes de Feynman de processus avec ISR ou FSR. . . . 94
3.14 Régions des calorimètres à exclure de l'analyse dans le plan (η, ϕ). . . . 97
3.15 Densités de probabilité de N_{PU} pour 2018. . . . 99
3.16 Densités de probabilité de N_{PU} pour 2017-UL. . . . 100
3.17 Observables d'événements $\gamma +$ jets en 2018. . . . 101
3.18 Réponses équilibrée et MPF en 2018. . . . 102
3.19 Extrapolation vers $\alpha = 0$ de la réponse des jets. . . . 103
3.20 Réponses équilibrées en 2018 avant extrapolation. . . . 104
3.21 Réponses MPF en 2018 avant extrapolation. . . . 105
3.22 Réponses équilibrées en 2018 après extrapolation. . . . 106
3.23 Réponses MPF en 2018 après extrapolation. . . . 107
3.24 Distributions de la réponse MPF moyenne en fonction de p_T en 2018. . . . 109
3.25 Comparaison des réponses MPF avant et après correction résiduelle absolue en 2018. . . . 109
3.26 Détermination de la résolution en énergie des jets. . . . 112
3.27 Facteurs correctifs de la résolution en énergie des jets. . . . 112
3.28 Réponses équilibrées avant et après correction de la résolution en énergie des jets. . . . 113

4.1 Production de boson de Higgs du MSSM par fusion de gluons. . . . 119
4.2 Illustration de la définition de $\hat{\zeta}$ [220]. Le plan de ce schéma est le plan transverse. . . . 124
4.3 Diagrammes de Feynman des signaux et principaux bruits de fond de l'analyse. . . . 125
4.4 Schéma récapitulatif de la méthode des données encapsulées. . . . 128
4.5 Distributions de ${E_T}^{miss}$ pour le canal $\mu\tau_h$ en 2018. . . . 129
4.6 Régions A, B, C et D utilisées pour estimer le bruit de fond QCD. . . . 129
4.7 Fractions des sources de faux τ_h dans le canal $\mu\tau_h$ en 2018. . . . 132
4.8 Illustration de la méthode des facteurs de faux. . . . 132
4.9 Ajustement de FF_W dans le canal $\mu\tau_h$ en 2018. . . . 133
4.10 Ajustement de FF_t dans le canal $\mu\tau_h$ en 2018. . . . 133

4.11 Corrections résiduelles des facteurs de faux dans le canal $\mu\tau_h$ en 2018. . . . 134
4.12 Distributions de la masse transverse de l'électron pour le canal $e\tau_h$ en 2016. . . . 135
4.13 Illustrations des catégorisations basées sur m_T^ℓ et D_ζ . . . 136
4.14 Catégories BSM pour les quatre canaux considérés. . . . 137
4.15 Diagrammes de Feynman d'un événement $H \to \tau\tau \to \mu\tau_h$. . . . 137
4.16 Distributions de m_T^{tot} en 2017 pour deux catégories et canaux. . . . 138
4.17 Distributions de NN_{score} en 2017 dans le canal $e\mu$. . . . 140
4.18 Réduction à une dimension de la catégorie xxh. . . . 141
4.19 Définition des deux régions utilisant des catégories différentes. . . . 142
4.20 Distributions de m_T^{tot} avec et sans regroupement des incertitudes. . . . 148
4.21 Limites d'exclusion indépendantes. . . . 153
4.22 Limites d'exclusion du scénario M_h^{125}. . . . 154
4.23 Limites d'exclusion du scénario $M_{H_1}^{125}$(CPV). . . . 154

5.1 Origine des limites haute (gauche) et basse (droite) de l'intervalle de masse utilisé. . . . 159
5.2 Distributions de la masse générée de $\mathcal{H}$. . . . 162
5.3 Diagramme de Feynman des événements d'entraînement des modèles. . . . 162
5.4 Illustrations du sous-entraînement et du surentraînement. . . . 165
5.5 Exemple d'un arbre de décision. . . . 166
5.6 Structure d'un neurone. . . . 167
5.7 Exemples de fonctions d'activation. . . . 168
5.8 Structure d'un réseau de neurones. . . . 170
5.9 Exemples de graphiques rendant compte des performances des modèles. . . . 173
5.10 Évaluations des XGBs regroupés selon les variables d'entrée. . . . 174
5.11 Évaluations des DNNs regroupés selon les variables d'entrée. . . . 175
5.12 Évaluations des XGBs et des DNNs par L_{MSE} et L_{MAPE}. . . . 176
5.13 Évaluations des XGBs et des DNNs par $\Delta_{1\sigma}$. . . . 176
5.14 Évaluations des XGBs et des DNNs par L_{MSE} et L_{MAPE} à basse masse. . . . 177
5.15 Évaluations des DNNs regroupés selon la fonction de coût. . . . 178
5.16 Évaluations des DNNs regroupés selon l'algorithme d'optimisation. . . . 179
5.17 Évaluations des DNNs regroupés selon N_{cc}. . . . 180
5.18 Évaluations des DNNs regroupés selon $N_{n/c}$. . . . 181
5.19 Évaluations des DNNs regroupés selon le mode d'initialisation des poids. . . . 182
5.20 Évaluations des DNNs regroupés selon la fonction d'activation. . . . 183
5.21 Réponse du modèle A. . . . 184
5.22 Réponse des modèles B à G. . . . 185
5.23 Réponses du modèle B^{0PU} sur les événements sans et avec empilement. . . . 186
5.24 Réponses du modèle B sur les événements sans et avec empilement. . . . 187
5.25 Réponses du modèle B^{gen} avec reconstruction parfaite ou réelle. . . . 187
5.26 Effet des faux τ_h sur la réponse du modèle B. . . . 189
5.27 Effet des faux τ_h sur l'écart des prédictions du modèle B à la valeur vraie. . . . 190
5.28 Comparaison des modèles entraînés par canal ($\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$) au modèle B. . . . 192
5.29 Comparaison des modèles entraînés par canal ($\mu\mu$, $e\mu$, ee) au modèle B. . . . 193
5.30 Comparaison du modèle entraîné sur les canaux semi-leptoniques au modèle B. . . . 194
5.31 Comparaison du modèle entraîné sur les canaux leptoniques au modèle B. . . . 194
5.32 Réponses du modèle B avec PUPPIMET ou PFMET. . . . 195
5.33 Réponses du modèle B^{PF} avec PUPPIMET ou PFMET. . . . 195
5.34 Performances du modèle $B^{200\text{-}500}$. . . . 196
5.35 Performances du modèle B. . . . 197
5.36 Distributions de y_{vraie} à $y_{préd}$ fixée. . . . 197
5.37 Mise en place de la coupure symétrique. . . . 198
5.38 Zones considérées pour l'entraînement. . . . 199
5.39 Comparaison des modèles B, B′ et B". . . . 200
5.40 Réponse de m_T^{tot} et comparaison de sa résolution relative à celle de m_{ML}. . . . 202

5.41 Limites d'exclusion attendues avec m_{ML} en 2017. 202
5.42 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} dans la catégorie `btag` du canal $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$. 203
5.43 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} dans la catégorie `no-btag` du canal $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$. 204
5.44 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} dans la catégorie `BSM no-btag loose-`m_{T} du canal $\mu\tau_{\mathrm{h}}$. . . 204
5.45 Distributions de m_{SVFIT} et m_{ML} dans les canaux $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $e\tau_{\mathrm{h}}$ et $e\mu$. 206
5.46 Distributions de m_{SVFIT} et m_{ML}, événements de haute masse, canaux $\mu\tau_{\mathrm{h}}$, $e\tau_{\mathrm{h}}$ et $e\mu$. . . 207
5.47 Distributions de m_{SVFIT} et m_{ML} sur des événements de topologie particulière. 208
5.48 Distributions de m_{SVFIT} et m_{ML} sur des événements de topologie particulière. 208
5.49 Distributions de m_{SVFIT} et m_{ML} sur des événements de topologie particulière. 209
5.50 Distributions de m_{SVFIT} et m_{ML} sur des événements de topologie particulière. 209
5.51 Temps nécessaire à l'obtention des prédictions de masse avec SVFIT et avec le DNN B". 210

C.1 Diagramme de Feynman de la diffusion électron-électron. 219
C.2 Prise en compte de plusieurs diagrammes de Feynman. 219
C.3 Types de particules et lignes correspondantes dans les diagrammes de Feynman. . . . 220
C.4 Diagramme de Feynman issu du terme (C.1). 220

G.1 Distribution de $|\eta|(\tau_{\mathrm{h}})$ en fonction de m_{ML}. 241
G.2 Variations de la réponse du réseau de neurones en fonction de ΔN_{PU}. 242
G.3 Distributions de contrôle, 2016 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, cinématique des leptons ($\tau_{\mathrm{h}}{}^{(1)}$, $\tau_{\mathrm{h}}{}^{(2)}$). 243
G.4 Distributions de contrôle, 2016 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, dilepton et énergie transverse manquante. 244
G.5 Distributions de contrôle, 2016 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, masses transverses, D_ζ et masses. 245
G.6 Distributions de contrôle, 2016 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, cinématique des deux jets principaux. 246
G.7 Distributions de contrôle, 2016 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, b-jets et activité hadronique additionnelle. 247
G.8 Distributions de contrôle, 2016 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, jets et empilement. 248
G.9 Distributions de contrôle, 2016 $\mu\tau_{\mathrm{h}}$, cinématique des leptons (μ, τ_{h}). 249
G.10 Distributions de contrôle, 2016 $\mu\tau_{\mathrm{h}}$, dilepton et énergie transverse manquante. 250
G.11 Distributions de contrôle, 2016 $\mu\tau_{\mathrm{h}}$, masses transverses, D_ζ et masses. 251
G.12 Distributions de contrôle, 2016 $\mu\tau_{\mathrm{h}}$, cinématique des deux jets principaux. 252
G.13 Distributions de contrôle, 2016 $\mu\tau_{\mathrm{h}}$, b-jets et activité hadronique additionnelle. 253
G.14 Distributions de contrôle, 2016 $\mu\tau_{\mathrm{h}}$, jets et empilement. 254
G.15 Distributions de contrôle, 2016 $e\tau_{\mathrm{h}}$, cinématique des leptons (e, τ_{h}). 255
G.16 Distributions de contrôle, 2016 $e\tau_{\mathrm{h}}$, dilepton et énergie transverse manquante. 256
G.17 Distributions de contrôle, 2016 $e\tau_{\mathrm{h}}$, masses transverses, D_ζ et masses. 257
G.18 Distributions de contrôle, 2016 $e\tau_{\mathrm{h}}$, cinématique des deux jets principaux. 258
G.19 Distributions de contrôle, 2016 $e\tau_{\mathrm{h}}$, b-jets et activité hadronique additionnelle. 259
G.20 Distributions de contrôle, 2016 $e\tau_{\mathrm{h}}$, jets et empilement. 260
G.21 Distributions de contrôle, 2016 $e\mu$, cinématique des leptons (e, μ). 261
G.22 Distributions de contrôle, 2016 $e\mu$, dilepton et énergie transverse manquante. 262
G.23 Distributions de contrôle, 2016 $e\mu$, masses transverses, D_ζ et masses. 263
G.24 Distributions de contrôle, 2016 $e\mu$, cinématique des deux jets principaux. 264
G.25 Distributions de contrôle, 2016 $e\mu$, b-jets et activité hadronique additionnelle. 265
G.26 Distributions de contrôle, 2016 $e\mu$, jets et empilement. 266
G.27 Distributions de contrôle, 2017 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, cinématique des leptons ($\tau_{\mathrm{h}}{}^{(1)}$, $\tau_{\mathrm{h}}{}^{(2)}$). 267
G.28 Distributions de contrôle, 2017 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, dilepton et énergie transverse manquante. 268
G.29 Distributions de contrôle, 2017 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, masses transverses, D_ζ et masses. 269
G.30 Distributions de contrôle, 2017 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, cinématique des deux jets principaux. 270
G.31 Distributions de contrôle, 2017 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, b-jets et activité hadronique additionnelle. 271
G.32 Distributions de contrôle, 2017 $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, jets et empilement. 272
G.33 Distributions de contrôle, 2017 $\mu\tau_{\mathrm{h}}$, cinématique des leptons (μ, τ_{h}). 273
G.34 Distributions de contrôle, 2017 $\mu\tau_{\mathrm{h}}$, dilepton et énergie transverse manquante. 274
G.35 Distributions de contrôle, 2017 $\mu\tau_{\mathrm{h}}$, masses transverses, D_ζ et masses. 275
G.36 Distributions de contrôle, 2017 $\mu\tau_{\mathrm{h}}$, cinématique des deux jets principaux. 276
G.37 Distributions de contrôle, 2017 $\mu\tau_{\mathrm{h}}$, b-jets et activité hadronique additionnelle. 277

G.38 Distributions de contrôle, 2017 $\mu\tau_h$, jets et empilement. . . . 278
G.39 Distributions de contrôle, 2017 $e\tau_h$, cinématique des leptons (e, τ_h). . . . 279
G.40 Distributions de contrôle, 2017 $e\tau_h$, dilepton et énergie transverse manquante. . . . 280
G.41 Distributions de contrôle, 2017 $e\tau_h$, masses transverses, D_ζ et masses. . . . 281
G.42 Distributions de contrôle, 2017 $e\tau_h$, cinématique des deux jets principaux. . . . 282
G.43 Distributions de contrôle, 2017 $e\tau_h$, b-jets et activité hadronique additionnelle. . . . 283
G.44 Distributions de contrôle, 2017 $e\tau_h$, jets et empilement. . . . 284
G.45 Distributions de contrôle, 2017 $e\mu$, cinématique des leptons (e, μ). . . . 285
G.46 Distributions de contrôle, 2017 $e\mu$, dilepton et énergie transverse manquante. . . . 286
G.47 Distributions de contrôle, 2017 $e\mu$, masses transverses, D_ζ et masses. . . . 287
G.48 Distributions de contrôle, 2017 $e\mu$, cinématique des deux jets principaux. . . . 288
G.49 Distributions de contrôle, 2017 $e\mu$, b-jets et activité hadronique additionnelle. . . . 289
G.50 Distributions de contrôle, 2017 $e\mu$, jets et empilement. . . . 290
G.51 Distributions de contrôle, 2018 $\tau_h\tau_h$, cinématique des leptons ($\tau_h{}^{(1)}$, $\tau_h{}^{(2)}$). . . . 291
G.52 Distributions de contrôle, 2018 $\tau_h\tau_h$, dilepton et énergie transverse manquante. . . . 292
G.53 Distributions de contrôle, 2018 $\tau_h\tau_h$, masses transverses, D_ζ et masses. . . . 293
G.54 Distributions de contrôle, 2018 $\tau_h\tau_h$, cinématique des deux jets principaux. . . . 294
G.55 Distributions de contrôle, 2018 $\tau_h\tau_h$, b-jets et activité hadronique additionnelle. . . . 295
G.56 Distributions de contrôle, 2018 $\tau_h\tau_h$, jets et empilement. . . . 296
G.57 Distributions de contrôle, 2018 $\mu\tau_h$, cinématique des leptons (μ, τ_h). . . . 297
G.58 Distributions de contrôle, 2018 $\mu\tau_h$, dilepton et énergie transverse manquante. . . . 298
G.59 Distributions de contrôle, 2018 $\mu\tau_h$, masses transverses, D_ζ et masses. . . . 299
G.60 Distributions de contrôle, 2018 $\mu\tau_h$, cinématique des deux jets principaux. . . . 300
G.61 Distributions de contrôle, 2018 $\mu\tau_h$, b-jets et activité hadronique additionnelle. . . . 301
G.62 Distributions de contrôle, 2018 $\mu\tau_h$, jets et empilement. . . . 302
G.63 Distributions de contrôle, 2018 $e\tau_h$, cinématique des leptons (e, τ_h). . . . 303
G.64 Distributions de contrôle, 2018 $e\tau_h$, dilepton et énergie transverse manquante. . . . 304
G.65 Distributions de contrôle, 2018 $e\tau_h$, masses transverses, D_ζ et masses. . . . 305
G.66 Distributions de contrôle, 2018 $e\tau_h$, cinématique des deux jets principaux. . . . 306
G.67 Distributions de contrôle, 2018 $e\tau_h$, b-jets et activité hadronique additionnelle. . . . 307
G.68 Distributions de contrôle, 2018 $e\tau_h$, jets et empilement. . . . 308
G.69 Distributions de contrôle, 2018 $e\mu$, cinématique des leptons (e, μ). . . . 309
G.70 Distributions de contrôle, 2018 $e\mu$, dilepton et énergie transverse manquante. . . . 310
G.71 Distributions de contrôle, 2018 $e\mu$, masses transverses, D_ζ et masses. . . . 311
G.72 Distributions de contrôle, 2018 $e\mu$, cinématique des deux jets principaux. . . . 312
G.73 Distributions de contrôle, 2018 $e\mu$, b-jets et activité hadronique additionnelle. . . . 313
G.74 Distributions de contrôle, 2018 $e\mu$, jets et empilement. . . . 314

H.1 Distributions de $\mathrm{NN}_{\mathrm{score}}$ en 2016 dans le canal $\tau_h\tau_h$. . . . 316
H.2 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2016 dans le canal $\tau_h\tau_h$. . . . 317
H.3 Distributions de $\mathrm{NN}_{\mathrm{score}}$ en 2016 dans le canal $\mu\tau_h$. . . . 318
H.4 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2016 dans le canal $\mu\tau_h$, catégories `btag`. . . . 319
H.5 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2016 dans le canal $\mu\tau_h$, catégories `no-btag`. . . . 320
H.6 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2016 dans le canal $\mu\tau_h$, catégories `BSM no-btag`. . . . 321
H.7 Distributions de $\mathrm{NN}_{\mathrm{score}}$ en 2016 dans le canal $e\tau_h$. . . . 322
H.8 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2016 dans le canal $e\tau_h$, catégories `btag`. . . . 323
H.9 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2016 dans le canal $e\tau_h$, catégories `no-btag`. . . . 324
H.10 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2016 dans le canal $e\tau_h$, catégories `BSM no-btag`. . . . 325
H.11 Distributions de $\mathrm{NN}_{\mathrm{score}}$ en 2016 dans le canal $e\mu$. . . . 326
H.12 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2016 dans le canal $e\mu$, catégories `btag`. . . . 327
H.13 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2016 dans le canal $e\mu$, catégories `no-btag`. . . . 328
H.14 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2016 dans le canal $e\mu$, catégories `BSM no-btag`. . . . 329
H.15 Distributions de $\mathrm{NN}_{\mathrm{score}}$ en 2017 dans le canal $\tau_h\tau_h$. . . . 330
H.16 Distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} en 2017 dans le canal $\tau_h\tau_h$. . . . 331

H.17 Distributions de NN_{score} en 2017 dans le canal $\mu\tau_h$. 332
H.18 Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal $\mu\tau_h$, catégories `btag`. 333
H.19 Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal $\mu\tau_h$, catégories `no-btag`. 334
H.20 Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal $\mu\tau_h$, catégories `BSM no-btag`. 335
H.21 Distributions de NN_{score} en 2017 dans le canal $e\tau_h$. 336
H.22 Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal $e\tau_h$, catégories `btag`. 337
H.23 Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal $e\tau_h$, catégories `no-btag`. 338
H.24 Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal $e\tau_h$, catégories `BSM no-btag`. 339
H.25 Distributions de NN_{score} en 2017 dans le canal $e\mu$. 340
H.26 Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal $e\mu$, catégories `btag`. 341
H.27 Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal $e\mu$, catégories `no-btag`. 342
H.28 Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal $e\mu$, catégories `BSM no-btag`. 343
H.29 Distributions de NN_{score} en 2018 dans le canal $\tau_h\tau_h$. 344
H.30 Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $\tau_h\tau_h$. 345
H.31 Distributions de NN_{score} en 2018 dans le canal $\mu\tau_h$. 346
H.32 Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $\mu\tau_h$, catégories `btag`. 347
H.33 Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $\mu\tau_h$, catégories `no-btag`. 348
H.34 Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $\mu\tau_h$, catégories `BSM no-btag`. 349
H.35 Distributions de NN_{score} en 2018 dans le canal $e\tau_h$. 350
H.36 Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $e\tau_h$, catégories `btag`. 351
H.37 Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $e\tau_h$, catégories `no-btag`. 352
H.38 Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $e\tau_h$, catégories `BSM no-btag`. 353
H.39 Distributions de NN_{score} en 2018 dans le canal $e\mu$. 354
H.40 Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $e\mu$, catégories `btag`. 355
H.41 Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $e\mu$, catégories `no-btag`. 356
H.42 Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $e\mu$, catégories `BSM no-btag`. 357

Liste des tableaux

1.1 Représentation des fermions selon leur chiralité et leur isospin faible. 12
1.2 Valeurs des hypercharges, isospins faibles et charges électriques pour les leptons. . . . 14
1.3 Valeurs expérimentales des 19 paramètres libres du modèle standard. 27
1.4 Couplages des bosons de Higgs neutres. 30
1.5 Particules et leurs superpartenaires. 32
1.6 Couplages des bosons de Higgs neutres dans la limite découplée. 33
1.7 Rapports de branchement des événements $h \to \tau\tau$. 39

2.1 *Runs* du LHC. 43
2.2 Caractéristiques du solénoïde supraconducteur de CMS. 52
2.3 Aire effective de correction de l'isolation de l'électron. 66
2.4 Coupures du *cut-based veto ID*. 68
2.5 Aires effectives de correction de l'isolation du photon. 68
2.6 Coupures utilisées pour l'identification des photons. 68
2.7 Critères d'identification des jets pour l'année 2016. 72
2.8 Critères d'identification des jets pour l'année 2017. 72
2.9 Critères d'identification des jets pour l'année 2018. 72
2.10 Critères d'identification des jets pour les années 2017-UL et 2018-UL. 72
2.11 Modes de désintégration du τ considérés. 75
2.12 Efficacités d'identification de l'algorithme DEEPTAU. 76
2.13 Corrections à l'énergie des électrons. 78
2.14 Corrections au taux d'identification des muons comme des τ_h. 79
2.15 Corrections au taux d'identification des électrons comme des τ_h. 79
2.16 Corrections à l'énergie des taus hadroniques. 80
2.17 Corrections à l'énergie des leptons identifiés comme des taus hadroniques. 80

3.1 Liste des périodes de prise de données considérées et luminosités correspondantes. . 95
3.2 Sections efficaces des différents événements γ + jets simulés. 96
3.3 Chemins de déclenchement. 97
3.4 Intervalles de ${p_T}^{\gamma}$. 98
3.5 Intervalles larges de $|\eta^{\text{jet}}|$. 98
3.6 Intervalles de α utilisés pour la JES. 101
3.7 Rapports des réponses équilibrée et MPF obtenus en 2018. 108
3.8 Rapports des réponses équilibrée et MPF obtenus en 2018 après extrapolation vers $\alpha = 0$. 109
3.9 Intervalles de ${p_T}^{\gamma}$ utilisés pour la JER. 111
3.10 Intervalles fins de $|\eta^{\text{jet}}|$ utilisés pour la JER. 111
3.11 Intervalles de α utilisés pour la JER. 111

4.1 Contributions en pourcent des bruits de fond aux canaux étudiés. 126
4.2 Valeurs prises par `gen_match`. 127
4.3 Remplacement des événements simulés par une estimation basée sur les données. . . 127
4.4 Catégories `SM` de bruit de fond pour les quatre canaux considérés. 141
4.5 Présences attendues des contributions au signal dans les catégories. 151

5.1 Liste des 7 modèles sélectionnés. . . . 184
5.2 Comparaison de nos hyper-paramètres à ceux de BÄRTSCHI & coll. . . . 184
5.3 Résolutions relatives de différents modèles. . . . 191

A.1 Équivalence entre les systèmes d'unités international et naturel. . . . 215

D.1 Jeux de données utilisés pour l'analyse 2018 des événements γ + jets. . . . 222
D.2 Jeux de données simulées utilisés pour l'analyse 2018 des événements γ + jets. . . . 222
D.3 Jeux de données utilisés pour l'analyse 2017UL des événements γ + jets. . . . 222
D.4 Jeux de données simulées utilisés pour l'analyse 2017UL des événements γ + jets. . . . 222

E.1 Fichiers de certification JSON. . . . 224
E.2 Jeux de données utilisés en 2016. . . . 224
E.3 Jeux de données utilisés en 2017. . . . 225
E.4 Jeux de données utilisés en 2018. . . . 225
E.5 Jeux de données simulées modélisant le boson de Higgs du modèle standard en 2016. 226
E.6 Jeux de données simulées modélisant les bosons de Higgs du MSSM en 2016. . . . 226
E.7 Jeux de données simulées modélisant le boson de Higgs du modèle standard en 2017. 227
E.8 Jeux de données simulées modélisant les bosons de Higgs du MSSM en 2017. . . . 227
E.9 Jeux de données simulées modélisant le boson de Higgs du modèle standard en 2018. 228
E.10 Jeux de données simulées modélisant les bosons de Higgs du MSSM en 2018. . . . 228
E.11 Jeux de données simulées modélisant le bruit de fond en 2016. . . . 229
E.12 Jeux de données simulées modélisant le bruit de fond en 2017. . . . 230
E.13 Jeux de données simulées modélisant le bruit de fond en 2018. . . . 231
E.14 Jeux de données encapsulées en 2016. . . . 232
E.15 Jeux de données encapsulées en 2017. . . . 232
E.16 Jeux de données encapsulées en 2018. . . . 233

F.1 Tableaux contenant les informations des chemins de déclenchement. . . . 235
F.2 Chemins de déclenchement utilisés en 2016 pour le canal $\tau_h\tau_h$. . . . 236
F.3 Chemins de déclenchement utilisés en 2016 pour le canal $\mu\tau_h$. . . . 236
F.4 Chemins de déclenchement utilisés en 2016 pour le canal $e\tau_h$. . . . 236
F.5 Chemins de déclenchement utilisés en 2016 pour le canal $e\mu$. . . . 237
F.6 Chemins de déclenchement utilisés en 2017 pour le canal $\tau_h\tau_h$. . . . 237
F.7 Chemins de déclenchement utilisés en 2017 pour le canal $\mu\tau_h$. . . . 237
F.8 Chemins de déclenchement utilisés en 2017 pour le canal $e\tau_h$. . . . 238
F.9 Chemins de déclenchement utilisés en 2017 pour le canal $e\mu$. . . . 238
F.10 Chemins de déclenchement utilisés en 2018 pour le canal $\tau_h\tau_h$. . . . 238
F.11 Chemins de déclenchement utilisés en 2018 pour le canal $\mu\tau_h$. . . . 239
F.12 Chemins de déclenchement utilisés en 2018 pour le canal $e\tau_h$. . . . 239
F.13 Chemins de déclenchement utilisés en 2018 pour le canal $e\mu$. . . . 239

Introduction

La physique est une science dont le but est de comprendre et modéliser les phénomènes régissant l'Univers. Les propriétés des objets dépendant de celles de leurs constituants, la description la plus fondamentale de l'Univers s'appuie sur la bonne compréhension du comportement des constituants de la matière et de leurs interactions.

Les avancées scientifiques, théoriques et expérimentales, ont dévoilé la nature de la matière à des échelles de plus en plus fines. Tout objet est un arrangement d'atomes, dont la taille est de l'ordre de 1 Å, soit 10^{-10} m. Ils sont eux-mêmes constitués d'un noyau atomique d'environ 10^{-15} m, autour duquel se trouve un nuage d'électrons. Ce noyau est lui-même constitué de protons et de neutrons, eux-même constitués de quarks u et d. À ce jour, il n'a pas été observé de sous-structure aux quarks ou à l'électron. De telles particules sont dites élémentaires.

La physique des particules, aussi nommée physique des hautes énergies, étudie les constituants élémentaires de la matière. Plus de 50 ans d'échanges entre prédictions théoriques et résultats expérimentaux ont permis de mettre en place dès les années 1960 le modèle standard, présenté dans le chapitre 1. Ce modèle a par exemple prédit l'existence du quark t, des gluons, des bosons W et Z ou encore du boson de Higgs, parfois des dizaines d'années avant leur observation.

Cependant, malgré plusieurs décennies de prédictions vérifiées expérimentalement, ce modèle n'explique pas toutes les observations réalisées. Par exemple, il ne prédit pas l'existence de la matière noire, alors que des études cosmologiques prouvent sa présence dans l'Univers. De plus, les neutrinos du modèle standard n'ont pas de masse, or leurs oscillations sont une preuve qu'ils en possèdent une. D'autres lacunes sont encore inhérentes à ce modèle et de nombreuses extensions sont proposées afin de les combler. L'une d'entre elles est l'extension supersymétrique minimale du modèle standard, ou MSSM, également présentée dans le chapitre 1.

Afin de tester la validité du modèle standard et des nouvelles théories allant au-delà, divers dispositifs expérimentaux ont été mis au point tels que le Grand Collisionneur de Hadrons (LHC, *Large Hadron Collider*) du CERN. Il permet, entre autres, de faire entrer en collision des protons. Lors de cet événement, leur énergie cinétique peut être convertie en masse, celle-ci correspondant alors à d'autres particules, par exemple un boson de Higgs.

Les particules issues des collisions sont pour la plupart instables. Elles se désintègrent alors en d'autres particules, qu'il est possible d'observer à l'aide de détecteurs tels que le Solénoïde Compact à Muons (CMS, *Compact Muon Solenoid*), présenté dans le chapitre 2. À partir de l'observation de leurs produits de désintégration, il est possible d'étudier la présence et les propriétés des particules issues des collisions. Le détecteur CMS a ainsi participé à la découverte du boson de Higgs en 2012.

La bonne caractérisation des objets physiques mesurés par CMS est un enjeu majeur pour réaliser des analyses de qualité. Le chapitre 3 présente la calibration en énergie des jets. Ces objets physiques complexes, constitués de nombreuses particules, sont une manifestation des quarks et des gluons. Les incertitudes systématiques liées aux jets sont dominantes dans plusieurs analyses, leur bonne calibration est donc cruciale. Lors de ma thèse, j'ai analysé des événements γ + jets dans lesquels l'énergie du photon (γ) permet de déterminer celle du jet. Les résultats obtenus concernent les données enregistrées en 2018 et en 2017-UL, c'est-à-dire dans un second traitement de l'année 2017. Ils sont utilisés dans la calibration officielle de CMS et donc dans toutes les analyses publiées par la collaboration exploitant les données récoltées ces années-là.

Le MSSM prédit cinq bosons de Higgs. L'un d'entre eux, h, doit correspondre au boson scalaire découvert en 2012. Parmi les quatre supplémentaires, deux sont chargés, H^+ et H^-. Les deux autres sont neutres, l'un scalaire, H, l'autre pseudo-scalaire, A. L'observation de ces deux bosons supplé-

mentaires serait ainsi une preuve directe d'une nouvelle physique au-delà du modèle standard. Leurs masses respectives sont plus élevées que celle de h et vraisemblablement proches, c'est pourquoi cette thèse porte sur la recherche de bosons de Higgs supplémentaires de haute masse. Leur canal de désintégration le plus prometteur, au vu de la phénoménologie de ces bosons au LHC, est la paire de leptons τ. Dans ce contexte, l'analyse des événements enregistrés par la collaboration CMS de 2016 à 2018 est présentée dans le chapitre 4.

Toutefois, les leptons τ ne sont pas stables et se désintègrent avant d'entrer dans le détecteur CMS. Lors de ce processus, des neutrinos sont produits. Or, ces derniers sont invisibles pour le détecteur, il est donc impossible de mesurer toutes les particules de l'état final. Les développements récents de l'intelligence artificielle et plus particulièrement du *machine learning* proposent de nouveaux outils tels que les réseaux de neurones afin de réaliser des tâches complexes. Le chapitre 5 étudie ainsi leur application à la reconstruction de la masse d'une particule se désintégrant en paire de leptons τ. L'application du modèle obtenu lors de cette thèse dans l'analyse du chapitre 4 est discutée. Une comparaison à l'algorithme actuellement utilisé par la collaboration CMS est également réalisée.

Chapitre 1
Particules, interactions et phénoménologie

Sommaire

1 Particules du modèle standard 4
1.1 Fermions 4
1.2 Bosons 5
2 Formalisme théorique et interactions 6
2.1 Introduction au formalisme du lagrangien 6
2.2 Interaction électromagnétique 9
2.3 Interaction électrofaible 10
2.4 Mécanisme de Higgs et brisure spontanée de symétrie 15
2.5 Interaction forte 18
2.6 Formation des jets 20
3 Succès et limites du modèle standard 24
3.1 Succès 24
3.2 Limites 25
4 Au-delà du modèle standard 27
4.1 Supersymétrie 28
4.2 Modèles à deux doublets de Higgs 28
4.3 Extension supersymétrique minimale du modèle standard 30
4.4 Violation de CP avec les bosons de Higgs 32
5 Phénoménologie des bosons de Higgs du MSSM 32
5.1 Production de bosons de Higgs 33
5.2 Désintégrations des bosons de Higgs 36
5.3 Désintégrations des leptons τ 38
6 Conclusion 39

Ce chapitre présente le contexte dans lequel s'inscrit cette thèse. Le modèle standard est le cadre théorique en place en physique des particules. Il permet de décrire les objets fondamentaux qui composent l'Univers, c'est-à-dire les particules et leurs interactions.

Les particules du modèle standard sont présentées dans la section 1. Le formalisme mathématique permettant de décrire leur comportement, faisant apparaître les forces fondamentales, est introduit dans la section 2. Le modèle standard ainsi construit propose une description de l'Univers à la fois précise et robuste.

Le boson de Higgs, dernière particule découverte à ce jour, a ainsi été postulé près de cinquante ans avant d'être observé. De nombreux succès, dont une présentation non exhaustive est proposée dans la section 3.1, couronnent ainsi le modèle standard. Cependant, malgré plusieurs décennies de prédictions correctement vérifiées, certaines observations montrent que le modèle standard ne saurait prétendre au titre de « théorie du tout ».

Les limites de ce modèle, dont certaines sont présentées dans la section 3.2, mènent à de nouvelles théories « au-delà du modèle standard », dont il est question dans la section 4. Parmi celles-ci se trouvent des modèles dit « à deux doublets de Higgs », c'est-à-dire avec un secteur de Higgs plus complexe, comme la supersymétrie.

Un cas particulier de modèle supersymétrique est l'extension supersymétrique minimale du modèle standard, ou MSSM, présentée dans la section 5. La phénoménologie du MSSM motive le choix du type d'événements d'intérêt pour la recherche de cette nouvelle physique.

1 Particules du modèle standard

Une particule est considérée comme élémentaire si elle ne possède pas de sous-structure observée à ce jour. Celles du modèle standard sont présentées en figure 1.1. Leur masse, comme les autres grandeurs physiques de cette thèse, sont exprimées en unités naturelles, différentes des unités du système international comme exposé dans l'annexe A. Ces particules peuvent être classées selon leurs propriétés. La première d'entre elles, le « spin », est une observable quantique intrinsèque aux particules.

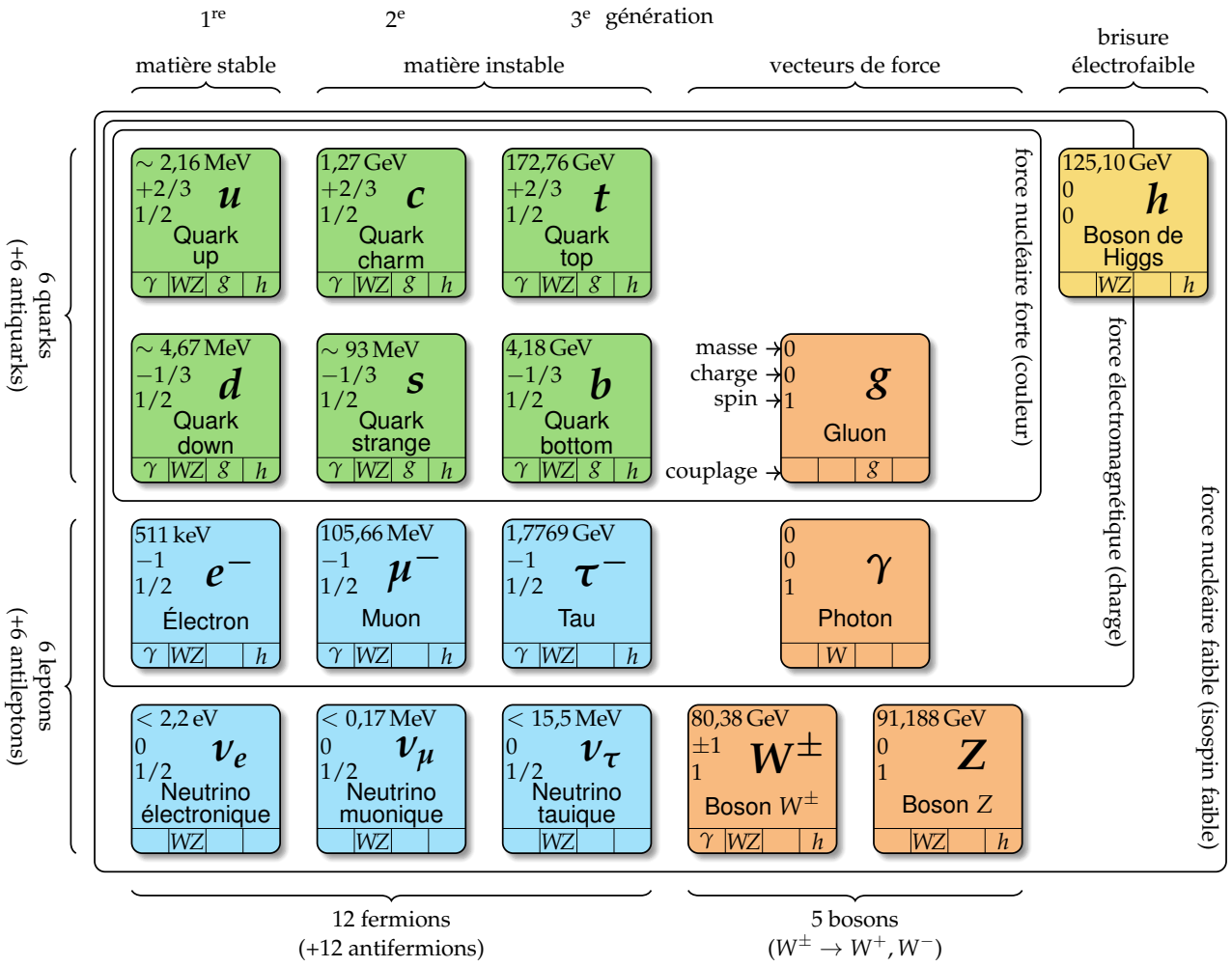


Figure 1.1 – *Particules élémentaires du modèle standard. Les quarks (en vert) et les leptons (en bleu) sont des fermions. Les bosons vecteurs (en orange) sont les médiateurs des forces fondamentales. Le boson de Higgs est quant à lui un boson scalaire (en jaune).*

1.1 Fermions

Les fermions sont les particules élémentaires de spin demi-entier et suivent donc la statistique de Fermi-Dirac. Ainsi, deux fermions ne peuvent pas occuper le même état quantique, c'est-à-dire avoir les mêmes nombres quantiques, comme exposé par le principe d'exclusion de Pauli. Le modèle standard comprend douze fermions constituant la matière, accompagnés de douze antifermions correspondants pour l'antimatière.

Les fermions peuvent se diviser d'une part en deux catégories, les quarks et les leptons, et d'autre part en trois catégories correspondant à trois « générations », comme illustré sur la figure 1.1. La première génération (quarks u et d, électron e^- et neutrino électronique ν_e) correspond aux particules les plus communes ; les deuxièmes et troisièmes générations contiennent des particules analogues, plus massives et instables.

1.1.1 Quarks

Les quarks sont les fermions possédant une charge de couleur. Il existe deux quarks par génération, un quark de type *up* et un quark de type *down*, formant un doublet d'isospin faible comme exposé en section 2.3. Il y a donc six quarks au total. Les quarks de type *up* (u, c et t) portent une charge électrique $+\frac{2}{3}e$ avec e la charge électrique élémentaire, les quarks de type *down* (d, s et b) une charge $-\frac{1}{3}e$. Les antiquarks possèdent une charge électrique opposée ($-\frac{2}{3}e$ et $+\frac{1}{3}e$). Ces particules sont donc sensibles à l'interaction électromagnétique.

À l'instar de la charge électrique pour l'interaction électromagnétique, la « couleur »rend les quarks sensibles à l'interaction forte. La charge de couleur peut prendre trois valeurs orthogonales, nommées par convention rouge, verte et bleue. Les particules portant une charge de couleur ne sont pas stables seules et se regroupent pour former des particules composites de charge de couleur nulle, ou de couleur « blanche ». C'est ce que l'on appelle le phénomène de « confinement de couleur », décrit dans la section 2.5.3.

Les particules composées de quarks sont les hadrons. Ces particules sont de couleur blanche, ce qui peut être obtenu de deux manières :

— par association d'un quark rouge, un vert et un bleu ; il s'agit d'un « baryon ». Le proton (uud) et le neutron (udd) sont deux exemples de baryons.

— par association d'un quark et d'un antiquark ; il s'agit d'un « méson ». En effet, un antiquark porte une « anticouleur ». Par exemple, un quark up (u) rouge et un antiquark down ($\bar{d}$) « antirouge » forment un pion chargé π^+. Une combinaison $u\bar{s}$ est un kaon K^+.

Enfin, comme tous les fermions, les quarks sont également sensibles à l'interaction faible. Les quarks sont ainsi les seules particules sensibles à toutes les interactions fondamentales décrites par le modèle standard.

1.1.2 Leptons

Les leptons sont les fermions ne possédant pas de charge de couleur. Ils sont donc insensibles à l'interaction forte. En revanche, ils sont tous sensibles à l'interaction faible. Sur le même principe que pour les quarks, il y a un doublet d'isospin faible de deux leptons par génération, soit six leptons au total. Les leptons d'isospin faible haut sont l'électron (e^-), le muon (μ^-) et le tau (τ^-), ils portent une charge électrique $-e$ ($+e$ pour les antiparticules correspondantes). Les leptons d'isospin faible bas sont les neutrinos. Ces derniers ne portent pas de charge électrique et interagissent donc uniquement par interaction faible, ce qui en fait des particules difficiles à détecter.

1.2 Bosons

Les bosons sont les particules élémentaires de spin entier. Ils suivent la statistique de Bose-Einstein qui n'interdit pas la présence de plusieurs bosons dans le même état quantique, contrairement à la statistique de Fermi-Dirac.

Les bosons de spin 1 sont les bosons de jauge, ou bosons vecteurs, et sont les médiateurs des interaction fondamentales. Ainsi, le photon (γ) est le boson vecteur de l'interaction électromagnétique. Il est de masse nulle et est électriquement neutre. Les bosons W^+, W^- et Z sont ceux de l'interaction faible. Le boson Z est électriquement neutre et de masse $m_Z = 91{,}19\,\text{GeV}$, les bosons W portent une charge électrique de $\pm e$, ont une masse de $m_W = 80{,}38\,\text{GeV}$ et n'interagissent qu'avec les particules de chiralité gauche et les antiparticules de chiralité droite. La chiralité est définie dans la section 2.3. Enfin, huit gluons (g) sont les médiateurs de l'interaction forte. Ils n'ont ni masse ni charge électrique, mais portent une charge de couleur et une charge d'anticouleur. Un gluon peut donc être

chargé « rouge et antibleu ». Si un tel gluon interagit avec un quark bleu, par conservation, ce quark devient rouge après interaction.

Les bosons de spin 0 sont dits scalaires. Le seul scalaire élémentaire observé à ce jour est le boson de Higgs, le modèle standard n'en prédit pas d'autre. Comme exposé en section 2.3, les forces électromagnétique et faible peuvent être unifiées en une seule force électrofaible. La symétrie associée à la force électrofaible est spontanément brisée par le mécanisme de Higgs, présenté en section 2.4. Les forces électromagnétique et faible sont alors retrouvées, les particules acquièrent une masse et le boson de Higgs émerge de ce mécanisme.

2 Formalisme théorique et interactions

Il ne suffit pas de lister les particules élémentaires pour obtenir un modèle, il faut également décrire leur dynamique, c'est-à-dire la façon dont ces particules interagissent. Pour cela, le modèle standard se base sur la théorie quantique des champs. Une particule n'est pas un « objet ponctuel » comme en mécanique classique, mais une excitation d'un champ quantique relativiste. Il s'agit alors de décrire l'évolution de ces excitations.

Le comportement des particules élémentaires est obtenu par application des équations d'Euler-Lagrange au lagrangien du modèle standard. Afin de comprendre pourquoi ce formalisme mathématique permet effectivement d'obtenir l'évolution des champs quantiques relativistes décrivant les particules, la section suivante consiste en une brève introduction du lagrangien dans le cas de la mécanique classique, suivie d'une généralisation au cas du modèle standard.

2.1 Introduction au formalisme du lagrangien

2.1.1 Lagrangien et principe de moindre action

Soit une particule de masse m, soumise à une force F prise ici constante à des fins d'illustration, se déplaçant dans le temps le long d'une dimension x, d'un point A à $t = 0$ à un point B à $t = \tau$, comme illustré sur la figure 1.2a.

La trajectoire de cette particule peut être déterminée à l'aide du principe fondamental de la dynamique, ou seconde loi de Newton, qui s'exprime dans ce cas sous la forme

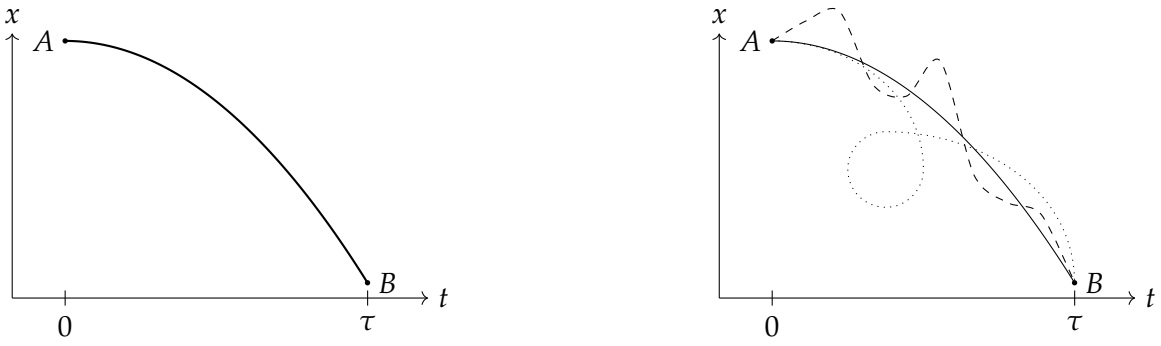
$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = F\,. \tag{1.1}$$

La position de la particule à tout instant est alors obtenue par double intégration selon le temps,

$$x(t) = x_0 + v_0 t + \frac{1}{2}\frac{F}{m}t^2 \tag{1.2}$$

avec x_0 et v_0 les position et vitesse de la particule à l'instant $t = 0$.

Or, cette méthode ne permet pas de décrire le comportement des particules élémentaires car à leur échelle la mécanique quantique prévaut. Il n'est alors pas possible, lorsque l'on observe une particule



(a) *Particule classique.* **(b)** *Particule quantique.*

Figure 1.2 – *Trajectoires d'une particule se déplaçant au cours du temps d'un point A à un point B.*

à un point A puis à un point B, de déterminer la trajectoire exacte suivie par cette particule. En effet, une particule quantique détectée en A à $t = 0$ puis en B à $t = \tau$ a suivi *toutes les trajectoires* reliant ces deux points aussi farfelues soient-elles, y compris celles revenant en arrière dans le temps, comme illustré en figure 1.2b. Utiliser une loi pouvant donner $x(t)$ n'est donc pas pertinent en mécanique quantique.

Il existe toutefois un autre principe physique toujours en place en mécanique quantique, la conservation de l'énergie. Dans le cas de la particule précédemment décrite, il s'agit de la somme de son énergie cinétique T et de son énergie potentielle V, c'est-à-dire

$$E = T + V = \mathrm{C^{te}} \tag{1.3}$$

où T dépend uniquement de la vitesse de la particule et V uniquement de sa position. Il en va ainsi de même pour les moyennes temporelles de ces grandeurs,

$$E = \langle E \rangle = \langle T \rangle + \langle V \rangle = \mathrm{C^{te}} \tag{1.4}$$

avec, en notant $\dot{x} = \frac{\mathrm{d}x}{\mathrm{d}t}$,

$$\langle T \rangle = \frac{1}{\tau} \int_0^\tau T(\dot{x}(t))\,\mathrm{d}t\ , \quad \langle V \rangle = \frac{1}{\tau} \int_0^\tau V(x(t))\,\mathrm{d}t\ . \tag{1.5}$$

Lorsque la trajectoire suivie par la particule varie par rapport à la trajectoire déterminée par la mécanique Newtonienne, ce qui est le cas en mécanique quantique, ces grandeurs sont modifiées. La variation de la valeur moyenne de l'énergie potentielle s'exprime à partir de la dérivée fonctionnelle

$$\frac{\delta\langle V \rangle}{\delta x(t)} = \frac{1}{\tau} \int_0^\tau \frac{\delta V(x(t'))}{\delta x(t)}\,\mathrm{d}t' = \frac{1}{\tau} \int_0^\tau \frac{\mathrm{d}V(x(t'))}{\mathrm{d}x(t')} \delta(t' - t)\,\mathrm{d}t' = \frac{1}{\tau} \left.\frac{\mathrm{d}V}{\mathrm{d}x(t')}\right|_{t'=t} = -\frac{1}{\tau} F(x(t)) \tag{1.6}$$

car la force F est reliée à V par $F = -\frac{\mathrm{d}V}{\mathrm{d}x}$. De même, l'énergie cinétique moyenne varie selon

$$\begin{aligned} \frac{\delta\langle T \rangle}{\delta x(t)} &= \frac{1}{\tau} \int_0^\tau \frac{\delta T(\dot{x}(t'))}{\delta x(t)}\,\mathrm{d}t' = \frac{1}{\tau} \int_0^\tau \frac{\mathrm{d}T(\dot{x}(t'))}{\mathrm{d}x(t')} \delta'(t' - t)\,\mathrm{d}t' = -\frac{1}{\tau} \int_0^\tau \delta(t' - t) \frac{\mathrm{d}}{\mathrm{d}t'} \left(\frac{\mathrm{d}T(\dot{x}(t'))}{\mathrm{d}\dot{x}(t')} \right) \mathrm{d}t' \\ &= -\frac{1}{\tau} \frac{\mathrm{d}}{\mathrm{d}t'} \left(\frac{\mathrm{d}T(\dot{x}(t'))}{\mathrm{d}\dot{x}(t')} \right) \Bigg|_{t'=t} = -\frac{1}{\tau} m \left.\frac{\mathrm{d}^2 x}{\mathrm{d}t'^2}\right|_{t'=t} = -\frac{1}{\tau} m \frac{\mathrm{d}^2 x}{\mathrm{d}t^2}(t) \end{aligned} \tag{1.7}$$

car pour une particule de masse m, en mécanique newtonienne, $T = \frac{1}{2} m \left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2$.

Le long de la trajectoire classique, le principe fondamental de la dynamique est vérifié. Alors, les variations autour de la trajectoire classique sont reliées par

$$m \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = F \Leftrightarrow \frac{\delta\langle T \rangle}{\delta x(t)} = \frac{\delta\langle V \rangle}{\delta x(t)} \Leftrightarrow \frac{\delta}{\delta x(t)} (\langle T \rangle - \langle V \rangle) = 0\,, \tag{1.8}$$

c'est-à-dire que la différence entre l'énergie cinétique et l'énergie potentielle du système étudié est stationnaire le long de la trajectoire classique. Cette différence est le lagrangien L,

$$L = T - V\,. \tag{1.9}$$

L'intégrale au cours du temps du lagrangien est appelée action et est définie comme

$$S = \int_0^\tau \mathrm{d}t\, L\,. \tag{1.10}$$

Compte-tenu de l'équation (1.8), l'action vérifie pour la trajectoire classique

$$\frac{\delta S}{\delta x(t)} = 0\,, \tag{1.11}$$

ce qui est connu sous le nom de principe de moindre action.

2.1.2 Cas de la mécanique quantique

En mécanique quantique, la probabilité $P(B,\tau|A,0)$ d'observer la particule en B à $t=\tau$ sachant qu'elle était en A à $t=0$ s'exprime à partir de l'opérateur $\hat{U}$ d'évolution temporelle,

$$P(B,\tau|A,0) = \left|\left\langle B\right|\hat{U}(\tau,0)\left|A\right\rangle\right|^2 . \tag{1.12}$$

Or, lors du calcul de l'amplitude $\left\langle B\right|\hat{U}(\tau,0)\left|A\right\rangle$, chacune des trajectoires possibles apporte une contribution avec un facteur de phase s'exprimant

$$\exp(\mathrm{i}S/\hbar) \tag{1.13}$$

avec S l'action de la trajectoire et $\hbar$ la constante de Planck réduite. Il est ainsi pertinent de déterminer les variations de S lorsque la trajectoire de la particule est modifiée.

Ces variations s'expriment

$$\frac{\delta S}{\delta x(t)} = \int_0^\tau \mathrm{d}t' \left[\frac{\delta L}{\delta x(t')}\delta(t'-t) + \frac{\delta L}{\delta \dot{x}(t')}\delta'(t'-t)\right] = \frac{\delta L}{\delta x(t)} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\delta L}{\delta \dot{x}(t)} . \tag{1.14}$$

Cette quantité est nulle autour de la trajectoire classique d'après le principe de moindre action (1.11). La phase est alors stationnaire, ce qui permet aux trajectoires quantiques proches de la trajectoire classique de contribuer plus fortement à l'amplitude. Le principe de moindre action peut se ré-exprimer avec le lagrangien L selon

$$\frac{\delta L}{\delta x(t)} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\delta L}{\delta \dot{x}(t)} = 0 . \tag{1.15}$$

Cette équation est l'équation d'Euler-Lagrange et permet d'obtenir toutes les équations du mouvement du système, c'est-à-dire de décrire son évolution au cours du temps. Connaître le lagrangien permet donc de décrire l'évolution des particules.

2.1.3 Lagrangien, champs et symétries

Le raisonnement précédent peut se généraliser à un espace à une dimension temporelle et trois dimensions spatiales. À partir du lagrangien, il est possible de définir la densité lagrangienne $\mathcal{L}$ telle que

$$L = \int \mathrm{d}^3x\, \mathcal{L} , \quad S = \int \mathrm{d}^4x\, \mathcal{L} \tag{1.16}$$

où x désigne la coordonnée dans l'espace-temps de Minkowski, c'est-à-dire l'espace-temps à quatre dimensions. Si la densité lagrangienne dépend d'un champ $\phi(x)$ et de ses dérivées $\partial_\mu\phi(x)$, avec ∂_μ la dérivée partielle par rapport à la coordonnée μ dans l'espace-temps de Minkowski, alors

$$S = \int \mathrm{d}^4x\, \mathcal{L}(\phi(x), \partial_\mu\phi(x)) \tag{1.17}$$

et du principe de moindre action résultent les équations d'Euler-Lagrange pour $\mathcal{L}$, une pour chaque valeur possible de μ,

$$\frac{\delta S}{\delta \phi} = \frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu\phi)} = 0 . \tag{1.18}$$

Il s'agit à présent de déterminer la densité lagrangienne $\mathcal{L}$, communément nommée lagrangien du modèle standard.

Un champ quantique peut subir une transformation de jauge locale. Une telle transformation doit laisser la physique inchangée, ainsi le lagrangien du modèle standard est construit pour être invariant sous les transformations de jauges locales du groupe de symétrie

$$SU(3)_C \times SU(2)_L \times U(1)_Y . \tag{1.19}$$

De cette construction résultent les interactions fondamentales, discutées ci-après.

2.2 Interaction électromagnétique

Le lagrangien libre d'un fermion, c'est-à-dire le lagrangien décrivant le comportement d'un fermion seul, s'exprime

$$\mathcal{L}_{\text{fermion libre}} = \bar{\psi}\left(\mathrm{i}\not{\partial} - m\right)\psi = \bar{\psi}\left(\mathrm{i}\gamma^{\mu}\partial_{\mu} - m\right)\psi = \mathrm{i}\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi \tag{1.20}$$

où la notation « slash » pour un objet k signifie $\not{k} = \gamma^{\mu}k_{\mu}$, ψ le spineur de Dirac correspondant au champ fermionique, $\bar{\psi} = \psi^{\dagger}\gamma^{0}$ son adjoint de Dirac, $\psi^{\dagger}$ étant l'adjoint de ψ, γ^{μ} les matrices de Dirac, définies dans l'annexe B et m la masse de la particule considérée. Le terme $\gamma^{\mu}\partial_{\mu}$, par convention de sommation d'Einstein exposée dans l'annexe B, correspond à une somme sur les différentes valeur de μ.

Le lagrangien $\mathcal{L}_{\text{fermion libre}}$ est invariant sous une transformation globale du groupe $U(1)_{em}$, c'est-à-dire lorsque la transformation suivante est appliquée au spineur ψ

$$\psi \to \mathrm{e}^{\mathrm{i}Q\alpha}\psi\,, \quad \bar{\psi} \to \bar{\psi}\mathrm{e}^{-\mathrm{i}Q\alpha} \tag{1.21}$$

où $\alpha \in \mathbb{R}$ et Q est l'opérateur de charge électrique. En effet, sous une telle transformation,

$$\bar{\psi}\psi \to \bar{\psi}\mathrm{e}^{-\mathrm{i}Q\alpha}\mathrm{e}^{\mathrm{i}Q\alpha}\psi = \bar{\psi}\psi \tag{1.22}$$

et

$$\mathrm{i}\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi \to \mathrm{i}\bar{\psi}\mathrm{e}^{-\mathrm{i}Q\alpha}\gamma^{\mu}\partial_{\mu}\left(\mathrm{e}^{\mathrm{i}Q\alpha}\psi\right) = \mathrm{i}\bar{\psi}\mathrm{e}^{-\mathrm{i}Q\alpha}\mathrm{e}^{\mathrm{i}Q\alpha}\gamma^{\mu}\partial_{\mu}\left(\psi\right) + \mathrm{i}\bar{\psi}\mathrm{e}^{-\mathrm{i}Q\alpha}\gamma^{\mu}\partial_{\mu}\cancel{\left(\mathrm{e}^{\mathrm{i}Q\alpha}\right)}\psi = \mathrm{i}\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi \tag{1.23}$$

car α ne dépend pas de l'espace-temps pour une transformation globale.

En revanche, pour une transformation locale,

$$\mathrm{i}\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi \to \mathrm{i}\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi + \mathrm{i}\mathrm{e}^{-\mathrm{i}Q\alpha}\bar{\psi}\gamma^{\mu}\partial_{\mu}\left(\mathrm{e}^{\mathrm{i}Q\alpha}\right)\psi = \mathrm{i}\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - \bar{\psi}\gamma^{\mu}Q\partial_{\mu}\alpha\psi \tag{1.24}$$

ce qui fait apparaître un terme supplémentaire, $\bar{\psi}\gamma^{\mu}Q\partial_{\mu}\alpha\psi$, provenant de la transformation du terme $\mathrm{i}\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi$ de $\mathcal{L}_{\text{fermion libre}}$ qui brise ainsi l'invariance de jauge du lagrangien. Afin de rendre le lagrangien invariant sous les transformations locales du groupe $U(1)_{em}$, il est possible de remplacer la dérivée usuelle ∂_{μ} par la dérivée covariante D_{μ}, telle que

$$\partial_{\mu} \to D_{\mu} = \partial_{\mu} - \mathrm{i}eQA_{\mu} \tag{1.25}$$

où e est la charge électrique élémentaire et A_{μ} un champ de jauge nouvellement introduit, dont la transformation de jauge permet de supprimer le terme supplémentaire qui brise l'invariance de jauge du lagrangien. En effet, le champ A_{μ} se transforme tel que

$$A_{\mu} \to A_{\mu} + \frac{1}{e}\partial_{\mu}\alpha\,. \tag{1.26}$$

Ainsi, en réécrivant le lagrangien du fermion de l'équation (1.20) avec la dérivée covariante,

$$\mathcal{L}'_{\text{fermion libre}} = \mathrm{i}\bar{\psi}\gamma^{\mu}D_{\mu}\psi - m\bar{\psi}\psi = \mathrm{i}\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi + \bar{\psi}\gamma^{\mu}eQA_{\mu}\psi = \mathcal{L}_{\text{fermion libre}} + \bar{\psi}\gamma^{\mu}eQA_{\mu}\psi\,, \tag{1.27}$$

le dernier terme se transforme en

$$\bar{\psi}\gamma^{\mu}eQA_{\mu}\psi \to \bar{\psi}\mathrm{e}^{-\mathrm{i}Q\alpha}\gamma^{\mu}eQ\left(A_{\mu} + \frac{1}{e}\partial_{\mu}\alpha\right)\mathrm{e}^{\mathrm{i}Q\alpha}\psi = \bar{\psi}\gamma^{\mu}eQA_{\mu}\psi + \bar{\psi}\gamma^{\mu}Q\partial_{\mu}\alpha\psi \tag{1.28}$$

et le dernier terme obtenu compense exactement celui brisant l'invariance de jauge dans l'équation (1.24).

Le nouveau terme introduit par l'utilisation de la dérivée covariante, $\bar{\psi}\gamma^{\mu}eQA_{\mu}\psi$, correspond à l'interaction entre un fermion et le champ de jauge A_{μ}, dont l'intensité est directement proportionnelle à la charge électrique du fermion. Toutefois, le champ A_{μ} ne représente pas encore le photon en l'état, il faut permettre au photon de se propager librement. Pour cela, un terme cinétique invariant de jauge dans le lagrangien doit être introduit, ce qui peut se faire avec

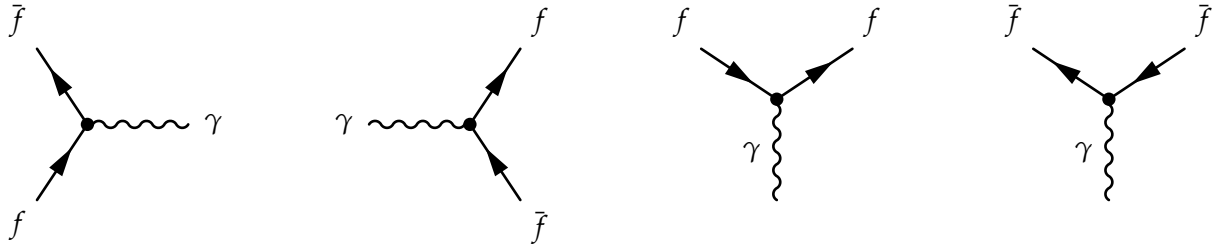
$$\mathcal{L}_{\text{photon libre}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}\,, \quad F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}\,. \tag{1.29}$$

Un terme de masse pour le champ A_μ devrait s'écrire sous la forme $\frac{1}{2}m^2 A^\mu A_\mu$, ce qui n'est pas invariant de jauge. Par conséquent, le champ A_μ est de masse nulle.

Le lagrangien complet pour l'interaction électromagnétique, aussi nommé QED (électrodynamique quantique, *Quantum Electro-Dynamics*), s'exprime alors

$$\mathcal{L}_{QED} = \underbrace{\bar{\psi}\left(\mathrm{i}\not{D} - m\right)\psi}_{\text{fermions}} \underbrace{-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{photons}} = \underbrace{\bar{\psi}\left(\mathrm{i}\not{\partial} - m\right)\psi}_{\mathcal{L}_{\text{fermion libre}}} \underbrace{+\bar{\psi}\gamma^\mu eQA_\mu\psi}_{\text{interaction}} \underbrace{-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\mathcal{L}_{\text{photon libre}}} . \tag{1.30}$$

Le terme d'interaction dans ce lagrangien permet de « connecter » les fermions aux photons dans les diagrammes de Feynman, dont le principe est décrit dans l'annexe C. La « connexion » ainsi obtenue est nommée « vertex ». La structure du terme d'interaction, $\bar{\psi}\gamma^\mu eQA_\mu\psi$, impose ainsi la présence au vertex d'un photon (A_μ), d'un fermion entrant ou d'un antifermion sortant (ψ) et d'un fermion sortant ou d'un antifermion entrant ($\bar{\psi}$). Ces possibilités donnent les diagrammes de la figure 1.3.



(a) *Un fermion f et un antifermion $\bar{f}$ s'annihilent en un photon γ.*

(b) *Un photon donne une paire de fermion et antifermion.*

(c) *Un fermion interagit avec un photon.*

(d) *Un antifermion interagit avec un photon.*

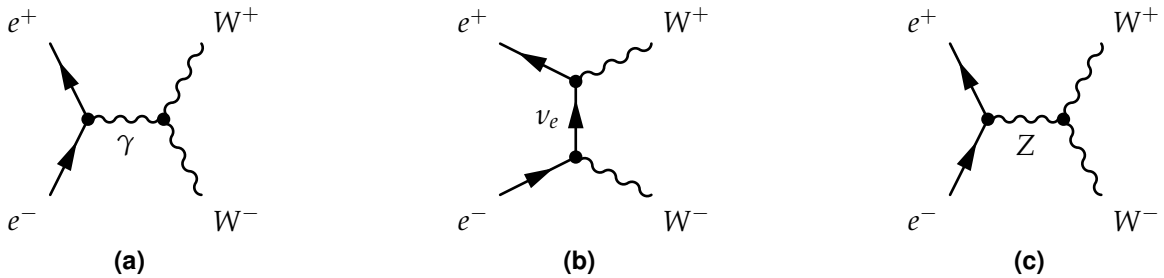
Figure 1.3 – *Diagrammes de Feynman possibles à partir du terme $\bar{\psi}\gamma^\mu eQA_\mu\psi$ du lagrangien $\mathcal{L}_{QED}$.*

Maintenir l'invariance de jauge locale à l'aide de la dérivée covariante fait émerger l'interaction électromagnétique dans le cas de l'invariance de jauge sous $U(1)_{em}$. Dans les sections suivantes, un raisonnement similaire est appliqué afin d'obtenir les interactions électrofaible et forte.

2.3 Interaction électrofaible

Le modèle standard décrit les interactions électromagnétique et faible comme deux facettes d'une seule et même interaction qui les unifie, l'interaction électrofaible, notée « EW » pour *electroweak*. Celle-ci a été développée à partir des travaux de GLASHOW [1], SALAM & WARD [2] et WEINBERG [3], récompensés par le prix Nobel de physique en 1979.

Une des raisons de l'unification de ces deux forces provient du calcul de la section efficace de production de paire W^+W^-. Le lien entre la section efficace d'un processus physique et le nombre d'événements de ce processus est introduit dans le chapitre 2. La section efficace ne peut pas être infinie, ce ne saurait correspondre à la réalité physique. Or, sans considérer chacun des diagrammes de la figure 1.4 pour ce calcul, la section efficace de production de paire W^+W^- est infinie. La prise en compte des interactions électromagnétique, figure 1.4a, et faible, figures 1.4b et 1.4c, pousse à les unifier.



(a) **(b)** **(c)**

Figure 1.4 – *Diagrammes de Feynman de production de paire W^+W^- à l'arbre.*

L'interaction électromagnétique repose sur l'invariance de jauge sous les transformations locales du groupe $U(1)_{em}$. Dans le cas de l'interaction électrofaible, ce groupe de symétrie est $SU(2)_L \times U(1)_Y$.

Dans un premier temps, seul le cas de $SU(2)_L$ avec les leptons est traité et permet de soulever toute la richesse de ce groupe par rapport à $U(1)$. Ensuite, $SU(2)_L \times U(1)_Y$ est abordé, toujours avec les leptons. Puis, le traitement des quarks est présenté, donnant alors une description de l'interaction électrofaible.

2.3.1 Symétrie $SU(2)_L$ et chiralité

Une des propriétés les plus importantes de l'interaction faible est de violer la symétrie de parité, notée P. La symétrie P consiste à remplacer une des trois coordonnées spatiales par son opposé, comme le fait un miroir. La violation de P par l'interaction faible, observée expérimentalement avec les désintégrations β [4], peut être expliquée en considérant que les couplages aux bosons W ne se font qu'avec certains états de chiralité.

La chiralité est une propriété des particules. Pour un fermion décrit par un spineur ψ, elle peut être droite (*right*) ou gauche (*left*). La composante de chiralité gauche ψ_L s'obtient à l'aide du projecteur chiral γ^5 selon

$$\psi_L = \frac{1}{2}(1-\gamma^5)\psi\,, \tag{1.31}$$

celle de chiralité droite ψ_R selon

$$\psi_R = \frac{1}{2}(1+\gamma^5)\psi\,. \tag{1.32}$$

Pour les antiparticules décrites par $\bar{\psi} = \psi^\dagger\gamma^0$,

$$\overline{\psi_L} = (\psi_L)^\dagger\gamma^0 = \left(\frac{1}{2}(1-\gamma^5)\psi\right)^\dagger\gamma^0 = \psi^\dagger\frac{1}{2}(1-\gamma^5)\gamma^0 = \psi^\dagger\gamma^0\frac{1}{2}(1+\gamma^5) = \bar{\psi}_R\,, \tag{1.33}$$

et de même, $\overline{\psi_R} = \bar{\psi}_L$.

Les expériences montrent que les bosons W ne sont couplés qu'aux fermions de chiralité gauche et aux antifermions de chiralité droite, ce qui correspond à introduire un facteur $\gamma^\mu(1-\gamma^5)$ aux termes de couplage correspondants dans le lagrangien. Dans la notation $SU(2)_L$, L signifie ainsi *left*. Les couplages étant différents selon l'état de parité, la symétrie P est violée. Les couplages du boson Z ne sont pas purement en $(1-\gamma^5)$, il est donc quand même couplé aux fermions de chiralité droite et aux antifermions de chiralité gauche. Ce comportement du Z peut être expliqué dans le cadre de la force électrofaible, la force faible à elle seule ne permet pas d'en rendre compte.

La violation de P par l'interaction faible a également pour conséquence la violation de C, la symétrie de charge. En effet, C change une particule de chiralité gauche en une antiparticule de même chiralité, dont les couplages aux bosons de l'interaction faible ne sont pas les mêmes. En revanche, la symétrie CP change une particule de chiralité gauche en une antiparticule de chiralité droite, ce qui semble être conservé par l'interaction faible. L'étude des désintégrations des kaons a toutefois montré que l'interaction faible viole également la symétrie CP [5], ce qui a pu être expliqué théoriquement en postulant l'existence d'une troisième génération de quarks [6] observée depuis.

L'introduction de la symétrie $SU(2)_L$ amène un nouveau nombre quantique, l'« isospin faible », noté I. Il se comporte mathématiquement comme le spin des particules, d'où son nom. Les fermions de chiralité gauche sont rassemblés en doublets d'isospin faible $I = \frac{1}{2}$, les fermions de chiralité droite en singlets d'isospin faible $I = 0$. Ces derniers sont ainsi invariants sous les transformations de $SU(2)_L$, ce qui se traduit physiquement par une insensibilité à l'interaction faible.

Les fermions peuvent être de chiralité droite ou gauche. Dans le cadre actuel du modèle standard, les neutrinos existent toutefois uniquement avec une chiralité gauche. Il n'y a à ce jour aucune raison pour les neutrinos de chiralité droite de ne pas exister. Cependant, ils n'interagissent pas, par construction, avec la matière dans le cadre du modèle standard. Ainsi, il est possible de les retirer du modèle tout en conservant une description cohérente du comportement des particules. Cela mène aux représentations du tableau 1.1.

I	Quarks gauches	Quarks droits	Leptons gauches	Leptons droits
$\frac{1}{2}$	$\begin{pmatrix} u_i \\ d_i \end{pmatrix}_L$	-	$\begin{pmatrix} \nu_i \\ \ell_i \end{pmatrix}_L$	-
0	-	$u_{i,R}$, $d_{i,R}$	-	$\ell_{i,R}$

Tableau 1.1 – *Représentation des fermions selon leur chiralité et leur isospin faible. L'indice $i \in \{1,2,3\}$ correspond à la génération des particules. Ainsi, les symboles u_i, d_i, ℓ_i et ν_i correspondent, respectivement, aux quarks d'isospin faible haut (u, c, t), d'isospin faible bas (d, s, b), aux leptons chargés (e, μ, τ) et aux neutrinos (ν_e, ν_μ, ν_τ).*

2.3.2 Symétrie $SU(2)$ et interactions entre bosons

Afin d'alléger les notations, le cas plus général d'un groupe de symétrie $SU(2)$ est traité. Pour étendre les résultats à $SU(2)_L$, il suffit de se souvenir que les couplages ont uniquement lieu entre fermions de chiralité gauche et antifermions de chiralité droite. La méthode reste la même que pour l'électromagnétisme. Sous une transformation de $SU(2)$, les spineurs se transforment selon

$$\psi \to \mathrm{e}^{\frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)}\psi\,, \quad \bar{\psi} \to \bar{\psi}\mathrm{e}^{-\frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)} \tag{1.34}$$

où $\boldsymbol{\alpha} \in \mathbb{R}^3$ et $\boldsymbol{\tau}$ est un vecteur dont les composantes τ_i sont les générateurs de $SU(2)$. Ces générateurs sont des matrices 2×2 s'identifiant aux matrices de Pauli σ_i définies dans l'annexe B. Toutefois, ces générateurs agissent dans le cas de $SU(2)_L$ sur les doublets d'isospin faible alors que les matrices de Pauli agissent sur le spin d'un fermion. Afin d'éviter les confusions, la notation τ est utilisée. L'équation (1.34) est l'analogue directe de l'équation (1.22).

Afin de simplifier les calculs qui suivent, seules les transformations infinitésimales sont considérées. En effet, $SU(2)$ est un groupe non abélien. Cela signifie que deux transformations successives a et b de ce groupe ne donnent pas le même résultat selon que soient appliquées a puis b ou b puis a, c'est-à-dire $ab - ba \neq 0$. Les transformations précédentes sous leurs formes infinitésimales, c'est-à-dire au premier ordre en $\boldsymbol{\alpha}$, s'expriment

$$\psi \to \left(1 + \frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)\right)\psi\,, \quad \bar{\psi} \to \bar{\psi}\left(1 - \frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)\right)\,. \tag{1.35}$$

Les termes du lagrangien du fermion libre, introduit dans l'équation (1.20), se transforment alors comme

$$-m\bar{\psi}\psi \to -m\bar{\psi}\left(1 - \frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)\right)\left(1 + \frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)\right)\psi = -m\bar{\psi}\psi + \mathcal{O}(\alpha^2) \tag{1.36}$$

et

$$\begin{aligned} \mathrm{i}\bar{\psi}\gamma^\mu\partial_\mu\psi &\to \mathrm{i}\bar{\psi}\left(1 - \frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)\right)\gamma^\mu\partial_\mu\left(\left(1 + \frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)\right)\psi\right) \\ &= \mathrm{i}\bar{\psi}\gamma^\mu\partial_\mu\psi - \bar{\psi}\frac{1}{2}\boldsymbol{\tau}\cdot\gamma^\mu\partial_\mu\boldsymbol{\alpha}(x)\psi + \mathcal{O}(\alpha^2)\,, \end{aligned} \tag{1.37}$$

ce qui fait apparaître, sur le même principe qu'avec l'interaction électromagnétique, un terme supplémentaire brisant l'invariance de jauge du lagrangien. Une nouvelle dérivée covariante peut être définie afin de rétablir l'invariance de jauge,

$$D_\mu = \partial_\mu - \frac{\mathrm{i}}{2}g_I\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu(x)\,, \tag{1.38}$$

où l'on introduit g_I la constante de couplage d'isospin faible, ainsi que trois champs de jauge vectoriels $W^i_\mu(x)$, $i \in \{1,2,3\}$ se transformant tels que

$$\boldsymbol{W}_\mu \to \boldsymbol{W}_\mu + \frac{1}{g_I}\partial_\mu\boldsymbol{\alpha} - (\boldsymbol{\alpha} \wedge \boldsymbol{W}_\mu)\,. \tag{1.39}$$

Dans ce cas, le lagrangien du fermion libre se réécrit sous la forme

$$\begin{aligned} \mathcal{L}'_{\text{fermion libre}} &= \mathrm{i}\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi = \mathrm{i}\bar{\psi}\gamma^\mu\partial_\mu\psi + \bar{\psi}\gamma^\mu\frac{1}{2}g_I\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu\psi - m\bar{\psi}\psi \\ &= \mathcal{L}_{\text{fermion libre}} + \bar{\psi}\gamma^\mu\frac{1}{2}g_I\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu\psi\,. \end{aligned} \tag{1.40}$$

Ainsi, le terme supplémentaire du lagrangien se transforme tel que

$$\begin{aligned}\bar{\psi}\gamma^\mu\frac{1}{2}g_I\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu\psi \to \bar{\psi}\left(1-\frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)\right)\gamma^\mu\frac{g_I}{2}\boldsymbol{\tau}\cdot\left(\boldsymbol{W}_\mu+\frac{1}{g_I}\partial_\mu\boldsymbol{\alpha}-(\boldsymbol{\alpha}\wedge\boldsymbol{W}_\mu)\right)\left(1+\frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)\right)\psi\\ =\bar{\psi}\gamma^\mu\frac{g_I}{2}\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu\psi-\bar{\psi}\frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)\gamma^\mu\frac{g_I}{2}\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu\psi+\bar{\psi}\gamma^\mu\frac{g_I}{2}\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu\frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(x)\psi\\ +\bar{\psi}\gamma^\mu\frac{1}{2}\boldsymbol{\tau}\cdot\partial_\mu\boldsymbol{\alpha}\psi-\bar{\psi}\gamma^\mu\frac{g_I}{2}\boldsymbol{\tau}\cdot(\boldsymbol{\alpha}\wedge\boldsymbol{W}_\mu)\psi+\mathcal{O}(\alpha^2)\,.\end{aligned}\tag{1.41}$$

Or,

$$(\boldsymbol{\tau}\cdot\boldsymbol{a})(\boldsymbol{\tau}\cdot\boldsymbol{b})=(\boldsymbol{a}\cdot\boldsymbol{b})+\mathrm{i}\boldsymbol{\tau}\cdot(\boldsymbol{a}\wedge\boldsymbol{b})\Leftrightarrow\boldsymbol{\tau}\cdot(\boldsymbol{a}\wedge\boldsymbol{b})=\mathrm{i}[(\boldsymbol{a}\cdot\boldsymbol{b})-(\boldsymbol{\tau}\cdot\boldsymbol{a})(\boldsymbol{\tau}\cdot\boldsymbol{b})]\,.\tag{1.42}$$

Ainsi,

$$\begin{aligned}\boldsymbol{\tau}\cdot(\boldsymbol{\alpha}\wedge\boldsymbol{W}_\mu)&=\frac{1}{2}\left[\boldsymbol{\tau}\cdot(\boldsymbol{\alpha}\wedge\boldsymbol{W}_\mu)-\boldsymbol{\tau}\cdot(\boldsymbol{W}_\mu\wedge\boldsymbol{\alpha})\right]\\ &=\frac{\mathrm{i}}{2}\left[[(\boldsymbol{\alpha}\cdot\boldsymbol{W}_\mu)-(\boldsymbol{\tau}\cdot\boldsymbol{\alpha})(\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu)]-[(\boldsymbol{W}_\mu\cdot\boldsymbol{\alpha})-(\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu)(\boldsymbol{\tau}\cdot\boldsymbol{\alpha})]\right]\\ &=\frac{\mathrm{i}}{2}\left[(\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu)(\boldsymbol{\tau}\cdot\boldsymbol{\alpha})-(\boldsymbol{\tau}\cdot\boldsymbol{\alpha})(\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu)\right]\,.\end{aligned}\tag{1.43}$$

En combinant les équations (1.41) et (1.43), il vient

$$\bar{\psi}\gamma^\mu\frac{1}{2}g_I\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu\psi\to\bar{\psi}\gamma^\mu\frac{g_I}{2}\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu\psi+\bar{\psi}\gamma^\mu\frac{1}{2}\boldsymbol{\tau}\cdot\partial_\mu\boldsymbol{\alpha}\psi+\mathcal{O}(\alpha^2)\,,\tag{1.44}$$

où le dernier terme obtenu compense exactement le terme brisant l'invariance de jauge dans l'équation (1.37).

À ce stade, l'analogie avec l'électromagnétisme pousse à introduire $\boldsymbol{W}_{\mu\nu}$ l'analogue à $F_{\mu\nu}$ tel que $\boldsymbol{W}_{\mu\nu}=\partial_\mu\boldsymbol{W}_\nu-\partial_\nu\boldsymbol{W}_\mu$. Or, les invariances de jauge imposées mènent à utiliser une définition légèrement différente,

$$\boldsymbol{W}_{\mu\nu}=\partial_\mu\boldsymbol{W}_\nu-\partial_\nu\boldsymbol{W}_\mu+g_I(\boldsymbol{W}_\mu\wedge\boldsymbol{W}_\nu)\,.\tag{1.45}$$

Le lagrangien pour $SU(2)$ s'écrit alors

$$\mathcal{L}_{SU(2)}=\bar{\psi}(\mathrm{i}\not{D}-m)\psi-\frac{1}{4}\boldsymbol{W}_{\mu\nu}\cdot\boldsymbol{W}^{\mu\nu}\,.\tag{1.46}$$

Une différence notable et importante vis-à-vis de $\mathcal{L}_{QCD}$ est la non linéarité de $\boldsymbol{W}_{\mu\nu}$ par rapport à $\boldsymbol{W}_\mu$ et $\boldsymbol{W}_\nu$. Cette composante non linéaire ouvre la porte aux interactions directes entre les champs W^i_μ, c'est-à-dire entre les bosons, ce qui était impossible avec QED. De nouveaux types de vertex, comme celui de la figure 1.5, sont donc possibles dans une théorie de jauge avec une symétrie locale $SU(2)$.

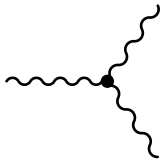


Figure 1.5 – *Diagramme de Feynman correspondant à l'interaction entre trois bosons.*

2.3.3 Symétrie $SU(2)_L\times U(1)_Y$ et unification électrofaible

Dans la notation $U(1)_Y$, Y est l'« hypercharge », reliée à Q la charge électrique et à I_3 la projection de l'isospin faible par la relation de Gell-Mann–Nishijima,

$$Q=I_3+\frac{Y}{2}\,,\tag{1.47}$$

dont les résultats pour les différents leptons sont présentés dans le tableau 1.2.

Les raisonnements réalisés précédemment peuvent ici être mis à profit. En effet, le cas de $U(1)_{em}$ est traité dans la section 2.2. Il est possible d'obtenir directement les mêmes résultats pour $U(1)_Y$ en procédant à l'analogie $U(1)_{em}\leftrightarrow U(1)_Y$, avec

$$A_\mu\leftrightarrow B_\mu\,,\quad F_{\mu\nu}\leftrightarrow F^{(B)}_{\mu\nu}\,,\quad e\leftrightarrow g_Y\,,\quad Q\leftrightarrow\frac{1}{2}Y\,.\tag{1.48}$$

De plus, sachant que $SU(2)_L$ couple les fermions de chiralité gauche et les antifermions de chiralité droite, les résultats pour $SU(2)$ sont directement utilisables en ajoutant les projections décrites par les équations (1.31) et (1.32). Cependant, les termes de masse des fermions, de la forme $\bar{\psi}\psi$, ne sont pas invariants de jauge sous $SU(2)_L\times U(1)_Y$. Les masses de ces particules sont donc nulles, à moins d'introduire un nouvel élément au lagrangien.

La dérivée covariante pour $SU(2)_L \times U(1)_Y$ est, d'après l'analogie $U(1)_{em} \leftrightarrow U(1)_Y$,

$$D_\mu = \partial_\mu - \mathrm{i}g_I I\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu - \frac{\mathrm{i}}{2}g_Y Y B_\mu\,. \tag{1.49}$$

Elle peut agir sur un doublet d'isospin faible, noté L, ou un singlet d'isospin faible, noté R, selon

$$D_\mu L = \left[\partial_\mu - \frac{\mathrm{i}}{2}g_I\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu + \frac{\mathrm{i}}{2}g_Y B_\mu\right] L\,, \tag{1.50}$$

$$D_\mu R = \left[\partial_\mu + \mathrm{i}g_Y B_\mu\right] R\,, \tag{1.51}$$

compte-tenu des différentes valeurs de Y et I données dans le tableau 1.2.

Champ	ν_e	e_L	e_R
Y	-1	-1	-2
I	$\frac{1}{2}$	$\frac{1}{2}$	0
I_3	$\frac{1}{2}$	$-\frac{1}{2}$	0
Q	0	-1	-1

Tableau 1.2 – *Valeurs des hypercharges, isospins faibles et charges électriques pour les leptons.*

Le lagrangien invariant sous $SU(2)_L \times U(1)_Y$ de l'interaction électrofaible s'écrit alors

$$\mathcal{L}_{EW} = \mathrm{i}\bar{\psi}\not{D}\psi - \frac{1}{4}\boldsymbol{W}_{\mu\nu}\cdot\boldsymbol{W}^{\mu\nu} - \frac{1}{4}F^{(B)}_{\mu\nu}\cdot F^{(B)\mu\nu}\,. \tag{1.52}$$

Attention toutefois, les champs B_μ et W^i_μ ne correspondent pas, respectivement, au photon et aux bosons $W^\pm$ et Z. Ces quatre bosons sont en fait des combinaisons de ces quatre champs, ce qui est décrit dans la section 2.4.2.

2.3.4 Interaction électrofaible pour les quarks

Le lagrangien électrofaible ainsi construit pour les leptons pourrait facilement être réutilisé dans le cas des quarks, ψ étant un champ décrivant un fermion. Cependant, le lagrangien de l'équation (1.52) ne couple entre eux que des fermions de même génération. Or, il a été observé expérimentalement que l'interaction faible peut également coupler des quarks de générations différentes.

Un mécanisme rendant possible de tels couplages a été introduit par Cabibbo, Kobayashi et Maskawa [6-8]. Le principe est de faire, pour les quarks, la distinction entre les états propres de masse, c'est-à-dire ceux que l'on observe, et les états propres de l'interaction faible. Ces deux ensembles d'états propres diffèrent ainsi pour les quarks d'isospin faible bas et sont reliés entre eux par la matrice CKM $\mathcal{M}_{CKM}$, matrice 3×3 unitaire complexe,

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}\,, \tag{1.53}$$

où d', s' et b' sont les états propres de l'interaction faible et d, s et b ceux de masse. L'élément de matrice V_{ij} ou son conjugué V^*_{ij} est ainsi un facteur appliqué au vertex pour le calcul de la section efficace des processus impliquant des quarks et l'interaction faible. Ces coefficients ne sont pas prédits par le modèle standard et sont donc mesurés expérimentalement. Les valeurs de leurs modules sont les suivantes [9]

$$|\mathcal{M}_{CKM}| = \begin{pmatrix} 0{,}973\,70 \pm 0{,}000\,14 & 0{,}2245 \pm 0{,}0008 & 0{,}003\,82 \pm 0{,}000\,24 \\ 0{,}221 \pm 0{,}004 & 0{,}987 \pm 0{,}011 & 0{,}0410 \pm 0{,}0014 \\ 0{,}0080 \pm 0{,}0003 & 0{,}0388 \pm 0{,}0011 & 1{,}013 \pm 0{,}030 \end{pmatrix}\,. \tag{1.54}$$

Cette matrice a une structure très prononcée, presque diagonale. Le couplage entre les quarks de générations différentes est faible, ce qui se traduit expérimentalement par des durées de vie de certains hadrons contenant des quarks de deuxième ou troisième génération suffisamment longs pour qu'ils se propagent sur quelques millimètres, voire quelques mètres.

Le modèle ainsi construit décrit l'interaction électrofaible pour tous les fermions. Cependant, il n'y a aucun terme de masse dans le lagrangien de l'équation (1.52). En effet, un terme de masse pour les fermions serait de la forme

$$-m\bar{\psi}\psi = -m(\bar{\psi}_R + \bar{\psi}_L)(\psi_R + \psi_L) = -m(\bar{\psi}_R\psi_R + \bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R + \bar{\psi}_L\psi_L) = -m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)\,. \tag{1.55}$$

Or, ce terme n'est pas invariant sous $SU(2)_L \times U(1)_Y$. Pour les champs W_μ^i et B_μ, des termes de masse violeraient également la symétrie de jauge. Dès lors, il semble difficile pour un tel lagrangien de décrire les forces électromagnétique et faible.

En réalité, ce lagrangien décrit l'interaction « électrofaible ». Les interactions électromagnétique et faible résultent d'un mécanisme de brisure spontanée de symétrie, qui se trouve dans ce cas être le mécanisme de Higgs. La section suivante montre comment l'introduction du champ de Higgs amène cette brisure de symétrie et comment sont obtenus des fermions massifs, le photon et les bosons $W^\pm$ et Z.

2.4 Mécanisme de Higgs et brisure spontanée de symétrie

2.4.1 Champ de Higgs et brisure de symétrie

Soit un champ complexe, scalaire, massif, le champ de Higgs [10-15], noté ϕ. Il comporte quatre composantes et s'écrit sous la forme d'un doublet d'isospin faible,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_3 + \mathrm{i}\phi_4 \\ \phi_1 + \mathrm{i}\phi_2 \end{pmatrix} . \tag{1.56}$$

Le champ de Higgs a pour hypercharge $Y = +1$ et pour isospin faible $I = \frac{1}{2}$. Il se transforme ainsi, respectivement sous $U(1)_Y$ et $SU(2)_L$, selon

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \to \begin{pmatrix} \mathrm{e}^{\mathrm{i}\frac{\beta}{2}} & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\frac{\beta}{2}} \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} , \quad \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \to \mathrm{e}^{\frac{\mathrm{i}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.57}$$

La dérivée covariante définie par l'équation (1.49) agit donc sur le champ de Higgs selon

$$D_\mu \phi = \left[\partial_\mu - \frac{\mathrm{i}}{2} g_I \boldsymbol{\tau} \cdot \boldsymbol{W}_\mu - \frac{\mathrm{i}}{2} g_Y B_\mu \right] \phi , \tag{1.58}$$

et ce dernier apporte les termes $\mathcal{L}_h$ au lagrangien du modèle standard, où

$$\mathcal{L}_h = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi) \tag{1.59}$$

avec

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 , \quad \lambda > 0 . \tag{1.60}$$

Il est possible de déterminer le comportement de ce champ autour du minimum du potentiel V, c'est-à-dire autour de sa position d'équilibre. Soit v tel que $v^2 = \phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2$. Alors, à la position d'équilibre,

$$\begin{cases} \left.\dfrac{\partial V}{\partial \phi}\right|_{\text{éq}} = 0 \\ \left.\dfrac{\partial^2 V}{\partial \phi^2}\right|_{\text{éq}} \geqslant 0 \end{cases} \Rightarrow \begin{cases} \left(2\mu^2 + 4\lambda \dfrac{v^2}{2}\right) \dfrac{v}{\sqrt{2}} = 0 \\ 2\mu^2 + 12\lambda \dfrac{v^2}{2} \geqslant 0 \end{cases} \Rightarrow \begin{cases} \mu^2 \geqslant 0 \Rightarrow v = 0 \\ \mu^2 < 0 \Rightarrow v^2 = \dfrac{-\mu^2}{\lambda} \end{cases} . \tag{1.61}$$

Ainsi, dans le cas où $\mu^2 < 0$, le potentiel possède une infinité de minimums, situés sur un cercle de rayon $\frac{|v|}{\sqrt{2}}$, comme cela est visible sur la figure 1.6. La forme de ce potentiel n'est pas sans rappeler celle d'un sombrero.

Le condensat de ce champ dans le vide correspond à la position de ce minimum de potentiel,

$$\langle \phi \rangle_0 = \frac{v}{\sqrt{2}} \neq 0 , \tag{1.62}$$

ce qui implique le phénomène de brisure spontanée de symétrie.

Le minimum de potentiel pour V est dégénéré, c'est-à-dire qu'il y a une infinité de points correspondant au minimum de V. Il est possible de briser la symétrie avec n'importe lequel de ces points. Les observables physiques s'obtiennent alors en réalisant un développement limité autour du point choisi. Le choix le plus simple est nommé « jauge unitaire ». Dans ce cas,

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.63}$$

où h correspond à un nouveau champ de Higgs, cette fois-ci physiquement réel.

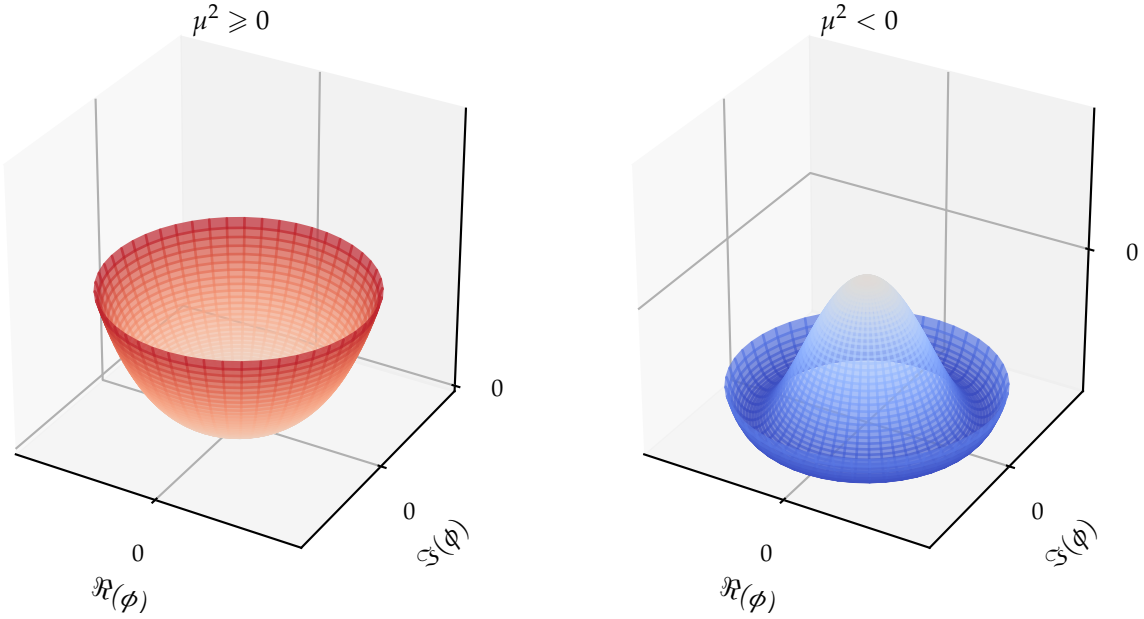


Figure 1.6 – *Forme du potentiel du Higgs V selon le signe de μ^2.*

2.4.2 Masses des bosons

Lorsque l'expression (1.63) de ϕ est injectée dans $\mathcal{L}_h$, le terme cinétique devient

$$\begin{aligned}(D^\mu\phi)^\dagger(D_\mu\phi) = &\frac{1}{2}\partial_\mu h\partial^\mu h \\ &+ \frac{(v+h)^2}{8}g_I^2\left(W_\mu^1 + \mathrm{i}W_\mu^2\right)\left(W^{\mu 1} - \mathrm{i}W^{\mu 2}\right) \\ &+ \frac{(v+h)^2}{8}\left(g_I W_\mu^3 - g_Y B_\mu\right)\left(g_I W^{\mu 3} - g_Y B^\mu\right) + \mathcal{O}(h^2)\,. \end{aligned} \tag{1.64}$$

Il est possible de développer cette expression et d'identifier les termes quadratiques qui correspondent à des termes de masse pour les bosons physiques, tout en obtenant les combinaisons des champs leur correspondant. Trois bosons massifs sont obtenus, W^+, W^- et Z,

$$\text{bosons } W^\pm : \qquad W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp \mathrm{i}W_\mu^2\right), \qquad m_W = \frac{1}{2}vg_I\,, \tag{1.65}$$

$$\text{boson } Z : \qquad Z_\mu = \frac{g_I W_\mu^3 - g_Y B_\mu}{\sqrt{g_I^2 + g_Y^2}}, \qquad m_Z = \frac{1}{2}v\sqrt{g_I^2 + g_Y^2}\,. \tag{1.66}$$

Or, il existe quatre champs quantiques, W_μ^i ($i \in \{1,2,3\}$) et B_μ. Un quatrième boson peut donc être défini. Il correspond au photon γ, sans masse,

$$\text{photon } \gamma : \qquad A_\mu = \frac{g_Y W_\mu^3 + g_I B_\mu}{\sqrt{g_I^2 + g_Y^2}}, \qquad m_A = 0\,. \tag{1.67}$$

Cette combinaison est fixée par la construction d'une matrice unitaire $\mathcal{M}_{\gamma Z}$ reliant les champs A_μ et Z_μ à B_μ et W_μ^3,

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \mathcal{M}_{\gamma Z}\begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \quad \mathcal{M}_{\gamma Z} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \tag{1.68}$$

avec θ_W l'angle de mélange faible [16], tel que

$$\cos\theta_W = \frac{g_I}{\sqrt{g_I^2 + g_Y^2}}, \quad \sin\theta_W = \frac{g_Y}{\sqrt{g_I^2 + g_Y^2}}\,. \tag{1.69}$$

Alors, les champs A_μ et Z_μ s'expriment

$$\text{photon } \gamma : \qquad A_\mu = B_\mu \cos\theta_W + W^3_\mu \sin\theta_W\,, \qquad m_A = 0\,, \tag{1.70}$$

$$\text{boson } Z : \qquad Z_\mu = W^3_\mu \cos\theta_W - B_\mu \sin\theta_W\,, \qquad m_Z = \frac{m_W}{\cos\theta_W}\,. \tag{1.71}$$

Pour le boson de Higgs lui-même, le terme de masse provient de $V(\phi)$ et donne $m_h = \sqrt{-2\mu^2}$.

Les masses des bosons $W^\pm$, γ et Z ainsi prédites [3] sont expérimentalement confirmées [9, 17-21]. Cependant, la masse du boson de Higgs dépendant de μ, paramètre libre de ce modèle, seule une détermination expérimentale permet de l'obtenir. En 2012, les collaborations ATLAS et CMS ont observé un boson [22-24], confirmé comme étant ce boson de Higgs [25, 26]. Sa masse est déterminée à $125{,}10 \pm 0{,}14\,\text{GeV}$ [9, 27, 28]. Les désintégrations de h en WW [29], ZZ [30, 31] et $\gamma\gamma$ [32] ont été observées, confirmant les propriétés attendues de h.

2.4.3 Masses des fermions

Le champ de Higgs peut également interagir avec les fermions. Une telle interaction, entre un champ scalaire et un champ de Dirac, est une interaction de Yukawa et permet d'introduire des termes de masse invariants de jauge pour ces fermions.

Avant la brisure de symétrie, les termes de Yukawa pour un champ ψ dont la composante d'isospin faible bas correspond à une particule de masse m s'expriment

$$\mathcal{L}_{\text{Yukawa}} = -\frac{m\sqrt{2}}{v}\bar{\psi}\phi\psi = -\frac{m\sqrt{2}}{v}\left(\bar{\psi}_L\phi\psi_R + \bar{\psi}_R\phi^\dagger\psi_L\right)\,, \tag{1.72}$$

donnant, après la brisure spontanée de symétrie précédemment décrite, les termes de masse pour les fermions d'isospin faible bas ainsi que le terme d'interaction entre le fermion et le boson de Higgs h,

$$\mathcal{L}_{\text{Yukawa}} = -m\bar{\psi}\psi + \frac{m}{v}\bar{\psi}h\psi\,. \tag{1.73}$$

En particulier, l'intensité du couplage entre le fermion et le boson de Higgs est proportionnelle à la masse m du fermion.

La brisure de symétrie dans le cas de la jauge unitaire (1.63) laisse la composante d'isospin faible haut du champ de Higgs nulle. Ce formalisme permet donc d'obtenir les termes de masse pour les leptons chargés. Les neutrinos étant considérés dans le lagrangien du modèle standard comme des particules de masses nulles, l'ensemble des leptons est donc traité à ce stade.

Dans le cas des quarks, en revanche, il faut obtenir des termes de masse pour les quarks d'isospin faible haut. Pour cela, il est possible d'introduire le conjugué de charge du champ de Higgs,

$$\phi^C = \mathrm{i}\sigma_2\phi^* = \begin{pmatrix}\phi^* \\ -\phi^-\end{pmatrix} \tag{1.74}$$

dont l'expression devient, après brisure de symétrie dans le cas de la jauge unitaire,

$$\phi^C = \frac{-1}{\sqrt{2}}\begin{pmatrix}v + h(x) \\ 0\end{pmatrix}\,. \tag{1.75}$$

Cette fois, la composante d'isospin faible haut du champ de Higgs n'est pas nulle et permet d'obtenir des termes de masse pour les quarks d'isospin faible haut.

Ainsi, les termes de Yukawa pour les fermions s'expriment

$$\mathcal{L}_{\text{Yukawa}} = y\bar{\psi}\phi\psi + \text{h.c.} = \sum_{i=1}^{3} y^\ell_i \bar{\ell}_i\phi\ell_i + \sum_{i=1}^{3} y^d_i \bar{d}_i\phi d_i + \sum_{i=1}^{3} y^u_i \bar{u}_i\phi^C u_i\,, \quad y^x_i = -\frac{\sqrt{2}}{v}m^x_i \tag{1.76}$$

où « h.c. » signifie conjugué hermitien, i correspond à la génération des fermions, ℓ_i, d_i et u_i aux champs listés dans le tableau 1.1 et m^x_i à la masse du fermion de type x et de génération i.

Les désintégrations de h en $b\bar{b}$ [33, 34] et $\tau\tau$ [35, 36] ont été observées, conformément aux couplages attendus. Un excès à 3σ est de plus obtenu sur la désintégration en $\mu\mu$ [37], ce qui est insuffisant pour déclarer une observation mais donne une preuve du couplage aux muons. Enfin, le cas du quark t est particulier. Cette particule est trop massive pour que h se désintègre en $t\bar{t}$, toutefois la production de h en association avec une paire de quarks t a été observée [38] et ces résultats sont en accord avec les prédictions du modèle standard.

2.5 Interaction forte

2.5.1 Charge de couleur

L'interaction forte est la troisième force fondamentale décrite par le modèle standard. L'analogue de la charge électrique pour l'interaction électromagnétique est, dans le cas de l'interaction forte, la « couleur », concept né de l'observation des baryons Δ^{++}, Δ^{-} [39-41] et Ω^{-} [42]. Dans le modèle des quarks, ces baryons sont composés comme

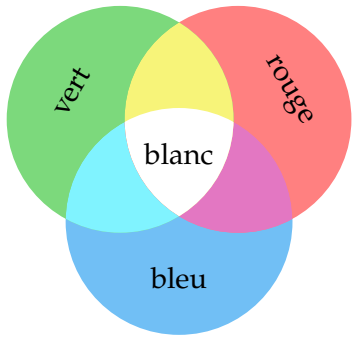
$$\Delta^{++} = (uuu)\,, \quad \Delta^{-} = (ddd)\,, \quad \Omega^{-} = (sss)\,. \tag{1.77}$$

Or, ces baryons sont de spin $\frac{3}{2}$. Les quarks possédant un spin $\frac{1}{2}$, il faudrait alors que pour chacun de ces baryons, les trois quarks les composant aient leurs nombres quantiques égaux, ce qui va à l'encontre du principe de Pauli.

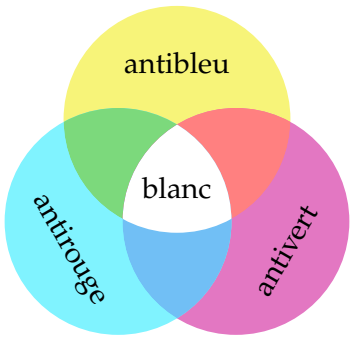
Il est possible de décrire ces baryons sans violer le principe d'exclusion de Pauli en introduisant un nouveau nombre quantique, la couleur. Les quarks portent ainsi une charge de couleur pouvant prendre trois valeurs orthogonales, nommées par convention rouge, verte et bleue. Les antiquarks portent une anticouleur. Chacun des trois quarks des baryons porte alors une couleur différente, c'est-à-dire

$$\Delta^{++} = (uuu)\,, \quad \Delta^{-} = (ddd)\,, \quad \Omega^{-} = (sss)\,. \tag{1.78}$$

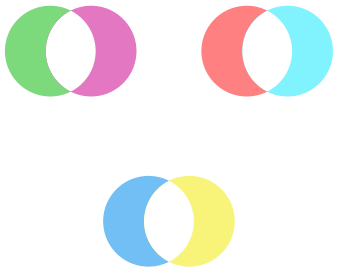
Les baryons ainsi formés portent une charge de couleur globale nulle, ils sont « blancs », comme cela est visible sur la figure 1.7a. Dans le cas des antibaryons formés de trois antiquarks, sur la figure 1.7b, c'est l'association des trois anticouleurs qui permet d'obtenir un baryon blanc. Il est également possible de former une particule composite blanche par association d'un quark avec un antiquark portant l'anticouleur correspondante. Les trois combinaisons possibles sont illustrées sur la figure 1.7c. Il s'agit alors de mésons.



(a) *Un baryon est constitué de trois quarks, un de chaque couleur.*



(b) *Un antibaryon est constitué de trois antiquarks, un de chaque anticouleur.*



(c) *Un méson est constitué d'un quark et d'un antiquark de l'anticouleur correspondante.*

Figure 1.7 – *Combinaisons des couleurs des quarks dans les hadrons. La couleur globale est toujours blanche, c'est-à-dire que la charge de couleur globale est nulle.*

Les quarks et antiquarks se regroupent ainsi en particules composites, les hadrons (baryons et mésons), dont la neutralité de couleur est confirmée expérimentalement. Ce phénomène est connu sous le nom de « confinement de couleur » et est abordé dans la section 2.5.3.

2.5.2 Symétrie $SU(3)_C$

Afin de décrire l'interaction forte dans le même formalisme que les autres interactions fondamentales, il faut un groupe de symétrie. Étant donné qu'il existe trois dimensions de couleur (rouge, verte, bleue), la théorie quantique des champs associée à l'interaction forte se base sur le groupe $SU(3)_C$, où C signifie « couleur ».

Tout comme $SU(2)$, $SU(3)$ est un groupe non abélien. Il est possible de reprendre exactement les mêmes calculs que ceux de la section 2.3.2, en procédant aux changements

$$\boldsymbol{\tau} \in \mathcal{M}_2(\mathbb{C})^3 \leftrightarrow \boldsymbol{\lambda} \in \mathcal{M}_3(\mathbb{C})^8\,, \quad \boldsymbol{\alpha} \in \mathbb{R}^3 \leftrightarrow \boldsymbol{\theta} \in \mathbb{R}^8\,, \quad g_I \leftrightarrow \alpha_s\,, \quad \boldsymbol{W}_\mu \leftrightarrow \boldsymbol{G}_\mu\,, \quad \boldsymbol{W}_{\mu\nu} \leftrightarrow \boldsymbol{G}_{\mu\nu} \tag{1.79}$$

où $\boldsymbol{\lambda}$ est un vecteur à huit composantes, chacune étant une matrice de Gell-Mann, définies dans l'annexe B, $\boldsymbol{G}_\mu$ décrit donc huit gluons, bosons vecteurs de l'interaction forte et α_s est la constante de couplage de l'interaction forte.

Les gluons portent une couleur et une anticouleur. Lors de chaque interaction, la charge de couleur est conservée, ainsi un quark rouge interagissant avec un gluon bleu-antirouge devient un quark bleu. Le flux de couleur ainsi conservé dans cet exemple est représenté sur la figure 1.8.



(a) *Diagramme de Feynman de l'interaction.* **(b)** *Représentation du flux de couleur conservé.*

Figure 1.8 – *Interaction entre un quark rouge et un gluon bleu-antirouge, donnant un quark bleu.*



(a) **(b)**

Figure 1.9 – *Diagrammes de Feynman correspondant à l'interaction entre trois et quatre gluons.*

Le terme non linéaire $\boldsymbol{G}_\mu \wedge \boldsymbol{G}_\nu$ dans l'expression de $\boldsymbol{G}_{\mu\nu}$, obtenue à partir de l'analogie (1.79) appliquée à l'équation (1.45), est lourd de conséquences. Il permet le couplage entre trois et quatre gluons, comme cela est illustré sur la figure 1.9, et donne à l'interaction forte toute sa singularité. En effet, ce terme est responsable du confinement de couleur et de l'initiation de la gerbe partonique qui donne naissance aux jets, discutés dans la section 2.6. Ce phénomène n'est pas présent dans le cas de l'interaction faible à cause de la brisure spontanée de symétrie donnant des bosons W et Z massifs.

2.5.3 Confinement de couleur et liberté asymptotique

Le confinement de couleur force les quarks, particules colorées, à s'associer en formant des particules composites, les hadrons, états liés de charge globale de couleur nulle. Ce phénomène empirique peut s'expliquer par la variation en fonction de l'échelle d'énergie de la constante de couplage de l'interaction forte α_s, représentée sur la figure 1.10.

Aux basses énergies, α_s diverge. Ainsi, séparer et isoler des particules colorées mène à une énergie potentielle de couleur suffisamment grande pour créer des paires quark-antiquark. Lorsqu'un quark est issu d'une collision en physique des particules, ce processus se réalise et produit un ensemble collimé de particules, un jet.

De plus, à cause de la valeur élevée de α_s aux basses énergies, il n'est pas possible de réaliser des calculs perturbatifs usuels en théorie quantique des champs. D'autres techniques sont toutefois utilisées, comme la méthode de QCD sur réseau. Son principe est de discrétiser l'espace-temps en un réseau de points. Bien que cette méthode requière d'importantes capacités de calcul et beaucoup de temps, elle permet d'obtenir avec succès les masses des hadrons comme cela se voit sur la figure 1.11 pour les hadrons légers.

Les valeurs de α_s à deux échelles d'énergie k et μ sont reliées par la relation

$$\alpha_s(k) = \frac{\alpha_s(\mu)}{1 + \frac{11n_c - 2n_f}{12\pi}\alpha_s(\mu)\ln\left(\frac{k^2}{\mu^2}\right)} \tag{1.80}$$

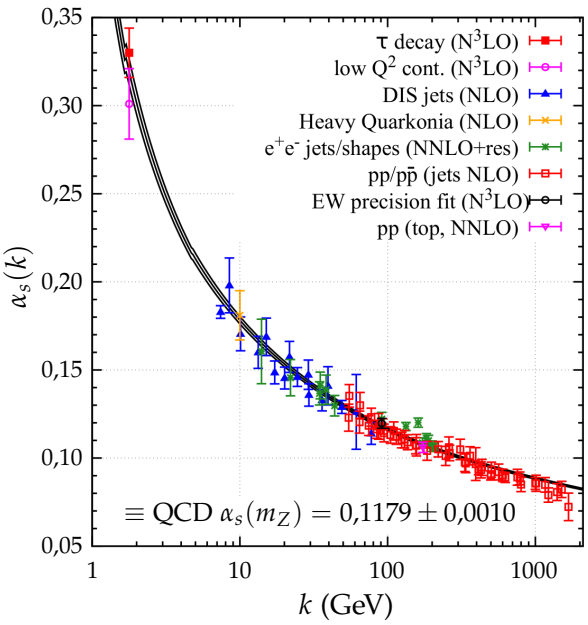


Figure 1.10 – *Mesures de α_s en fonction de l'échelle d'énergie k (points) et prédiction théorique (courbe) [9]. Le degré des calculs perturbatifs de QCD utilisés pour extraire α_s est indiqué entre parenthèses (NLO :* next-to-leading order, *c'est-à-dire jusqu'à l'ordre suivant le premier degré non nul ; NNLO : un ordre de plus que NLO ; etc.).*

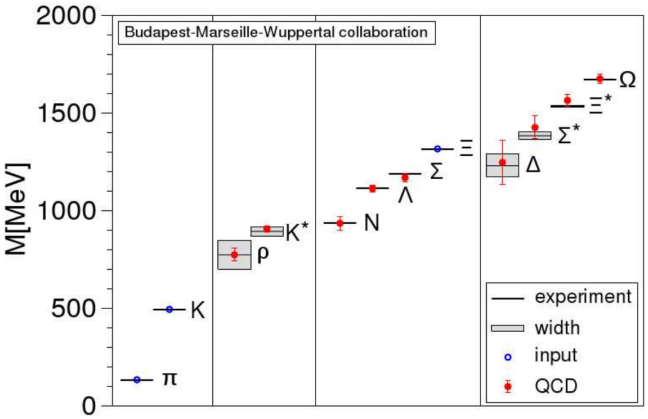


Figure 1.11 – *Spectre de masse des hadrons légers [43]. Les lignes horizontales ainsi que les zones grisées sont les valeurs expérimentales et les largeurs de désintégration. Les résultats obtenus en utilisant des calculs de QCD sur réseau sont représentés par des cercles, avec les erreurs associées. Seules les masses des hadrons π, K et Ξ ne présentent pas de barres d'erreurs, celles-ci sont utilisées pour fixer les paramètres libres du modèle.*

avec n_c le nombre de couleurs et n_f le nombre de saveurs de quarks, c'est-à-dire $n_c = 3$ et $n_f = 6$ [44]. Cette relation peut ainsi se réécrire, en prenant $\mu = \Lambda_{\text{QCD}}$ l'échelle d'énergie à laquelle α_s diverge,

$$\alpha_s(k) = \frac{6\pi}{21 \ln\left(\frac{k}{\Lambda_{\text{QCD}}}\right)}, \quad \Lambda_{\text{QCD}} = 218 \pm 24\,\text{MeV}\,. \tag{1.81}$$

Il ressort que α_s décroît lorsque l'échelle d'énergie augmente. Cette diminution de α_s aux hautes énergies est la « liberté asymptotique » [45, 46], régime dans lequel les particules colorées ne sont plus confinées et peuvent se propager comme des particules libres. Au LHC, les énergies de collision permettent d'atteindre ce régime.

2.6 Formation des jets

Lorsqu'un parton (quark ou gluon) est issu d'une collision de particules, il possède une haute énergie et émet alors, par interaction forte, d'autres partons. La « gerbe partonique », sujet de la prochaine section, est ainsi créée. Par conservation, l'énergie portée individuellement par chaque parton

de cette gerbe diminue au fur et à mesure des nouvelles émissions de partons et par conséquent, α_s augmente. Tant que l'échelle d'énergie est suffisamment grande pour que $\alpha_s \ll 1$, ce qui correspond à des énergies supérieures à la centaine de MeV, il est possible de réaliser des calculs perturbatifs. En deçà d'une centaine de MeV, ce n'est plus possible. Des modèles paramétriques sont alors utilisés pour caractériser le phénomène d'« hadronisation ».

2.6.1 Gerbe partonique

Chaque parton issu d'une collision au LHC se trouve dans un premier temps dans le régime de liberté asymptotique. Ainsi, pour un événement $Z \to q\bar{q}$ comme celui de la figure 1.12a avec deux quarks dans l'état final, il est possible d'obtenir par émission d'un gluon un état $q\bar{q}g$ comme ceux illustrés sur les figures 1.12b et 1.12c, par exemple.

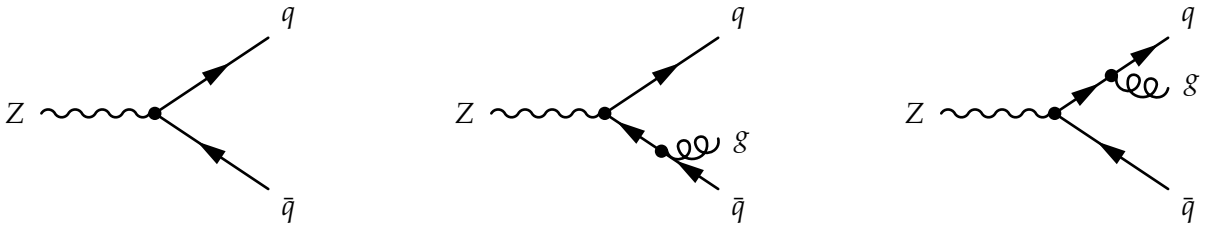


Figure 1.12 – *Un boson Z se désintègre en paire quark-antiquark. Dans les cas des figures 1.12b et 1.12c, un gluon supplémentaire est émis.*

Il est légitime de se demander quelle est la probabilité d'obtenir un état $q\bar{q}g$ à partir d'un état $q\bar{q}$. Des calculs de section efficace permettent d'obtenir [44], pour un état initial contenant X partons dont i qui émet j, donnant un état final à $X+1$ partons,

$$\mathrm{d}\sigma_{X+j} \simeq \sigma_X \sum_{i \in \{X\}} \frac{\alpha_s}{2\pi} \frac{\mathrm{d}\theta^2}{\theta^2} \mathrm{d}z\, P_{ij}(z) \tag{1.82}$$

où θ est l'angle entre j et i. La grandeur $P_{ij}(z)$ est la probabilité que j émis par i emporte une fraction z de l'énergie initiale de i, qui s'exprime en fonction de la nature de i et j (quark ou gluon) selon

$$P_{qq}(z) = C_F \frac{1+z^2}{1-z}, \qquad P_{qg}(z) = C_F \frac{1+(1-z)^2}{z}, \tag{1.83}$$

$$P_{gg}(z) = C_A \frac{z^4+1+(1-z)^4}{z(1-z)}, \qquad P_{gq}(z) = T_R(z^2+(1-z)^2), \tag{1.84}$$

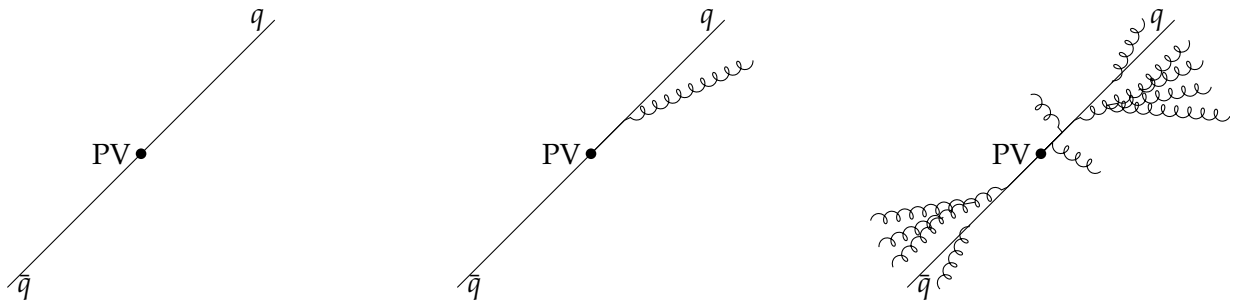
et $P_{g\bar{q}}(z) = P_{gq}(z)$, avec $C_F = \frac{4}{3}$, $C_A = 3$ et $T_R = \frac{1}{2}$. La probabilité d'émettre un parton supplémentaire diverge dans deux cas :

— le parton émis a une énergie faible devant celle du parton émetteur, c'est la limite infrarouge ;
— l'angle entre le parton émis et le parton émetteur est petit, c'est la limite colinéaire.

Les nouveaux partons ainsi émis et les partons initiaux continuent chacun ces processus jusqu'à ce que le phénomène de confinement de couleur réapparaisse. Pour chaque parton directement issu de la collision ayant lieu au vertex primaire (PV), une gerbe partonique est formée, c'est-à-dire un ensemble collimé de partons, comme illustré sur la figure 1.13. Ce sont ces particules qui vont participer au phénomène d'hadronisation dû au confinement de couleur.

2.6.2 Hadronisation

Lorsque des partons en émettent d'autres, la conservation de l'énergie implique que chaque particule possède individuellement une énergie de plus en plus petite. Or α_s augmente lorsque l'échelle d'énergie diminue et en-deçà de quelques centaines de MeV, α_s diverge. Le phénomène de confinement de couleur réapparaît et la gerbe partonique subit alors un phénomène d'hadronisation. Un flux collimé de hadrons est obtenu. Certains d'entre eux peuvent comporter des quarks de deuxième



(a) *Deux quarks sont initialement produits, ce qui correspond au diagramme de la figure 1.12a.*

(b) *Un des quarks peut émettre un gluon, ce qui correspond au diagramme de la figure 1.12c.*

(c) *Le processus est réitéré, donnant un ensemble de particules colorées.*

Figure 1.13 – *Formation de deux gerbes partoniques à partir d'une paire de quarks formée au vertex primaire (PV).*

ou troisième génération. Ces hadrons sont alors instables et peuvent être amenés à se désintégrer. Dans ce cas, leurs produits de désintégration sont observés dans le détecteur.

Le phénomène d'hadronisation ayant lieu lorsque $\alpha_s \gtrsim 1$, il n'est pas possible de réaliser des calculs perturbatifs. Afin de décrire ce phénomène, il faut avoir recours à des modèles paramétriques comme le modèle d'agglomération hadronique [47] ou le modèle des cordes de Lund [48].

Agglomération hadronique Ce modèle [47] repose sur l'hypothèse de conservation des nombres quantiques ainsi que de l'énergie-impulsion entre les partons issus de la gerbe partonique et les hadrons obtenus après hadronisation.

Dans un premier temps, les gluons de la gerbe partonique se désintègrent en paires $q\bar{q}$. Les partons, uniquement des quarks à ce stade donc, se rassemblent dans un second temps en agglomérats de charge de couleur nulle, c'est le « pré-confinement ». Deux cas de figure se présentent alors pour chaque agglomérat :

— si la masse de l'agglomérat est proche de celle d'un hadron, il produit ce hadron ;
— sinon, il se désintègre en agrégats plus petits et forme plusieurs hadrons.

Ce processus est illustré sur la figure 1.14.

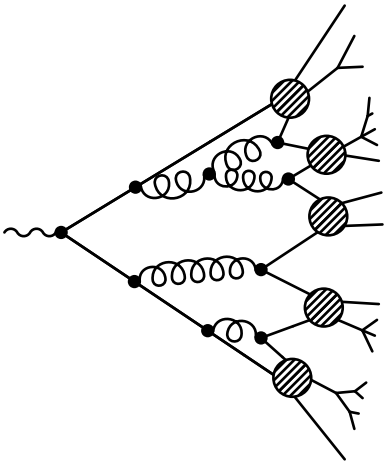


Figure 1.14 – *Schématisation de l'hadronisation dans le cadre du modèle d'agglomération hadronique.*

Cordes de Lund Dans le modèle des cordes de Lund [48], les quarks sont reliés en paires $q\bar{q}$ par des « cordes » de couleur, de tension $\kappa \simeq 1\,\mathrm{GeV{\cdot}fm^{-1}}$, comme sur la figure 1.15b. Les gluons sont décrits comme des nœuds des cordes de couleur.

Lorsque deux charges colorées s'éloignent, l'énergie potentielle augmente. Une fois que l'énergie potentielle est suffisamment grande, une nouvelle paire $q'\bar{q}'$ est créée (fig. 1.15c) avec une probabilité proportionnelle à $\exp\left(-\frac{\pi}{\kappa} m_{q'}\right)$. La probabilité d'obtenir des quarks lourds par ce processus est donc très faible. Le partage de l'énergie entre les paires de quarks est régi par une fonction de partition dont les paramètres sont estimés expérimentalement.

2.6.3 Parton initial et caractéristiques du jet

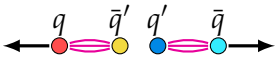
Un jet présente certaines caractéristiques suivant le type de parton à son origine. Le parton n'est pas visible expérimentalement, mais les caractéristiques du jet obtenu permettent d'en estimer la



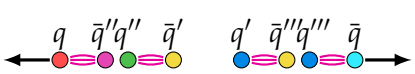
(a) *Les deux quarks issus de la collision se séparent à grande vitesse.*



(b) *Une « corde » de flux de couleur se forme entre les deux quarks.*



(c) *L'énergie potentielle de la corde est suffisamment grande pour former de nouvelles paires de quarks.*



(d) *Le processus se répète tant qu'il y a suffisamment d'énergie pour générer une paire de quarks.*



(e) *Des hadrons non colorés sont formés à partir des quarks de basse énergie.*

Figure 1.15 – *Processus de formation de deux jets dans le cadre du modèle des cordes de Lund.*

saveur, comme exposé au chapitre 2.

Le quark t possède une durée de vie trop courte pour participer à l'hadronisation. Il se désintègre par interaction faible en un autre quark, généralement un quark b, et un boson W. Le nouveau quark issu de cette désintégration forme alors un jet.

Les autres quarks (d, u, s, c et b) participent à l'hadronisation. Ils se retrouvent donc confinés au sein des hadrons qui en résultent. Ceux-ci sont éventuellement instables.

Le quark b forme un hadron instable qui donne d'autres hadrons lors de la désintégration du quark b en quark c ou u par interaction faible. Dans 70 % des cas, cette désintégration se fait avec émission d'une nouvelle paire de quarks $q\bar{q}$ selon

$$b \to c\, q_d \bar{q}_d\,, \quad \bar{b} \to \bar{c}\, q_u \bar{q}_u\,, \quad b \to u\, q_d \bar{q}_d\,, \quad \bar{b} \to \bar{u}\, q_u \bar{q}_u\,, \tag{1.85}$$

où q_d et q_u désignent respectivement des quarks d'isospin faible bas et haut. Le nombre de constituants du jet, ainsi que le nombre de traces provenant d'un vertex secondaire (SV), est alors plus important. Dans 30 % des cas, la désintégration du quark b se fait avec émission d'une paire de leptons, l'un étant électriquement chargé et l'autre correspondant au neutrino associé, c'est-à-dire

$$b \to c\, \ell^- \bar{\nu}_\ell\,, \quad \bar{b} \to \bar{c}\, \ell^+ \nu_\ell\,, \quad b \to u\, \ell^- \bar{\nu}_\ell\,, \quad \bar{b} \to \bar{u}\, \ell^+ \nu_\ell\,. \tag{1.86}$$

Un lepton chargé au sein d'un jet donne une signature caractéristique de la présence d'un quark b lors des collisions proton-proton du LHC.

Les désintégrations (1.85) font intervenir les modules des coefficients V_{cb} ou V_{ub} de la matrice CKM, introduite dans la section 2.3.4, dont les valeurs sont faibles ; elles sont donc fortement supprimées. Les hadrons contenant un quark b ont ainsi une durée de vie de l'ordre de la picoseconde [49, 50] et peuvent voyager sur une distance de l'ordre du millimètre. Les traces des particules chargées issues de cette nouvelle désintégration proviennent donc d'un vertex secondaire (SV), différent du vertex primaire (PV). Ces traces sont « déplacées ». Pour chacune d'entre elles, il est possible de déterminer le paramètre d'impact (IP) au vertex primaire, dont la valeur est typiquement plus grande que pour des traces provenant du vertex primaire, comme cela est illustré sur la figure 1.16.

Le quark c possède une phénoménologie similaire au quark b. Cependant, la désintégration du quark c en quark s selon

$$c \to s\, q_u \bar{q}_u\,, \quad \bar{c} \to \bar{s}\, q_d \bar{q}_d\,, \quad c \to s\, \ell^+ \nu_\ell\,, \quad \bar{c} \to \bar{s}\, \ell^- \bar{\nu}_\ell\,, \tag{1.87}$$

fait intervenir le module du coefficient V_{cs} de la matrice CKM, proche de 1. Les hadrons contenant un quark c ont ainsi une durée de vie inférieure à la picoseconde [50] et il est donc plus difficile d'identifier les jets issus de quarks c que ceux issus de quarks b.

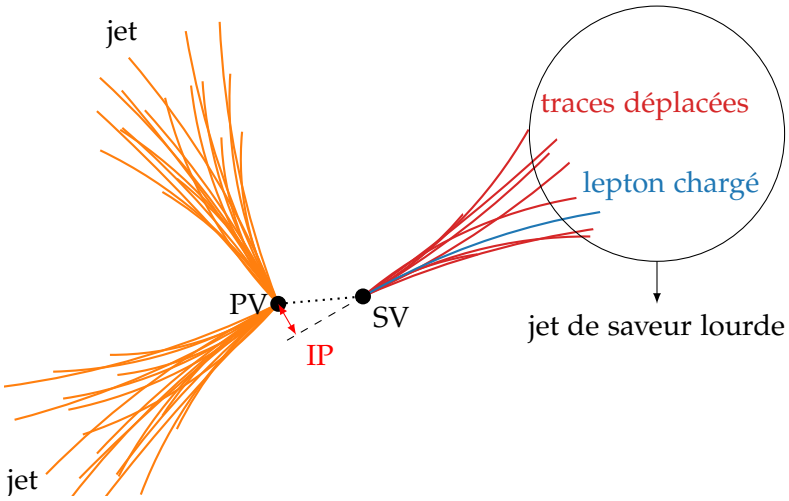


Figure 1.16 – *Trois jets, dont un de saveur lourde. Les particules composant ce jet proviennent d'un vertex secondaire (SV), différent du vertex primaire (PV) où a lieu la collision entre les protons et la formation du hadron lourd à l'origine du SV. Le paramètre d'impact (IP) est également indiqué. Réalisé à l'aide de CMSTransverseTikZ [51].*

Les quarks *d*, *u* et *s* forment des hadrons :

— très instables, par exemple les π^0, dont seuls les produits de désintégration sont observés ;
— faiblement instables, tels que les K^+ et les K^0, qui se propagent généralement jusque dans les parties sensibles du détecteur et peuvent donc être observés directement ;
— stables, par exemple les protons, qui sont directement observés dans le détecteur.

Dans tous les cas, des traces de particules chargées proviennent du PV, lieu de formation du quark initial. Certains hadrons neutres tels que les Λ^0 ou les K^0 peuvent se désintégrer en particules chargées au sein du trajectographe, mais le nombre de traces déplacées reste limité. Le phénomène décrit précédemment pour les quarks b et c n'est donc pas observable. Les jets issus de ces trois types de quarks, les plus légers, sont ainsi regroupés sous la dénomination de « jets légers ».

Les gluons portent une charge de couleur plus importante que les quarks. Les quarks portent en effet une couleur, les antiquarks une anticouleur et les gluons portent une couleur et une anticouleur. Les jets initiés par des gluons comportent typiquement plus de particules électriquement chargées et sont moins collimés que les jets légers [52].

3 Succès et limites du modèle standard

3.1 Succès

Le modèle standard a été conçu il y a plus de 50 ans et a évolué avec les nouvelles observations expérimentales. Ses succès sont nombreux, notamment de par son pouvoir prédictif. L'existence de bosons massifs W et Z est ainsi prédite dès la fin des années 60 [3] et leur observation est confirmée expérimentalement en 1983 [17-21]. De même, le quark top postulé en 1973 [6] est observé 22 ans plus tard [53, 54]. Dernier tour de force en date, le boson de Higgs, prédit en 1964 [10-13], est découvert en 2012 [22, 23].

Le lagrangien du modèle standard s'écrit à ce jour sous la forme

$$\mathcal{L}_{SM} = \underbrace{-\frac{1}{4}\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}}_{\substack{\text{bosons}\\\text{vecteurs}}} + \underbrace{i\bar{\psi}\not{D}\psi}_{\substack{\text{fermions et}\\\text{interactions}}} + \underbrace{(D^\mu\phi)^\dagger(D_\mu\phi) - V(\phi)}_{\substack{\text{boson de Higgs et}\\\text{interactions avec les autres bosons}}} + \underbrace{(y\bar{\psi}\phi\psi + \text{h.c.})}_{\substack{\text{termes de Yukawa}\\\text{(interactions fermions-Higgs)}}} \tag{1.88}$$

où « h.c. » signifie conjugué hermitien. Dans cette forme simplifiée sont compris de nombreux termes,

car

$$\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu} = \boldsymbol{G}_{\mu\nu}\cdot\boldsymbol{G}^{\mu\nu} + \boldsymbol{W}_{\mu\nu}\cdot\boldsymbol{W}^{\mu\nu} + F^{(B)}_{\mu\nu}F^{(B)\mu\nu}\,, \tag{1.89}$$

$$\not{D} = \gamma^\mu D_\mu = \gamma^\mu\left[\partial_\mu - \mathrm{i}g_I I\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu - \mathrm{i}g_Y\frac{Y}{2}B_\mu - \mathrm{i}\alpha_s\frac{C}{2}\boldsymbol{\lambda}\cdot\boldsymbol{G}_\mu\right]\,, \tag{1.90}$$

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2\,. \tag{1.91}$$

Le modèle standard ainsi construit permet d'expliquer la plupart des phénomènes observés avec une grande précision. Cependant, il comporte certaines lacunes discutées dans la section suivante.

3.2 Limites

Certaines observations présentées de manière non exhaustive ci-après laissent à penser que le modèle standard est une théorie effective à basse énergie et qu'il existe une autre plus fondamentale permettant de les expliquer.

Nombre de générations Le modèle standard ne permet pas de prédire le nombre de générations, aujourd'hui égal à trois. Ainsi, il est tout à fait possible qu'il existe une quatrième génération de fermions (quarks t' et b', leptons τ' et ${\nu_\tau}'$). Seule contrainte connue à ce jour, les mesures sur la largeur du Z impliquent que seulement trois générations peuvent présenter des neutrinos de masse inférieure à $\frac{1}{2}m_Z$ [55]. Il faudrait ainsi nécessairement ${\nu_\tau}' > \frac{1}{2}m_Z$. Mesurer avec précision les paramètres de la matrice CKM permet de tester la présence d'une génération supplémentaire de fermions, cette matrice devant être unitaire.

Masse des neutrinos Dans le lagrangien du modèle standard (1.88), il n'existe pas de terme de masse pour les neutrinos. Or, des oscillations entre saveurs de neutrinos ont été observées [56, 57]. Celles-ci impliquent, d'une part, que les neutrinos possèdent en réalité une masse et, d'autre part, que les états propres de l'interaction faible ne sont pas les états propres de masse des neutrinos.

Les états propres de l'interaction faible des neutrinos peuvent être reliés à leurs états propres de masse à l'aide de la matrice PMNS [58], pour Pontecorvo, Maki, Nakagawa et Sakata, analogue à la matrice CKM introduite dans la section 2.3.4. Dans ce cas,

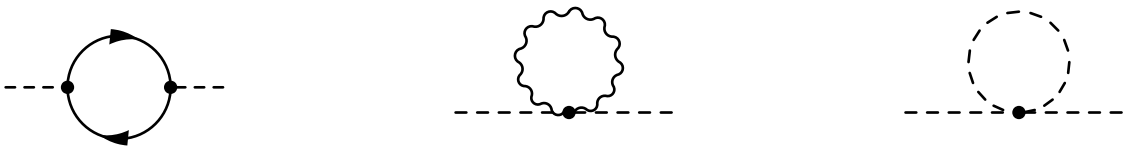
$$\begin{pmatrix}\nu_e\\ \nu_\mu\\ \nu_\tau\end{pmatrix} = \begin{pmatrix}U_{e1} & U_{\mu 2} & U_{\tau 3}\\ U_{e1} & U_{\mu 2} & U_{\tau 3}\\ U_{e1} & U_{\mu 2} & U_{\tau 3}\end{pmatrix}\begin{pmatrix}\nu_1\\ \nu_2\\ \nu_3\end{pmatrix}\,, \tag{1.92}$$

où ν_e, ν_μ et ν_τ sont les états propres de l'interaction faible et ν_1, ν_2 et ν_3 ceux de masse. La masse des neutrinos peut être obtenue par deux extensions différentes du modèle standard.

La première consiste à introduire directement des neutrinos de chiralité droite. Des telles particules, par construction, n'interagissent ni par interaction forte (pas de charge de couleur), ni par interaction électromagnétique (pas de charge électrique), ni par interaction faible (particules de chiralité droite). Plusieurs hypothèses, comme le mécanisme de Seesaw [59-63], sont avancées afin de décrire de tels neutrinos stériles. Il n'existe à ce jour aucun résultat expérimental permettant de conclure sur la validité de ces hypothèses.

La seconde revient à considérer que les neutrinos sont des particules de Majorana, c'est-à-dire qu'ils sont leurs propres antiparticules [64, 65]. Dans ce cas, les antineutrinos sont des neutrinos. Des neutrinos de chiralité droite sont ainsi naturellement introduits, rendant possible le terme de masse dans le lagrangien. Cependant, des neutrinos de Majorana peuvent violer la conservation du nombre de leptons lors de certains processus physiques. En particulier, la double désintégration β sans neutrinos [66, 67] est alors possible, mais ce phénomène n'a pas été observé.

Ajustement fin Les couplages du boson de Higgs aux particules rendent possible l'existence des diagrammes de la figure 1.17. Le calcul de la masse du boson de Higgs doit les prendre en compte.



(a) *Quark top.* **(b)** *Bosons vecteurs.* **(c)** *Couplage quadratique du Higgs.*

Figure 1.17 – *Diagrammes de Feynman des contributions aux corrections radiatives à la masse du boson de Higgs.*

Dans ce cas, la masse effective du boson de Higgs m_h s'exprime à partir de la masse « nue » m_{h0} à laquelle sont apportées des corrections [68] :

$$\text{du top, figure 1.17a} \qquad -\frac{3}{8\pi^2}y_t^2\Lambda^2 \qquad \sim -(2\,\text{TeV})^2 \tag{1.93}$$

$$\text{des bosons vecteurs, figure 1.17b} \qquad +\frac{1}{16\pi^2}g^2\Lambda^2 \qquad \sim +(0{,}7\,\text{TeV})^2 \tag{1.94}$$

$$\text{du Higgs, figure 1.17c} \qquad +\frac{1}{16\pi^2}\lambda^2\Lambda^2 \qquad \sim +(0{,}5\,\text{TeV})^2 \tag{1.95}$$

où y_t^2 est la constante de couplage de Yukawa du quark top défini équation (1.76), g la constante de couplage de l'interaction associée au boson vecteur considéré, λ le paramètre du terme quadratique du potentiel de Higgs (1.60) et Λ l'échelle d'énergie de la « nouvelle physique », c'est-à-dire l'échelle d'énergie où le modèle standard n'est plus valable.

La masse du boson de Higgs s'exprime alors

$$m_h^2 = m_{h0}^2 - \frac{3}{8\pi^2}y_t^2\Lambda^2 + \frac{1}{16\pi^2}g^2\Lambda^2 + \frac{1}{16\pi^2}\lambda^2\Lambda^2 + \dots \tag{1.96}$$

Le boson de Higgs a été observé avec une masse de $125{,}10 \pm 0{,}14\,\text{GeV}$. Or, les corrections précédentes sont de l'ordre du TeV. De plus, si d'autres particules plus massives existent, les corrections supplémentaires sont une à une encore plus importantes mais doivent toujours donner une masse de 125 GeV. Les paramètres dont toutes ces corrections dépendent doivent alors être ajustés au moins jusqu'à la 32$^{\text{e}}$ décimale, ce qui semble peu naturel. Une théorie qui ne nécessite pas un tel ajustement est préférable.

Nombre de paramètres libres Le modèle standard tel que décrit dans la section 2 comporte 19 paramètres libres, listés dans le tableau 1.3. Une théorie en comportant moins propose plus de prédictions, sur la masse des particules par exemple, ce qui permet de réaliser plus de comparaisons aux données expérimentales. Parmi ces 19 paramètres libres, 15 (les masses des 6 quarks et des 3 leptons chargés, la masse du Higgs, son condensat dans le vide et les quatre paramètres de la matrice CKM) sont reliés au boson de Higgs. C'est pourquoi l'étude du boson de Higgs est un enjeu majeur en physique des particules.

L'introduction de neutrinos massifs, précédemment évoquée, apporte trois nouveaux paramètres libres, les masses des trois neutrinos. De plus, la matrice PMNS permettant de décrire leurs oscillations demande quatre paramètres libres, analogues aux quatre paramètres reliés à la matrice CKM dans le tableau 1.3. Il y a donc potentiellement 26 paramètres libres pour un modèle standard incluant les neutrinos massifs.

Gravitation Le modèle standard ne propose aucune description de la gravitation. La relativité générale permet de décrire la gravitation à l'échelle macroscopique avec succès, mais cette théorie n'est pas applicable avec la mécanique quantique. En l'état actuel, l'ajout de la gravitation au modèle standard introduit des infinis qui le rendent non renormalisable. Le modèle standard est donc valide uniquement pour des échelles où la gravitation est négligeable face aux autres interactions (électromagnétique, faible et forte). C'est le cas dans les expériences de physique des particules ; aucune d'entre elles n'a observé d'effet gravitationnel dans ses résultats. Toutefois, le modèle standard ne peut pas être une théorie décrivant complètement l'Univers sans prendre en compte la gravitation.

Grandeur	Symbole	Valeur	
Masse du quark up	m_u	$2{,}16^{+0{,}49}_{-0{,}26}$	MeV
Masse du quark down	m_d	$4{,}67^{+0{,}48}_{-0{,}17}$	MeV
Masse du quark strange	m_s	93^{+11}_{-5}	MeV
Masse du quark charm	m_c	$1{,}27^{+0{,}02}_{-0{,}035}$	GeV
Masse du quark bottom	m_b	$4{,}18^{+0{,}03}_{-0{,}02}$	GeV
Masse du quark top	m_t	$172{,}76 \pm 0{,}30$	GeV
Masse de l'électron	m_e	$0{,}510\,998\,946\,1 \pm 0{,}000\,000\,003\,1$	MeV
Masse du muon	m_μ	$105{,}658\,374\,5 \pm 0{,}000\,002\,4$	MeV
Masse du tau	m_τ	$1776{,}86 \pm 0{,}12$	MeV
Angle de mélange CKM I-II	θ_{12}	$13{,}091 \pm 0{,}028$	°
Angle de mélange CKM II-III	θ_{23}	$2{,}323^{+0{,}048}_{-0{,}035}$	°
Angle de mélange CKM I-III	θ_{13}	$0{,}2068^{+0{,}0063}_{-0{,}0052}$	°
Phase de violation CP CKM	δ_{CKM}	$68{,}53^{+2{,}58}_{-2{,}46}$	°
Phase de violation CP forte	θ_{QCD}	$< 10^{-10}$	
Constante de couplage $U(1)_Y$	g_Y	$0{,}349\,70 \pm 0{,}000\,19$	
Constante de couplage $SU(2)_L$	g_I	$0{,}652\,95 \pm 0{,}000\,12$	
Constante de couplage $SU(3)_C$	α_s	$0{,}1179 \pm 0{,}0010$	
Condensat du champ de Higgs	v	$246 \pm 6 \times 10^{-5}$	GeV
Masse du boson de Higgs	m_h	$125{,}10 \pm 0{,}14$	GeV

Tableau 1.3 – *Valeurs expérimentales des 19 paramètres libres du modèle standard [9].*

Matière noire et énergie noire La matière noire est, par hypothèse, de la matière interagissant gravitationnellement mais pas électromagnétiquement. Son existence est postulée suite à l'observation des galaxies, dont les courbes de vitesses de rotation ne peuvent être expliquées à ce jour avec la relativité générale et la matière ordinaire seulement. Des effets de lentille gravitationnelle vont également en ce sens [69]. Le modèle standard ne propose aucun candidat pour la matière noire. D'autres observations cosmologiques montrent que l'expansion de l'Univers s'accélère. Cette accélération implique l'existence d'une énergie, l'énergie noire, inexpliquée par le modèle standard. Les mesures cosmologiques [70] indiquent que l'Univers est ainsi composé de 68 % d'énergie noire et 27 % de matière noire. Le modèle standard ne décrit donc que 5 % du contenu de l'Univers.

Asymétrie matière-antimatière Pour chaque particule de matière, il existe une antiparticule, dont les nombres quantiques sont opposés. Or, l'Univers est composé presque exclusivement de matière. La force électrofaible propose un mécanisme violant la symétrie CP pouvant mener à une asymétrie entre matière et antimatière, lié à la phase complexe de la matrice CKM, δ_{CKM}. Cependant, cet effet est trop faible pour expliquer à lui seul l'asymétrie effectivement observée.

4 Au-delà du modèle standard

Le modèle standard souffre ainsi de lacunes malgré ses prédictions précises. Des modèles sont développés afin de les combler, ils sont dits « au-delà » du modèle standard (BSM, *Beyond Standard Model*). Un de ces modèles BSM est la supersymétrie (SUSY). La section 4.1 présente la SUSY et motive l'introduction d'un second doublet de Higgs. La SUSY est en effet un cas particulier de modèle à deux doublets de Higgs (2HDM, *2 Higgs Doublets Model*). Les 2HDM sont abordés dans la section 4.2. Puis, le modèle le plus simple de SUSY, l'extension supersymétrique minimale du modèle standard (MSSM, *Minimal Supersymmetric extension of Standard Model*), est présenté section 4.3.

4.1 Supersymétrie

La supersymétrie (SUSY) [71-73] introduit une nouvelle symétrie entre fermions et bosons. Ces deux types de particules ne sont plus indépendants, ce sont des saveurs, ou manifestations, d'un champ quantique plus complexe. Du fait de cette symétrie, à chaque particule du modèle standard correspond une nouvelle particule ou « superpartenaire ». Les fermions du modèle standard ont des superpartenaires de spin entier, c'est-à-dire des bosons, les « sfermions ». Les bosons du modèle standard ont des superpartenaires de spin demi-entier, c'est-à-dire des fermions, les « bosinos ». Une particule et son superpartenaire ont les mêmes nombres quantiques à l'exception de leurs spins.

De nouvelles interactions sont possibles, dans lesquelles les nombres baryonique B et leptonique L ne sont pas conservés et $B - L$ non plus. Le proton est alors instable, ce qui n'est pas observé expérimentalement. Une nouvelle symétrie est ainsi introduite afin de restaurer la conservation de $B - L$, la parité R. L'opérateur de parité R est défini comme

$$P_R = (-1)^{3(B-L)-2s} \tag{1.97}$$

où s correspond au spin de la particule. La parité R est égale à 1 pour les particules du modèle standard, -1 pour leurs superpartenaires. La conservation de cette nouvelle parité permet non seulement de garder le proton stable, mais rend également stable la particule supersymétrique de plus basse masse, notée LSP (*Lightest Supersymmetric Particle*).

La SUSY permet de résoudre de nombreuses lacunes du modèle standard. Elle permet par exemple d'unifier trois forces fondamentales (électromagnétique, faible et forte). Dans la section 2.3, l'unification des forces électromagnétique et faible est déjà réalisée. Toutefois, la force électrofaible et la force forte ne semblent pas s'unifier à haute énergie. Or, les interactions avec les superpartenaires introduits par la SUSY modifient le comportement des constantes de couplage des trois forces fondamentales de manière à les unifier à haute énergie. La SUSY propose également un candidat pour la matière noire dans le cas où la LSP est de charge électrique nulle, il peut être un neutralino ou un sneutrino. De plus, la SUSY permet de résoudre le problème de l'ajustement fin. La divergence quadratique de la masse du boson de Higgs est naturellement supprimée par les diagrammes à boucles des superpartenaires dont les contributions ont des signes opposés à celles des particules, les fermions ayant des contributions positives et les bosons des contributions négatives [74].

Toutefois, il est impossible de mettre en place la SUSY sans un second doublet de Higgs. Dans le modèle standard, la masse des quarks d'isospin faible haut est obtenue dans la section 2.4.3 à l'aide du conjugué de charge du doublet de Higgs. Cependant, le potentiel supersymétrique contenant les termes de Yukawa, nécessaires à l'obtention des masses des fermions, n'autorise pas l'utilisation de ce conjugué afin de donner une masse aux quarks d'isospin faible haut [68]. Un second doublet de Higgs, couplé aux fermions d'isospin faible haut, doit nécessairement être introduit [68, 74]. La SUSY est donc un cas particulier de modèle à deux doublets de Higgs.

4.2 Modèles à deux doublets de Higgs

Dans les modèles à deux doublets de Higgs (2HDM), le doublet de Higgs du modèle standard ϕ défini par (1.56) est remplacé par deux doublets ϕ_1 et ϕ_2 et le potentiel de Higgs (1.60) par le potentiel scalaire le plus général possible brisant spontanément $SU(2)_L \times U(1)_Y$ [74, 75],

$$\begin{aligned} V(\phi_1,\phi_2) = {} & \lambda_1 \left(\phi_1^\dagger\phi_1 - \frac{1}{2}v_1^2\right)^2 + \lambda_2 \left(\phi_2^\dagger\phi_2 - \frac{1}{2}v_2^2\right)^2 \\ & + \lambda_3 \left[\left(\phi_1^\dagger\phi_1 - \frac{1}{2}v_1^2\right) + \left(\phi_2^\dagger\phi_2 - \frac{1}{2}v_2^2\right)\right]^2 + \lambda_4 \left[(\phi_1^\dagger\phi_1)(\phi_2^\dagger\phi_2) - (\phi_1^\dagger\phi_2)(\phi_2^\dagger\phi_1)\right] \\ & + \lambda_5 \left[\Re\left(\phi_1^\dagger\phi_2\right) - \frac{1}{2}v_1v_2\cos\xi\right]^2 + \lambda_6 \left[\Im\left(\phi_1^\dagger\phi_2\right) - \frac{1}{2}v_1v_2\sin\xi\right]^2 \\ & + \lambda_7 \left[\Re\left(\phi_1^\dagger\phi_2\right) - \frac{1}{2}v_1v_2\cos\xi\right]\left[\Im\left(\phi_1^\dagger\phi_2\right) - \frac{1}{2}v_1v_2\sin\xi\right], \end{aligned} \tag{1.98}$$

où λ_i, $i \in \{1,\dots,7\}$ et ξ sont des paramètres du modèle à valeurs réelles. Dans le cas de la SUSY, $\lambda_5 = \lambda_6$. Le dernier terme de (1.98) peut être éliminé en redéfinissant les phases des champs scalaires.

Dans le cas $\sin\xi \neq 0$, le secteur de Higgs du modèle viole la symétrie CP, ce cas est abordé dans la section 4.4. Le minimum du potentiel ainsi construit est tel que

$$\langle\phi_1\rangle_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v_1\end{pmatrix}, \quad \langle\phi_2\rangle_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v_2 e^{i\xi}\end{pmatrix}. \tag{1.99}$$

Il est possible de définir, à ce stade, une variable importante dans la suite, le rapport des condensats des doublets de Higgs dans le vide,

$$\tan\beta = \frac{\langle\phi_2\rangle_0}{\langle\phi_1\rangle_0} = \frac{v_2}{v_1} \tag{1.100}$$

avec $0 \leqslant \beta \leqslant \pi/2$. Il est aussi possible de définir

$$v^2 = v_1^2 + v_2^2. \tag{1.101}$$

De ce formalisme découle l'existence de cinq bosons de Higgs massifs,

deux Higgs chargés : $$H^\pm = -\phi_1^\pm \sin\beta + \phi_2^\pm \cos\beta, \qquad m_{H^\pm}^2 = \frac{1}{2}\lambda_4 v^2, \tag{1.102}$$

un Higgs pseudo-scalaire : $$A = \sqrt{2}\left(-\Im(\phi_1^0)\sin\beta + \Im(\phi_2^0)\cos\beta\right), \qquad m_A^2 = \frac{1}{2}\lambda_6 v^2, \tag{1.103}$$

ainsi que deux bosons de Higgs scalaires neutres dont les champs quantiques sont mélangés par la matrice

$$\mathcal{M} = \frac{1}{2}\begin{pmatrix}4v_1^2(\lambda_1+\lambda_3)+v_2^2\lambda_5 & (4\lambda_3+\lambda_5)v_1v_2\\ (4\lambda_3+\lambda_5)v_1v_2 & 4v_2^2(\lambda_2+\lambda_3)+v_1^2\lambda_5\end{pmatrix}. \tag{1.104}$$

Ces deux bosons de Higgs sont

$$h = \sqrt{2}\left(-\Re\left(\phi_1^0 - v_1/\sqrt{2}\right)\sin\alpha + \Re\left(\phi_2^0 - v_2/\sqrt{2}\right)\cos\alpha\right), \tag{1.105}$$

$$H = \sqrt{2}\left(\Re\left(\phi_1^0 - v_1/\sqrt{2}\right)\cos\alpha + \Re\left(\phi_2^0 - v_2/\sqrt{2}\right)\sin\alpha\right), \tag{1.106}$$

où l'angle de mélange α s'obtient par

$$\sin 2\alpha = \frac{2\mathcal{M}_{12}}{\sqrt{(\mathcal{M}_{11}-\mathcal{M}_{22})^2+4\mathcal{M}_{12}^2}}, \quad \cos 2\alpha = \frac{\mathcal{M}_{11}-\mathcal{M}_{22}}{\sqrt{(\mathcal{M}_{11}-\mathcal{M}_{22})^2+4\mathcal{M}_{12}^2}} \tag{1.107}$$

avec $-\pi/2 \leqslant \alpha \leqslant 0$ et dont les masses à l'ordre le plus bas s'expriment, avec $m_h \leqslant m_H$,

$$m_{h,H}^2 = \frac{1}{2}\left(\mathcal{M}_{11}+\mathcal{M}_{22} \mp \sqrt{(\mathcal{M}_{11}-\mathcal{M}_{22})^2+4\mathcal{M}_{12}^2}\right). \tag{1.108}$$

Enfin, v est fixée par la masse du W,

$$m_W = \frac{1}{2}v g_I. \tag{1.109}$$

Le 2HDM ainsi construit possède 6 paramètres libres :
— m_h, m_H, m_A, $m_{H^\pm}$ les masses des bosons de Higgs ;
— $\tan\beta$ le rapport des condensats des doublets de Higgs dans le vide ;
— α l'angle de mélange des Higgs.

Ce modèle peut être affiné par les observations expérimentales. Par exemple, le changement de saveur par courant neutre (FCNC, *Flavor-Changing Neutral Currents*), n'est pas observé expérimentalement. Afin d'être compatible avec ce fait expérimental [74],
— soit les masses des bosons de Higgs sont élevées, de l'ordre du TeV, supprimant ainsi suffisamment le FCNC pour rester dans les limites observées ;
— soit tous les fermions portant une même charge électrique ne sont couplés qu'à un seul doublet de Higgs au plus.

La masse du boson de Higgs du modèle standard n'étant pas de l'ordre du TeV, la seconde option est choisie.

Dans les modèle de type I, les fermions ne sont pas couplés à ϕ_1, mais le sont à ϕ_2. Dans le cas des modèles de type II, les fermions d'isospin faible bas sont couplés à ϕ_1 et ceux d'isospin faible haut à ϕ_2. Les intensités des couplages des fermions et des bosons avec h, H et A ainsi obtenues, par rapport à celles avec le boson de Higgs du modèle standard, sont présentées dans le tableau 1.4.

Couplage avec	h	H	A
Bosons vecteurs	$\sin(\beta-\alpha)$	$\cos(\beta-\alpha)$	0
Fermions hauts	$\frac{\cos\alpha}{\sin\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$\cot\beta$
Fermions bas	$\frac{-\sin\alpha}{\cos\beta}$	$\frac{\cos\alpha}{\cos\beta}$	$\tan\beta$

Tableau 1.4 – *Couplages des bosons de Higgs neutres des modèles de type II par rapport aux couplages du boson de Higgs du modèle standard [74].*

Les modèles à deux doublets de Higgs sont donc une extension du modèle standard ajoutant une nouvelle physique, en particulier l'existence de nouveaux bosons de Higgs. Ils doivent nécessairement être introduits dans les modèles supersymétriques pour que ceux-ci respectent les observations expérimentales.

4.3 Extension supersymétrique minimale du modèle standard

L'extension supersymétrique minimale du modèle standard ou MSSM [76, 77] est le modèle le plus simple permettant d'introduire la SUSY tout en étant compatible avec les observations expérimentales à ce jour. Dans le MSSM, les deux doublets de Higgs s'expriment en fonction de ϕ_1 et ϕ_2 introduits dans la section 4.2 comme [74]

$$H_d = \begin{pmatrix} \phi_1^{0*} \\ -\phi_1^- \end{pmatrix}, \quad H_u = \begin{pmatrix} \phi_2^+ \\ \phi_2^0 \end{pmatrix}. \tag{1.110}$$

L'expression du potentiel de Higgs général des 2HDM (1.98) devient

$$\begin{aligned} V(H_d, H_u) = {} & \mu_d^2 H_d^\dagger H_d + \mu_u^2 H_u^\dagger H_u - \mu^2 (H_d \wedge H_u + \text{h.c.}) \\ & + \frac{g_I^2 + g_Y^2}{8} (H_d^\dagger H_d - H_u^\dagger H_u)^2 + \frac{g_I^2}{2} (H_d^\dagger H_u)^2 \end{aligned} \tag{1.111}$$

en posant [68, 74, 75]

$$\lambda_2 = \lambda_1, \tag{1.112a}$$

$$\lambda_3 = \frac{1}{8}(g_I^2 + g_Y^2) - \lambda_1, \tag{1.112b}$$

$$\lambda_4 = 2\lambda_1 - \frac{1}{2} g_Y^2, \tag{1.112c}$$

$$\lambda_5 = \lambda_6 = 2\lambda_1 - \frac{1}{2}(g_I^2 + g_Y^2), \tag{1.112d}$$

$$\mu_d^2 = \mu^2 \tan\beta - \frac{1}{2} m_Z^2 \cos(2\beta), \tag{1.112e}$$

$$\mu_u^2 = \mu^2 \cot\beta + \frac{1}{2} m_Z^2 \cos(2\beta), \tag{1.112f}$$

$$\mu^2 = -\frac{1}{2} v_1 v_2 (g_I^2 + g_Y^2 - 4\lambda_1). \tag{1.112g}$$

Afin d'assurer la stabilité du vide, le potentiel ne doit pas pouvoir être infiniment bas, ce qui implique $\mu_u^2 + \mu_d^2 > 2\mu^2$ si $|\phi_1^0| = |\phi_2^0|$. La brisure spontanée de symétrie donnant leurs masses aux bosons de l'interaction faible est présente si $\mu^4 > \mu_u^2 \mu_d^2$. Alors, les condensats dans le vide des deux doublets de Higgs sont

$$\langle H_d \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} v_1 \\ 0 \end{pmatrix}, \quad \langle H_u \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2 \end{pmatrix}. \tag{1.113}$$

Les masses des bosons de Higgs s'expriment alors à l'ordre le plus bas

$$m_A^2 = \mu^2(\tan\beta + \cot\beta) = \frac{2\mu^2}{\sin 2\beta}, \tag{1.114}$$

$$m_{H^\pm}^2 = m_A^2 + m_W^2, \tag{1.115}$$

$$m_{h,H}^2 = \frac{1}{2}\left(m_A^2 + m_Z^2 \mp \sqrt{(m_A^2 + m_Z^2)^2 - 4 m_A^2 m_Z^2 \cos^2 2\beta} \right), \tag{1.116}$$

et l'angle de mélange des Higgs scalaires neutres vérifie

$$\cos 2\alpha = -\frac{m_A^2 - m_Z^2}{m_H^2 - m_h^2} \cos 2\beta, \quad \sin 2\alpha = -\frac{m_A^2 + m_Z^2}{m_H^2 - m_h^2} \sin 2\beta. \tag{1.117}$$

Enfin, la masse du W vérifie toujours (1.109) et celle du Z peut être exprimée en fonction des paramètres du MSSM. Ainsi,

$$m_W = \frac{1}{2} v g_I \Leftrightarrow m_W^2 = \frac{1}{4} v^2 g_I^2 , \quad m_Z^2 = \frac{\mu_d^2 - \mu_u^2 \tan^2 \beta}{\tan^2 \beta - 1} . \tag{1.118}$$

À l'ordre le plus bas, les masses des bosons de Higgs dépendent donc seulement de deux paramètres libres, m_A et $\tan\beta$ défini par (1.100). Il est à noter que l'équation (1.116) implique l'existence d'un boson de Higgs neutre de masse inférieure à $m_Z = 91{,}19\,\text{GeV}$. Toutefois, ceci n'est vrai qu'à l'ordre le plus bas. La prise en compte des corrections radiatives change les expressions de ces masses, en particulier celles dues au quark top et à son superpartenaire le stop. La masse du boson de Higgs léger est ainsi modifiée selon [68]

$$m_h^2 \to m_h^2 + \delta m_h^2 \simeq m_Z^2 + \frac{3 m_t^4}{2\pi^2 v^2} \left[\ln \frac{m_S^2}{m_t^2} + \frac{X_t^2}{m_S^2} \left(1 - \frac{X_t^2}{12 m_S^2} \right) \right] \tag{1.119}$$

où

$$X_t = A_t - \mu \cot \beta \tag{1.120}$$

est le paramètre de mélange du stop, avec A_t sa constante de couplage aux Higgs, et

$$m_S = \sqrt{m_{\tilde{t}_1} m_{\tilde{t}_2}} . \tag{1.121}$$

est l'échelle d'énergie de la SUSY, définie comme la moyenne géométrique des masses des stops. Il existe en effet deux états propres de masse pour ces particules, $\tilde{t}_1$ et $\tilde{t}_2$, mélanges des stops de chiralité droite et gauche. La présence de nombreux paramètres libres du MSSM mène à proposer des scénarios dans lesquels les paramètres intervenant dans les corrections d'ordres supérieurs sont fixés [78, 79]. Il ne reste alors que deux paramètres libres, m_A et $\tan\beta$. Les valeurs des masses corrigées de h, H et $H^\pm$ sont tracées sur la figure 1.18 en fonction de m_A pour $\tan\beta = 3$ et 30 dans le cas de mélange maximal du stop avec $m_{\tilde{t}} = 2\,\text{TeV}$ et les autres paramètres de la SUSY fixés à 1 TeV [68]. Ainsi, pour $m_A \lesssim 125\,\text{GeV}$, H joue le rôle du modèle standard et il existe un Higgs léger. Pour $m_A \gtrsim 125\,\text{GeV}$, h joue le rôle du modèle standard et les bosons de Higgs supplémentaires sont de masses plus élevées. Dans ce cas, $m_H \simeq m_A$.

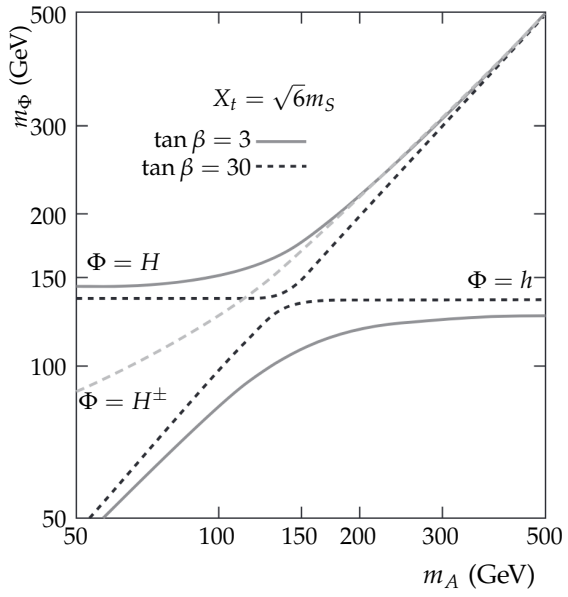


Figure 1.18 – *Masses des bosons de Higgs du MSSM en fonction de m_A pour $\tan\beta = 3$ et 30 dans le cas de mélange maximal du stop avec $m_{\tilde{t}} = 2\,\text{TeV}$ et les autres paramètres de la SUSY fixés à 1 TeV [68].*

Les particules du MSSM et leurs superpartenaires sont résumés dans le tableau 1.5. Un test expérimental est possible par la recherche d'un signal correspondant aux bosons de Higgs supplémentaires, ce qui fait l'objet de cette thèse. L'étude de la phénoménologie de ces bosons de Higgs, présentée ci-après, permet de déterminer les conditions favorables à la recherche d'un tel signal.

Particules				Superpartenaires			
Type	Spin	Particules	Symboles	Type	Spin	Particules	Symboles
Fermions	$\frac{1}{2}$	quarks	q	Sfermions	0	squarks	$\tilde{q}$
		leptons	ℓ			sleptons	$\tilde{\ell}$
Bosons	1	gluon	g	Bosinos	$\frac{1}{2}$	gluino	$\tilde{g}$
		bosons $W^{\pm}$	W^+, W^-			winos	$\tilde{W}^+$, $\tilde{W}^-$
		photon	γ			photino	$\tilde{\gamma}$
		boson Z	Z			zino	$\tilde{Z}$
	0	Higgs	h, H, A, $H^{\pm}$			Higgsinos	$\tilde{h}$, $\tilde{H}$, $\tilde{A}$, $\tilde{H}^{\pm}$

Tableau 1.5 – *Particules et leurs superpartenaires. La présence de plusieurs bosons de Higgs est justifiée par la nécessité d'un second doublet de Higgs. Ce formalisme est décrit dans la section 4.2.*

4.4 Violation de CP avec les bosons de Higgs

L'asymétrie matière-antimatière ne peut être expliquée par la seule violation de CP de l'interaction faible. Or, une violation de CP dans les couplages de Yukawa est possible lorsque le paramètre ξ de l'équation (1.98) vérifie $\sin\xi \neq 0$.

Les termes de Yukawa s'écrivent en général, pour un boson de Higgs Φ,

$$\mathcal{L}_{\text{Yukawa}} = -\frac{m\sqrt{2}}{v}\left(\cos(\varphi)\bar{\psi}\Phi\psi + \sin(\varphi)\bar{\psi}i\gamma^5\Phi\psi\right) \tag{1.122}$$

avec φ l'angle de mélange CP de Φ. Lorsque $\varphi = 0$, Φ est CP-pair, c'est-à-dire scalaire avec un nombre quantique $J^{CP} = 0^{++}$. Pour $\varphi = \frac{\pi}{2}$, Φ est CP-impair, c'est-à-dire pseudo-scalaire avec un nombre quantique $J^{CP} = 0^{+-}$. Pour toutes les autres valeurs de φ, l'état est mixte et permet une violation de CP.

Dans le MSSM sans violation de CP, les états propres de masse des bosons de Higgs neutres sont également des états propres de CP. Ainsi h et H sont CP-pairs et A est CP-impair. Avec violation de CP, les états propres de masse sont des états mixtes de CP. Afin de les distinguer de h, H et A, ils sont notés, par masse croissante, H_1, H_2 et H_3.

Les sections qui suivent traitent du cas sans violation de CP. Le cas avec violation correspond, dans le cadre de ma thèse, au passage $\{h, H, A\} \to \{H_1, H_2, H_3\}$.

5 Phénoménologie des bosons de Higgs du MSSM

Pour concevoir une analyse de physique des particules à même de tester le MSSM, il faut dans un premier temps déterminer la manifestation de ce modèle à observer. Comme cela a été développé dans la section précédente, le MSSM implique l'existence de quatre bosons de Higgs supplémentaires dont deux neutres, H et A. S'ils existent, un signal leur correspondant doit pouvoir être observé.

Au premier ordre, les masses des bosons de Higgs s'expriment en fonction de deux paramètres uniquement, m_A et $\tan\beta$. Les couplages des trois bosons de Higgs neutres du MSSM aux autres particules, par rapport aux couplages du boson de Higgs du modèle standard, sont présentés dans le tableau 1.4 en fonction de α et β. Or, ces deux paramètres sont reliés par les équations (1.117), donnant

$$\tan 2\alpha = \frac{m_A^2 + m_Z^2}{m_A^2 - m_Z^2}\tan 2\beta\,. \tag{1.123}$$

Les observations expérimentales semblent favoriser $m_A \gg m_Z$ [25, 26, 80, 81]. Dans ce cas, H et A sont de masses similaires et h prend le rôle du boson de Higgs du modèle standard observé expérimentalement en 2012 [22-24]. Cette situation correspond à la limite découplée, dans laquelle

$$\lim_{m_A \gg m_Z} \tan 2\alpha = \tan 2\beta \tag{1.124}$$

d'après (1.123). Alors, dans cette limite, $\alpha \sim \beta$ ou $\alpha \sim \beta \pm \frac{\pi}{2}$. Or, $\beta \geqslant 0$ et $\alpha \leqslant 0$ et $\tan\beta$ est contraint par [82]

$$1 < \tan\beta \lesssim \frac{m_t}{m_b} \simeq 42\,. \tag{1.125}$$

Il ne reste donc plus que la possibilité $\alpha \sim \beta - \frac{\pi}{2}$. Les couplages du tableau 1.4 deviennent alors ceux du tableau 1.6.

Couplage avec	h	H	A
Bosons vecteurs	~ 1	~ 0	0
Fermions hauts	~ 1	$\sim -\cot\beta$	$\cot\beta$
Fermions bas	~ 1	$\sim \tan\beta$	$\tan\beta$

Tableau 1.6 – *Couplages des bosons de Higgs neutres dans la limite découplée du MSSM par rapport aux couplages du boson de Higgs du modèle standard.*

Les couplages ainsi obtenus dans le tableau 1.6 présentent trois caractéristiques d'intérêt :
— h se comporte comme le boson de Higgs du modèle standard, ce qui le rend tout à fait cohérent avec les observations actuelles ;
— H et A présentent peu voire aucun couplage aux bosons vecteurs, par exemple la désintégration $A \to ZZ$ est impossible mais $A \to Zh$ est possible ;
— H et A sont couplés de manières similaires aux fermions.

Lorsque $\tan\beta$ augmente, les couplages de H et A aux fermions d'isospin faible bas sont augmentés et ceux aux fermions d'isospin faible haut supprimés. La production et les désintégrations de H et A sont donc intrinsèquement liées à la présence de fermions d'isospin faible bas.

5.1 Production de bosons de Higgs

La production de bosons de Higgs au LHC peut être réalisée selon plusieurs modes dont un calcul précis des sections efficaces a pu être fait dans le cadre du modèle standard [83-86]. Ces sections efficaces sont représentées en fonction de l'énergie de collision dans le centre de masse sur la figure 1.19. Les processus correspondant à ces différents modes sont présentés dans la section 5.1.1 ci-après. Puis, la production des bosons de Higgs du MSSM est discutée dans la section 5.1.2.

5.1.1 Dans le cadre du modèle standard

Le mode de production principal du boson de Higgs du modèle standard h au LHC est la fusion de gluon. Près de 85 % des bosons de Higgs sont produits ainsi. Ce mode est noté ggh et est représenté figure 1.20a. L'interaction entre gluons et Higgs est réalisée par une boucle de quarks. Or, comme cela est vu dans la section 2.4.3 le couplage du Higgs aux fermions est proportionnel à leur masse. Le quark top est donc dominant dans cette boucle.

Le second mode de production de Higgs le plus important au LHC est la fusion de boson vecteur, noté VBF (*Vector Boson Fusion*) et représenté sur les figures 1.20b et 1.20c. Deux quarks produisent chacun un boson vecteur (W^+ et W^- ou deux Z). Ces deux bosons fusionnent en un boson de Higgs. Bien que la section efficace du VBF soit dix fois plus faible que celle du ggh, les deux quarks de l'état final donnent deux jets très caractéristiques, dont le processus de formation à partir des quarks est abordé dans la section 2.6. Le calcul de la section efficace de ce processus inclu les corrections QCD (chromodynamique quantique, *Quantum ChromoDynamics*), abordée dans la section 2.5, au NNLO et les corrections électrofaibles au NLO [86, 87]. Les notations NLO, NNLO, N^{3}LO, etc. signifient *next-to-leading order*, c'est-à-dire jusqu'à l'ordre suivant le premier degré non nul ; *next-to-next-to-leading order*, un ordre de plus que NLO ; etc.

La production d'un boson de Higgs peut également se faire en association avec un boson vecteur, c'est le mode VH. Une paire quark-antiquark produit un boson vecteur de haute énergie (W sur la figure 1.21a ou Z sur la figure 1.21b). Ce boson émet alors un Higgs, d'où la dénomination « Higgs-strahlung » parfois utilisée pour le mode VH. Les sections efficaces de ces processus sont calculées

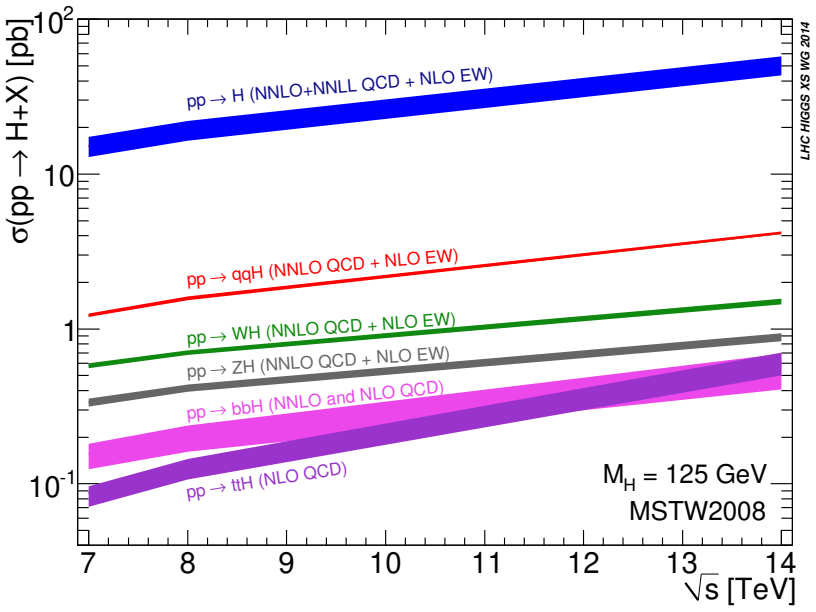
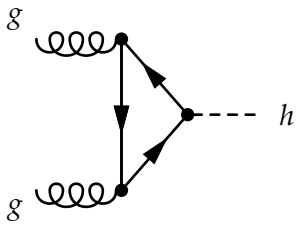
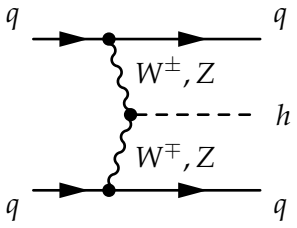


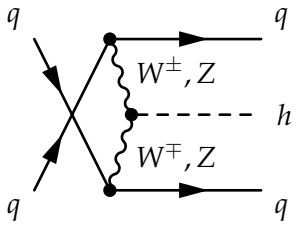
Figure 1.19 – *Sections efficaces des modes de production du boson de Higgs du modèle standard [83-86].*



(a) *Production par fusion de gluons (ggh).*

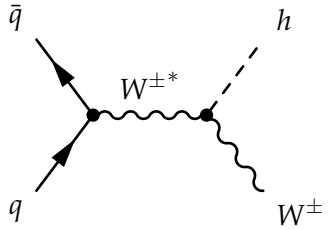


(b) *Production par fusion de bosons vecteurs (VBF h) en voie t.*

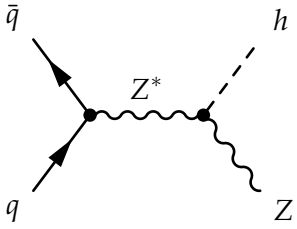


(c) *Production par fusion de bosons vecteurs (VBF h) en voie u.*

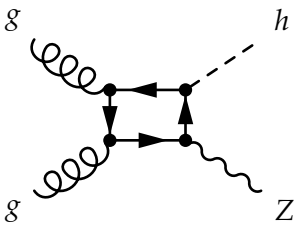
Figure 1.20 – *Diagrammes de Feynman de production de boson de Higgs dans le cadre du modèle standard par fusion de gluons (ggh) et fusion de bosons vecteurs (VBF).*



(a) *Production en association avec un boson W (Wh).*

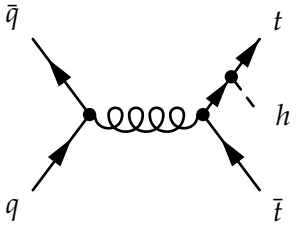


(b) *Production en association avec un boson Z (Zh).*

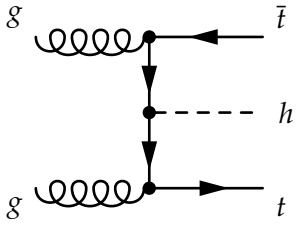


(c) *Production par fusion de gluons associée à un boson Z (ggZh).*

Figure 1.21 – *Diagrammes de Feynman de production de boson de Higgs dans le cadre du modèle standard en association avec un boson vecteur (VH).*



(a)



(b)

Figure 1.22 – *Diagrammes de Feynman de production de boson de Higgs dans le cadre du modèle standard en association avec un quark top ($t\bar{t}h$).*

en prenant en compte les corrections QCD NNLO et les corrections électrofaibles au NLO [86]. Une fusion de gluons peut également amener à la production d'un Higgs en association avec un Z, c'est le cas du processus de la figure 1.21c.

Enfin, il est possible de produire un Higgs en association avec des fermions lourds, en particulier des quarks top ($t\bar{t}h$) ou bottom ($b\bar{b}h$). Dans ces cas, des jets sont également formés. Le mode $t\bar{t}h$ est illustré sur la figure 1.22. Ces processus contribuent peu à la production de bosons de Higgs au LHC dans le cadre du modèle standard. Cependant, la phénoménologie du MSSM peut rendre les modes de production en association avec des quarks b significatifs voire dominants.

5.1.2 Dans le cadre du MSSM

Dans la limite découplée du MSSM, compte-tenu des couplages des bosons de Higgs h, H et A du tableau 1.6, les processus présentés dans la section précédente sont modifiés. Ainsi, la fusion de gluons de la figure 1.20a permet, dans le MSSM, de produire h, H et A. Il s'agit toujours du mode dominant si $\tan\beta$ ne prend pas de valeur élevée. Dans le cas de la production de h et H, la boucle peut également contenir des contributions des squarks stop et sbottom [88]. Le mode VBF, dont les processus sont présentés sur les figures 1.20b et 1.20c, permet de produire h et H, mais pas A. Les corrections aux ordres supérieurs de ces diagrammes dues à la QCD supersymétrique sont faibles et celles dues à la force électrofaible supersymétrique de l'ordre du pourcent [83]. Ces nouveaux processus sont représentés sur la figure 1.23.



(a) *Production par fusion de gluons.* **(b)** *Production par fusion de bosons vecteurs en voie t.*

Figure 1.23 – *Diagrammes de Feynman de production de boson de Higgs dans le cadre du MSSM par fusion de gluons (ggH) et fusion de bosons vecteurs (VBF).*

Dans le mode VH, le Higgs émis peut également être un H. Les processus de la figure 1.21 sont ainsi modifiés en ceux de la figure 1.24. Les corrections aux ordres supérieurs de ces diagrammes dues à la QCD supersymétrique sont faibles et celles dues à la force électrofaible supersymétrique ne sont pas connues [83].



(a) *Production en association avec un boson W.* **(b)** *Production en association avec un boson Z.*

Figure 1.24 – *Diagrammes de Feynman de production de boson de Higgs dans le cadre du MSSM en association avec un boson.*

Pour de grandes valeurs de $\tan\beta$, la production de Higgs lourds en association avec des quarks b est un mode dominant. Plusieurs processus participent à ce mode. Sur les figures 1.25a et 1.25b se trouvent les processus analogues à ceux du mode $t\bar{t}h$ du modèle standard présentés figures 1.22a et 1.22b. Des processus comme ceux des figures 1.25c et 1.25d sont également envisageables si le quark b est considéré comme présent au sein du proton, c'est le « schéma à cinq saveurs » ou 5fs (*5-flavor scheme*). Les sections efficaces des processus du mode $b\bar{b}h$ sont calculées au NLO dans le 4fs et au NNLO pour le 5fs.

Les processus des figures 1.25a et 1.25b présentent deux jets de quarks b associés à la production d'un boson de Higgs, celui de la figure 1.25d un jet de quark b. L'identification de ces jets est donc un enjeu dans les analyses testant les cas de hautes valeurs de $\tan\beta$.

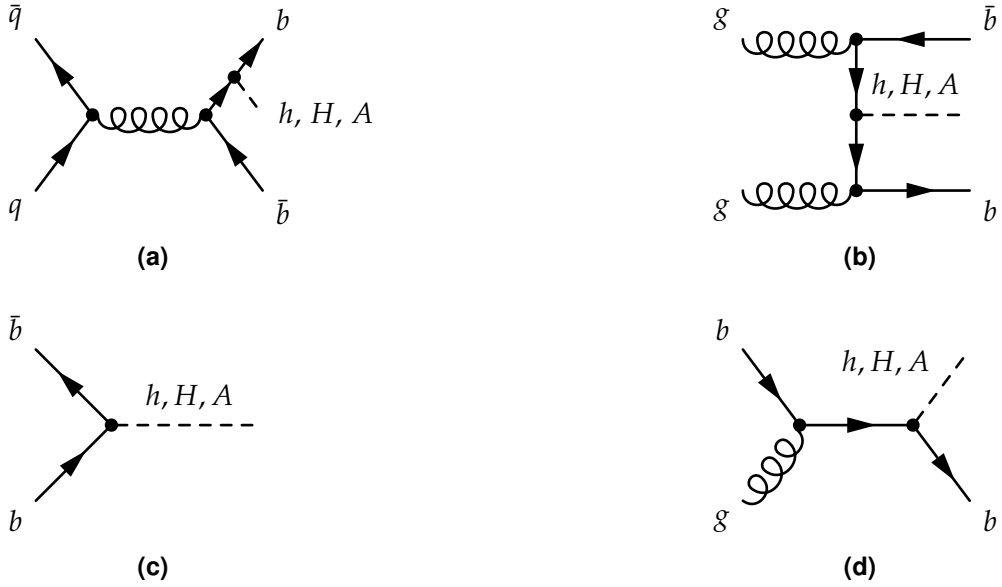


Figure 1.25 – *Diagrammes de Feynman de production de boson de Higgs dans le cadre du MSSM en association avec un quark b.*

5.2 Désintégrations des bosons de Higgs

Les bosons de Higgs ont une durée de vie très courte, de l'ordre de 10^{-22} s pour le boson de Higgs du modèle standard par exemple [9]. Il est donc impossible d'observer directement la présence d'un boson de Higgs, comme cela peut se faire avec d'autres particules plus stables comme les kaons, les muons, les électrons ou les protons par exemple. Pour étudier ces bosons, il faut donc observer leurs produits de désintégration.

La désintégration des bosons de Higgs peut se faire sous différentes formes ayant différents rapports de branchement ou $\mathcal{BR}$ (*Branching Ratio*), c'est-à-dire différentes probabilités de survenir. La topologie des événements correspondants est également fortement affectée par les produits de désintégration des bosons de Higgs.

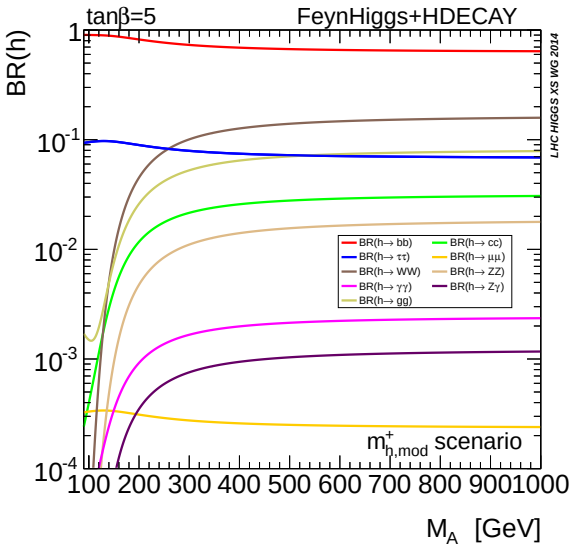
Dans la limite découplée du MSSM, les couplages des bosons de Higgs supplémentaires H et A aux bosons vecteurs sont supprimés et ceux aux fermions sont proportionnels :

— aux masses de ces derniers ;
— à $1/\tan\beta = \cot\beta$ pour ceux d'isospin faible haut ;
— à $\tan\beta$ pour ceux d'isospin faible bas.

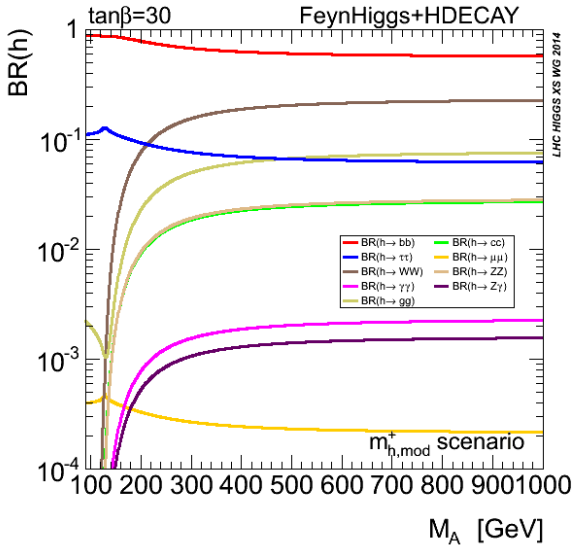
Les rapports de branchement qui en découlent sont représentés sur la figure 1.26 pour les trois bosons de Higgs neutres du MSSM et pour $\tan\beta = 5$ et 30. Pour des masses de H et A suffisamment grandes pour leur permettre de se désintégrer en paire de quarks top, malgré la masse élevée de ce dernier, la suppression des couplages aux fermions d'isospin faible haut par $\tan\beta$ laisse le quark b et le lepton τ avec les rapports de branchement les plus élevés à haut $\tan\beta$. Les rapports de branchement du boson de Higgs h correspondant au boson de Higgs du modèle standard sont peu affectés par $\tan\beta$.

La valeur de $\tan\beta$ est un paramètre libre du MSSM pouvant être grand. À haut $\tan\beta$ le quark b et le lepton τ proposent les rapports de branchement les plus grands à H et A. Pour des valeurs modérées voire basses de $\tan\beta$, le quark top peut éventuellement proposer un rapport de branchement plus grand, mais seulement pour $m_A \gtrsim 350\,\text{GeV}$. Les désintégrations en $b\bar{b}$ et en $\tau^+\tau^-$ sont donc les plus prometteuses pour la recherche de bosons de Higgs supplémentaires de haute masse.

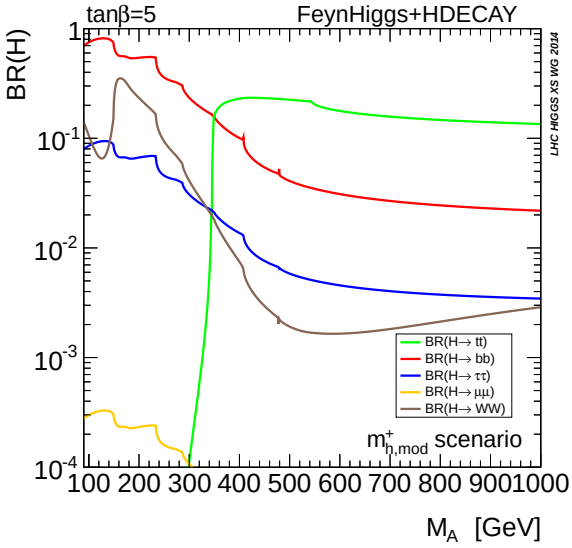
Bien que le canal de désintégration $h, H, A \to b\bar{b}$ possède un rapport de branchement 5 à 10 fois supérieur à celui du canal $h, H, A \to \tau^+\tau^-$, il est sujet à de nombreuses sources de bruit de fond au LHC où les collisions ont lieu entre protons. C'est pour cette raison que cette thèse s'intéresse



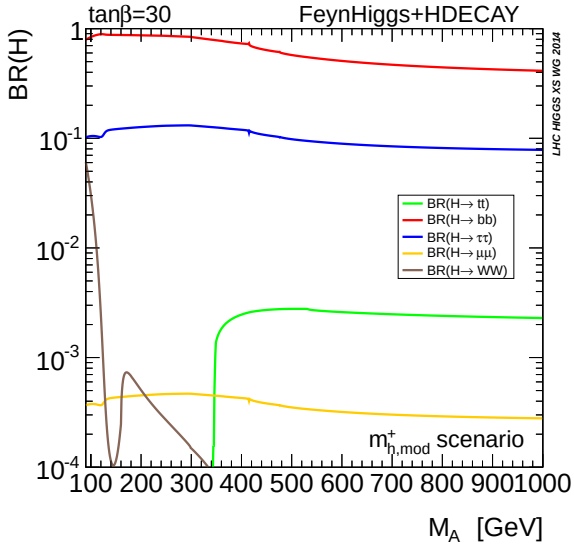
(a) *Rapports de branchement à* $\tan\beta = 5$ *pour* h.



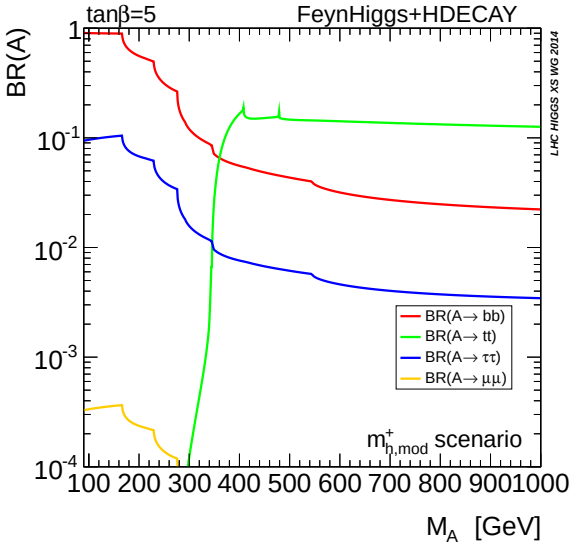
(b) *Rapports de branchement à* $\tan\beta = 30$ *pour* h.



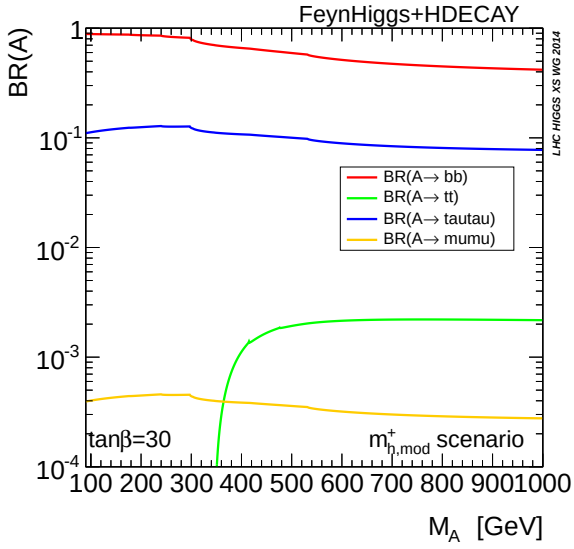
(c) *Rapports de branchement à* $\tan\beta = 5$ *pour* H.



(d) *Rapports de branchement à* $\tan\beta = 30$ *pour* H.



(e) *Rapports de branchement à* $\tan\beta = 5$ *pour* A.



(f) *Rapports de branchement à* $\tan\beta = 30$ *pour* A.

Figure 1.26 – *Rapports de branchement des bosons de Higgs du MSSM en fonction de* m_A *pour* $\tan\beta = 5$ *et* 30 *[84, 85]. L'effet de* $\tan\beta$ *sur les couplages de* H *et* A *aux fermions est bien visible avec l'augmentation des rapports de branchement au bottom (en rouge) et au tau (en bleu) et la diminution du rapport de branchement au top (en vert).*

à la recherche de bosons de Higgs supplémentaires de haute masse se désintégrant en paire de τ, dont l'accessibilité expérimentale est meilleure. La présence de deux leptons τ de haute énergie dans l'état final est en effet une signature bien plus claire que la présence de quarks b. Le diagramme de Feynman correspondant à cette désintégration est présenté sur la figure 1.27. Toutefois, les τ ne sont pas des particules stables et se désintègrent avant d'entrer dans les parties sensibles du détecteur. Seuls leurs produits de désintégration sont observés.

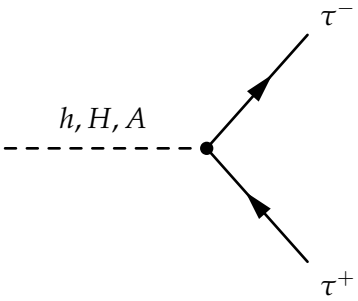


Figure 1.27 – *Diagramme de Feynman d'une désintégration $h, H, A \to \tau^+\tau^-$.*

5.3 Désintégrations des leptons τ

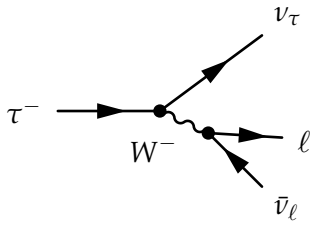
La durée de vie des leptons τ est de 290 fs, ce qui est trop court pour qu'ils puissent se propager jusqu'aux parties sensibles du détecteur. Un τ n'est donc pas directement observé, seuls ses produits de désintégration le peuvent.

Les leptons τ se désintègrent par interaction faible selon $\tau^- \to W^-\nu_\tau$ [9]. Le boson W, virtuel, se désintègre immédiatement :

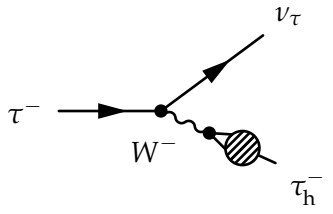
— leptoniquement selon $W^- \to e^-\bar{\nu}_e$ dans 17,82 % des cas ;
— leptoniquement selon $W^- \to \mu^-\bar{\nu}_\mu$ dans 17,39 % des cas ;
— hadroniquement selon $W^- \to q\bar{q}'$ dans 64,79 % des cas.

Dans ce dernier cas, l'hadronisation présentée dans la section 2.6 a lieu et les deux quarks donnent un ensemble constitué de quelques hadrons, en général trois ou moins, et éventuellement des particules neutres comme des π^0, ces derniers se désintégrant presque toujours en deux photons. L'ensemble des particules issues de la désintégration du W forme ainsi un petit jet. Il s'agit d'un « tau hadronique », noté τ_h dans la suite.

Les diagrammes de Feynman correspondant aux désintégrations leptonique et hadronique du τ sont représentés figures 1.28a et 1.28b. Le tableau 1.7a résume plus en détail les rapports de branchement des différents modes de désintégration du τ.



(a) *Désintégration leptonique d'un τ^-. Le lepton ℓ peut être un électron ou un muon.*



(b) *Désintégration hadronique d'un τ^-.*

Figure 1.28 – *Diagrammes de Feynman de désintégration d'un τ^-.*

La désintégration d'un τ peut donc se faire selon trois modes différents, $\tau \to \tau_\text{h}\,\nu_\tau$, $\tau \to \mu\,\nu_\mu\nu_\tau$ et $\tau \to e\,\nu_e\nu_\tau$. Dans tous les cas, les neutrinos sont invisibles pour le détecteur. Les deux leptons τ issus de la désintégration du boson de Higgs se désintègrent indépendamment l'un de l'autre. Il existe donc six canaux différents pour les événements dans lesquels un boson de Higgs se désintègre en paire de τ, listés avec leurs rapports de branchement dans le tableau 1.7b.

Mode de désintégration	$\mathcal{BR}$ (%)
$\tau^- \to e^- \bar{\nu}_e \bar{\nu}_\tau$	17,82
$\tau^- \to \mu^- \bar{\nu}_\mu \bar{\nu}_\tau$	17,39
$\tau^- \to h^- \bar{\nu}_\tau$	11,51
$\tau^- \to h^- \pi^0 \bar{\nu}_\tau$	25,93
$\tau^- \to h^- \pi^0 \pi^0 \bar{\nu}_\tau$	9,48
$\tau^- \to h^- h^- h^+ \bar{\nu}_\tau$	9,80
$\tau^- \to h^- h^- h^+ \pi^0 \bar{\nu}_\tau$	4,76
Autres modes hadroniques	3,31
$\tau^- \to \tau_h^- \bar{\nu}_\tau$	64,79

(a) *Rapports de branchement des différents modes de désintégration du τ [9].*

Canal	$\mathcal{BR}$ (%)
$\tau_h \tau_h$	41,98
$\mu \tau_h$	22,53
$e \tau_h$	23,09
$\mu\mu$	3,02
ee	3,18
$e\mu$	6,20

(b) *Rapports de branchement des six canaux des événements $h \to \tau\tau$.*

Tableau 1.7 – *Rapports de branchement des différents modes de désintégration du τ [9] et des différents canaux des événements $h \to \tau\tau$.*

6 Conclusion

Les constituants fondamentaux de la matière usuelle sont les fermions de la première génération, c'est-à-dire les quarks u et d pouvant former les protons (uud) et les neutrons (udd), l'électron et le neutrino électronique. Cependant, la liste des particules élémentaires est bien plus longue. Il existe ainsi trois générations de fermions, portant leur nombre à douze. Chacun de ces fermions est accompagné d'un antifermion correspondant.

Aux fermions s'ajoutent les bosons. Les bosons de jauge sont les vecteurs des forces fondamentales, il s'agit du photon, des bosons W^+, W^- et Z et des gluons. Le boson de Higgs, quant à lui, est scalaire.

Le modèle standard, reposant sur la théorie quantique des champs, permet de décrire le comportement de ces particules. L'invariance de jauge locale sous les transformations du groupe $SU(3)_C \times SU(2)_L \times U(1)_Y$ fait émerger naturellement les forces électrofaible et forte. Le mécanisme de brisure spontanée de symétrie introduit le champ de Higgs et donne une masse aux particules.

Ce modèle a permis de prédire l'existence de particules comme les bosons W et Z, le quark top ou encore le boson de Higgs de nombreuses années avant leurs observations. Malgré ces prouesses et les décennies de prédictions correctement vérifiées, la communauté scientifique sait que le modèle standard n'est qu'une étape vers une théorie plus complète. En effet, il ne permet pas d'expliquer certains faits expérimentaux comme la présence de la matière noire. D'autres phénomènes, comme l'ajustement fin, laissent à penser qu'il s'agit d'une théorie effective à basse énergie.

De nombreuses extensions au modèle standard sont alors proposées. Par exemple, la supersymétrie introduit les « sparticules », partenaires des particules. Dans sa version la plus simple, l'extension supersymétrique minimale du modèle standard ou MSSM, il n'y a pas un mais cinq bosons de Higgs, ainsi que leurs partenaires. La recherche d'un signal associé à ces bosons de Higgs supplémentaires est un test expérimental du MSSM.

Dans le cas où ces bosons de Higgs supplémentaires sont de haute masse, le canal de désintégration en paire de leptons τ est le plus prometteur. Les leptons τ se désintègrent eux-mêmes en électron, muon ou tau hadronique. Il existe donc six états finaux différents.

Grâce à la collaboration CMS (*Compact Muon Solenoid*) et au détecteur du même nom installé au LHC (*Large Hadron Collider*) de l'organisation européenne pour la recherche nucléaire ou CERN (Conseil Européen pour la Recherche Nucléaire), les conditions expérimentales sont réunies pour procéder à la recherche de bosons de Higgs supplémentaires de haute masse se désintégrant en paire de leptons τ. Le chapitre 2 présente ce dispositif expérimental.

Chapitre 2
Dispositif expérimental

Sommaire

1 Introduction 41
2 CERN : Conseil Européen pour la Recherche Nucléaire 42
2.1 Origines du CERN 42
2.2 Réalisations du CERN 42
2.3 Accélérateurs du CERN 43
3 LHC : *Large Hadron Collider* 43
3.1 Exploitation du LHC 43
3.2 Accélération de protons 45
3.3 Collisions de protons 45
3.4 Luminosité, nombre d'événements et statistiques 47
3.5 Empilement 49
3.6 Expériences du LHC 49
4 CMS : *Compact Muon Solenoïd* 50
4.1 Vue d'ensemble et système de coordonnées 50
4.2 Solénoïde 51
4.3 Trajectographe 53
4.4 Calorimètre électromagnétique ou ECAL 54
4.5 Calorimètre hadronique ou HCAL 57
4.6 Chambres à muons 58
4.7 Prise de données à CMS 60
5 Reconstruction des événements 61
5.1 Éléments d'identification du *Particle Flow* 62
5.2 Identification et reconstruction des particules 64
5.3 Reconstruction des jets 69
5.4 Énergie transverse manquante 71
5.5 Taus hadroniques 74
6 Simulation d'événements 76
6.1 Génération d'événements 76
6.2 Simulation du détecteur 76
6.3 Corrections apportées aux simulations 77
7 Conclusion 81

1 Introduction

L'étude des particules élémentaires présentées au chapitre 1 nécessite des conditions expérimentales particulières. En effet, hormis le photon et les fermions de la première génération, la plupart d'entre elles ont une durée de vie inférieure à la nanoseconde. Il faut donc dans un premier temps les créer. Dans certains cas, les conditions naturelles le permettent. Par exemple, les rayons cosmiques issus du Soleil produisent de nombreuses particules lors de leur interaction avec l'atmosphère. Leur

étude a ainsi permis la découverte des muons [89]. Toutefois, les particules les plus massives nécessitent des gammes d'énergies bien plus élevées afin d'être produites et certains processus du modèle standard ont une faible section efficace, c'est-à-dire une faible probabilité de survenir. La caractérisation des particules de l'Univers ne peut donc se faire uniquement par des observations de processus naturels.

Le Grand Collisionneur de Hadrons (LHC, *Large Hadron Collider*) [90-92] de l'organisation européenne pour la recherche nucléaire ou CERN (Conseil Européen pour la Recherche Nucléaire) [93] permet de réaliser des collisions entre particules. Cet environnement expérimental, contrôlé, permet d'atteindre des échelles d'énergies suffisantes pour produire des particules de haute masse et ce en très grand nombre. Ces conditions sont nécessaires afin d'observer et de caractériser certaines particules élémentaires comme le boson de Higgs [22-26].

Le CERN, présenté dans la section 2, possède plusieurs collisionneurs de particules dont le plus grand et le plus puissant à ce jour, le LHC, est introduit dans la section 3. La section 4 présente le détecteur CMS, installé au LHC, dont les données expérimentales sont utilisées dans cette thèse. Les particules ne sont pas observées en tant que telles dans le détecteur, seuls des signaux de leur passage sont récoltés. La reconstruction des particules physiques à partir de ces signaux est décrite dans la section 5. La compréhension et la validation des phénomènes fondamentaux liés au modèle théorique étudié nécessite quand à lui la comparaison des données récoltées expérimentalement avec des données simulées, ce qui est abordé dans la section 6.

2 CERN : Conseil Européen pour la Recherche Nucléaire

2.1 Origines du CERN

L'acronyme « CERN » signifie Conseil Européen pour la Recherche Nucléaire. Sa création est motivée par l'état de la recherche scientifique en Europe après la Seconde Guerre Mondiale [93]. Des scientifiques comme Raoul DAUTRY, Pierre AUGER ou Niels BOHR envisagent la création d'un laboratoire européen de physique atomique, considérée comme fondamentale à cette époque, n'ayant aucun but militaire. Leur objectif est de stopper la fuite des cerveaux vers l'Amérique, d'unifier l'Europe et de lui donner les moyens d'avoir une infrastructure de recherche de calibre mondial.

Le 9 décembre 1949, Louis DE BROGLIE propose officiellement la création d'un laboratoire européen. C'est en décembre 1951, lors d'une conférence de l'Unesco à Paris, qu'est adoptée une résolution pour la mise en place d'un Conseil Européen pour la Recherche Nucléaire ayant pour objectif de créer une convention pour un laboratoire européen sous 18 mois. Le site de Meyrin, au Nord-Ouest de Genève, est choisi en octobre 1952 pour sa position centrale vis-à-vis des pays européens et la neutralité militaire suisse. Les travaux commencent dès le printemps 1954.

Lors de la sixième session du Conseil Européen pour la Recherche Nucléaire, la convention établissant l'organisation européenne pour la recherche nucléaire est adoptée par les douze pays membres fondateurs : l'Allemagne (ex-RFA), la Belgique, le Danemark, la France, la Grèce, l'Italie, la Norvège, les Pays-Bas, le Royaume-Uni, la Suède, la Suisse et l'ex-Yougoslavie. La ratification est terminée le 29 septembre 1954. Le Conseil Européen pour la Recherche Nucléaire est alors dissous, mais l'acronyme CERN est resté attaché à l'organisation européenne pour la recherche nucléaire.

2.2 Réalisations du CERN

Le CERN a permis de réaliser de nombreuses découvertes en physique fondamentale, comme celle des courants neutres (1973) [94-96], des bosons W et Z (1983, Nobel 1984) [17-21] et dernièrement du boson de Higgs (2012, Nobel 2013) [22, 23].

En plus de ces découvertes scientifiques majeures, le CERN apporte également des innovations technologiques importantes comme les techniques de hadronthérapie pour le traitement des tumeurs cancéreuses, le Web [97] ou encore les écrans tactiles dont le développement dans les années 70 au CERN a permi de réduire le nombre de boutons dans la salle de contrôle du Supersynchrotron à Protons [98].

2.3 Accélérateurs du CERN

Le premier accélérateur de particules du CERN est le synchrocyclotron, mis en service en 1957 à une énergie de 600 MeV.

À la fin des années 50, le Synchrotron à Protons (PS) permet d'accélérer des protons et d'atteindre une énergie de 28 GeV, ce qui en fait l'accélérateur le plus puissant à l'époque. Avec l'arrivée de nouveaux anneaux au CERN, le PS sert également de pré-accélérateur.

En 1976, le Supersynchrotron à Protons (SPS) est mis en service. Son tunnel circulaire de 7 km de circonférence permet de faire collisionner deux faisceaux de particules circulant en sens inverse avec une énergie dans le centre de masse allant jusqu'à 450 GeV pour des protons. Le SPS a permis entre autres d'étudier la structure interne du proton et de découvrir les bosons W [17-19] et Z [20, 21].

Une nouvelle étape est franchie à la fin des années 80 avec la mise en service du Grand Collisionneur Électron-Positron (LEP, *Large Electron-Positron collider*), dans un tunnel sous-terrain de 27 km de circonférence. Aujourd'hui remplacé par le LHC, il reste le plus grand collisionneur de leptons au monde. Quatre grandes expériences étaient installées sur le LEP, ALEPH (*Apparatus for LEP PHysics at CERN*) [99], DELPHI (*DEtector with Lepton, Photon and Hadron Identification*) [100], L3 [101] et OPAL (*Omni-Purpose Apparatus at LEP*) [102], dont les emplacements sont visibles sur la figure 2.1 avec ceux des grandes expériences du LHC. Le LEP a permis de réaliser des mesures de précision sur les bosons W et Z précédemment découverts avec le SPS. Il a été mis en arrêt en 2000 afin de construire le LHC, sujet de la section suivante.

De nombreuses autres installations expérimentales sont présentes au CERN dont le complexe d'accélérateurs s'étend sur près de 10 km. Sur la figure 2.2 se trouve une carte de ce complexe avec les différentes structures encore en fonctionnement à ce jour.

3 LHC : *Large Hadron Collider*

Le Grand Collisionneur de Hadrons (LHC, *Large Hadron Collider*) [90-92] est le plus grand et le plus puissant accélérateur de particules au monde. Son tracé ainsi que ceux du *Booster*, du PS et du SPS sont illustrés sur la figure 2.1. Le LHC est installé dans le même tunnel que le LEP, il s'agit donc d'un accélérateur circulaire de 27 km de circonférence, situé entre 50 et 100 m de profondeur sous la frontière franco-suisse. Le LHC permet de réaliser des collisions proton-proton, proton-ion lourd et ion lourd-ion lourd. Dans tous les cas, deux faisceaux de particules sont accélérés en sens inverses. Seules les collisions de protons sont considérées dans cette thèse.

3.1 Exploitation du LHC

Le fonctionnement du LHC peut être divisé en plusieurs périodes ou *runs*. Chaque *run* présente différentes caractéristiques, en particulier l'énergie dans le centre de masse des collisions. Le tableau 2.1 résume les différents *runs* du LHC, passés et à venir. Chaque *run* est lui-même divisé par année civile, des arrêts techniques ayant lieu en période hivernale. Enfin, une année civile est subdivisée en plusieurs périodes entre lesquelles les conditions expérimentales comme la nature des particules entrant en collision peuvent varier.

Run	Période	Énergie dans le centre de masse	Luminosité proton-proton
I	2011-2012	7 à 8 TeV	$30\,\text{fb}^{-1}$
II	2016-2018	13 TeV	$160\,\text{fb}^{-1}$
III	2022-2024	13 à 14 TeV	$350\,\text{fb}^{-1}$?
IV	2027-2030	14 TeV	$3000\,\text{fb}^{-1}$?
V	2032-2034	14 TeV	

Tableau 2.1 – Runs *du LHC avec les énergies dans le centre de masse et les luminosités correspondantes des collisions de protons [104]. La luminosité est présentée dans la section 3.4.*

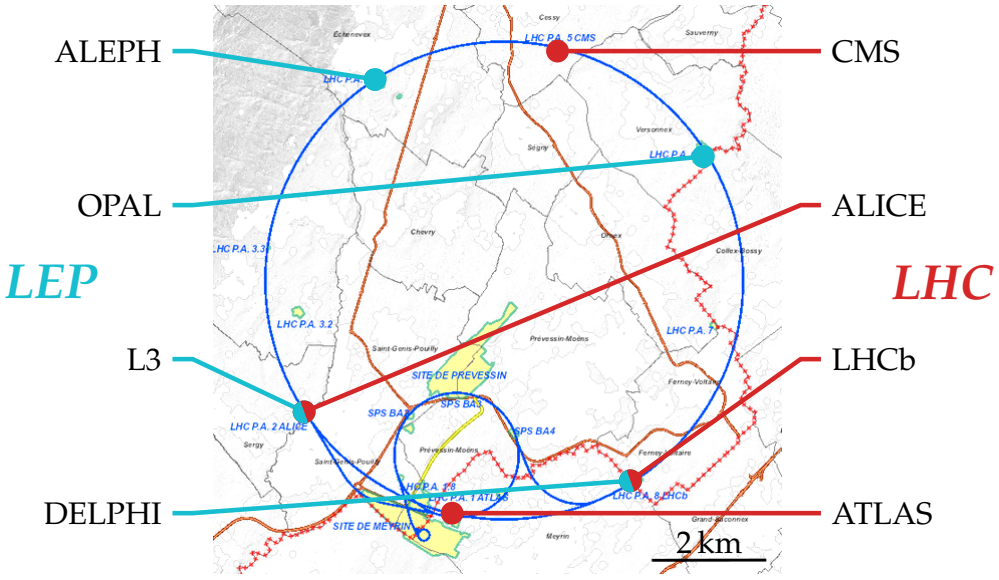


Figure 2.1 – *Emplacements des grandes expériences du LEP (1989-2000) et du LHC (depuis 2008) [103]. Les tracés des* booster, *PS et SPS sont également visibles.*

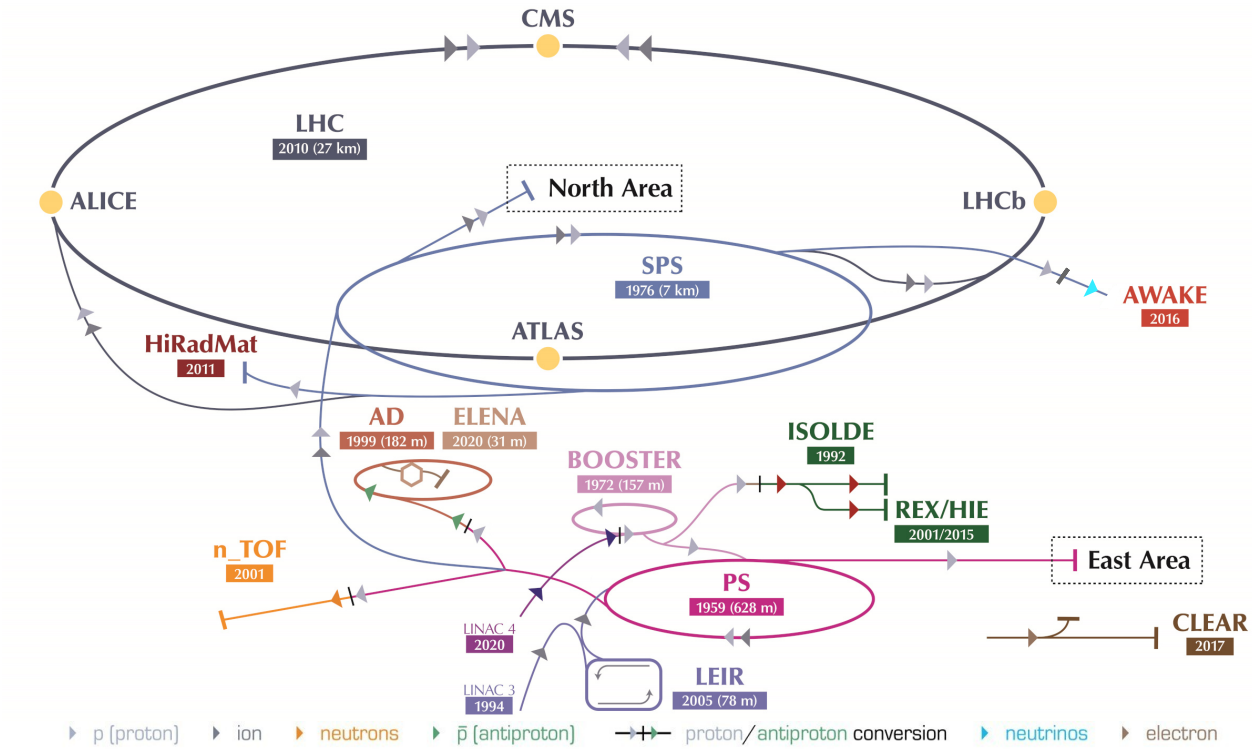


Figure 2.2 – *Complexe des accélérateurs du CERN [93]. De nombreuses expériences y sont installées : AD, Décélérateur d'Antiprotons ; AWAKE,* Advanced WAKefield Experiment ; *BOOSTER, Booster du Synchrotron à Protons ; CLEAR,* CERN Linear Electron Accelerator for Research ; *ELENA,* Extra Low Energy Antiproton ; *HiRadMat,* High-Radiation to Materials ; *ISOLDE,* Isotope mass Separator On-Line ; *LEIR, Anneau d'Ions de Basse Énergie ; LHC, Grand Collisionneur de Hadrons ; LINAC 3, Accélérateur Linéaire 3 ; LINAC 4, Accélérateur Linéaire 4, remplace le LINAC 2 ; n_TOF,* Neutrons Time Of Flight ; *PS, Synchrotron à Protons ; REX/HIE,* Radioactive EXperiment/High Intensity and Energy ; *SPS, Supersynchrotron à Protons ; ALICE,* A Large Ion Collider Experiment ; *ATLAS,* A Toroidal LHC ApparatuS ; *CMS,* Compact Muon Solenoid ; *LHCb,* Large Hadron Collider beauty.

3.2 Accélération de protons

Les protons sont obtenus par ionisation de dihydrogène, directement issu d'une bouteille. Ils sont alors progressivement accélérés à travers différentes installations du CERN, illustrées sur la figure 2.2, menant les protons à des niveaux d'énergie de plus en plus hauts avant de pouvoir être injectés dans le LHC [92] :

— l'accélérateur linéaire 2 (LINAC 2), remplacé pour le Run III du LHC par le LINAC 4, permet d'accélérer les protons à une énergie de 50 MeV ;
— le *Booster*, premier élément circulaire, amène les protons à 1,4 GeV ;
— le PS permet d'atteindre 25 GeV ;
— le SPS, dernier élément avant le LHC, accélère les protons jusqu'à 450 GeV.

Le LHC accélère alors les protons jusqu'à 6,5 TeV lors du Run II et ira jusqu'à 7 TeV lors du Run III, permettant de réaliser des collisions avec une énergie dans le centre de masse, $\sqrt{s}$, de 13 et 14 TeV, respectivement.

Les particules sont accélérées grâce à 16 cavités radiofréquences, 8 par faisceau. Ces cavités créent un champ électrique oscillant. Cette technique présente l'avantage d'accélérer les particules différemment selon leurs positions respectives, ce qui permet de les conserver en plusieurs paquets dont l'intérêt est développé dans la section suivante.

Afin de maintenir les particules dans le tube circulaire du LHC, leur trajectoire est courbée, c'est-à-dire qu'une accélération orthogonale au faisceau est appliquée. Cette courbure s'obtient grâce à un champ magnétique de 8,33 T généré par 1232 aimants dipolaires supraconducteurs répartis tout au long des 27 km du LHC et refroidis à l'hélium superfluide à 1,8 K.

3.3 Collisions de protons

3.3.1 Structure du proton

Les protons ne sont pas des particules élémentaires, ils possèdent une sous-structure. En tant que baryons, les protons sont des hadrons composés en moyenne de trois quarks dits « de valence », deux *up* et un *down*. Cependant, ces quarks échangent des gluons virtuels par l'intermédiaire de l'interaction forte. Ces gluons peuvent se matérialiser en paires $q\bar{q}$, comme illustré sur la figure 2.3. Au sein des protons, d'autres partons (quarks et gluons) que les trois quarks u, u et d de valence peuvent donc être présents. Ces particules additionnelles forment la « mer » dans laquelle se trouvent des gluons, des antiquarks et des quarks, y compris de saveurs lourdes.

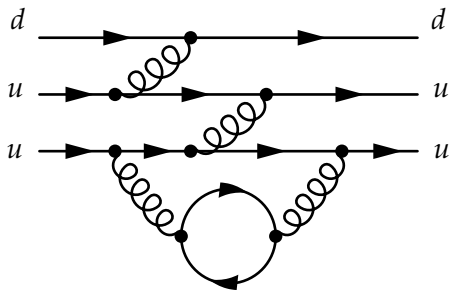
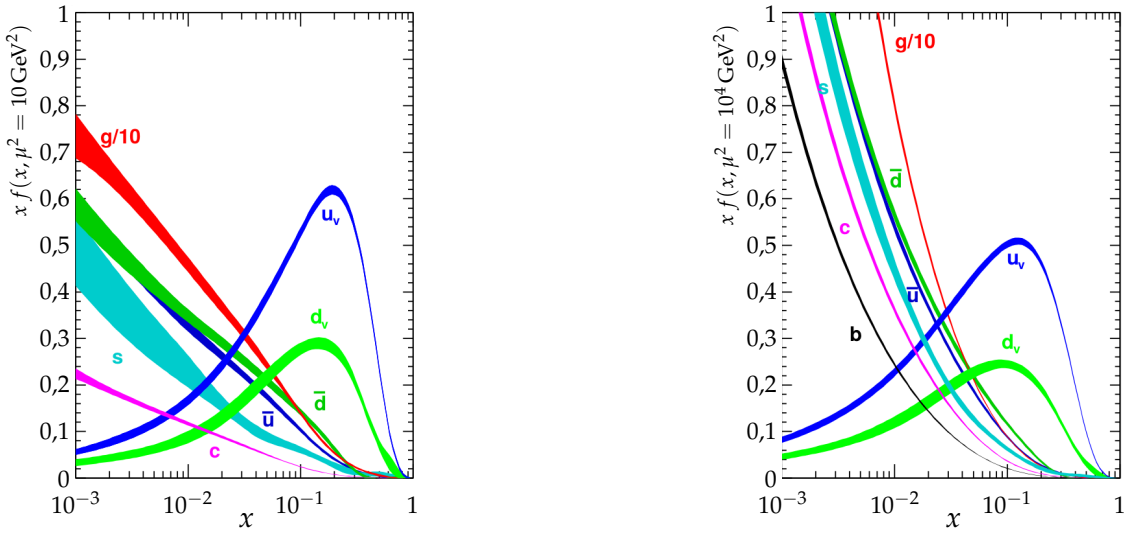


Figure 2.3 – *Diagramme de Feynman de la propagation d'un proton avec apparition d'une paire $q\bar{q}$.*

Chacun des constituants du proton porte une fraction x de son énergie totale. La probabilité pour qu'un parton i du proton porte une fraction x_i de l'énergie peut être décrite à l'aide des fonctions de distribution partoniques (PDF, *Parton Distribution Functions*). Ces fonctions sont obtenues par ajustement des données expérimentales [105, 106] et sont représentées sur la figure 2.4. Les partons de la mer portent individuellement une faible fraction de l'énergie. Un quark de valence porte statistiquement 10 à 20 % de l'énergie du proton. L'ensemble des quarks de valence porte ainsi 30 à 60 % seulement de l'énergie totale du proton. Dans le cas du proton, les quarks de valence u portent environ deux fois plus d'énergie que les quarks de valence d, ce qui est cohérent avec la structure moyenne (uud) du proton.

Lorsque deux protons entrent en collision, ce sont en fait leurs constituants respectifs qui interagissent. Or, le constituant du premier proton n'a aucune raison de porter une fraction d'énergie x identique à celle que porte celui du second. Ainsi, bien que les impulsions des deux protons soient opposées, c'est-à-dire que le référentiel du centre de masse coïncide avec le référentiel du détecteur,



(a) *Pour* $\mu^2 = 10\,\mathrm{GeV}^2$.

(b) *Pour* $\mu^2 = 10^4\,\mathrm{GeV}^2$.

Figure 2.4 – *Fonctions de distribution partoniques (PDF,* Parton Distribution Functions*) à différentes échelles d'énergie* μ *obtenues au NNLO NNPDF3.0 [107] avec* $\alpha_s(m_Z^2) = 0{,}118$ *[9]. Les bandes tracées correspondent aux PDF* f*, avec incertitude, multipliées par la fraction d'impulsion* x*, où* f *peut être les quarks u et d de valence (*u_v*,* d_v*) ou les gluons (*g*), quarks et antiquarks de la mer (*$\bar{u}$*,* $\bar{d}$*,* $s \simeq \bar{s}$*,* $c = \bar{c}$*,* $b = \bar{b}$*).*

cela n'est pas vrai pour les partons impliqués dans la collision. Seule l'impulsion totale dans le plan transverse aux faisceaux est donc nulle et l'impulsion selon l'axe des faisceaux est inconnue. De plus, l'énergie dans le centre de masse des protons n'est pas totalement utilisée. Une collision de partons à 13 TeV est ainsi peu probable au LHC mais il est ainsi possible de balayer une large gamme d'énergies effectives de collision, ce qui est favorable aux recherches de nouvelles particules. Enfin, les constituants des protons n'interagissent pas tous forcément lors de la collision. Des fragments résiduels continuent ainsi leur trajet le long du tube de faisceau et sont inaccessibles pour des détecteurs comme CMS.

3.3.2 Faisceaux et paquets de protons

La dimension spatiale des protons, de l'ordre du femtomètre (10^{-15} m), ne permet pas de les faire entrer en collision un à un de manière efficace. Au LHC, deux faisceaux de protons sont ainsi accélérés, chacun dans un sens. Les protons y sont regroupés par paquets espacés temporellement de 25 ns lors du Run II.

La formation et le maintien de ces paquets sont rendus possibles par l'utilisation des cavités radiofréquence. Elles produisent un champ électrique dont la norme est plus importante au niveau des queues des paquets qu'à leurs têtes. Alors, les protons « en queue de peloton » sont plus accélérés que les protons en tête et les rattrapent.

Chacun des faisceaux du LHC est formé de 2808 paquets. Avant les premières collisions, un paquet comporte $1{,}15 \times 10^{11}$ protons. Les paquets font environ 30 cm de long. Lorsqu'ils circulent dans le LHC, leur diamètre est de l'ordre du millimètre mais au niveau des points de collisions, un ensemble de champs magnétiques réduit ce diamètre à 16 µm.

Le passage des $1{,}15 \times 10^{11}$ protons d'un paquet d'un faisceau du LHC à travers la surface de 16 µm de diamètre, combiné au passage d'un paquet de l'autre faisceau, permet d'obtenir des collisions entre protons. Au cours de celles-ci, les constituants des protons interagissent et créent de nouvelles particules par conservation de l'énergie. Les faisceaux du LHC sont stables une dizaine d'heures, pendant lesquelles des collisions surviennent 40 millions de fois par seconde. La luminosité, introduite dans la section 3.4 ci-après, permet de rendre compte de la quantité de collisions réalisées.

3.4 Luminosité, nombre d'événements et statistiques

La quantité d'événements $\mathrm{d}N_i$ issus d'un processus physique i de section efficace σ_i pendant une durée $\mathrm{d}t$ s'exprime

$$\mathrm{d}N_i = \mathcal{L}_{\text{inst}}\sigma_i \,\mathrm{d}t \tag{2.1}$$

où $\mathcal{L}_{\text{inst}}$ est la luminosité instantanée du dispositif expérimental, exprimée par unité de surface et de temps. La luminosité instantanée au LHC peut s'exprimer en fonction des propriétés des faisceaux selon

$$\mathcal{L}_{\text{inst}} = \frac{\gamma \nu n_p N_p^2}{4\pi\epsilon_n\beta^*} = \frac{\nu n_p N_p^2}{4\pi\ell_x\ell_y} \tag{2.2}$$

où γ est le boost de Lorentz des paquets de protons, ν la fréquence de révolution des paquets dans l'anneau du LHC, n_p le nombre de paquets, N_p le nombre de protons par paquet, ϵ_n l'émittance transverse, qui permet de mesurer le parallélisme des faisceaux, β^* la fonction d'amplitude mesurant la distance entre le point de croisement des faisceaux et le lieu où un faisceau est deux fois plus large et ℓ_x et ℓ_y les dimensions transverses du faisceau au point d'interaction. La luminosité instantanée est donc favorisée par une faible largeur du faisceau au niveau des points d'interaction.

La luminosité intégrée ou « totale », $\mathcal{L}$, s'obtient par intégration temporelle de $\mathcal{L}_{\text{inst}}$. Elle s'exprime donc par unité de surface vue l'équation (2.1). Le nombre d'événements N_i produit durant une période donnée par un processus physique i de section efficace σ_i est donc

$$N_i = \mathcal{L}\sigma_i \,. \tag{2.3}$$

Certaines expériences de physique des particules se ramènent donc à des expériences de comptage dans lesquelles les données sont traitées statistiquement. Dans le cas des recherches de nouvelles particules, l'écart entre les observations effectives, d'une part, et attendues dans l'hypothèse où aucun phénomène nouveau n'est présent, d'autre part, est quantifié en termes de nombre de déviations standard. En l'absence de nouvelle physique, un écart serait dû à des fluctuations statistiques. Il est alors peu probable d'obtenir une grande déviation. Il est ainsi considéré qu'un phénomène, par exemple une nouvelle particule, est observé à partir de cinq déviations standard (5σ) ou plus, ce qui correspond à une probabilité de réaliser cette observation uniquement à cause de fluctuations d'une sur deux millions.

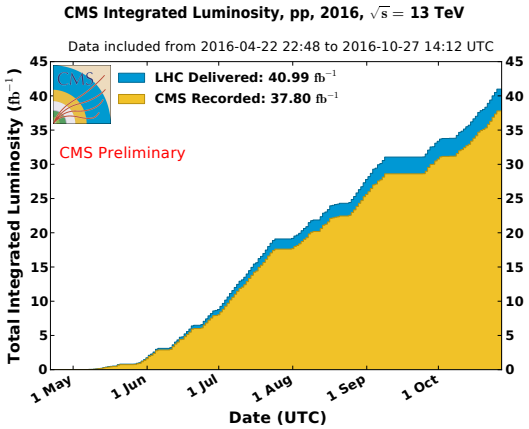
Les incertitudes jouent un rôle majeur dans la quantification du nombre de déviations standard observées. Il est important de bien les déterminer et de les minimiser. L'incertitude systématique peut être limitée en concevant des expériences les plus précises possible. L'incertitude statistique relative peut être diminuée en augmentant le nombre d'événements correspondant au phénomène étudié. Les conditions favorables aux découvertes sont donc celles dans lesquelles la luminosité ou la section efficace du processus étudié sont importantes. En l'absence d'un écart significatif entre prédictions et observations, il est possible de proposer des limites sur les valeurs des sections efficaces des processus en jeu, limites au-delà desquelles l'écart aurait été observé.

Les figures 2.5a, 2.6a et 2.7a présentent les luminosités totales délivrées par le LHC et enregistrées par le détecteur CMS en fonction du temps lors du Run II. Les luminosités y sont exprimées en inverse femtobarn. Le barn ($1\,\mathrm{b} = 10^{-28}\,\mathrm{m}^2$) est une unité qui permet d'obtenir des valeurs numériques pour la luminosité plus abordables qu'avec les unités usuelles du système international. En effet pour le Run II du LHC, la luminosité totale des collisions de protons est de

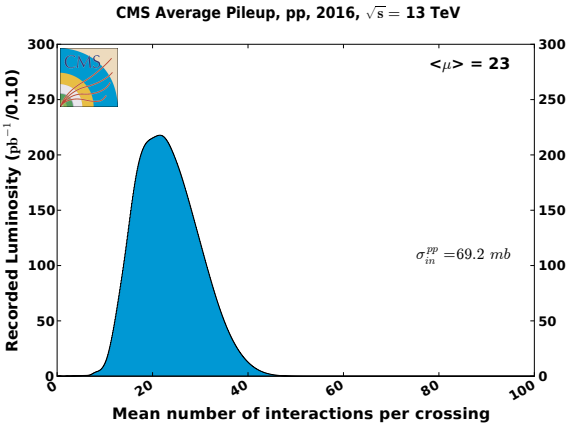
$$160\,\mathrm{fb}^{-1} = 160\times10^{15}\,\mathrm{b}^{-1} = 160\times10^{43}\,\mathrm{m}^{-2}\,, \tag{2.4}$$

son expression en m^{-2} n'est donc pas pratique à cause du facteur 10^{43}.

Pendant le Run II, les bonnes performances du détecteur CMS lui ont permis d'enregistrer 92,39 % de la luminosité délivrée par le LHC (92,22 % en 2016, 90,34 % en 2017 et 93,83 % en 2018) [108-110]. Les projets de développements futurs du LHC s'orientent dans un premier temps vers une augmentation de la luminosité, il s'agit du « HL-LHC » (LHC Haute Luminosité). Les performances actuelles du détecteur devront être encore améliorées d'ici-là et la collaboration y travaille d'ores et déjà.

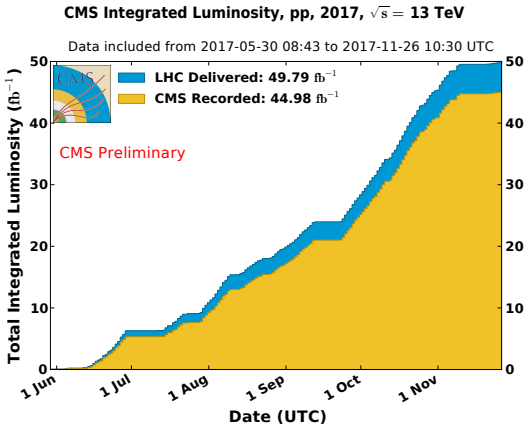


(a) *Luminosité totale délivrée par le LHC et enregistrée par le détecteur CMS en fonction du temps en 2016.*

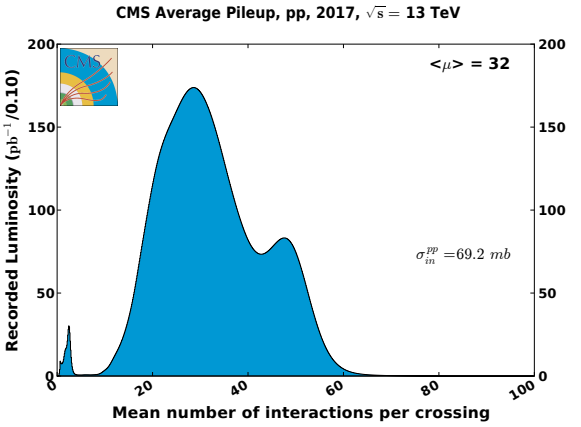


(b) *Distribution du nombre moyen d'interactions par croisement de faisceaux en 2016.*

Figure 2.5 – *Luminosité totale et empilement en 2016 [108, 111].*

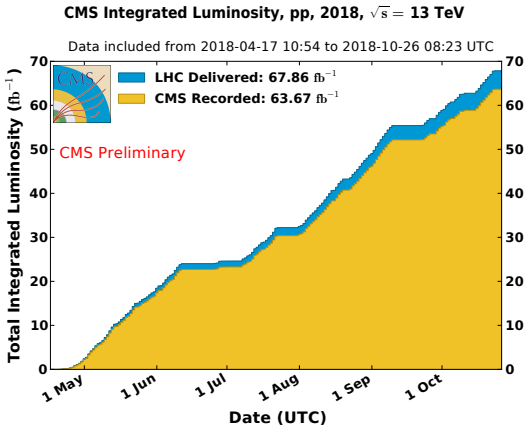


(a) *Luminosité totale délivrée par le LHC et enregistrée par le détecteur CMS en fonction du temps en 2017.*

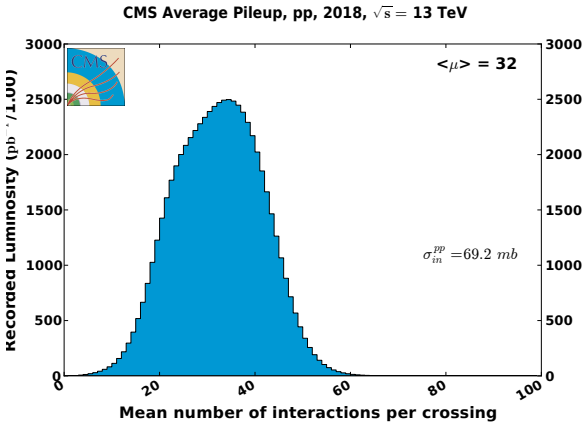


(b) *Distribution du nombre moyen d'interactions par croisement de faisceaux en 2017.*

Figure 2.6 – *Luminosité totale et empilement en 2017 [109, 111].*



(a) *Luminosité totale délivrée par le LHC et enregistrée par le détecteur CMS en fonction du temps en 2018.*



(b) *Distribution du nombre moyen d'interactions par croisement de faisceaux en 2018.*

Figure 2.7 – *Luminosité totale et empilement en 2018 [110, 111].*

3.5 Empilement

L'empilement (PU, *Pile-Up*) est le phénomène de superposition des signaux issus de différents processus physiques. Il existe deux types d'empilement, synchrone et asynchrone.

3.5.1 Empilement synchrone

Plusieurs collisions de protons peuvent avoir lieu lors d'un même croisement de paquet de protons ou « événement ». Le nombre d'interactions d'empilement synchrone moyen dépend des conditions du LHC. Les figures 2.5b, 2.6b et 2.7b présentent ces distributions pour les trois années du Run II.

À chaque collision correspond un vertex, lieu effectif de la rencontre entre les deux protons. Pour chacun des vertex de l'événement, la somme des impulsions transverses (p_{T}) au carré des traces de particules chargées détectées en provenance de ce vertex est calculée. Le vertex principal est choisi comme étant le vertex présentant la plus grande valeur de cette somme. Les autres sont considérés comme des vertex d'empilement et sont en général peu intéressants dans les analyses expérimentales.

Toutefois, les particules provenant des vertex d'empilement laissent elles aussi des signaux dans le détecteur, en plus des signaux d'intérêts. Par conséquent, elles viennent perturber la mesure des phénomènes se produisant au vertex principal.

3.5.2 Empilement asynchrone

L'empilement asynchrone est dû au temps de réponse des composants du détecteur. Des croisements de paquets ont en effet lieu toutes les 25 ns au LHC. Or, le détecteur peut présenter ponctuellement un temps de réponse supérieur, selon les signaux qu'il doit traiter. Les signaux de sortie du détecteur pour un événement donné peuvent donc être superposés avec ceux de l'événement suivant, ou précédant selon le point de vue.

3.6 Expériences du LHC

Il existe sept expériences au LHC. Parmi elles, quatre sont de « grandes expériences » et se situent chacune à un des points d'interaction de l'anneau afin d'étudier les collisions qui y sont produites.

ALICE [112], *A Large Ion Collider Experiment*, est une expérience conçue pour étudier le déconfinement des quarks et des gluons à l'aide de collisions d'ions lourds. Elle est installée au point 2, indiqué en figure 2.1.

ATLAS [113], *A Toroidal LHC ApparatuS*, est une expérience généraliste avec un éventail d'études très large, allant des mesures de précision des paramètres du modèle standard à la recherche de nouvelle physique. Ce détecteur se trouve au point 1 du LHC.

CMS [114], *Compact Muon Solenoid*, est également une expérience généraliste dont les objectifs sont similaires à ceux d'ATLAS. Les détecteurs d'ATLAS et de CMS étant conçus différemment, ces deux expériences peuvent valider leurs résultats de manière indépendante. Le détecteur CMS est installé au point 5 du LHC, à l'exact opposé d'ATLAS.

LHCb [115], *Large Hadron Collider beauty*, se concentre sur l'étude de la violation de la symétrie *CP* dans le secteur du quark *b*. Cette expérience réalise également des mesures de précision de certains paramètres du modèle standard. L'expérience LHCb se situe au point 8.

Les trois autres expériences du LHC sont LHCf, TOTEM et MoEDAL. L'expérience LHCf (*Large Hadron Collider forward*) [116], installée à 140 m de part et d'autre du détecteur ATLAS, observe les particules issues des collisions et presque alignées avec le faisceau du LHC afin de simuler des rayons cosmiques. La plus « longue » des expériences du CERN, TOTEM (*Total, elastic and diffractive cross-section measurement*) [117], est quant à elle installée sur un demi kilomètre autour de CMS et étudie les protons grâce aux particules alignées avec le faisceau. Enfin, MoEDAL (*Monopole and Exotics Detector At the LHC*) [118] cherche à détecter l'existence de monopoles magnétiques et de particules ionisantes massives prédites par certains modèles au-delà du modèle standard grâce à des détecteurs installés près de LHCb.

4 CMS : *Compact Muon Solenoïd*

4.1 Vue d'ensemble et système de coordonnées

Le détecteur CMS est installé dans la caverne du point d'interaction numéro 5 du LHC, visible sur la figure 2.1 au Nord de l'installation, dans la commune de Cessy, en France. Il a été pensé avec pour but premier l'étude de la brisure de symétrie électrofaible et la recherche du boson de Higgs [119]. Pour cela, la conception du détecteur repose sur :

— un système de détection des muons de haute performance ;
— le meilleur calorimètre électromagnétique possible compatible avec le point précédent ;
— un système de trajectographie central entièrement basé sur des détecteurs au silicium ;
— un calorimètre hadronique avec une résolution suffisante et une bonne herméticité.

Son design généraliste permet de nombreuses autres analyses de physique, comme des mesures de précision, la recherche d'une nouvelle physique ou encore les collisions d'ions lourds.

La figure 2.8 présente une vue ouverte du détecteur CMS. Il possède une forme cylindrique de 28,7 m de long et 15 m de diamètre pour un poids total de 14 000 tonnes. Il est structuré en couches concentriques, chacune ayant un rôle spécifique détaillé dans les sections qui suivent. À partir du centre du détecteur, lieu des collisions, se trouvent dans l'ordre le trajectographe [121], le calorimètre électronique [122], le calorimètre hadronique [123], le solénoïde [124] donnant son « S » à CMS et les chambres à muons [125] donnant son « M » à CMS, encastrées dans la culasse d'acier. Des calorimètres « vers l'avant » se trouvent aux extrémité du détecteur, le long de l'axe du faisceau. Le détecteur propose ainsi une couverture d'un angle solide de presque 4π sr, c'est-à-dire de presque toutes les directions, ce qui est capital afin de reconstruire les collisions.

Le détecteur peut de plus être divisé en trois grandes parties de par sa forme cylindrique. La première, centrale, est le « tonneau », dans laquelle les sous-parties ont une géométrie cylindrique. Les parties sensibles du détecteur y sont orientées vers l'axe du faisceau. Aux deux extrémités du

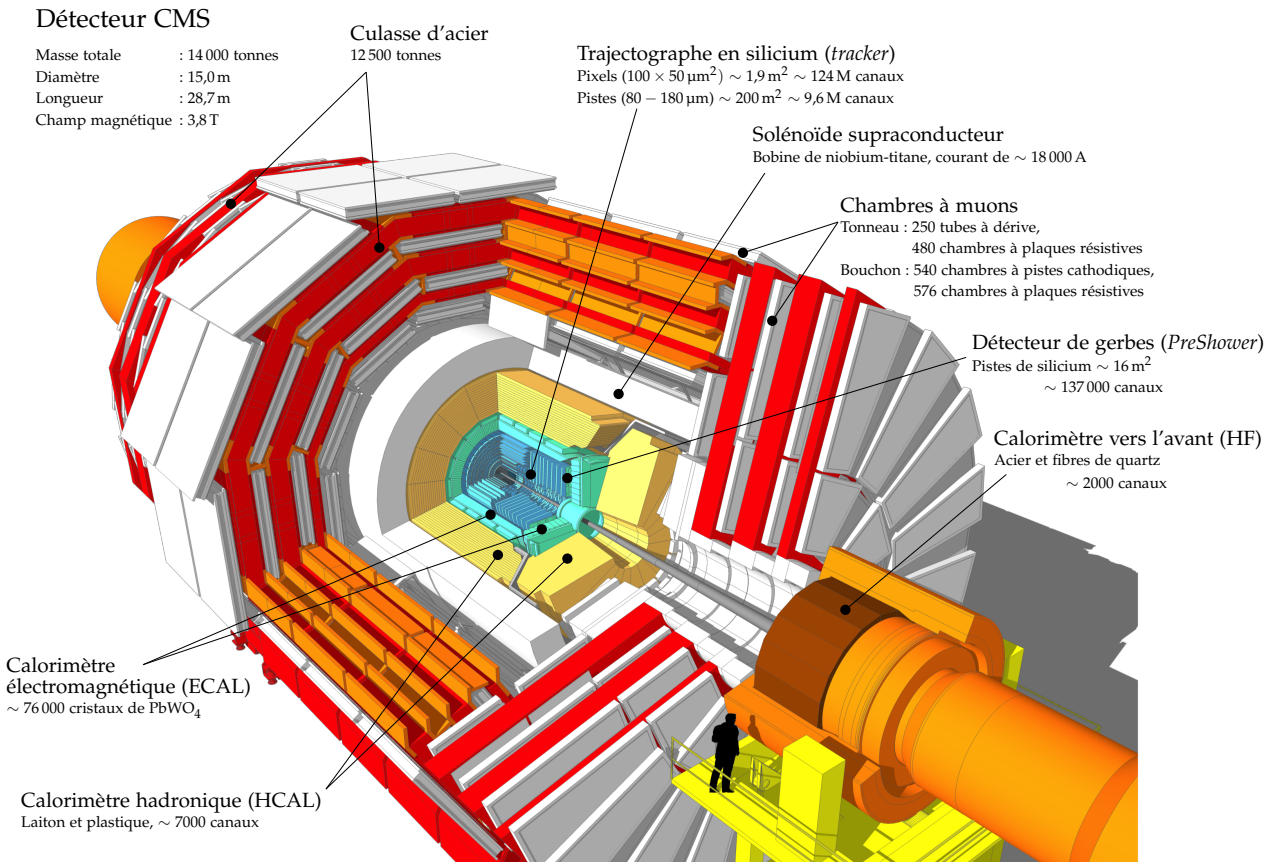


Figure 2.8 – *Vue ouverte du détecteur CMS [120].*

détecteur se trouvent les « bouchons », dont l'orientation des parties sensibles du détecteur se fait dans le plan transverse au faisceau. Ces différentes orientations sont bien visibles sur la figure 2.8.

L'acronyme CMS signifie *Compact Muon Solenoïd*, c'est-à-dire Solénoïde Compact à Muons. La structure du détecteur, conçue à partir de celle du solénoïde, mène en effet à un design compact pour le système à muons (chambres à muons et culasse d'acier), d'où le qualificatif [119].

La géométrie cylindrique du détecteur pousse à définir un système de coordonnées également cylindriques en complément d'un repère cartésien. Le schéma de la figure 2.9 illustre la définition de ces systèmes de coordonnées. L'origine O des repères est le centre du détecteur où les protons entrent en collision. Le vecteur de base $\vec{e}_x$ pointe vers le centre du LHC, $\vec{e}_y$ vers le haut ($\vec{g} \cdot \vec{e}_y < 0$) et $\vec{e}_z$ est colinéaire au tube de faisceau. Le trièdre $(\vec{e}_x, \vec{e}_y, \vec{e}_z)$ est direct. Le plan (Oxy) est nommé « plan transverse », il est orthogonal aux faisceaux. Le système de coordonnées cylindriques est défini par la distance à l'origine et deux angles $\theta \in [0, \pi]$ et $\phi \in [-\pi, \pi]$. L'angle entre le vecteur $\vec{a}$ à caractériser et $\vec{e}_z$ est θ. L'angle entre $\vec{a}$ et $\vec{e}_x$ dans le plan transverse est ϕ.

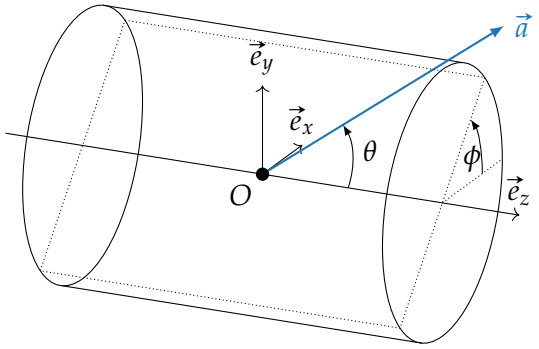


Figure 2.9 – *Système de coordonnées.*

L'angle θ n'est généralement pas utilisé directement et est remplacé par la « pseudo-rapidité » η définie comme

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{2.5}$$

La pseudo-rapidité est ainsi égale à zéro dans le plan transverse. L'usage de cette variable est motivé par la densité de production de particules qui est constante suivant η et non selon θ. De plus, dans la limite « ultra-relativiste » c'est-à-dire $|\vec{p}| \gg m$, condition remplie au LHC, la pseudo-rapidité tend vers la rapidité y (à ne pas confondre avec la coordonnée y) des particules,

$$y = \frac{1}{2}\ln\left(\frac{E + p_z c}{E - p_z c}\right). \tag{2.6}$$

Or la rapidité est un invariant de Lorentz, ainsi au LHC η est en très bonne approximation un invariant de Lorentz, contrairement à θ. La figure 2.10 présente un quadrant du détecteur CMS sur lequel figurent quelques valeurs de pseudo-rapidité et les directions correspondantes dans le plan (Oyz).

Du fait de la structure des protons discutée dans la section 3.3, lors de la collision, l'impulsion totale selon l'axe des faisceaux est inconnue. Seule l'impulsion totale dans le plan transverse, c'est-à-dire le plan (Oxy), est nulle. C'est pourquoi des variables relatives au plan transverse sont définies, en particulier l'impulsion transverse $\vec{p}_\mathrm{T}$, sa norme p_T et l'« énergie transverse » E_T,

$$\vec{p}_\mathrm{T} = p_x\vec{e}_x + p_y\vec{e}_y\,, \quad p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}\,, \quad E_\mathrm{T} = E\sin\theta = \frac{E}{\cosh\eta}\,. \tag{2.7}$$

4.2 Solénoïde

Le solénoïde supraconducteur est une des parties les plus importantes du détecteur. Il aide à caractériser les particules électriquement chargées issues des collisions. En effet, un champ magnétique courbe les trajectoires des particules électriquement chargées. Pour une particule d'impulsion $\vec{p}$, de charge q et de vitesse $\vec{v}$ soumise à un champ magnétique $\vec{B}$,

$$\frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = q\vec{v} \wedge \vec{B}\,. \tag{2.8}$$

Il s'agit de la composante magnétique de la force de Lorentz. Avec un champ magnétique dirigé « vers le haut », une particule chargée positivement est ainsi déviée vers la droite et une particule chargée négativement vers la gauche.

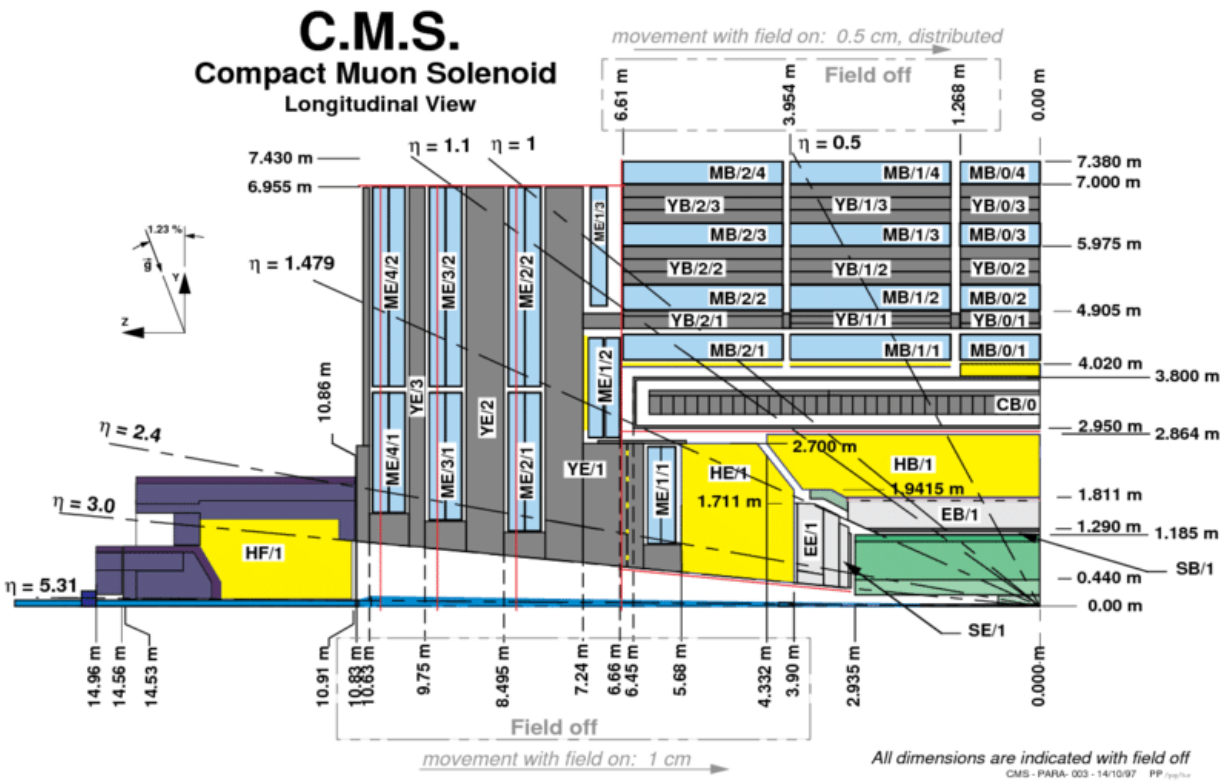


Figure 2.10 – *Vue longitudinale d'un quadrant du détecteur CMS [126]. Les directions correspondant à quelques valeurs de pseudo-rapidité sont illustrées et des mesures de distances par rapport au centre du détecteur, lieu des collisions, sont indiquées. Le sol de la caverne présente une inclinaison de* 1,23 % *par rapport à la direction de la gravité locale* $\vec{g}$*, ce que montre le schéma à gauche.*

Le rayon de courbure r de la trajectoire de la particule est

$$r = \frac{p_{\mathrm{T}}}{|q|B} \tag{2.9}$$

avec p_{T} l'impulsion dans le plan transverse au champ magnétique. Dans le cas du détecteur CMS, le champ magnétique est aligné avec l'axe du faisceau, il s'agit donc également de l'impulsion transverse définie dans la section 4.1. Les particules chargées se propagent ainsi selon une trajectoire hélicoïdale autour de l'axe du faisceau et les particules neutres en ligne droite.

Champ	3,8 T
Diamètre interne	5,9 m
Longueur	12,9 m
Nombre de tours	2168
Courant	19,5 kA
Énergie stockée	2,7 GJ

Tableau 2.2 – *Caractéristiques du solénoïde supraconducteur de CMS [127].*

Afin d'assurer de bonnes performances sur l'identification des particules, en particulier sur la détermination du signe de la charge des muons (permettant de savoir s'il s'agit d'un muon ou d'un antimuon) et de leur impulsion jusqu'à l'ordre du TeV, la collaboration CMS a choisi d'utiliser un solénoïde supraconducteur [114, 124, 127, 128] dont les caractéristiques sont résumées dans le tableau 2.2. Le rapport rayon sur longueur du cylindre défini par le solénoïde permet ainsi une identification efficace des muons de pseudo-rapidité inférieure à 2,4 [124].

À l'extérieur du solénoïde, une culasse d'acier, visible en rouge sur la figure 2.8, permet de contenir le retour du champ magnétique. La culasse d'acier est composée de plusieurs couches séparées par les chambres à muons. Elle présente ainsi une épaisseur cumulée d'environ 1,5 m. Le champ magnétique, au lieu d'être presque nul hors du solénoïde, atteint ainsi 1 à 2 T dans la culasse d'acier, selon l'endroit considéré. Les trajectoires des muons sont alors courbées dans un sens dans le volume interne du solénoïde, puis dans l'autre sens hors du solénoïde. La carte du champ magnétique obtenu est représentée sur la figure 2.11.

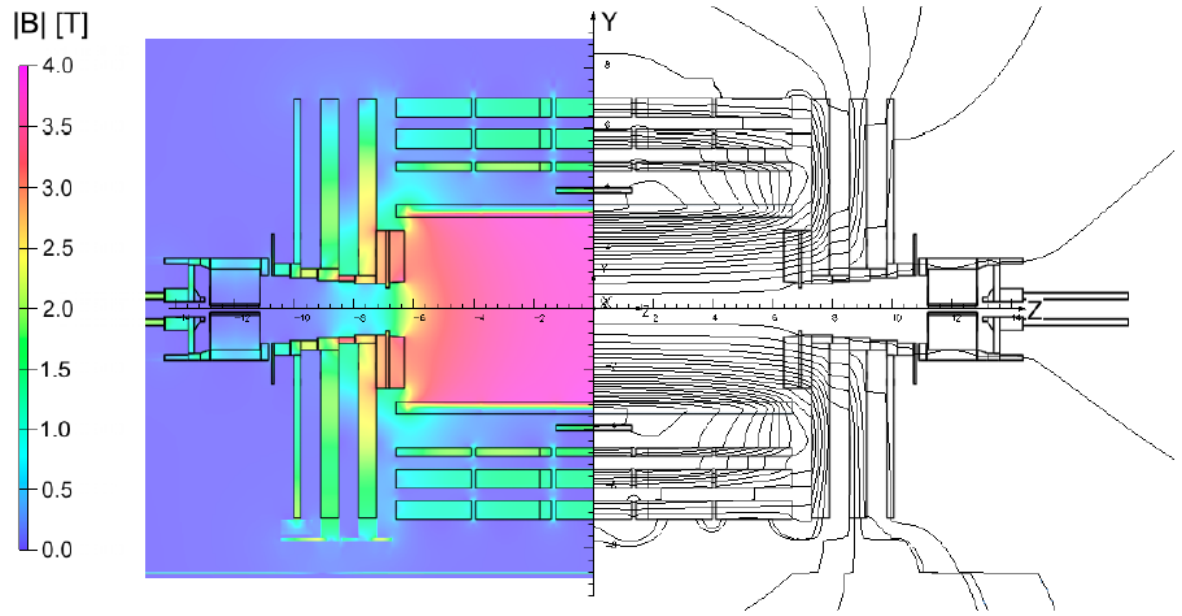


Figure 2.11 – *Valeurs de la norme du champ magnétique (à gauche) et lignes de champ (à droite) prédites dans la section longitudinale du déteteur CMS avec une valeur du champ au centre de* 3,8 T *[129]. Chaque ligne de champ correspond à un flux magnétique de* 6 Wb.

4.3 Trajectographe

Le trajectographe couvre la partie centrale du détecteur, son acceptation correspondant à la région $|\eta| < 2{,}5$. Les particules chargées laissent des signaux de leur passage dans les différents modules du trajectographe en les traversant. Il est ainsi essentiel à la reconstruction des vertex des événements du LHC. Les particules chargées se trouvant hors de son acceptation mais visibles dans d'autres sous-détecteurs sont reconstruites comme des particules neutres.

Deux types de modules composent le trajectographe de CMS [114, 121, 130-132]. Dans sa partie interne, c'est-à-dire la plus centrale en rayon, des modules à pixels de silicium sont utilisés. Au début du Run II du LHC, le tonneau du trajectographe interne était composé de trois couches de pixels à 4,4, 7,3 et 10,2 cm de rayon [114] et les bouchons de deux disques de pixels. Le trajectographe interne est visible en rouge sur la figure 2.12. Les pixels ont une surface de $100 \times 150\,\mu\text{m}^2$. La résolution spatiale de cette partie du trajectographe est de l'ordre de 10 µm [114] et permet d'obtenir une bonne reconstruction des vertex. En mars 2017, cette partie du trajectographe a été remplacée [130, 133]. Elle comporte à présent quatre couches de pixels dans la partie tonneau à 2,9, 6,8, 10,9 et 16,0 cm de rayon. Dans les bouchons, les disques ont également été repensés afin d'obtenir quatre points de passage pour les traces telles que $|\eta| < 2{,}5$. Une comparaison des trajectographes à pixels utilisés en 2016 (Phase-0) et à partir de 2017 (Phase-1) est illustrée sur la figure 2.13. Cette modification du détecteur permet d'améliorer la reconstruction des vertex ainsi que l'efficacité de l'identification de jets issus de quarks *b*.

De 20 à 116 cm de rayon se trouve le trajectographe à piste, lui-même composé de trois sous-parties. La première de ces sous-parties comporte un tonneau (TIB, *Tracker Inner Barrel*) de quatre couches et trois disques (TID, *Tracker Inner Disks*) vers l'avant. Les pistes dans ces couches sont parallèles au faisceau dans le TIB et axiales dans les TID. Elles permettent d'obtenir une résolution de 23 µm et 35 µm respectivement [114]. Le TIB et les TID sont entourés par le tonneau du trajectographe externe (TOB, *Tracker Outer Barrel*) de six couches. Le TOB a une résolution de 53 µm pour ses quatre premières couches et de 35 µm ensuite [114]. Enfin, les bouchons du trajectographe externe (TEC, *Tracker EndCaps*) se situent aux extrémités du dispositif.

Les trajectoires des particules chargées sont alors reconstruites par un ajustement aux différents points de passage dans le trajectographe. À partir de ces trajectoires, il est possible de déterminer le signe de la charge électrique et l'impulsion transverse des particules à l'aide de l'équation (2.9). Les

résolutions obtenues sur les impulsions transverses des muons et des particules chargées à l'aide du trajectographe sont présentées sur la figure 2.14.

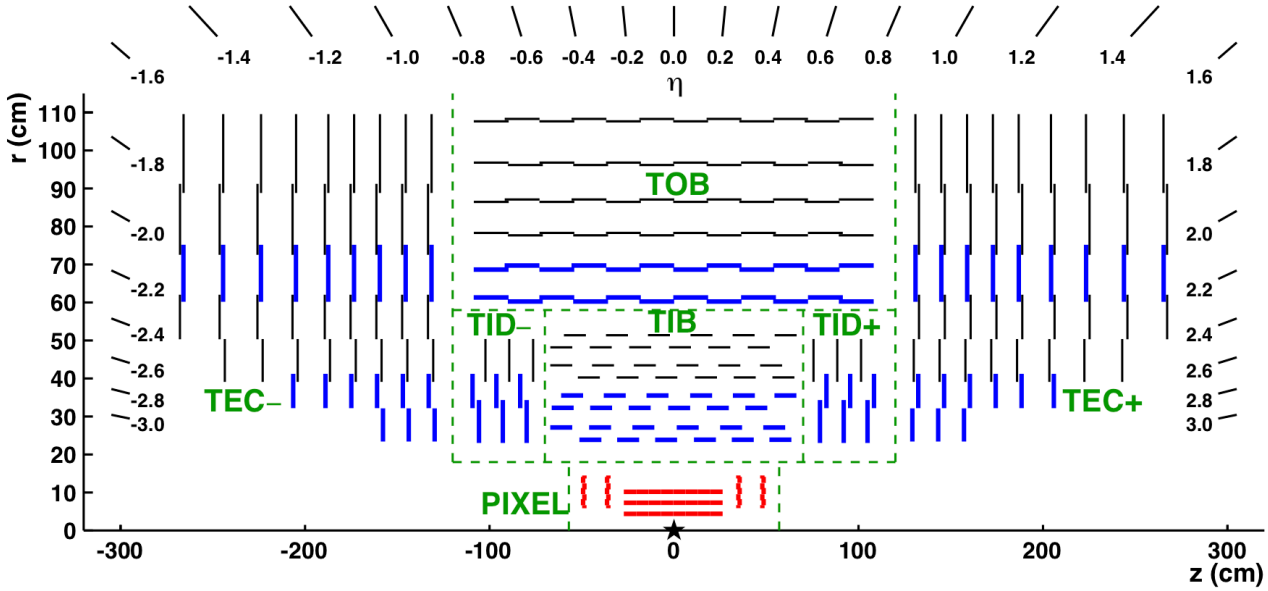


Figure 2.12 – *Schéma détaillé du trajectographe du détecteur CMS dans le plan (Orz) [131, 132]. Le trajectographe est symétrique par rapport à l'axe (Oz), axe du faisceau. Le centre du trajectographe, correspondant approximativement au lieu des collisions, est indiqué par une étoile. Les différentes sous-parties du trajectographe sont délimitées par les pointillés verts. Les modules à piste donnant des signaux en deux dimensions sont indiqués en lignes noires fines et ceux donnant des signaux en trois dimensions en lignes bleues épaisses. Ces derniers sont en fait constitués de deux modules à piste dos à dos dont l'un est tourné de 90°. Les modules à pixels, en rouge, permettent également d'obtenir des signaux à trois dimensions. Les légers décalages des modules leur permettent d'éviter tout angle mort dans la zone d'acceptation du détecteur.*

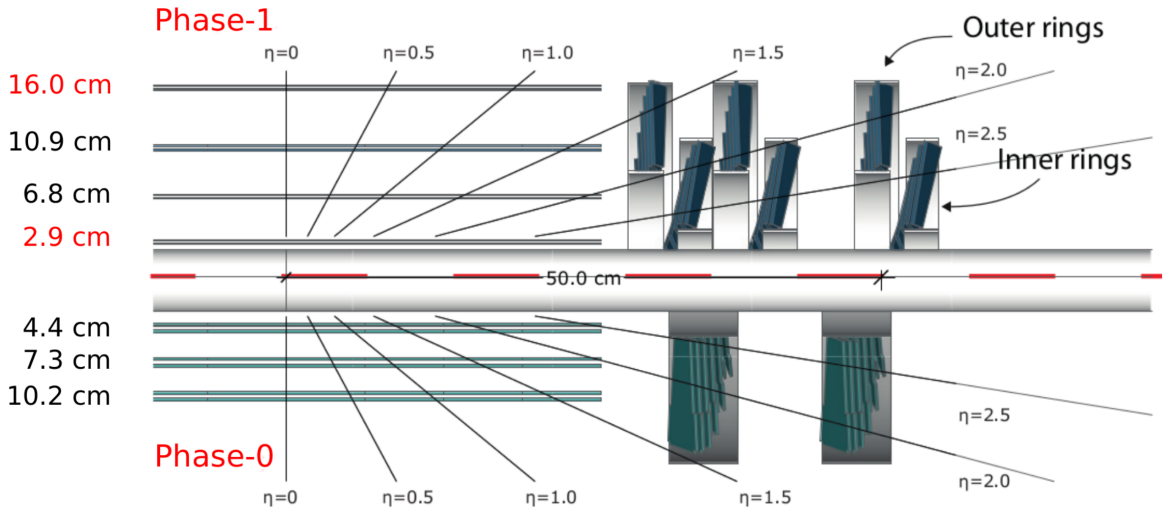
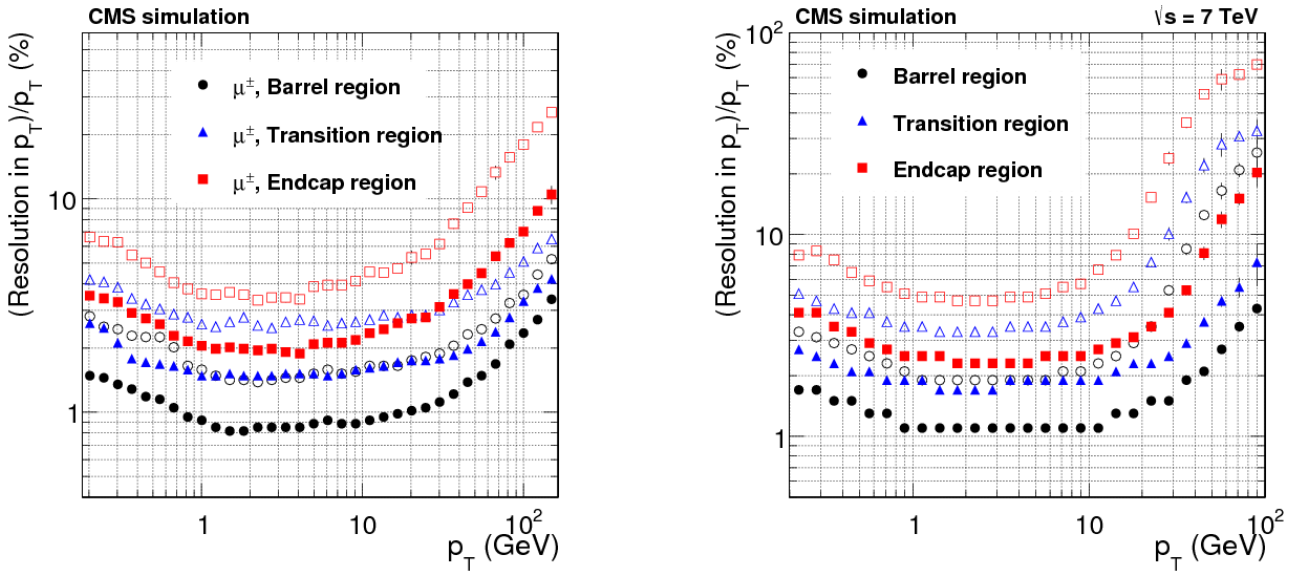


Figure 2.13 – *Comparaison des trajectographes à pixels utilisés en 2016 (Phase-0, en bas) et à partir de 2017 (Phase-1, en haut) [130, 133].*

4.4 Calorimètre électromagnétique ou ECAL

Le calorimètre électromagnétique (ECAL) [114, 122, 134-136] permet de mesurer l'énergie des électrons et des photons par un processus destructif. Le ECAL est composé d'environ 76 000 cristaux d'oxyde de tungstate de plomb ($PbWO_4$) dans lesquels les électrons et les photons explosent en une gerbe électromagnétique. Celle-ci donne naissance dans les cristaux à une impulsion lumineuse dont l'intensité permet d'obtenir l'énergie de la particule initiale.



(a) *Résolution pour des muons seuls et isolés.*

(b) *Résolution pour des particules chargées.*

Figure 2.14 – *Résolution en p_T, en fonction de p_T, du trajectographe pour différentes particules [131]. Les symboles pleins correspondent à la demi-largeur à 68 % de la distribution, les symboles creux à 90 %.*

Le ECAL se divise en trois sous-parties, schématisées sur la figure 2.15. Le tonneau du ECAL (EB) couvre la région $|\eta| < 1{,}479$. La face frontale des cristaux du EB se trouve à 1,29 m du faisceau. Les bouchons du ECAL (EE) couvrent la région $1{,}479 < |\eta| < 3{,}0$ et se trouvent à 315,4 cm du point de collision le long de l'axe du faisceau.

Devant les EE se trouvent les détecteurs de gerbes (*PreShower*). Leur rôle est d'identifier les pions neutres dans la région $1{,}653 < |\eta| < 2{,}6$. Ces particules se propagent sur des distances moyennes de 26 nm puis se désintègrent dans 99 % des cas en deux photons [9]. Ce sont donc ces deux photons

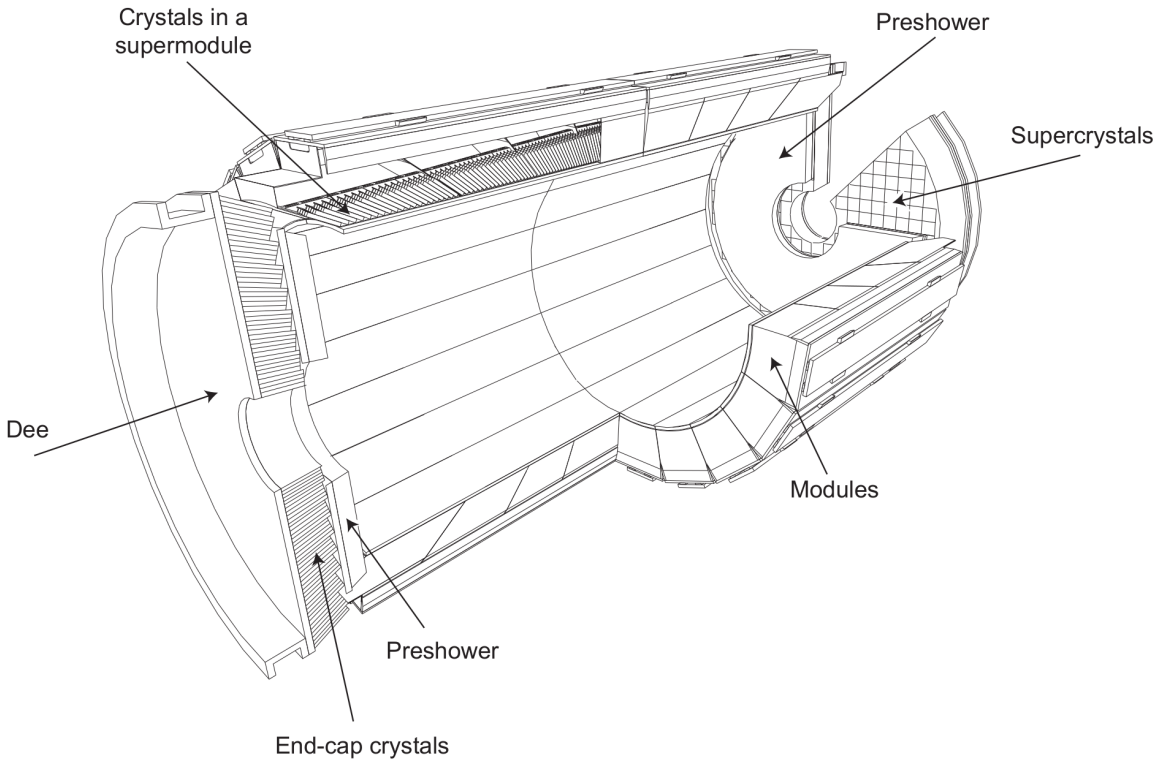


Figure 2.15 – *Schéma du calorimètre électronique de CMS [114, 134] montrant le positionnement des cristaux, modules et supermodules dans le tonneau, des supercristaux et du détecteur de gerbes dans les bouchons.*

que le *PreShower* doit identifier. Ce dernier aide à discriminer les électrons vis-à-vis des particules ionisantes ainsi qu'à déterminer les positions des photons et électrons. Il est composé d'une couche de plomb initiant la gerbe électromagnétique suivie d'un détecteur à pistes de silicium mesurant les dépôts d'énergie.

L'oxyde de tungstate de plomb est très dense, 8,29 g·cm^{-3}, et transparent. Ce matériau possède également une faible longueur d'interaction, $X_0 = 0{,}89$ cm, ainsi qu'un rayon de Molière de 2,19 cm. Le ECAL présente une réponse rapide, une bonne granularité et une résistance suffisante aux radiations. Près de 80 % du signal lumineux émis dans les cristaux du ECAL par les électrons et photons se trouve en effet dans une fenêtre temporelle de 25 ns correspondant à la durée entre deux événements successifs au LHC [114]. Dans le cas des hadrons, la traversée des cristaux du ECAL correspond approximativement à une longueur d'interaction. Près des deux tiers des hadrons initient donc une gerbe hadronique dans le ECAL, c'est-à-dire avant d'arriver dans le calorimètre hadronique.

Pour les électrons et les photons, la longueur des cristaux, 23 cm dans le tonneau et 22 dans les bouchons, correspond respectivement à 25,8 et 24,7 longueurs d'interaction, permettant d'absorber 98 % de leur énergie jusqu'à 1 TeV. Ces particules ne se propagent donc pas, en bonne approximation, dans les parties suivantes du détecteur.

La réponse des cristaux du ECAL est contrôlée régulièrement par l'injection de signaux lumineux issus de lasers [135]. La figure 2.16 présente l'évolution de la réponse des cristaux du ECAL depuis le début de l'exploitation du LHC. La réponse se dégrade au cours du temps car les cristaux, bien que peu sensibles aux radiations, se trouvent dans un environnement à très fortes radiations. Une perte de la transparence des cristaux est ainsi inévitable, diminuant leur réponse. Des corrections sont alors appliquées afin d'assurer une stabilité temporelle de la réponse du ECAL. De plus, la réponse des cristaux présente une forte dépendance thermique, de l'ordre de 2 %/°C. Un système de refroidissement assure une stabilité de la température des cristaux à ±0,5 °C.

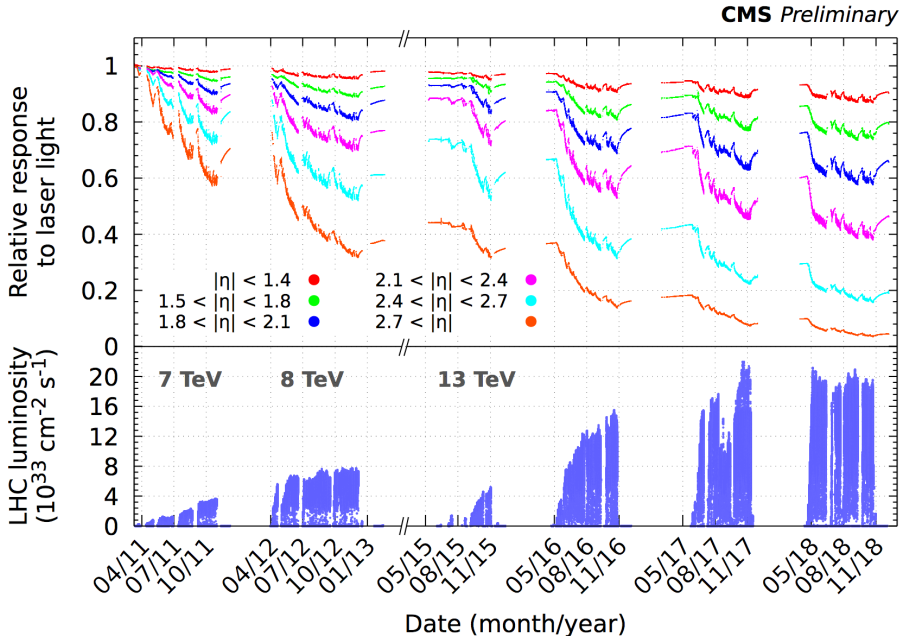


Figure 2.16 – *Évolution temporelle de la réponse du ECAL [135] (haut) et luminosité instantanée du LHC (bas).*

La résolution σ du ECAL est paramétrée selon

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C \tag{2.10}$$

où $\oplus$ désigne une somme quadratique, S est un terme stochastique prenant en compte la largeur latérale de la gerbe électronique, N le terme de bruit des composants électroniques et C une constante rendant compte des erreurs de calibration. Des tests en faisceau réalisés en 2006 [114] ont permis de mesurer $S = 0{,}028\,\text{GeV}^{1/2}$, $N = 0{,}12$ GeV et $C = 3{,}0\times10^{-3}$. La figure 2.17 présente la résolution relative du ECAL sur l'énergie des électrons lors du Run II.

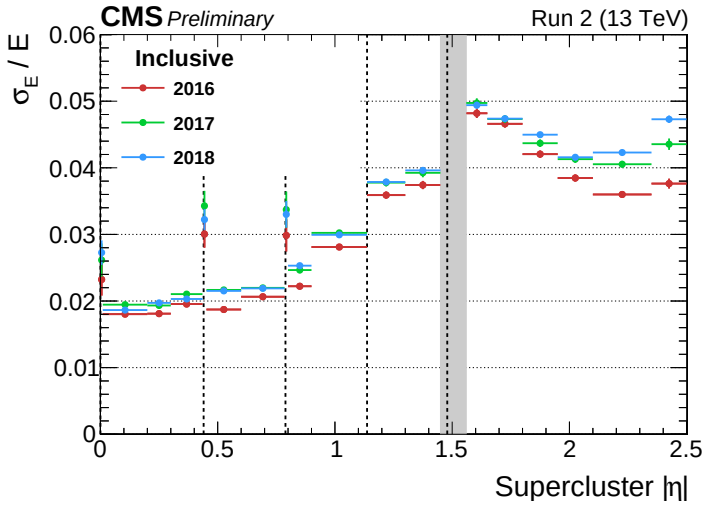


Figure 2.17 – *Résolution relative de l'énergie des électrons dans le ECAL lors du Run II en fonction de η [136]. La résolution est obtenue à partir d'événements $Z \to e^+e^-$.*

4.5 Calorimètre hadronique ou HCAL

Le calorimètre hadronique (HCAL) [114, 123, 137] permet de mesurer l'énergie des hadrons par un processus destructif. Le HCAL étant situé à l'intérieur du solénoïde de CMS, les particules y déposant leur énergie ne sont donc pas perturbées par une traversée du solénoïde. La figure 2.18 représente l'agencement du HCAL.

Tout comme le ECAL, il comporte un tonneau (HB) couvrant la région $|\eta| < 1{,}3$ et deux bouchons (HE) couvrant $1{,}3 < |\eta| < 3$. Le HCAL est composé de couches alternées d'absorbeurs et de scintillateurs. L'absorbeur, du laiton, permet d'initier la gerbe hadronique. Le scintillateur est fait en plastique. Des fibres optiques permettent de recueillir la lumière émise par les gerbes hadroniques. La mesure de ce signal lumineux donne une mesure de l'énergie des hadrons.

Cependant, le nombre de longueurs d'interaction combinées des ECAL et HCAL dans le tonneau, de l'ordre de dix, est insuffisant pour contenir toutes les gerbes hadroniques [114]. Le HB est ainsi complété par un calorimètre hadronique externe (HO) installé sur la face interne de la culasse, c'est-

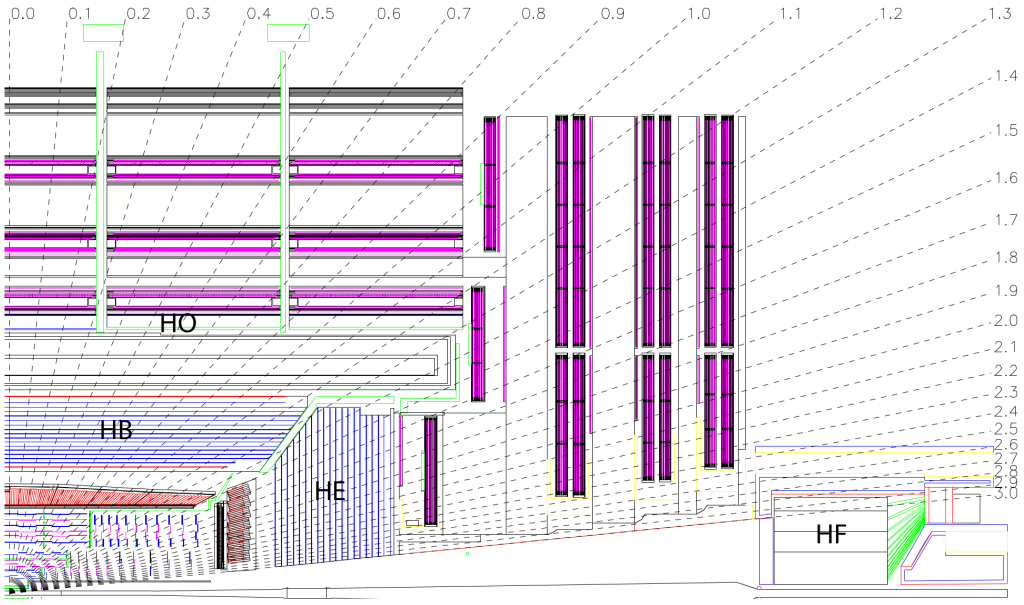


Figure 2.18 – *Schéma d'un cadrant du détecteur CMS [114] montrant la localisation des calorimètres hadroniques du tonneau (HB), externe (HO), du bouchon (HE) et vers l'avant (HF). Certaines valeurs de η et les directions associées sont indiquées.*

à-dire de l'autre côté du solénoïde et avant les chambres à muons.

Enfin, une couverture plus large en η est assurée par le calorimètre hadronique vers l'avant (HF) couvrant $2{,}9 < |\eta| < 5{,}2$. Les deux HF, un à chaque extrémité de CMS, sont des détecteurs cylindriques ayant des absorbeurs en acier dans lesquels passent des fibres optiques de quartz. Les particules incidentes émettent de la lumière Cherenkov lors de leur passage, cette lumière est alors recueillie par les fibres optiques.

La réponse relative du HCAL, c'est-à-dire l'énergie reconstruite dans le HCAL par rapport à l'énergie effective du dépôt a été estimée par simulation et est représentée en fonction de l'énergie simulée du dépôt sur la figure 2.19. Elle ne dévie pas de plus de 5 % au-delà de 10 GeV [138] une fois que l'empilement asynchrone, défini dans la section 3.5, est retiré [138, 139]. Comme dans le cas du ECAL, la réponse du HCAL évolue au cours du temps [140] et doit être contrôlée. La résolution σ obtenue sur l'énergie des hadrons, par combinaison avec les signaux du ECAL, a été déterminée à l'aide d'un faisceau test de pions comme étant

$$\frac{\sigma}{E} = \frac{1{,}1}{\sqrt{E}} + 0{,}09 \qquad (2.11)$$

où E est l'énergie mesurée en GeV.

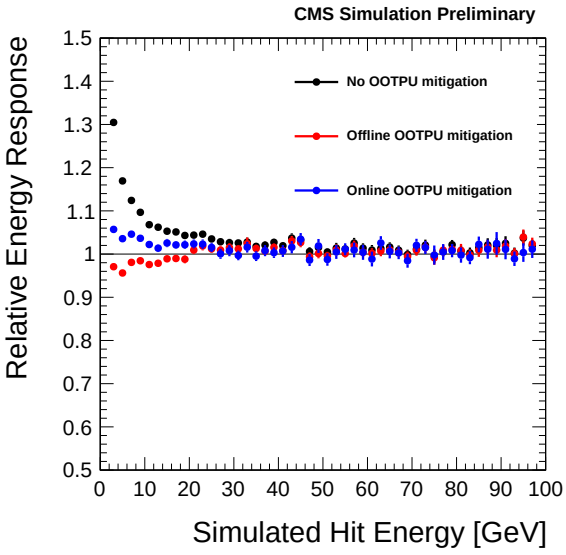


Figure 2.19 – *Réponse relative du calorimètre hadronique de CMS [138] en fonction de l'énergie simulée du dépôt, estimée par simulation. En noir, sans correction de l'empilement asynchrone (OOTPU). En bleu, avec des corrections en ligne, c'est-à-dire un ajustement des amplitudes et temps d'arrivée des signaux en prenant en compte jusqu'à trois signaux avant et après le signal d'intérêt. En rouge, avec l'ensemble des corrections.*

4.6 Chambres à muons

La détection des muons est un enjeu important pour la discrimination entre les signatures de processus physiques d'intérêt et le bruit de fond au LHC [114]. La mesure la plus précise possible des muons ainsi qu'une large couverture angulaire pour leur détection se trouve au cœur de la conception du détecteur CMS, comme l'indique le « M » de l'acronyme.

Les chambres à muons [114, 125, 141, 142], destinées à la mesure de ces particules, sont encastrées dans la culasse de retour du champ magnétique, c'est-à-dire dans la partie la plus externe du détecteur. Cette dernière, décrite dans la section 4.2, permet d'obtenir, à l'aide du solénoïde, un champ magnétique de 1 à 2 T dans cette zone [129], donnant une bonne résolution sur l'impulsion des muons. Les autres types de particules ayant été absorbés dans les couches précédentes, seuls les muons atteignent cette partie du détecteur. La figure 2.20 schématise la structure des chambres à muons dans un cadrant du détecteur.

Dans le tonneau (MB), cinq roues le long de l'axe du faisceau (numérotées de -2 à 2) couvrent la région $|\eta| < 1{,}2$. Une roue est composée de douze segments réalisant un tour complet en ϕ. Chacun d'entre eux comporte quatre stations ou couches successives de chambres à muons. Les conditions expérimentales permettent d'y utiliser des chambres à tubes à dérive (DT, *Drift Tubes*). Les trois premières stations contiennent douze chambres à muons. Deux groupes de quatre chambres mesurent la position du muon dans le plan transverse et sont séparés autant que possible afin d'obtenir la meilleure résolution angulaire. Les quatre autres donnent cette position le long de l'axe du faisceau. La quatrième station ne comporte pas de mesure selon cet axe.

Les bouchons (ME) sont soumis à une quantité plus importante de muons. Au lieu de tubes à dérive, la technologie utilisée est celle des chambres à pistes cathodiques (CSC, *Cathode Strip Chambers*). Les CSC présentent un temps de réponse plus court, une segmentation fine ainsi qu'une bonne résistance aux radiations. Elles couvrent la région $0{,}9 < |\eta| < 2{,}4$. Quatre stations de CSC successives

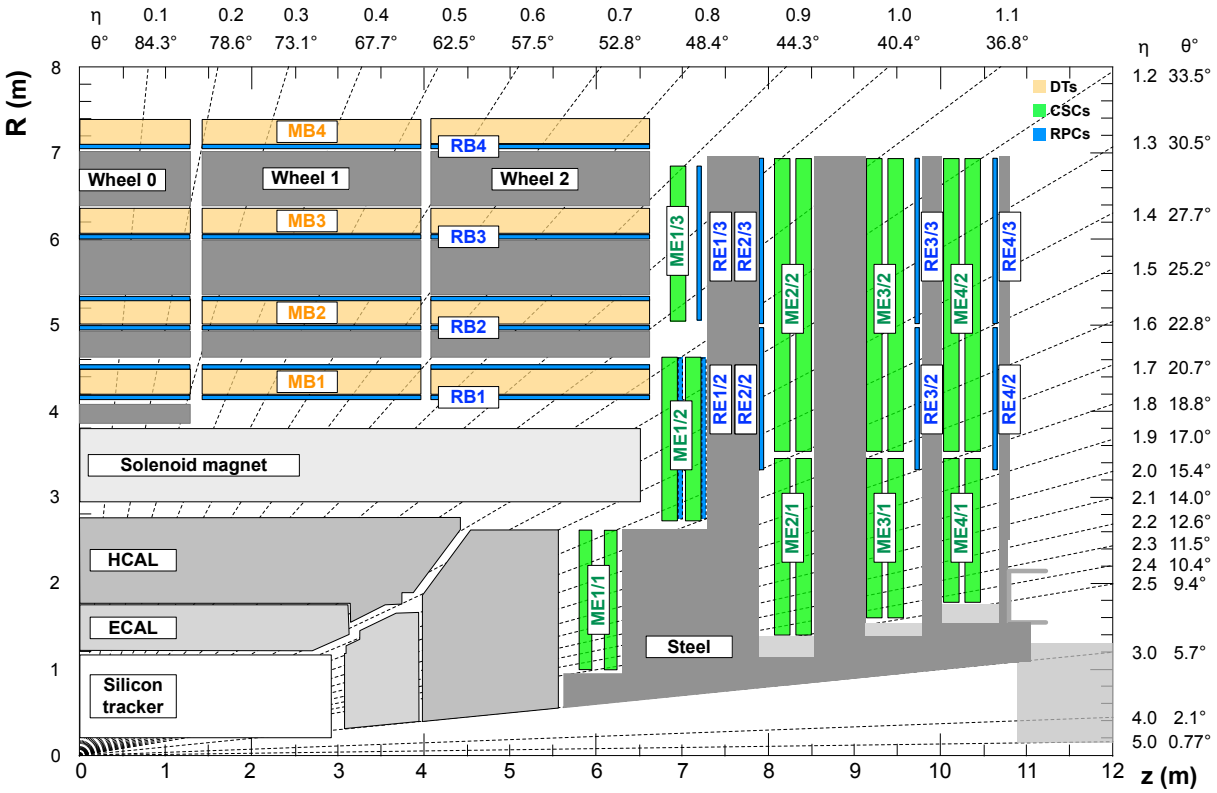


Figure 2.20 – *Schéma d'un cadrant du détecteur CMS [142] montrant la localisation des chambres à muons et leur nature : tubes à dérive (DT, en jaune), chambres à pistes cathodiques (CSC, en vert) et chambres à plaques résistives (RPC, en bleu). Certaines valeurs de η et les directions associées sont indiquées.*

sont installées dans les bouchons et sont orientées radialement par rapport au faisceau, donnant une mesure de précision dans le plan transverse de la position des muons.

Outre l'identification et la mesure des muons, les chambres à muons sont également utilisées afin de déclencher l'enregistrement des données. Les modalités du rejet ou de l'enregistrement d'un événement sont discutées plus en détail dans la section 4.7. Un système de déclenchement complémentaire aux DT et CSC est ajouté, il s'agit du RB dans le tonneau et du RE dans les bouchons. Il est composé de chambres à plaques résistives (RPC, *Resistive Plate Chambers*) fournissant des signaux indépendants des DT et CSC, rapides (moins de 25 ns) et proposant une bascule en fonction de l'impulsion transverse bien définie. La bascule est la densité de probabilité de donner un signal ou non. Il est en général favorable d'en avoir une à une valeur fixée afin de limiter le domaine de réponse aléatoire. La figure 2.21 montre la distribution de la masse invariante des dimuons, c'est-à-dire d'un ensemble de deux muons, sélectionnés par ce système de déclenchement. Ces résultats montrent la capacité du détecteur CMS à identifier les muons, se déclencher vis-à-vis de leur présence, de reconstruire leurs propriétés cinématiques et d'identifier ainsi sans ambiguïté les particules dont la désintégration donne ces muons sur une large gamme d'énergie [142]. La caractérisation complète des muons, abordée dans la section 5.2, utilise conjointement les informations du trajectographe et des chambres à muons.

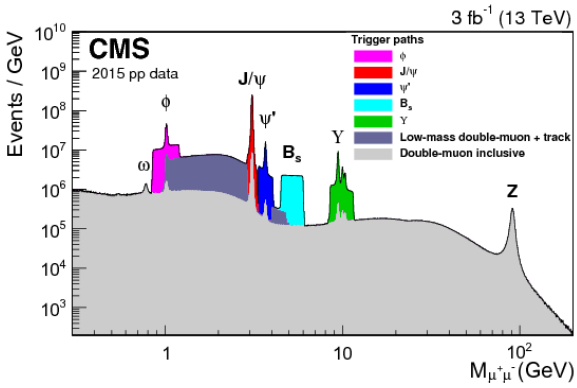


Figure 2.21 – *Distribution de la masse invariante des dimuons obtenus à partir du système de déclenchement des chambres à muons [142]. Les données ont été récoltées en 2015 à l'aide d'un déclenchement global (gris) ainsi que plusieurs déclenchements spécifiques (en couleur). Les résonances de diverses particules apparaissent distinctement.*

4.7 Prise de données à CMS

Pour un événement, le détecteur CMS décrit dans les sections précédentes produit une quantité de données de l'ordre de 1 Mo. Si cette grandeur peut paraître raisonnable, il faut la mettre en relation avec la fréquence des événements. Des collisions ont lieu toutes les 25 ns au LHC, soit à une fréquence de 40 MHz. Le détecteur CMS produit alors un flux de données de 40 To·s^{-1}, ce qui est bien trop important, tant pour l'électronique d'acquisition que pour la reconstruction des événements, abordée section 5, et pour la quantité de données à stocker elle-même. Afin d'en réduire le débit, la collaboration CMS s'appuie sur un système de déclenchement (*trigger*) [114, 143-146] dont le rôle est de supprimer les événements présentant peu d'intérêt physique.

Le système de déclenchement de CMS comporte deux niveaux, le niveau « L1 » (*Level-1*) et le niveau « HLT » (*High Level Trigger*). Le L1 est constitué d'un système électronique programmable utilisant les signaux issus des chambres à muons et des calorimètres pour estimer l'intérêt d'un événement, par exemple par la présence de muons ou d'activité calorimétrique. Il doit analyser chacun des événements. Ceux ne présentant ni muons ni activité calorimétrique suffisante sont directement rejetés. En revanche, jusqu'à 3,2 µs sont nécessaires au traitement des événements les plus complexes [114]. Un système de mise en attente est donc utilisé. Le L1 permet ainsi de réduire la fréquence des événements à analyser de 40 MHz à 100 kHz, limite imposée par l'électronique de lecture complète du détecteur [145]. La structure du L1 est illustrée sur la figure 2.22.

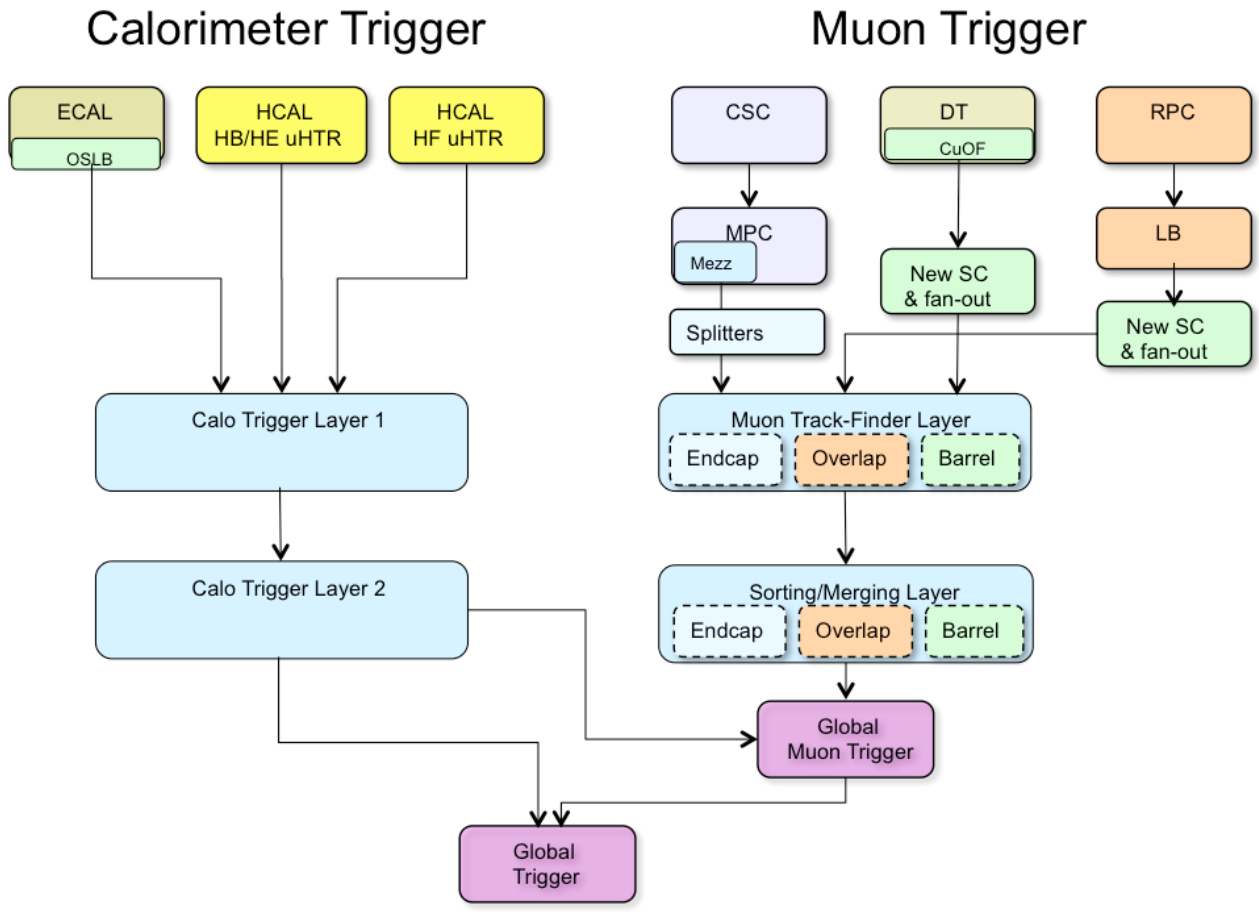


Figure 2.22 – *Architecture du système de déclenchement L1 [146].*

Le second niveau de déclenchement, le HLT, a accès à l'ensemble des signaux issus du détecteur. Une reconstruction simple des événements est réalisée sur une ferme de calcul et permet d'identifier les photons, muons, électrons, jets et taus hadroniques de l'événement. Des calculs plus complexes, similaires à ceux utilisés dans les analyses finales, sont alors réalisés [114]. Il est possible de concevoir de nombreuses conditions de sélection sur les particules présentes, leur énergie, etc. Il existe ainsi de nombreux chemins de déclenchement (HLT *paths*). La reconstruction simple des événements et leur

sélection se fait en un temps maximal de 50 ms. La fréquence moyenne des événements conservés est ainsi abaissée à 100 Hz [114, 145], fréquence raisonnable pour enregistrer les données.

5 Reconstruction des événements

L'analyse des événements à partir des signaux bruts issus du détecteur n'est pas aisée, qu'il s'agisse de données réelles ou simulées. Une interprétation de ces signaux en termes de particules physiques donne un point de départ beaucoup plus accessible. Pour y parvenir, un algorithme de reconstruction est utilisé. Son rôle est de déterminer quelles particules sont issues de la collision étant donnés les signaux dans le détecteur. Cette reconstruction est réalisée de manière identique dans les données réelles et simulées.

Les signaux caractéristiques des différents types de particules dans le plan transverse du détecteur CMS sont illustrés sur la figure 2.23. Il s'agit de la conséquence directe de la description du détecteur de la section 4. La plupart des particules laissent des signaux dans plusieurs sous-détecteurs. Pour une particule donnée, ces signaux doivent donc présenter une corrélation.

La reconstruction des événements se fait ainsi en combinant les informations issues des différents sous-détecteurs. Un algorithme spécialement développé afin d'optimiser cette combinaison, l'algorithme de reconstruction du flux de particules (PF, *Particle Flow*) [147, 148], a pour rôle de réaliser cette reconstruction. Dans un premier temps, l'algorithme de PF reconstruit les éléments d'identification des particules à partir des signaux de chaque sous-détecteur. Dans un second temps, ces éléments d'identification sont combinés afin de reconstruire les particules de l'événement. Des objets physiques plus complexes, dits de « haut niveau », sont également définis à partir de ces particules individuelles. Il s'agit des jets, de l'énergie transverse manquante et des taus hadroniques.

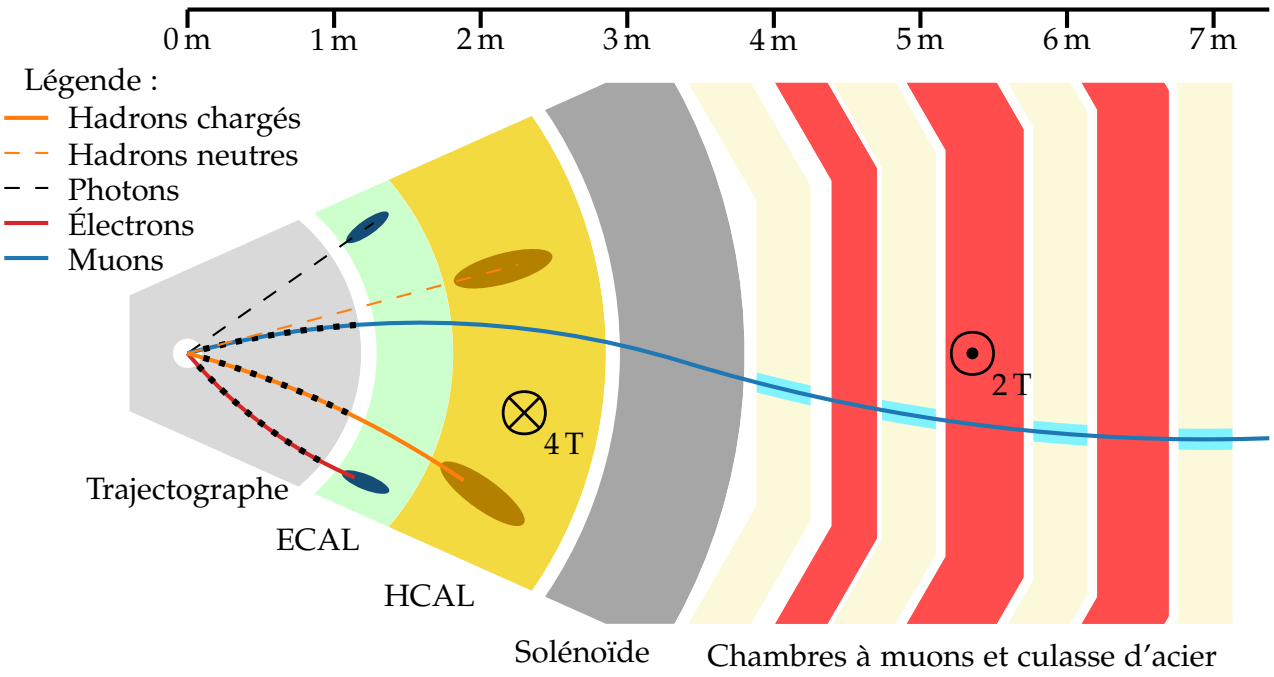


Figure 2.23 – *Coupe transverse schématique du détecteur CMS et signaux caractéristiques laissés par les particules. Les hadrons forment un dépôt dans le HCAL. Les hadrons chargés présentent de plus une trace dans le trajectographe dont l'extrapolation doit passer par ce dépôt. De même, les photons et électrons forment un dépôt dans le ECAL et les électrons présentent une trace dans le trajectographe dont l'extrapolation doit passer par ce dépôt. Enfin, les muons se propagent à travers tout le détecteur et laissent une trace dans les chambres à muons dont l'extrapolation doit correspondre à une trace du trajectographe. Dans toute la suite de ce manuscrit, les schémas des événements dans le détecteur utilisent la même présentation, en particulier pour les couleurs des sous-détecteurs et des particules. Réalisé à l'aide de CMSTransverseTikZ [51].*

5.1 Éléments d'identification du *Particle Flow*

5.1.1 Traces des particules chargées et vertex

Traces des particules chargées Les particules chargées laissent des traces de leur passage dans le trajectographe et, dans le cas des muons, dans les chambres à muons. Le trajectographe, présenté section 4.3, ne donne pas de traces continues mais des points de passage de ces particules.

Les traces des particules sont alors reconstruites à l'aide d'une méthode itérative [147, 149]. Un détecteur de traces par combinaison basé sur un filtre de Kalman (KF) [150] reconstruit les traces selon la procédure suivante :

1. Une trajectoire initiale est générée à partir d'un ajustement à quelques points de passages. Cette trajectoire doit être compatible avec celle d'une particule chargée.
2. L'ensemble des points de passage devant correspondre à cette trajectoire est déterminé. Ces points sont obtenus en extrapolant la trajectoire à travers le trajectographe. À chaque ajout d'un point de passage, l'ajustement de la trajectoire à l'ensemble des points qui lui sont associés est mis à jour.
3. Une fois que l'ensemble des points pouvant être associés à la trajectoire a été trouvé, un ajustement final est réalisé. Il permet de déterminer l'ajustement optimal de la trajectoire aux points de passage. Les propriétés de la particule chargée (origine, impulsion transverse, direction) sont ainsi déterminées. La trace est acceptée ou rejetée selon des critères portant sur les paramètres de l'ajustement final. Les points de passage associés à une trace acceptée sont retirés de la liste des points à traiter.

Cette procédure est alors répétée en utilisant des critères de plus en plus précis afin d'augmenter l'efficacité de la reconstruction des traces. Le taux de mauvaise reconstruction est quant à lui réduit en appliquant des critères de qualité sur les traces reconstruites.

Cas des électrons La reconstruction des électrons de haute énergie et isolés vis-à-vis des autres particules est naturellement basée sur les données du ECAL [151]. La position et la valeur du dépôt d'énergie sont utilisées afin de déterminer les points de passage attendus dans le trajectographe pour l'électron (ou le positron). Cependant, à cause de l'épaisseur du trajectographe, les électrons émettent, avant de parvenir au ECAL, une fraction de leur énergie par *bremsstrahlung*, c'est-à-dire sous forme de photons. Les performances de reconstruction des électrons dépendent ainsi fortement de la capacité à identifier ces photons et mesurer leurs énergies. Les dépôts d'énergie dans le ECAL (*clusters*) dus à l'électron et aux photons potentiellement issus du *bremsstrahlung* sont regroupés en un *supercluster*. Le *supercluster* est construit à partir des dépôts du ECAL se situant dans une fenêtre plus large en ϕ qu'en η afin de prendre en compte la courbure de la trajectoire de l'électron due au champ magnétique.

Dans le cas des électrons contenus dans les jets, en revanche, un biais apparaît à cause des autres particules du jet situées à proximité de l'électron. Cet effet mène à un fort taux de mauvaise reconstruction. L'identification des électrons à partir du ECAL est ainsi limitée aux cas d'électrons isolés des autres particules. Afin de reconstruire les électrons exclus par cette limitation, une approche basée sur les données du trajectographe a été développée [147]. Dès qu'une trace reconstruite selon la procédure exposée précédemment vérifie $p_{\mathrm{T}} > 2\,\mathrm{GeV}$, elle est considérée comme candidat électron.

Lorsque l'électron émet peu d'énergie par *bremsstrahlung*, l'ajustement de la trace correspondante est de bonne qualité à travers l'ensemble du trajectographe. De plus, l'extrapolation de cette trace jusqu'à la surface du ECAL permet de lui faire correspondre un dépôt d'énergie. L'énergie contenue dans le dépôt doit alors être compatible avec celle déterminée à partir de la trace.

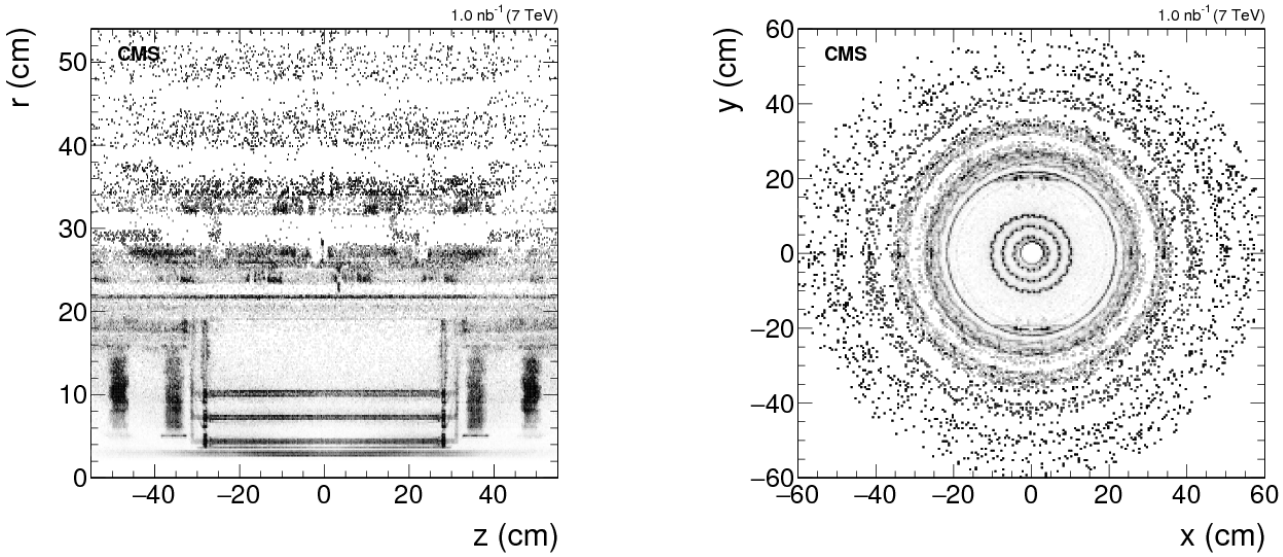
Plus l'électron émet de l'énergie par *bremsstrahlung*, moins l'ajustement de la trace correspondante est de qualité, car cet ajustement ne prend pas en compte la perte d'énergie au cours de la propagation de l'électron. Les traces correspondant à cette situation sont sélectionnées à partir du nombre de points de passage associés et du χ^2 de l'ajustement. Elles sont réajustées à partir d'un filtre de somme de gaussiennes (GSF, *Gaussian Sum Filter*) [152] utilisant cinq paramètres, plus adapté que le KF pour ces électrons [147]. Un critère d'identification basé sur un arbre de décision (BDT, *Boosted Decision Tree*) est appliqué sur cet ajustement, il est détaillé dans la section 5.2.

Cas des muons Les muons laissent des signaux de leur passage dans le trajectographe et dans les chambres à muons. Trois types de muons peuvent être définis :

— les muons seuls (*standalone muons*), reconstruits uniquement à partir des signaux des chambres à muons ;
— les muons globaux (*global muons*), obtenus par la correspondance d'une trace dans le trajectographe avec l'extrapolation de la trace d'un muon seul ;
— les muons du trajectographe (*tracker muons*) sont les traces du trajectographe d'impulsion transverse supérieure à 0,5 GeV dont l'extrapolation passe par une des chambres à muons ayant détecté le passage d'une particule.

Lorsque des hadrons chargés arrivent au HCAL, ils se désintègrent en une gerbe hadronique. Si des éléments de cette gerbe parviennent à traverser le HCAL (*punch-through*), ils donnent un signal dans les chambres à muons. Des critères d'identification sont alors appliqués aux traces afin de réduire la quantité de muons reconstruits à partir de hadrons. Ils sont abordés dans la section 5.2.

Interactions avec le matériau du trajectographe Des interactions entre les particules issues des collisions et le matériau du trajectographe peuvent mener à une déviation (*kink*) [153, 154] voire à la production de vertex secondaires [132, 147], comme illustré sur la figure 2.24. Un algorithme a été développé [155] afin d'identifier les traces de particules ayant interagit avec le matériau du trajectographe.



(a) *Dans le plan longitudinal* (Orz). **(b)** *Dans le plan transverse* (Oxy).

Figure 2.24 – *Carte des points d'interaction entre particules des événements et matière composant le détecteur [132, 147] à partir de données prises en 2011 à* $\sqrt{s} = 7\,\text{TeV}$.

Vertex La combinaison des traces permet de reconstruire les vertex d'interactions de l'événement. Plusieurs vertex sont présents du fait de l'empilement. Le vertex principal est choisi comme étant le vertex dont la somme des impulsions transverses au carré des traces en provenant est la plus élevée, les autres sont considérés comme des vertex de l'empilement. L'efficacité de reconstruction du vertex principal est ainsi de l'ordre de 100 %, celle des vertex de l'empilement de 70 % [156].

5.1.2 Dépôts dans les calorimètres

Agglomération L'énergie d'une particule unique se répartit dans plusieurs cellules des calorimètres. Les dépôts dans chacune des cellules sont alors regroupés de proche en proche en agglomérats (*clusters*) [147].

La construction de ces agglomérats commence par l'identification des cellules des calorimètres mesurant une énergie supérieure à un seuil donné, défini pour chaque sous-partie des calorimètres.

Les seuils sont fixés à partir d'une optimisation sur des simulations de photons, π^0, K^0 et jets. Les cellules adjacentes sont ajoutées à l'agglomérat. Puis, toute cellule avec au moins un coin en commun avec une cellule déjà dans l'agglomérat et mesurant une énergie supérieure à deux fois le niveau moyen du bruit est ajoutée à l'agglomérat.

Calibration Les photons et les hadrons neutres sont reconstruits à l'aide de leurs dépôts dans les calorimètres. Des dépôts isolés vis-à-vis des traces de particules chargées sont une signature claire des particules neutres. Cependant, un dépôt de particule neutre situé au même endroit qu'un dépôt de particule chargée ne peut être détecté que comme étant un excès d'énergie pour la particule chargée par rapport à l'énergie déterminée à l'aide du trajectographe. Une bonne calibration de la réponse des calorimètres aux photons et aux hadrons est donc cruciale pour la bonne reconstruction des particules neutres. Cette calibration a été réalisé dans un premier temps avant les premières collisions à l'aide de faisceaux de test. Une fois les collisions commencées, les données issues des celles-ci sont exploitées afin de calibrer plus finement les calorimètres.

5.2 Identification et reconstruction des particules

Une seule particule donne généralement lieu à plusieurs éléments d'identification du PF dans différents sous-détecteurs. La reconstruction des particules se fait alors par association de ces éléments. L'association des éléments dus à une particule est limitée par [147] :

— la granularité des sous-détecteurs ;
— le nombre de particules par unité d'angle solide ;
— la quantité de matière traversée par les particules en amont des calorimètres ou, le cas échéant, des chambres à muons, pouvant dévier la particule [153, 154].

Un algorithme teste les associations possibles d'éléments de reconstruction provenant de différents sous-détecteurs. Afin de limiter les temps de calcul, seules les associations d'éléments les plus proches entre eux selon un arbre de recherche multidimentionnel [157] sont considérées. Des conditions supplémentaires sont requises afin d'associer deux éléments et sont détaillées dans les sections suivantes. La distance définie par l'arbre de recherche multidimentionnel permet de quantifier la qualité de cette association. Une association d'élément constitue un « bloc » du PF, dont chacun des éléments est relié aux autres, directement par une association ou indirectement par un élément associé commun. Les blocs servent de base à la reconstruction des objets physiques, dont l'identification en tant que particule d'un type donné dépend de critères spécifiques. Ces différents types de particules sont détaillés dans les sections qui suivent.

5.2.1 Muons

Isolation des muons L'isolation permet de quantifier la présence d'activité autour de l'objet physique considéré comme un muon. Elle est définie à partir des autres particules reconstruites situées dans un cône de rayon

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < R_\mu = 0{,}4 \tag{2.12}$$

autour de la direction du muon au niveau du vertex primaire principal, avec $\Delta\eta$ et $\Delta\phi$ les distances angulaires des particules au muon dans les directions η et ϕ, selon

$$I^\mu = \sum_{\text{ch,PV}} {p_\text{T}}^{\text{ch}} + \max\left(0, \sum_{\text{h}^0} {E_\text{T}}^{\text{h}^0} + \sum_{\gamma} {E_\text{T}}^{\gamma} - \Delta\beta \sum_{\text{ch,PU}} {p_\text{T}}^{\text{ch}}\right)\Bigg|_{\Delta R < R_\mu} \tag{2.13}$$

où $\sum_{\text{ch,PV}} {p_\text{T}}^{\text{ch}}$ est la somme scalaire des impulsions transverses des particules chargées provenant du vertex primaire principal à l'exception de ce muon, $\sum_{\text{h}^0} {E_\text{T}}^{\text{h}^0}$ est la somme des énergies dans le plan transverse de tous les hadrons neutres, $\sum_{\gamma} {E_\text{T}}^{\gamma}$ est la somme des énergies dans le plan transverse de tous les photons, $\sum_{\text{ch,PU}} {p_\text{T}}^{\text{ch}}$ est la somme scalaire des impulsions transverses des particules chargées provenant de l'empilement et $\Delta\beta$ est une estimation du rapport entre particules neutres et particules chargées créées lors des collisions de protons. Le second terme de l'équation (2.13) permet ainsi d'estimer la contribution des particules neutres à l'isolation. La variable d'isolation ainsi

construite est basse pour des particules isolées, haute pour des particules non isolées. Il est possible de définir l'isolation relative comme étant le rapport entre l'isolation et l'impulsion transverse de la particule,

$$I^i_{\text{rel}} = \frac{1}{{p_{\text{T}}}^i} I^i \,. \tag{2.14}$$

Lorsqu'un objet physique est considéré à tort comme étant un muon, la valeur de I^μ est plus élevée. C'est le cas par exemple pour un hadron chargé présentant un *punch-through*, dont le dépôt dans le HCAL augmente I^μ. De plus, des muons peuvent être produits lors de la désintégration de quarks de saveur lourde. Ces désintégrations sont accompagnées de jets, comme exposé dans le chapitre 1. Ces muons font alors partie du jet et ne sont pas issus de la collision initiale. La présence du jet mène également à de plus hautes valeurs de I^μ. L'isolation permet donc de rejeter des muons mal identifiés ou issus de jets.

Critères d'identification des muons Il est possible de définir un critère de qualité sur l'objet reconstruit devant correspondre à un muon à partir des propriétés des éléments d'identification du PF. Il s'agit du *muonID* [142, 158]. Le χ^2 de l'ajustement de la trajectoire ainsi que la fraction des signaux du trajectographe valides associés au muon sont des métriques utilisées pour le *muonID*. De plus, un algorithme (*kink finder*) sépare la trajectoire du muon et détermine un χ^2 afin de vérifier si cette trajectoire reconstruite peut en réalité provenir de deux traces réelles distinctes. Ce dernier cas de figure peut survenir suite à une déviation du muon par le matériel constituant le détecteur, par exemple.

Pour des critères d'identification stricts, il est possible d'utiliser le nombre de points de passage (*hits*) dans les chambres à muons utilisés pour l'ajustement global de la trajectoire du muon, $N^{\text{MD}}_{\text{hits}}$. Le nombre de stations de chambres à muons associées à la trajectoire, N_{MS}, est aussi exploité. Les informations issues du trajectographe sont également utilisées. Il s'agit du nombre de *hits* dans la partie à pixels, $N^{\text{pixel}}_{\text{hits}}$, et du nombre total de *hits* dans le trajectographe, $N^{\text{tracker}}_{\text{hits}}$.

Trois niveaux d'exigence ou points de fonctionnement (WP, *Working Point*) sont définis, de plus en plus exigeants. En particulier, le *Medium muonID* est utilisé dans l'analyse présentée chapitre 4, comme le recommande le *POG* (*Physics Object Group*, groupe responsable d'un objet physique) Muons [158].

Loose muonID (exigence lâche)
- le muon est issu du PF ;
- le muon est reconstruit comme muon global ou du trajectographe.

Medium muonID (exigence moyenne)
- le muon passe le *loose muonID* ;
- au moins 80 % des signaux du trajectographe associés au muon sont valides.
- s'il s'agit d'un muon global :
  - l'ajustement de la trajectoire vérifie $\chi^2/N_{\text{dof}} < 3$, avec N_{dof} le nombre de degrés de liberté de l'ajustement,
  - l'accord entre le muon seul et le muon du trajectographe issus des mêmes éléments de reconstruction que le muon global vérifie $\chi^2 < 12$,
  - la compatibilité avec une déviation du muon due au matériel du détecteur (*kink finder*) vérifie $\chi^2 < 20$,
  - la compatibilité du segment est supérieure à 0,303 ;
- s'il s'agit d'un muon du trajectographe :
  - la compatibilité du segment est supérieure à 0,451.

Tight muonID (exigence stricte)
- le muon est issu du PF ;
- le muon est reconstruit comme muon global ;
- l'ajustement de la trajectoire vérifie $\chi^2/N_{\text{dof}} < 10$;
- les chambres à muon vérifient $N^{\text{MD}}_{\text{hits}} > 0$ et $N_{\text{MS}} > 1$;

— le trajectographe vérifie $N_{\text{hits}}^{\text{pixel}} > 0$ et $N_{\text{hits}}^{\text{tracker}} > 5$;
— les paramètres d'impact du muon vis-à-vis du vertex primaire principal vérifient $d_{xy} < 2\,\text{mm}$ et $d_z < 5\,\text{mm}$.

Reconstruction des muons Les muons sont reconstruits à partir des éléments d'identifications que sont les muons globaux, seuls et du trajectographe définis dans la section 5.1.1.

Tout d'abord, les muons globaux isolés, c'est-à-dire sans autre activité dans le voisinage de la trajectoire correspondante, sont sélectionnés [147]. Les traces additionnelles et les dépôts d'énergie dans les calorimètres se situant dans un cône de rayon ΔR inférieur à 0,3 dans le plan (η, ϕ) sont également associés au muon global. Il est requis que l'isolation relative du muon global soit inférieure à 0,1. Ce critère est suffisant pour rejeter les hadrons dont la gerbe de désintégration traverse le HCAL. Ensuite, les muons globaux non isolés sont sélectionnés à l'aide du critère d'identification strict (*Tight muonID*).

Les muons non identifiés à ce stade peuvent l'être en utilisant les muons seuls et les muons du trajectographe. Les muons seuls présentant un grand nombre de signaux dans les chambres à muons, au moins 23 dans les DT (pour un maximum possible de 32) ou 15 dans les CSC (pour un maximum possible de 24), et dont l'ajustement de la trace à ces signaux est de bonne qualité sont ainsi retenus. Les muons du trajectographe sont également retenus s'ils contiennent au moins 13 points de passage dans le trajectographe et que les agglomérats dans les calorimètres sont compatibles avec la traduction de la trace correspondante en tant que muon.

La résolution sur l'impulsion transverse des muons reconstruits est de 1 % dans le tonneau et 3 % dans les bouchons pour les muons d'impulsion transverse inférieure à 100 GeV et inférieure à 7 % dans le tonneau jusqu'à $p_{\text{T}} = 1\,\text{TeV}$ [142]. L'efficacité de reconstruction est de 95 % et le taux d'identification de hadrons en tant que muons inférieur à 1 %. Les éléments d'identification du PF utilisés pour reconstruire les muons sont retirés dans la suite du processus de reconstruction de l'événement.

5.2.2 Électrons et photons isolés

Isolation des électrons L'isolation des électrons est définie de manière similaire à celle des muons. Elle est quantifiée à partir des particules situées dans un cône de rayon

$$\Delta R < R_e = 0{,}3 \tag{2.15}$$

autour de la direction de l'électron au niveau du vertex primaire principal, selon

$$I^e = \sum_{\text{ch,PV}} {p_{\text{T}}}^{\text{ch}} + \max\left(0, \sum_{\text{h}^0} {E_{\text{T}}}^{\text{h}^0} + \sum_{\gamma} {E_{\text{T}}}^{\gamma} - \rho \times \mathcal{E}_{\mathcal{A}}\right)\Bigg|_{\Delta R < R_e} \tag{2.16}$$

où $\sum_{\text{ch,PV}} {p_{\text{T}}}^{\text{ch}}$ est la somme scalaire des impulsions transverses des particules chargées provenant du vertex primaire principal à l'exception de cet électron, $\sum_{\text{h}^0} {E_{\text{T}}}^{\text{h}^0}$ est la somme des énergies dans le plan transverse de tous les hadrons neutres, $\sum_{\gamma} {E_{\text{T}}}^{\gamma}$ est la somme des énergies dans le plan transverse de tous les photons, ρ est la densité d'énergie issue de l'empilement dans l'événement et $\mathcal{E}_{\mathcal{A}}$ est l'aire effective, c'est-à-dire la fraction de l'espace (η, ϕ) correspondant à la zone d'isolation à corriger pour l'empilement. Les valeurs des aires effectives utilisées sont présentées dans le tableau 2.3.

Région	$\mathcal{E}_{\mathcal{A}}$
$\lvert\eta\rvert \leqslant 1{,}0$	0,1440
$1{,}0 < \lvert\eta\rvert \leqslant 1{,}479$	0,1562
$1{,}479 < \lvert\eta\rvert \leqslant 2{,}0$	0,1032
$2{,}0 < \lvert\eta\rvert \leqslant 2{,}2$	0,0859
$2{,}2 < \lvert\eta\rvert \leqslant 2{,}3$	0,1116
$2{,}3 < \lvert\eta\rvert \leqslant 2{,}4$	0,1321
$\lvert\eta\rvert > 2{,}4$	0,1654

Tableau 2.3 – *Valeurs de l'aire effective $\mathcal{E}_{\mathcal{A}}$ utilisée pour corriger la contribution de l'empilement aux isolations des électrons vis-à-vis des autres particules.*

Critères d'identification des électrons Deux critères d'identification des électrons existent. Le premier est issu d'une analyse multivariée (MVA, *MultiVariate Analysis*), le second est obtenu à partir de coupures (*cuts*) sur certaines variables.

Electron ID MVA Ce critère est basé sur un arbre de décision (BDT, *Bossted Decision Tree*) [159]. Le BDT est entraîné à l'aide de XGBOOST [160] sur des événements Drell-Yan (Z/γ^*) + jets simulés à l'aide de MADGRAPH [161]. Les variables prises en compte sont :

- l'impulsion transverse de l'électron ${p_T}^e$;
- la pseudo-rapidité du *supercluster* ;
- la densité d'énergie issue de l'empilement dans l'événement ρ ;
- les étalements en η et en ϕ du dépôt d'énergie dans le ECAL, $\sigma_{i\eta i\eta}$ et $\sigma_{i\phi i\phi}$ où $i\eta$ et $i\phi$ correspondent au nombre entier désignant le cristal du calorimètre ;
- la circularité du dépôt d'énergie, $1 - E_{1\times5}/E_{5\times5}$, où $E_{1\times5}$ et $E_{5\times5}$ correspondent respectivement aux énergies dans une région de 1×5 et 5×5 cristaux centrée sur le cristal contenant le plus d'énergie dans le *supercluster* ;
- $R_9 = \frac{E_{3\times3}}{E_{SC}}$, où $E_{3\times3}$ correspond à l'énergie dans une région de 3×3 cristaux centrée sur le cristal contenant le plus d'énergie dans le *supercluster* et E_{SC} est l'énergie contenue dans le *supercluster* ;
- H/E_{SC} où H est l'énergie hadronique située dans un cône de $\Delta R < 0{,}15$ autour de l'électron ;
- $E_{SC}^{\mathrm{PS}}/E_{SC}^{\mathrm{raw}}$ le rapport de l'énergie du *supercluster* située dans le *PreShower* sur son énergie totale non corrigée ;
- la largeur du *supercluster*, $\Delta\eta_{SC}$ et $\Delta\phi_{SC}$;
- le χ^2 de l'ajustement de la trajectoire ;
- le nombre de *hits* valides utilisés pour l'ajustement de la trajectoire ;
- le χ^2 de l'ajustement de la trajectoire GSF (*GSFtrack*) ;
- le nombre de *hits* utilisés pour l'ajustement de la trajectoire GSF, N^{GSF} ;
- le nombre attendu de *hits* manquants, $N_{\mathrm{lost}}^{\mathrm{GSF}}$;
- la fraction d'énergie perdue par *bremsstrahlung*, $f_{\mathrm{brem}} = 1 - p_{\mathrm{out}}/p_{\mathrm{in}}$, avec p_{in} et p_{out} les impulsions de l'électron obtenue d'après la courbe de sa trajectoire près du vertex primaire et près de la surface du ECAL respectivement ;
- E_{SC}/p_{in} ;
- $E_{\mathrm{PF}}/p_{\mathrm{in}}$, où E_{PF} est l'énergie du *supercluster* le plus proche du point d'entrée de l'électron dans le ECAL ;
- les écarts $\Delta\eta_{\mathrm{in}}$ et $\Delta\phi_{\mathrm{in}}$ entre le *supercluster* et la direction de la trace associée à l'électron au niveau du vertex primaire ;
- l'écart $\Delta\eta_{\mathrm{out}}$ entre le *supercluster* et la direction de la trace associée à l'électron au niveau de la surface du ECAL ;
- $1/E_e - 1/p_e$ où E_e est l'énergie de l'électron et p_e son impulsion ;
- la probabilité que l'électron soit issu d'une conversion $\gamma \to e^+e^-$;

Le point de fonctionnement à 90 % d'efficacité est défini à partir d'une valeur minimale de sortie du BDT. Cette valeur dépend de ${p_T}^e$ et η^e ainsi que de l'année de prise de données.

Cut-based ID Ce critère d'identification consiste en une liste de coupures (*cut*) sur certaines variables. Les valeurs de ces coupures dépendent du point de fonctionnement. Dans l'analyse présentée chapitre 4, seul le point de fonctionnement « veto » est utilisé, les coupures associées sont listées dans le tableau 2.4. Les variables utilisées sont définies précédemment, à l'exception de :

- $|\Delta\eta_{\mathrm{in}}^{\mathrm{seed}}|$ l'écart en η entre le point d'entrée de l'électron dans le ECAL et la position du *supercluster* identifié par l'algorithme de PF ;
- $I_{\mathrm{rel}}^{\Delta\beta}$ l'isolation relative de l'électron obtenue de façon analogue à celle des muons (2.14), à l'exception de la taille du cône valant ici $R_e = 0{,}3$.

Isolation des photons L'isolation des photons est définie vis-à-vis des hadrons chargés, des hadrons neutres et des autres photons selon le même principe que l'isolation des électrons équation (2.16), certains termes étant nuls suivant l'isolation déterminée. Les aires effectives correspondantes sont données dans le tableau 2.5.

Critères d'identification des photons À l'instar du *cut-based ID*, la collaboration CMS propose des critères d'identification des photons (lâche, moyen et strict) à partir de coupures sur diverses propriété-

Variable	$\lvert\eta^e\rvert < 1{,}479$	$\lvert\eta^e\rvert \geqslant 1{,}479$
$\sigma_{i\eta i\eta}$	$< 0{,}0126$	$< 0{,}0457$
$\lvert\Delta\eta_{\text{in}}^{\text{seed}}\rvert$	$< 0{,}004\,63$	$< 0{,}008\,14$
$\lvert\Delta\phi_{\text{in}}\rvert$	$< 0{,}148$	$< 0{,}19$
H/E_{SC}	$< 0{,}05 + \frac{1{,}16}{E_{SC}[\text{GeV}]} + 0{,}0324\,\frac{\rho}{E_{SC}}$	$< 0{,}05 + \frac{2{,}54}{E_{SC}[\text{GeV}]} + 0{,}183\,\frac{\rho}{E_{SC}}$
$I_{\text{rel}}^{\Delta\beta}$	$< 0{,}198 + \frac{0{,}506}{p_{\text{T}}{}^e[\text{GeV}]}$	$< 0{,}203 + \frac{0{,}96}{p_{\text{T}}{}^e[\text{GeV}]}$
$\lvert 1/E_{SC} - 1/p_{\text{in}}\rvert$	$< 0{,}209\,\text{GeV}^{-1}$	$< 0{,}132\,\text{GeV}^{-1}$
$N_{\text{lost}}^{\text{GSF}}$	$\leqslant 2$	$\leqslant 3$
veto de conversion	passé	passé

Tableau 2.4 – *Coupures du* cut-based veto ID *pour les deux régions en η du* supercluster *possibles. Les variables sont détaillées dans le texte.*

Région	Hadrons chargés	Hadrons neutres	Photons
$\lvert\eta\rvert \leqslant 1{,}0$	0,0112	0,0668	0,1113
$1{,}0 < \lvert\eta\rvert \leqslant 1{,}479$	0,0108	0,1054	0,0953
$1{,}479 < \lvert\eta\rvert \leqslant 2{,}0$	0,0106	0,0786	0,0619
$2{,}0 < \lvert\eta\rvert \leqslant 2{,}2$	0,010 02	0,0233	0,0837
$2{,}2 < \lvert\eta\rvert \leqslant 2{,}3$	0,0098	0,0078	0,1070
$2{,}3 < \lvert\eta\rvert \leqslant 2{,}4$	0,0089	0,0028	0,1212
$\lvert\eta\rvert > 2{,}4$	0,0087	0,0137	0,1466

Tableau 2.5 – *Valeurs des aires effectives $\mathcal{E}_{\mathcal{A}}$ utilisées pour corriger la contribution de l'empilement aux isolations des photons vis-à-vis des autres particules.*

tés du « candidat » photon. Les variables exploitées sont :

— H/E_{SC} où H est l'énergie hadronique située dans un cône de $\Delta R < 0{,}15$ autour du photon et E_{SC} l'énergie contenue dans le *supercluster* ;
— $\sigma_{i\eta i\eta}$ l'étalement en η du dépôt d'énergie dans le ECAL ;
— $I_{\text{rel}}^{/CH}$ l'isolation relative vis-à-vis des hadrons chargés ;
— $I_{\text{rel}}^{/NH}$ l'isolation relative vis-à-vis des hadrons neutres ;
— $I_{\text{rel}}^{/\gamma}$ l'isolation relative vis-à-vis des photons.

Les coupures correspondant aux différents critères d'identification des photons ainsi que leurs efficacités d'identification et de réjection sont résumées dans le tableau 2.6.

Critère		Lâche		Moyen		Strict	
Région		Tonneau	Bouchon	Tonneau	Bouchon	Tonneau	Bouchon
Efficacité		90,08 %	90,65 %	80,29 %	80,08 %	70,24 %	70,13 %
Réjection		86,25 %	76,72 %	89,36 %	81,85 %	90,97 %	84,55 %
H/E_{SC}		0,045 96	0,0590	0,021 97	0,0326	0,021 48	0,0321
$\sigma_{i\eta i\eta}$		0,0106	0,0272	0,010 15	0,0272	0,009 96	0,0271
$I_{\text{rel}}^{/CH}$		1,694	2,089	1,141	1,051	0,65	0,517
$I_{\text{rel}}^{/NH}$		24,032	19,722	1,189	2,718	0,317	2,716
		$+0{,}015\,12\,p_{\text{T}}$	$+0{,}011\,p_{\text{T}}$	$+0{,}015\,12\,p_{\text{T}}$	$+0{,}0117\,p_{\text{T}}$	$+0{,}015\,12\,p_{\text{T}}$	$+0{,}0117\,p_{\text{T}}$
		$+2{,}259 p_{\text{T}}{}^2/10^5$	$+2{,}3 p_{\text{T}}{}^2/10^5$	$+2{,}259 p_{\text{T}}{}^2/10^5$	$+2{,}3 p_{\text{T}}{}^2/10^5$	$+2{,}259 p_{\text{T}}{}^2/10^5$	$+2{,}3 p_{\text{T}}{}^2/10^5$
$I_{\text{rel}}^{/\gamma}$		2,876	4,162	2,08	3,867	2,044	3,032
		$+0{,}004\,017\,p_{\text{T}}$	$+0{,}0037\,p_{\text{T}}$	$+0{,}004\,017\,p_{\text{T}}$	$+0{,}0037\,p_{\text{T}}$	$+0{,}004\,017\,p_{\text{T}}$	$+0{,}0037\,p_{\text{T}}$

Tableau 2.6 – *Valeurs maximales des observables considérées pour l'identification des photons selon le critère utilisé et la région du détecteur dans laquelle se trouve le candidat photon (tonneau pour $\lvert\eta\rvert < 1{,}479$, bouchon sinon).*

Reconstruction des électrons et des photons L'identification des électrons et des photons isolés se base sur les éléments d'identification du PF du trajectographe et du ECAL. Lors de leur traversée du trajectographe, les électrons émettent des photons par *bremsstrahlung*. De plus, les photons se convertissent en paires e^+e^-, ces électrons étant également sujets au *bremsstrahlung*, etc. Pour cette raison, électrons et photons isolés sont traités de manières similaires pour leur reconstruction.

Un candidat électron est défini lorsqu'une trace du trajectographe, extrapolée jusqu'au ECAL, est associée à un dépôt d'énergie, si ce dépôt n'est pas lui-même relié à trois autres traces ou plus. Les candidats photons isolés correspondent aux dépôts du ECAL avec une énergie transverse supérieure à 10 GeV n'étant pas associés à une trace. Pour tous ces candidats, la somme des énergies mesurées dans les cellules du HCAL se situant dans un cône de rayon ΔR inférieur à 0,15 dans le plan (η, ϕ) ne doit pas correspondre à plus de 10 % de l'énergie du dépôt du ECAL. Les traces identifiées comme celles de conversions de photons et les dépôts du ECAL associés sont de plus rattachées au candidat initial.

Les électrons et photons isolés sont alors obtenus en soumettant aux candidats les critères d'identification définis précédemment [147]. Les éléments d'identification du PF utilisés pour les reconstruire sont retirés dans la suite du processus de reconstruction de l'événement.

5.2.3 Hadrons et photons non isolés

Une fois les muons, électrons et photons isolés identifiés et reconstruits restent les hadrons et les photons non isolés issus de la formation des jets et de l'hadronisation. Ces particules sont généralement détectées comme des hadrons chargés ($\pi^\pm$, $K^\pm$, protons), des hadrons neutres (K_L, neutrons), des photons non isolés (désintégrations des π^0) et plus rarement comme des muons (désintégrations de hadrons lourds).

Dans la région d'acceptation du trajectographe ($|\eta| < 2{,}5$), les photons non isolés et les hadrons neutres sont reconstruits respectivement à partir des dépôts d'énergie dans les ECAL et HCAL non associés à une trace. Une priorité est donnée aux photons dans la mesure où 25 % de l'énergie des jets est portée par ces particules alors que seulement 3 % de l'énergie des jets est déposée dans le ECAL par les hadrons neutres. Au-delà de l'acceptation du trajectographe, il n'est pas possible de faire la distinction entre hadrons neutres et chargés. Près de 25 % de l'énergie des jets est ainsi déposée dans le ECAL et les agglomérats du ECAL se situant dans la même région qu'un agglomérat du HCAL sont considérés comme dus à la même gerbe hadronique, c'est-à-dire au même hadron. Les autres dépôts du ECAL sont considérés comme dus à des photons.

Les hadrons chargés sont identifiés à partir des agglomérats restant dans le HCAL, associés aux traces dans le trajectographe non utilisées pour l'identification des particules précédentes. Ces traces peuvent elles-mêmes être reliées à un agglomérat résiduel du ECAL. Pour chaque bloc du PF ainsi construit, l'énergie dans les calorimètres est comparée à la somme des impulsions des traces. Un excès d'énergie d'après les calorimètres est interprété comme la présence d'une particule neutre supplémentaire. Si la valeur de cet excès est inférieure à l'énergie contenue dans le ECAL et supérieure à 500 MeV, alors cette particule neutre supplémentaire est considérée comme étant un photon dont l'énergie est égale à l'excès. Sinon, l'énergie dans le ECAL est identifiée comme un photon et si la partie de l'excès contenue dans le HCAL est supérieure à 1 GeV, un hadron neutre est également identifié. Puis, à partir de l'énergie calorimétrique restante, chaque trace du bloc du PF donne un hadron chargé.

5.3 Reconstruction des jets

La phénoménologie des jets est présentée dans le chapitre 1. Les partons, c'est-à-dire les quarks et les gluons, ne peuvent pas être directement observés dans le détecteur. Leur signature expérimentale est un jet, c'est-à-dire un flux collimé de particules stables composé de hadrons, de leptons et de photons. Afin de pouvoir étudier le processus initial dont sont issus les partons à l'origine des jets observés, il est nécessaire de reconstruire ces jets en regroupant l'ensemble des particules les constituant. Des critères d'identification en tant que jet sont imposés à ces regroupements. Les jets ainsi obtenus sont alors calibrés en énergie selon la procédure détaillée au chapitre 3.

5.3.1 Regroupement des particules

Le regroupement des particules en jets est réalisé par l'algorithme anti-k_T [162]. Il s'agit d'un algorithme de recombinaison séquentielle, dans lequel chaque particule forme initialement un pseudo-jet d'une seule particule [163].

La distance d_{ij} entre deux jets i et j est définie par

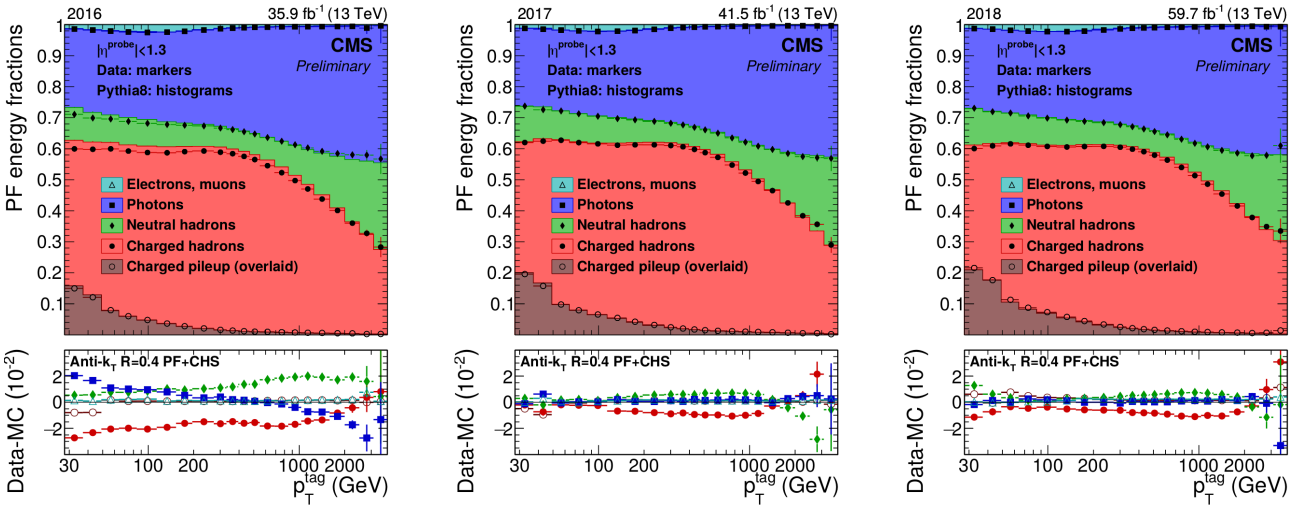
$$d_{ij} = \min\left(\frac{1}{p_{T_i}^2}, \frac{1}{p_{T_j}^2}\right) \frac{\Delta R_{ij}^2}{R^2}, \tag{2.17}$$

où

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2 \tag{2.18}$$

avec η_x la pseudo-rapidité, ϕ_x l'angle azimutal et p_{T_x} l'impulsion transverse du jet x et R un paramètre libre. De manière itérative et à l'aide de la distance d_{ij}, la paire composée des deux jets les plus proches fusionne tant que la distance entre eux est en-deçà d'une valeur seuil. Les jets fusionnés donnent la liste des jets de l'événement.

Sur la figure 2.25 sont illustrées les compositions des jets reconstruits lors des trois années du Run II. Les hadrons sont les composants majoritaires des jets, avec une contribution de 30 à 60 % au contenu total provenant des hadrons chargés et de 5 à 25 % des hadrons neutres. Pour des jets de bas p_T, la fraction de particules issues de l'empilement peut être de l'ordre de 20 %. Pour $p_T > 50$ GeV, elles représentent mois de 10 % du contenu des jets. La proportion de photons augmente avec l'impulsion transverse des jets, de 25 % à bas p_T à environ 40 % pour $p_T \gtrsim 1$ TeV. Les électrons et muons représentent moins de 5 % des particules. L'écart entre données réelles et simulées n'excède généralement pas 2 %.



(a) *Année 2016.* **(b)** *Année 2017.* **(c)** *Année 2018.*

Figure 2.25 – *Composition des jets reconstruits à l'aide de l'algorithme anti-k_T lors du Run II [164] en fonction de l'impulsion transverse du jet dans les données réelles (Data, histogrammes avec des points) et simulées (MC, histogrammes en couleurs). La partie* Charged pileup (overlaid) *en brun correspond à la fraction du jet retirée par la procédure CHS décrite dans le chapitre 3.*

5.3.2 Identification des jets

Les jets obtenus par regroupement des particules sont en réalité des « candidats » jets. À l'instar des particules individuelles, des critères d'identification leur sont appliqués afin de rejeter le bruit de fond et de s'assurer de la qualité des jets utilisés dans les analyses.

Ces critères reposent sur les caractéristiques des candidats jets tels que la fraction d'énergie provenant de leurs constituants neutres ou encore le nombre de ces constituants. Ils dépendent des années de prise de données et de la pseudo-rapidité du jet, c'est-à-dire de la région du détecteur dans laquelle il se trouve.

Les critères utilisés pour les années 2016, 2017, 2018, 2017-UL et 2018-UL, listés page 72, permettent d'obtenir une efficacité d'identification des jets supérieure à 99 % dans chacune des régions en η du détecteur. La dénomination « UL » signifie *Ultra-Legacy* et correspond à une réinterprétation des données récoltées une fois que la collaboration peut prendre plus de recul sur l'obtention de celles-ci. La réjection du bruit de fond est supérieure à 98 % pour $|\eta| \leqslant 3{,}0$ et supérieure à 36 % pour $|\eta| > 3{,}0$.

5.3.3 Saveur des jets

Il est impossible de connaître avec certitude le parton à l'origine d'un jet, mais ce dernier possède des propriétés caractéristiques dépendantes du parton, comme exposé dans le chapitre 1.

En utilisant ces propriétés, des algorithmes d'identification de la saveur des jets ont été mis au point par la collaboration CMS [165]. Les avancées récentes dans le domaine du *Deep Learning*, appliquées à l'identification des jets [166], ont permis l'amélioration de ces algorithmes. DEEPCSV [167] en est un exemple.

Les variables utilisées dans cet algorithme, décrites dans la référence [167], sont traitées par un réseau de neurones profond composé de quatre couches cachées de 100 nœuds connectés les uns aux autres. Le principe des réseaux de neurones est abordé plus en détail dans le chapitre 5. Ce réseau a été entraîné à l'aide des librairies KERAS [168] et TENSORFLOW [169] sur un ensemble d'événements simulés $t\bar{t}$, présentant de nombreux jets de quarks b, et multijet.

Les performances ainsi obtenues pour l'algorithme DEEPCSV sont comparées à d'autres algorithmes d'identification de la saveur des jets sur la figure 2.26. Les algorithmes CMVAV2 et DEEPCSV présentent les meilleures performances en termes d'identification des jets de quark b (*b-tagging*). Pour le traitement des jets de quark c, l'algorithme DEEPCSV propose les meilleures performances. Dans les analyses présentées dans les chapitres 3 et 4, c'est cet algorithme qui est utilisé afin d'identifier les jets issus de quarks c ou b.

5.4 Énergie transverse manquante

Des neutrinos peuvent être produits lors des collisions. Or, ces particules se propagent sans laisser de signal dans le détecteur, elles sont donc invisibles. Toutefois, lorsque de telles particules sont produites en association avec des particules détectées, leur présence peut être déduite du déséquilibre dans l'impulsion totale des particules de l'événement [170].

Lors d'une collision de protons, dans l'état initial, la composante longitudinale de l'impulsion est inconnue et l'impulsion totale dans le plan transverse est nulle comme discuté dans la section 3.3. Par conservation, l'impulsion totale dans le plan transverse est nulle dans l'état final. Les neutrinos n'étant pas détectés, leurs impulsions transverses sont manquantes dans le bilan de l'état final. L'observable définie afin de quantifier ce manque est l'énergie transverse manquante (MET, *Missing Transverse Energy*). Bien que son nom mentionne une énergie, il s'agit bien d'une impulsion.

La MET peut être déterminée à partir de l'algorithme de PF (PF MET) ou par l'algorithme PUPPI (PUPPI MET). Il s'agit de la MET brute, à laquelle la calibration en énergie des jets détaillée au chapitre 3 doit être propagée.

MET brute issue de l'algorithme de PF La somme des impulsions transverses des particules invisibles doit compenser celle des particules reconstruites, c'est-à-dire

$$\sum_{\substack{\text{toutes les}\\ \text{particules}}} \vec{p}_{\mathrm{T}} = \vec{0} \Leftrightarrow \sum_{\substack{\text{particules}\\ \text{invisibles}}} \vec{p}_{\mathrm{T}} + \sum_{\substack{\text{particules}\\ \text{reconstruites}}} \vec{p}_{\mathrm{T}} = \vec{0} \Leftrightarrow \sum_{\substack{\text{particules}\\ \text{invisibles}}} \vec{p}_{\mathrm{T}} = - \sum_{\substack{\text{particules}\\ \text{reconstruites}}} \vec{p}_{\mathrm{T}}\,. \tag{2.19}$$

La MET brute issue de l'algorithme de PF représente ainsi l'impulsion transverse totale des particules invisibles et est définie comme

$$\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}(\mathrm{PF}) = - \sum_{i \in \{\text{ particules }\}} \vec{p}_{\mathrm{T}}^{\,i}\,, \tag{2.20}$$

où les particules sont celles reconstruites par l'algorithme de PF. Cette définition, simple, est toutefois sensible aux particules issues de l'empilement. Afin de réduire l'effet de l'empilement, l'algorithme PUPPI a été développé.

Propriété du jet à identifier	$\lvert\eta\rvert \leqslant 2{,}4$	$2{,}4 < \lvert\eta\rvert \leqslant 2{,}7$	$2{,}7 < \lvert\eta\rvert \leqslant 3{,}0$	$3{,}0 < \lvert\eta\rvert$
Fraction d'énergie				
— hadronique neutre	$< 0{,}90$	$< 0{,}90$	$< 0{,}98$	
— électromagnétique neutre	$< 0{,}90$	$< 0{,}90$	$> 0{,}01$	$< 0{,}90$
— hadronique chargée	> 0			
— électromagnétique chargée	$< 0{,}90$			
— muonique	$< 0{,}8$	$< 0{,}8$		
Nombre de constituants	> 1	> 1		
— neutres			> 2	> 10
— chargés	> 0			

Tableau 2.7 – *Critères d'identification des jets à CMS pour l'analyse des données de 2016.*

Propriété du jet à identifier	$\lvert\eta\rvert \leqslant 2{,}4$	$2{,}4 < \lvert\eta\rvert \leqslant 2{,}7$	$2{,}7 < \lvert\eta\rvert \leqslant 3{,}0$	$3{,}0 < \lvert\eta\rvert$
Fraction d'énergie				
— hadronique neutre	$< 0{,}90$	$< 0{,}90$		$> 0{,}02$
— électromagnétique neutre	$< 0{,}90$	$< 0{,}90$	$< 0{,}99$ et $> 0{,}02$	$< 0{,}90$
— hadronique chargée	> 0			
— électromagnétique chargée	$< 0{,}8$			
— muonique	$< 0{,}8$	$< 0{,}8$		
Nombre de constituants	> 1	> 1		
— neutres			> 2	> 10
— chargés	> 0			

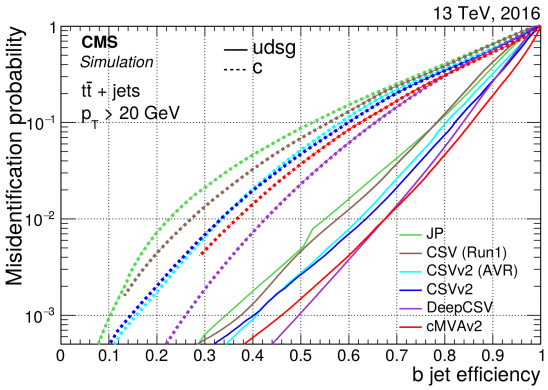
Tableau 2.8 – *Critères d'identification des jets à CMS pour l'analyse des données de 2017.*

Propriété du jet à identifier	$\lvert\eta\rvert \leqslant 2{,}6$	$2{,}6 < \lvert\eta\rvert \leqslant 2{,}7$	$2{,}7 < \lvert\eta\rvert \leqslant 3{,}0$	$3{,}0 < \lvert\eta\rvert \leqslant 5{,}0$
Fraction d'énergie				
— hadronique neutre	$< 0{,}90$	$< 0{,}90$		$> 0{,}2$
— électromagnétique neutre	$< 0{,}90$	$< 0{,}99$	$< 0{,}99$ et $> 0{,}02$	$< 0{,}90$
— hadronique chargée	> 0			
— électromagnétique chargée	$< 0{,}8$	$< 0{,}8$		
— muonique	$< 0{,}8$	$< 0{,}8$		
Nombre de constituants	> 1			
— neutres			> 2	> 10
— chargés	> 0	> 0		

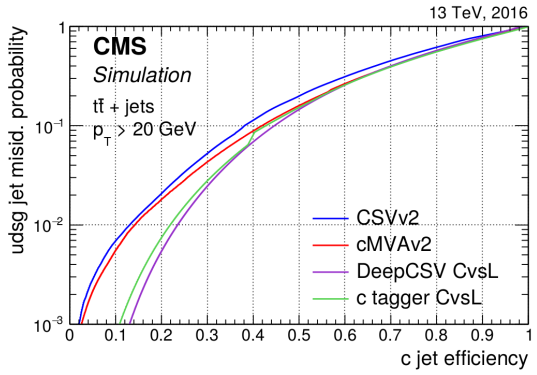
Tableau 2.9 – *Critères d'identification des jets à CMS pour l'analyse des données de 2018.*

Propriété du jet à identifier	$\lvert\eta\rvert \leqslant 2{,}6$	$2{,}6 < \lvert\eta\rvert \leqslant 2{,}7$	$2{,}7 < \lvert\eta\rvert \leqslant 3{,}0$	$3{,}0 < \lvert\eta\rvert \leqslant 5{,}0$
Fraction d'énergie				
— hadronique neutre	$< 0{,}90$	$< 0{,}90$		$> 0{,}2$
— électromagnétique neutre	$< 0{,}90$	$< 0{,}99$	$< 0{,}99$ et $> 0{,}01$	$< 0{,}90$
— hadronique chargée	> 0			
— électromagnétique chargée	$< 0{,}8$	$< 0{,}8$		
— muonique	$< 0{,}8$	$< 0{,}8$		
Nombre de constituants	> 1			
— neutres			> 1	> 10
— chargés	> 0	> 0		

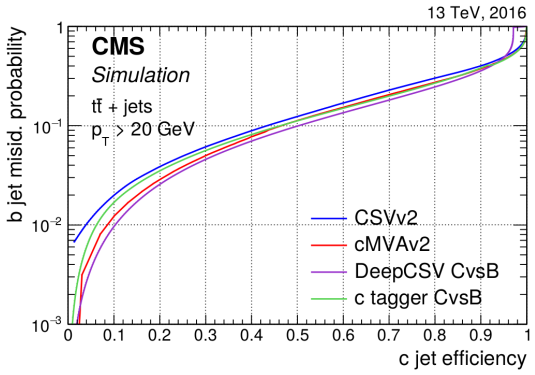
Tableau 2.10 – *Critères d'identification des jets à CMS pour l'analyse des données de 2017-UL et 2018-UL.*



(a) *Probabilité de mauvaise identification en tant que jet de quark b de jets de gluon ou quarks légers (traits pleins) ou de jets de quark c (pointillés) en fonction de l'efficacité d'identification des jets de quark b.*



(b) *Probabilité de mauvaise identification en tant que jet de quark c de jets de gluon ou quarks légers en fonction de l'efficacité d'identification des jets de quark c.*



(c) *Probabilité de mauvaise identification en tant que jet de quark b de jets de quark c en fonction de l'efficacité d'identification des jets de quark c.*

Figure 2.26 – *Comparaison des performances des algorithmes d'identification de la saveur des jets [167].*

MET brute issue de l'algorithme PUPPI La MET peut également être estimée par l'algorithme PUPPI (*PileUp Per Particle Identification*) [171]. La « PUPPIMET » obtenue est moins sensible à l'empilement (*pileup*) que la MET issue de l'algorithme de PF (PFMET). L'algorithme PUPPI exploite en effet des informations sur :

— l'environnement de chaque particule identifiée par l'algorithme de PF
— les propriétés de l'empilement dans l'événement ;
— les données issues du trajectographe ;

afin d'associer un poids w_i à chaque particule i, lié à la probabilité que celle-ci provienne de l'empilement au lieu du vertex primaire principal. Ce poids varie entre 0 pour des particules issues de l'empilement et 1 pour des particules provenant du vertex primaire principal. Plus de détails dans la détermination de w_i sont disponibles dans les références [170, 171].

La MET issue de l'algorithme PUPPI est définie comme

$$\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}(\mathrm{PUPPI}) = -\sum_{i\in\{\text{ particules }\}} w_i\,\vec{p}_{\mathrm{T}}^{\,i}\,. \tag{2.21}$$

Propagation de la calibration en énergie des jets à $E_{\mathrm{T}}^{\mathrm{miss}}$ La calibration en énergie des jets, abordée dans le chapitre 3, doit être propagée à $E_{\mathrm{T}}^{\mathrm{miss}}$ afin de conserver une description cohérente des événements. Il s'agit de la correction dite de « type I » [172], réalisée selon

$$\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}(\text{Type I}) = \vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}(\text{brute}) - \sum_{i\in\{\text{ jets }\}} \left(\vec{p}_{\mathrm{T\,corr}}^{\,i}) - \vec{p}_{\mathrm{T\,reco}}^{\,i}\right) \tag{2.22}$$

où « reco » correspond aux observables avant calibration et « corr » après.

5.5 Taus hadroniques

Lors d'une désintégration hadronique d'un lepton τ, une paire de quarks est émise. Il s'ensuit donc un processus d'hadronisation, phénomène à l'origine de la formation des jets. Du lepton τ résulte alors un ensemble de hadrons, comme illustré sur la figure 2.27. Ces hadrons, en général trois ou moins, sont éventuellement accompagnés de particules neutres, principalement des π^0. Ces derniers se désintégrant majoritairement en deux photons. L'ensemble de ces particule forme un « tau hadronique », noté τ_h, et est initialement identifié comme un jet.

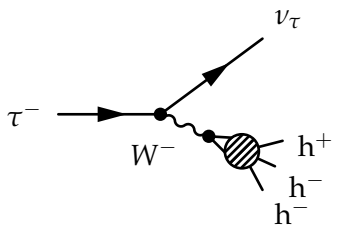


Figure 2.27 – *Diagramme de Feynman de désintégration hadronique d'un τ^-.*

5.5.1 Obtention de candidats

L'identification des τ_h est réalisée par l'algorithme *Hadrons Plus Strips* (HPS) [173, 174] à partir des jets reconstruits par l'algorithme de PF contenant au moins un hadron chargé et vérifiant $p_T > 14\,\text{GeV}$ et $|\eta| < 2{,}5$. Les hadrons chargés contenus dans le jet initial tels que $p_T > 0{,}5\,\text{GeV}$ et de paramètre d'impact transverse $d_{xy} < 0{,}1\,\text{cm}$ vis-à-vis du vertex primaire principal sont utilisés pour former des candidats τ_h.

Des π^0 sont souvent présents dans les désintégrations hadroniques des τ. Or, ces derniers se désintègrent à leur tour selon $\pi^0 \to \gamma\gamma$. Les photons sont eux-mêmes susceptibles de se convertir en paires e^+e^-. Afin d'estimer les π^0 présents, les photons et les électrons contenus dans le jet initial sont regroupés en bandes (*strips*). La construction d'une bande est un procédé itératif :

1. Une bande est créée à partir de l'électron ou du photon (e/γ) de plus haut p_T contenu dans le jet initial et n'ayant pas déjà été associé à une bande. La position de cette particule dans le plan (η, ϕ), ainsi que son p_T, sont associés à la bande.
2. L'électron ou photon de plus haut p_T restant est ajouté à la bande s'il est situé à une distance par rapport à la bande dans le plan (η, ϕ) telle que

$$\Delta\eta < f\left({p_T}^{(e/\gamma)}\right) + f\left({p_T}^{\text{bande}}\right), \quad f(p_T) = 0{,}20(p_T[\text{GeV}])^{-0{,}66} \tag{2.23}$$

$$\Delta\phi < g\left({p_T}^{(e/\gamma)}\right) + g\left({p_T}^{\text{bande}}\right), \quad g(p_T) = 0{,}35(p_T[\text{GeV}])^{-0{,}71} \tag{2.24}$$

   avec ${p_T}^{(e/\gamma)}$ l'impulsion transverse de l'électron ou du photon à ajouter à la bande et ${p_T}^{\text{bande}}$ l'impulsion transverse associée à la bande avant ajout de l'électron ou du photon.
   Si l'ajout se fait, la bande est mise à jour selon

$${p_T}^{\text{bande}} = \sum_{e/\gamma \in \text{bande}} {p_T}^{(e/\gamma)}, \tag{2.25}$$

$$\eta^{\text{bande}} = \frac{1}{{p_T}^{\text{bande}}} \sum_{e/\gamma \in \text{bande}} {p_T}^{(e/\gamma)} \eta_{(e/\gamma)}, \tag{2.26}$$

$$\phi^{\text{bande}} = \frac{1}{{p_T}^{\text{bande}}} \sum_{e/\gamma \in \text{bande}} {p_T}^{(e/\gamma)} \phi_{(e/\gamma)}, \tag{2.27}$$

   ce qui rend la bande dynamique lors de sa construction. Les dimensions de la bande sont limitées à $0{,}05 < \Delta\eta < 0{,}15$ et $0{,}05 < \Delta\phi < 0{,}3$.
3. L'étape précédente est répétée jusqu'à ce qu'une limite de taille de la bande soit atteinte ou qu'il ne reste plus d'électron ni de photon tels que $p_T > 0{,}5\,\text{GeV}$ dans la zone de la bande.
4. Les éléments associés à la bande sont retirés de la liste des électrons et photons en attente d'association à une bande.
5. Le procédé reprend à l'étape 1.

Toute bande vérifiant $p_T > 2{,}5\,\text{GeV}$ est considérée comme un candidat π^0.

Des candidats τ_h compatibles avec un des modes de désintégration hadronique du τ sont ainsi formés à partir de toutes les combinaisons possibles de hadrons chargés et de candidats π^0.

5.5.2 Modes de désintégration

Les modes de désintégration (DM, *Decay Modes*) principaux et physiquement possibles sont listés dans le tableau 2.11. Une liste exhaustive est disponible dans la référence [9]. À chaque DM correspond une valeur afin de le désigner, définie comme

$$\mathrm{DM} = 5 \times (N_{\mathrm{h}^\pm} - 1) + N_{\pi^0} \tag{2.28}$$

où $N_{\mathrm{h}^\pm}$ est le nombre de hadrons chargés et N_{π^0} le nombre de π^0 contenus dans le τ_h. Lorsqu'un des hadrons chargés n'est pas reconstruit, il est possible d'obtenir les DM 5, 6 ou 7. Ces cas de figure sont largement contaminés par le bruit de fond « QCD multijet », ils sont donc généralement rejetés dans les analyses.

Code	Mode de désintégration	$\mathcal{BR}$ (%)
0	$\tau^- \to \mathrm{h}^-\bar{\nu}_\tau$	11,51
1	$\tau^- \to \mathrm{h}^-\pi^0\bar{\nu}_\tau$	25,93
2	$\tau^- \to \mathrm{h}^-\pi^0\pi^0\bar{\nu}_\tau$	9,48
10	$\tau^- \to \mathrm{h}^-\mathrm{h}^-\mathrm{h}^+\bar{\nu}_\tau$	9,80
11	$\tau^- \to \mathrm{h}^-\mathrm{h}^-\mathrm{h}^+\pi^0\bar{\nu}_\tau$	4,76

Tableau 2.11 – *Modes de désintégration du τ considérés. Le mode de désintégration d'un τ^- correspondant au DM, ainsi que le rapport de branchement $\tau^- \to \tau_\mathrm{h}^-$ correspondant [9] sont également donnés.*

Certains DM présentent des contraintes supplémentaires sur la masse du τ_h :

$$\begin{aligned}
\text{DM 1 :} \quad & 0{,}3\,\mathrm{GeV} < m_{\tau_\mathrm{h}} < 1{,}3\sqrt{\frac{p_\mathrm{T}[\mathrm{GeV}]}{100}}\,\mathrm{GeV}\,, \\
\text{DM 2 :} \quad & 0{,}4\,\mathrm{GeV} < m_{\tau_\mathrm{h}} < 1{,}2\sqrt{\frac{p_\mathrm{T}[\mathrm{GeV}]}{100}}\,\mathrm{GeV}\,, \\
\text{DM 10 et 11 :} \quad & 0{,}8\,\mathrm{GeV} < m_{\tau_\mathrm{h}} < 1{,}5\,\mathrm{GeV}\,,
\end{aligned}$$

et, dans le cas du DM 10, les traces des hadrons chargés doivent provenir du même vertex dans la limite de $\Delta z < 0{,}4\,\mathrm{cm}$.

5.5.3 Sélection d'un candidat

Il est possible d'obtenir plusieurs candidats τ_h au sein d'un même jet. Des critères de qualité sur les candidats leur sont alors imposés.

Par conservation, la somme des charges électriques des hadrons contenus dans le candidat τ_h doit valoir ± 1. Ces hadrons chargés doivent de plus être contenus dans le cône dit « de signal », centré sur la direction de l'impulsion du candidat τ_h et dont la largeur est définie et contrainte selon

$$\Delta R_\mathrm{sig} = \frac{3\,\mathrm{GeV}}{p_\mathrm{T}^{\tau_\mathrm{h}}}\,, \quad 0{,}05 < \Delta R_\mathrm{sig} < 0{,}1\,. \tag{2.29}$$

Les centres des bandes du candidat τ_h doivent également se situer dans ce cône. S'il reste plusieurs candidats à ce stade, celui de plus haut p_T est retenu. Il reste donc au plus un τ_h par jet.

5.5.4 Mauvaises reconstructions

Un τ_h peut être reconstruit à tort à partir d'autres jets, d'électrons ou de muons. Afin de réduire la quantité de mauvais τ_h (faux τ_h), un réseau de neurones profond convolutionnel (DNN) [175] a été développé à CMS. Il s'agit de l'algorithme DeepTau [176] qui fournit les discriminateurs `deepTau vs jet`, `deepTau anti-electron` et `deepTau anti-muon` utilisés dans cette analyse.

Les efficacités d'identification de chacun des points de fonctionnement existants sont données dans le tableau 2.12. Les taux de mauvaise identification de jet, électron ou muon comme τ_h, c'est-à-dire les faux positifs, dépendent de la nature des événements sur lesquels ces discriminateurs sont appliqués et se situent entre 10^{-4} et 10^{-2}.

Discriminateur	*VVTight*	*VTight*	*Tight*	*Medium*	*Loose*	*VLoose*	*VVLoose*	*VVVLoose*
`vs jet`	40	50	60	70	80	90	95	98
`anti-electron`	60	70	80	90	95	98	99	99,5
`anti-muon`	-	-	99,5	99,8	99,9	99,95	-	-

Tableau 2.12 – *Efficacités d'identification en % de l'algorithme* DeepTau *pour chacun des points de fonctionnement disponibles [176, 177].*

6 Simulation d'événements

La simulation d'événements permet de comparer les résultats expérimentaux aux prédictions théoriques. Elle se déroule en plusieurs étapes.

Premièrement, les processus physiques prédits par le modèle théorique à tester sont simulés. La nature probabiliste de ces processus mène à utiliser un générateur d'événements Monte-Carlo. Cette étape est détaillée en section 6.1. Les particules issues des collisions simulées sont alors obtenues ainsi que tout l'historique de leurs formations à partir des particules initiales entrant en collision.

Deuxièmement, la propagation de ces particules dans le détecteur, leurs interactions avec ses différents composants et les signaux qui en résultent sont également simulés. Cette simulation du détecteur est l'objet de la section 6.2. Cette méthode permet d'obtenir une estimation des signaux devant être observés avec le détecteur si le modèle testé correspond à celui décrivant effectivement l'Univers.

6.1 Génération d'événements

La description analytique de l'interaction entre les constituants des protons lors des collisions est réalisée grâce à la théorie des perturbations. À l'aide des règles de Feynman, il est possible de calculer l'« élément de matrice » permettant de décrire le passage d'un état initial à un état final. Les événements sont alors générés à un ordre perturbatif donné à l'aide de MadGraph [161] ou pythia [178, 179] par exemple. La plupart des processus sont ainsi disponibles à l'ordre dominant (LO, *Leading Order*). Dans certains cas, les ordres supérieurs (NLO, *Next-to-Leading Order*, NNLO, *Next-to-Next-to-Leading Order*, etc.) sont également disponibles grâce à des générateurs NLO tels que powheg [180] et mc@nlo [181].

Le processus initial de la collision duquel sont issues de nouvelles particules est ainsi simulé. Cependant, les particules possédant une charge de couleur comme les quarks ne peuvent subsister seules à cause du confinement de couleur, abordé dans le chapitre 1. Des étapes de génération supplémentaires sont alors nécessaires afin de décrire l'évolution ultérieure de ces particules. Il s'agit de la formation de la gerbe partonique et de l'hadronisation, détaillées dans le chapitre 1. Des générateurs comme pythia [178, 179] et herwig [182] permettent de simuler ces étapes ultérieures.

De plus, lors d'une collision de protons et plus généralement de hadrons, plusieurs interactions entre les constituants de ces derniers peuvent survenir, comme illustré sur la figure 2.28. L'interaction emportant la plus grande fraction de l'énergie des hadrons est l'« événement dur ». Les autres interactions constituent l'« événement sous-jacent ».

6.2 Simulation du détecteur

Une fois simulée la physique de l'événement, indépendante du détecteur utilisé, il faut modéliser la réponse du détecteur. La propagation des particules est alors simulée. Certaines d'entre elles se désintègrent dans le volume du détecteur, ce qui est également pris en compte. La modélisation du détecteur donne de plus les déviations des particules dues à la traversée de la matière le constituant [153, 154] ainsi que les interactions propres à la détection de ces particules comme les traces ou les gerbes électromagnétiques et hadroniques. Enfin, la modélisation de l'électronique et du système de déclenchement donnent une simulation de la réponse complète du détecteur.

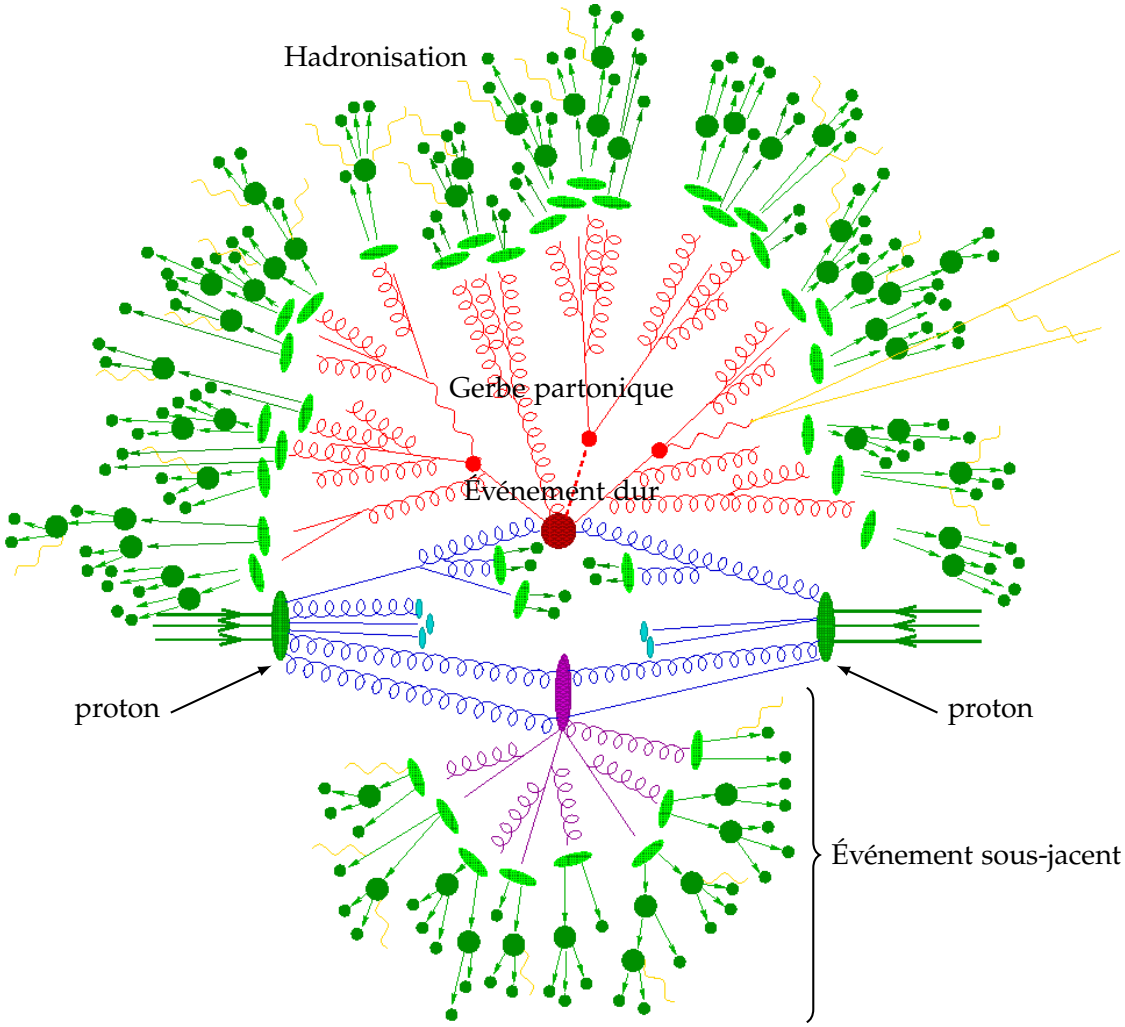


Figure 2.28 – *Représentation d'une collision de protons [183]. Les interactions initiales entre les constituants des deux protons sont représentées en bleu. Celle de plus haute énergie est l'événement dur. Les particules en étant issues sont représentées en rouge. Le reste des interactions entre les constituants des protons est l'événement sous-jacent, dont sont issues les particules en violet. Les gerbes partoniques formées dans les événements dur et sous-jacent subissent la hadronisation dont sont issues les particules en vert observées dans le détecteur. Des radiations de photons, en jaune, peuvent survenir.*

Cette simulation du détecteur est basée sur le logiciel GEANT4 [184-186]. La prise en compte de la totalité des éléments du détecteur (câblage interne, système de refroidissement, éléments de structure, etc.) permet d'obtenir des résultats fidèles à la réalité. Les signaux simulés ainsi obtenus sont alors traités, comme dans le cas des données réelles, par l'algorithme de *Particle Flow* permettant de reconstruire l'événement physique, puis par les algorithmes de reconstruction des objets de haut niveau, comme exposé dans la section 5.

6.3 Corrections apportées aux simulations

Des écarts résiduels entre les données simulées et la réalité qu'elles doivent décrire existent et sont mesurés par des analyses dédiées. Ces analyses fournissent alors des corrections à appliquer aux simulations. Il existe également des données hybrides, dont les événements sont construits à partir de données réelles et simulées. Il s'agit des données dites « encapsulées » (*embedded*) présentées dans le chapitre 4, pour lesquelles certaines de ces corrections peuvent être différentes de celles à appliquer aux données 100 % simulées.

6.3.1 Pondérations dues aux collisions

Pondération de l'empilement (*Pileup reweighting*) Les données simulées sont générées avec un réglage donné de luminosité instantanée, relié à la quantité d'empilement obtenu. Or, la production de ces jeux de données est souvent faite avant la mesure de ces observables dans les données réelles. Afin de corriger la différence sur le profil d'empilement obtenu, un poids est appliqué aux événements simulés afin que ce profil soit cohérent avec celui des données réelles.

Pondération du pré-déclenchement En 2016 et 2017, le niveau L1 du système de déclenchement de CMS présentait un défaut. Dans la partie à haute η du ECAL, des objets physiques responsables du déclenchement du L1 étaient associés à l'événement précédent, c'est-à-dire celui avant déclenchement. Seul un événement sur trois consécutifs pouvant être enregistré, l'efficacité de la prise de données est ainsi moindre qu'attendue. Les événements avec des jets de hautes valeurs de η sont particulièrement touchés par ce phénomène. Une pondération est alors appliquée afin de corriger cet effet. Le poids correspondant dépend de la topologie de l'événement et varie généralement entre 0,95 et 1,0.

6.3.2 Reconstruction et identification des particules individuelles

Efficacité d'identification et isolation des muons et des électrons (*muon/electron ID/iso efficiency*) Des facteurs correctifs sont déterminés par le groupe $H\tau\tau$. Ils sont appliqués individuellement à chaque muon et électron utilisé dans les analyses et dépendent de l'année ainsi que de la nature des données, simulées ou encapsulées.

Efficacité du trajectographe (*tracking efficiency*) L'efficacité de la reconstruction des traces des particules n'est pas la même selon la nature des données, réelles ou simulées, comme l'ont constaté les *POG*s *EGamma* (électrons et photons) et *tracking* dans le cas des électrons et des muons. Des facteurs d'échelle, que ces *POG*s fournissent, sont appliqués afin de corriger cet effet.

Énergie des électrons (*electron energy scale*) La mesure de l'énergie des électrons dans les données simulées est corrigée selon les recommandations du *POG EGamma* [187], résumées dans le tableau 2.13.

Région du détecteur	2016	2017	2018
Tonneau ($\lvert\eta\rvert < 1{,}479$)	$-0{,}24 \pm 0{,}5$	$-0{,}07 \pm 0{,}5$	$-0{,}33 \pm 0{,}5$
Bouchons ($\lvert\eta\rvert > 1{,}479$)	$-0{,}70 \pm 1{,}25$	$-1{,}13 \pm 1{,}25$	$-0{,}56 \pm 1{,}25$

Tableau 2.13 – *Corrections à l'énergie des électrons en % avec incertitude pour les trois années du Run II.*

6.3.3 Jets

Énergie des jets (*Jet Energy Calibration*) La mesure de l'énergie des jets et la résolution sur celle-ci sont corrigées. Le chapitre 3 est dédié à ces corrections, elle y sont détaillées.

Correction de l'efficacité du *b-tagging* Le *POG BTV* fournit des facteurs correctifs *SF* à l'efficacité du *b-tagging* en fonction de la saveur du jet au niveau généré, des propriétés cinématiques du jet et du point de fonctionnement du discriminateur de *b-tagging* utilisé [167, 188]. Le taux de mauvaise identification est également corrigé.

Pour cela, une méthode de promotion-relégation (*promote-demote*) est utilisée. Une fraction des jets tagués b, c'est-à-dire identifiés comme issus d'un quark b, est relégué à l'état de jet non tagué b et une fraction des jets non tagués b est promue à l'état de jet tagué b. Un jet peut être promu si

son facteur correctif SF est supérieur à 1. Sinon, il peut être relégué. La probabilité d'être promu ou relégué s'exprime

$$P(\text{promu}) = \frac{SF-1}{\frac{1}{\epsilon}-1}, SF > 1\,, \quad P(\text{relégué}) = 1 - SF, SF < 1\,, \tag{2.30}$$

avec ϵ l'efficacité du *b-tagging*.

6.3.4 Taus hadroniques

Efficacité d'identification et isolation des τ_h (τ_h *ID/iso scale factors*) L'efficacité d'identification des τ_h n'est pas la même dans les données réelles et simulées [189]. Des facteurs correctifs sont déterminés par le *POG tau* à partir d'événements Drell-Yan dans le canal $\mu\tau_h$, c'est-à-dire lorsqu'un des leptons τ issus du Z se désintègre en muon et l'autre en τ_h. Ils sont de plus fournis séparément pour les données simulées et encapsulées. De même, la mesure de l'isolation des τ_h est ajustée dans les simulations.

Taux de mauvaise identification $\mu \to \tau_h$ ($\mu \to \tau_h$ *fake rate*) L'efficacité de la réjection des faux τ_h issus de muons diffère entre données réelles et simulées [189]. Un facteur d'échelle à appliquer aux simulations est fourni par le *POG tau* en fonction de la pseudo-rapidité du faux τ_h comme exposé dans le tableau 2.14. Les points de fonctionnement donnés sont ceux utilisés dans le chapitre 4.

Région du détecteur	WP	2016	2017	2018
$(0 < \lvert\eta\rvert < 0{,}4)$	*VLoose*	$1{,}25 \pm 0{,}08$	$1{,}12 \pm 0{,}09$	$1{,}00 \pm 0{,}08$
	Tight	$0{,}38 \pm 0{,}12$	$0{,}92 \pm 0{,}17$	$0{,}81 \pm 0{,}15$
$(0{,}4 < \lvert\eta\rvert < 0{,}8)$	*VLoose*	$0{,}96 \pm 0{,}15$	$0{,}76 \pm 0{,}12$	$1{,}08 \pm 0{,}14$
	Tight	$0{,}72 \pm 0{,}30$	$0{,}79 \pm 0{,}25$	$1{,}02 \pm 0{,}35$
$(0{,}8 < \lvert\eta\rvert < 1{,}2)$	*VLoose*	$1{,}29 \pm 0{,}11$	$0{,}99 \pm 0{,}10$	$1{,}04 \pm 0{,}10$
	Tight	$1{,}34 \pm 0{,}27$	$0{,}67 \pm 0{,}26$	$0{,}92 \pm 0{,}22$
$(1{,}2 < \lvert\eta\rvert < 1{,}7)$	*VLoose*	$0{,}92 \pm 0{,}20$	$0{,}75 \pm 0{,}14$	$0{,}95 \pm 0{,}16$
	Tight	$1{,}03 \pm 0{,}65$	$1{,}07 \pm 0{,}45$	$0{,}83 \pm 0{,}47$
$(1{,}7 < \lvert\eta\rvert < 2{,}3)$	*VLoose*	$5{,}01 \pm 0{,}38$	$4{,}44 \pm 0{,}30$	$5{,}58 \pm 0{,}40$
	Tight	$5{,}05 \pm 0{,}88$	$4{,}08 \pm 0{,}85$	$4{,}52 \pm 0{,}92$

Tableau 2.14 – *Corrections à appliquer au taux d'identification des muons comme des τ_h en % avec incertitude pour les trois années du Run II.*

Taux de mauvaise identification $e \to \tau_h$ ($e \to \tau_h$ *fake rate*) À l'instar des muons, un facteur d'échelle à appliquer aux simulations est fourni par le *POG tau* en fonction de la pseudo-rapidité du faux τ_h comme indiqué dans le tableau 2.15. Les points de fonctionnement donnés sont ceux utilisés dans le chapitre 4.

Région du détecteur	WP	2016	2017	2018
Tonneau $(\lvert\eta\rvert < 1{,}479)$	*VVLoose*	$1{,}38 \pm 0{,}08$	$1{,}11 \pm 0{,}09$	$0{,}91 \pm 0{,}06$
	Tight	$1{,}22 \pm 0{,}38$	$1{,}22 \pm 0{,}32$	$1{,}47 \pm 0{,}27$
Bouchons $(\lvert\eta\rvert > 1{,}479)$	*VVLoose*	$1{,}29 \pm 0{,}08$	$1{,}03 \pm 0{,}09$	$0{,}91 \pm 0{,}07$
	Tight	$1{,}47 \pm 0{,}32$	$0{,}93 \pm 0{,}38$	$0{,}66 \pm 0{,}20$

Tableau 2.15 – *Corrections à appliquer au taux d'identification des électrons comme des τ_h en % avec incertitude pour les trois années du Run II.*

Énergie des τ_h (τ_h *energy scale*) L'énergie mesurée des τ_h peut différer entre les τ_h réels et simulés, ainsi que selon le DM du τ_h [189]. Le *POG tau* fournit les corrections à appliquer aux τ_h simulés, elles sont données dans le tableau 2.16. Ces corrections sont obtenues à partir d'événements du canal $\mu\tau_h$, par exploitation de la masse du τ_h et de la masse visible du système $\mu\tau_h$. Elles sont dépendantes de l'année, du DM et du type de données, simulées ou encapsulées.

DM	2016	2017	2018
0	$-0{,}6 \pm 1{,}0$	$0{,}7 \pm 0{,}8$	$-1{,}3 \pm 1{,}1$
1	$-0{,}5 \pm 0{,}9$	$-0{,}2 \pm 0{,}8$	$-0{,}5 \pm 0{,}9$
10	$0{,}0 \pm 1{,}1$	$0{,}1 \pm 0{,}9$	$-1{,}2 \pm 0{,}8$
11	$0{,}1 \pm 1{,}0$	$-0{,}5 \pm 1{,}6$	$0{,}1 \pm 1{,}0$

(a) *Pour les données simulées.*

DM	2016	2017	2018
0	$-0{,}2 \pm 0{,}5$	$0{,}0 \pm 0{,}4$	$-0{,}3 \pm 0{,}4$
1	$-0{,}2 \pm 0{,}3$	$-1{,}2 \pm 0{,}5$	$-0{,}6 \pm 0{,}4$
10	$-1{,}3 \pm 0{,}5$	$-0{,}8 \pm 0{,}5$	$-0{,}7 \pm 0{,}3$
11	$-1{,}3 \pm 0{,}5$	$-0{,}8 \pm 0{,}5$	$-0{,}7 \pm 0{,}3$

(b) *Pour les données encapsulées.*

Tableau 2.16 – *Corrections à appliquer à l'énergie des taus hadroniques en % avec incertitude pour les trois années du Run II.*

Énergie des muons identifiés comme τ_h ($\mu \to \tau_h$ *energy scale*) À l'instar des vrais τ_h discutés dans le paragraphe précédent, l'énergie mesurée des faux τ_h issus de muons peut différer entre les données réelles et simulées. Dans ce cas, le quadrivecteur du faux τ_h est directement corrigé selon le DM du τ_h identifié. Cette correction, généralement inférieure au pourcent, est appliquée uniquement aux DMs 0 et 1 et pour des τ_h correspondant au niveau généré à un muon. La quantité de muons identifiés comme des τ_h avec un DM plus élevé, en particulier les DMs 10 et 11, est négligeable, c'est pourquoi aucune correction n'est prévu dans ce cas. Les valeurs des corrections à appliquer aux données simulées sont données dans le tableau 2.17a.

Énergie des électrons identifiés comme τ_h ($e \to \tau_h$ *energy scale*) La correction correspondante est similaire au cas des muons, mais peut être de l'ordre de 5 % selon le DM et la pseudo-rapidité. Les valeurs des corrections à appliquer aux données simulées sont données dans les tableaux 2.17b et 2.17c.

DM	2016	2017	2018
0	0,0	−0,2	−0,2
1	−0,5	−0,8	−1,0

(a) *Muons.*

DM	2016	2017	2018
0	0,7	0,9	1,4
1	3,4	1,2	1,9

(b) *Électrons du tonneau ($|\eta| <$ 1,479).*

DM	2016	2017	2018
0	−0,4	−2,6	−3,1
1	5,0	1,5	−1,5

(c) *Électrons des bouchons ($|\eta| >$ 1,479).*

Tableau 2.17 – *Corrections à appliquer à l'énergie des électrons et des muons identifiés comme des taus hadroniques en % avec incertitude pour les trois années du Run II.*

6.3.5 Énergie transverse manquante

Propagation des corrections à E_T^{miss} À l'instar de la calibration en énergie des jets, la correction en énergie des différentes particules et des jets doit être propagée à E_T^{miss} afin de conserver une description cohérente des événements. Cette propagation est faite selon

$$\vec{E}_T^{miss}(\text{corr.}) = \vec{E}_T^{miss}(\text{non corr.}) - \sum_{i \in \{\text{ particules }\}} \left(\vec{p}_T{}^{i}(\text{corr.}) - \vec{p}_T{}^{i}(\text{non corr.}) \right) \tag{2.31}$$

où « non corr. » correspond aux observables avant correction et « corr. » après.

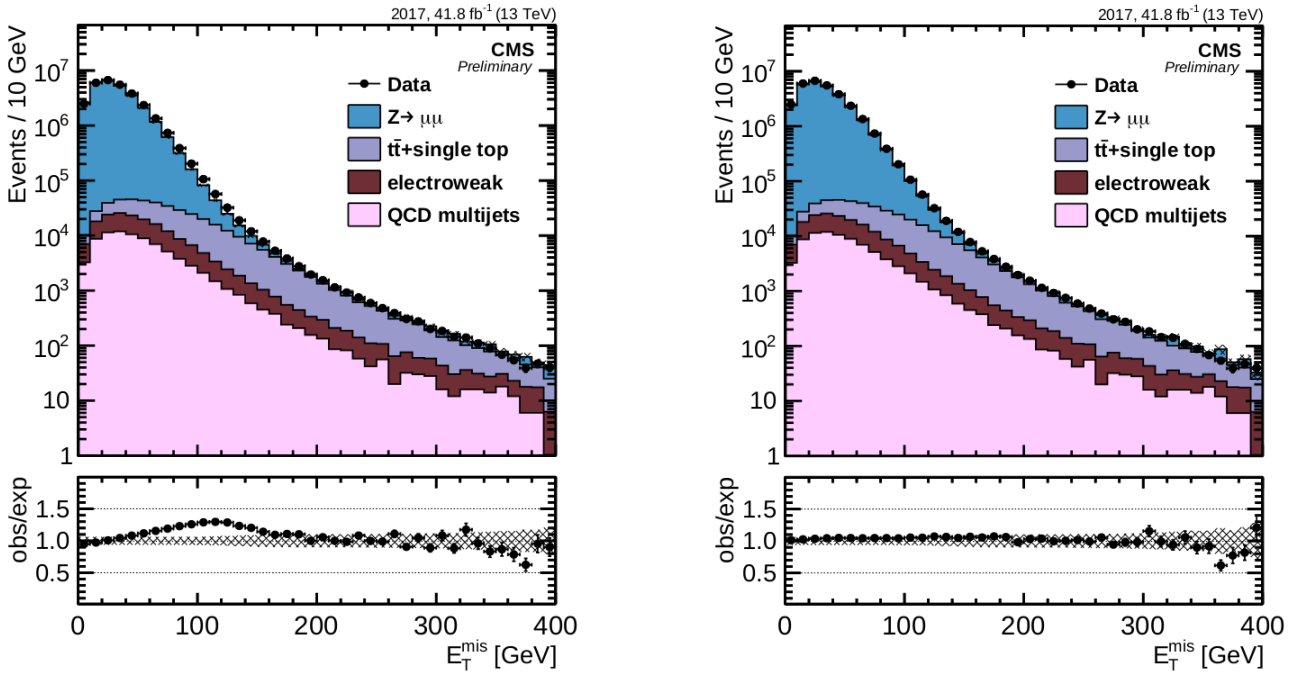
Recul de E_T^{miss} (*MET recoil corrections*) La modélisation de E_T^{miss} dans certains jeux de données simulées (production du boson de Higgs, Drell-Yan (boson Z) et W + jets) ne correspond pas aux observations dans les données réelles. Des corrections sur $\vec{U}$, défini comme la différence entre E_T^{miss} et la somme des impulsions des neutrinos provenant de la désintégration du boson de Higgs, Z ou W, c'est-à-dire

$$\vec{U} = \vec{E}_T^{miss} - \sum_{\nu_i \leftarrow h,Z,W} \vec{p}_T^{\,\nu_i}, \tag{2.32}$$

sont appliquées pour corriger cet effet.

Les composantes colinéaire U_1 et orthogonale U_2 du vecteur $\vec{U}$ à l'impulsion du boson sont déterminées dans des événements $Z \to \mu\mu$ dans lesquels il n'y a pas de neutrino provenant de la désintégration du Z, ce qui permet de mesurer précisément son impulsion. L'écart à zéro de U_1 ainsi que la résolution sur U_1 et U_2 sont déterminés dans les données réelles et simulées. Les données simulées sont alors corrigées afin de faire correspondre en moyenne ces valeurs à celles observées dans les données réelles. Ces moyennes sont déterminées sur des intervalles d'impulsion du Z ($[0, 10[$, $[10, 20[$, $[20, 30[$, $[30, 50[$ et > 50 GeV) et du nombre de jets ($N_{jets} \in \{0, 1, \geqslant 2\}$).

L'effet de cette correction sur une sélection d'événements $Z \to \mu\mu$ en 2017 est présenté sur la figure 2.29. Les distributions observées (données réelles) et modélisées (données simulées) de E_T^{miss} y sont tracées. L'accord entre observation et modélisation, décrit par le rapport (obs/exp), est sensiblement amélioré par cette correction.



(a) *Distribution de E_T^{miss} sans correction.*

(b) *Distribution de E_T^{miss} avec correction.*

Figure 2.29 – *Effet de la correction de recul de E_T^{miss} sur une sélection d'événements $Z \to \mu\mu$ en 2017 [190].*

7 Conclusion

Le détecteur CMS, exploité par la collaboration du même nom, est le dispositif expérimental utilisé dans cette thèse. Il s'agit d'une des quatre grandes expériences installées au LHC, le plus grand des collisionneurs de hadrons au monde à ce jour, qui se trouve au CERN.

L'acronyme « CERN » signifie Conseil Européen pour la Recherche Nucléaire, mais le CERN correspond aujourd'hui à l'organisation européenne pour la recherche nucléaire. Basé au Nord-Ouest de Genève, il s'étend des deux côtés de la frontière franco-suisse et comporte de nombreuses installations expérimentales de physique nucléaire et des particules, entre autres. Des innovations majeures sont issues des recherches menées CERN, comme les écrans tactiles, le Web et la hadronthérapie.

Le LHC, le Grand Collisionneur de Hadrons, est un accélérateur circulaire de 27 km de circonférence. Il permet d'atteindre des énergies de collision dans le centre de masse de 13 TeV pour des protons et il est prévu de passer à 14 TeV dès le Run III prévu en 2022. Les expériences ALICE, ATLAS, CMS, LHCb, LHCf, TOTEM et MoEDAL y sont installées.

La composition et le fonctionnement du détecteur CMS ont été détaillées. Ce détecteur est de forme cylindrique et possède une structure en couches concentriques, chacune étant un sous-détecteur ayant un rôle précis. Tout d'abord, le trajectographe permet d'obtenir les trajectoires des particules chargées. Puis, le calorimètre électromagnétique stoppe les électrons et les photons en mesurant leurs énergies. Le calorimètre hadronique fait de même avec les hadrons. La couche suivante, le solénoïde, ne détecte pas les particules mais produit un champ magnétique de 4 T afin de courber les trajectoires des particules chargées. Enfin, la couche externe est constituée d'une culasse d'acier pour le retour du champ magnétique dans laquelle sont insérées des chambres à muons, détectant le passage de ces particules.

Les 40 millions d'événements par seconde du LHC donnent une quantité de données bien trop importante pour toutes les stocker. Un système de déclenchement à deux niveaux est utilisé afin de n'en conserver que 100 par seconde environ. Le premier niveau se base sur les signaux bruts du détecteur tandis que le second procède à une reconstruction simple de l'événement pour l'analyser plus en détail.

Un algorithme de reconstruction permet de déterminer quelles particules sont issues des collisions. Cet algorithme se base sur les signaux du détecteur et leurs corrélations, en particulier spatiales, afin d'estimer la nature et les propriétés des particules présentes lors de l'événement. Des objets physiques de haut niveau sont définis à partir des particules reconstruites. Il s'agit de l'énergie transverse manquante, des jets et des taus hadroniques.

Une simulation d'événements physiques et du détecteur lui-même permettent d'obtenir une estimation des observations attendues pour un modèle théorique donné. Cette simulation est corrigée de divers effets, mesurés par des analyses dédiées. Une de ces corrections porte sur l'énergie des jets et est détaillée dans le chapitre 3 qui présente une analyse dédiée à son obtention.

Chapitre 3
Calibration en énergie des jets

Sommaire

1 Introduction 83
2 Étapes de la calibration 84
2.1 Correction de l'empilement 85
2.2 Correction de la réponse du détecteur en p_T et en η 87
2.3 Corrections résiduelles 87
2.4 Incertitudes sur la correction en énergie des jets 90
2.5 Correction de la résolution en énergie 92
3 Phénoménologie des événements γ + jets 92
3.1 Principe des événements γ + jets et réponse équilibrée 92
3.2 Activité additionnelle des jets et réponse MPF 94
4 Correction résiduelle absolue en p_T avec les événements γ + jets 95
4.1 Événements utilisés 95
4.2 Analyse 98
4.3 Résultats 103
5 Correction de la résolution en énergie avec les événements γ + jets 110
5.1 Définition de la résolution en énergie des jets 110
5.2 Analyse 110
5.3 Résultats 111
6 Conclusion 113

1 Introduction

Lors des collisions de protons, comme expliqué au chapitre 2, ce sont en réalité leurs constituants respectifs qui interagissent, c'est-à-dire les quarks et les gluons, regroupés sous le terme de « partons ». Il est donc très probable que l'interaction forte intervienne. De plus, les particules les plus lourdes issues de ces collisions, instables, peuvent émettre des quarks en se désintégrant. De nombreux partons sont ainsi émis lors des collisions de protons au LHC. Comme exposé dans le chapitre 1, les partons issus des collisions forment des jets, objets physiques de haut niveau dont la reconstruction est introduite au chapitre 2. Ces jets sont omniprésents dans les analyses réalisées par la collaboration CMS, leur caractérisation est donc un point essentiel. Or, les jets sont des objets physiques composés de nombreuses particules. Leur calibration en énergie est ainsi nécessaire afin d'obtenir une estimation de leur énergie réelle la plus fidèle possible. Les étapes de cette calibration dans l'expérience CMS sont abordées dans la section 2. Lors de ma thèse, j'ai contribué à la détermination de cette calibration. L'étape correspondante est discutée plus en détail. Elle utilise des événements γ + jets, où un photon et au moins un jet sont présents. La section 3 en présente la phénoménologie et la section 4 aborde l'obtention de la calibration. La correction de la résolution en énergie des jets à l'aide de ces mêmes événements est présentée dans la section 5.

2 Étapes de la calibration

Les jets sont des objets physiques composites complexes qu'il est nécessaire de calibrer. La précision apportée à la mesure des jets est capitale dans de nombreuses analyses dans lesquelles il s'agit d'une source majeure d'incertitude systématique. Les avancées réalisées récemment sur la calibration des jets ont ainsi permis d'améliorer la précision sur la mesure de la section efficace inclusive de production de jets et de la masse du quark t [156].

À partir des jets reconstruits par les méthodes décrites précédemment, un procédé de correction de l'énergie des jets (JEC, *Jet Energy Correction*) est réalisé. Il permet de corriger l'échelle en énergie des jets (JES, *Jet Energy Scale*) ainsi que la résolution sur cette énergie (JER, *Jet Energy Resolution*). La collaboration CMS utilise une approche factorisée en plusieurs étapes, chacune corrigeant un effet en particulier [156]. La figure 3.1 résume ces étapes, décrites dans les sections qui suivent.

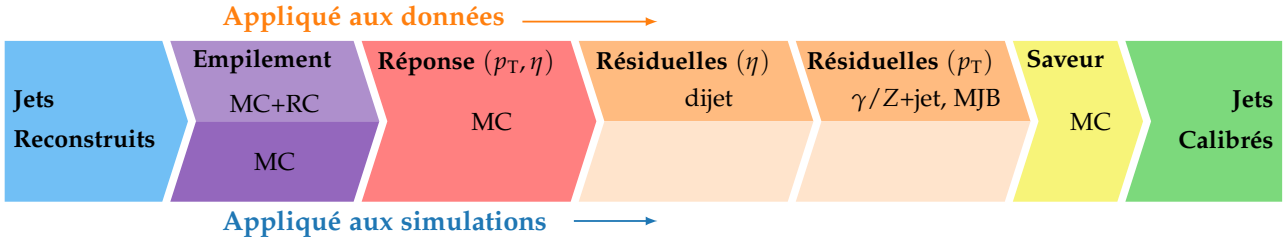


Figure 3.1 – *Étapes successives de la JEC pour les données réelles (« données ») et simulées (« simulations ») [156]. Les corrections des étapes marquées « MC » sont obtenues par l'étude de simulations, celles marquées « RC » par une méthode de cône aléatoire* (Random Cone) *sur les données réelles. Les types d'événements utilisés dans les corrections résiduelles sont également indiqués.*

Trois stades ou « niveaux » de correction sur les particules peuvent être définis :

— le niveau « particule », noté ptcl, ou niveau « vrai », se réfère aux objets et variables après hadronisation mais avant interaction avec le détecteur. Il s'agit donc des grandeurs recherchées, uniquement accessibles dans les événements simulés ;

— le niveau « reconstruit », noté reco, correspond aux objets et variables après interaction avec le détecteur et reconstruction par l'algorithme de PF ;

— le niveau « corrigé » ou calibré, noté corr, correspond aux objets et variables corrigés, c'est-à-dire ceux du niveau reconstruit auxquels ont été appliquées les corrections.

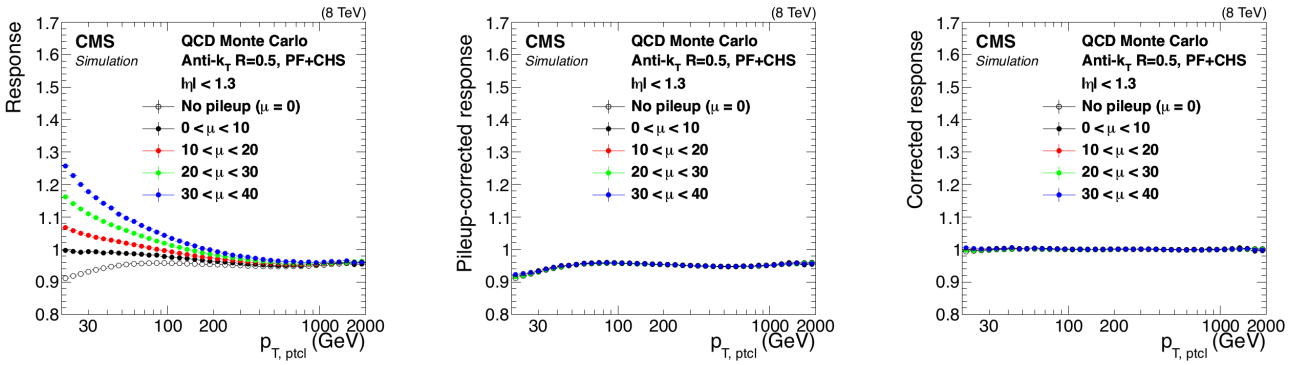
La réponse d'un jet, variable importante dans ce chapitre, est définie comme

$$R = \frac{p_T}{p_{T_{ptcl}}} . \tag{3.1}$$

La réponse peut être définie à différents niveaux, et par définition $R_{ptcl} = 1$. Si la JEC est correcte, alors l'impulsion transverse du jet corrigé doit correspondre sensiblement à l'impulsion transverse au niveau particule, c'est-à-dire $R_{corr} \simeq 1$. Sur la figure 3.2 sont représentées les réponses de jets d'événements QCD simulés à différentes étapes de la JEC. Après avoir appliqué toutes les corrections (figure 3.2c), la réponse est sensiblement égale à 1, ce qui montre que la JEC est correcte.

Les jets au niveau particule sont reconstruits en appliquant la procédure de recombinaison à toutes les particules de durée de vie τ telle que $c\tau > 1\,\text{cm}$, à l'exception des neutrinos [156]. Les hadrons contenant des quarks lourds, c ou b, ne rentrent pas dans cette catégorie et ce sont donc leurs produits de désintégration qui sont pris en compte pour la recombinaison. Exclure les neutrinos permet de définir la réponse des jets d'une manière qui soit accessible expérimentalement et similaire entre jets légers ou de gluons et jets lourds dans lesquels des neutrinos sont produits lors des désintégrations des quarks lourds.

La JEC est déterminée pour chaque année de prise de données et pour différentes subdivisions de ces années. En effet, les conditions expérimentales changent au cours du temps. Ce peut être dû au vieillissement du détecteur, à une modification des caractéristiques des faisceaux de protons ou encore à une modification du détecteur lui-même. Par exemple, entre 2016 et 2017, la partie centrale du trajectographe a été remplacée par un système d'architecture différente [130, 133]. De 2017 à 2018, les



(a) *Avant correction (R_{reco}).* **(b)** *Correction de l'empilement.* **(c)** *Après correction (R_{corr}).*

Figure 3.2 – *Valeur moyenne de la réponse de jets d'événements QCD simulés en fonction de $p_{T_{ptcl}}$ à différentes étapes de la JEC [156] et pour différentes valeurs d'interactions d'empilement μ.*

photodiodes des bouchons du HCAL ont été changées pour des photomultiplicateurs à silicium [137, 191]. La réponse en énergie des jets dépend fortement de la nature du détecteur, c'est pourquoi des évolutions importantes peuvent apparaître sur la JEC d'une année à une autre. C'est en particulier le cas pour la réponse du détecteur dont la correction, présentée dans la section 2.2, est réalisée après la correction de l'empilement abordée ci-après.

2.1 Correction de l'empilement

Des contributions additionnelles à l'énergie et à l'impulsion des jets peuvent apparaître du fait de l'empilement, décrit dans le chapitre 2. La correction de l'empilement a pour but de soustraire ces contributions et est appliquée aux données réelles et simulées. Elle permet d'améliorer la résolution en énergie des jets et d'obtenir une JES plus précise.

L'empilement asynchrone est réduit par l'analyse temporelle des signaux des calorimètres, l'empilement synchrone par la méthode de soustraction des hadrons chargés (CHS, *pile-up Charged Hadron Subtraction*) [192], décrite ci-après.

Pour chacun des vertex primaires de l'événement, la somme des carrés des impulsions transverses des traces associées au vertex est calculée. Le vertex primaire principal est choisi comme étant le vertex présentant la plus grande valeur ainsi obtenue. Les autres vertex primaires sont considérés comme des vertex d'empilement. La méthode CHS consiste à retirer de l'événement les hadrons chargés dont les traces associées proviennent de vertex d'empilement. Ces traces permettent en effet de s'assurer que ces hadrons ne sont pas issus du vertex primaire principal. La reconstruction des jets est alors réalisée à partir de l'événement nettoyé, ce qui permet d'améliorer la résolution en p_T des jets.

La correction C_{PU} de l'empilement résiduel, principalement dû aux hadrons neutres, aux photons, aux traces non associées à un vertex et à l'empilement asynchrone qui n'a pas pu être corrigé totalement, s'exprime en fonction de

— l'impulsion transverse du jet avant application de cette correction et après CHS, $p_{T_{reco}}^{CHS}$;
— la pseudo-rapidité du jet, η ;
— l'aire du jet dans le plan (η, ϕ), A ;
— la densité en énergie dans le plan (η, ϕ) de l'événement contenant ce jet, notée ρ ;

sous la forme

$$C_{PU}(p_{T_{reco}}^{CHS}, \eta, A, \rho) = 1 - \frac{\langle p_{T_{ptcl}}^{add} \rangle}{p_{T_{reco}}^{CHS}} \tag{3.2}$$

où $\langle p_{T_{ptcl}}^{add} \rangle$ est la contribution additionnelle de l'empilement au niveau particule, estimée à l'aide de la méthode de l'aire hybride (*hybrid jet area*) à partir d'événements QCD multijet simulés avec et sans empilement, c'est-à-dire

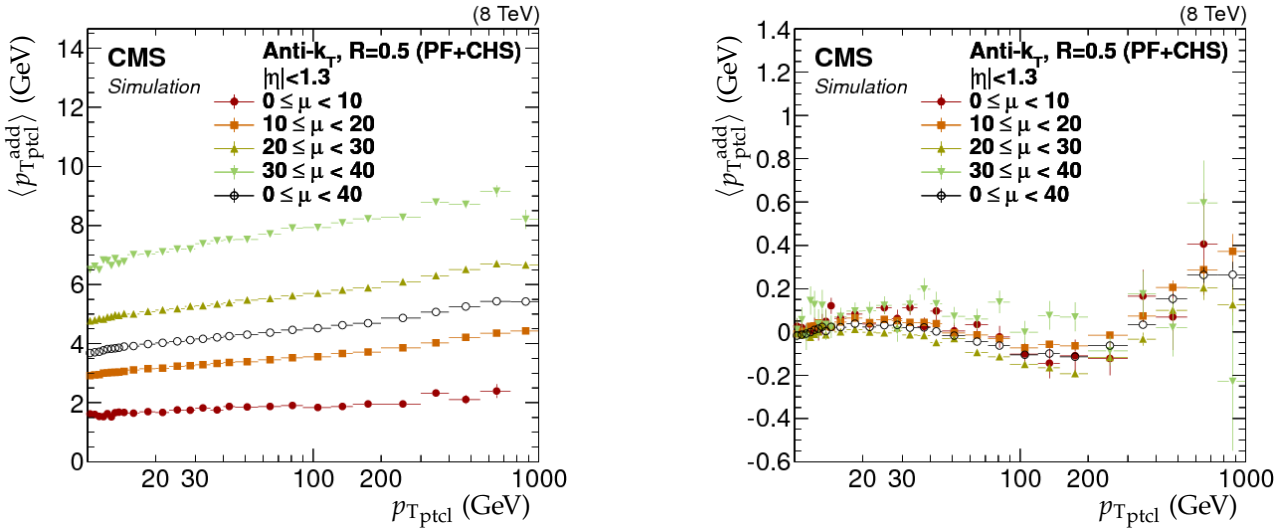
$$\langle p_{T_{ptcl}}^{add} \rangle (\rho, \eta, p_{T_{reco}}^{CHS}) = \langle p_{T_{ptcl}}^{\text{avec PU}} - p_{T_{ptcl}}^{\text{sans PU}} \rangle , \tag{3.3}$$

avec $p_{\mathrm{T}_{\mathrm{ptcl}}}^{\text{avec PU}}$ et $p_{\mathrm{T}_{\mathrm{ptcl}}}^{\text{sans PU}}$ les impulsions du jet au niveau particule avec et sans empilement. La contribution additionnelle de l'empilement au niveau particule est alors paramétrée en fonction de $p_{\mathrm{T}_{\mathrm{reco}}}^{\mathrm{CHS}}$, η, A et ρ et la correction de l'empilement résiduel $\mathcal{C}_{\mathrm{PU}}$ définie par (3.2) peut se réécrire

$$\mathcal{C}_{\mathrm{PU}}(p_{\mathrm{T}_{\mathrm{reco}}}^{\mathrm{CHS}}, \eta, A, \rho) = 1 - \frac{\left[\rho_0(\eta) + \rho\,\beta(\eta)(1 + \gamma(\eta)\log p_{\mathrm{T}_{\mathrm{reco}}}^{\mathrm{CHS}})\right] A}{p_{\mathrm{T}_{\mathrm{reco}}}^{\mathrm{CHS}}} \tag{3.4}$$

où $\rho_0(\eta)$, $\beta(\eta)$ et $\gamma(\eta)$ sont les paramètres de cette correction, dépendants de η.

La figure 3.3 montre $\langle p_{\mathrm{T}_{\mathrm{ptcl}}}^{\mathrm{add}} \rangle$ en fonction de l'impulsion transverse du jet au niveau particule, avant et après application de la correction de l'empilement. Les résultats de la figure 3.3b sont cohérents avec l'absence d'énergie supplémentaire due à l'empilement à $\pm 0{,}2\,\mathrm{GeV}$ jusqu'à $p_{\mathrm{T}_{\mathrm{ptcl}}} = 500\,\mathrm{GeV}$ et à $\pm 0{,}6\,\mathrm{GeV}$ au-delà.



(a) *Avant correction.*

(b) *Après correction.*

Figure 3.3 – *Contribution additionnelle de l'empilement au niveau particule telle que définie dans l'équation (3.3) pour $|\eta| < 1{,}3$ en fonction de l'impulsion du jet au niveau particule pour différentes valeurs du nombre d'interaction d'empilement (μ) [156].*

La correction ainsi décrite doit être légèrement adaptée pour pouvoir l'appliquer aux données réelles à cause des biais de simulation du détecteur. Pour cela, un ajustement en fonction de η est déterminé à l'aide de la méthode de cône aléatoire (RC, *Random Cone*). La méthode RC consiste à placer aléatoirement des cônes couvrant l'ensemble de l'acceptation du détecteur dans le plan (η, ϕ). La reconstruction des jets se fait alors cône par cône, par regroupement des particules contenues dans chacun d'entre eux. L'impulsion transverse moyenne de ces jets, dans chaque événement, est une mesure de la densité d'énergie moyenne se retrouvant dans les jets [156]. L'étude est réalisée sur des événements dits de « zéro biais ». Il s'agit d'événements sélectionnés par un déclenchement aléatoire pendant que les faisceaux de protons se croisent. Le déclenchement n'étant pas dû à un dépôt d'énergie en particulier, ces événements ne comportent pas, en général, de contribution provenant d'une interaction dure, c'est-à-dire d'une collision effective entre les protons. Dans ce cas, l'impulsion transverse moyenne des jets reconstruits par la méthode RC permet d'estimer la moyenne de la contribution additionnelle de l'empilement, c'est-à-dire

$$\langle p_{\mathrm{T}}^{\mathrm{add}} \rangle^{\mathrm{RC}} = \langle p_{\mathrm{T}_{\mathrm{cône}}} \rangle\,. \tag{3.5}$$

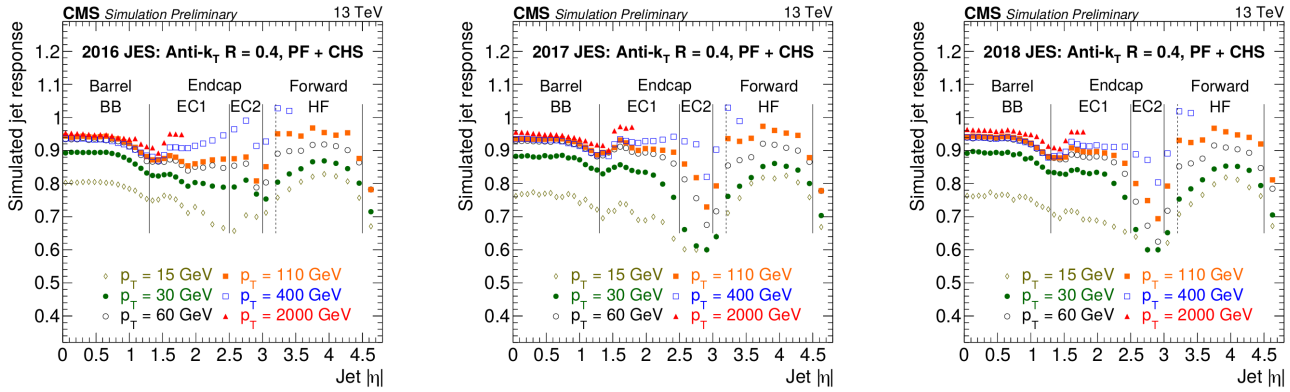
Il est alors possible de définir un facteur d'échelle à appliquer aux paramètres ρ_0 et β de l'équation (3.4) lorsque cette correction est appliquée aux données réelles. Ce facteur d'échelle s'exprime

$$\frac{\langle p_{\mathrm{T}}^{\mathrm{add}} \rangle^{\mathrm{RC}}_{\text{données}}(\eta, \rho_{\text{données}})}{\langle p_{\mathrm{T}}^{\mathrm{add}} \rangle^{\mathrm{RC}}_{\text{simulation}}(\eta, \rho_{\text{simulation}})}\,. \tag{3.6}$$

La contribution additionnelle de l'empilement est ainsi corrigée dans les simulations et les données réelles.

2.2 Correction de la réponse du détecteur en p_T et en η

La réponse du détecteur CMS à un jet n'est pas uniforme selon la valeur de p_T et η du jet. La réponse au niveau reconstruit des jets simulés R_{reco}, déterminée grâce à une simulation du détecteur CMS basée sur GEANT4 [184-186], combinée à PYTHIA 6.4 [178] avec les réglages Z2* [193], est représentée sur la figure 3.4 pour les trois années du Run II du LHC. Il apparaît, par exemple, qu'un jet de $p_T = 30\,\text{GeV}$ nécessite une correction allant de 10 % dans la région centrale $|\eta| < 0{,}7$ à plus de 30 % lorsque $|\eta| \simeq 3$ en 2017 et 2018.



(a) *Année 2016.* **(b)** *Année 2017.* **(c)** *Année 2018.*

Figure 3.4 – *Réponse des jets reconstruits en fonction de p_T et η lors du Run II [164]. La chute de la réponse des jets dans la région $|\eta| \simeq 3$ est due à la transition entre le bouchon et la partie vers l'avant du détecteur. Pour $|\eta| > 4{,}5$, les limites d'acceptation du détecteur expliquent la chute de la réponse des jets. La dégradation au cours du temps du détecteur dans la région « EC2 » s'observe par la baisse de la réponse des jets dans cette région de 2016 à 2017.*

Afin de corriger la réponse du détecteur en p_T et en η, la correction $\mathcal{C}_{\text{Rép}}$ à appliquer s'exprime

$$\mathcal{C}_{\text{Rép}}(p_{T_{\text{reco}}}', \eta) = \frac{\langle p_{T_{\text{ptcl}}} \rangle}{\langle p_{T_{\text{reco}}}' \rangle} = \frac{1}{\langle R_{\text{reco}}' \rangle} \tag{3.7}$$

où $p_{T_{\text{reco}}}'$ est l'impulsion transverse du jet après correction de l'empilement. Les moyennes sont réalisées sur les jets appartenant à la même cellule d'une grille en $(p_{T_{\text{ptcl}}}, \eta)$ prédéfinie [156].

2.3 Corrections résiduelles

Les corrections décrites dans les sections précédentes permettent d'obtenir une bonne calibration en énergie des jets. Toutefois, des différences dans leurs réponses, de l'ordre du pourcent, subsistent entre données réelles et simulées. Des corrections résiduelles à appliquer aux données réelles sont ainsi déterminées afin de réduire ces écarts, définies telles que

$$\mathcal{C}_{\text{Res}} = \frac{R_{\text{simulations}}}{R_{\text{données}}}\,. \tag{3.8}$$

Le principe est d'estimer la réponse du jet en s'appuyant sur un objet de référence pouvant être un boson Z (événements $Z+$jets), un photon (événements $\gamma+$jets) ou un autre jet (événements dijet et multijet). Deux méthodes existent et sont utilisées de manière complémentaire.

La méthode équilibrée estime que l'objet de référence et le jet sont équilibrés (*balanced*) au niveau particule, c'est-à-dire d'impulsion transverse totale nulle, soit

$$\vec{p}_{T_{\text{ptcl}}}^{\,\text{réf}} + \vec{p}_{T_{\text{ptcl}}}^{\,\text{jet}} = \vec{0}\,. \tag{3.9}$$

L'objet de référence étant fidèlement reconstruit,

$$\vec{p}_{T_{\text{reco}}}^{\,\text{réf}} \simeq \vec{p}_{T_{\text{ptcl}}}^{\,\text{réf}} = \vec{p}_T^{\,\text{réf}}\,. \tag{3.10}$$

Ainsi, l'équation (3.9) peut se réécrire à l'aide de (3.10) sous la forme

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{réf}} + \vec{p}_{\mathrm{T\,ptcl}}^{\,\mathrm{jet}} = \vec{0}. \tag{3.11}$$

La réponse d'un jet définie par (3.1) permet alors de faire apparaître la réponse équilibrée du jet, notée R_{bal} (*balanced*),

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{réf}} + \frac{1}{R_{bal}}\vec{p}_{\mathrm{T\,reco}}^{\,\mathrm{jet}} = \vec{0} \Rightarrow R_{bal}(p_{\mathrm{T}}, \eta) = \frac{p_{\mathrm{T\,reco}}^{\mathrm{jet}}}{p_{\mathrm{T}}^{\mathrm{réf}}}. \tag{3.12}$$

La méthode « MPF » (*MET Projection Fraction*) prend en compte l'ensemble de l'activité hadronique de l'événement et considère l'impulsion de recul vis-à-vis de l'objet de référence, c'est-à-dire

$$\vec{p}_{\mathrm{T\,ptcl}}^{\,\mathrm{réf}} + \vec{p}_{\mathrm{T\,ptcl}}^{\,\mathrm{recul}} = \vec{0}. \tag{3.13}$$

Au niveau reconstruit, l'énergie transverse manquante (MET), introduite dans le chapitre 2, doit être prise en compte dans le recul. Afin de garder une description cohérente de l'événement, les corrections précédentes apportées aux jets sont d'abord propagées à $\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}$. L'équation précédente, valable au niveau particule, s'écrit alors

$$\vec{p}_{\mathrm{T\,reco}}^{\,\mathrm{réf}} + \vec{p}_{\mathrm{T\,reco}}^{\,\mathrm{recul}} = -\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}} \tag{3.14}$$

La réponse d'un jet définie par (3.1) permet alors de faire apparaître la réponse MPF du jet, notée R_{MPF},

$$\vec{p}_{\mathrm{T\,reco}}^{\,\mathrm{réf}} + R_{MPF}\vec{p}_{\mathrm{T\,ptcl}}^{\,\mathrm{recul}} = -\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}. \tag{3.15}$$

En appliquant (3.10) et (3.13) à l'équation précédente, il est possible d'écrire

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{réf}} - R_{MPF}\vec{p}_{\mathrm{T}}^{\,\mathrm{réf}} = -\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}. \tag{3.16}$$

Par produit scalaire avec $\vec{p}_{\mathrm{T}}^{\,\mathrm{réf}}$, il vient

$$\left|\vec{p}_{\mathrm{T}}^{\,\mathrm{réf}}\right|^2 (1 - R_{MPF}) = -\vec{p}_{\mathrm{T}}^{\,\mathrm{réf}} \cdot \vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}, \tag{3.17}$$

ce qui permet de définir R_{MPF} comme

$$R_{MPF}(p_{\mathrm{T}}, \eta) = 1 + \frac{\vec{p}_{\mathrm{T}}^{\,\mathrm{réf}} \cdot \vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}}{|\vec{p}_{\mathrm{T}}^{\,\mathrm{réf}}|^2}. \tag{3.18}$$

Ces deux réponses, équilibrée et MPF, sont exploitées afin de déterminer les corrections résiduelles décrites ci-après.

2.3.1 Correction résiduelle relative en η

La première correction résiduelle s'exprime en fonction de η. Elle est obtenue à partir de la comparaison données-simulations sur une sélection d'événements dijet. Son but est de rendre indépendant de η le rapport données réelles sur simulées de la réponse des jets. Cette correction s'appuie sur la bonne reconstruction des jets dans le tonneau du détecteur. Lorsqu'un événement présente un premier jet avec $|\eta| < 1{,}3$, c'est-à-dire dans la région de référence du tonneau, et un second avec $|\eta| > 1{,}3$ et de p_{T} similaire, le premier sert d'objet de référence afin de calibrer le second. Pour cette raison, cette correction est qualifiée de « relative ». La correction à appliquer aux données réelles ainsi obtenue est illustrée sur la figure 3.5 dans le cas des jets d'impulsion transverse égale à 120 GeV.

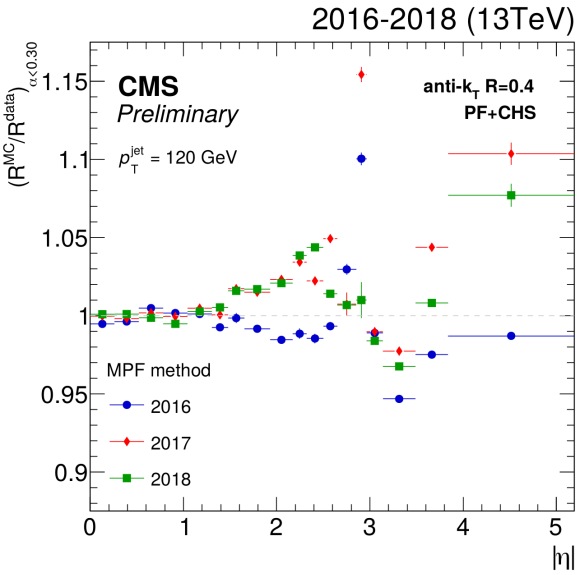


Figure 3.5 – *Correction résiduelle relative en η lors du Run II [164].*

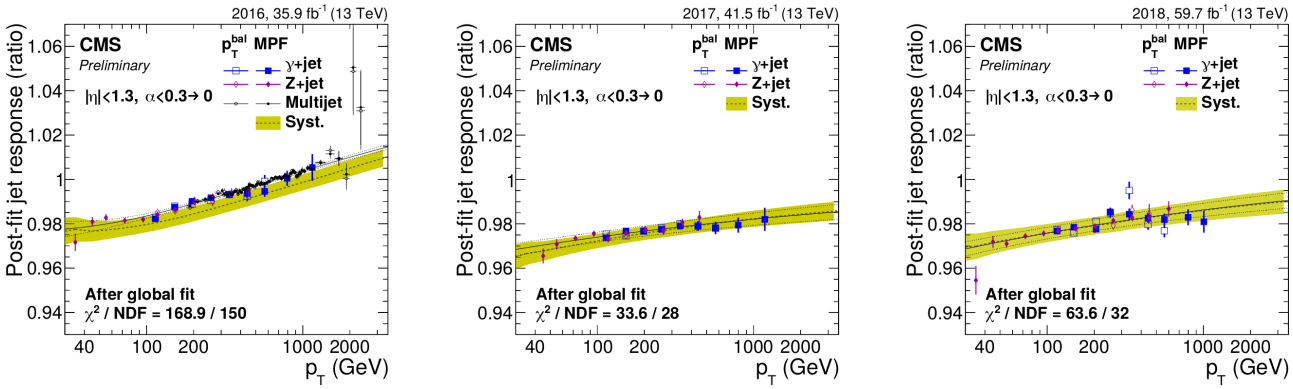
2.3.2 Correction résiduelle absolue en p_T

Cette correction, fonction de p_T, a pour but de rendre indépendant de p_T le rapport données réelles sur simulées de la réponse des jets. Elle combine, à l'aide d'un ajustement global, les comparaisons données-simulations de plusieurs types d'événements afin de couvrir un large spectre de valeurs de p_T. Chaque type d'événement est en effet dominant, de par sa statistique, dans une gamme de p_T donnée :

- événements Z + jets : il s'agit d'événements $Z(\to \mu^+\mu^-)$ + jets et $Z(\to e^+e^-)$ + jets, sélectionnés par la présence d'une paire de muons ou d'électrons compatibles avec la désintégration d'un Z, ils couvrent la région $p_T < 400\,\text{GeV}$;
- événements γ + jets : sélectionnés dans les données à l'aide d'un déclenchement basé sur la présence d'un photon, ils permettent de traiter la région $100\,\text{GeV} < p_T < 1000\,\text{GeV}$;
- événements multijet : ces événements contiennent au moins deux jets dans l'état final et couvrent la région $p_T > 200\,\text{GeV}$.

En 2017 et 2018, les événements multijet n'ont pas pu être exploités. La correction résiduelle absolue en p_T dans la région $p_T > 800\,\text{GeV}$ est alors contrainte par l'analyse des événements γ + jets.

Cette correction corrige l'échelle en énergie absolue des jets, d'où son qualificatif, à partir d'un objet de référence pouvant être un boson Z (Z + jets), un photon (γ + jets) ou un autre jet (multijet). La correction à appliquer aux données réelles ainsi obtenue est illustrée sur la figure 3.6 dans le cas des jets de pseudo-rapidité $|\eta| < 1{,}3$.



(a) *Année 2016.* **(b)** *Année 2017.* **(c)** *Année 2018.*

Figure 3.6 – *Correction résiduelle absolue en p_T pour $|\eta| < 1{,}3$ lors du Run II obtenue par ajustement global sur les événements γ + jets, Z + jets et multijet [164].*

Durant ma thèse, j'ai été responsable de la mesure de cette correction avec les événements γ + jets pour les années 2018, utilisés dans la figure 3.6c afin de réaliser un ajustement global avec les autres analyses, et 2017-UL. Cette responsabilité s'inscrit dans la continuité de ce savoir-faire au sein du groupe CMS de l'IP2I, la détermination de cette correction en 2016 et 2017 y ayant déjà été réalisée [194]. La phénoménologie de ces événements ainsi que leur analyse sont détaillées dans les sections 3 et 4.

2.3.3 Correction résiduelle de saveur

Les propriétés des jets dépendent de leur saveur, c'est-à-dire du parton initial, comme exposé au chapitre 1. Il existe ainsi une différence de réponse selon la saveur du jet, majoritairement due à la fragmentation en énergie et la composition du jet qui en dépendent [156]. Par exemple, les particules de bas p_T se retrouvent hors de l'acceptation du détecteur. Or, des jets initiés par des gluons présentent de nombreuses particules de bas p_T par rapport à ceux issus de quarks légers. Dans une moindre mesure, les jets lourds possèdent également plus de particules de bas p_T que les jets de quarks légers suite à la désintégration du hadron lourd. La proportion de particules neutres dans le jet est également un des paramètres affectant le plus sa réponse.

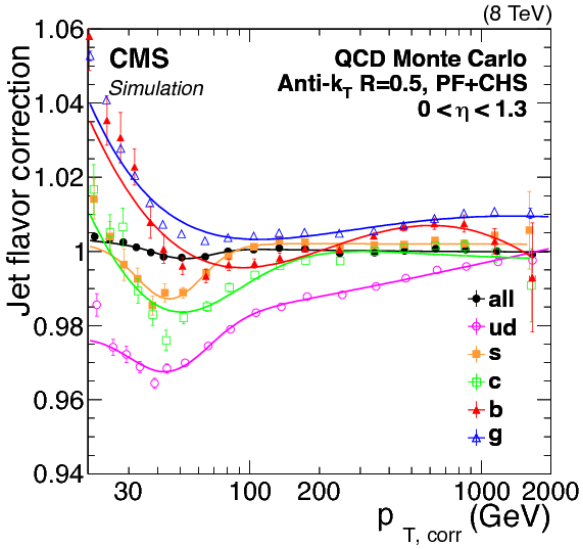


Figure 3.7 – *Correction résiduelle de saveur en fonction de l'impulsion du jet à laquelle ont été aplliquées les corrections décrites dans les sections précédantes, $p_{T_{corr}}$, pour des jets de pseudo-rapidité $|\eta| <$ 1,3 [156].*

La correction résiduelle de saveur $\mathcal{C}_{Sav}$ à appliquer aux données réelles et simulées est obtenue à l'aide de PYTHIA 6.4 [178] avec les réglages Z2* [193] sur des événements dijet, Z + jets et γ + jets simulés et est représentée sur la figure 3.7. Elle est de moins de 2 % au-delà de 100 GeV mais peut atteindre 4 % à bas p_T. Dans les analyses de physique des particules, cette correction ne peut être appliquée qu'à condition de connaître la saveur du jet. Elle n'est donc utilisée en pratique que sur les jets issus de quarks b.

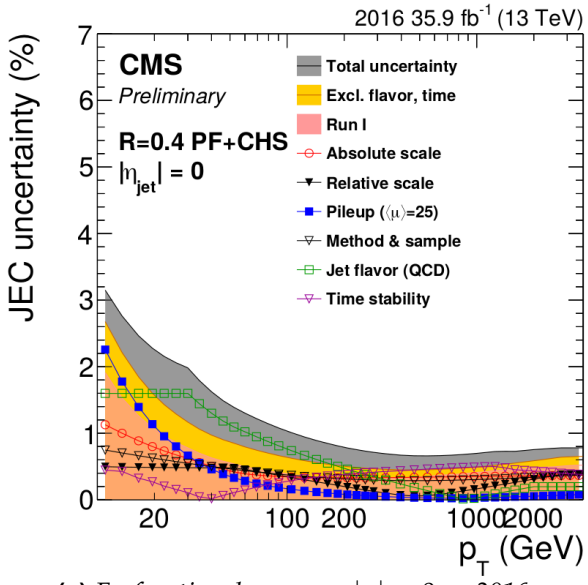
2.4 Incertitudes sur la correction en énergie des jets

Des incertitudes liées aux effets systématiques et, dans une moindre mesure, statistiques sont présentes dans chacune des étapes de la JEC. Elles sont parmi les plus importantes dans de nombreuses analyses de la collaboration CMS, leur caractérisation est donc un enjeu important. Elles sont réparties en six groupes dans le cadre de la JEC [164] :

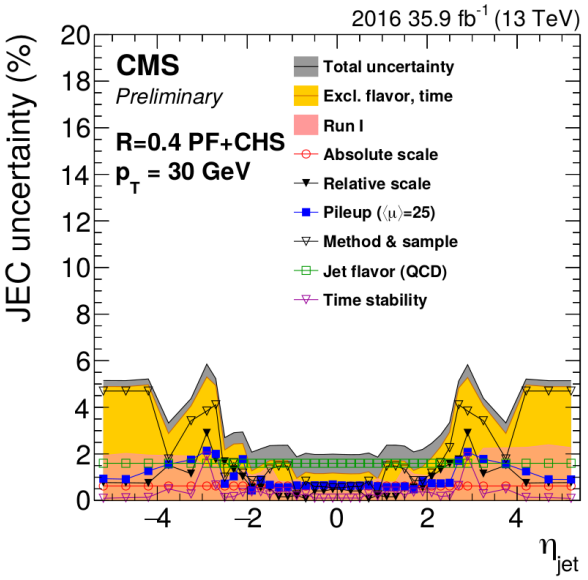
Échelle absolue ou *Absolute scale* sur les légendes des figures 3.8a à 3.8f. Elle rend compte de l'incertitude sur l'échelle en énergie des objets de référence utilisés pour obtenir la correction résiduelle absolue en p_T décrite dans la section 2.3.2 ainsi que des effets de l'ISR (*Initial State Radiation*) et du FSR (*Final State Radiation*) abordés dans la section 3.

Échelle relative ou *Relative scale* sur les légendes des figures 3.8a à 3.8f. Cette incertitude est principalement due à la JER qui s'applique à l'objet de référence dans la correction résiduelle relative en η décrite dans la section 2.3.1 ainsi qu'aux effets de l'ISR et du FSR.

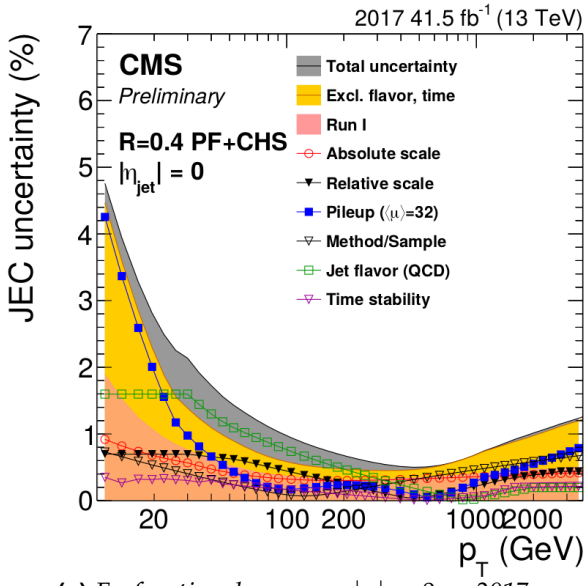
Empilement ou *Pileup* sur les légendes des figures 3.8a à 3.8f. Il s'agit de rendre compte de l'incertitude sur la détermination de la contribution additionnelle de l'empilement au niveau particule. Une incertitude de 5 % sur le rapport données réelles sur simulées de cette contribution, obtenue à l'aide de la méthode de cône aléatoire, est considérée. La différence résiduelle entre la contribution obtenue par cône aléatoire et la contribution réelle extraite des simulations est également prise en compte.



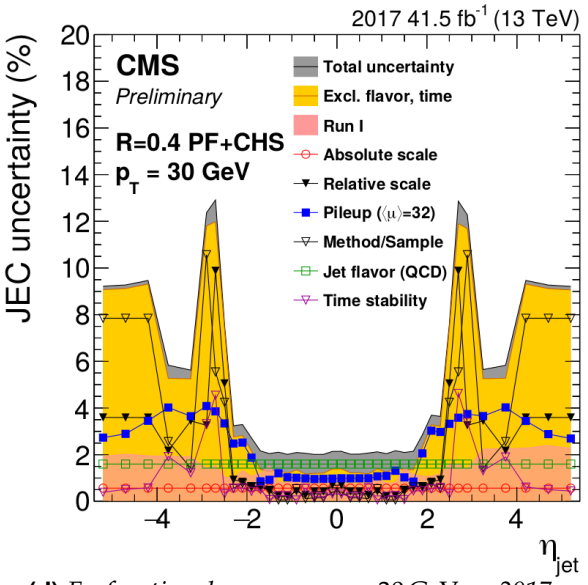
(a) *En fonction de p_T pour $|\eta| = 0$ en 2016.*



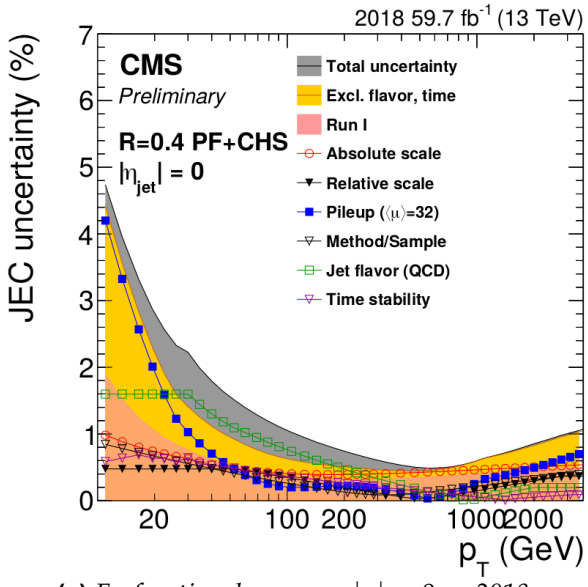
(b) *En fonction de η pour $p_T = 30\,\text{GeV}$ en 2016.*



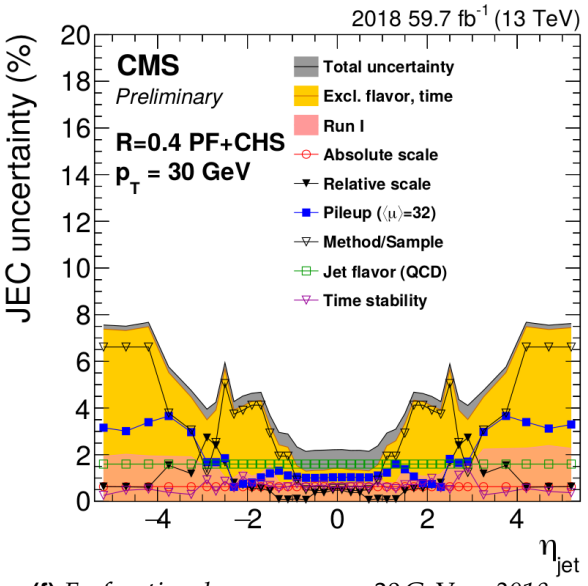
(c) *En fonction de p_T pour $|\eta| = 0$ en 2017.*



(d) *En fonction de η pour $p_T = 30\,\text{GeV}$ en 2017.*



(e) *En fonction de p_T pour $|\eta| = 0$ en 2018.*



(f) *En fonction de η pour $p_T = 30\,\text{GeV}$ en 2018.*

Figure 3.8 – *Incertitudes relatives sur la JEC lors du Run II [164].*

Méthode et jeux de données ou *Method & sample* sur les légendes des figures 3.8a à 3.8f. Cette incertitude correspond aux écarts observés entre les méthodes utilisant les réponses R_{bal} et R_{MPF} d'une part et entre les analyses utilisant les événements $Z+$ jets, $\gamma+$ jets et dijet d'autre part.

Saveur ou *Jet flavor* sur les légendes des figures 3.8a à 3.8f. L'incertitude sur la dépendance en saveur de la réponse des jets dans les simulations est estimée à partir de la différence entre deux générateurs, PYTHIA [178] et HERWIG [182].

Stabilité temporelle ou *Time stability* sur les légendes des figures 3.8a à 3.8f. La JEC est déterminée pour chaque période de prise de données chaque année. Les écarts observés entre ces périodes sont inclus dans cette source d'incertitude.

La figure 3.8 résume les valeurs de ces incertitudes pour les trois années du Run II. L'incertitude globale sur la JEC est généralement inférieure à 2 %, excepté pour les cas $p_T \leqslant 30\,\text{GeV}$ ou $|\eta| \geqslant 2$ où elle peut être de l'ordre de 10 %. Pour $|\eta| \in [2{,}5; 3{,}0]$, cette incertitude est bien plus importante en 2017 que pour les autres années. Il s'agit de la région de transition entre les bouchons du HCAL (HE) et le HCAL vers l'avant (HF), c'est-à-dire deux parties distinctes du détecteur. Or, la qualité de la modélisation de cette région du détecteur est moindre pour l'année 2017, c'est pourquoi l'incertitude y est plus importante.

2.5 Correction de la résolution en énergie

La résolution en énergie des jets, notée JER, est de l'ordre de 20 % pour des jets à $p_T = 30\,\text{GeV}$ et de 10 % à $p_T = 100\,\text{GeV}$ [156]. Cette résolution est donc bien moins bonne que celles d'autres objets physiques tels que les électrons (2 à 5 %), les muons (1 à 6 %) et les photons (environ 1 %). La JER joue ainsi un rôle important dans les analyses cherchant des résonances étroites, par exemple. Il est donc nécessaire de maîtriser cette grandeur.

La JER est définie comme la largeur de la gaussienne obtenue par un ajustement sur la distribution de R_{corr} des jets, c'est-à-dire $p_{T_{\text{corr}}}/p_{T_{\text{ptcl}}}$. Sa mesure est réalisée à l'aide d'événements $\gamma+$ jets et $Z+$ jets et les résultats obtenus lors du Run I sont présentés sur la figure 3.9a. Elle dépend de $p_{T_{\text{ptcl}}}$, η et μ.

La JER observée dans les simulations diffère de celle observée dans les données réelles, elle est légèrement meilleure. Afin de pouvoir réaliser des analyses comparant données réelles et simulées, il est nécessaire d'avoir une JER comparable dans ces deux catégories d'événements. La JER des simulations est ainsi détériorée par un facteur d'échelle (JER SF), déterminé à partir d'événements $\gamma+$ jets et dijet et défini en fonction de η. Les résultats obtenus lors du Run II sont présentés sur la figure 3.9b. Le principe est le même que pour les corrections résiduelles décrites dans les sections 2.3.1 et 2.3.2. Au lieu de s'intéresser à la moyenne de la distribution, c'est sa largeur qui est étudiée.

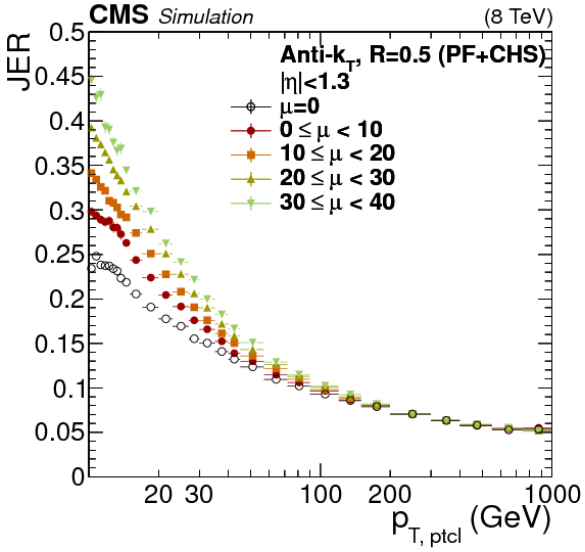
3 Phénoménologie des événements $\gamma+$ jets

Les événements $\gamma+$ jets peuvent être utilisés afin d'obtenir la correction résiduelle absolue en p_T des jets, introduite dans la section 2.3.2, ainsi que la résolution en énergie des jets. Les analyses correspondantes sont abordées dans les sections 4 et 5.

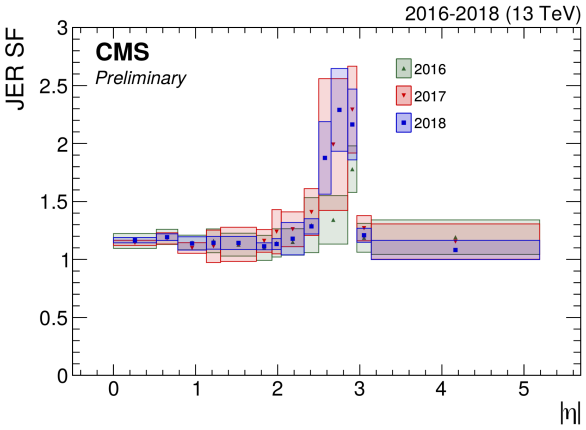
3.1 Principe des événements $\gamma+$ jets et réponse équilibrée

L'état final d'un événement $\gamma+$ jets comporte, d'une part, un jet à calibrer et, d'autre part, un photon utilisé comme objet de référence. Sur la figure 3.10, la résolution sur les photons est inférieure à 4 % et de l'ordre du pourcent dans le tonneau. Dans le cas des jets, sur la figure 3.9a, la résolution minimale est de 5 %. L'utilisation de photons comme objet de référence est donc justifiée.

Des diagrammes de Feynman correspondant à des événements $\gamma+$ jets sont présentés sur la figure 3.11. Ces événements ne comportent pas de neutrino issu de l'interaction dure, il n'y a donc pas d'énergie transverse manquante due à la physique de ces événements. L'impulsion transverse étant nulle dans l'état initial, par conservation, elle est nulle dans l'état final. Le photon et le jet sont



(a) *JER en fonction de p_T dans le tonneau de CMS ($|\eta| < 1{,}3$) pour différentes valeurs d'interactions d'empilement μ lors du Run I [156].*



(b) *Facteurs d'échelle de la résolution en énergie des jets en fonction de η lors du Run II [164].*

Figure 3.9 – *Résolution en énergie des jets dans les simulations et facteurs d'échelle à leur appliquer.*

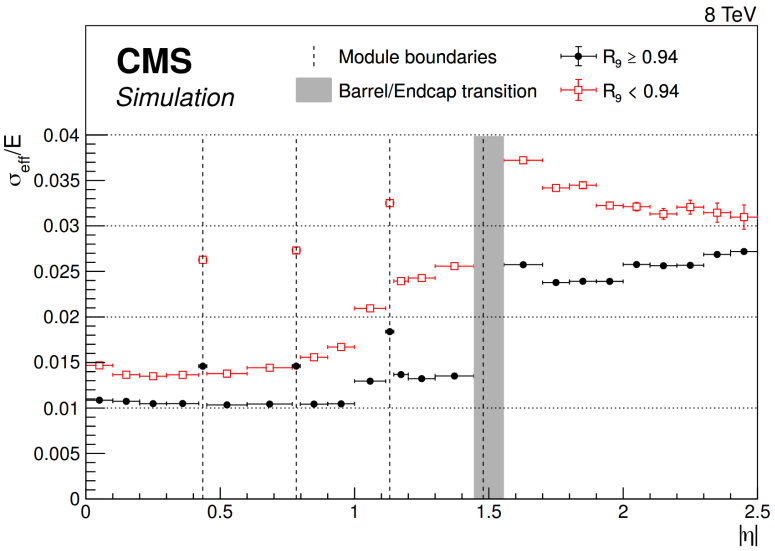


Figure 3.10 – *Résolution relative en énergie des photons en fonction de η pour des événements simulés $h \to \gamma\gamma$ [195]. La variable R_9 est définie au chapitre 2.*

donc équilibrés, c'est-à-dire

$$\vec{p}_{\mathrm{T}\,\mathrm{ptcl}}^{\,\gamma} + \vec{p}_{\mathrm{T}\,\mathrm{ptcl}}^{\,\mathrm{jet}} = \vec{0} \Rightarrow p_{\mathrm{T}\,\mathrm{ptcl}}^{\,\gamma} = p_{\mathrm{T}\,\mathrm{ptcl}}^{\,\mathrm{jet}}\,. \tag{3.19}$$

L'impulsion transverse du jet doit donc être égale à celle du photon, objet de référence. La bonne résolution en énergie sur les photons permet de considérer que leur impulsion transverse au niveau reconstruit est égale à leur impulsion transverse au niveau particule. Ainsi, la méthode équilibrée introduite dans la section 2.3 permet de définir

$$R_{bal} = \frac{p_{\mathrm{T}\,\mathrm{reco}}^{\,\mathrm{jet}}}{p_{\mathrm{T}}^{\,\gamma}}\,, \tag{3.20}$$

qui doit valoir 1 après correction. Cette méthode est performante pour les événements à un photon et un jet dont la topologie est représentée sur la figure 3.12a.

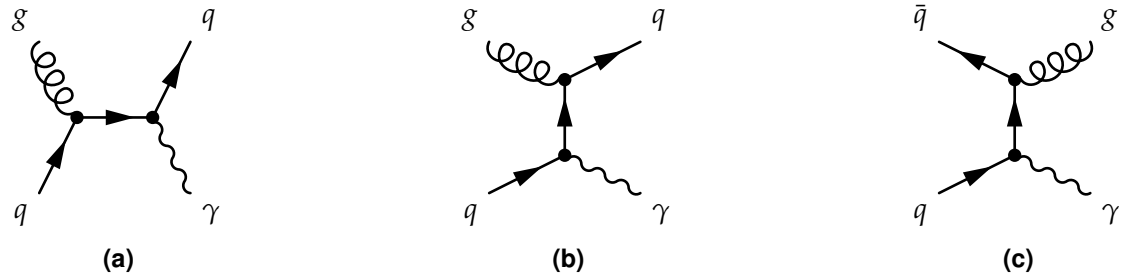
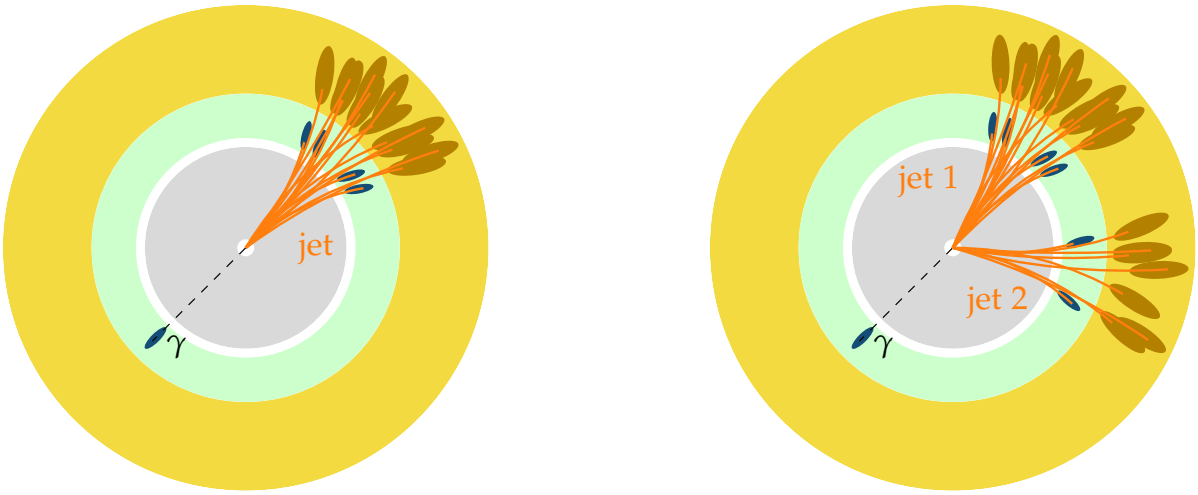


Figure 3.11 – *Exemples de diagrammes de Feynman de processus physiques donnant un photon et un jet dans l'état final.*



(a) *Topologie typique des événements correspondant aux diagrammes de la figure 3.11.*

(b) *Topologie typique des événements correspondant au diagramme de la figure 3.13d.*

Figure 3.12 – *Topologies typiques des événements γ + jets. Réalisé à l'aide de CMSTransverseTikZ [51].*

3.2 Activité additionnelle des jets et réponse MPF

Une activité additionnelle des jets peut survenir et altérer la topologie des événements γ + jets. Un photon peut être émis dans l'état initial (ISR, *Initial State Radiation*) ou dans l'état final (FSR, *Final State Radiation*), ce qui correspond aux diagrammes de Feynman des figures 3.13a et 3.13b. Un système composé d'un des photons et du jet n'est donc pas équilibré dans ce cas. Il est possible de supprimer ce biais en imposant la présence d'un seul photon dans l'événement. La section efficace de production d'événements γ + jets à 13 TeV est importante [196], il est donc possible de sélectionner de manière stricte les événements afin d'obtenir une bonne pureté tout en conservant une statistique suffisante.

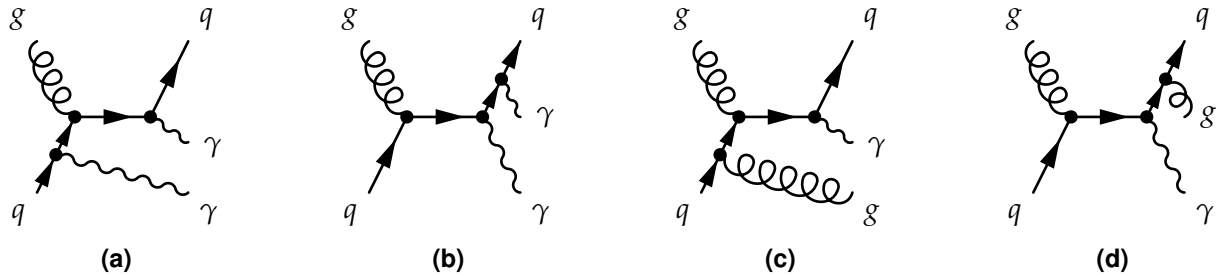


Figure 3.13 – *Exemples de diagrammes de Feynman de processus correspondant à des événements avec deux photons (3.13a, 3.13b) ou deux jets (3.13c, 3.13d) dans l'état final, dus à l'ISR (3.13a, 3.13c) ou au FSR (3.13b, 3.13d).*

L'ISR et le FSR peuvent aussi produire un gluon, ce qui correspond aux diagrammes de Feynman des figures 3.13c et 3.13d. Plusieurs jets sont alors présents dans l'état final et sont ordonnés par impulsion transverse décroissante. La topologie d'un tel événement est illustrée sur la figure 3.12b. Il est possible de réduire le nombre d'événements avec ISR ou FSR en imposant une condition sur

les directions du photon et du premier jet qui doivent être opposées. Toutefois, la plupart des événements présentent plusieurs jets dans la direction opposée au photon. La réponse équilibrée est alors considérée entre le photon et le premier jet, c'est-à-dire le jet d'impulsion transverse la plus grande. Ainsi,

$$R_{bal} = \frac{p_{\mathrm{Treco}}^{\text{jet 1}}}{{p_{\mathrm{T}}}^{\gamma}} . \tag{3.21}$$

En complément de la réponse équilibrée, la réponse MPF, définie comme

$$R_{MPF} = 1 + \frac{{\vec{p}_{\mathrm{T}}}^{\gamma} \cdot \vec{E}_{\mathrm{T}}^{\mathrm{miss}}}{|{\vec{p}_{\mathrm{T}}}^{\gamma}|^2} , \tag{3.22}$$

est également analysée. Les impulsions de toutes les particules présentes étant considérées, R_{MPF} est moins sensible à l'activité additionnelle que R_{bal}. L'utilisation conjointe de R_{MPF} avec R_{bal} permet d'obtenir des résultats complémentaires. Des écarts significatifs observés entre les deux méthodes indiqueraient ainsi des effets incompris, nécessitant de plus amples investigations.

4 Correction résiduelle absolue en p_T avec les événements γ + jets

Durant ma thèse, j'étais responsable de la mesure de cette correction avec les événements γ + jets pour les années 2018 et 2017-UL pour la collaboration CMS. Cette section présente la sélection des événements γ + jets, leur analyse et les derniers résultats obtenus pour l'année 2018.

4.1 Événements utilisés

4.1.1 Ensembles d'événements analysés

Données réelles Les jeux de données réelles utilisés pour 2018 et 2017-UL sont basés sur la présence d'un photon dans l'état final. Pour chacune de ces années, plusieurs périodes (*runs*) sont considérées, celles des collisions pp, dont la liste et les luminosités correspondantes sont présentées dans le tableau 3.1. La liste détaillée de ces jeux de données se trouve dans l'annexe D.

Run	Luminosité (fb^{-1})	
	2018	2017-UL
A	13,98	-
B	7,06	4,823
C	6,90	9,664
D	31,75	4,252
E	-	9,278
F	-	13,54
Total	59,69	41,56

Tableau 3.1 – *Liste des périodes de prise de données considérées et luminosités correspondantes.*

Données simulées Les simulations utilisées contiennent des événements γ + jets de type $qg \to q\gamma$, comme ceux des figures 3.11a et 3.11b, et $q\bar{q} \to g\gamma$, comme celui de la figure 3.11c. Pour l'année 2018, les événements sont générés en un seul jeu de données à l'aide de PYTHIA 8 [179] avec les réglages CP5-Flat [197] et une énergie dans le centre de masse de 13 TeV. Dans l'état final, un photon d'impulsion transverse comprise entre 15 et 6000 GeV est généré. Pour l'année 2017-UL, les événements sont générés conjointement à l'aide de PYTHIA 8 [179] avec les réglages CP5 [197] et MADGRAPH 5 [161] et une énergie dans le centre de masse de 13 TeV. Dans l'état final, la somme scalaire des impulsions transverses des jets, notée HT, appartient à un intervalle, définissant ainsi cinq jeux de données. Les sections efficaces des événements simulés ainsi obtenus sont présentées dans le tableau 3.2. La liste détaillée de ces jeux de données simulées se trouve dans l'annexe D.

Année	Caractéristique	Section efficace (pb)
2018	$p_T{}^{\gamma} \in [15, 6000]$ GeV	283 000,0
2017-UL	HT $\in [40, 100]$ GeV	18 700,0
2017-UL	HT $\in [100, 200]$ GeV	8640,0
2017-UL	HT $\in [200, 400]$ GeV	2185,0
2017-UL	HT $\in [400, 600]$ GeV	259,9
2017-UL	HT > 600 GeV	85,31

Tableau 3.2 – *Sections efficaces des différents événements γ + jets simulés.*

4.1.2 Sélection des événements

Une sélection des événements à considérer est réalisée lors de l'analyse. Les événements souhaités sont ceux contenant un photon avec un ou plusieurs jets. Un des principaux bruits de fond provient d'événements multijet dans lesquels l'un des jets est identifié à tort comme un photon. Cette situation peut arriver lorsqu'une fraction importante de l'énergie de ce jet est portée par un ou plusieurs pions neutres, les π^0. Les π^0 se propagent sur des distances moyennes de 26 nm puis se désintègrent dans 99 % des cas en deux photons [9]. Ces particules ne laissent donc aucune trace dans le trajectographe et un dépôt d'énergie dans le ECAL, tout comme les photons issus de l'interaction initiale. Un tel jet comporte ainsi une signature similaire à celle d'un photon d'un événement γ + jet autour duquel une activité hadronique existe.

Sélection sur les photons Une sélection des photons est appliquée afin de réduire le bruit de fond. Le critère d'identification des photons retenu dans l'analyse est le critère strict, introduit dans le chapitre 2. Il est demandé que $R_9 > 0{,}90$, où R_9 est définie dans le chapitre 2. De plus, un veto est appliqué aux photons pouvant être reliés à une trace dans le trajectographe. Ce veto permet de limiter la contamination des photons considérés dans l'analyse par des électrons reconstruits à tort comme des photons. Seuls les photons situés dans le tonneau sont utilisés car ils présentent la meilleure résolution. Une coupure sur leur pseudo-rapidité est donc appliquée telle que $|\eta| < 1{,}3$, excluant également les bords du tonneau proches des bouchons.

Une étude interne a montré que l'utilisation des photons des bouchons ($|\eta| > 1{,}479$) permettrait de rajouter près de 30 % d'événements dans les données et de réduire ainsi les incertitudes statistiques. Les corrections résiduelles obtenues pourraient alors être plus précises. Toutefois, la résolution en énergie des photons, représentée en fonction de $|\eta|$ sur la figure 3.10 est de l'ordre de 1 % dans le tonneau contre 2,5 % dans les bouchons. Cette différence sur la résolution en énergie de l'objet de référence rend délicate la combinaison des photons du tonneau et des bouchons dans cette analyse. Des travaux sont actuellement menés dans cette direction.

Sélection sur les jets Les événements présentant un unique photon sélectionné d'après les critères précédents sont retenus. Avec ce photon doit être présent au moins un jet reconstruit à l'aide de l'algorithme anti-k_T [162] avec un paramètre $R = 0{,}4$ et respectant les critères définis dans le chapitre 2. Ces critères permettent de rejeter les jets issus du bruit de fond avec une efficacité de 99 %.

Les jets ainsi sélectionnés sont calibrés en énergie en suivant la procédure décrite dans la section 2 jusqu'à la correction résiduelle relative en η incluse. Ils sont alors triés par impulsion transverse décroissante. Pour s'assurer d'un bon équilibre dans le plan transverse entre le photon et le premier jet, c'est-à-dire celui d'impulsion transverse la plus grande, seuls les événements proposant un écart angulaire $\Delta\phi$ entre le photon et ce jet supérieur à 2,8 rad sont considérés dans la suite. Le photon et le jet sont donc dos à dos dans le plan transverse. Il est de plus requis que ce jet soit d'impulsion transverse supérieure à 15 GeV.

Si un second jet d'impulsion transverse supérieure à 10 GeV et à 30 % de $p_T{}^{\gamma}$ est présent, l'événement est rejeté. L'événement est également rejeté si un lepton (électron ou muon) isolé, c'est-à-dire hors des jets, est présent.

Il a été observé que des régions des calorimètres produisent des signaux menant à des reconstructions biaisées de jets. De plus, un système de lecture expérimental « SiPM » a été installé dans le bouchon du HCAL [191]. Il en résulte des calibrations qui ne sont pas optimales dans les régions correspondantes du détecteur. Afin d'éviter ces effets, tout événement est rejeté s'il comporte un jet sélectionné dans ces régions des calorimètres, résumées sur la figure 3.14.

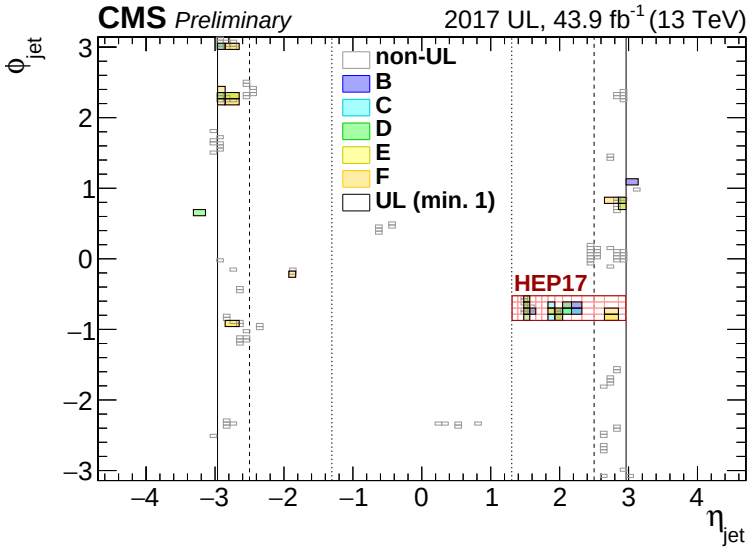


Figure 3.14 – *Régions des calorimètres à exclure de l'analyse dans le plan (η, ϕ) pour les événements de 2017-UL. Certaines régions ne concernent que certaines périodes de l'année (en couleur). La région « HEP17 » correspond à l'emplacement du système de lecture expérimental « SiPM » [191].*

Sélection sur le chemin de déclenchement Comme expliqué dans le chapitre 2, un événement observé par le détecteur CMS est sauvegardé si un chemin de déclenchement (HLT *path*) est activé. Dans cette analyse, seuls les événements dont le photon servant d'objet de référence pour la calibration des jets correspond à celui ayant activé le chemin de déclenchement sont retenus. Or, plus l'impulsion transverse du photon ${p_T}^\gamma$ est faible, plus le nombre d'événements pouvant être sauvegardés est importante, si bien que la chaîne d'acquisition arrive à saturation. Pour pallier cela, il existe différents chemins de déclenchement selon ${p_T}^\gamma$. Pour chacun d'entre eux, seule une fraction des événements les déclenchant est sauvegardée, déterminée par le nombre d'événements observés pour un sauvegardé ou « *prescale* ». Un *prescale* de 100 signifie qu'un événement sur 100 est enregistré. Un intervalle en p_T du photon retenu est défini pour chaque chemin de déclenchement utilisé afin de se placer à son plateau d'efficacité. Le tableau 3.3 résume ces intervalles et *prescales*. Par exemple, un photon tel que ${p_T}^\gamma = 95\,\text{GeV}$ doit avoir activé le chemin nommé `HLT_Photon75_R9Id90_HE10_IsoM`.

Chemin de déclenchement	${p_T}^\gamma$ (GeV)	*Prescale* 2018	*Prescale* 2017-UL
`HLT_Photon33`	$[40, 60[$	24 928,1	2911,33
`HLT_Photon50_R9Id90_HE10_IsoM`	$[60, 85[$	253,333	135,050
`HLT_Photon75_R9Id90_HE10_IsoM`	$[85, 105[$	63,8340	31,2939
`HLT_Photon90_R9Id90_HE10_IsoM`	$[105, 130[$	31,9591	15,7153
`HLT_Photon120_R9Id90_HE10_IsoM`	$[130, 175[$	8	5,322 78
`HLT_Photon165_R9Id90_HE10_IsoM`	$[175, 230[$	4	1,465 51
`HLT_Photon200`	$[230, +\infty[$	1	1

Tableau 3.3 – *Chemins de déclenchement, intervalles d'impulsion transverse du photon et* prescales *utilisés.*

4.2 Analyse

Intervalles de $p_T{}^{\gamma}$ L'analyse a pour but de déterminer la correction résiduelle absolue en p_T des jets, définie dans la section 2.3. Pour cela, l'écart à l'unité du rapport moyen des réponses des jets dans les données réelles et simulées est déterminé dans différents intervalles de $p_T{}^{\gamma}$, listés dans le tableau 3.4. Ils sont une subdivision des intervalles définis pour les chemins de déclenchement dans le tableau 3.3, ce qui permet de séparer le traitement des événements correspondant à différents chemins de déclenchement.

[40, 50[	[50, 60[	[60, 85[	[85, 105[
[105, 130[	[130, 175[	[175, 230[	[230, 300[
[300, 400[	[400, 500[	[500, 700[	[700, 1000[
[1000, 3000]			

Tableau 3.4 – *Intervalles de $p_T{}^{\gamma}$ en* GeV.

Intervalles de $|\eta^{\text{jet}}|$ La calibration en énergie des jets dépend fortement de la région du détecteur dans laquelle le jet laisse un signal, comme le montre la figure 3.4. Cet effet est dû aux différentes technologies utilisées ainsi qu'au vieillissement non uniforme du détecteur. Des intervalles de pseudo-rapidité du jet sont ainsi définis dans le tableau 3.5 afin de séparer le traitement de ces différentes régions.

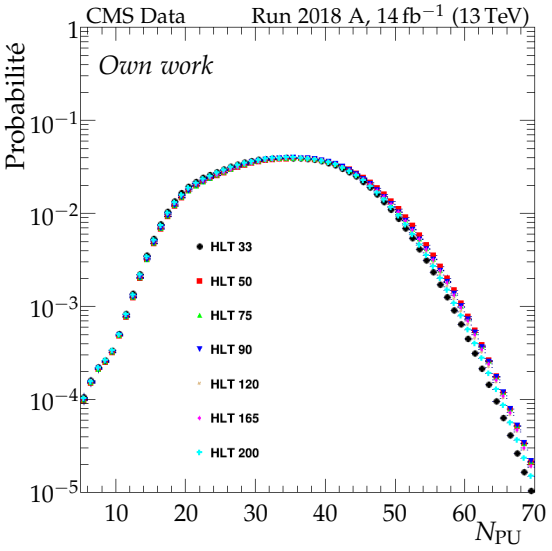
[0; 0,783[	[0,783; 1,305[	[1,305; 1,93[	[1,93; 2,5[
[2,5; 2,964[	[2,964; 3,2[	[3,2; 5,191[	

Tableau 3.5 – *Intervalles larges de $|\eta^{jet}|$.*

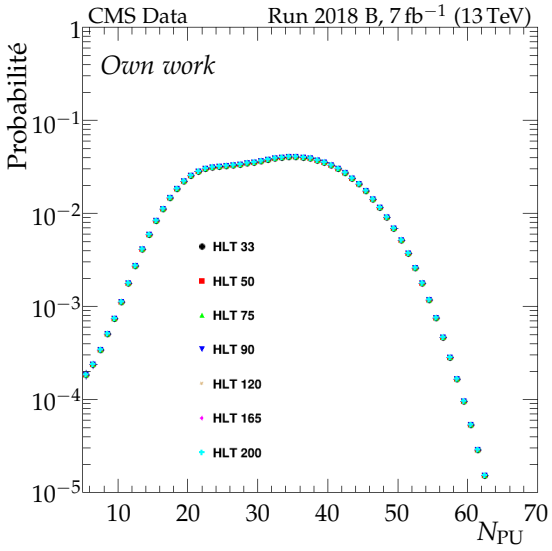
Pondération par l'empilement Le profil d'empilement, c'est-à-dire la densité de probabilité du nombre d'interactions d'empilement, dépend de la période de la prise de données et du chemin de déclenchement par lequel l'événement est retenu. Ces dépendances sont illustrées sur les graphiques des figures 3.15 et 3.16. Les événements simulés sont ainsi pondérés pour faire correspondre leur profil d'empilement à celui des données réelles, en prenant en compte la double dépendance avec la période de prise de données et le chemin de déclenchement.

Accord données-simulations Pour comparer les distributions des observables dans les événements, les distributions des événements simulés sont normalisées à la luminosité mesurée pour le jeu de données considéré. Les comparaisons étant faites entre les données réelles et les événements simulés γ + jets, un désaccord dû à la contamination à bas p_T d'événements multijet est attendu, ces événements n'étant pas présents dans les simulations utilisées. De plus, l'utilisation d'une simulation au premier ordre perturbatif seulement influe sur le nombre de jets dans l'état final qui s'en trouve plus faible, en particulier dans les queues des distributions. Ces désaccords se constatent sur les graphiques de la figure 3.17 présentant les distributions de l'impulsion transverse du photon, l'énergie transverse manquante et les impulsions transverses du premier et du second jet. Afin de déterminer la correction résiduelle absolue en p_T des jets ainsi que la correction de leur résolution en énergie, seule la comparaison des distributions de R_{bal} et R_{MPF} est nécessaire. L'accord ainsi obtenu entre données réelles et simulées est considéré comme suffisant.

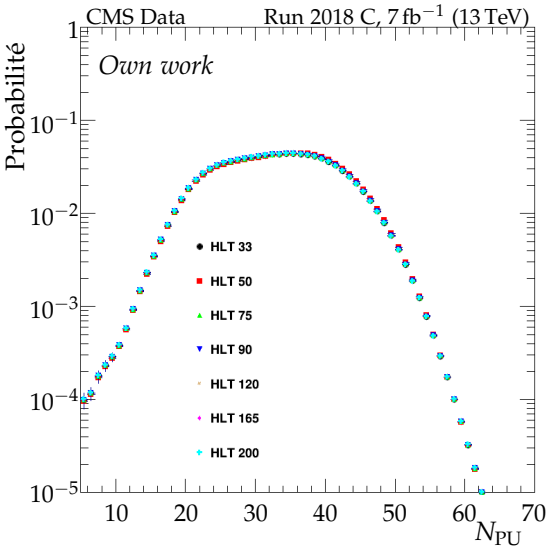
Activité additionnelle des jets supplémentaires La présence d'un jet secondaire, comme sur la figure 3.12b, crée un déséquilibre dans R_{bal} dû à la physique de l'événement et non à la JES. Il ne faut donc pas corriger cet effet. Pour cela, il faut pouvoir se ramener au cas où un seul jet est présent, comme dans l'événement de la figure 3.12a. L'activité additionnelle liée aux jets supplémentaires est



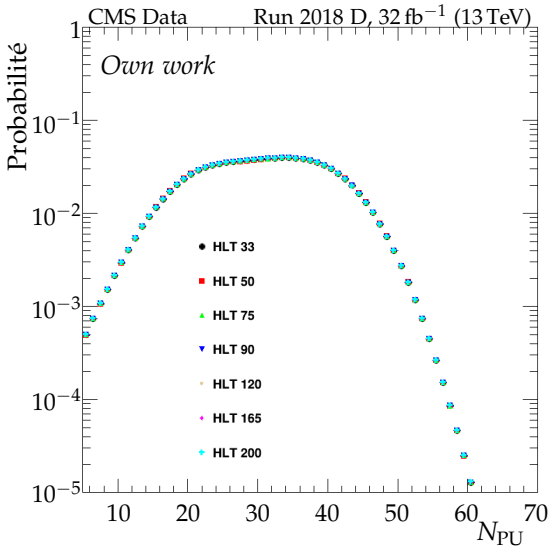
(a) *Run 2018 A.*



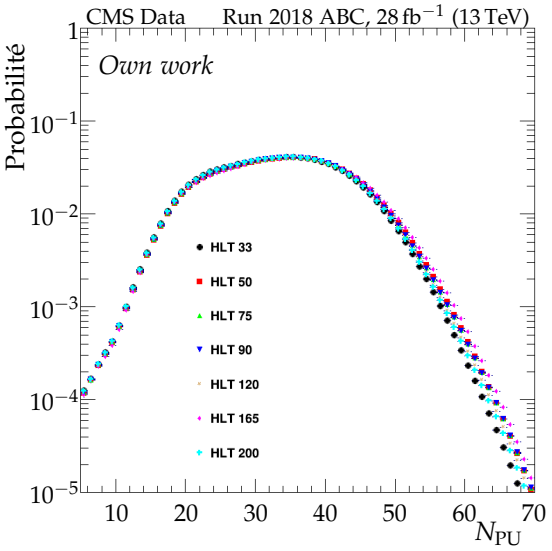
(b) *Run 2018 B.*



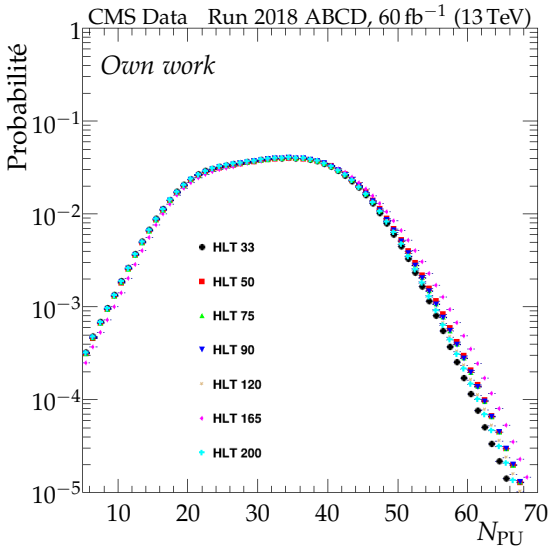
(c) *Run 2018 C.*



(d) *Run 2018 D.*

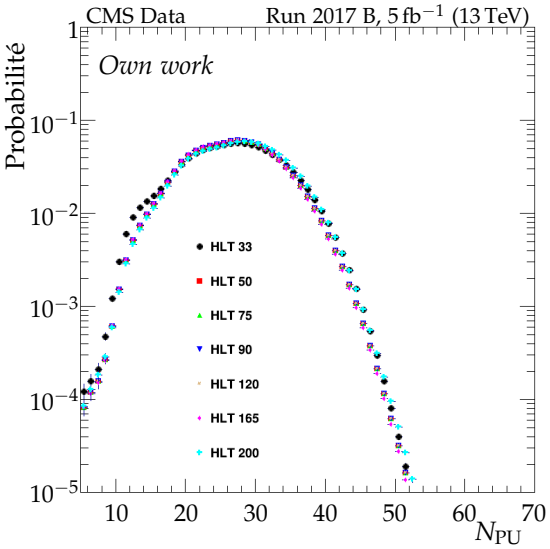


(e) *Run 2018 ABC.*

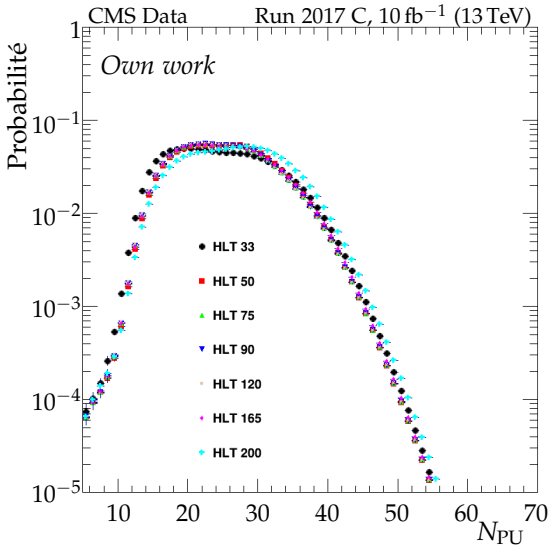


(f) *Run 2018 ABCD.*

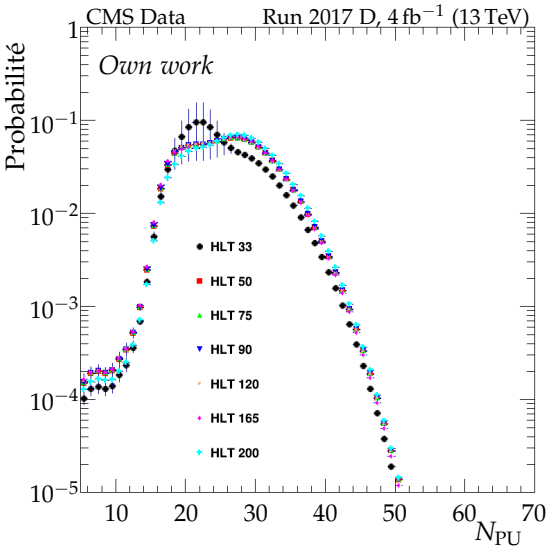
Figure 3.15 – *Densités de probabilité du nombre d'interactions d'empilement N_{PU} pour les périodes de prises de données de 2018.*



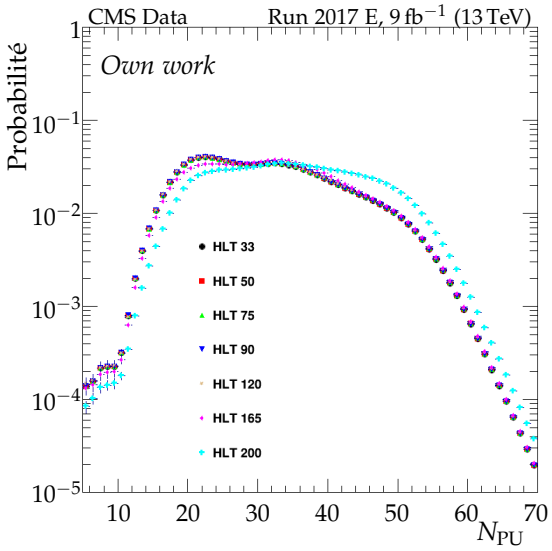
(a) *Run 2017-UL B.*



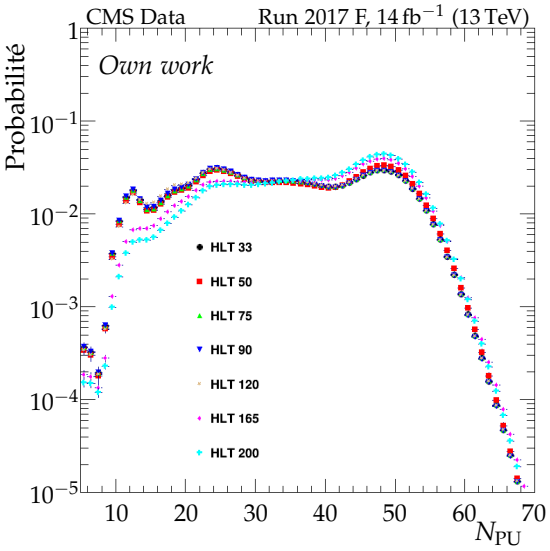
(b) *Run 2017-UL C.*



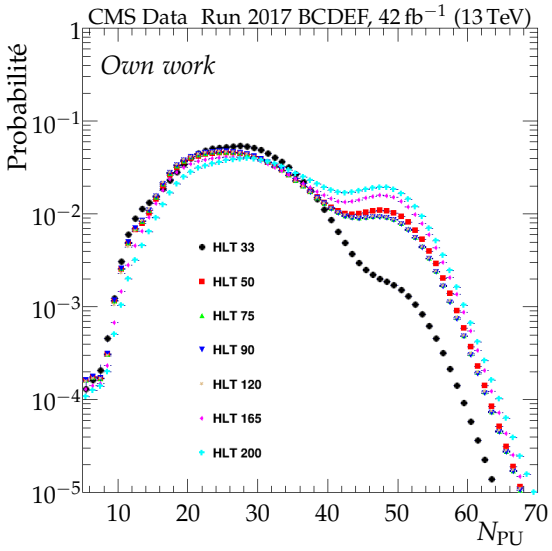
(c) *Run 2017-UL D.*



(d) *Run 2017-UL E.*

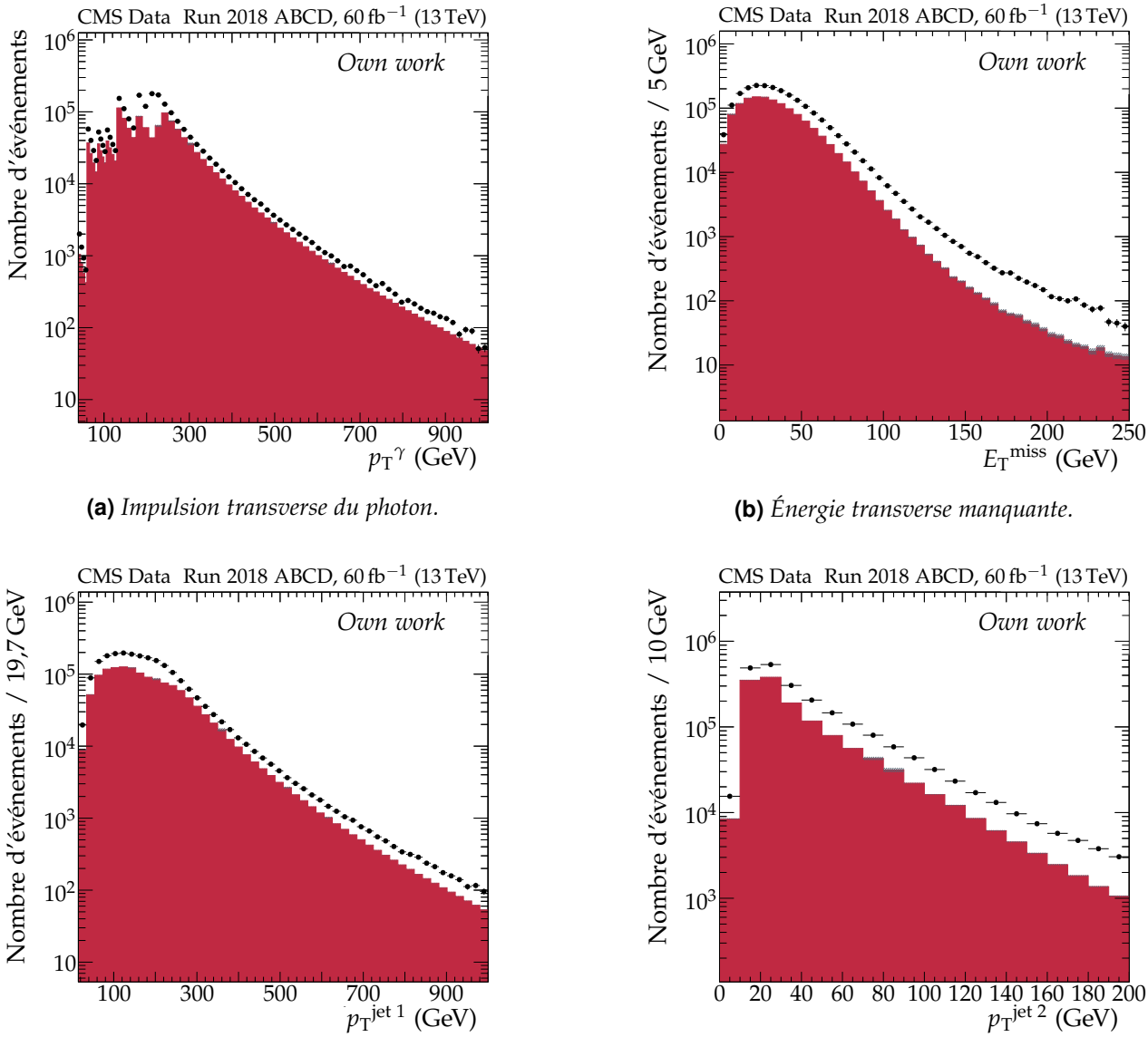


(e) *Run 2017-UL F.*



(f) *Run 2017-UL BCDEF.*

Figure 3.16 – *Densités de probabilité du nombre d'interactions d'empilement N_{PU} pour les périodes de prises de données de 2017-UL.*



(a) *Impulsion transverse du photon.*

(b) *Énergie transverse manquante.*

(c) *Impulsion transverse du premier jet.*

(d) *Impulsion transverse du second jet.*

Figure 3.17 – *Distributions d'observables dans les données réelles (points noirs) et simulées (histogramme en rouge) pour l'année 2018. Sur la figure 3.17a, l'effet des* prescales *donnant une distribution en dents de scie est clairement visible.*

quantifiée par la variable

$$\alpha = \frac{p_{T\text{reco}}^{\text{jet 2}}}{p_T^{\gamma}}. \tag{3.23}$$

L'analyse des événements γ + jets est alors réalisée pour différents intervalles de α afin de pouvoir réaliser par la suite une extrapolation à $\alpha = 0$, correspondant au cas idéal d'événements γ + jet. Les intervalles utilisés sont présentés dans le tableau 3.6. Il s'agit d'intervalles inclusifs, c'est-à-dire que chaque intervalle contient l'intervalle précédent. L'évolution des réponses moyennes en fonction de α y est linéaire *a posteriori*. Cela se retrouve dans les résultats de la figure 3.19. Ce comportement linéaire rend possible une extrapolation simple vers $\alpha = 0$.

[0; 0,10[	[0; 0,15[	[0; 0,20[	[0; 0,25[	[0; 0,30[

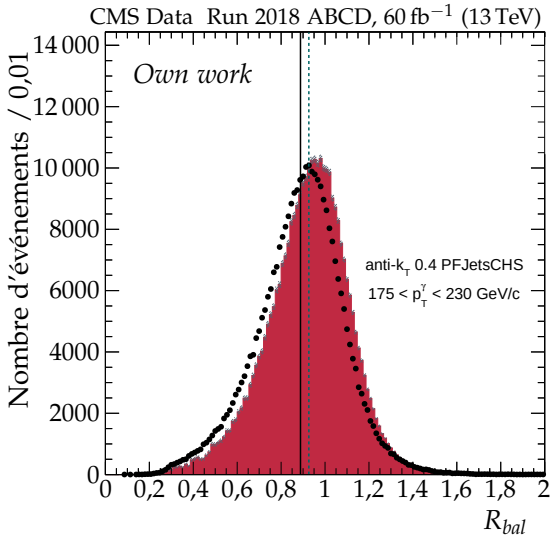
Tableau 3.6 – *Intervalles de α utilisés pour la JES.*

Des études sont en cours afin d'inclure des valeurs de α allant jusqu'à 1. L'exploitation des événements tels que $\alpha > 0{,}3$ est doublement motivée. Ces événements permettraient d'améliorer les

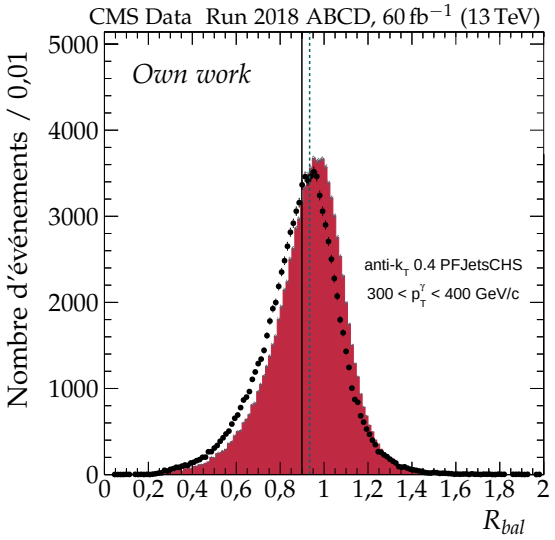
corrections vis-à-vis du FSR et les corrections à bas p_T. En effet, pour $p_T{}^{\gamma} < 100\,\text{GeV}$, imposer $\alpha < 0{,}3$ implique $p_T{}^{\text{jet 2}} < 30\,\text{GeV}$, ce qui limite fortement le nombre d'événements exploitables.

Obtention des corrections pour $(p_T{}^{\gamma}, \eta^{\text{jet}}, \alpha^{\text{max}})$ donnés Pour chaque domaine de $p_T{}^{\gamma}$ défini dans le tableau 3.4, de η^{jet} défini dans le tableau 3.5 et de α défini dans le tableau 3.6, les distributions des réponses équilibrée et MPF des données réelles et simulées sont déterminées. Certaines de ces distributions sont représentées sur la figure 3.18.

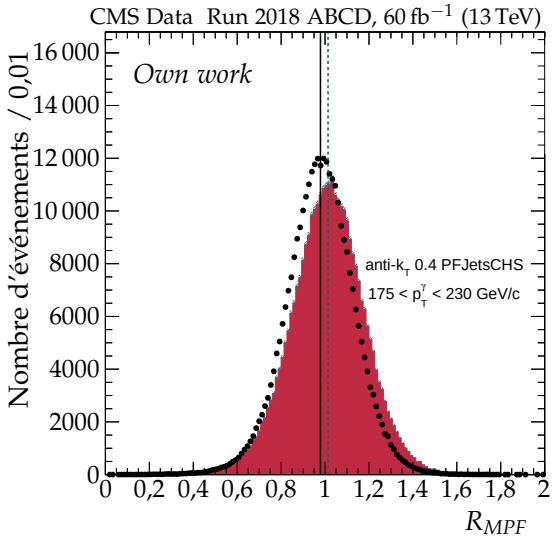
Afin de limiter les effets des queues de ces distributions, en particulier dans le cas de la réponse équilibrée, une troncature leur est appliquée pour n'en conserver que les parties centrales. Pour cela, un ajustement à une gaussienne est réalisé pour chaque distribution. Les points considérés dans la suite sont alors ceux appartenant à un intervalle $[\bar{R} - \Delta R, \bar{R} + \Delta R]$ où $\bar{R}$ est le centre de la gaussienne obtenue et ΔR est fixé tel que l'intégrale de la distribution tronquée représente 98,5 % de l'intégrale de la distribution initiale. Une estimation des moyennes de ces distributions tronquées est alors obtenue ; ces dernières sont représentées sur la figure 3.18. Un écart est effectivement observé entre données réelles et simulées, il s'agit précisément de celui que la correction résiduelle absolue en p_T des jets doit corriger.



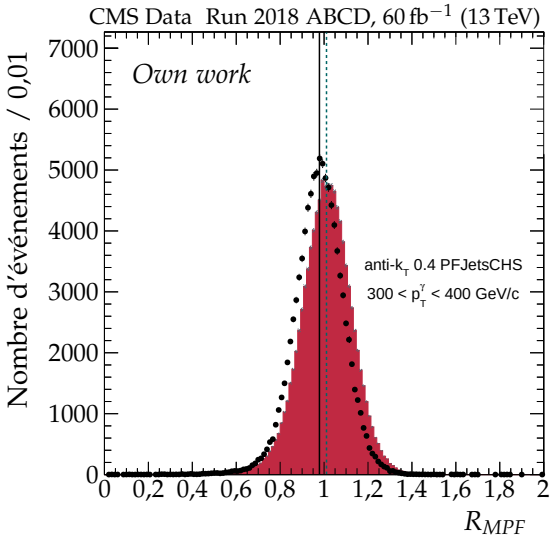
(a) *Réponse équilibrée pour* $p_T{}^{\gamma} \in [175, 230[$ GeV.



(b) *Réponse équilibrée pour* $p_T{}^{\gamma} \in [300, 400[$ GeV.



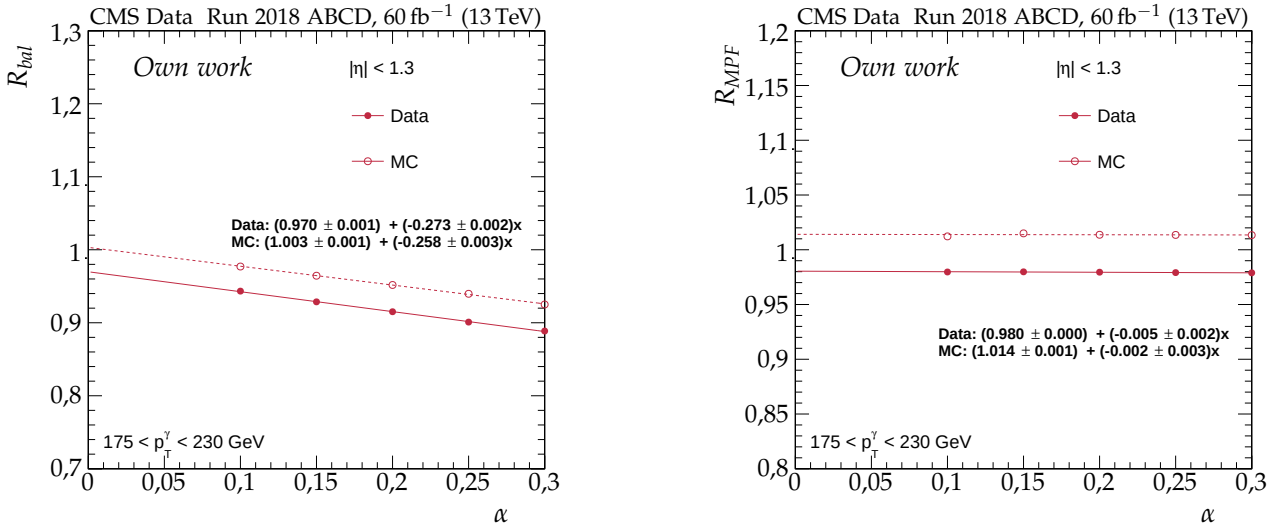
(c) *Réponse MPF pour* $p_T{}^{\gamma} \in [175, 230[$ GeV.



(d) *Réponse MPF pour* $p_T{}^{\gamma} \in [300, 400[$ GeV.

Figure 3.18 – *Réponses équilibrée et MPF dans les données réelles (points noirs) et simulées (histogramme en rouge) pour* $\alpha < 0{,}3$, $|\eta^{jet}| < 1{,}3$ *et deux intervalles de* $p_T{}^{\gamma}$ *en 2018.*

Extrapolation vers $\alpha = 0$ Une extrapolation vers $\alpha = 0$ est réalisée afin de s'affranchir de l'activité additionnelle des jets décrite dans la section 3.2. Les intervalles de α utilisés pour la JES sont présentés dans le tableau 3.6. L'utilisation des ces intervalles inclusifs permet une extrapolation linéaire en α, ce qui est réalisé sur la figure 3.19.



(a) *Réponse équilibrée.* **(b)** *Réponse MPF.*

Figure 3.19 – *Extrapolation vers $\alpha = 0$ de la réponse des jets pour $|\eta| < 1{,}3$ et $175 < p_T{}^\gamma < 230\,\text{GeV}$ en 2018.*

4.3 Résultats

La correction à appliquer aux données réelles, définie par la formule (3.8) d'après la démarche exposée dans la section 2.3, s'obtient en calculant la valeur moyenne des réponses R_{bal} ou R_{MPF} pour les données et les simulations dans chacun des intervalles de $p_T{}^\gamma$ défini dans le tableau 3.4 et de η^{jet} défini dans le tableau 3.5. Elle permet de ramener la réponse moyenne des jets dans les données réelles à celle constatée dans les simulations.

Les résultats ainsi obtenus à l'aide des méthodes équilibrée et MPF, avant et après extrapolation vers $\alpha = 0$, sont présentés dans les sections 4.3.1 et 4.3.2. Les distributions moyennes des réponses en fonction de $p_T{}^\gamma$ dans les données réelles et simulées, ainsi que leurs rapports, y sont représentés. Un ajustement constant est réalisé dans chaque intervalle de η^{jet} afin d'obtenir un ordre de grandeur de la correction à appliquer dans cet intervalle. La dépendance en p_T de la correction est déterminée grâce à un ajustement global réalisé avec les résultats d'autres analyses, présenté dans la section 4.3.3. Enfin, une vérification de la bonne mise en œuvre de la correction ainsi déterminée est présentée dans la section 4.3.4.

4.3.1 Résultats avant extrapolation

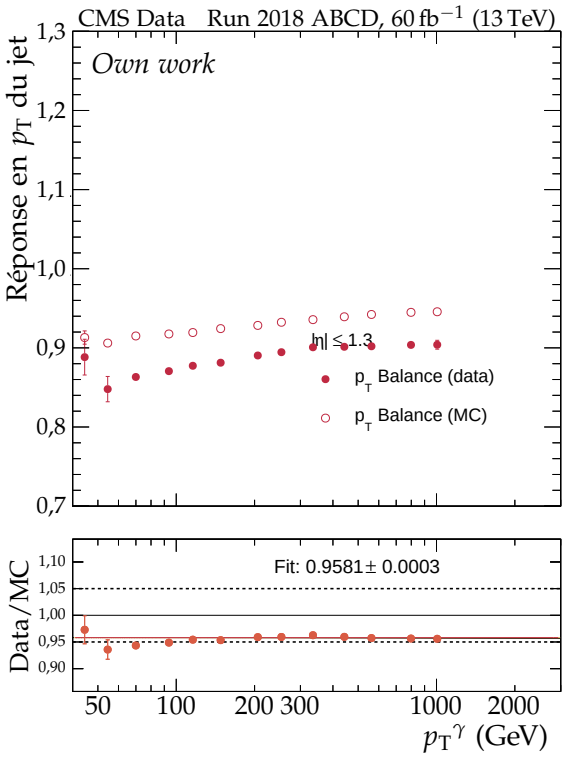
Les distributions des réponses équilibrées avant extrapolation se trouvent figure 3.20, page 104. Les corrections à appliquer sont de l'ordre de 4 % pour $|\eta^{\text{jet}}| < 1{,}3$, de 6 % pour $1{,}3 \leqslant |\eta^{\text{jet}}| < 2{,}5$ et de 3 % pour $|\eta^{\text{jet}}| > 2{,}5$.

Les distributions des réponses MPF avant extrapolation se trouvent figure 3.21, page 105. Les corrections à appliquer sont de l'ordre de 3 % pour $|\eta^{\text{jet}}| < 1{,}3$, de 5 % pour $1{,}3 \leqslant |\eta^{\text{jet}}| < 2{,}5$, soit environ 1 % de moins qu'avec la méthode équilibrée, et de 3 % pour $|\eta^{\text{jet}}| > 2{,}5$,

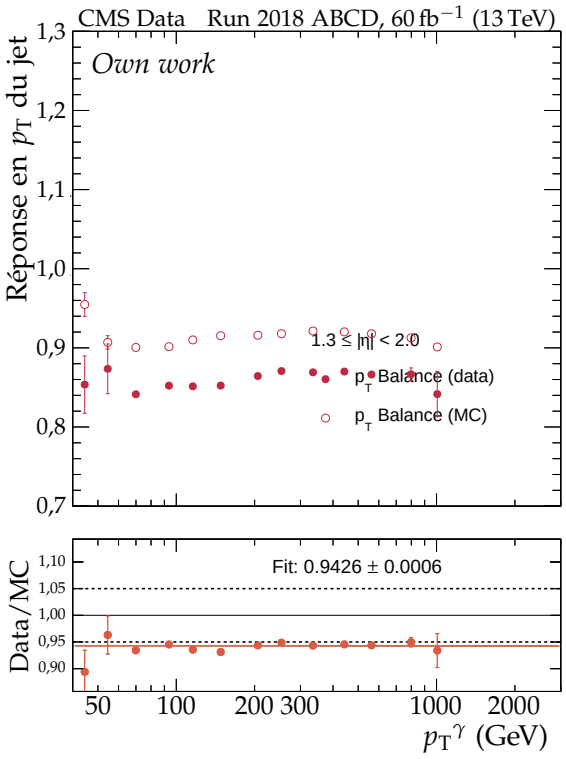
Il est à noter que ces résultats sont obtenus avant extrapolation vers $\alpha = 0$. Or, cette extrapolation a un effet beaucoup plus important sur la réponse équilibrée que sur la réponse MPF, comme le montre la figure 3.19.

4.3.2 Résultats après extrapolation

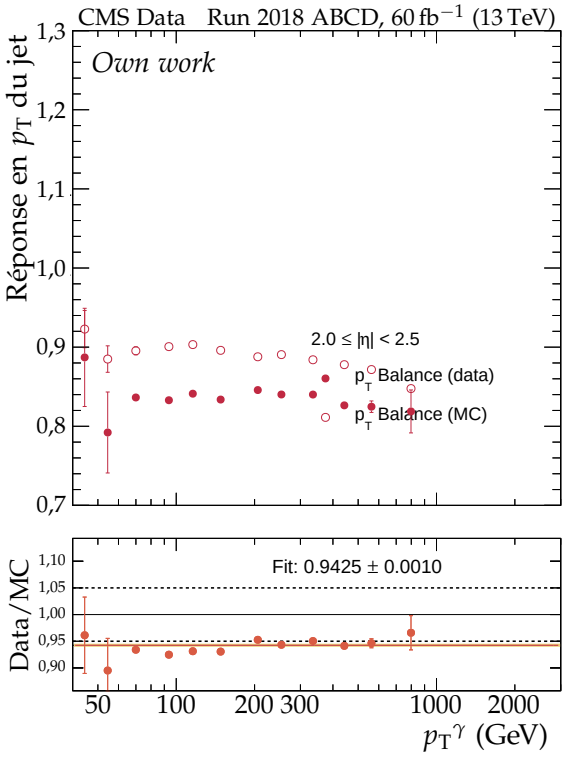
L'extrapolation des réponses vers $\alpha = 0$ est réalisée comme expliqué dans la section 4.2.



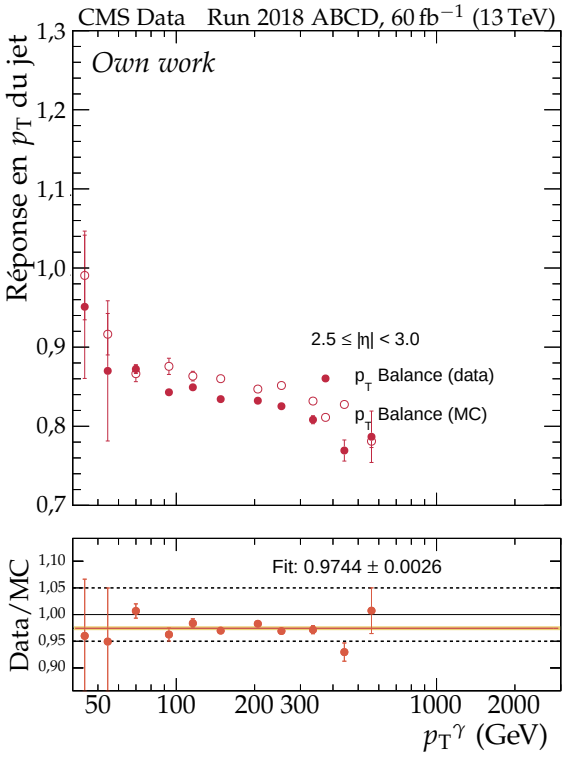
(a) $|\eta^{jet}| < 1{,}3$.



(b) $1{,}3 \leqslant |\eta^{jet}| < 2{,}0$.

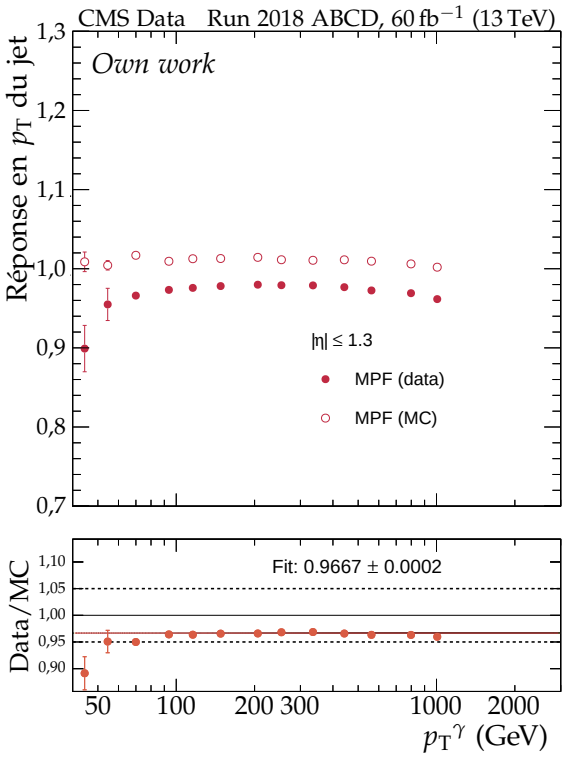


(c) $2{,}0 \leqslant |\eta^{jet}| < 2{,}5$.

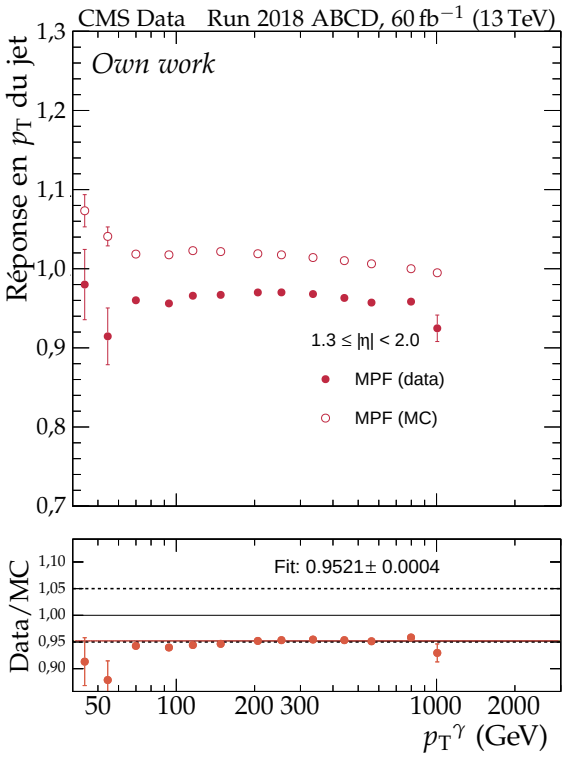


(d) $2{,}5 \leqslant |\eta^{jet}| < 3{,}0$.

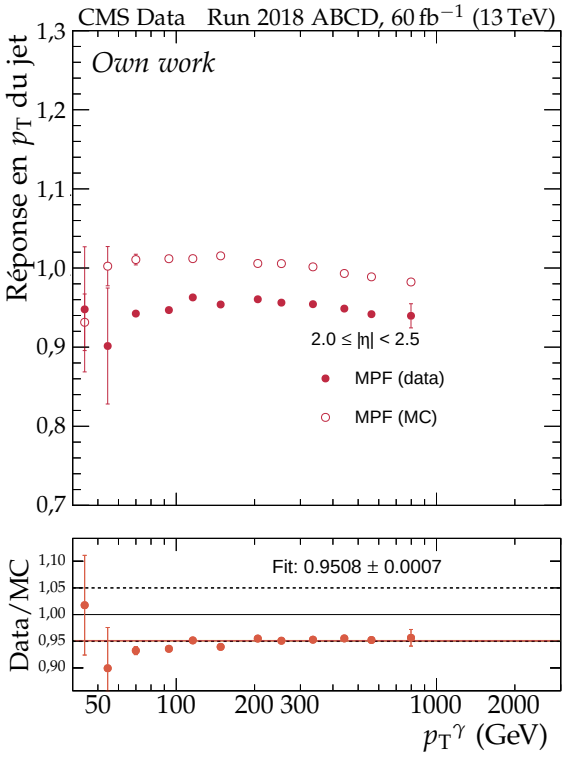
Figure 3.20 – *Distributions des réponses équilibrées moyennes en fonction de p_T^γ pour différents intervalles de η^{jet} en 2018 avant extrapolation. Le rapport données réelles sur simulées est présenté dans chaque cas ainsi qu'un ajustement à une constante donnant l'ordre de grandeur de la correction résiduelle à appliquer.*



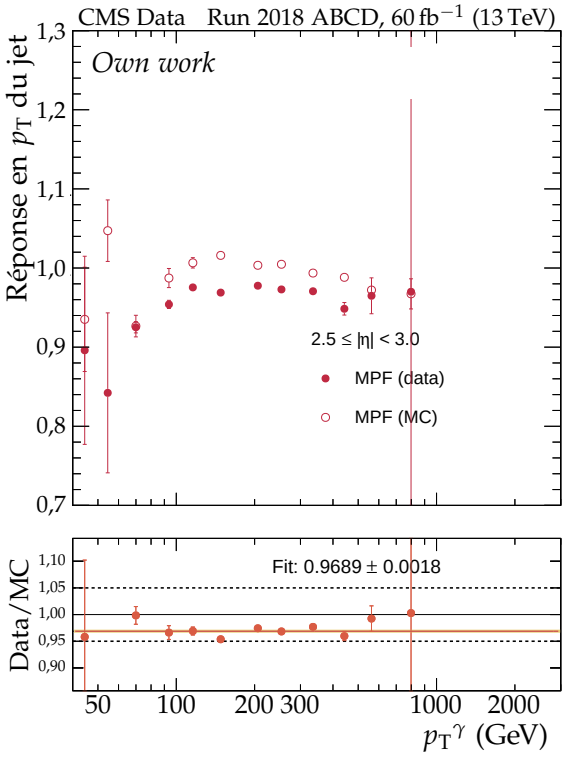
(a) $|\eta^{jet}| < 1{,}3$.



(b) $1{,}3 \leqslant |\eta^{jet}| < 2{,}0$.

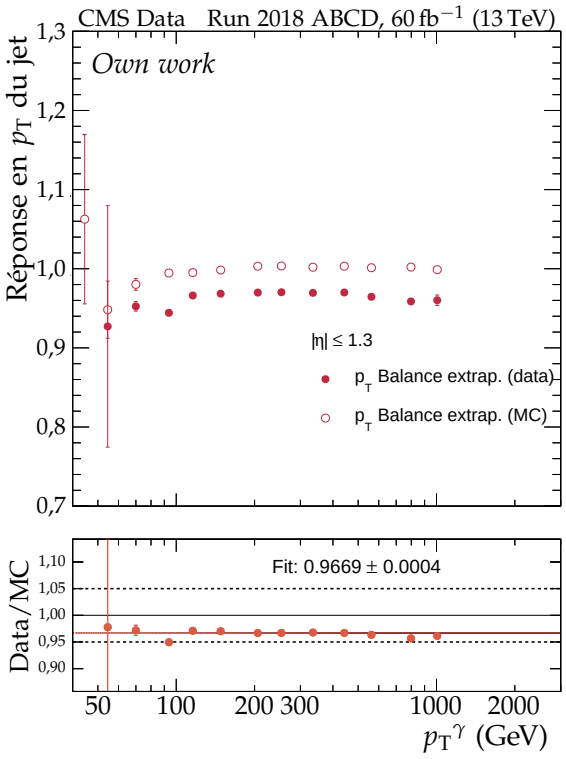


(c) $2{,}0 \leqslant |\eta^{jet}| < 2{,}5$.

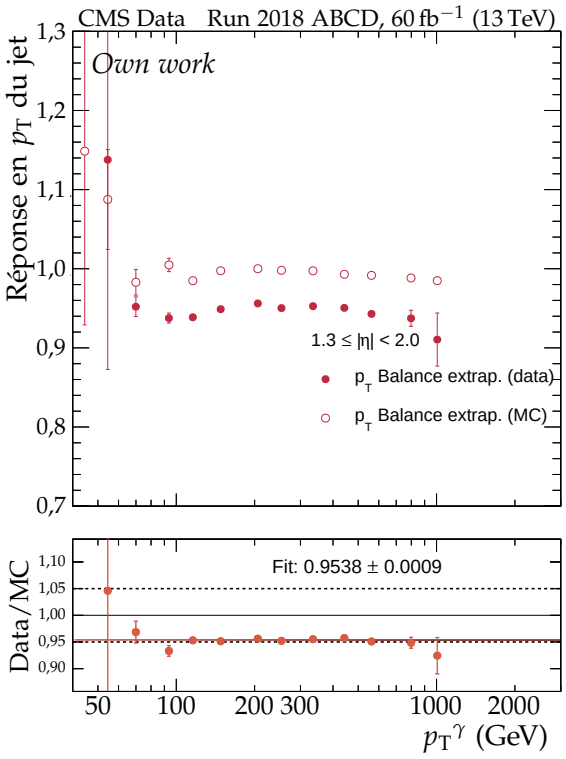


(d) $2{,}5 \leqslant |\eta^{jet}| < 3{,}0$.

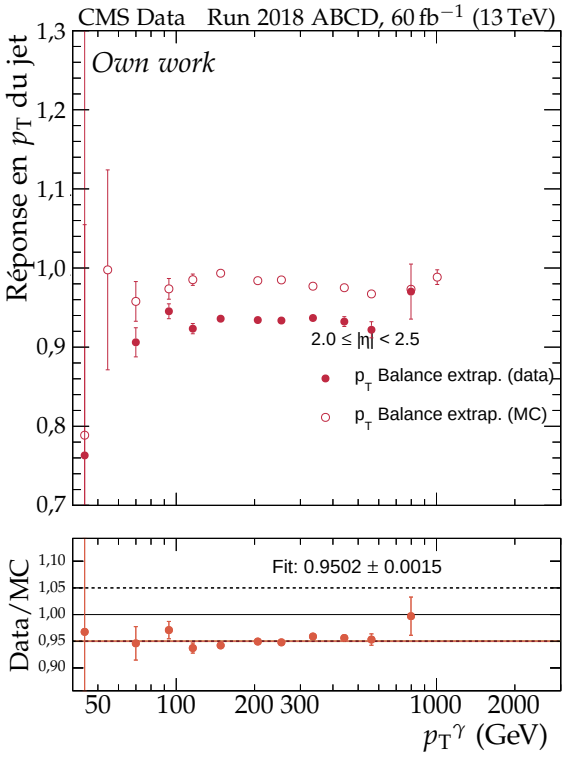
Figure 3.21 – *Distributions des réponses MPF moyennes en fonction de p_T^γ pour différents intervalles de η^{jet} en 2018 avant extrapolation. Le rapport données réelles sur simulées est présenté dans chaque cas ainsi qu'un ajustement à une constante donnant l'ordre de grandeur de la correction résiduelle à appliquer.*



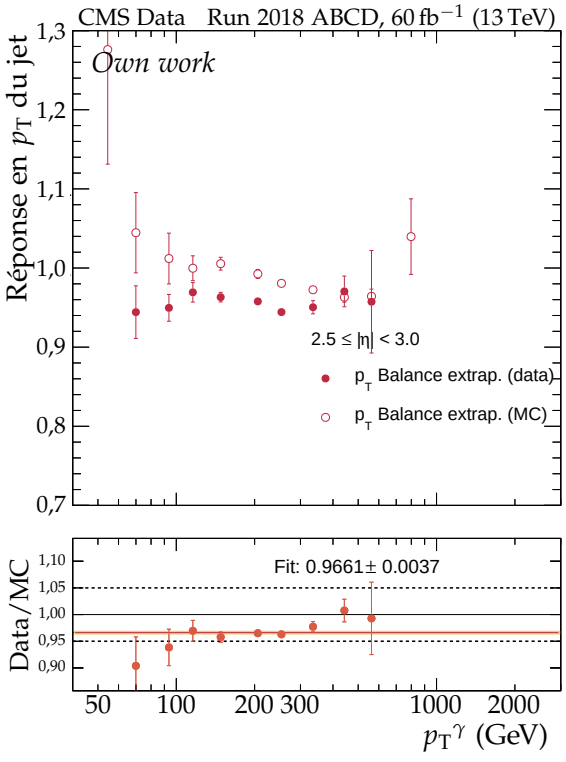
(a) $|\eta^{jet}| < 1{,}3$.



(b) $1{,}3 \leqslant |\eta^{jet}| < 2{,}0$.

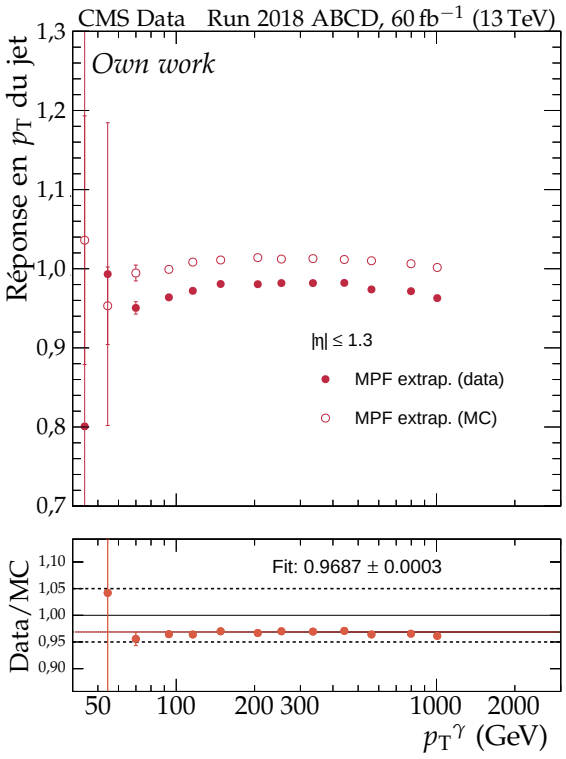


(c) $2{,}0 \leqslant |\eta^{jet}| < 2{,}5$.

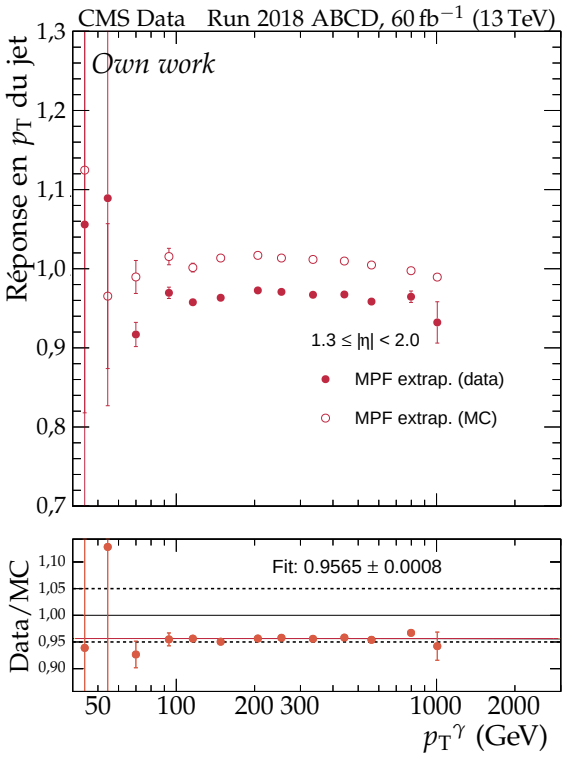


(d) $2{,}5 \leqslant |\eta^{jet}| < 3{,}0$.

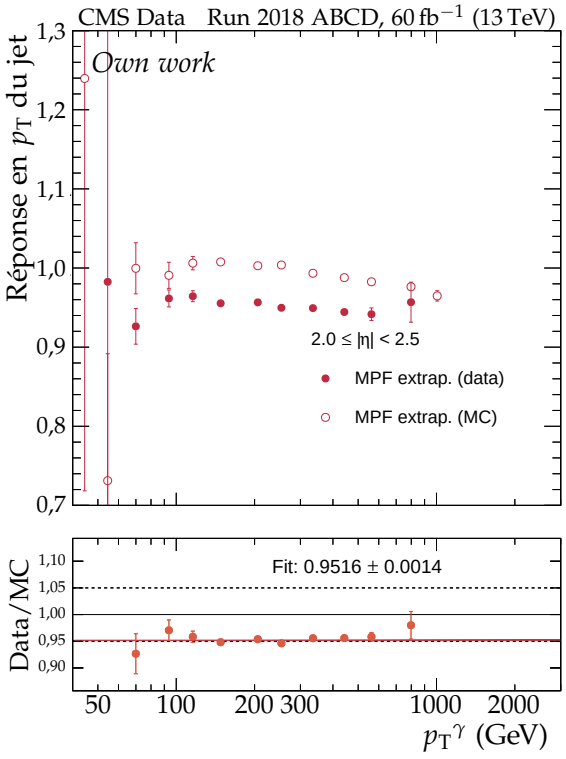
Figure 3.22 – *Distributions des réponses équilibrées moyennes en fonction de p_T^γ pour différents intervalles de η^{jet} en 2018 après extrapolation. Le rapport données réelles sur simulées est présenté dans chaque cas ainsi qu'un ajustement à une constante donnant l'ordre de grandeur de la correction résiduelle à appliquer.*



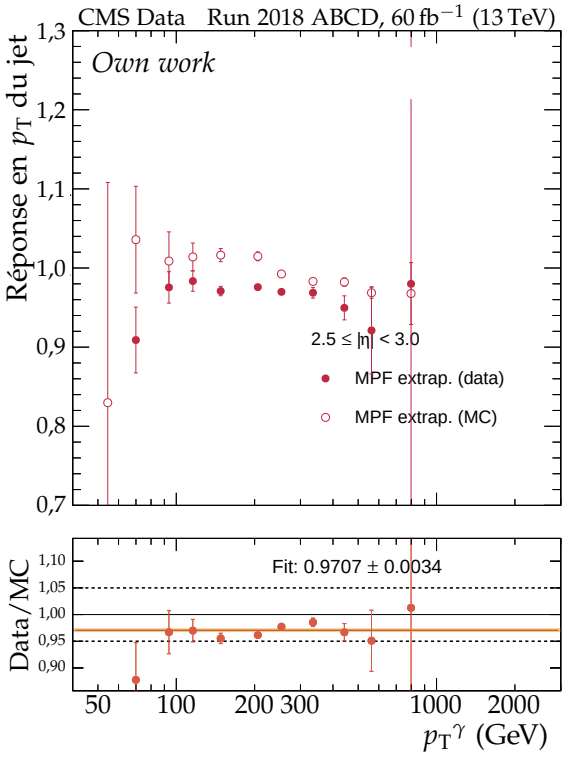
(a) $|\eta^{jet}| < 1{,}3$.



(b) $1{,}3 \leqslant |\eta^{jet}| < 2{,}0$.



(c) $2{,}0 \leqslant |\eta^{jet}| < 2{,}5$.



(d) $2{,}5 \leqslant |\eta^{jet}| < 3{,}0$.

Figure 3.23 – *Distributions des réponses MPF moyennes en fonction de p_T^γ pour différents intervalles de η^{jet} en 2018 après extrapolation. Le rapport données réelles sur simulées est présenté dans chaque cas ainsi qu'un ajustement à une constante donnant l'ordre de grandeur de la correction résiduelle à appliquer.*

Les distributions des réponses équilibrées après extrapolation se trouvent figure 3.22, page 106. Les corrections à appliquer sont de l'ordre de 3 % pour $|\eta^{\text{jet}}| < 1{,}3$, de 5 % pour $1{,}3 \leqslant |\eta^{\text{jet}}| < 2{,}5$, soit environ 1 % de moins qu'avant extrapolation, et de 3 % pour $|\eta^{\text{jet}}| > 2{,}5$.

Les distributions des réponses MPF après extrapolation se trouvent figure 3.23, page 107. Les corrections à appliquer sont de l'ordre de 3 % pour $|\eta^{\text{jet}}| < 1{,}3$, de 5 % pour $1{,}3 \leqslant |\eta^{\text{jet}}| < 2{,}5$ et de 3 % pour $|\eta^{\text{jet}}| > 2{,}5$, soit du même ordre qu'avant extrapolation. L'extrapolation a donc bien un effet très faible sur R_{MPF}.

Les valeurs des rapports données réelles sur simulées des réponses équilibrée et MPF obtenus sont résumés dans le tableau 3.7. L'extrapolation vers $\alpha = 0$ permet de rétablir l'accord entre les rapport des réponses équilibrée et MPF. Cet accord permet de valider l'utilisation de ces méthodes afin d'estimer la JES.

$\lvert\eta^{\text{jet}}\rvert \in$	Réponse équilibrée		Réponse MPF	
	$\alpha < 0{,}3$	$\alpha \to 0$	$\alpha < 0{,}3$	$\alpha \to 0$
$[0;1{,}3[$	$0{,}9581 \pm 0{,}0003$	$0{,}9669 \pm 0{,}0004$	$0{,}9667 \pm 0{,}0002$	$0{,}9687 \pm 0{,}0003$
$[1{,}3;2{,}0[$	$0{,}9426 \pm 0{,}0006$	$0{,}9538 \pm 0{,}0009$	$0{,}9521 \pm 0{,}0004$	$0{,}9565 \pm 0{,}0008$
$[2{,}0;2{,}5[$	$0{,}9425 \pm 0{,}0010$	$0{,}9502 \pm 0{,}0015$	$0{,}9508 \pm 0{,}0007$	$0{,}9516 \pm 0{,}0014$
$[2{,}5;3{,}0[$	$0{,}9744 \pm 0{,}0026$	$0{,}9661 \pm 0{,}0037$	$0{,}9689 \pm 0{,}0018$	$0{,}9707 \pm 0{,}0034$

Tableau 3.7 – *Rapports données réelles sur simulées des réponses équilibrée et MPF obtenus en 2018.*

4.3.3 Ajustement global

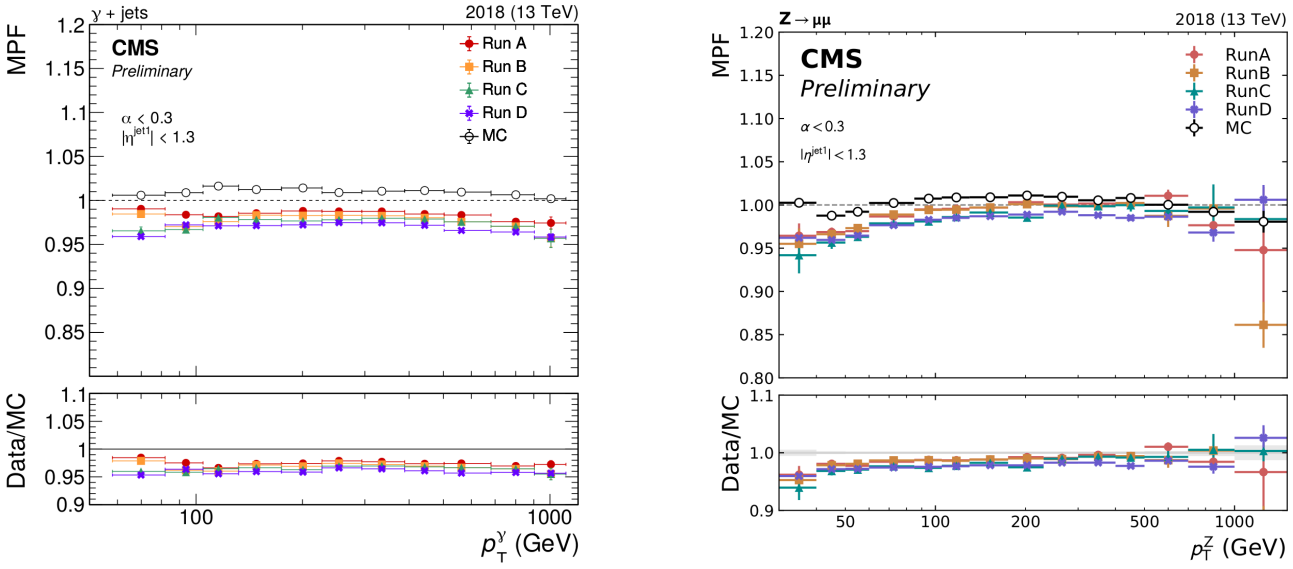
Les événements γ + jets ne permettent pas à eux seuls de couvrir avec une statistique suffisante l'ensemble de la gamme d'impulsions transverses à calibrer. De plus, l'utilisation de différentes catégories d'événements permet de valider *a posteriori* les résultats des analyses entre elles. Un ajustement global est alors réalisé sur les événements Z + jets, γ + jets et multijet afin d'obtenir la correction finale à appliquer aux données réelles.

Cet ajustement est réalisé en minimisant un χ^2 prenant en compte les contraintes de chaque catégorie d'événements. La correction résiduelle absolue en p_{T} des jets correspond ainsi à l'ajustement d'une fonction paramétrique. Les incertitudes présentes dans les différentes analyses sont considérées comme des paramètres de nuisance pour l'ajustement. Ces incertitudes sont :

— 4,6 % sur la section efficace de collision inélastique pp utilisée pour estimer les profils d'empilement ;
— les incertitudes de la JEC, décrites section 2.4 ;
— l'échelle en énergie des objets de référence, 0,2 % pour les photons et les muons, 0,5 % pour les électrons ;
— les effets de l'ISR et du FSR se retrouvant dans l'incertitude de l'extrapolation en α ;
— la propagation des calibrations des photons et des électrons dans l'énergie transverse manquante.

La figure 3.24 compare les résultats produits lors de ma thèse avec les événements γ + jets à ceux de l'analyse basée sur les événements $Z(\to \mu^+\mu^-)$ + jets pour l'année 2018. La réponse des jets dans les données réelles diminue du Run A au Run D dans les deux analyses, ce qui est dû à l'évolution des conditions d'acquisition des données au cours du temps. Le vieillissement du détecteur est une des sources de dépendance temporelle de la réponse des jets. La calibration en énergie des jets est ainsi déterminée à la fois pour une année entière, pour les différents *runs* individuellement et éventuellement pour des ensembles de *runs* successifs, ce qui permet d'améliorer la précision obtenue sur l'énergie des jets.

L'ajustement global sur les résultats des différentes analyses est illustré, pour les trois années du Run II, sur la figure 3.6. La correction résiduelle absolue en p_{T} des jets utilisée par la collaboration CMS est ainsi obtenue.

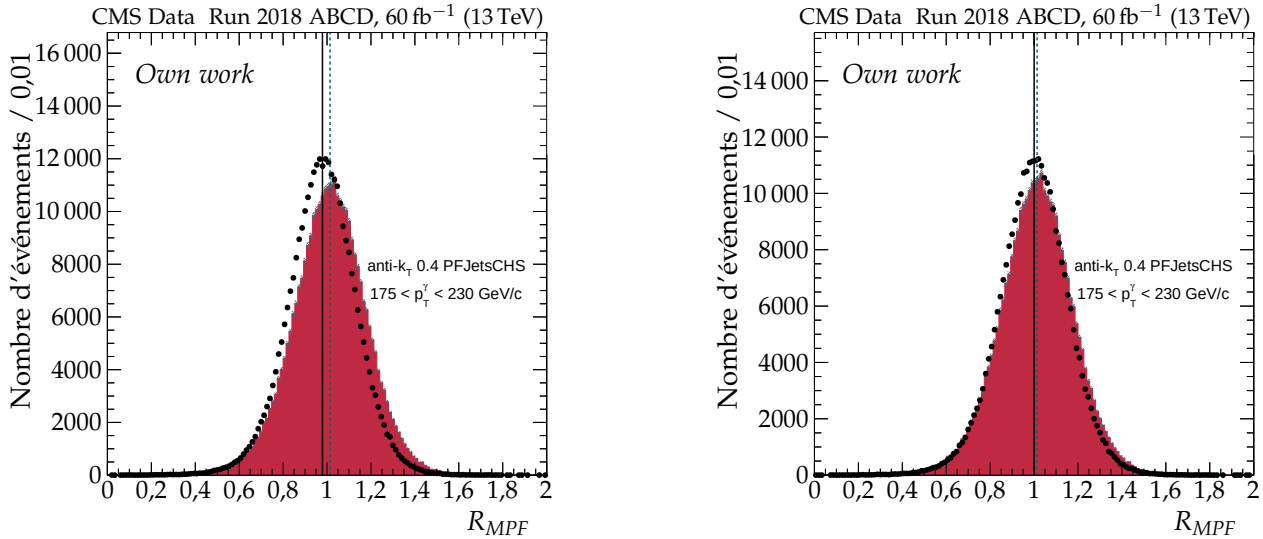


(a) *Avec les événements γ + jets.* **(b)** *Avec les événements $Z(\to \mu^+\mu^-)$ + jets.*

Figure 3.24 – *Distributions de la réponse MPF moyenne en fonction de p_T dans les événements avec $|\eta^{jet}| < 1{,}3$ et $\alpha < 0{,}3$ pour chaque période de prise de données et pour les simulations en 2018 [164].*

$\lvert\eta^{\text{jet}}\rvert \in$	Réponse équilibrée		Réponse MPF	
	avant $\mathcal{C}_{\text{Res}}$	après $\mathcal{C}_{\text{Res}}$	avant $\mathcal{C}_{\text{Res}}$	après $\mathcal{C}_{\text{Res}}$
$[0;1{,}3[$	$0{,}9669 \pm 0{,}0004$	$0{,}9867 \pm 0{,}0004$	$0{,}9687 \pm 0{,}0003$	$0{,}9877 \pm 0{,}0003$
$[1{,}3;2{,}0[$	$0{,}9538 \pm 0{,}0009$	$0{,}9739 \pm 0{,}0009$	$0{,}9565 \pm 0{,}0008$	$0{,}9753 \pm 0{,}0008$
$[2{,}0;2{,}5[$	$0{,}9502 \pm 0{,}0015$	$0{,}9698 \pm 0{,}0016$	$0{,}9516 \pm 0{,}0014$	$0{,}9724 \pm 0{,}0014$
$[2{,}5;3{,}0[$	$0{,}9661 \pm 0{,}0037$	$0{,}9884 \pm 0{,}0039$	$0{,}9707 \pm 0{,}0034$	$0{,}9922 \pm 0{,}0035$

Tableau 3.8 – *Rapports données réelles sur simulées des réponses équilibrée et MPF obtenus en 2018 après extrapolation vers $\alpha = 0$.*



(a) *Avant correction (figure 3.18c).* **(b)** *Après correction.*

Figure 3.25 – *Comparaison des réponses MPF avant et après correction résiduelle absolue pour $p_T{}^{\gamma} \in [175, 230[$ GeV et $|\eta| < 1{,}3$ en 2018.*

4.3.4 Test d'intégrité

Il est possible de vérifier que la correction résiduelle absolue en p_T des jets déterminée permet bien de rapprocher les réponses des jets entre données réelles et simulées. Pour cela, l'analyse est à nouveau réalisée en appliquant la correction résiduelle absolue en p_T des jets lors de leur calibration. Les valeurs des rapports données réelles sur simulées des réponses équilibrée et MPF obtenus avant et après utilisation de cette correction sont présentés dans le tableau 3.8. Ces rapports se rapprochent de 1, ce qui montre que la correction améliore l'accord données-simulations. Cette amélioration peut également se constater sur les distributions des réponses des jets, dont une comparaison est proposée sur la figure 3.25 où les deux distributions sont plus proches l'une de l'autre après correction complète.

5 Correction de la résolution en énergie avec les événements γ + jets

Déterminer la correction de la résolution en énergie des jets, ou JER, en 2018 et 2017-UL avec les événements γ + jets a été un des mes travaux de thèse. La méthode est sensiblement la même que pour déterminer la correction résiduelle absolue en p_T des jets, ou JES.

Dans le cas de la JES, la moyenne de la distribution des réponses des jets est corrigée. Pour la JER, c'est la largeur de cette distribution qui doit être corrigée. La sélection des événements est ainsi faite comme dans le cas de la JES décrite section 4.1, à ceci près que la correction résiduelle absolue en p_T des jets est appliquée.

5.1 Définition de la résolution en énergie des jets

La résolution en énergie des jets se détermine à l'aide de leur réponse équilibrée R_{bal}. À partir de la définition de R_{bal}, équation (3.12), il est possible d'écrire dans le cas des événements γ + jets

$$R_{bal} = \frac{p_{\mathrm{T}\,\mathrm{reco}}^{\text{jet 1}}}{p_{\mathrm{T}}{}^{\gamma}} = \frac{p_{\mathrm{T}\,\mathrm{reco}}^{\text{jet 1}}}{p_{\mathrm{T}\,\mathrm{reco}}^{\gamma}} = \frac{p_{\mathrm{T}\,\mathrm{reco}}^{\text{jet 1}}}{p_{\mathrm{T}\,\mathrm{ptcl}}^{\text{jet 1}}} \times \frac{p_{\mathrm{T}\,\mathrm{ptcl}}^{\text{jet 1}}}{p_{\mathrm{T}\,\mathrm{ptcl}}^{\gamma}} \times \frac{p_{\mathrm{T}\,\mathrm{ptcl}}^{\gamma}}{p_{\mathrm{T}\,\mathrm{reco}}^{\gamma}}\,, \tag{3.24}$$

ce qui se traduit en terme des largeurs des distributions de chacune de ces fractions sous la forme

$$\sigma_{R_{bal}} = \sigma\left(\frac{p_{\mathrm{T}\,\mathrm{reco}}^{\text{jet 1}}}{p_{\mathrm{T}\,\mathrm{ptcl}}^{\text{jet 1}}}\right) \oplus \sigma\left(\frac{p_{\mathrm{T}\,\mathrm{ptcl}}^{\text{jet 1}}}{p_{\mathrm{T}\,\mathrm{ptcl}}^{\gamma}}\right) \oplus \sigma\left(\frac{p_{\mathrm{T}\,\mathrm{ptcl}}^{\gamma}}{p_{\mathrm{T}\,\mathrm{reco}}^{\gamma}}\right)\,, \tag{3.25}$$

où $\oplus$ désigne une somme quadratique. Des termes de cette dernière équation, le premier rend compte de la résolution en énergie des jets au niveau reconstruit et est noté σ_{JER} dans la suite. Il s'agit de la grandeur d'intérêt dans cette analyse. Le second terme est lié à la physique de l'événement sous-jacent, c'est-à-dire de l'empilement, des émissions de particules secondaires et des neutrinos. Après extrapolation vers $\alpha = 0$, la contribution des émissions de particules secondaires devient négligeable. Ce terme est noté σ_{PLI} dans la suite ; « PLI » signifie interaction au niveau particule (*Particle Level Interaction*). Enfin, le dernier terme est lié à la résolution en énergie des photons, noté σ_γ.

L'équation (3.25) se réécrit alors, en utilisant les notations introduites,

$$\sigma_{R_{bal}} = \sigma_{\mathrm{JER}} \oplus \sigma_{\mathrm{PLI}} \oplus \sigma_\gamma\,, \tag{3.26}$$

ce qui peut se réarranger afin d'exprimer σ_{JER} sous la forme

$$\sigma_{\mathrm{JER}} = \sigma_{R_{bal}} \ominus \sigma_{\mathrm{PLI}} \ominus \sigma_\gamma\,. \tag{3.27}$$

La bonne qualité de reconstruction des photons permet de négliger le terme σ_γ dans la suite.

5.2 Analyse

Similitudes avec l'analyse menée pour la JES L'analyse des événements γ + jets dans le cas de la JER est semblable à celle pour la JES, décrite dans la section 4.2. Les intervalles de $p_T{}^{\gamma}$, $|\eta^{\text{jet}}|$ et α sont toutefois différents. Les intervalles de ces grandeurs utilisés pour la JER sont définis dans les tableaux 3.9, 3.10 et 3.11. En particulier, les intervalles de α sont exclusifs, contrairement aux intervalles inclusifs utilisés pour la JES.

[105, 130[	[130, 175[	[175, 200[	[200, 230[	[230, 300[
[300, 400[	[400, 500[	[500, 700[	[700, 3000[	

Tableau 3.9 – *Intervalles de p_T^γ en* GeV *utilisés pour la JER.*

[0,0; 0,522[	[0,522; 0,783[	[0,783; 1,131[	[1,131; 1,305[	[1,305; 1,740[
[1,740; 1,930[	[1,930; 2,043[	[2,043; 2,322[	[2,322; 2,5[	[2,5; 2,853[
	[2,853; 2,954[	[2,954; 3,139[	[3,139; 5,191[	

Tableau 3.10 – *Intervalles fins de $|\eta^{jet}|$ utilisés pour la JER.*

[0; 0,10[	[0,10; 0,15[	[0,15; 0,20[	[0,20; 0,25[	[0,25; 0,30[

Tableau 3.11 – *Intervalles de α utilisés pour la JER.*

Obtention de $\sigma_{R_{bal}}$ et σ_{PLI} pour $(p_T^\gamma, \eta^{\text{jet}}, \alpha)$ donnés Pour chaque domaine de p_T^γ défini dans le tableau 3.9, de η^{jet} défini dans le tableau 3.10 et de α défini dans le tableau 3.11, les distributions de la réponse équilibrée dans les données réelles et simulées sont déterminées.

Comme dans le cas de la JES, seuls les centres de ces distributions sont considérés afin de limiter les effets des leurs queues. Alors, $\sigma_{R_{bal}}$ s'obtient à partir des points restant comme étant le rapport de la variance de la distribution de ces points divisée par leur valeur moyenne.

Dans ces mêmes domaines de p_T^γ, η^{jet} et α, les distributions de $p_{T\,\text{ptcl}}^{\text{jet 1}}$ et $p_{T\,\text{ptcl}}^{\gamma}$ sont estimées à partir des événements simulés. Il est alors possible d'obtenir σ_{PLI}.

Extrapolation vers $\alpha = 0$ Une extrapolation vers $\alpha = 0$ est réalisée afin de s'affranchir de l'activité additionnelle des jets décrits dans la section 3.2. Les intervalles de α utilisés pour la JER sont présentés dans le tableau 3.11. L'extrapolation n'est plus linéaire comme dans le cas de la JES. Une fonction de la forme

$$\sigma_i = \sqrt{(a_i \alpha)^2 + b_i^2} \tag{3.28}$$

est utilisée pour $\sigma_{R_{bal}}^{\text{données}}$, $\sigma_{R_{bal}}^{\text{simulations}}$ et σ_{PLI}. L'extrapolation est illustrée sur la figure 3.26a.

Détermination de σ_{JER} en fonction de η^{jet} Dans chacun des intervalles de p_T^γ et η^{jet}, pour les données réelles et simulées, la contribution estimée de l'événement sous-jacent est soustraite en quadrature à $\sigma_{R_{bal}}$ afin d'obtenir σ_{JER} en suivant l'équation (3.27). Puis, un ajustement constant en p_T des rapports de σ_{JER} extrapolés à $\alpha = 0$ entre données réelles et simulées est déterminé dans chaque intervalle de η^{jet}, comme sur la figure 3.26b. La valeur de cet ajustement correspond au facteur à appliquer à la JER dans l'intervalle de η^{jet}.

Incertitudes Les incertitudes prises en compte dans la mesure de la JER sont :

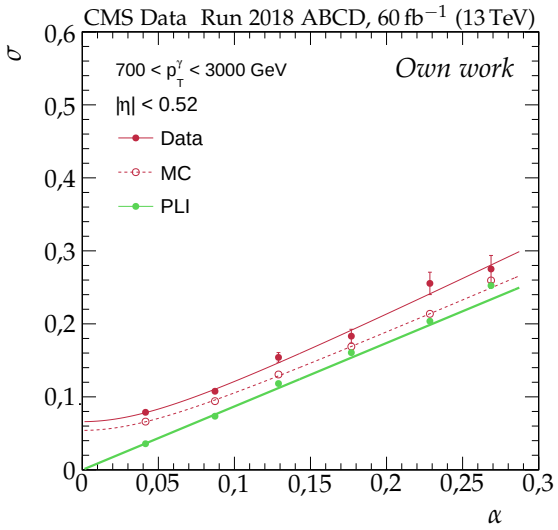
— 4,6 % sur la section efficace de collision inélastique pp utilisée pour estimer les profils d'empilement ;

— les incertitudes de la JEC, décrites section 2.4.

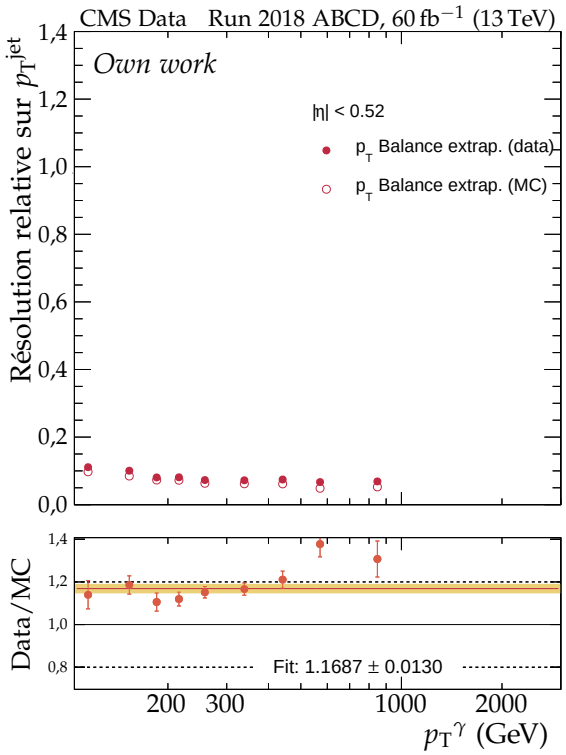
Les incertitudes sur l'échelle en énergie des photons ainsi que leur résolution sont négligées face aux autres incertitudes considérées.

5.3 Résultats

Les résultats issus de l'analyse des événements γ + jets pour l'année 2018 sont présentés sur la figure 3.27a. La combinaison avec l'analyses des événements dijet permet d'obtenir les facteurs correctifs utilisés par la collaboration, présentés sur la figure 3.27b. Ces facteurs sont de l'ordre de 1,2 dans le tonneau et peuvent atteindre 2,3 dans les bouchons.

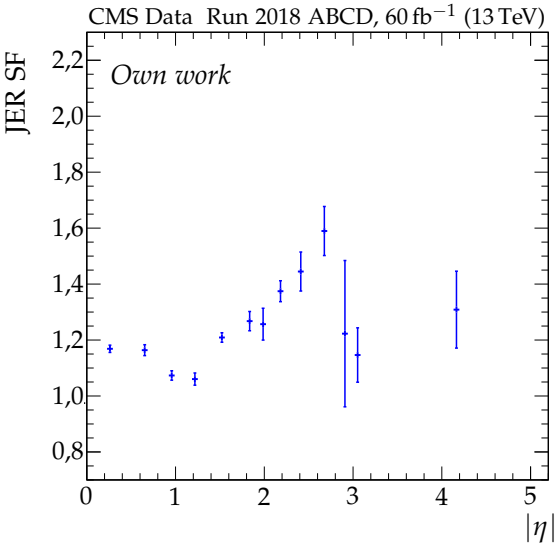


(a) *Extrapolation vers $\alpha = 0$ de la résolution des jets pour $|\eta| < 0{,}52$ et $700 < p_T{}^\gamma < 3000\,\text{GeV}$ en 2018.*

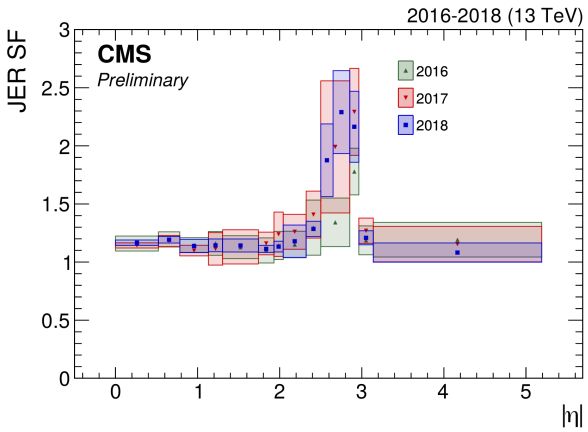


(b) *Résolution en énergie des jets extrapolée à $\alpha = 0$ pour $|\eta| < 0{,}52$ en 2018.*

Figure 3.26 – *Détermination de la résolution en énergie des jets.*



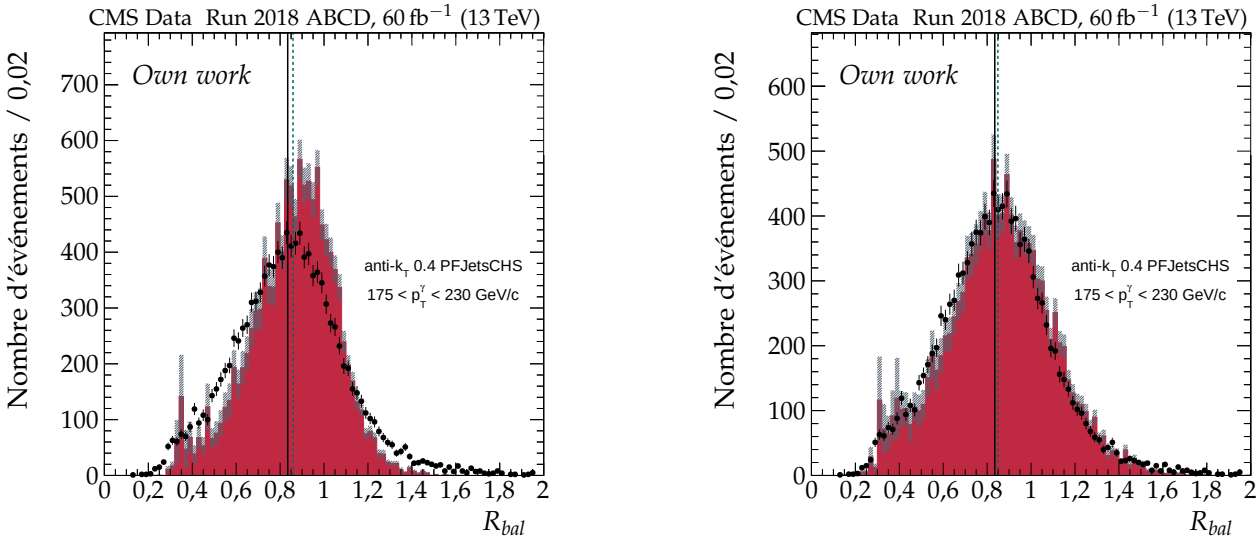
(a) *Facteurs correctifs déterminés avec les événements γ + jets en 2018.*



(b) *Facteurs correctifs utilisés par la collaboration lors du Run II [164].*

Figure 3.27 – *Facteurs correctifs de la résolution en énergie des jets.*

L'effet de l'application des facteurs correctifs de la résolution en énergie des jets est particulièrement visible dans la région $2{,}5 \leqslant |\eta| < 3{,}0$ où sa valeur est de l'ordre de 2. Les distributions des réponses équilibrées des jets dans les données réelles et simulées avant et après application de ces facteurs sont illustrées pour cette région sur la figure 3.28. Les réponses des jets dans les données réelles ne sont pas modifiées par la correction de la résolution en énergie des jets, ces distributions sont donc identiques dans les figures 3.28a et 3.28b.



(a) *Avant application des JER SF.*

(b) *Après application des JER SF.*

Figure 3.28 – *Comparaison des réponses équilibrées avant et après correction de la résolution en énergie des jets pour* $p_T{}^\gamma \in [175, 230[$ GeV *et* $2{,}5 \leqslant |\eta| < 3{,}0$ *en 2018.*

6 Conclusion

Ce chapitre a présenté la calibration en énergie des jets. Elle permet d'obtenir des jets de la meilleure qualité possible pour les analyses de physique menées par l'ensemble de la collaboration CMS. Il s'agit d'une approche factorisée de plusieurs corrections, chacune ayant pour but de corriger un effet en particulier.

L'obtention d'une de ces corrections pour les années 2018 et 2017-UL a fait partie de mon travail de thèse et a été développée plus en détail, ainsi que la phénoménologie des événements utilisés. Il s'agit d'événements γ + jets dans lesquels l'équilibre entre le photon et un jet permet d'estimer l'échelle en énergie du jet connaissant celle du photon. Pour l'année 2018, la réponse des jets dans ces événements est inférieure dans les données par rapport aux simulations de 3 % pour $|\eta^{\text{jet}}| < 1{,}3$ à 5 % pour $1{,}3 \leqslant |\eta^{\text{jet}}| < 2{,}5$. L'incertitude absolue sur ces mesures est inférieure à 0,3 %.

La résolution en énergie des jets doit également être corrigée. À l'aide d'une étude similaire, menée sur les mêmes événements et également détaillée dans ce chapitre, les facteurs d'échelle ont été obtenus lors de ma thèse pour les années 2018 et 2017-UL. Pour l'année 2018, la résolution en énergie des jets dans ces événements est inférieure dans les données par rapport aux simulations de (6 ± 3) % pour $1{,}131 \leqslant |\eta^{\text{jet}}| < 1{,}305$ à (60 ± 9) % pour $2{,}5 \leqslant |\eta^{\text{jet}}| < 2{,}8$.

Les jets sont omniprésents dans les collisions du LHC. Ces travaux permettant la bonne caractérisation des jets sont donc essentiels pour réaliser des analyses de qualité. Parmi elles se trouve celle présentée dans le chapitre 4.

Chapitre 4
Recherche de bosons de Higgs supplémentaires de haute masse se désintégrant en paire de leptons τ

Sommaire

1 Introduction . . . 116
2 Corrections . . . 117
2.1 Efficacité des chemins de déclenchement . . . 117
2.2 Impulsions des particules générées . . . 118
3 Sélection des événements . . . 119
3.1 Sélection pour l'enregistrement des données . . . 120
3.2 Obtention du dilepton et sélection finale . . . 120
3.3 Sélection des jets . . . 124
4 Modélisation du bruit de fond . . . 125
4.1 Méthode des données encapsulées ou *embedding* . . . 127
4.2 Estimation du bruit de fond QCD dans le canal $e\mu$ . . . 129
4.3 Méthode des facteurs de faux ou *fake factors* . . . 130
5 Catégorisation des événements et variables discriminantes . . . 135
5.1 Catégories « BSM » . . . 135
5.2 Catégories « SM » . . . 138
5.3 Combinaison des catégories SM et BSM . . . 141
6 Incertitudes systématiques . . . 142
6.1 Incertitudes de normalisation . . . 142
6.2 Incertitudes de forme . . . 143
7 Extraction du signal . . . 146
7.1 Modèle de vraisemblance . . . 146
7.2 Méthode CL_s . . . 149
7.3 Limites indépendantes du modèle . . . 150
7.4 Interprétation dans les scénarios du MSSM . . . 150
8 Résultats et interprétations . . . 152
8.1 Limites indépendantes du modèle . . . 152
8.2 Limites du scénario M_h^{125} . . . 153
8.3 Limites du scénario $M_{H_1}^{125}$ avec violation de CP . . . 154
9 Conclusion . . . 154

1 Introduction

Dans le chapitre 1, il a été montré que le modèle standard (SM, *Standard Model*) souffre de lacunes quant à l'explication à apporter à certaines observations. Des modèles allant au-delà (BSM, *Beyond Standard Model*), comme l'extension supersymétrique minimale du modèle standard ou « MSSM », peuvent combler certaines d'entre elles. Le MSSM prédit l'existence de cinq bosons de Higgs, dont trois neutres, h, H et A. L'un d'entre eux doit correspondre au boson découvert en 2012 et interprété comme étant le boson de Higgs du SM [22-26]. L'existence des deux bosons de Higgs neutres supplémentaires peut être testée expérimentalement avec des accélérateurs de particules, comme cela a été fait au LEP [198]. Ces bosons se désintègrent préférentiellement en paire de quarks b ou de leptons τ. Bien que le rapport de branchement ($\mathcal{BR}$) de ces bosons aux b soit 5 à 10 fois supérieur que celui aux τ, ces derniers offrent une meilleure accessibilité expérimentale dans les collisionneurs hadroniques comme le Tevatron, où ces désintégrations en τ ont été étudiées [199, 200].

L'expérience CMS installée au LHC et présentée dans le chapitre 2 permet elle aussi de tester expérimentalement le MSSM, dans des conditions de collision inédites. La recherche de bosons de Higgs supplémentaires se désintégrant en paire de leptons τ a été menée dans les collisions de protons avec une énergie dans le centre de masse de $\sqrt{s} = 7$ et 8 TeV (Run I) [80, 201-203] ainsi qu'avec les données récoltées en 2016 avec une énergie de $\sqrt{s} = 13$ TeV [81]. La désintégration en paire de b est également exploitée [33, 204], ainsi que celle en paire de muons [205]. L'expérience ATLAS mène des recherches similaires [206-208].

Ce chapitre présente la recherche de bosons de Higgs supplémentaires de haute masse se désintégrant en paire de τ avec les données récoltées par l'expérience CMS lors du Run II du LHC (années 2016, 2017 et 2018), correspondant à une luminosité intégrée de 137 fb^{-1} ($35{,}9 + 41{,}5 + 59{,}7\,\text{fb}^{-1}$) à une énergie dans le centre de masse de $\sqrt{s} = 13$ TeV. Sur les six canaux de désintégration de la paire de leptons τ introduits dans le chapitre 1, les quatre présentant les plus grands $\mathcal{BR}$ sont considérés dans l'analyse. Il s'agit des canaux hadronique ($\tau_\text{h}\tau_\text{h}$), semi-leptoniques ($\mu\tau_\text{h}$, $e\tau_\text{h}$) et leptonique asymétrique ($e\mu$). Les canaux leptoniques symétriques ($\mu\mu$, ee) ne sont pas exploités.

Dans les données réelles, les particules doivent forcément être reconstruites à partir des signaux qu'elles produisent dans le détecteur. Dans le cas des données simulées, la réponse du détecteur aux particules est modélisée. À partir des signaux réels comme simulés, les particules individuelles et les objets physiques de haut niveau sont reconstruits comme exposé dans le chapitre 2. Les simulations n'étant pas exemptes de défauts, des corrections déterminées à l'aide d'analyses annexes leur sont appliquées. Les corrections génériques sont présentées au chapitre 2, à l'exception de la calibration en énergie des jets détaillée dans le chapitre 3. Les corrections spécifiques à la présente analyse sont introduites dans la section 2. Les objets reconstruits et corrigés permettent de sélectionner les événements d'intérêt pour l'analyse selon la procédure explicitée en section 3. Des processus physiques différents de ceux du signal recherché passent cette sélection et constituent le bruit de fond. Afin d'interpréter les observations, il est nécessaire de modéliser ce bruit de fond. Cette modélisation est présentée section 4. En plus de l'utilisation de données simulées, des techniques basées sur les données réelles sont exploitées. Des données dites « encapsulées » (*embedded*) sont ainsi produites selon la procédure exposée section 4.1 et décrivent les événements contenant une vraie paire de leptons τ. Une estimation du bruit de fond dû aux jets identifiés à tort comme des taus hadroniques (faux τ_h) est quant à elle obtenue grâce à la méthode des facteurs de faux (*fake factors*) introduite section 4.3. Les événements sont par la suite catégorisés afin d'augmenter la sensibilité de l'analyse. Les catégories utilisées sont présentées en section 5. Les sources d'incertitudes systématiques sont données section 6. Leur prise en compte dans l'extraction du signal ainsi que la modélisation de celui-ci sont exposées dans la section 7. Enfin, les résultats obtenus sont disponibles section 8. Certains sont indépendants de tout modèle, d'autres sont obtenus dans le cadre de scénarios spécifiques du MSSM [79].

Une note d'analyse [209] interne à la collaboration CMS est déjà disponible et un article est en préparation [210]. Ces travaux sont réalisés au sein d'une équipe regroupant :

— l'Institut de Physique des 2 Infinis (IP2I) de l'Université Claude Bernard de Lyon ;
— l'*Institut für Experimentelle Teilchenphysik* (ETP) du *Karlsruher Institut für Technologie* (KIT) de Karlsruhe ;
— le *Deutsches Elektronen-Synchrotron* (DESY) de Hambourg ;

— l'*Imperial College* de Londres ;
— l'*Institut für Hochenergiephysik* (HEPHY) de Vienne.

En début de thèse, j'ai travaillé sur les données de l'année 2017 en équipe avec Gaël TOUQUET qui a exploité le canal $\tau_h\tau_h$ dans sa thèse [211]. Je me suis concentré sur les canaux semi-leptoniques et plus particulièrement le canal $\mu\tau_h$ [212]. Les événements étaient analysés à l'aide d'un code basé sur HEPPY [213], indépendant de celui utilisé par les autres instituts listés précédemment, ce qui a permis à l'ensemble des acteurs de cette analyse de valider la bonne implémentation des différentes corrections et sélections détaillées dans ce chapitre. À cette occasion, j'ai découvert une erreur dans le code de COMBINE, l'outil de combination statistique utilisé par la collaboration CMS et basé sur ROOSTATS [214]. Cette erreur a été comprise et corrigée comme présenté dans la section 7.1.3. Le correctif [215] a été transmis à la collaboration CMS qui l'a pris en compte.

J'ai par la suite travaillé directement avec le groupe de Karlsruhe dans le cadre de l'analyse du Run II. J'ai implémenté le traitement du scénario avec violation de la symétrie CP et participé au traitement des jeux de données utilisés, listés dans l'annexe E. Il s'agissait de s'assurer du bon déroulement de plusieurs milliers de tâches informatiques et du regroupement de leurs résultats. Enfin, j'ai activement participé à la rédaction de la note d'analyse CMS [209] et de la publication [210] correspondantes. J'ai ainsi apporté une contribution significative à ces travaux.

2 Corrections

Dans le but d'améliorer la description des données réelles par les données simulées et encapsulées, des corrections obtenues à partir d'analyses annexes leur sont appliquées. Outre les corrections usuelles introduites au chapitre 2, des corrections spécifiques à cette analyse, présentées ci-après, sont également appliquées.

2.1 Efficacité des chemins de déclenchement

Efficacité des chemins de déclenchement des muons et des électrons (*lepton trigger scale factors*) L'efficacité des chemins de déclenchement des leptons (muons et électrons) est déterminée dans les données réelles et simulées en fonction de l'impulsion transverse et de la pseudo-rapidité du lepton par une méthode « balise et sonde » (*tag and probe*) sur des événements $Z \to \ell\ell$ [216] dans lesquels ℓ est un électron ou bien un muon. La balise, un des deux ℓ, permet de déterminer la quantité totale d'événements. La sonde est l'autre ℓ. L'efficacité ϵ d'un chemin de déclenchement « électron seul » ou « muon seul » alors est donnée par

$$\epsilon = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}} \tag{4.1}$$

avec N_{pass} le nombre d'événement dans lesquels balise et sonde activent le chemin de déclenchement et N_{fail} où seule la balise l'active. Ces critères sont donnés dans la référence [216].

Les efficacités des chemins de déclenchement « muon et τ_h » et « électron et τ_h » sont déterminées par combinaison avec celles des « τ_h seul » tel qu'exposé ci-après.

Efficacité des chemins de déclenchement des τ_h (τ_h *trigger scale factors*) L'efficacité des chemins de déclenchement est mesurée à partir d'une méthode similaire à celle utilisée pour la détermination de l'efficacité d'identification des τ_h [217]. Les événements exploités sont ceux contenant un boson W hautement virtuel, tel que $m_{W^*} \geqslant 200\,\text{GeV}$, se désintégrant selon $W^* \to \tau\nu_\tau$. Il est requis que le W^* soit produit avec une faible impulsion transverse, ce qui permet de rejeter les jets directement issus de l'événement dur. Les événements sélectionnés comportent ainsi un τ_h d'impulsion équilibrée avec $\vec{E}_{\text{T}}^{\text{miss}}$. Le bruit de fond principal provient des événements dans lesquels un jet est identifié à tort comme un τ_h. Il est estimé par la méthode des facteurs de faux décrite en section 4.3. Les autres bruits de fond sont estimés par des simulations.

Les données réelles sont sélectionnées à partir d'un chemin de déclenchement basé sur $E_{\text{T}}^{\text{miss}}$ [209, 217]. Les modélisations du signal $W^* \to \tau\nu_\tau$ et des bruits de fond sont ajustées à ces données réelles

par COMBINE. Puis, l'estimation des bruits de fond leur est retirée afin d'obtenir uniquement la contribution des événements $W^* \to \tau\nu_\tau$. Les quantités d'événements $W^* \to \tau\nu_\tau$ dans les données réelles et simulées sont ainsi obtenues.

Les chemins de déclenchement « τ_h seul » utilisés dans l'analyse sont donnés en annexe F. Leur efficacité ϵ est donnée par

$$\epsilon = \frac{N_{\text{pass}}}{N_{\text{total}}} \tag{4.2}$$

avec N_{pass} le nombre d'événements $W^* \to \tau\nu_\tau$ dans lesquels le τ_h correspond à l'objet physique ayant activé au moins un des chemins de déclenchement « τ_h seul » utilisés et N_{total} le nombre total d'événements. Le facteur d'échelle correctif SF à appliquer aux événements simulés est ainsi

$$SF = \frac{\epsilon(\text{données réelles})}{\epsilon(\text{données simulées})}\,. \tag{4.3}$$

Les efficacités et le facteur d'échelle sont déterminés en fonction de l'impulsion transverse du τ_h. L'efficacité des chemins de déclenchement « double τ_h » est obtenue comme étant le produit des efficacités pour chacun des deux τ_h.

L'efficacité combinée des chemins de déclenchement « τ_h seul » et « muon seul » dans le canal $\mu\tau_h$ (« électron seul » dans le canal $e\tau_h$) est obtenue selon

$$\epsilon = \epsilon(1\tau_h) + \epsilon(1\ell) - \epsilon(1\tau_h) \times \epsilon(1\ell) \tag{4.4}$$

où ℓ correspond au muon (à l'électron). Dans le cas du canal $\tau_h\tau_h$, la présence de deux τ_h impose une formule plus complexe pour déterminer l'efficacité combinée des chemins de déclenchement « τ_h seul » et « double τ_h »,

$$\begin{aligned}\epsilon = {} & \epsilon(2\tau_h) + \epsilon(\tau_h 1) + \epsilon(\tau_h 2) \\ & - \epsilon(2\tau_h + \tau_h 1) - \epsilon(2\tau_h + \tau_h 1) - \epsilon(\tau_h 1 + \tau_h 2) \\ & + \epsilon(2\tau_h + \tau_h 1 + \tau_h 2)\,,\end{aligned} \tag{4.5}$$

où $2\tau_h$ signifie « double τ_h », $\tau_h 1$ « τ_h seul » appliqué au τ_h de plus haut p_T et $\tau_h 2$ « τ_h seul » appliqué à l'autre τ_h.

2.2 Impulsions des particules générées

Repondération de l'impulsion transverse et de la masse du boson Z (*DY p_T-mass reweighting*) Les impulsions transverses ainsi que la masse invariante des leptons issus de la désintégration du boson Z sont corrigées dans les événements simulés Drell-Yan. Ces corrections sont déterminées dans une région de contrôle $Z \to \mu\mu$ et n'introduisent pas de modification du nombre total d'événements.

Repondération de l'impulsion transverse du quark t (*top p_T reweighting*) La modélisation du bruit de fond $t\bar{t}$ est corrigée afin que les données simulées au NLO correspondent au NNLO. Pour cela, la distribution en p_T des quarks t est pondérée. Le poids à appliquer à un quark t, déterminé par le groupe $t\bar{t}H$, s'exprime en fonction de l'impulsion transverse du quark t en GeV selon

$$\omega = \exp\left(0{,}088 - 8{,}7\times10^{-4} \times p_T + 9{,}2\times10^{-7} \times {p_T}^2\right). \tag{4.6}$$

Le poids total à appliquer aux événements $t\bar{t}$ contenant deux quarks t est alors

$$\omega(\text{total}) = \sqrt{\omega(1) \times \omega(2)}\,. \tag{4.7}$$

Repondération de l'impulsion transverse du boson de Higgs Au premier ordre non nul (LO, *Leading Order*), les propriétés cinématiques du signal ne dépendent que de la masse du boson de Higgs, ce qui est couvert par la variété des jeux de données utilisés, listés dans l'annexe E. Cependant, à l'ordre supérieur (NLO, *Next-to Leading Order*), ce n'est plus vrai dans le cas du processus ggH, illustré figure 4.1.

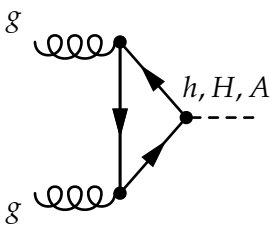


Figure 4.1 – *Diagramme de Feynman de production de boson de Higgs dans le cadre du MSSM par fusion de gluons (ggH).*

La boucle fermionique du diagramme comporte des contributions provenant des quarks, les plus importantes étant celles des quarks t et b ainsi que de leur interférence notée tb. Pour chacune d'entre elles, les distributions des impulsions transverses des bosons de Higgs (h, H ou A) ne sont pas dépendantes de $\tan\beta$, paramètre introduit dans le chapitre 1. Toutefois, les proportions de ces contributions le sont. Ainsi, les propriétés cinématiques des bosons de Higgs et par conséquent celles des leptons τ varient avec $\tan\beta$.

L'effet de $\tan\beta$ étant uniquement dû aux proportions des contributions des quarks à la boucle, il est possible d'obtenir les distributions en p_{T} des bosons de Higgs au NLO à partir des neuf contributions $(h, H, A) \times (t, b, tb)$ pour chaque point de masse utilisé selon les méthodes introduites dans les références [218, 219].

D'une part, une simulation de référence est réalisée par point de masse. La génération de ces événements est faite au NLO à l'aide du module `gg_H_2HDM` de POWHEG [180], les gerbes partoniques, l'hadronisation et l'événement sous-jacent sont simulés par PYTHIA [179] et la modélisation du détecteur est traitée par GEANT4 [184-186]. Cette simulation se fait dans le cadre d'un modèle général à deux doublets de Higgs (2HDM), dont le MSSM est un cas particulier comme exposé dans le chapitre 1. POWHEG permet alors d'obtenir les neuf contributions $(h, H, A) \times (t, b, tb)$. En principe, toutes valeurs des paramètres α et $\tan\beta$ peuvent être utilisées. En pratique, pour éviter d'obtenir un terme d'interférence presque nul menant à de faibles statistiques, ces paramètres sont fixés à $\alpha = \pi/4$ et $\tan\beta = 15$.

D'autre part, des simulations annexes sont réalisées pour différentes valeurs de $\tan\beta$, également avec le module `gg_H_2HDM` de POWHEG, mais uniquement au niveau générateur c'est-à-dire sans propagation dans le détecteur. Pour chaque valeur de $\tan\beta$, les distributions en p_{T} des neuf contributions considérées sont pondérées dans la simulation de référence de manière à correspondre à celles obtenues dans la simulation annexe correspondante.

Le signal complet du MSSM est obtenu à partir des contributions individuelles et pondérées du 2HDM utilisé ($\alpha = \pi/4$, $\tan\beta = 15$) selon

$$\begin{aligned}\sigma_{\mathrm{MSSM}} = &\left(\frac{y_{t,\mathrm{MSSM}}}{y_{t,\mathrm{2HDM}}}\right)^2 \sigma_{t,\mathrm{2HDM}}(Q_t) + \left(\frac{y_{b,\mathrm{MSSM}}}{y_{b,\mathrm{2HDM}}}\right)^2 \sigma_{b,\mathrm{2HDM}}(Q_b) \\ &+ \left(\frac{y_{t,\mathrm{MSSM}}}{y_{t,\mathrm{2HDM}}}\frac{y_{b,\mathrm{MSSM}}}{y_{b,\mathrm{2HDM}}}\right)\left[\sigma_{t+b,\mathrm{2HDM}}(Q_{tb}) - \sigma_{t,\mathrm{2HDM}}(Q_{tb}) - \sigma_{b,\mathrm{2HDM}}(Q_{tb})\right] \end{aligned} \tag{4.8}$$

où σ peut correspondre à la section efficace inclusive ou différentielle selon une variable donnée, y_t et y_b sont les constantes de couplage de Yukawa pour les quarks t et b introduites dans le chapitre 1 et Q l'échelle d'énergie [218, 219]. Les trois termes de cette formule correspondent aux contributions t, b et tb. Les valeurs de y_t et y_b dépendent de m_A et $\tan\beta$ et sont définies pour chacun des bosons de Higgs (h, H, A).

3 Sélection des événements

Cette section décrit la sélection des événements dans la région de signal. D'autres régions, de contrôle, d'application ou de détermination, sont également utilisées dans l'analyse, en particulier pour la méthode des facteurs de faux présentée section 4.3. Leurs définitions diffèrent généralement peu de celle de la région de signal, il s'agit souvent du changement d'un des critères de sélection. Ces changements sont explicités lors de la première évocation de ces régions.

3.1 Sélection pour l'enregistrement des données

Les chemins de déclenchement (HLT *paths*) permettent une présélection en temps réel des événements observés à CMS afin de les enregistrer, comme abordé dans le chapitre 2. Des listes de chemins de déclenchement sont ainsi définies, pour chacune des trois années (2016, 2017 et 2018) et chacun des états finaux considérés ($\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ et $e\mu$). Elles sont données dans l'annexe F. Les événements, une fois enregistrés sur décision d'au moins un des chemins de déclenchement, sont sélectionnés plus précisément selon la procédure détaillée dans la section 3.2.

3.2 Obtention du dilepton et sélection finale

La phénoménologie des événements $H \to \tau\tau$ est décrite dans le chapitre 1. Les leptons τ peuvent se désintégrer hadroniquement en tau hadronique (τ_h) ou leptoniquement en électron (e) ou en muon (μ). Ces désintégrations s'accompagnent de l'émission de un (cas hadronique) ou deux (cas leptoniques) neutrinos. Il existe ainsi six états finaux différents ou canaux pour ces événements dont seulement quatre sont considérés dans l'analyse : le canal hadronique ($\tau_h\tau_h$), les deux canaux semi-leptoniques ($\mu\tau_h$, $e\tau_h$) et un canal leptonique ($e\mu$).

Les produits de désintégration visibles des leptons τ sont notés L_1 et L_2. La correspondance de L_1 et L_2 aux τ_h, muons ou électrons suit l'ordre de notation du canal, par exemple $\mu\tau_h = L_1L_2$. Dans le cas d'un canal symétrique ($\tau_h\tau_h$, $\mu\mu$, ee), ${p_T}^{L_1} > {p_T}^{L_2}$. Pour L_1 comme L_2, une liste de candidats est obtenue à partir des particules reconstruites auxquelles sont appliquées des coupures détaillées dans les sections qui suivent pour chacun des canaux.

Des dileptons, c'est-à-dire des paires L_1L_2, sont formés à partir de ces deux listes de candidats. Pour être compatibles avec le cas de figure $H \to \tau\tau \to L_1L_2$, L_1 et L_2 doivent être de charges électriques opposées. Il est de plus requis que L_1 et L_2 soient séparés d'une distance ΔR dans le plan (η, ϕ) supérieure à 0,5 dans les canaux $\tau_h\tau_h$, $\mu\tau_h$ et $e\tau_h$ et supérieure à 0,3 dans le canal $e\mu$. Si aucun dilepton L_1L_2 n'est formé, l'événement est rejeté. Si plusieurs dileptons peuvent être formés, un seul d'entre eux est sélectionné selon la procédure suivante :

1. préférer une paire avec L_1 le plus isolé possible (L_2 pour le canal $e\mu$) ;
2. en cas d'égalité, préférer une paire avec L_1 de plus haut p_T (L_2 pour le canal $e\mu$) ;
3. en cas d'égalité, préférer une paire avec L_2 le plus isolé possible (L_1 pour le canal $e\mu$) ;
4. en cas d'égalité, préférer une paire avec L_2 de plus haut p_T (L_1 pour le canal $e\mu$).

Il est nécessaire de s'assurer, à cause de l'utilisation de canaux différents, qu'un événement donné ne peut être sélectionné dans le traitement de plusieurs canaux. Pour cela, après avoir sélectionné le dilepton, des vetos sur la présence de leptons supplémentaires à ceux du dilepton sont appliqués. Ces vetos sont explicités pour chaque canal dans les sections ci-après et sont au moins aussi lâches que les coupures de sélection les plus lâches des leptons de signal, c'est-à-dire ceux utilisés pour les dileptons, parmi tous les canaux. Ainsi, si un lepton est sélectionné dans un canal pour former un dilepton, alors il déclenche forcément le veto correspondant dans les autres canaux.

3.2.1 Canal $\tau_h\tau_h$

Sélection des taus hadroniques Tout τ_h respectant les critères listés ci-après est retenu pour jouer le rôle de L_1 ou L_2 dans le dilepton :

— ${p_T}^{\tau_h} > 40\,\text{GeV}$;
— $|\eta^{\tau_h}| < 2{,}1$;
— $d_z < 0{,}2\,\text{cm}$ avec d_z la distance longitudinale entre la trace principale du τ_h et le vertex primaire ;
— passer le discriminateur `NewDecayModeFinding` (modes de désintégration 5, 6, et 7 interdits) ;
— passer les points de fonctionnement :
  — *very very loose* du discriminateur `deepTau anti-electron`,
  — *very loose* du discriminateur `deepTau anti-muon`,
  — *medium* du discriminateur `deepTau vs jet`.

Correspondance du dilepton avec les chemins de déclenchement Les deux τ_h de la paire sélectionnée doivent correspondre, le cas échéant, aux objets ayant activé un des chemins de déclenchement

utilisés pour enregistrer l'événement. Les objets considérés pour chacun des chemins de déclenchement sont donnés dans l'annexe F. La correspondance est établie lorsque la particule reconstruite et l'objet du chemin de déclenchement sont séparés de moins de 0,5 dans le plan (η, ϕ), c'est-à-dire $\Delta R < 0{,}5$. Trois catégories sont définies pour les événements enregistrés en 2016 (respectivement 2017 et 2018) :

— bas p_T : ${p_T}^{\tau_h 1} < 120\,(180)$ GeV, ${p_T}^{\tau_h 2} < 120\,(180)$ GeV. Seuls les chemins de déclenchement « double τ_h » sont considérés pour L_1 et L_2 ;
— moyen p_T : ${p_T}^{\tau_h 1} > 120\,(180)$ GeV, ${p_T}^{\tau_h 2} < 120\,(180)$ GeV. Une combinaison logique « ou » des chemins de déclenchement « τ_h seul » et « double τ_h » est considérée pour L_1 et seulement les « double τ_h » pour L_2 ;
— haut p_T : ${p_T}^{\tau_h 1} > 120\,(180)$ GeV, ${p_T}^{\tau_h 2} > 120\,(180)$ GeV. Une combinaison logique « ou » des chemins de déclenchement « τ_h seul » et « double τ_h » est considérée pour L_1 et L_2.

Les τ_h ne sont considérés comme correspondant aux objets HLT qu'à condition que leurs impulsions soient supérieures d'au moins 5 GeV au seuil d'activation du chemin de déclenchement.

Vetos de leptons supplémentaires Les vetos de leptons supplémentaires doivent être respectés, c'est-à-dire que l'événement ne contient pas :

— de muon tel que ${p_T}^{\mu} > 10$ GeV, $|\eta^{\mu}| < 2{,}4$, passant le point de fonctionnement *medium* du *muonID* et d'isolation $I^{\mu} < 0{,}3\,{p_T}^{\mu}$;
— d'électron tel que ${p_T}^{e} > 10$ GeV, $|\eta^{e}| < 2{,}5$, passant le point de fonctionnement à 90 % d'efficacité de l'*electron ID MVA* et d'isolation $I^{e} < 0{,}3\,{p_T}^{e}$.

3.2.2 Canal $\mu\tau_h$

Sélection des muons Tout muon respectant les critères listés ci-après est retenu pour jouer le rôle de L_1 dans le dilepton :

— ${p_T}^{\mu} > 23\,(25)$ GeV en 2016 (2017, 2018) et correspondant à l'objet HLT du chemin de déclenchement « muon seul » ou $20\,(21) < {p_T}^{\mu} \leqslant 23\,(25)$ GeV en 2016 (2017, 2018) et correspondant à l'objet HLT de type muon du chemin de déclenchement « muon et τ_h » ;
— $|\eta^{\mu}| < 2{,}1$;
— paramètres d'impact $d_z < 0{,}2$ cm et $d_{xy} < 0{,}045$ cm ;
— $I^{\mu} < 0{,}15\,{p_T}^{\mu}$;
— passer le point de fonctionnement *medium* du *muonID*.

Sélection des taus hadroniques Tout τ_h respectant les critères listés ci-après est retenu pour jouer le rôle de L_2 dans le dilepton :

— ${p_T}^{\tau_h} > 30$ GeV ;
— $|\eta^{\tau_h}| < 2{,}3$;
— $d_z < 0{,}2$ cm avec d_z la distance longitudinale entre la trace principale du τ_h et le vertex primaire ;
— passer le discriminateur `NewDecayModeFinding` (modes de désintégration 5, 6, et 7 interdits) ;
— passer les points de fonctionnement :
  — *very very loose* du discriminateur `deepTau anti-electron`,
  — *tight* du discriminateur `deepTau anti-muon`,
  — *medium* du discriminateur `deepTau vs jet`.

Correspondance du dilepton avec les chemins de déclenchement Le muon et le τ_h de la paire sélectionnée doivent correspondre, le cas échéant, aux objets ayant activé un des chemins de déclenchement utilisés pour enregistrer l'événement. Les objets considérés pour chacun des chemins de déclenchement sont donnés dans l'annexe F. La correspondance est établie lorsque la particule reconstruite et l'objet du chemin de déclenchement sont séparés de moins de 0,5 dans le plan (η, ϕ), c'est-à-dire $\Delta R < 0{,}5$. Deux catégories sont définies pour les événements enregistrés en 2016 (respectivement 2017 et 2018) :

— bas p_T : ${p_T}^{\tau_h} < 120\,(180)$ GeV en 2016 (2017, 2018). Une combinaison logique « ou » des chemins de déclenchement « muon seul » et « muon et τ_h » est considérée ;

— haut p_T : ${p_T}^{\tau_h} > 120\,(180)$ GeV en 2016 (2017, 2018). Une combinaison logique « ou » des chemins de déclenchement « muon seul » et « τ_h seul » est considérée.

Dans le cas de l'utilisation du chemin de déclenchement « muon et τ_h », le τ_h doit de plus :

— correspondre à l'objet HLT ;
— vérifier ${p_T}^{\tau_h} > 25\,(32)$ GeV en 2016 (2017, 2018).

Masse transverse du muon La masse transverse du muon, définie par

$$m_T^{\mu} = m_T(\mu, {E_T}^{\text{miss}}) = \sqrt{2\,{p_T}^{\mu}\,{E_T}^{\text{miss}}\,(1 - \cos\Delta\phi)} \tag{4.9}$$

avec $\Delta\phi = \phi^{\mu} - \phi^{{E_T}^{\text{miss}}}$ doit vérifier $m_T < 70$ GeV. Cette coupure permet de s'assurer que la région de signal soit orthogonale à la région de détermination des facteurs de faux des événements W + jets tout en rejetant ce bruit de fond dans la région de signal. Les facteurs de faux sont abordés dans la section 4.3.

Vetos de leptons supplémentaires Les vetos de leptons supplémentaires doivent être respectés, c'est-à-dire que l'événement ne contient pas :

— de second muon tel que ${p_T}^{\mu} > 10$ GeV, $|\eta^{\mu}| < 2{,}4$, passant le point de fonctionnement *medium* du *muonID* et d'isolation $I^{\mu} < 0{,}3\,{p_T}^{\mu}$;
— d'électron tel que ${p_T}^{e} > 10$ GeV, $|\eta^{e}| < 2{,}5$, passant le point de fonctionnement à 90 % d'efficacité de l'*electron ID MVA* et d'isolation $I^{e} < 0{,}3\,{p_T}^{e}$;
— de paire de muons de charges opposées avec $\Delta R > 0{,}15$, tous deux vérifiant ${p_T}^{\mu} > 15$ GeV, $|\eta^{\mu}| < 2{,}4$, passant le point de fonctionnement *loose* du *muonID*, de paramètres d'impact $d_z <$ 0,2 cm et $d_{xy} < 0{,}045$ cm et d'isolation $I^{\mu} < 0{,}3\,{p_T}^{\mu}$.

3.2.3 Canal $e\tau_h$

Sélection des électrons Tout électron respectant les critères listés ci-après est retenu pour jouer le rôle de L_1 dans le dilepton :

— ${p_T}^{e} > 26$ GeV en 2016, 28 en 2017 et 33 en 2018 et correspondant à l'objet HLT du chemin de déclenchement « électron seul » ou ${p_T}^{e}$ entre 25 GeV et la valeur précédente et correspondant à l'objet HLT de type électron du chemin de déclenchement « électron et τ_h » ;
— $|\eta^{e}| < 2{,}1$;
— paramètres d'impact $d_z < 0{,}2$ cm et $d_{xy} < 0{,}045$ cm ;
— $I^{e} < 0{,}15\,{p_T}^{e}$;
— passer le point de fonctionnement à 90 % d'efficacité de l'*electron ID MVA*.

Sélection des taus hadroniques Tout τ_h respectant les critères listés ci-après est retenu pour jouer le rôle de L_2 dans le dilepton :

— ${p_T}^{\tau_h} > 30$ GeV,
— $|\eta^{\tau_h}| < 2{,}3$,
— $d_z < 0{,}2$ cm avec d_z la distance longitudinale entre la trace principale du τ_h et le vertex primaire,
— passer le discriminateur `NewDecayModeFinding` (modes de désintégration 5, 6, et 7 interdits),
— passer les points de fonctionnement :
  — *tight* du discriminateur `deepTau anti-electron`,
  — *very loose* du discriminateur `deepTau anti-muon`,
  — *medium* du discriminateur `deepTau vs jet`.

Correspondance du dilepton avec les chemins de déclenchement L'électron et le τ_h de la paire sélectionnée doivent correspondre, le cas échéant, aux objets ayant activé un des chemins de déclenchement utilisés pour enregistrer l'événement. Les objets considérés pour chacun des chemins de déclenchement sont donnés dans l'annexe F. La correspondance est établie lorsque la particule reconstruite et l'objet du chemin de déclenchement sont séparés de moins de 0,5 dans le plan (η, ϕ),

c'est-à-dire $\Delta R < 0{,}5$. Deux catégories sont définies pour les événements enregistrés en 2016 (respectivement 2017 et 2018) :

— bas p_{T} : ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} < 120\,(180)$ GeV en 2016 (2017, 2018). Une combinaison logique « ou » des chemins de déclenchement « électron seul » et « électron et τ_{h} » est considérée ;
— haut p_{T} : ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} > 120\,(180)$ GeV en 2016 (2017, 2018). Une combinaison logique « ou » des chemins de déclenchement « électron seul » et « τ_{h} seul » est considérée.

Dans le cas de l'utilisation du chemin de déclenchement « électron et τ_{h} », le τ_{h} doit de plus :

— correspondre à l'objet HLT ;
— vérifier ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} > 25\,(35)$ GeVGeV en 2016 (2017, 2018) ;
— vérifier $|\eta^{\tau_{\mathrm{h}}}| < 2{,}1$.

Masse transverse de l'électron La masse transverse de l'électron, définie par

$$m_{\mathrm{T}}^{e} = m_{\mathrm{T}}(e, {E_{\mathrm{T}}}^{\mathrm{miss}}) = \sqrt{2\,{p_{\mathrm{T}}}^{e}\,{E_{\mathrm{T}}}^{\mathrm{miss}}\,(1 - \cos\Delta\phi)} \tag{4.10}$$

avec $\Delta\phi = \phi^{e} - \phi^{{E_{\mathrm{T}}}^{\mathrm{miss}}}$ doit vérifier $m_{\mathrm{T}} < 70\,\mathrm{GeV}$. Cette coupure permet de s'assurer que la région de signal soit orthogonale à la région de détermination des facteurs de faux des événements $W+$jets tout en rejetant ce bruit de fond dans la région de signal. Les facteurs de faux sont abordés dans la section 4.3.

Vetos de leptons supplémentaires Les vetos de leptons supplémentaires doivent être respectés, c'est-à-dire que l'événement ne contient pas :

— de muon tel que ${p_{\mathrm{T}}}^{\mu} > 10\,\mathrm{GeV}$, $|\eta^{\mu}| < 2{,}4$, passant le point de fonctionnement *medium* du *muonID* et d'isolation $I^{\mu} < 0{,}3\,{p_{\mathrm{T}}}^{\mu}$;
— de second électron tel que ${p_{\mathrm{T}}}^{e} > 10\,\mathrm{GeV}$, $|\eta^{e}| < 2{,}5$, passant le point de fonctionnement à 90 % d'efficacité de l'*electron ID MVA* et d'isolation $I^{e} < 0{,}3\,{p_{\mathrm{T}}}^{e}$;
— de paire d'électrons de charges opposées avec $\Delta R > 0{,}15$, tous deux vérifiant ${p_{\mathrm{T}}}^{e} > 15\,\mathrm{GeV}$, $|\eta^{e}| < 2{,}5$, passant le *cut-based veto ID*, de paramètres d'impact $d_z < 0{,}2\,\mathrm{cm}$ et $d_{xy} < 0{,}045\,\mathrm{cm}$ et d'isolation $I^{e} < 0{,}3\,{p_{\mathrm{T}}}^{e}$.

3.2.4 Canal $e\mu$

Sélection des muons Tout muon respectant les critères listés ci-après est retenu pour jouer le rôle de L_1 dans le dilepton :

— ${p_{\mathrm{T}}}^{\mu} > 15\,\mathrm{GeV}$;
— $|\eta^{\mu}| < 2.4$;
— paramètres d'impact $d_z < 0{,}2\,\mathrm{cm}$ et $d_{xy} < 0{,}045\,\mathrm{cm}$;
— $I^{\mu} < 0{,}2\,{p_{\mathrm{T}}}^{\mu}$;
— passer le point de fonctionnement *medium* du *muonID*.

Sélection des électrons Tout électron respectant les critères listés ci-après est retenu pour jouer le rôle de L_2 dans le dilepton :

— ${p_{\mathrm{T}}}^{e} > 15\,\mathrm{GeV}$;
— $|\eta^{e}| < 2.4$;
— paramètres d'impact $d_z < 0{,}2\,\mathrm{cm}$ et $d_{xy} < 0{,}045\,\mathrm{cm}$;
— $I^{e} < 0{,}15\,{p_{\mathrm{T}}}^{e}$;
— passer le point de fonctionnement à 90 % d'efficacité de l'*electron ID MVA* sans utilisation des variables d'isolation ;
— présenter moins de deux points de passage manquants dans le trajectographe ;
— passer le veto d'électron de conversion.

Correspondance du dilepton avec les chemins de déclenchement L'électron et le muon de la paire sélectionnée doivent correspondre, le cas échéant, aux objets ayant activé un des chemins de déclenchement utilisés pour enregistrer l'événement. Les objets considérés pour chacun des chemins de déclenchement sont donnés dans l'annexe F. La correspondance est établie lorsque la particule reconstruite et l'objet du chemin de déclenchement sont séparés de moins de 0,5 dans le plan (η, ϕ), c'est-à-dire $\Delta R < 0{,}5$.

Cas de la période 2016GH Pour les *runs* G et H de l'année 2016, le filtre en d_z des chemins de déclenchement utilisés pour le canal $e\mu$ n'est pas modélisé dans les données simulées. Il ne peut donc être appliqué que sur les données réelles. Afin de prendre en compte ce filtre manquant, un autre moins strict, sans le critère portant sur d_z, est appliqué sur les données simulées qui sont renormalisées selon l'efficacité du filtre sur d_z manquant.

La mesure sur un jeu de données simulées $t\bar{t}$ avec un électron et un muon dans l'état final donne une efficacité de 95,3 %. Aucune dépendance en p_{T} ou η de l'électron ou du muon n'a été observée. Pour les *runs* B à F, ce filtre en d_z n'est pas utilisé. Ainsi, sur l'ensemble de l'année 2016, le facteur $SF(d_z)$ à appliquer aux événements simulés vaut

$$SF(d_z) = \frac{\mathcal{L}_{\text{B-F}}}{\mathcal{L}_{\text{B-H}}} + 0{,}953\,\frac{\mathcal{L}_{\text{G-H}}}{\mathcal{L}_{\text{B-H}}} = 0{,}979 \tag{4.11}$$

avec $\mathcal{L}_x$ la luminosité intégrée sur la période x, notion introduite dans le chapitre 2.

Coupure sur D_ζ La variable D_ζ est définie selon

$$D_\zeta = p_\zeta^{\text{miss}} - 0{,}85 p_\zeta^{(\tau\tau)} \tag{4.12}$$

avec

$$p_\zeta^{\text{miss}} = \vec{E}_{\mathrm{T}}^{\text{miss}} \cdot \hat{\zeta}\,, \quad p_\zeta^{(\tau\tau)} = \vec{p}_{\mathrm{T}}{}^{(\tau\tau)} \cdot \hat{\zeta} \tag{4.13}$$

où $\hat{\zeta}$ est la direction bisectionnelle entre l'électron et le muon dans le plan transverse [220] et

$$\vec{p}_{\mathrm{T}}{}^{(\tau\tau)} = \vec{p}_{\mathrm{T}}{}^{e} + \vec{p}_{\mathrm{T}}{}^{\mu} \tag{4.14}$$

comme illustré sur la figure 4.2. Il est requis que $D_\zeta \geqslant -35$ afin de s'assurer que la région de signal soit orthogonale à la région de contrôle (CR) du bruit de fond $t\bar{t}$. Cette coupure permet de plus de rejeter ce bruit de fond dans la région de signal.

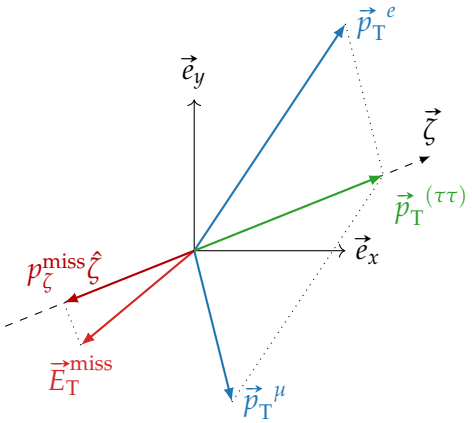


Figure 4.2 – *Illustration de la définition de $\hat{\zeta}$ [220]. Le plan de ce schéma est le plan transverse.*

Vetos de leptons supplémentaires Les vetos de leptons supplémentaires doivent être respectés, c'est-à-dire que l'événement ne contient pas :

- de second muon tel que $p_{\mathrm{T}}{}^{\mu} > 10\,\text{GeV}$, $|\eta^{\mu}| < 2{,}4$, passant le point de fonctionnement *medium* du *muonID* et d'isolation $I^{\mu} < 0{,}3\,p_{\mathrm{T}}{}^{\mu}$;
- de second électron tel que $p_{\mathrm{T}}{}^{e} > 10\,\text{GeV}$, $|\eta^{e}| < 2{,}5$, passant le point de fonctionnement à 90 % d'efficacité de l'*electron ID MVA* et d'isolation $I^{e} < 0{,}3\,p_{\mathrm{T}}{}^{e}$, l'électron devant passer le veto d'électron de conversion et présenter moins de deux points de passage manquants dans le trajectographe.

3.3 Sélection des jets

Les événements sont répartis en catégories d'après les différents mécanismes de production des bosons de Higgs. Cette catégorisation est détaillée dans la section 5. À cette fin, les jets présents dans l'événement sont exploités.

Pour réduire la contamination par les jets issus de l'empilement, les hadrons chargés sont soumis à la procédure de *pile-up Charged Hadron Subtraction* (CHS) [192] décrite dans le chapitre 3. Les jets

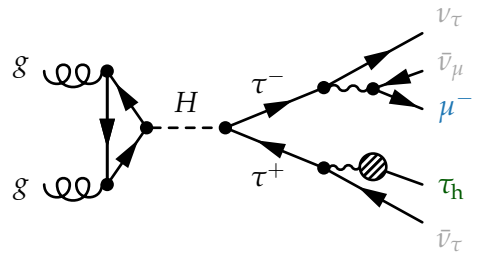
utilisés sont ceux obtenus à partir des particules restantes à l'aide de l'algorithme anti-k_T [162] avec un paramètre $R = 0{,}4$.

Ces jets doivent également passer les critères d'identification présentés dans le chapitre 2. L'identification des jets issus de quarks b (*b-tagging*) est réalisée par l'algorithme DEEPCSV [167, 221]. Les jets tels que $p_T > 20\,\text{GeV}$ et $|\eta| < 2{,}4(2{,}5)$ en 2016 (2017, 2018) sont considérés comme issus d'un b si leur score est supérieur à 0,3093 en 2016, 0,3033 en 2017 et 0,2770 en 2018. De plus, tout jet tel que $p_T > 30\,\text{GeV}$ et $|\eta| < 4{,}7$ est retenu. Afin d'exclure les électrons, muons et taus hadroniques de la liste des jets, il est requis que les jets soient distants du dilepton de $\Delta R > 0{,}5$.

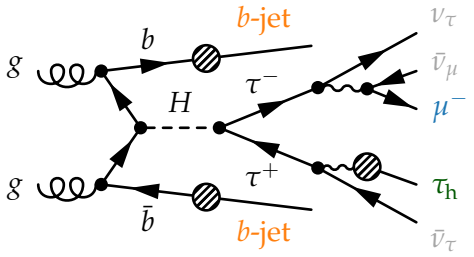
Lors de la prise de données en 2017, le bouchon du ECAL présentait un bruit important, perturbant la reconstruction des jets. Conformément aux recommandations du *POG* (*Physics Object Group*, groupe responsable d'un objet physique) *JetMET*, les jets reconstruits tels que $2{,}65 < |\eta| < 3{,}139$ ayant une impulsion transverse avant correction inférieure à 50 GeV sont rejetés. L'énergie transverse manquante est corrigée en conséquence.

4 Modélisation du bruit de fond

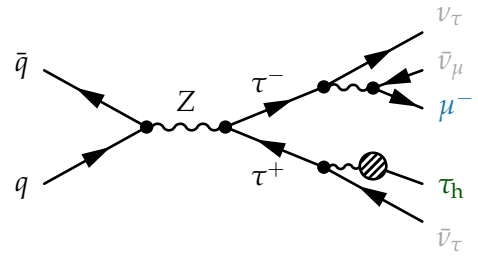
Le bruit de fond est constitué de tous les événements sélectionnés selon la procédure décrite section 3 mais ne correspondant pas au signal recherché, c'est-à-dire hors $h, H, A \to \tau\tau$. Plusieurs processus y contribuent. En effet, ces derniers peuvent donner des états finaux similaires à ceux attendus avec le signal recherché, comme illustré sur la figure 4.3. Ils peuvent également produire des objets physiques pouvant être interprétés comme des produits de désintégration de leptons τ.



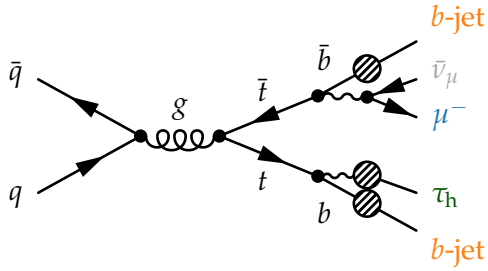
(a) *Signal* $gg \to H \to \tau\tau \to \mu\tau_h$.



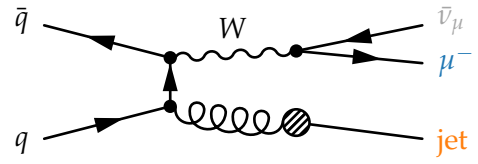
(b) *Signal* $gg \to b\bar{b}H \to b\bar{b}\tau\tau \to \mu\tau_h + 2b\text{-jets}$.



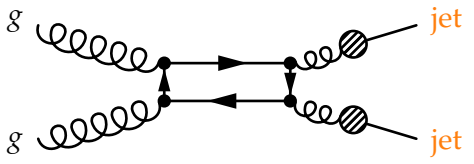
(c) *Drell-Yan* $q\bar{q} \to Z \to \tau\tau \to \mu\tau_h$.



(d) $t\bar{t}$ *Drell-Yan* $q\bar{q} \to g \to t\bar{t} \to \mu\tau_h + 2b\text{-jets}$.



(e) W *+ jets avec un muon dans l'état final.*



(f) *QCD.*

Figure 4.3 – *Diagrammes de Feynman complets des signaux ggH (4.3a) et bb̄H (4.3b) et bruits de fond Drell-Yan (4.3c), tt̄ (4.3d), W + jets (4.3e) et QCD (4.3f) de l'analyse illustrés dans le cas du canal* $\mu\tau_h$.

Ces processus, résumés dans le tableau 4.1 avec les pourcentages de leurs contributions au bruit de fond total, sont :

$Z \to \tau\tau$, $Z \to \ell\ell$ La désintégration du boson Z en paire de leptons τ ($Z \to \tau\tau$), ainsi qu'en paire de muons ou d'électrons ($Z \to \ell\ell$) lorsque l'un de ces leptons est mal identifié (les canaux $\mu\mu$ et ee n'étant pas exploités). La production du Z peut se faire par annihilation d'une paire de quarks, comme illustré sur la figure 4.3c. Il s'agit des processus « Drell-Yan ». Le Z peut également être produit par fusion de bosons électrofaibles (EWK, *ElectroWeaK*). Dans ce cas, deux jets supplémentaires sont présents dans l'état final.

W + jets La production d'un boson W, en particulier dans les canaux semi-leptoniques, avec le muon ou l'électron issu de la désintégration du W associé à un jet identifié à tort comme un τ_h. Ce processus est illustré figure 4.3e. Le W peut être produit par annihilation d'une paire de quarks, comme sur la figure 4.3e, ou par fusion de bosons électrofaibles (EWK).

$t\bar{t}$ La production d'une paire de quarks t, en particulier pour les événements contenant des jets issus de quarks b. Ce cas est illustré figure 4.3d. Les désintégrations par interaction faible des quarks t forment des bosons W, comme lors des désintégrations des leptons τ, d'où la contribution au bruit de fond de ces processus $t\bar{t}$.

Diboson Les productions de paires de bosons vecteurs ainsi que de quark t seul (*Single top*) contribuent également au bruit de fond, en particulier dans le canal $e\mu$.

QCD Enfin, les événements contenant des jets produits par interaction forte (QCD), lorsque ces jets sont identifiés à tort comme des éléments de désintégration d'une paire de leptons τ, forment la dernière source de bruit de fond considérée. Cette source de bruit de fond est particulièrement importante dans le canal $\tau_h\tau_h$.

Les contenus exacts en processus physiques de ces six sortes de bruit de fond sont détaillés dans l'annexe E. Plusieurs techniques sont utilisées afin de modéliser leurs contributions.

De plus, la désintégration du boson de Higgs du SM en paire de bosons W, $h \to W^+W^-$, constitue également un bruit de fond vis-à-vis de l'analyse $h \to \tau\tau$. En effet, les leptons τ se désintègrent par interaction faible en produisant un neutrino, invisible dans le détecteur, et un W virtuel. La désintégration d'un τ forme ainsi un état final très similaire à celle d'un W.

	Canal			
Bruit de fond	$\tau_h\tau_h$	$\mu\tau_h$	$e\tau_h$	$e\mu$
$Z \to \tau\tau$	33	46	27	20
$Z \to \ell\ell, \ell \in \{e, \mu\}$	~ 1	2	9	1
$t\bar{t}$	< 1	13	18	54
W + jets	< 1	36	42	3
QCD	66			11
Diboson	< 1	3	4	11

Tableau 4.1 – *Contributions en pourcent des bruits de fond aux canaux étudiés.*

Pour tous les processus à part QCD, des jeux de données simulées par générateur Monte-Carlo sont disponibles. Toutefois, une large partie des bruits de fond est estimée à partir des données réelles, ce qui permet d'améliorer l'accord entre ces dernières et l'estimation du bruit de fond tout en réduisant les incertitudes systématiques. Tous les événements simulés contenant deux leptons τ authentiques (*genuine*) sont ainsi remplacés par les données encapsulées (*embedded*) présentées dans la section 4.1. Les événements $Z \to \tau\tau$ sont ainsi couverts par cette méthode mais également une partie des bruits de fond $t\bar{t}$ et Diboson. De plus, la contribution du bruit de fond QCD dans le canal $e\mu$ est estimée à partir d'une région de contrôle où les charges électriques de l'électron et du muon sont de même signe. Cette méthode dénommée « QCD SS » (*Same Sign*) est exposée dans la section 4.2. Enfin, la proportion d'événements contenant au moins un jet identifié à tort comme provenant d'un lepton τ est estimée par la méthode des facteurs de faux (*Fake Factors*) décrite section 4.3. Tous les autres bruits de fond sont modélisés par des données simulées. Les jeux de données ainsi utilisés dans l'analyse sont présentés dans l'annexe E.

Afin de séparer les contributions estimées à partir des différentes techniques et de procéder à ces remplacements de manière cohérente, les événements simulés sont répartis selon la provenance des produits de désintégration visibles des leptons τ au niveau générateur. Pour cela, un *generator matching* est appliqué. Les particules reconstruites sélectionnées (électrons, muons et taus hadroniques) sont associées à l'objet physique généré le plus proche dans le plan (η, ϕ) et à moins de $\Delta R = 0{,}2$.

Si aucun objet généré ne respecte cette condition, l'objet reconstruit est considéré comme provenant d'un jet. Il est ainsi possible de déterminer l'origine de l'objet reconstruit en connaissant la provenance de l'objet généré correspondant. Il existe six cas de figure différents :

— électron natif (*prompt electron*), c'est-à-dire un électron ne provenant pas de la désintégration d'un lepton τ ;
— muon natif (*prompt muon*), c'est-à-dire un muon ne provenant pas de la désintégration d'un lepton τ ;
— électron provenant de la désintégration d'un lepton τ ;
— muon provenant de la désintégration d'un lepton τ ;
— tau hadronique ;
— jet ou particule issue de l'empilement.

Les définitions exactes de chacun de ces cas de figure sont données dans le tableau 4.2. Un τ_h généré est reconstruit à partir des produits de désintégration générés visibles hors électrons et muons. Il est de plus requis que l'impulsion transverse de ce τ_h généré reconstruit soit supérieure à 15 GeV afin d'éviter la limite de reconstruction des τ_h et d'éliminer des faux électrons et muons issus des τ_h. Dans le cas des électrons et muons natifs, la coupure $p_T > 8$ GeV permet de supprimer les leptons issus du FSR (*Final State Radiation*) $\gamma \to \ell^+\ell^-$. Le FSR est introduit au chapitre 3. Les remplacements des événements simulés se font ainsi sur la base des valeurs de `gen_match`, donnés dans le tableau 4.2, pour L_1 et L_2 selon les coupures données dans le tableau 4.3.

`gen_match`	Type de particule	Propriétés de l'objet au niveau générateur
1	électron natif	$\|\text{pdgID}\| = 11$, $p_T > 8$ GeV, `IsPrompt == True`
2	muon natif	$\|\text{pdgID}\| = 13$, $p_T > 8$ GeV, `IsPrompt == True`
3	$\tau \to e$	$\|\text{pdgID}\| = 11$, $p_T > 8$ GeV, `IsDirectPromptTauDecayProduct == True`
4	$\tau \to \mu$	$\|\text{pdgID}\| = 13$, $p_T > 8$ GeV, `IsDirectPromptTauDecayProduct == True`
5	$\tau \to \tau_h$	Tau hadronique généré
6	Faux τ_h, τ_h de l'empilement	Tout objet ne rentrant pas dans les catégories 1 à 5

Tableau 4.2 – *Valeurs prises par* `gen_match`*. Les particules telles que* `IsPrompt == True` *proviennent directement de l'événement dur et celles avec* `IsDirectPromptTauDecayProduct == True` *sont issues de la désintégration d'un lepton* τ *lui-même vérifiant* `IsPrompt == True`*.*

Canal	`gen_match` L_1	`gen_match` L_2	Simulations remplacées par la méthode
$\tau_h\tau_h$	5	5	Données encapsulées
$\tau_h\tau_h$	?	6	Facteurs de faux
$\tau_h\tau_h$	6	?	Facteurs de faux
$\mu\tau_h$	4	5	Données encapsulées
$\mu\tau_h$	?	6	Facteurs de faux
$e\tau_h$	3	5	Données encapsulées
$e\tau_h$	?	6	Facteurs de faux
$e\mu$	3	4	Données encapsulées

Tableau 4.3 – *Remplacement des événements simulés par une estimation basée sur les données. Un « ? » signifie « toute valeur possible ».*

4.1 Méthode des données encapsulées ou *embedding*

La méthode des données encapsulées (*embedding*) permet d'estimer le bruit de fond issu du SM donnant une paire de leptons τ dans l'état final en minimisant l'utilisation de simulations. La tech-

nique, présentée en détail dans la référence [222], se déroule en quatre étapes, résumées sur la figure 4.4 et listées ci-après :

1. Sélection d'une paire de muons :
   Dans les données réelles, des paires de muons sont formées. La paire de masse invariante la plus proche de celle du boson Z est choisie pour la suite. Il existe ainsi des contributions issues des processus $Z \to \mu\mu$, $t\bar{t}$ et Diboson.
2. Suppression de la paire de muons :
   Les signaux dans le détecteur correspondant aux muons sont retirés. Les autres signaux sont conservés pour la reconstruction de l'événement.
3. Génération d'une paire de taus :
   Deux leptons τ sont générés. Les propriétés cinématiques des muons initiaux sont utilisées afin d'obtenir celles des leptons τ. Leurs valeurs exactes sont modifiées afin de rendre compte de la différence de masse entre les muons et les taus. Plus de détails sont disponibles dans la section 5.3 de la référence [222]. Les désintégrations respectives des taus en électron, muon ou tau hadronique et leurs propagations dans le détecteur sont simulées.
4. Assemblage des données sans la paire de muons et des taus générés :
   Les traces et dépôts d'énergie des objets simulés à l'étape précédente sont ajoutés à ceux de l'événement réel, auquel les signaux associés à la paire de muons initiaux ont été retirés. La reconstruction des événements présentée au chapitre 2 peut alors être réalisée.

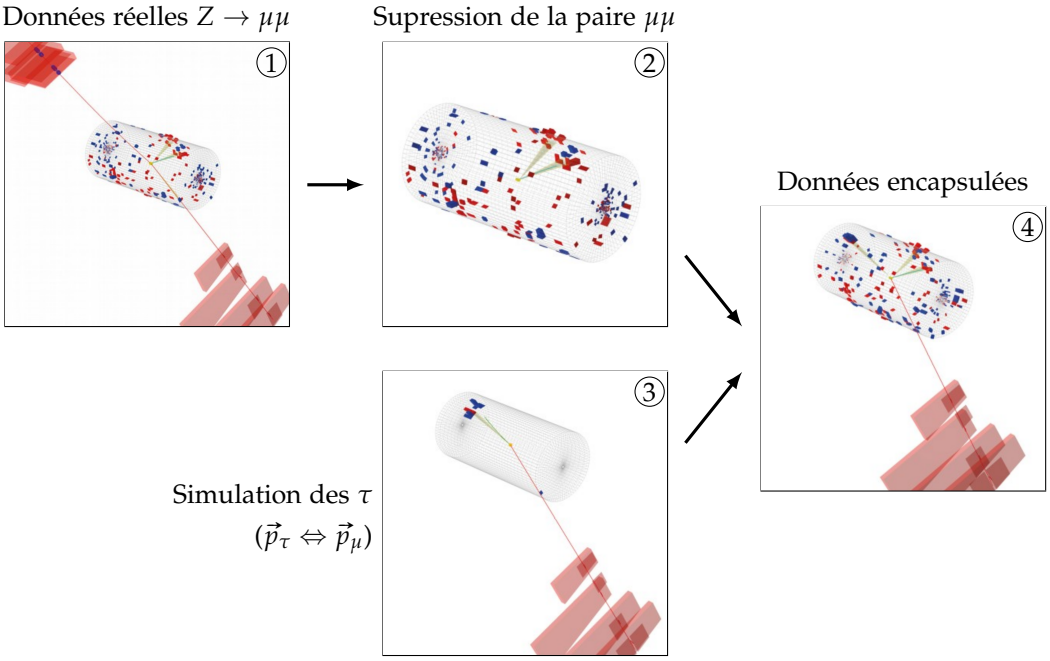
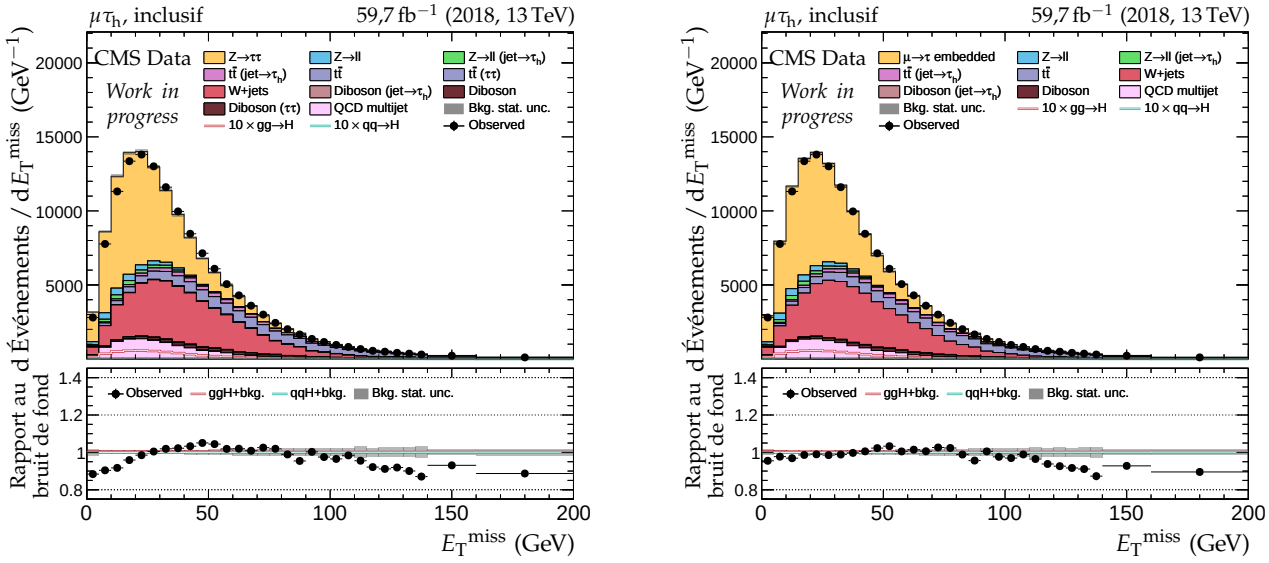


Figure 4.4 – *Schéma récapitulatif de la méthode des données encapsulées [222], illustrée dans le cas de l'état final* $\mu\tau_h$.

Les données encapsulées nécessitent ainsi l'utilisation de simulation uniquement pour la paire de leptons τ et leurs désintégrations. Tous les autres objets présents sont issus de données réelles. L'empilement, l'événement sous-jacent et les jets de l'événement principal sont donc décrits de manière parfaitement identique à la réalité, dans la mesure où ils ne sont pas simulés. L'énergie transverse manquante est modifiée, le contenu en particules de l'événement n'étant plus le même. De plus, l'incertitude sur la luminosité est supprimée pour les données encapsulées, car leur quantité est directement reliée à celle des données réelles, ce qui n'est pas le cas pour les données entièrement simulées. Enfin, les effets dus au détecteur tels que le bruit inhérent à la mesure, les pièces défectueuses et son vieillissement sont naturellement inclus dans les données encapsulées.

L'amélioration de la description des données ainsi obtenue grâce à l'encapsulation est visible sur la figure 4.5, sur laquelle les distributions de l'énergie transverse manquante dans les données et dans l'estimation du bruit de fond sans et avec cette méthode sont tracées à titre d'illustration. L'accord est sensiblement amélioré pour $E_T^{\text{miss}} < 60\,\text{GeV}$.



(a) *Sans encapsulation.*

(b) *Avec encapsulation.*

Figure 4.5 – *Distributions de E_T^{miss} pour le canal $\mu\tau_h$ en 2018. Les données observées (points noirs) sont comparées à la modélisation des bruits de fond (histogrammes remplis en couleur et empilés). Les bandes grisées correspondent à l'incertitude statistique totale sur le bruit de fond. Le rapport au bruit de fond est donné dans la partie inférieure des graphiques.*

4.2 Estimation du bruit de fond QCD dans le canal $e\mu$

Dans le cas du canal $e\mu$, le bruit de fond QCD contribue à la sélection des événements lorsqu'au moins un jet est identifié à tort comme un électron ou un muon. Une estimation de cette contribution est réalisée à partir des données réelles en suivant le principe de la méthode « ABCD ».

Quatre régions pouvant se résumer schématiquement comme illustré sur la figure 4.6 sont définies :

A région de signal (SR), définie dans la section 3 ;

B définie comme la SR mais avec les charges électriques de l'électron et du muon de même signe (SS, *Same Signs*) et non de signes opposés (OS, *Opposite Signs*) comme dans la SR ;

C définie comme la SR mais avec un muon « anti-isolé », c'est-à-dire que la coupure sur son isolation est inversée, $0{,}2 \leqslant I^{\mu}_{\text{rel}} < 0{,}5$ au lieu de $I^{\mu}_{\text{rel}} < 0{,}2$;

D définie comme la SR mais avec muon anti-isolé et SS.

μ :	isolé	anti-isolé
OS	A = SR	C
SS	B = AR	D

Figure 4.6 – *Définition schématique des régions A, B, C et D pour l'estimation du bruit de fond QCD.*

Les hypothèses d'application de cette méthode sont :

— la forme de la distribution de la variable v issue du bruit de fond QCD est identique dans la région A à déterminer et dans la région B connue ;

— le rapport du nombre d'événements entre A et B est le même qu'entre C et D.

Les contributions des bruits de fond autres que QCD aux régions B, C et D sont soustraits à partir de données simulées.

La méthode ABCD permet alors d'obtenir le bruit de fond QCD dans la région de signal A selon ce qui s'assimile à un produit en croix,

$$A = B \times \frac{C}{D} \Leftrightarrow h_v^{\text{A}} = h_v^{\text{B}} \times \frac{\int h_v^{\text{C}}}{\int h_v^{\text{D}}} \tag{4.15}$$

où h_v^{X} correspond à la distribution de la variable v dans la région X et $\int h_v^{\text{X}}$ à son intégrale, c'est-à-dire la quantité d'événements (indépendante de v). La région B est ainsi également nommée région d'application (AR, *Application Region*) du facteur C/D.

Afin d'augmenter la quantité d'événements exploités, et donc de réduire l'incertitude statistique, la coupure sur D_ζ n'est pas appliquée dans les régions C et D. Cependant, un facteur C/D global donne une estimation trop peu précise [223] car l'hypothèse d'indépendance de la forme de la distribution n'est pas vérifiée. Afin de corriger cet effet, le facteur C/D est déterminé en fonction de :

— la distance entre l'électron et le muon dans le plan (η, ϕ), ΔR ;
— le nombre de jets N_{jets} ;
— l'impulsion transverse de l'électron, ${p_{\text{T}}}^e$;
— l'impulsion transverse du muon, ${p_{\text{T}}}^\mu$;

La dépendance en ΔR est majoritairement due à la contribution $b\bar{b}$ au bruit de fond QCD. Elle est modélisée par un polynôme de degré 2.

Afin de corriger le biais introduit par le changement de critère d'isolation du muon, le facteur C/D est également déterminé dans les cas de figure suivants :

— électron anti-isolé ($0{,}15 \leqslant I^e_{\text{rel}} < 0{,}5$ au lieu de $I^e_{\text{rel}} < 0{,}15$) et muon isolé ;
— électron et muon anti-isolés.

Le rapport de ces facteurs donne la correction relative au passage des muons isolés à anti-isolés.

4.3 Méthode des facteurs de faux ou *fake factors*

La méthode des facteurs de faux (*fake factors*) a pour but de fournir, en se basant presque exclusivement sur les données collectées, une estimation des bruits de fond dans lesquels des jets, provenant de quarks ou de gluons, sont identifiés à tort comme des taus hadroniques (τ_{h}). De tels jets sont notés « faux τ_{h} » dans la suite.

Cette méthode est ainsi appliquée aux canaux contenant des τ_{h} dans l'état final, c'est-à-dire les canaux complètement hadronique ($\tau_{\text{h}}\tau_{\text{h}}$) et semi-leptoniques ($\ell\tau_{\text{h}}$ où $\ell \in \{ e, \mu \}$). Après la sélection détaillée dans la section 3, les faux τ_{h} représentent près de 70 % des événements dans le canal $\tau_{\text{h}}\tau_{\text{h}}$, 38 % dans le canal $\mu\tau_{\text{h}}$ et 68 % dans le canal $e\tau_{\text{h}}$ [224, 225]. Les processus physiques responsables des faux τ_{h} sont majoritairement QCD, W + jets et $t\bar{t}$. Dans le canal $\tau_{\text{h}}\tau_{\text{h}}$, près de 93 % des faux τ_{h} proviennent du bruit de fond QCD. Dans les canaux $\ell\tau_{\text{h}}$, environ 70 % des faux τ_{h} sont issus du bruit de fond W + jets. Les autres sources de faux τ_{h}, non traitées par cette méthode, sont les événements $Z \to \tau\tau$ avec un jet identifié comme un τ_{h}, couverts par la méthode décrite section 4.1, et Diboson, ce dernier type de processus ne contribuant que de l'ordre du pourcent au total des faux τ_{h}.

Les faux τ_{h} sont particulièrement difficiles à modéliser dans les simulations [224, 225]. De plus, le faible taux de mauvaise identification des τ_{h}, inférieur à 1 %, impliquerait l'utilisation de larges jeux de données simulées afin d'obtenir de faibles incertitudes statistiques. C'est en particulier le cas dans les régions de l'espace des phases contenant des bosons de Higgs lourds, recherchés dans cette analyse. La méthode des facteurs de faux se basant presque exclusivement sur les données collectées, les incertitudes inhérentes aux simulations deviennent négligeables face aux autres sources d'incertitudes. De plus, l'efficacité statistique de cette modélisation est directement liée à la luminosité intégrée, sans nécessiter de données simulées correspondantes.

4.3.1 Principe de base

Cette méthode suit le même principe de produit en croix que la méthode « ABCD » présentée section 4.2 mais va plus loin dans la détermination du facteur C/D nommé ici facteur de faux. Dans une région de contrôle, détaillée dans la section suivante, le rapport des quantités de τ_{h} isolés sur ceux anti-isolés est déterminé. Il s'agit du facteur de faux noté FF, défini comme

$$\text{FF} = \frac{n_{\text{iso}}}{n_{\text{anti-iso}}} = \frac{n(\texttt{Medium})}{n(\texttt{VVVLoose \&\& !Medium})} \tag{4.16}$$

avec :

— $n(\texttt{Medium})$ est la quantité d'événements dans la région de contrôle (CR, *Control Region*) passant le point de fonctionnement *medium* du discriminateur `deepTau vs jet`, utilisé également pour sélectionner les événements de signal ;
— $n(\texttt{VVVLoose \&\& !Medium})$ est la quantité d'événements dans la région de contrôle passant le point de fonctionnement le plus lâche de ce discriminateur, mais pas le moyen.

Le facteur de faux FF est déterminé de manière indépendante pour chaque canal ($\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$), chaque année (2016, 2017, 2018) et dépend :
— de l'impulsion transverse de l'objet physique identifié comme un τ_h, ${p_T}^{\tau_h}$;
— de l'impulsion transverse du jet le plus proche du τ_h, ${p_T}^{\text{jet}}$;
— du nombre de jets tels que $|\eta^{\text{jet}}| < 2{,}4$ et ${p_T}^{\text{jet}} > 20\,\text{GeV}$, $N_{\text{pre } b\text{-jets}}$.

Une région d'application du facteur de faux (AR, *Application Region*) est définie de manière similaire à la région de signal (SR, *Signal Region*), seul le critère d'isolation des τ_h passe de « isolé » à « anti-isolé ». La AR est ainsi riche en faux τ_h. La quantité d'événements contenant des faux τ_h dans la SR, notée $n_{j\to\tau_h}$, est alors obtenue par produit en croix avec la quantité d'événements dans la AR, notée n_{AR}, selon

$$n_{j\to\tau_h} = n_{\text{AR}} \cdot \text{FF}\,, \tag{4.17}$$

l'hypothèse étant l'universalité, c'est-à-dire que le facteur de faux mesuré dans la CR est supposé identique à celui de la AR.

4.3.2 Prise en compte des différentes sources de faux τ_h

La composition des jets n'est pas la même selon le processus physique dont proviennent les faux τ_h. Il existe donc différentes probabilités pour un jet de donner un faux τ_h suivant le processus physique. Dans le cas du canal $\tau_h\tau_h$, seul le bruit de fond QCD est traité par les facteurs de faux. Pour les canaux semi-leptoniques, trois sources de faux τ_h sont considérées, QCD, $W+\text{jets}$ et $t\bar{t}$. Pour chacune de ces sources, un facteur de faux est alors déterminé, selon l'équation (4.16), à partir d'une région de détermination (DR, *Determination Region*) dédiée, définie ci-après. Du fait de la séparation de la CR en plusieurs DR, l'universalité n'est alors plus complètement garantie. Des corrections résiduelles sont appliquées afin de corriger le biais introduit par la séparation en DR.

Le facteur de faux global est ainsi la moyenne de ceux obtenus pour chaque source, avec comme coefficients les fractions f d'événements de ces sources dans la AR, c'est-à-dire

$$\text{FF} = \sum_i f_i \cdot \text{FF}_i\,, \quad f_i = \frac{n^i_{\text{AR}}}{\sum_j n^j_{\text{AR}}}\,, \quad i,j \in \{\,\text{QCD},\, W+\text{jets},\, t\bar{t}\,\}\,. \tag{4.18}$$

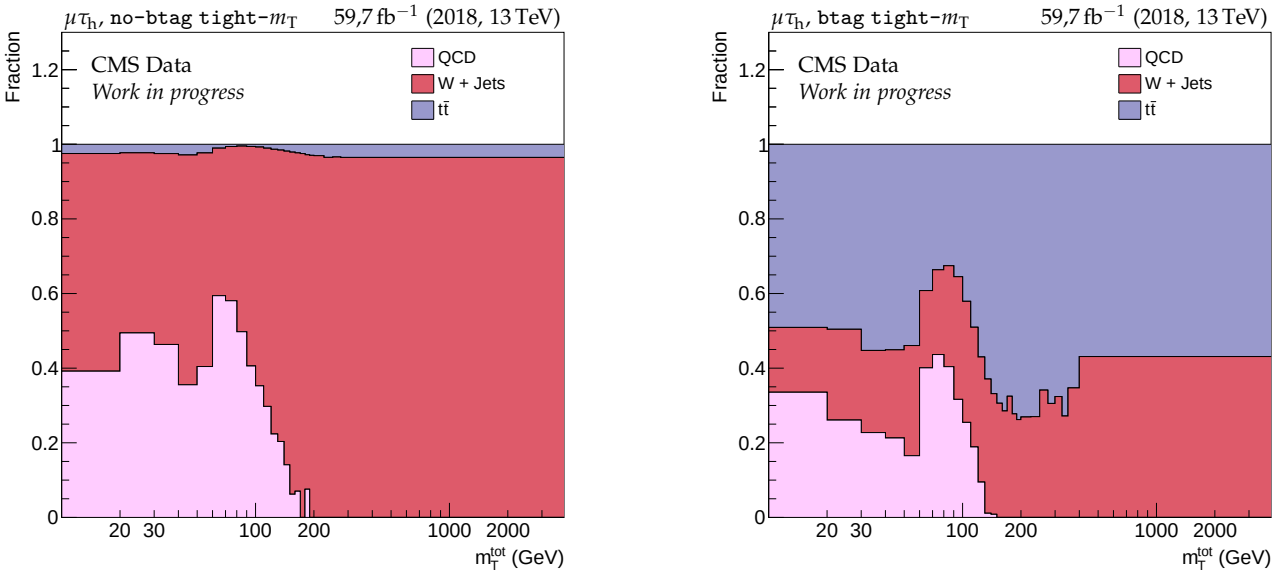
Les fractions f_i sont déterminées à partir de simulations et dépendent :
— de la masse transverse du lepton $\ell \in \{\,e,\mu\,\}$, m_T^ℓ ;
— du nombre de jets $N_{\text{pre } b\text{-jets}}$ identifiés comme issus de quarks b, $N_{b\text{-jets}}$;
— de la masse transverse totale, définie équation (4.24), m_T^{tot}.

La figure 4.7 présente ces fractions en fonction de m_T^{tot} pour $m_T^\mu < 40\,\text{GeV}$ et $N_{b\text{-jets}} \in \{\,0, \geqslant 1\,\}$ pour le canal $\mu\tau_h$ en 2018. La méthode des facteurs de faux est résumée figure 4.8.

4.3.3 Régions de détermination

QCD La DR QCD est définie de la même manière que la SR à l'exception du critère sur les charges électriques des éléments du dilepton. En effet, ceux-ci doivent être de charges opposées (OS, *Opposite Signs*) dans la SR car, les bosons de Higgs recherchés étant neutres, la charge globale du dilepton doit être nulle par conservation. Pour la DR QCD, ces charges doivent être de même signe (SS, *Same Sign*). Dans le cas des canaux $\ell\tau_h$, il est de plus requis que $I^\ell_{\text{rel}} > 0{,}05$ afin de réduire les contributions de processus donnant des électrons ou des muons sans objet physique pertinent pour les facteurs de faux. Les contributions d'autres processus à la DR sont soustraites grâce à l'utilisation de données simulées. Pour le canal $\tau_h\tau_h$, le facteur de faux FF_Q est déterminé uniquement pour le premier τ_h (de plus haut p_T).

Le facteur de faux FF_Q est mesuré séparément pour $N_{\text{pre } b\text{-jets}} = 0$ et $N_{\text{pre } b\text{-jets}} \geqslant 1$ et dans chacun de ces deux cas pour :
— ${p_T}^{\text{jet}} / {p_T}^{\tau_h} < 1{,}25$;
— $1{,}25 \leqslant {p_T}^{\text{jet}} / {p_T}^{\tau_h} < 1{,}5$;
— $1{,}5 \leqslant {p_T}^{\text{jet}} / {p_T}^{\tau_h}$.



(a) *Pour* $N_{b\text{-}jets} = 0$ *et* $m_{\mathrm{T}}^{\mu} < 40\,\mathrm{GeV}$

(b) *Pour* $N_{b\text{-}jets} \geqslant 1$ *et* $m_{\mathrm{T}}^{\mu} < 40\,\mathrm{GeV}$

Figure 4.7 – *Fractions des sources de faux* τ_{h} *dans le canal* $\mu\tau_{\mathrm{h}}$ *en 2018 [209].*

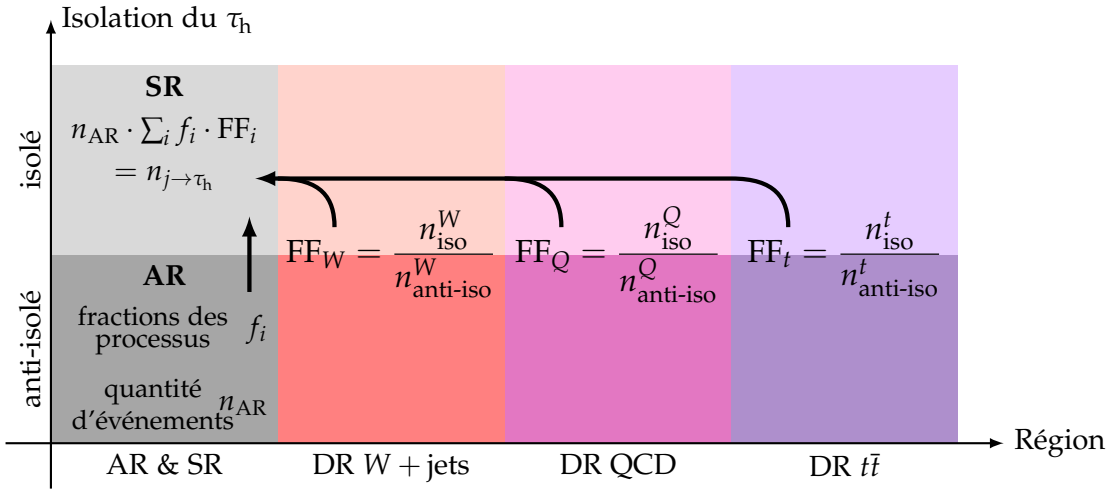


Figure 4.8 – *Illustration de la méthode des facteurs de faux. Les facteurs de faux sont obtenus à partir du nombre d'événements avec des* τ_{h} *isolés et anti-isolés dans les régions de détermination (DR) de chaque processus contribuant significativement au bruit de fond contenant des faux* τ_{h}*. La quantité de faux* τ_{h} *dans la région de signal (SR) est estimée à partir des fractions de ces processus et du nombre d'événements présents dans la région d'application (AR).*

Pour ces six catégories, la dépendance en ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}}$ de FF_Q est modélisée par un polynôme de degré 3 ajusté aux mesures pour ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} < 200\,\mathrm{GeV}$.

Dans le cas du canal $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, peu d'événements sont disponibles pour ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} \geqslant 200\,\mathrm{GeV}$. Le facteur de faux FF_Q est ainsi fixé à la valeur mesurée pour ${p_{\mathrm{T}}}^{\mathrm{jet}} / {p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} \geqslant 1{,}5$ et à la valeur à 200 GeV du polynôme pour ${p_{\mathrm{T}}}^{\mathrm{jet}} / {p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} < 1{,}5$. Pour les canaux $\ell\tau_{\mathrm{h}}$, la situation est similaire à partir de ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} \geqslant 140\,\mathrm{GeV}$. La valeur utilisée à haut ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}}$ suit la logique suivante :

- si l'erreur relative sur FF_Q pour ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} \geqslant 200\,\mathrm{GeV}$ est inférieure à 0,5 :
  - si l'erreur relative sur FF_Q pour ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} \in [140, 200]\,\mathrm{GeV}$ est inférieure à 0,5, les valeurs mesurées sont utilisées sur les intervalles $[140, 200]\,\mathrm{GeV}$ et $[200, \infty[\,\mathrm{GeV}$,
  - si l'erreur relative sur FF_Q pour ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} \in [140, 200]\,\mathrm{GeV}$ est supérieure à 0,5, la valeur mesurée est utilisée sur l'intervalle $[200, \infty[\,\mathrm{GeV}$;
- si l'erreur relative sur FF_Q pour ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} \geqslant 200\,\mathrm{GeV}$ est supérieure à 0,5 et inférieure à 0,5 pour ${p_{\mathrm{T}}}^{\tau_{\mathrm{h}}} \in [140, 200]\,\mathrm{GeV}$, la valeur mesurée est utilisée sur l'intervalle $[140, \infty[\,\mathrm{GeV}$;
- sinon, la valeur obtenue par l'ajustement est utilisée.

W + jets La DR W + jets ne concerne que les canaux semi-leptoniques. Elle est définie de la même manière que la SR à l'exception de la coupure sur la masse transverse du lepton qui doit ici être supérieure à 70 GeV, alors qu'elle est inférieure à cette même valeur dans la SR. Il est de plus requis que $N_{b\text{-jets}} = 0$ afin de supprimer la contamination par les événements $t\bar{t}$. Les contributions d'autres processus physiques à la DR sont soustraites par l'utilisation directe de données simulées. Le bruit de fond QCD à retirer est obtenu à partir des données réelles avec les charges électriques des éléments du dilepton de même signe, auxquelles sont soustraits les autres bruits de fond avec charges électriques de même signe, y compris W + jets, obtenus par simulation directe. Un facteur correctif de 1,1 est appliqué aux données à retirer, correspondant au rapport observé d'événements avec charges opposées sur événements avec charges de même signe.

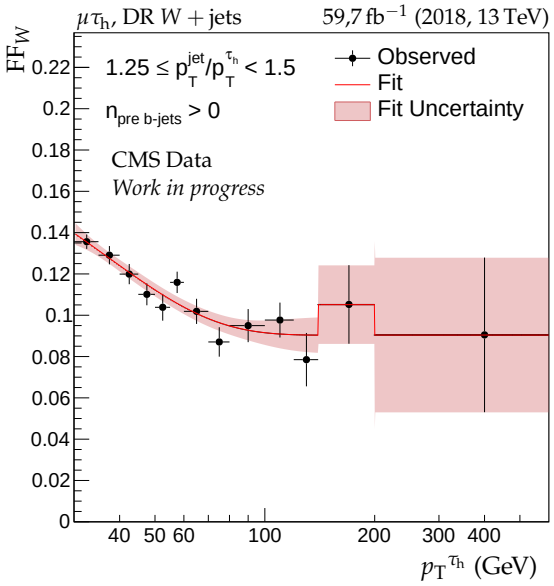


Figure 4.9 – *Ajustement de* FF$_W$ *dans le canal* $\mu\tau_h$ *en 2018 [209].*

Le facteur de faux FF$_W$ est mesuré séparément pour $N_{\text{pre } b\text{-jets}} = 0$ et $N_{\text{pre } b\text{-jets}} \geqslant 1$ et dans chacun de ces deux cas pour :

— $p_T^{jet}/p_T^{\tau_h} < 1{,}25$;
— $1{,}25 \leqslant p_T^{jet}/p_T^{\tau_h} < 1{,}5$;
— $1{,}5 \leqslant p_T^{jet}/p_T^{\tau_h}$.

Pour ces six catégories, la dépendance en $p_T^{\tau_h}$ de FF$_W$ est modélisée par un polynôme de degré 3 ajusté aux mesures pour $p_T^{\tau_h} < 140$ GeV.

Dans cette DR également, peu d'événements sont disponibles pour $p_T^{\tau_h} \geqslant 140$ GeV. La même logique que pour FF$_Q$, détaillée précédemment, est suivie sur la valeur de FF$_W$ à utiliser. L'ajustement obtenu pour FF$_W$ sur le canal $\mu\tau_h$ en 2018 est illustré figure 4.9 pour $N_{\text{pre } b\text{-jets}} \geqslant 1$ et $1{,}25 \leqslant p_T^{jet}/p_T^{\tau_h} < 1{,}5$. Il y apparaît l'effet du traitement de la région à haut $p_T^{\tau_h}$.

$t\bar{t}$ La DR $t\bar{t}$ ne concerne que les canaux semi-leptoniques. Il n'est pas possible de définir une DR issue des données suffisamment pure pour mesurer FF$_t$. Ce facteur de faux est alors obtenu à partir de données simulées.

Le facteur de faux FF$_W$ peut également être mesuré à partir de données simulées uniquement. Un écart de 10 à 20 % avec le facteur de faux obtenu à partir des données réelles est observé. La contribution $t\bar{t}$ étant faible par rapport aux autres bruits de fond, le biais introduit par l'utilisation de données simulées est négligeable face aux incertitudes sur les facteurs de faux.

Le facteur de faux FF$_t$ est mesuré pour :

— $p_T^{jet}/p_T^{\tau_h} < 1{,}25$;
— $1{,}25 \leqslant p_T^{jet}/p_T^{\tau_h} < 1{,}5$;
— $1{,}5 \leqslant p_T^{jet}/p_T^{\tau_h}$;

sans séparation en $N_{\text{pre } b\text{-jets}}$, la majorité des événements $t\bar{t}$ vérifiant $N_{\text{pre } b\text{-jets}} \geqslant 1$. L'ajustement obtenu pour FF$_t$ sur le canal $\mu\tau_h$ en 2018 est illustré figure 4.10 pour $p_T^{jet}/p_T^{\tau_h} < 1{,}25$.

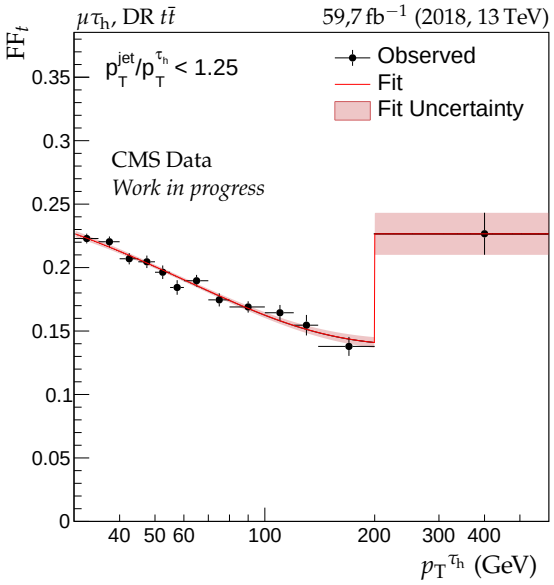


Figure 4.10 – *Ajustement de* FF$_t$ *dans le canal* $\mu\tau_h$ *en 2018 [209].*

4.3.4 Corrections résiduelles

Afin de valider les facteurs de faux obtenus, ces derniers sont appliqués aux événements des DRs sélectionnés avec les points de fonctionnement de l'algorithme DEEPTAU utilisés dans la SR. Les distributions des données réelles et des prédictions obtenues par les facteurs de faux appliqués aux données réelles doivent alors coïncider. Les écarts résiduels donnent la correction à appliquer, paramétrisée en fonction :

— du nombre de jets $N_{\text{pre } b\text{-jets}}$ identifiés comme issus de quarks b, $N_{b\text{-jets}}$;
— de l'impulsion transverse du lepton $\ell \in \{ e, \mu \}$ pour les canaux semi-leptoniques, ${p_{\mathrm{T}}}^{\ell}$;
— de l'isolation du lepton $\ell \in \{ e, \mu \}$ pour les canaux semi-leptoniques, I^{ℓ} ;
— de la fraction d'énergie transverse manquante alignée avec le τ_{h} pour les événements QCD, C_Q,

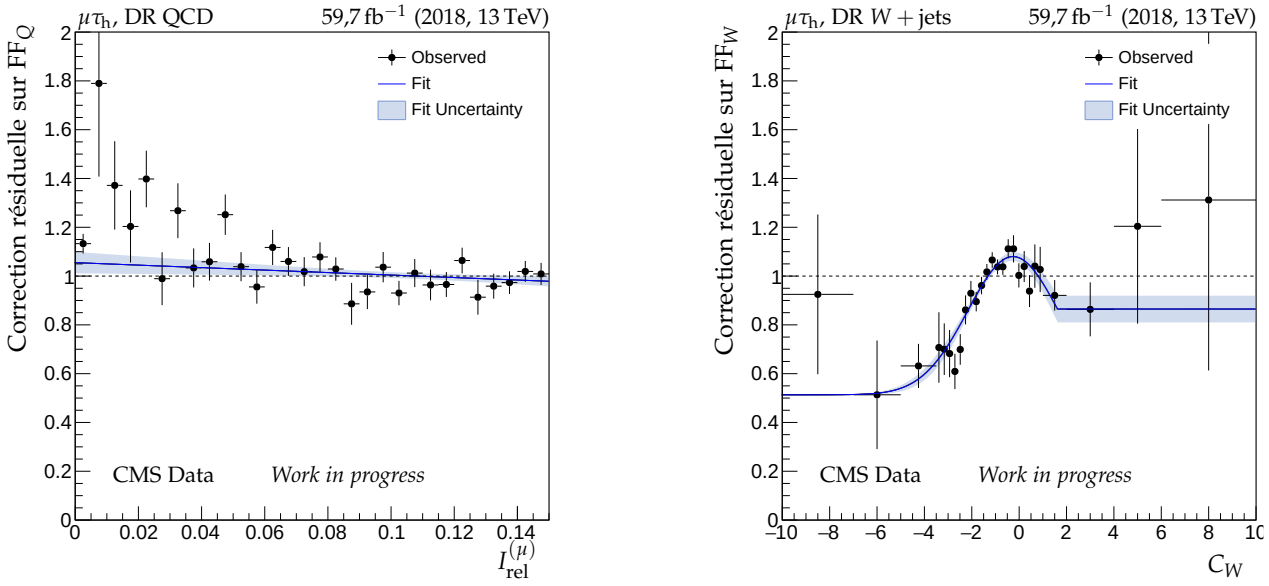
$$C_Q = \frac{{E_{\mathrm{T}}}^{\text{miss}}}{{p_{\mathrm{T}}}^{\tau_{\mathrm{h}}}} \cos\Big(\Delta\phi(\vec{E}_{\mathrm{T}}^{\text{miss}}, {\vec{p}_{\mathrm{T}}}^{\tau_{\mathrm{h}}})\Big) \; ; \tag{4.19}$$

— de la fraction d'énergie transverse manquante alignée avec le τ_{h} pour les événements W + jets, C_W,

$$C_W = \frac{\| \vec{E}_{\mathrm{T}}^{\text{miss}} + {\vec{p}_{\mathrm{T}}}^{\ell} \|}{{p_{\mathrm{T}}}^{\tau_{\mathrm{h}}}} \cos\Big(\Delta\phi(\vec{E}_{\mathrm{T}}^{\text{miss}} + {\vec{p}_{\mathrm{T}}}^{\ell}, {\vec{p}_{\mathrm{T}}}^{\tau_{\mathrm{h}}})\Big) \, , \tag{4.20}$$

dont la définition est semblable à celle de C_Q mais où $\vec{E}_{\mathrm{T}}^{\text{miss}}$ est remplacé par $\vec{E}_{\mathrm{T}}^{\text{miss}} + {\vec{p}_{\mathrm{T}}}^{\ell}$ afin de prendre en compte la contribution à ${E_{\mathrm{T}}}^{\text{miss}}$ du neutrino issu de la désintégration du boson W. Il est ici considéré comme dos-à-dos avec ℓ, ce qui n'est strictement vrai que pour un W au repos.

Deux de ces corrections résiduelles obtenues sur le canal $\mu\tau_{\mathrm{h}}$ en 2018 sont illustrées sur la figure 4.11.

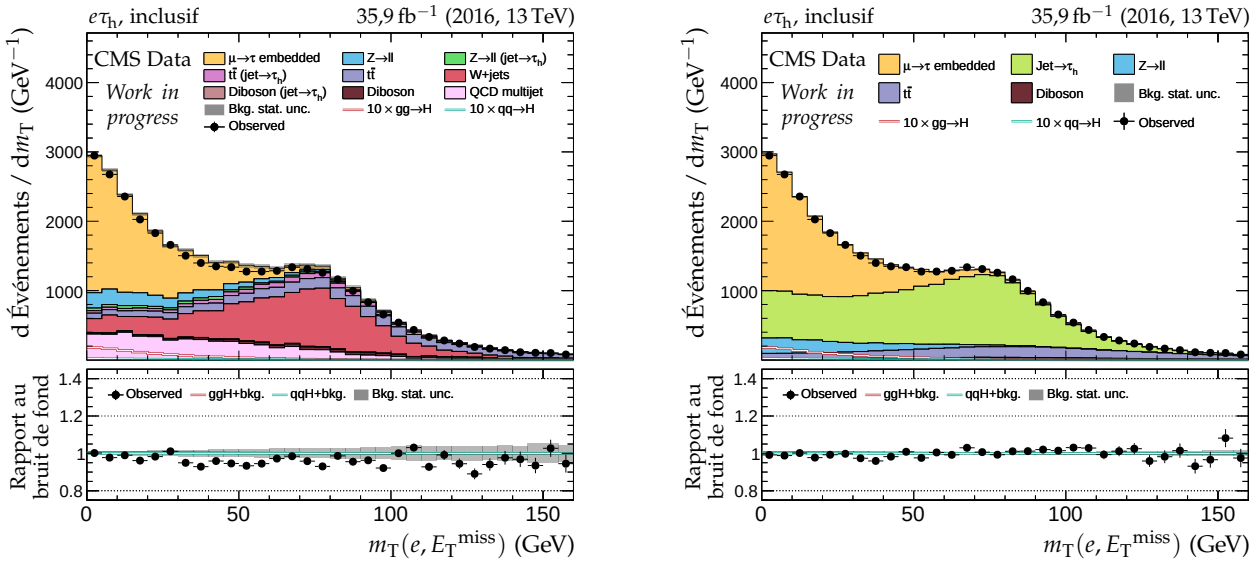


(a) *En fonction de* $I_{\text{rel}}^{(\mu)}$ *pour* FF_Q.

(b) *En fonction de* C_W *pour* FF_W

Figure 4.11 – *Corrections résiduelles des facteurs de faux dans le canal* $\mu\tau_{\mathrm{h}}$ *en 2018 [209].*

L'amélioration de la description des données ainsi obtenue grâce aux facteurs de faux est visible sur la figure 4.12, où les distributions de la masse transverse de l'électron dans le canal $e\tau_{\mathrm{h}}$ dans les données et dans l'estimation du bruit de fond sans et avec cette méthode sont tracées à titre d'illustration. Outre un meilleur accord entre observations et estimation du bruit de fond, l'incertitude statistique est également réduite.



(a) *Sans facteurs de faux.* **(b)** *Avec facteurs de faux.*

Figure 4.12 – *Distributions de la masse transverse de l'électron pour le canal $e\tau_h$ en 2016.*

5 Catégorisation des événements et variables discriminantes

Afin d'augmenter la sensibilité de l'analyse à un signal particulier, il est possible de définir des catégories. Par exemple, comme exposé dans le chapitre 1, le mode de production dominant des bosons H et A peut être celui en association avec des quarks b. Le signal correspondant, dans ce cas, se trouve dans les événements comportant des jets issus de quarks b. Il est donc pertinent de séparer les événements en deux groupes, avec et sans jets de quarks b.

Des catégories conçues pour la recherche de H et A ont déjà été exploitées avec les données enregistrées en 2016 par l'expérience CMS [81]. Ces catégories, utilisées dans le cadre du MSSM c'est-à-dire au-delà du SM (*Beyond Standard Model*), sont notées « BSM » et sont détaillées dans la section 5.1. La recherche d'un signal supplémentaire, indépendamment de tout modèle, est réalisée avec ces catégories.

En plus de la recherche d'un signal correspondant à de nouvelles particules, il est possible d'exploiter les signaux de particules déjà connues afin de tester la validité d'un modèle. En effet, les modèles supersymétriques tels que le MSSM doivent en premier lieu être compatibles avec les propriétés des particules déjà connues, comme h. Pour obtenir une meilleure sensibilité au boson de Higgs du SM h, il est possible de combiner les catégories BSM avec un jeu de catégories issu de l'analyse des événements $h \to \tau\tau$ dans le cadre du SM [226-228]. Il s'agit des catégories « SM », présentées dans la section 5.2. La combinaison des catégories SM et BSM est détaillée dans la section 5.3.

Afin de séparer signal et bruit de fond, il est nécessaire de définir une variable discriminante. Les variables discriminantes utilisées dans les catégories BSM et SM sont définies dans les sections correspondantes.

5.1 Catégories « BSM »

5.1.1 Définition des catégories

Les catégories `BSM`, introduites dans la référence [81], sont construites dans le but de chercher une résonance correspondant à un boson de Higgs lourd.

Une première catégorisation est basée sur la présence de jets issus de quarks b. Deux catégories sont ainsi définies :

— `no-btag` : $N_{b\text{-jets}} = 0$;
— `btag` : $N_{b\text{-jets}} \geqslant 1$.

Dans le cas des canaux $\mu\tau_h$, $e\tau_h$ et $e\mu$, chacune de ces deux catégories est à nouveau subdivisée.

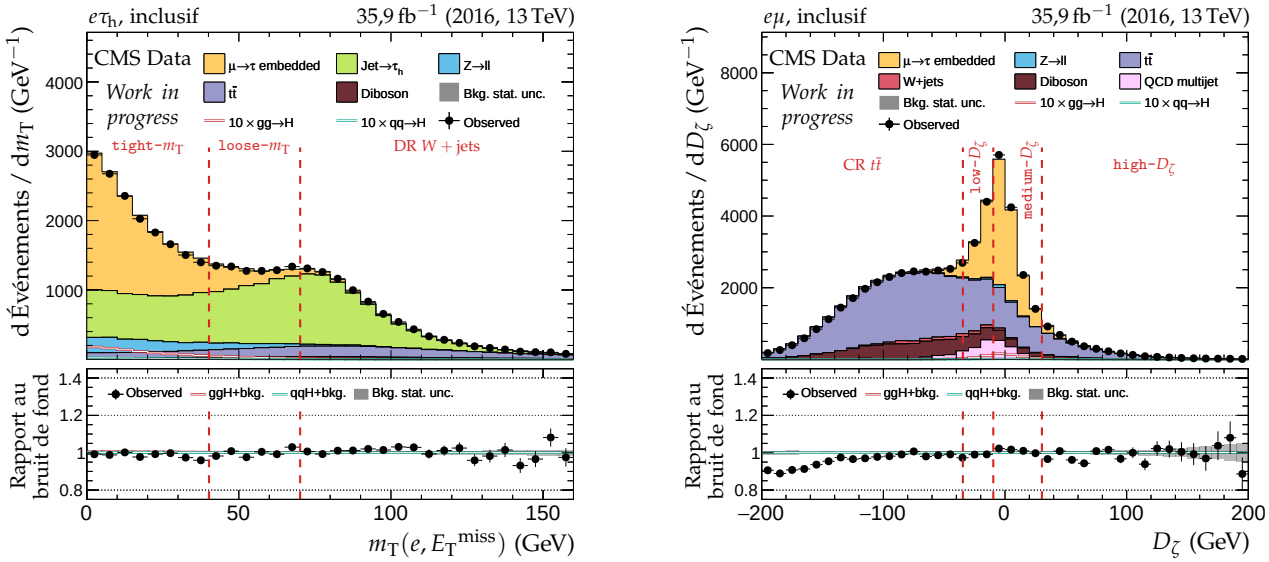
Canaux $\mu\tau_h$ et $e\tau_h$ Dans ces deux canaux, la masse transverse de L_1 (le muon ou l'électron, notés ℓ), définie par

$$m_T^\ell = m_T(\ell, {E_T}^{\text{miss}}) = \sqrt{2\, p_T{}^\ell\, {E_T}^{\text{miss}}\, (1 - \cos\Delta\phi)} \tag{4.21}$$

avec $\Delta\phi = \phi^\ell - \phi^{{E_T}^{\text{miss}}}$, est utilisée afin de définir deux catégories :

— `tight-`m_T : $m_T^\ell < 40\,\text{GeV}$;
— `loose-`m_T : $40\,\text{GeV} \leqslant m_T^\ell < 70\,\text{GeV}$;

la limite haute sur m_T pour la catégorie `loose-`m_T étant appliquée afin de s'assurer que la région de signal soit orthogonale à la région de détermination (DR) des facteurs de faux des événements W + jets. La majorité des événements de signal, en particulier pour H et A de basse masse, se trouve dans la catégorie `tight-`m_T. La catégorie `loose-`m_T permet quant à elle d'augmenter l'acceptation du signal pour $m_{H,A} > 700\,\text{GeV}$. La figure 4.13a illustre ces coupures sur m_T^ℓ dans le cas du canal $e\tau_h$ pour l'année 2016.



(a) *Catégorisation basée sur m_T^ℓ.*

(b) *Catégorisation basée sur D_ζ.*

Figure 4.13 – *Illustrations des catégorisations basées sur m_T^ℓ et D_ζ, respectivement sur les événements des canaux $e\tau_h$ et $e\mu$ de l'année 2016.*

Canal $e\mu$ Trois catégories sont définies selon la valeur de D_ζ définie équation (4.12) :

— `low-`D_ζ : $-35\,\text{GeV} \leqslant D_\zeta < -10\,\text{GeV}$;
— `medium-`D_ζ : $-10\,\text{GeV} \leqslant D_\zeta < 30\,\text{GeV}$;
— `high-`D_ζ : $30\,\text{GeV} \leqslant D_\zeta$;

la limite basse sur D_ζ pour la catégorie `low-`D_ζ étant appliquée afin de s'assurer que la région de signal soit orthogonale à la région de contrôle (CR) du bruit de fond $t\bar{t}$. Ces trois catégories permettent d'obtenir diverses puretés de signal et fractions de bruit de fond $t\bar{t}$. La majorité des événements de signal se trouve dans la catégorie `medium-`D_ζ. La figure 4.13b illustre ces coupures sur D_ζ.

Catégories obtenues Les catégories `BSM` correspondant à la région de signal (SR), c'est-à-dire en dehors des régions de détermination (DR) et de contrôle (CR), sont résumées sur la figure 4.14 pour les quatre canaux considérés.

5.1.2 Variable discriminante

La masse invariante d'une particule peut être obtenue par un calcul de relativité restreinte à partir des propriétés cinématiques de chacun des ses produits de désintégration. Cette observable doit

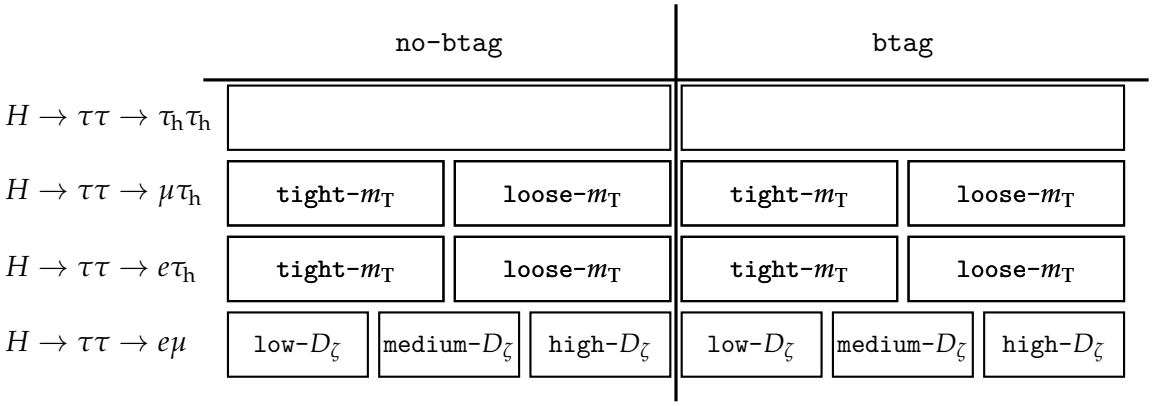
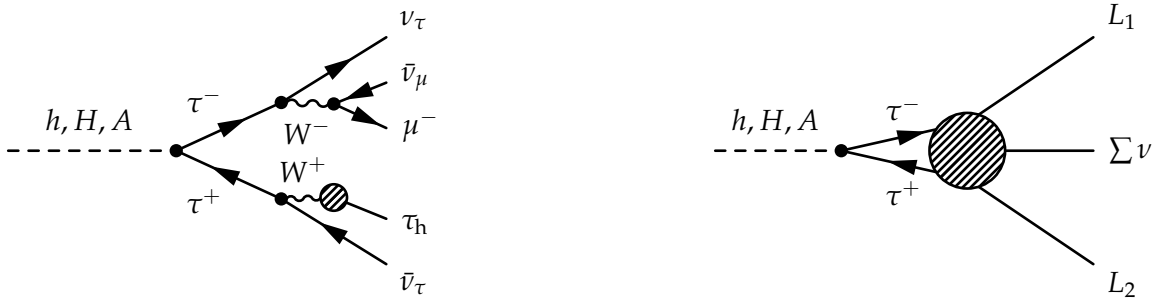


Figure 4.14 – *Catégories BSM pour les quatre canaux considérés.*

correspondre à la masse de cette particule et est donc un choix pertinent de variable discriminante. Elle est ainsi utilisée, par exemple, dans l'analyse $h \to ZZ \to 4\ell$ [30].

Cependant, dans l'analyse menée ici, l'état final comporte deux à quatre neutrinos issus des désintégrations des leptons τ. La figure 4.15a illustre le cas du canal $\mu\tau_h$ dans lequel trois neutrinos sont ainsi présents. Or, les neutrinos sont invisibles dans le détecteur CMS. Il est donc impossible de déterminer la masse invariante totale de ce système.



(a) *Événement réel.* **(b)** *Événement approximé.*

Figure 4.15 – *Diagrammes de Feynman d'un événement $H \to \tau\tau \to \mu\tau_h$, avec trois neutrinos dans l'état final.*

L'énergie transverse manquante, introduite dans le chapitre 2, correspond à la somme des impulsions transverses de toutes les particules invisibles de l'événement. Alors, pour un détecteur sans défaut dont l'acceptation spatiale est totale, c'est-à-dire avec une reconstruction parfaite de toutes les particules hormis les neutrinos,

$$\vec{E}_T^{\text{miss}} = \sum_{\nu_i} \vec{p}_T^{\,\nu_i}\,, \tag{4.22}$$

la somme se faisant sur tous les neutrinos de l'événement dont ceux de l'état final du processus $H \to \tau\tau$.

Ainsi, il est possible de considérer le système à trois corps suivant :

— L_1 la partie visible de la désintégration d'un des leptons τ, par exemple le muon de la figure 4.15a ;

— L_2 la partie visible de la désintégration de l'autre lepton τ, par exemple le $\tau_h$résultant de $W^+ \to u\bar{d}$ sur la figure 4.15a ;

— $\sum \nu \simeq E_T^{\text{miss}}$ l'ensemble des neutrinos issus des désintégrations des leptons τ, eux-mêmes issus de la désintégration du boson de Higgs, dont une estimation des propriétés cinématiques dans le plan transverse est donnée par l'énergie transverse manquante $\vec{E}_T^{\text{miss}}$;

comme illustré figure 4.15b. En se restreignant donc au plan transverse car E_T^{miss} est utilisée, pour des particules relativistes ($m \ll E$), le calcul de la « masse invariante » de ce système à trois corps

donne

$$
\begin{aligned}
m^2 = E^2 - p^2 &= \left(\sum_{i\in\{L_1,L_2,\sum\nu\}} E_i\right)^2 - \left(\sum_{i\in\{L_1,L_2,\sum\nu\}} \vec{p}_{\mathrm{T}}{}^{i}\right)^2 \\
&= \left(E_{L_1} + E_{L_2} + E_{\mathrm{T}}{}^{\mathrm{miss}}\right)^2 - \left(\vec{p}_{\mathrm{T}}{}^{L_1} + \vec{p}_{\mathrm{T}}{}^{L_2} + \vec{E}_{\mathrm{T}}^{\mathrm{miss}}\right)^2 \\
&= E_{L_1}^2 + E_{L_2}^2 + E_{\mathrm{T}}{}^{\mathrm{miss}\,2} + 2\left(E_{L_1}E_{L_2} + E_{L_1}E_{\mathrm{T}}{}^{\mathrm{miss}} + E_{L_2}E_{\mathrm{T}}{}^{\mathrm{miss}}\right) \\
&\quad - \left(\vec{p}_{\mathrm{T}}{}^{L_1\,2} + \vec{p}_{\mathrm{T}}{}^{L_2\,2} + \vec{E}_{\mathrm{T}}^{\mathrm{miss}\,2}\right) - 2\left(\vec{p}_{\mathrm{T}}{}^{L_1}\cdot\vec{p}_{\mathrm{T}}{}^{L_2} + \vec{p}_{\mathrm{T}}{}^{L_1}\cdot\vec{E}_{\mathrm{T}}^{\mathrm{miss}} + \vec{p}_{\mathrm{T}}{}^{L_2}\cdot\vec{E}_{\mathrm{T}}^{\mathrm{miss}}\right) \\
&= 2\Big[p_{\mathrm{T}}{}^{L_1}p_{\mathrm{T}}{}^{L_2}(1-\cos\Delta\phi^{(L_1,L_2)}) \\
&\qquad + p_{\mathrm{T}}{}^{L_1}E_{\mathrm{T}}{}^{\mathrm{miss}}(1-\cos\Delta\phi^{(L_1,E_{\mathrm{T}}{}^{\mathrm{miss}})}) + p_{\mathrm{T}}{}^{L_2}E_{\mathrm{T}}{}^{\mathrm{miss}}(1-\cos\Delta\phi^{(L_2,E_{\mathrm{T}}{}^{\mathrm{miss}})})\Big] \\
&= m_{\mathrm{T}}^2(L_1,L_2) + m_{\mathrm{T}}^2(L_1,E_{\mathrm{T}}{}^{\mathrm{miss}}) + m_{\mathrm{T}}^2(L_2,E_{\mathrm{T}}{}^{\mathrm{miss}}) = m_{\mathrm{T}}^{\mathrm{tot}\,2}\,.
\end{aligned}
\tag{4.23}
$$

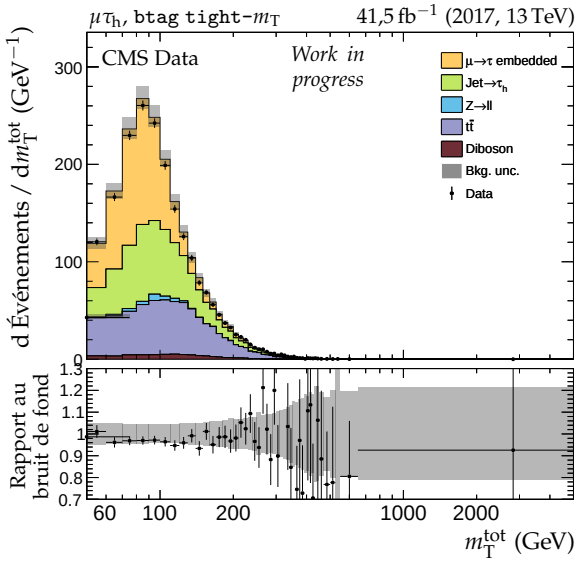
La variable discriminante utilisée dans les catégories `BSM` est ainsi $m_{\mathrm{T}}^{\mathrm{tot}}$, la masse transverse totale, définie telle que

$$
m_{\mathrm{T}}^{\mathrm{tot}} = \sqrt{m_{\mathrm{T}}^2(L_1,L_2) + m_{\mathrm{T}}^2(L_1,E_{\mathrm{T}}{}^{\mathrm{miss}}) + m_{\mathrm{T}}^2(L_2,E_{\mathrm{T}}{}^{\mathrm{miss}})} \tag{4.24}
$$

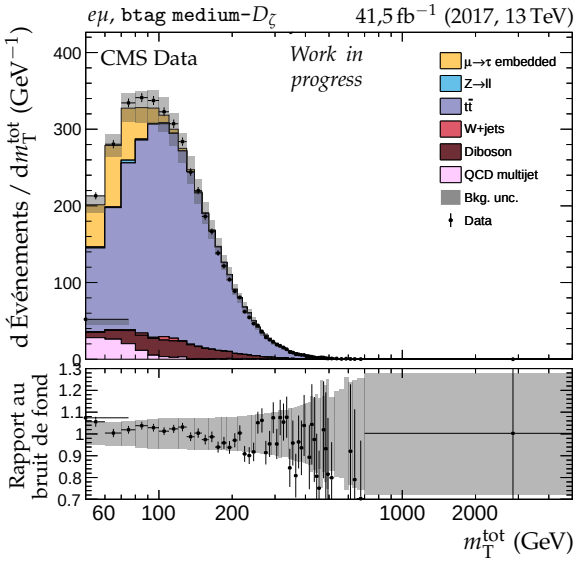
avec

$$
m_{\mathrm{T}}(A,B) = \sqrt{2\,p_{\mathrm{T}}{}^{A}\,p_{\mathrm{T}}{}^{B}\,(1-\cos\Delta\phi^{(A,B)})}\,. \tag{4.25}
$$

À titre d'illustration, les distributions obtenues pour la catégorie `btag tight-`m_{T} du canal $\mu\tau_{\mathrm{h}}$ et `btag medium-`D_ζ du canal $e\mu$ en 2017 sont représentées en figure 4.16.



(a) *Canal* $\mu\tau_{\mathrm{h}}$, *catégorie* `btag tight-`m_{T}.



(b) *Canal* $e\mu$, *catégorie* `btag medium-`D_ζ.

Figure 4.16 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *en 2017 pour deux catégories et canaux. Les données observées (points noirs) sont comparées à la modélisation des bruits de fond (histogrammes remplis en couleur et empilés). Les bandes grisées correspondent à l'incertitude totale (statistique et systématique) sur le bruit de fond avant ajustement des paramètres de nuisance par* Combine. *Le rapport au bruit de fond est donné dans la partie inférieure des graphiques.*

5.2 Catégories « SM »

5.2.1 Définition des catégories

Les catégories `SM`, introduites dans les références [227, 228], sont construites dans le but d'étudier le boson de Higgs du SM h de masse 125 GeV. Cette catégorisation est faite à l'aide d'un réseau de

neurones dont l'objectif est de définir différentes catégories d'événements, chacune contenant un processus physique dominant. Le principe des réseaux de neurones est abordé dans le chapitre 5. Le réseau utilisé est ici décrit succinctement, plus de détails sont disponibles dans la référence [228].

Structure du réseau de neurones Les principales variables d'entrée du réseau sont :
- les impulsions transverses des éléments du dilepton ;
- la masse transverse du dilepton dans le cas du canal $e\mu$ ($m_{\mathrm{T}}(\vec{p}_{\mathrm{T}}{}^{e} + \vec{p}_{\mathrm{T}}{}^{\mu}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}})$) ;
- les impulsions transverses des deux principaux jets de l'événement ;
- le nombre de jets N_{jets} ;
- le nombre de jets de quarks b $N_{b\text{-jets}}$;
- la masse invariante du système des deux jets principaux m_{jj} ;
- la distance dans le plan (η, ϕ) entre les deux jets principaux $\Delta\eta_{\mathrm{jj}}$;
- l'impulsion transverse totale des deux principaux jets de l'événement ;
- la masse du dilepton estimée par SVFIT [229], un estimateur de la masse d'une particule se désintégrant en paire de leptons τ par ajustement d'un profil de vraisemblance, m_{SVFIT} ;
- la masse invariante du dilepton, $m^{(\tau\tau)}$;
- l'impulsion transverse du dilepton, $p_{\mathrm{T}}{}^{(\tau\tau)}$.

Le réseau est constitué de deux couches cachées de 200 neurones chacune, complètement connectées. Leur fonction d'activation est la tangente hyperbolique.

Afin de permettre une catégorisation plus poussée qu'une simple discrimination signal ou bruit de fond, la couche de sortie du réseau contient autant de neurones que de catégories souhaitées. Ce réseau fournit donc un vecteur et non un scalaire. La fonction d'activation de ces neurones de sortie est la fonction exponentielle normalisée ou *Softmax*,

$$Softmax(\vec{z})_j = \frac{\exp(z_j)}{\sum_{k=1}^{n} \exp(z_k)}, \quad j \in \{1, \ldots, K\}, \tag{4.26}$$

chaque composante de ce vecteur est alors la probabilité que l'événement appartienne à la catégorie correspondante.

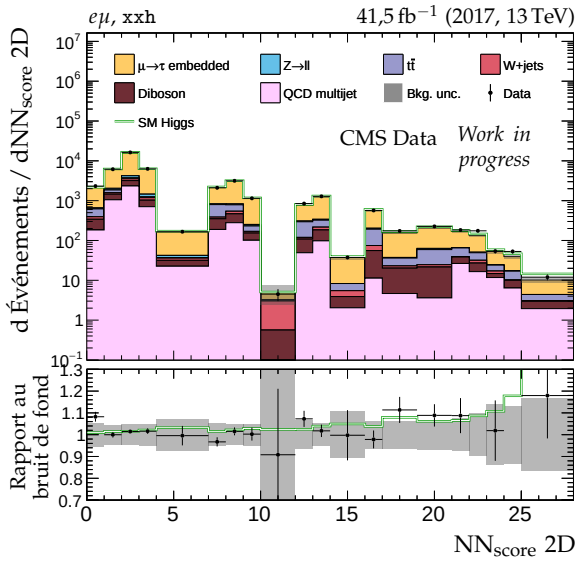
Catégories obtenues Pour chaque canal, deux catégories de signal existent visant chacune certains modes de production du boson de Higgs :
- `ggh` : production par fusion de gluons (ggh) ;
- `qqh` : production par fusion de bosons vecteurs (VBF) ou en association avec un boson (VH).

Ces modes de production sont introduits dans le chapitre 1.

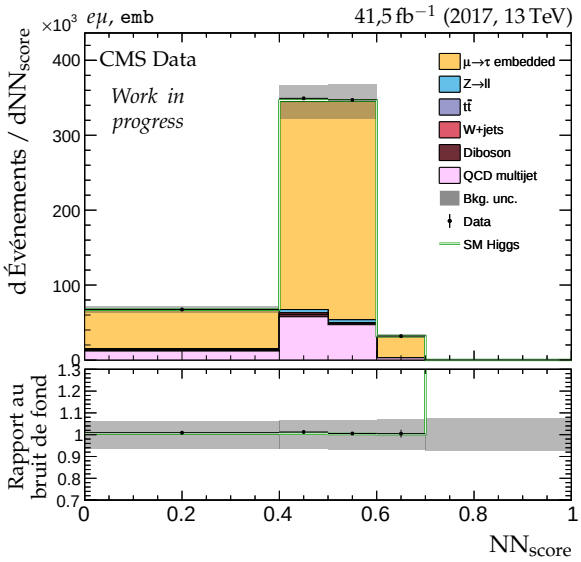
Une catégorie est également définie pour chacun des principaux bruits de fond, présentés dans la section 4. La catégorie `emb` doit correspondre aux données encapsulées décrites dans la section 4.1. La catégorie `fake`, quant à elle, doit contenir les événements décrits par la méthode des facteurs de faux présentée section 4.3. Pour les bruits de fond ayant une faible contribution ou étant peu différentiables d'autres bruits de fond, une catégorie `misc` est également définie. Les différentes catégories ainsi possibles sont listées dans le tableau 4.4. Le canal $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ ne devant pas contenir d'électron ni de muons, les processus $Z \to \ell\ell$ ($\ell \in \{e, \mu\}$), $t\bar{t}$ et W + jets contribuent peu au bruit de fond, c'est pourquoi il n'existe pas de catégories leur étant dédiées dans ce canal. Ils sont donc associés à la catégorie `misc` pour le canal $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$. Celle-ci couvre donc les processus $Z \to \ell\ell$, $t\bar{t}$, W + jets et Diboson dans le canal $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$; Diboson dans les canaux $\mu\tau_{\mathrm{h}}$ et $e\tau_{\mathrm{h}}$; $Z \to \ell\ell$ et W + jets dans le canal $e\mu$. Les processus EWK $Z \to LL$ et EWK $Z \to \nu\nu$, introduits dans l'annexe E, sont également associés à la catégorie `misc` dans les canaux $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $\mu\tau_{\mathrm{h}}$ et $e\tau_{\mathrm{h}}$.

5.2.2 Variable discriminante

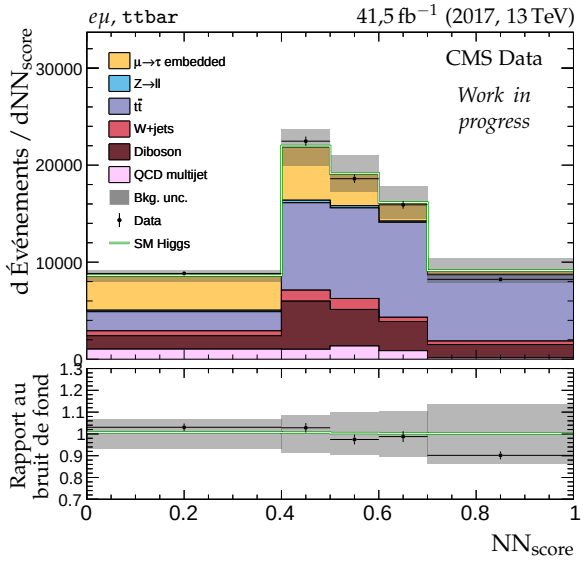
Le réseau de neurones utilisé a pour but de classer les événements selon leur nature dans les différentes catégories définies précédemment. De l'entraînement de ce réseau résultent les frontières entre les différentes catégories. Les régions frontalières, c'est-à-dire les régions de l'espace des phases dans lesquelles les événements ont de proches probabilités d'appartenir à deux catégories ou plus,



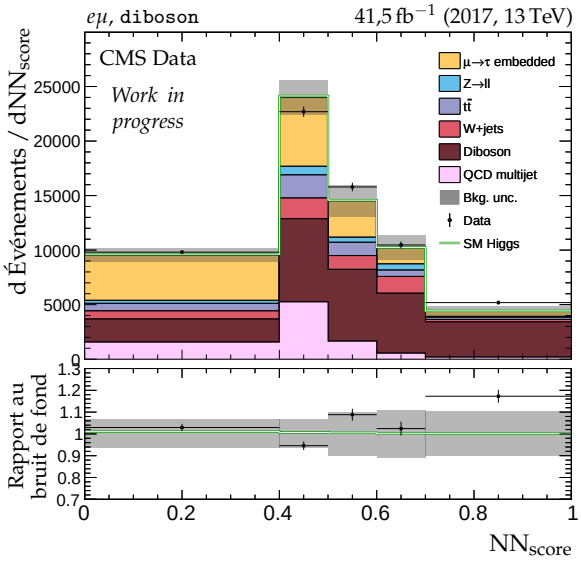
(a) *Catégorie* `xxh`.



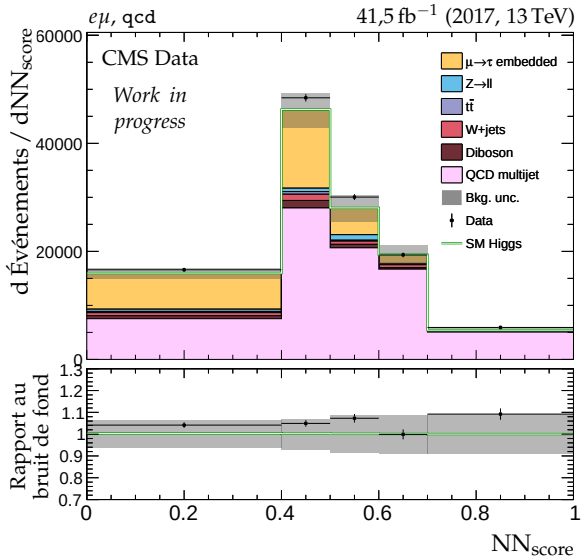
(b) *Catégorie* `emb`.



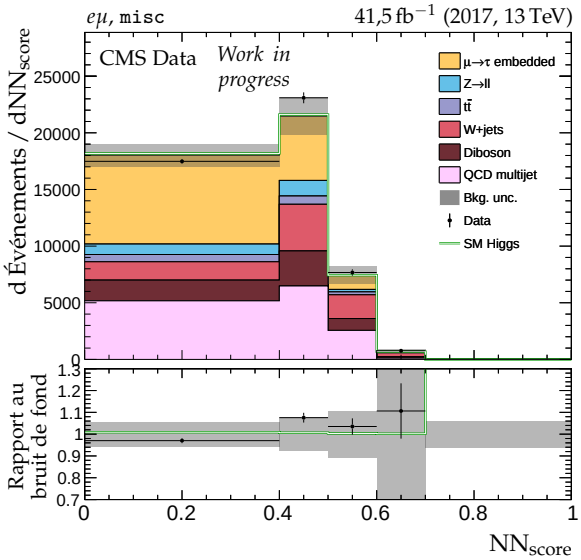
(c) *Catégorie* `ttbar`.



(d) *Catégorie* `diboson`.



(e) *Catégorie* `qcd`.



(f) *Catégorie* `misc`.

Figure 4.17 – *Distributions de* NN$_{score}$ *en 2017 dans le canal eμ.*

Canal	Catégories de bruit de fond possibles					
$\tau_h\tau_h$	emb				fake	misc
$\mu\tau_h$	emb	zll	ttbar		fake	misc
$e\tau_h$	emb	zll	ttbar		fake	misc
$e\mu$	emb		ttbar	diboson	qcd	misc

Tableau 4.4 – *Catégories SM de bruit de fond pour les quatre canaux considérés.*

sont ainsi délibérément fixées lors de l'entraînement et en dépendent. Modifier les paramètres du réseau ou de l'entraînement mène ainsi à des migrations d'événements frontaliers d'une catégorie à une autre. Or, ces derniers sont ceux dont la probabilité d'appartenir à une catégorie ne peut être grande, car dans ce cas ils seraient très caractéristiques de cette catégorie.

Il est donc pertinent d'utiliser les valeurs de sortie du réseau pour définir une variable discriminante. Le choix fait est d'utiliser la plus grande probabilité parmi celles données par le réseau, c'est-à-dire la probabilité d'appartenir à la catégorie dans laquelle le réseau estime que cet événement fait partie. Cette variable est dénommée « score » et notée $\mathrm{NN}_{\mathrm{score}}$. À titre d'illustration, les distributions obtenues pour le canal $e\mu$ en 2017 sont représentées en figure 4.17.

Certains événements de signal peuvent être difficilement classés dans une unique catégorie de signal, ggh ou qqh. Dans ce cas, leurs scores sont bas pour ces catégories et ils se retrouvent dans un segment de l'histogramme contenant d'avantage de bruit de fond que s'ils avaient obtenu un score plus élevé. Afin de limiter cet effet, la solution trouvée [227] est de créer une catégorie globale xxh = ggh + qqh. Le score dans cette catégorie xxh est alors bidimensionnel, chaque dimension correspondant à un des deux scores des catégories ggh et qqh. Pour obtenir un histogramme à une dimension, une réduction est réalisée tel qu'illustré figure 4.18.

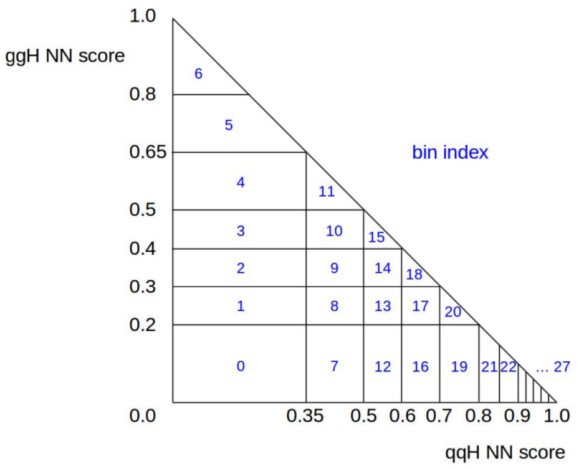


Figure 4.18 – *Réduction à une dimension de la catégorie xxh [227].*

La segmentation est à peu près uniforme en fonction du score ggh. Aux bas scores qqh, elle est plus large à cause des larges contributions du bruit de fond ainsi que de la quantité réduite d'événements. Aux hauts scores qqh, elle est plus fine car le signal y est fortement présent. L'indice du $\mathrm{NN}_{\mathrm{score}}$ 2D ainsi obtenu est utilisé comme variable discriminante dans la catégorie xxh. Comme pour toutes les autres distributions, une resegmentation automatique est réalisée afin de s'assurer que chaque segment contienne au moins dix événements de bruit de fond. C'est pourquoi les distributions peuvent montrer des segmentations variables, en particulier moins fines que celles initialement définies.

5.3 Combinaison des catégories SM et BSM

Les catégories BSM sont définies pour être sensibles au signal de H et A, tandis que les SM le sont pour celui de h. L'utilisation combinée des catégories SM et BSM rend alors l'analyse plus sensible aux propriétés du boson de Higgs du SM h par rapport à l'utilisation des catégories BSM uniquement.

Cette catégorisation combinée est une innovation importante et non triviale par rapport à la catégorisation classique « BSM uniquement » utilisée dans les précédentes analyses $H \to \tau\tau$ dans le cadre du MSSM [81, 202, 203]. En effet, les propriétés de h sont modifiées dans le MSSM par rapport au SM, comme exposé dans le chapitre 1. La modélisation de h dans le cadre du MSSM est décrite dans la section 7. La complémentarité de la recherche du signal de H et A avec le test des propriétés de h permet donc d'obtenir de plus fortes contraintes sur les modèles testés, comme cela a déjà été constaté dans des travaux récents [230].

Afin d'éviter tout recouvrement entre les catégories SM et BSM lors de leur utilisation combinée, deux régions sont définies, chacune utilisant les catégories correspondantes. La région SM concerne les événements ne comportant pas de jets issus de quark b ($N_{b\text{-jets}} = 0$) et tels que $m_{\text{SVFIT}} < 250\,\text{GeV}$ où m_{SVFIT} est la masse du dilepton estimée par SVFIT [229]. La région BSM, quant à elle, concerne les événements contenant des jets issus de quark b ($N_{b\text{-jets}} \geqslant 1$) ou tels que $m_{\text{SVFIT}} \geqslant 250\,\text{GeV}$. Les deux régions ainsi obtenues ne se recouvrent pas et peuvent se résumer selon le schéma de la figure 4.19.

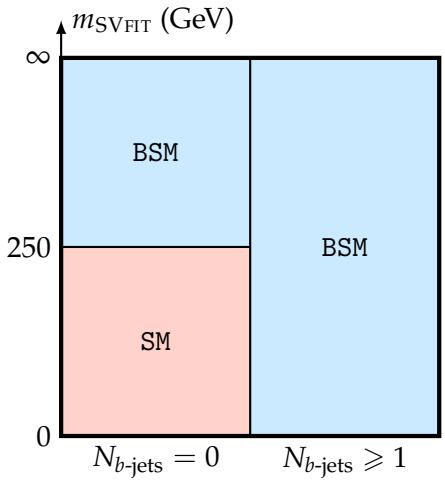


Figure 4.19 – *Définition des deux régions utilisant des catégories différentes.*

6 Incertitudes systématiques

La statistique n'est pas la seule source d'incertitudes sur les distributions de l'estimation des bruits de fond et du signal. Des incertitudes expérimentales, liées à la reconstruction des objets physiques et leur identification par exemple, sont à prendre en compte. Des incertitudes théoriques existent également. Chaque incertitude peut affecter les distributions sous la forme d'un facteur de normalisation, d'une modification des formes de celles-ci voire des deux. Les incertitudes de normalisation sont présentées dans la section 6.1, celles pouvant modifier les formes des distributions dans la section 6.2. Il en résulte des paramètres de nuisance, exploités dans la section 7.

6.1 Incertitudes de normalisation

Luminosité L'incertitude sur la luminosité enregistrée est de 2,5 % en 2016 [108], 2,3 % en 2017 [109] et 2,5 % en 2018 [110]. Elle est donnée par le *POG* Lumi. Plus de détails sont disponibles dans la référence [231]. L'incertitude pour l'année 2016 a récemment été abaissée à 1,2 % [232], les résultats présentés ici n'en tiennent pas compte.

Pondération du pré-déclenchement Les variations hautes et basses dues à cet effet introduit dans la section 2 sont données par le *POG* L1 DPG. Elles concernent tous les jeux de données simulées des années 2016 et 2017. L'incertitude obtenue est de l'ordre de 1 %. Les paramètres de nuisance associés sont corrélés entre les années et les canaux.

Taux de mauvaise identification $\ell \to \tau_h$ ($\ell \to \tau_h$ *fake rate*) L'incertitude, fournie par le *POG* tau, dépend de la pseudo-rapidité η du lepton ℓ. Elle est décorrélée entre les différentes parties du détecteur. L'effet sur la forme des distributions est négligeable face à l'effet de normalisation, c'est pourquoi cette incertitude est traitée comme une normalisation. Les paramètres de nuisance associés sont non corrélés.

Efficacité d'identification des muons et des électrons (*muon/electron ID efficiency*) Une incertitude estimée à 2 % sur le facteur d'échelle introduit dans la section 2 est considérée. Elle est appliquée à tous les processus estimés à partir de données simulées ou encapsulées, corrélées à 50 %. Les paramètres de nuisance associés sont corrélés.

Efficacité du *b-tagging* (*Btag efficiency*) Les facteurs d'échelle fournis par le *POG* BTV [188] comportent des incertitudes dépendantes de la région du détecteur. L'efficacité d'identification et le taux de positifs donnent ainsi deux paramètres de nuisance, dont l'effet de forme est négligeable face à l'effet de normalisation. Les paramètres de nuisance associés sont non corrélés.

Bruits de fond simulés Les incertitudes de normalisation sur les bruits de fond simulés sont :
- 2 % sur les processus $Z \to \ell\ell$, due à l'incertitude sur la section efficace Drell-Yan [233], corrélée entre les années ;
- 4 % sur les processus W + jets, due à l'incertitude sur leurs sections efficaces [233], corrélée entre les années ;
- 5 % sur les processus Diboson et *Single top*, due à l'incertitude sur leurs sections efficaces [233], corrélée entre les années ;
- l'extrapolation sur l'acceptation des simulations dans la CR $t\bar{t}$, prise à 1 % ;
- 4 % (2 % par muon) dans les données encapsulées pour rendre compte de l'efficacité du chemin de déclenchement « double muon », corrélée entre les canaux et décorrélée entre les années ;
- lorsque le boson de Higgs du SM h est considéré comme faisant partie des bruits de fond, les incertitudes sur les sections efficaces de sa production recommandées dans la référence [86] sont appliquées ;

Incertitudes théoriques Pour les limites dépendantes d'un modèle dans le plan $(m_A, \tan\beta)$, les incertitudes théoriques sur la section efficace de production des bosons de Higgs du MSSM sont inclues. Elles sont fournies par le groupe LHC Higgs [234].

Dans le cas du processus $gg \to b\bar{b}\Phi \to \tau\tau$ avec $\Phi \in \{ H, A \}$, l'incertitude sur l'acceptation en fonction de $N_{b\text{-jets}}$ est estimée en faisant varier le paramètre `hdamp` du générateur de POWHEG par des facteurs $\frac{1}{2}$ et 2. Une incertitude sur la QCD est également considérée afin de rendre compte des termes au-delà du NLO non traités. Elle est estimée en faisant varier les échelles de renormalisation (μ_R) et de refactorisation (μ_F) par des facteurs $\frac{1}{2}$ et 2 indépendamment tout en conservant $\frac{1}{2} \leqslant \frac{\mu_R}{\mu_F} \leqslant 2$. L'incertitude sur l'acceptation du signal dans les catégories `btag` ainsi obtenue est de l'ordre de 1 à 6 % selon le point de masse. Enfin, des incertitudes sur les PDFs, introduites dans le chapitre 2, sont prises en comptes. Elles sont de l'ordre de 1 à 2 %.

6.2 Incertitudes de forme

Efficacité des chemins de déclenchement des τ_h (τ_h *trigger efficiency*) Cette incertitude est déterminée à partir des facteurs d'échelle définis section 2. Des paramètres de nuisance sont définis pour chaque chemin de déclenchement et différents modes de désintégration (DM, *Decay Modes*). L'efficacité du chemin de déclenchement « double τ_h » est de plus déterminée pour les impulsions transverses supérieures et inférieures à 100 GeV afin que l'ajustement réalisé section 7 ait plus de liberté vis-à-vis des régions à bas et haut p_T. Dans le cas du chemin de déclenchement « τ_h seul », le manque de statistiques mène à ne définir qu'un seul paramètre de nuisance commun à tous les DMs.

Tous les processus déterminés par simulation ou encapsulation sont concernés. Une corrélation de 50 % entre les paramètres de nuisance des données simulées et encapsulées est utilisée, l'encapsulation étant un hybride entre données réelles et simulées. Les paramètres de nuisance sont non corrélés entre les canaux et les années.

Efficacité des chemins de déclenchement des muons et des électrons (*lepton trigger efficiency*) L'incertitude sur l'efficacité des chemins de déclenchement des muons et des électrons est de 2 % par lepton. Il s'agit en première approximation d'une incertitude de normalisation, traitée ici comme une incertitude de forme car elle ne concerne que les événements dans lesquels un chemin de déclenchement basé sur ces leptons est utilisé.

Tous les processus déterminés par simulation ou encapsulation sont concernés. Les paramètres de nuisance associés sont non corrélés entre données simulées et encapsulées. Ils ne le sont pas non plus entre les canaux et années, car différents chemins de déclenchement et sélections sont utilisés.

Efficacité d'identification des τ_h (τ_h *ID efficiency*) L'incertitude sur l'efficacité d'identification des τ_h est donnée par le *POG tau* [189] en fonction de l'impulsion transverse du τ_h et de son DM. À chaque DM correspond un paramètre de nuisance, corrélé entre les différents canaux ($\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$) et non corrélé entre les années. Afin de rendre compte des différents points de fonctionnement des

discriminateurs anti-lepton, une incertitude supplémentaire de 3 % par τ_h est appliquée de manière non corrélée entre les canaux. Dans le cas des données encapsulées, la même procédure est suivie mais une corrélation de 50 % avec les données simulées est utilisée.

Efficacité du trajectographe pour les τ_h des données encapsulées (*embedded τ_h tracking efficiency*) L'incertitude sur cette efficacité, corrélée entre les canaux, corrélée à 50 % entre les années et décorrélée entre les DMs 0, 1, 2 et 10, 11, est fournie par le *POG tau*.

Repondération de l'impulsion transverse et de la masse du boson Z (*DY p_T-mass reweighting*) L'incertitude est déterminée à partir de la variation entre zéro et deux fois la correction correspondante, introduite dans la section 2. L'incertitude à 1σ correspond alors à 10 % de cette variation et est appliquée sur les événements $Z \to \ell\ell$ dans tous les canaux. Les paramètres de nuisance des années 2017 et 2018 sont corrélés, car les mêmes réglages de simulation sont utilisés (CP5 [197]). Celui de l'année 2016 est décorrélé.

Repondération de l'impulsion transverse du quark t (*top p_T reweighting*) L'incertitude considérée est la variation entre zéro et deux fois la correction correspondante, introduite dans la section 2. Les paramètres de nuisance des différentes années sont corrélés.

Recul de E_T^{miss} (*MET recoil correction uncertainty*) Dans les processus physiques concernés par la correction de recul de E_T^{miss}, c'est-à-dire ceux de production de bosons de Higgs, de Drell-Yan (boson Z) et de W + jets, la réponse en énergie des hadrons est modifiée selon l'incertitude déterminée sur cette correction. Les paramètres de nuisance des différentes années sont non corrélés.

E_T^{miss} non regroupée (*MET unclustered uncertainty*) L'algorithme de PF introduit dans le chapitre 2 peut fournir des objets physiques candidats n'étant pas identifiés comme des muons, électrons, photons, hadrons ou jets. Il s'agit par exemple de particules de très bas p_T. Ces candidats sont toutefois utilisés dans le calcul de E_T^{miss}, il s'agit de l'énergie transverse manquante non regroupée (*MET unclustered*). L'incertitude sur cette observable est appliquée à toutes les données simulées non concernées par la correction de recul de E_T^{miss} comme le recommande le *POG JetMET* [235]. Les différentes années sont non corrélées.

Énergie des jets (*jet energy scale*) Comme proposé par le *POG JetMET*, au lieu d'une seule source d'incertitude globale, 11 paramètres de nuisance sont considérés. Certains d'entre eux sont corrélés entre les années. Pour les processus physiques non concernés par la correction de recul de E_T^{miss}, c'est-à-dire $t\bar{t}$, Diboson et *Single top*, la variation en énergie des jets est propagés à E_T^{miss}, ainsi qu'aux variables en dépendant comme m_T^{tot}.

Résolution sur l'énergie des jets (*jet energy resolution*) L'incertitude sur la résolution en énergie des jets donnée par le module fourni par la collaboration CMS [236] est appliquée aux jeux de données simulées. Comme pour l'incertitude sur l'énergie des jets, la propagation à E_T^{miss} est effectuée pour les processus physiques non concernés par la correction de recul de E_T^{miss}. Cette incertitude est non corrélée entre les années.

Énergie des τ_h (τ_h *energy scale*) Une incertitude de forme est appliquée et dépend du DM du τ_h ainsi que du type de données, simulées ou encapsulées. Un paramètre de nuisance par DM est obtenu.

Dans les données encapsulées, les événements hybrides peuvent présenter des τ_h contenant des dépôts dans les calorimètres provenant du muon initial. Une corrélation de 50 % entre données simulées et encapsulées est alors appliquée. Les années ne sont pas corrélées, comme le suggère le *POG tau* [189].

Énergie des leptons identifiés comme τ_h ($\ell \to \tau_h$ *energy scale*) Une variation sur l'impulsion transverse des leptons identifiés à tort comme des τ_h est appliquée. Elle est de l'ordre de 1 % pour les muons. Pour les électrons, elle dépend de l'année et de la région du détecteur et peut aller de 0,5 à 6,6 %. Les années sont non corrélées.

Contamination $t\bar{t}$ dans les données encapsulées Une partie du bruit de fond $t\bar{t}$ est couvert par les données encapsulées. Il s'agit des événements contenant une paire de leptons τ issus de la désintégration des quarks t. L'incertitude haute (basse) sur cette contamination est obtenue en ajoutant (soustrayant) 10 % de la fraction d'événements simulés $t\bar{t}$ contenant une paire de leptons τ aux événements encapsulés. Cette incertitude est corrélée entre les canaux mais pas entre les années, car le chemin de déclenchement de sélection de la paire de muons pour les données encapsulées change d'une année à l'autre.

Bruits de fond estimés par les facteurs de faux Différentes sources d'incertitudes sont considérées selon le canal. Dans le canal $\tau_h\tau_h$, ces incertitudes sont :

— statistique sur la mesure de FF_Q, déterminée pour chaque région de N_{jets} et p_T^{jet}, non corrélée entre les années ;
— statistique sur les corrections résiduelles de FF_Q, déterminée pour chaque région de N_{jets}, non corrélée entre les années ;
— systématique sur l'extrapolation OS/SS de FF_Q, l'incertitude haute (basse) est obtenue en appliquant deux (zéro) fois la correction, corrélée entre les années ;
— systématique sur l'utilisation dans ce canal de FF_Q comme FF global, c'est-à-dire appliqué aussi aux événements W + jets et $t\bar{t}$. L'incertitude est de 20 % pour les événements W + jets et 40 % pour $t\bar{t}$, corrélée entre les années.

Dans les canaux $\mu\tau_h$ et $e\tau_h$, ces incertitudes sont :

— statistique sur les mesures des FF_i, déterminées pour chaque région de N_{jets} et p_T^{jet}, non corrélées entre les années ni entre elles ;
— statistique sur les corrections résiduelles des FF_i, déterminées pour chaque région de N_{jets}, non corrélées entre les années ni entre elles ;
— systématique sur l'extrapolation OS/SS de FF_Q, l'incertitude haute (basse) est obtenue en appliquant deux (zéro) fois la correction, corrélée entre les années ;
— systématique sur l'extrapolation isolé/anti-isolé de FF_Q, corrélée entre les années ;
— systématique sur l'extrapolation de $m_T^\ell > 70\,\text{GeV}$ à $m_T^\ell < 70\,\text{GeV}$ de FF_W, l'incertitude haute (basse) est obtenue en appliquant deux (zéro) fois la correction, corrélée entre les années ;
— systématique sur la détermination de FF_t à partir de données simulées, obtenue à partir de la différence entre FF_W déterminé avec des données réelles et simulées, corrélée entre les années.

Estimation du bruit de fond QCD Cette estimation, utilisée dans le canal $e\mu$, comporte dix sources d'incertitudes, dues aux facteurs d'extrapolations des ajustement à des polynômes d'ordre 2. Elles sont non corrélées entre les années.

Incertitudes de segmentation (*bin-by-bin uncertainties*) Les incertitudes de forme dues à la statistique des bruits de fond suite à l'utilisation conjointe des données simulées, encapsulées et réelles dans leurs estimations sont prises en compte par la fonction `autoMCstats` de COMBINE.

Incertitudes théoriques Dans le cas du processus $gg \to \Phi \to \tau\tau$, en particulier dans les cas de basse masse, les variations du paramètre `hdamp` du générateur de POWHEG ainsi que celles de μ_R/μ_F sont utilisées afin d'obtenir l'incertitude sur la distribution en p_T de chacune des contributions au signal NLO de production de bosons de Higgs par fusion de gluon, dont l'estimation est présentée dans la section 2.

7 Extraction du signal

Afin de déterminer si un signal est présent ou non compte-tenu des observations, un ajustement segmenté de maximum de vraisemblance (*binned maximum likelihood fit*) est réalisé sur les catégories présentées section 5 pour deux hypothèses :

— aucun signal, c'est-à-dire uniquement des bruits de fond, notée $\mathfrak{b}$;
— présence d'un signal en plus des bruits de fond, notée $\mathfrak{sb}$.

Le test statistique de ces deux hypothèses est fait par une approche fréquentiste modifiée connue sous le nom de méthode $CL_\mathfrak{s}$ [237-239], implémentée dans COMBINE.

Le modèle de vraisemblance utilisé est détaillé dans la section 7.1. La méthode $CL_\mathfrak{s}$ est présentée dans la section 7.2. L'application de cette méthode pour l'obtention de limites indépendantes d'un modèle est introduite section 7.3. Enfin, la section 7.4 expose l'interprétation de l'analyse dans le cas de scénarios spécifiques du MSSM.

7.1 Modèle de vraisemblance

La fonction de vraisemblance $\mathfrak{L}$ à maximiser est définie par le produit des probabilités poissoniennes $\mathfrak{P}(n_i|\nu_i(\mu,\vec{\theta}))$ d'observer n_i événements dans chaque segment i de l'histogramme de la variable discriminante utilisée selon

$$\mathfrak{L}(\{n_i\}|\mu,\vec{\theta}) = \prod_i \mathfrak{P}(n_i|\nu_i(\mu,\vec{\theta})) \cdot \prod_j \mathfrak{C}(\theta_j,\tilde{\theta}_j)\,, \quad \mathfrak{P}(n_i|\nu_i(\mu,\vec{\theta})) = \frac{\nu_i^{n_i}}{n_i!}\mathrm{e}^{-\nu_i} \tag{4.27}$$

où

— ν_i est le nombre d'événements attendus dans ce segment dans l'hypothèse $\mathfrak{sb}$, c'est-à-dire

$$\nu_i(\mu,\vec{\theta}) = \mu\, s_i(\vec{\theta}) + b_i(\vec{\theta}) \tag{4.28}$$

avec s le nombre d'événements de signal et b le nombre d'événements de bruit de fond. Lorsque $\mu = 0$, ν_i correspond donc au nombre d'événements attendus dans l'hypothèse $\mathfrak{b}$;
— μ est le modificateur d'intensité du signal (*signal strength modifier*). Il représente la fréquence du signal, indéterminée, par rapport à une section efficace de référence, par exemple la section efficace de production du boson de Higgs h ;
— $\vec{\theta}$ sont les paramètres de nuisance correspondant aux sources d'incertitude présentées section 6. Les variations de ces paramètres changent la quantité d'événements de signal s_i et de bruit de fond b_i attendus dans le segment i ;
— j est un indice courant sur les différentes contraintes $\mathfrak{C}$ connues sur les paramètres de nuisance, discutées ci-après. Chacune de ces contraintes représente la probabilité que ce paramètre prenne la valeur θ_j, sachant que la meilleure estimation de ce dernier est $\tilde{\theta}_j$, obtenue par des mesures annexes.

Pour une valeur de μ donnée, les paramètres de nuisance sont ajustés de manière à maximiser $\mathfrak{L}$. Il peut être intéressant de faire varier l'un des paramètres de plusieurs déviations standard pour maximiser le produit des poissoniennes $\mathfrak{P}$, mais les contraintes $\mathfrak{C}$ rendent compte de la faible probabilité que ce choix soit juste. Les variations des paramètres de nuisance sont généralement inférieures à une déviation standard.

7.1.1 Incertitudes de normalisation

Les contraintes sur les incertitudes correspondant à des facteurs multiplicatifs sur la quantité d'événements de signal ou de bruit de fond, par exemple les facteurs d'échelle, sont représentées par des fonctions de densité de probabilité log-normales,

$$\mathfrak{C}(\theta,\tilde{\theta})\big|_{\text{facteurs}} = \frac{1}{\sqrt{2\pi}\ln\kappa}\,\frac{1}{\theta}\,\exp\left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln\kappa)^2}\right) \tag{4.29}$$

où κ vaut $1+x$ avec x l'incertitude relative sur l'observable contrainte. Par exemple, pour une incertitude de 10 %, $\kappa = 1{,}10$.

Les contraintes sur les incertitudes d'origine statistique, par exemple les quantités d'événements observés dans les régions de contrôle, sont représentées par des fonctions de densité de probabilité Gamma,

$$\mathfrak{C}(\theta,\tilde{\theta})\big|_{\text{stat}} = \frac{1}{\kappa\,\Gamma(\tilde{\theta}+1)}\left(\frac{\theta}{\kappa}\right)^{\tilde{\theta}}\exp\left(-\frac{\theta}{\kappa}\right) \tag{4.30}$$

avec Γ la fonction gamma et κ le rapport attendu entre θ et $\tilde{\theta}$. La valeur de κ a sa propre incertitude, généralement traitée comme une contrainte log-normale supplémentaire.

7.1.2 Incertitudes de forme

Les incertitudes systématiques de forme sur les distributions des variables discriminantes du signal ainsi que du bruit de fond sont traitées par la technique du « morphing vertical ». Pour chaque source d'incertitude, une distribution centrale (ou nominale) ainsi que celles correspondant à des variations de $\pm1\sigma$ de l'incertitude sont déterminées. Un paramètre de nuisance λ est ajouté au modèle de vraisemblance afin d'interpoler entre ces différentes distributions.

Les effets de plusieurs incertitudes de forme sont additifs. Soient h_0 la distribution centrale, h_j^+ (h_j^-) la distribution correspondant à une variation de $+1\sigma$ (-1σ) de l'incertitude j et λ_j le paramètre de nuisance ainsi obtenu. Le modèle de distribution est donné par

$$h(\vec{\lambda}) = h_0 + \sum_j \left(a(\lambda_j)h_j^+ + b(\lambda_j)h_0 + c(\lambda_j)h_j^-\right) \tag{4.31}$$

avec

$$a = \begin{cases} \lambda(\lambda+1)/2\,, & |\lambda| \leqslant 1\,, \\ 0\,, & \lambda < -1\,, \\ \lambda\,, & \lambda > +1\,, \end{cases} \qquad b = \begin{cases} -\lambda^2\,, & |\lambda| \leqslant 1\,, \\ -|\lambda|\,, & |\lambda| > 1\,, \end{cases} \qquad c = \begin{cases} \lambda(\lambda-1)/2\,, & |\lambda| \leqslant 1\,, \\ |\lambda|\,, & \lambda < -1\,, \\ 0\,, & \lambda > +1\,. \end{cases} \tag{4.32}$$

L'interpolation (4.31) est réalisée lors de la maximisation de la fonction de vraisemblance.

7.1.3 Incertitudes statistiques

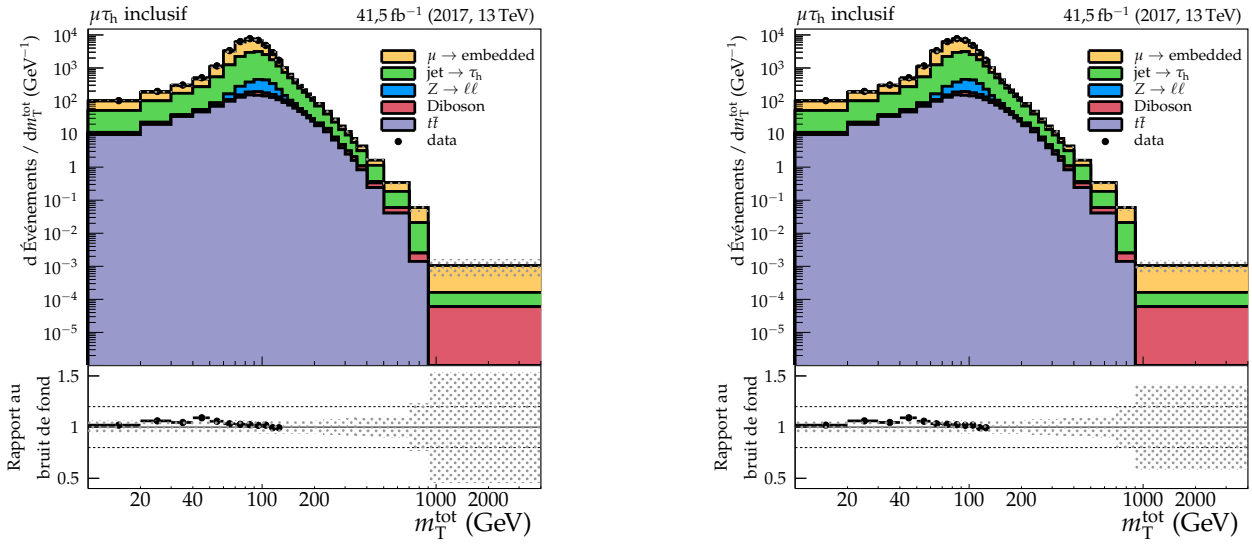
Principe L'incertitude statistique dans les distributions des variables discriminantes est prise en compte par la méthode de Barlow-Beeston [240, 241]. La quantité d'événements dans chaque segment peut varier dans l'incertitude statistique type, ce qui revient à créer une incertitude de forme.

Afin de réduire la quantité de paramètres de nuisance, et donc le temps de calcul, la procédure suivante est suivie dans chaque segment :

1. Les processus i contenant x_i événements et une incertitude statistique e_i tels que e_i/x_i est supérieur à une valeur `AddThreshold` choisie sont sélectionnés.
2. L'incertitude totale e_{tot} sur l'ensemble de ces processus est déterminée selon
$$e_{\text{tot}}^2 = \sum_{j\in\{\,i\,\}} e_j^2\,. \tag{4.33}$$
3. Les processus i sont classés par valeur croissante de e_i^2/e_{tot}^2.
4. Dans l'ordre des processus obtenu, les incertitudes statistiques sont supprimées tant que la somme de leurs carrés est inférieure à une fraction `merge_threshold` choisie de l'incertitude totale au carré.
5. Les incertitudes restantes sont multipliées par un facteur permettant de conserver une incertitude totale constante.

Il s'agit donc de regrouper les incertitudes.

Contribution personnelle Lors de ma thèse, j'ai observé que l'incertitude totale pouvait varier lors de cette procédure, comme cela est illustré sur la figure 4.20. Dans le dernier segment, il apparaît sur le rapport données sur bruit de fond que l'incertitude totale sur le bruit de fond est modifiée par le regroupement. Il s'agissait d'un bug que j'ai identifié et corrigé [215] dans le code de COMBINE.



(a) *Sans regroupement.*

(b) *Avec regroupement.*

Figure 4.20 – *Distributions de m_T^{tot} avec et sans regroupement des incertitudes pour le canal $\mu\tau_h$ en 2017. Le tracé des données s'arrête à* 130 GeV*, avant la zone où le signal est attendu.*

Dans le code initial, pour chaque segment des distributions dans chaque catégorie, les processus i peuvent être classés dans cinq groupes :

groupe Z $x_i = 0$ et $e_i = 0$ (processus non présent dans le segment) ou e_i / x_i inférieur à `AddThreshold`, non traités par la procédure de regroupement ;

groupe A incertitude à supprimer et $0 < e_i < x_i$;

groupe B incertitude à conserver et $0 < e_i < x_i$;

groupe C incertitude à supprimer et $0 < x_i \leqslant e_i$;

groupe D incertitude à conserver et $0 < x_i \leqslant e_i$.

Les processus tels que $0 < e_i < x_i$ (groupes A et B) possèdent un attribut `can_expand = true` et sont ceux dont l'incertitude statistique est renormalisée (*expand*) à l'étape 5 par un facteur

$$\texttt{expand = std::sqrt(1. / (1. - (removed / tot_bbb_added)))} \Leftrightarrow E = \sqrt{\frac{1}{1 - \frac{R}{T}}} \tag{4.34}$$

avec

$$R = \texttt{removed} = \sum_{i \in \{\,\mathrm{A,C}\,\}} e_i^2\,, \quad T = \texttt{tot_bbb_added} = \sum_{i \in \{\,\mathrm{A,B}\,\}} e_i^2\,. \tag{4.35}$$

Ainsi, l'incertitude totale après regroupement s'exprime en fonction des incertitudes de chaque processus i avant regroupement selon

$$\begin{aligned} e^2_{\text{tot,après}} &= \sum_{i \in \{\,\mathrm{A}\,\}} (E \times 0 \times e_i)^2 + \sum_{i \in \{\,\mathrm{B}\,\}} (E \times e_i)^2 + \sum_{i \in \{\,\mathrm{C}\,\}} (0 \times e_i)^2 + \sum_{i \in \{\,\mathrm{D,Z}\,\}} (e_i)^2 \\ &= E^2 \sum_{i \in \{\,\mathrm{B}\,\}} e_i^2 + \sum_{i \in \{\,\mathrm{D,Z}\,\}} e_i^2\,. \end{aligned} \tag{4.36}$$

Or,

$$E^2 = \frac{1}{1 - \frac{R}{T}} = \frac{T}{T - R} = \frac{\sum_{i \in \{\,\mathrm{A,B}\,\}} e_i^2}{\sum_{i \in \{\,\mathrm{A,B}\,\}} e_i^2 - \sum_{i \in \{\,\mathrm{A,C}\,\}} e_i^2} = \frac{\sum_{i \in \{\,\mathrm{A,B}\,\}} e_i^2}{\sum_{i \in \{\,\mathrm{B}\,\}} e_i^2 - \sum_{i \in \{\,\mathrm{C}\,\}} e_i^2} \tag{4.37}$$

soit

$$e^2_{\text{tot,après}} = \frac{\sum_{i \in \{\,\mathrm{A,B}\,\}} e_i^2}{\sum_{i \in \{\,\mathrm{B}\,\}} e_i^2 - \sum_{i \in \{\,\mathrm{C}\,\}} e_i^2} \times \sum_{i \in \{\,\mathrm{B}\,\}} e_i^2 + \sum_{i \in \{\,\mathrm{D,Z}\,\}} e_i^2 \tag{4.38}$$

ce qui est différent de l'erreur initiale dans le cas général. Cette formule a été testée numériquement, ce qui a permis de confirmer la bonne compréhension du code initial.

Le problème vient du traitement du groupe C, c'est-à-dire des processus dont l'incertitude est supprimée mais dont la quantité d'événement est inférieure à celle-ci. En effet, ils ne sont pas pris en compte dans le calcul de T. Le correctif proposé [215] est de refuser le cas du groupe C et de rediriger ces processus dans le groupe D. Alors, le groupe C étant forcément un ensemble vide, l'équation (4.38) se réécrit

$$e^2_{\text{tot,après}} = \frac{\sum_{i\in\{\,\text{A,B}\,\}} e_i^2}{\sum_{i\in\{\,\text{B}\,\}} e_i^2} \times \sum_{i\in\{\,\text{B}\,\}} e_i^2 + \sum_{i\in\{\,\text{D,Z}\,\}} e_i^2 = \sum_{i\in\{\,\text{A,B}\,\}} e_i^2 + \sum_{i\in\{\,\text{D,Z}\,\}} e_i^2 = e^2_{\text{tot,avant}}\,, \tag{4.39}$$

l'incertitude totale est donc bien conservée.

Dans le cas d'étude de la figure 4.20, ce correctif introduit six paramètres de nuisance supplémentaires, ce qui reste raisonnable en terme de charge computationnelle. En effet, les processus devant entrer dans le groupe C sont peu nombreux.

7.2 Méthode $CL_{\mathfrak{s}}$

7.2.1 Approche fréquentiste classique

Afin de déterminer quantitativement quelle hypothèse, entre $\mathfrak{b}$ et $\mathfrak{sb}$, est la plus compatible avec les résultats de l'analyse, il faut réaliser un test statistique. Plusieurs tests existent, celui utilisé pour les expériences du LHC est le profil du rapport de vraisemblance (*profile likelihood ratio*),

$$q_\mu = -2\ln\left(\frac{\mathfrak{L}(\text{données}|\mu,\hat{\hat{\theta}}_\mu)}{\mathfrak{L}(\text{données}|\hat{\mu},\hat{\hat{\theta}}_{\hat{\mu}})}\right), \quad 0 \leqslant \hat{\mu} \leqslant \mu \tag{4.40}$$

où « données » se réfère aux quantités d'événements n_i dans chaque segment i des distributions des variables discriminantes dans chaque catégorie et $\hat{\hat{\theta}}_x$ est l'ensemble des paramètres de nuisance maximisant $\mathfrak{L}$ pour $\mu = x$. L'ensemble $(\hat{\mu},\hat{\hat{\theta}}_{\hat{\mu}})$ donne le maximum global de $\mathfrak{L}$. La contrainte $0 \leqslant \hat{\mu}$ impose une fréquence du signal positive, c'est-à-dire que μ a une interprétation physique. De plus, $\hat{\mu} \leqslant \mu$ interdit de rejeter μ plus petit que $\hat{\mu}$, valeur la plus probable du modificateur d'intensité du signal. Lorsqu'une valeur de μ est rejetée, toutes les valeurs plus élevées le sont donc également.

Les grandes valeurs de q_μ correspondent ainsi aux cas où la valeur de μ est incompatible avec les données. À l'inverse, lorsque $q_\mu \simeq 0$, les données sont compatibles avec μ dans le cadre de l'hypothèse $\mathfrak{sb}$. La probabilité d'obtenir une valeur de q_μ plus élevée que celle observée q_μ^{obs}, c'est-à-dire de réaliser une observation moins compatible avec l'hypothèse $\mathfrak{sb}$ que celle effectivement réalisée, est définie par

$$CL_{\mathfrak{sb}} = \int_{q_\mu^{\text{obs}}}^{+\infty} f(q_\mu|\mu,\hat{\hat{\theta}}_\mu)\,\mathrm{d}q_\mu \tag{4.41}$$

où f est la fonction de densité de probabilité pour q_μ, obtenue en tirant au sort de nombreuses combinaisons des valeurs des paramètres de nuisance et de μ.

Une valeur de μ est ainsi considérée exclue avec un niveau de confiance α tel que $\alpha = 1 - CL_{\mathfrak{sb}}$. Un niveau de confiance de 95 % est généralement utilisé. Toutefois, en prenant $\mu \simeq 0$, cette approche mène statistiquement 5 % des analyses de physique à exclure la présence d'un signal. Pour éviter ce cas de figure, une approche fréquentiste modifiée est utilisée.

7.2.2 Approche fréquentiste modifiée

Cette méthode est introduite pour traiter le cas d'un signal très faible par rapport au bruit de fond [237-239]. En plus de $CL_{\mathfrak{sb}}$ définie précédemment, la probabilité de réaliser une observation moins compatible avec l'hypothèse $\mathfrak{b}$ que celle effectivement réalisée, définie comme

$$CL_{\mathfrak{b}} = \int_{q_\mu^{\text{obs}}}^{+\infty} f(q_\mu|0,\hat{\hat{\theta}}_0)\,\mathrm{d}q_\mu \tag{4.42}$$

est déterminée. La quantité CL_s est le rapport

$$CL_s = \frac{CL_{s+b}}{CL_b}. \tag{4.43}$$

Dans l'approche fréquentiste modifiée, l'exclusion à 95 % de confiance est obtenue lorsque $CL_s \leqslant 0{,}05$. Plus de détails sur la méthode CL_s sont disponibles dans la référence [242].

7.3 Limites indépendantes du modèle

En l'absence de signal, des limites hautes sont déterminées sur la section efficace σ de production des bosons de Higgs neutres du MSSM H ou A, multipliée par le rapport de branchement $\mathcal{BR}$ à la désintégration en paire de leptons τ. La limite d'exclusion à 95 % de confiance sur $\sigma \times \mathcal{BR}$ est déterminée à partir de la valeur de μ telle que $CL_s = 0{,}05$. Il s'agit donc de la valeur maximale de $\sigma \times \mathcal{BR}$ au-delà de laquelle un signal aurait été considéré comme observé.

La modélisation du signal consiste en un unique boson de Higgs $\Phi \in \{ H, A \}$ avec pour modes de production :

— $gg \to \Phi \to \tau\tau$;
— $gg \to b\bar{b}\Phi \to \tau\tau$.

Lorsque les limites sont déterminées pour l'un de ces modes, aucune hypothèse n'est faite sur l'autre, en particulier sur sa normalisation.

Dans le cas du processus $gg \to \Phi \to \tau\tau$, les contributions à la boucle fermionique des quarks t, b et de leur interférence sont fixées à celles attendues dans le SM. Aucune modification due à $\tan\beta$ n'est donc considérée.

En plus des processus $Z \to \tau\tau$, $Z \to \ell\ell$, W + jets, $t\bar{t}$, Diboson, *Single top* et QCD, le boson de Higgs du SM h dans ses désintégrations $h \to \tau\tau$ et $h \to WW$ est également considéré comme un bruit de fond. Le signal du MSSM est ainsi constitué des processus $gg \to \Phi \to \tau\tau$ et $gg \to b\bar{b}\Phi \to \tau\tau$ avec les coupages du SM aux quarks t et b et $m_\Phi \in [110, 3200]$ GeV.

Il s'agit donc de la recherche d'un boson de Higgs neutre en plus du boson déjà observé. Les catégories `BSM` introduites section 5.1 sont utilisées, sans combinaison avec les catégories `SM`, le boson de Higgs du SM étant considéré comme un bruit de fond.

7.4 Interprétation dans les scénarios du MSSM

Les résultats de l'analyse sont interprétés dans le cadre de scénarios de référence [79] respectant les limites fixées par les expériences du LEP, du Tevatron et du LHC. En particulier, un des bosons de Higgs scalaire doit jouer le rôle du boson découvert en 2012 avec une masse de 125 ± 3 GeV.

7.4.1 Scénario M_h^{125}

Dans le scénario M_h^{125}, les masses des superpartenaires sont suffisamment élevées pour que les processus physiques de production et de désintégration des bosons de Higgs soient peu affectés par leur présence [79]. En particulier, les couplages de h aux superpartenaires sont faibles par rapport à ceux aux particules du SM. De plus, si leurs masses sont inférieures à 2 TeV, H et A se désintègrent uniquement en particules du SM dans ce scénario.

Comme exposé dans le chapitre 1, les deux paramètres permettant de décrire au premier ordre les bosons de Higgs dans le MSSM sont m_A et $\tan\beta$, pris comme paramètres libres. Les autres sont fixés [79] :

— les paramètres de masse des squarks de troisième génération M_{Q_3}, M_{U_3} et M_{D_3} pris à 1,5 TeV ;
— les paramètres de masse des sleptons de troisième génération M_{L_3} et M_{E_3} pris à 2 TeV ;
— le paramètre de masse des Higgsinos $\mu = 1$ TeV ;
— les paramètres de masse des *gauginos* $M_1 = 1$ TeV, $M_2 = 1$ TeV et $M_3 = 2{,}5$ TeV ;
— le paramètre de mélange du stop, $X_t = 2{,}8$ TeV ;
— les constantes de couplage trilinéaire entre les Higgs et le stop, le sbottom et le stau, respectivement A_t, A_b et A_τ, prises égales entre elles.

Le groupe en charge de la physique des bosons de Higgs au LHC fournit en fonction de m_A et $\tan\beta$ les masses des bosons de Higgs, sections efficaces de production, rapports de branchement et incertitudes théoriques associés de ce scénario [234].

Le signal du MSSM attendu (hypothèse $\mathfrak{sb}$) est celui de $h_{\text{MSSM}} + H + A$, h_{MSSM} étant le boson de Higgs léger du MSSM jouant le rôle du boson découvert en 2012 et interprété comme celui du SM (h_{SM}). En particulier, les propriétés de h_{MSSM} peuvent différer de celles de h_{SM}. Pour h_{MSSM}, H et A, la forme du signal est obtenue à partir des valeurs de m_A et $\tan\beta$. Une combinaison linéaire donne le signal total $\Phi \in \{ h_{\text{MSSM}}, H, A \} \to \tau\tau$ attendu.

Les modes de production considérés sont :

— ggh, $b\bar{b}h$, VBF et VH pour h_{MSSM} ;
— ggH, $b\bar{b}H$ pour H ;
— ggA, $b\bar{b}A$ pour A.

Dans le cas des modes de production de h_{MSSM} VBF et VH, le signal attendu est pris comme celui du SM multiplié par

$$\sin^2(\beta-\alpha) \times \frac{\mathcal{BR}(h_{\text{MSSM}} \to \tau\tau)}{\mathcal{BR}(h_{\text{SM}} \to \tau\tau)}\,. \tag{4.44}$$

Dans la limite découplée, $\sin^2(\beta-\alpha) \simeq 1$, seuls les rapports de branchement ont donc un effet significatif.

L'hypothèse $\mathfrak{b}$, c'est-à-dire sans signal du MSSM, correspond donc au cas où seul le boson de Higgs du SM h_{SM} est présent. Or, le signal correspondant n'est pas compris dans les bruits de fonds du MSSM, car dans ce dernier le boson h est décrit par h_{MSSM}. Ainsi, le paramètre $\nu_i(\mu,\vec{\theta})$ donnant le nombre d'événements attendus,

$$\nu_i(\mu,\vec{\theta}) = \mu\, s_i(\vec{\theta}) + b_i(\vec{\theta})\,, \tag{4.45}$$

est réécrit sous la forme

$$\nu_i(\mu,\vec{\theta}) = \mu\, s_i^{\text{MSSM}}(\vec{\theta}) + (1-\mu) s_i^{\text{SM}}(\vec{\theta}) + b_i(\vec{\theta}) \tag{4.46}$$

avec s_i^{MSSM} le signal $\Phi \in \{ h_{\text{MSSM}}, H, A \} \to \tau\tau$ attendu selon le MSSM dans le segment i de l'histogramme de la variable discriminante et s_i^{SM} cette même quantité pour le signal $h_{\text{SM}} \to \tau\tau$ du SM. Le modificateur d'intensité du signal μ joue ainsi le rôle de distinction entre MSSM et SM. En effet, l'existence simultanée de ces deux modèles n'est pas physique, l'hypothèse du MSSM ($\mu = 1$) doit donc être testée par rapport à celle du SM ($\mu = 0$).

Cependant, le profil du rapport de vraisemblance défini section 7.2 ne permet pas de tester $\mu = 1$ par rapport à $\mu = 0$, contrairement à celui utilisé au Tevatron,

$$q_\mu = -2\ln\left(\frac{\mathfrak{L}(\text{données}|\mu,\hat{\hat{\theta}}_\mu)}{\mathfrak{L}(\text{données}|0,\hat{\hat{\theta}}_0)}\right), \quad 0 \leqslant \mu\,. \tag{4.47}$$

C'est donc celui-ci qui est utilisé pour les limites dépendantes d'un modèle.

Dans le MSSM, en plus de bosons de Higgs neutres supplémentaires, les propriétés du boson de Higgs correspondant à celui découvert en 2012 sont modifiées. L'utilisation conjointe des catégories `SM` et `BSM`, introduite dans la section 5.3, peut donc permettre d'obtenir des limites plus contraignantes sur le MSSM. La présence attendue d'une contribution des différents modes de production considérés pour h, H et A au signal dans ces différentes catégories est donnée dans le tableau 4.5. L'utilisation des catégories `SM` permet d'avoir une sensibilité spécifique à h avec les modes ggh, VBF et VH.

Catégorie		ggh, VBF, VH	$b\bar{b}h$	H, A
`no-btag`, $m_{\text{SVFIT}} < 250\,\text{GeV}$	(`SM`)	✓	✗	✗
`no-btag`, $m_{\text{SVFIT}} \geqslant 250\,\text{GeV}$	(`BSM`)	✗	✗	✓
`btag`	(`BSM`)	✓	✓	✓

Tableau 4.5 – *Présences attendues des contributions au signal dans les catégories.*

7.4.2 Scénario $M_{H_1}^{125}(\text{CPV})$

Le boson de Higgs du SM h_{SM} est prédit comme étant purement CP-pair, ce qui est également le cas du boson de Higgs h_{MSSM} du MSSM sans violation de CP. Dans ce cas, H est également CP-pair et A est CP-impair, tous ces bosons sont donc des états propres de CP.

Cependant, comme exposé au chapitre 1, une violation de CP (CPV, *CP Violation*) peut apparaître avec les bosons de Higgs. Le scénario $M_{H_1}^{125}(\text{CPV})$ correspond à ce cas de figure. Les paramètres fixes sont [79] :

- les paramètres de masse des squarks de troisième génération M_{Q_3}, M_{U_3} et M_{D_3} pris à 2 TeV ;
- les paramètres de masse des sleptons de troisième génération M_{L_3} et M_{E_3} pris à 2 TeV ;
- le paramètre de masse des Higgsinos $\mu = 1{,}65\,\text{TeV}$;
- les paramètres de masse des *gauginos* $M_1 = 1\,\text{TeV}$, $M_2 = 1\,\text{TeV}$ et $M_3 = 2{,}5\,\text{TeV}$;
- les constantes de couplage trilinéaire entre les Higgs et le stop, le sbottom et le stau, respectivement A_t, A_b et A_τ, prises telles que

$$|A_t| = \mu \cot\beta + 2{,}8\,\text{TeV}\,, \quad \phi_{A_t} = \frac{2\pi}{15}\,, \quad A_b = A_\tau = |A_t|\,. \tag{4.48}$$

La phase ϕ_{A_t} non nulle mène à des états propres de masse pour les bosons de Higgs neutres différents des états propres de CP. Les états propres de masse de ces bosons sont, par masses croissantes, H_1, H_2 et H_3. Il s'agit donc d'états CP mixtes, c'est-à-dire avec une composante CP-paire et une CP-impaire. Dans le cadre de la recherche de bosons supplémentaires de haute masse, le rôle du boson de Higgs déjà observé h est joué par H_1.

La recherche d'une composante CP-impaire peut être réalisée directement sur h. Les événements avec une paire de leptons τ peuvent être étudiés à cette fin [243] à l'aide des méthodes du paramètre d'impact [244], du plan de désintégration [245] ou du vecteur polarimétrique [246], basées sur les propriétés cinématiques des particules de l'état final.

Dans le cadre de l'analyse menée dans cette thèse, le scénario $M_{H_1}^{125}(\text{CPV})$ est testé de manière similaire à M_h^{125}. Cependant, le signal du MSSM considéré est $\Phi \in \{H_1, H_2, H_3\} \to \tau\tau$. De plus, comme A n'est pas un état propre de masse, le paramètre m_A est remplacé par la masse des bosons de Higgs chargés $m_{H^\pm}$. Enfin, des interférences entre H_2 et H_3 sont attendues dans une partie de l'espace des phases et peuvent mener à des limites d'exclusion plus faibles. L'implémentation du traitement de ce scénario à partir des données fournies en fonction de $m_{H^\pm}$ et $\tan\beta$ [234] est une de mes contributions à cette analyse.

8 Résultats et interprétations

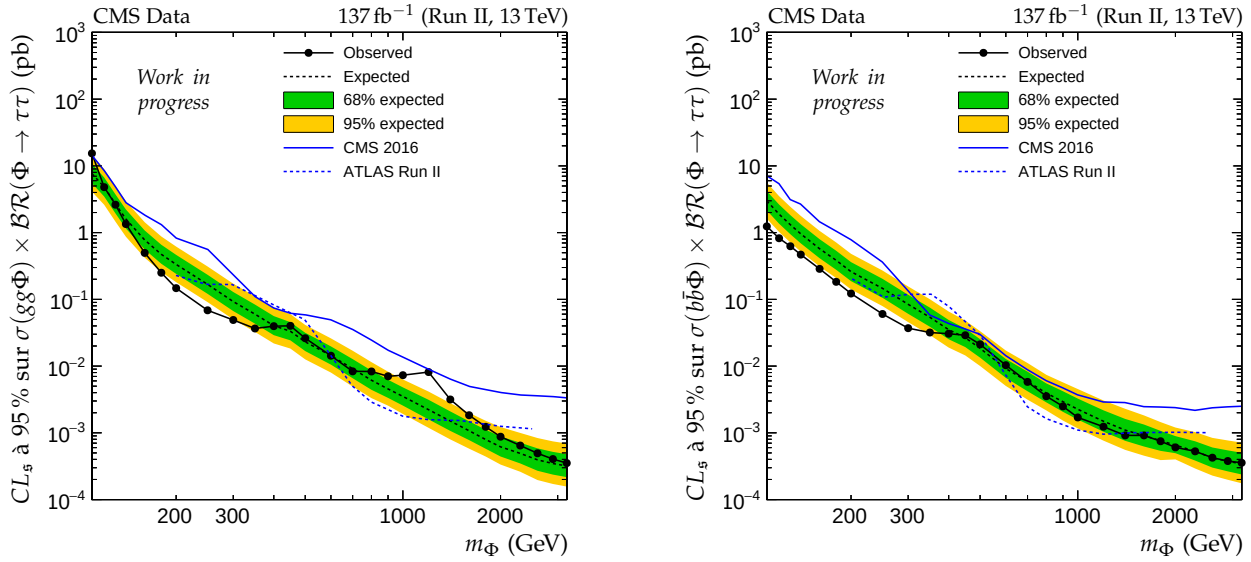
Les résultats présentés dans cette thèse, obtenus dans le cadre de l'analyse collaborative MSSM $H/A \to \tau\tau$, sont préliminaires et peuvent donc encore évoluer avant leur publication officielle. Ils comportent :

- des distributions de variables de contrôle, données en annexe G ;
- les distributions des variables discriminantes dans les différentes catégories utilisées, données en annexe H ;
- des limites d'exclusion obtenues indépendamment d'un modèle ;
- les valeurs ajustées des paramètres de nuisance associées à ces limites, données en annexe I ;
- des contours d'exclusions pour des scénarios du MSSM.

8.1 Limites indépendantes du modèle

Les limites d'exclusion indépendantes du modèle sont données en figure 4.21 pour les processus $gg \to \Phi \to \tau\tau$ (figure 4.21a) et $gg \to b\bar{b}\Phi \to \tau\tau$ (figure 4.21b). Les précédents résultats obtenus par la collaboration CMS avec les données récoltées en 2016 [81] ainsi que les derniers résultats de la collaboration ATLAS sur l'intégralité du Run II [208] y sont également affichés.

Les limites attendues (*expected*) correspondent au cas dans lequel les données observées correspondent exactement à la description des bruits de fond. Ces limites permettent d'interpréter celles



(a) *Processus* $gg \to \Phi \to \tau\tau$.

(b) *Processus* $gg \to b\bar{b}\Phi \to \tau\tau$.

Figure 4.21 – *Limites d'exclusion indépendantes du modèle physique obtenues avec l'intégralité des données du Run II récoltées par CMS. Les précédents résultats [81] obtenus avec les données de l'année 2016 sont également donnés (CMS 2016) ainsi que ceux de la collaboration ATLAS sur l'intégralité du Run II [208] (ATLAS Run II).*

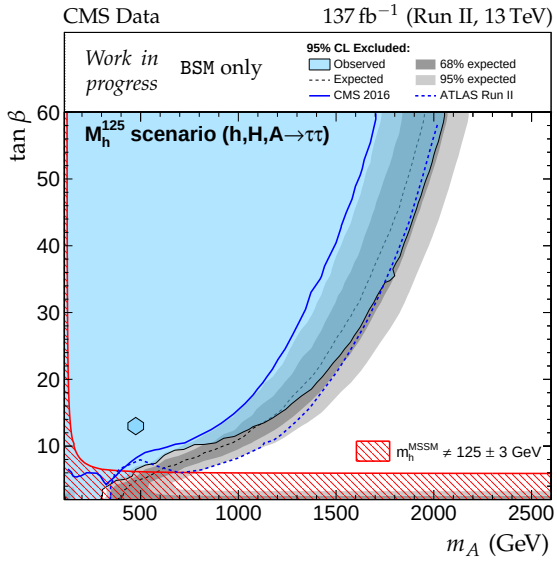
effectivement observées. Lorsqu'elles sont égales, comme c'est le cas pour $m_\Phi \gtrsim 400\,\text{GeV}$ en figure 4.21b, les observations sont en accord avec les prédictions du modèle. Lorsque les limites observées sont plus basses que les limites attendues, cela signifie que les données observées permettent de mieux exclure la présence d'un signal que la modélisation du bruit de fond elle-même. Ce cas de figure peut survenir lorsque cette modélisation est imparfaite. Par exemple, la méthode des données encapsulées surestimait initialement la quantité d'événements pour $m_\text{T}^\text{tot} \gtrsim 300\,\text{GeV}$ [211] et donnait des limites d'exclusion observées plus basses que celles attendues, ce qui a été corrigé depuis. Sur les figures 4.21a et 4.21b, les limites observées sont plus basses d'environ 2σ que celles attendues à basse masse ($m_\Phi \lesssim 300\,\text{GeV}$). La déviation de 2σ n'est toutefois pas statistiquement significative.

En revanche, lorsque les limites observées sont plus hautes que les limites attendues, cela signifie que les observations ne permettent pas d'exclure la présence d'un signal aussi bien que les prédictions. Ce peut être dû à un manque de statistiques dans la région correspondante ou à la présence effective d'un signal. Sur la figure 4.21a, la limite d'exclusion observée à $m_\Phi = 1200\,\text{GeV}$ est de $8\times10^{-3}\,\text{pb}$ contre $2\times10^{-3}\,\text{pb}$ attendus. Cet écart est d'environ 4σ. Cependant, les derniers résultats de la collaboration ATLAS [208] excluent cette valeur. La présence d'un signal pour expliquer cet excès est donc peu probable.

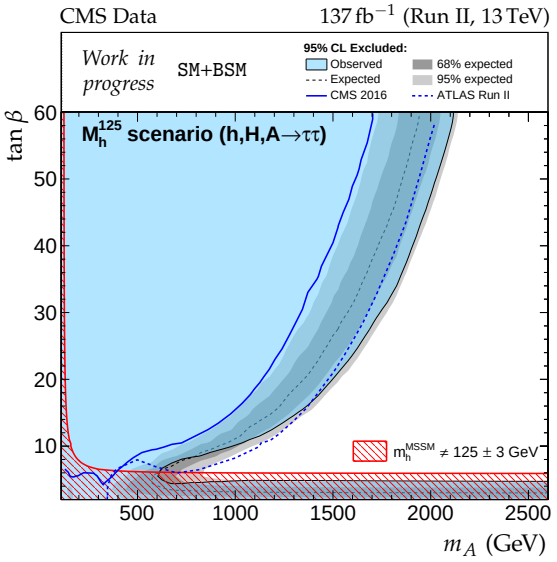
8.2 Limites du scénario M_h^{125}

Les limites d'exclusion dans le plan $(m_A, \tan\beta)$ du scénario M_h^{125} sont données en figure 4.22. Sur la figure 4.22a, seules les catégories `BSM` sont exploitées sans combinaison avec les catégories `SM`, à l'instar de ce qui a été fait lors de l'analyse des données enregistrées en 2016 [81]. Dans le cas de la figure 4.22b, ces limites sont obtenues tel que décrit dans la section 7.4.1, c'est-à-dire avec combinaison des catégories `BSM` et `SM`. La zone hachurée en rouge ($m_{h_\text{MSSM}} \neq 125 \pm 3\,\text{GeV}$) est forcément exclue, car le boson h_MSSM n'y possède pas une masse compatible avec le boson découvert en 2012. La zone bleue correspond à la région de l'espace des phases pour laquelle l'hypothèse $\mathfrak{sb}$ correspondant au MSSM est rejetée en faveur de l'hypothèse $\mathfrak{b}$ correspondant au SM d'après les données observées.

La comparaison des figures 4.22a et 4.22b montre l'effet de la combinaison des catégories `BSM` et `SM`. La conséquence majeure de cette combinaison est l'extension de la zone d'exclusion observée aux basses valeurs de $\tan\beta$. Dans cette région de l'espace des phases, la masse de h_MSSM est en effet incompatible avec les données expérimentales. Bien que cette région soit exclue par la zone hachurée, cela montre que la prise en compte des propriétés de h_MSSM et leur comparaison à celles de h



(a) *Limites d'exclusion obtenues avec la catégorisation classique* `BSM` *décrite section 5.1.*



(b) *Limites d'exclusion obtenues avec la catégorisation combinée* `SM` $+$ `BSM` *décrite section 5.3.*

Figure 4.22 – *Limites d'exclusion du scénario M_h^{125} obtenues avec l'intégralité des données du Run II récoltées par CMS. Les précédents résultats [81] obtenus avec les données de l'année 2016 sont également donnés (CMS 2016) ainsi que ceux de la collaboration ATLAS sur l'intégralité du Run II [208] (ATLAS Run II).*

observé permet d'obtenir de meilleurs résultats. De plus, les limites observées à $\tan\beta \simeq 10$ sont un peu plus étendues, bien que toujours compatibles avec les limites attendues à 2σ peu sensibles à la combinaison des catégories dans cette région de l'espace de phases. Sur la figure 4.22b, les valeurs de m_A inférieures à 600 GeV sont exclues. Cette limite passe à 1 TeV pour $\tan\beta \gtrsim 10$, et 2 TeV pour $\tan\beta \gtrsim 50$.

8.3 Limites du scénario $M_{H_1}^{125}$(CPV)

Les limites d'exclusion du scénario $M_{H_1}^{125}$(CPV) dans le plan $(m_{H^\pm}, \tan\beta)$ sont données en figure 4.23. Dans la région $\tan\beta \simeq 9$ et $m_{H^\pm} \simeq 700$ GeV, elles sont affectées par les interférences entre H_2 et H_3. Ces interférences destructives donnent un signal atténué, réduisant la sensibilité de l'analyse. Les limites observées sont en bon accord avec les limites attendues. Les valeurs inférieures à 400 GeV pour la masse $m_{H^\pm}$ sont exclues. Pour $\tan\beta > 10$, hormis la zone d'interférences à $m_{H^\pm} \simeq 650$ GeV, $m_{H^\pm} < 1$ TeV est rejeté. Lorsque $\tan\beta \simeq 20$, l'exclusion s'étend jusqu'à $m_{H^\pm} \simeq 1{,}4$ TeV.

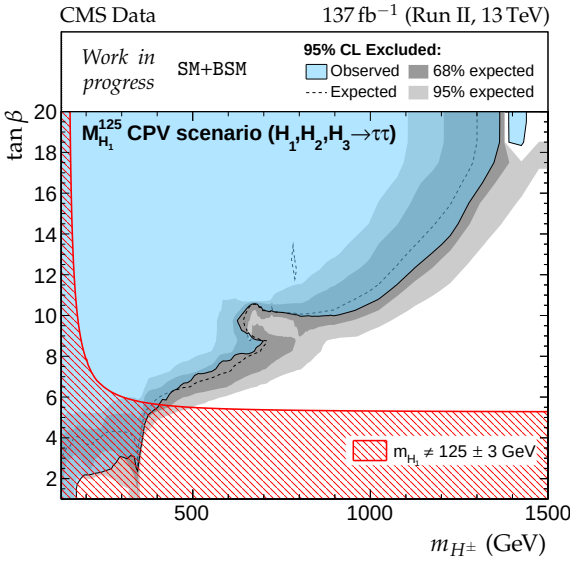


Figure 4.23 – *Limites d'exclusion du scénario $M_{H_1}^{125}$ (CPV) dans le plan $(m_{H^\pm}, \tan\beta)$ obtenues avec l'intégralité des données du Run II récoltées par CMS.*

9 Conclusion

La recherche de bosons de Higgs supplémentaires de haute masse se désintégrant en paire de taus a été présentée. Ces particules correspondent aux deux bosons neutres H et A prédits par le MSSM. Les données analysées sont celles récoltées par CMS lors des collisions de protons avec une énergie dans le centre de masse de $\sqrt{s} = 13$ TeV du Run II du LHC, correspondant à une luminosité intégrée de 137 fb^{-1}. Parmi les six canaux ou états finaux possibles pour la paire de leptons τ, quatre ont été exploités ($\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$, $e\mu$).

Afin d'interpréter les données observées, une modélisation des processus physiques attendus, tels que les bruits de fond, est nécessaire. La méthode des données encapsulées permet de limiter l'utilisation de simulations afin de modéliser certains processus contenant une véritable paire de leptons τ, en particulier $Z \to \tau\tau$. L'estimation des faux τ_{h}, jets identifiés à tort comme étant des τ_{h}, est également basée presque exclusivement sur les données réelles. Ces méthodes permettent de réduire les incertitudes inhérentes à la simulation.

Le signal correspondant aux bosons supplémentaires est modélisé par les processus de fusion de gluons ($gg \to \Phi \to \tau\tau$) et en association avec des quarks b ($gg \to b\bar{b}\Phi \to \tau\tau$). Afin de maximiser la sensibilité de l'analyse à ces deux modes de production, une catégorisation des événements basée sur le nombre de jets issus de quarks b est utilisée. Une catégorisation supplémentaire, basée sur la masse transverse du lepton ℓ dans les canaux $\ell\tau_{\mathrm{h}}$ ($\mu\tau_{\mathrm{h}}$, $e\tau_{\mathrm{h}}$) et sur D_ζ dans le canal $e\mu$, permet d'augmenter encore cette sensibilité. Des limites d'exclusion sur le produit de la section efficace de production des bosons supplémentaires avec leur rapport de branchement à la désintégration en paire de leptons τ ont été données pour les deux modes de production étudiés pour des masses comprises entre 110 GeV et 3,2 TeV.

Le scénario M_h^{125} fixe les valeurs de certains paramètres du MSSM. Des limites d'exclusion de ce modèle en faveur du SM ont été données dans le plan $(m_A, \tan\beta)$. Les valeurs de m_A inférieures à 600 GeV sont exclues et cette valeur augmente à 2 TeV lorsque $\tan\beta \gtrsim 50$. Un second scénario, $M_{H_1}^{125}(\mathrm{CPV})$, a également été étudié. Dans celui-ci, le choix des valeurs des paramètres du MSSM autorise la violation de CP par les bosons de Higgs car les états propres de CP ne sont plus les mêmes que ceux de masse. Ces derniers, pour les bosons neutres, sont H_1, H_2 et H_3. En particulier, A n'est plus un état propre de masse et c'est donc dans le plan $(m_{H^\pm}, \tan\beta)$ que sont données les limites d'exclusion en faveur du SM. Pour $\tan\beta \simeq 9$ et $m_{H^\pm} \simeq 700$ GeV, les interférences entre H_2 et H_3 réduisent fortement la sensibilité de l'analyse, limitant alors l'étendue de la région d'exclusion du MSSM. L'analyse menée permet toutefois d'exclure $m_{H^\pm} < 400$ GeV et cette valeur augmente à 1,4 TeV lorsque $\tan\beta \simeq 20$.

Cependant, les résultats présentés dans cette thèse sont donnés à titre d'illustration de l'état actuel de l'analyse. L'ajustement des paramètres de nuisance ainsi que les résultats obtenus nécessitent de plus amples investigations.

À ce jour, aucun des modèles BSM proposés ne permet de mieux décrire les résultats expérimentaux que le SM. Augmenter l'échelle d'énergie des collisions ou la luminosité intégrée exploitable permettra d'aller plus loin dans l'analyse des processus physiques. Les modèles proposés par les théoriciens sont alors mieux contraints et sont soit rejetés, soit affinés. L'énergie dans le centre de masse des collisions de protons au LHC doit passer à 14 TeV lors du Run III, dont le début est prévu en 2022, avec une luminosité intégrée de l'ordre de 300 fb^{-1}. Puis, de 2027 à 2035, le LHC à haute luminosité (HL-LHC) générera près de 3000 fb^{-1} de données. Cette augmentation considérable de la statistique permettra d'améliorer les mesures de précision des paramètres du SM ainsi que les analyses telles que celle présentée dans ce chapitre. Cependant, le tracé du LHC ne permet pas, avec la technologie actuelle, d'espérer augmenter encore l'énergie de collision. Le projet FCC (Future Collisionneur Circulaire) [247] pourrait proposer une énergie de collision de 100 TeV à l'horizon 2040 avec des protons.

Au-delà de la limitation imposée par les données obtenues à ce jour au LHC, l'analyse effectuée dans ce chapitre est compliquée par la présence de neutrinos dans l'état final des événements exploités. Elle rend difficile l'estimation de la masse invariante d'une résonance se désintégrant en deux τ. Cet effet, inhérent à cette analyse, limite sa sensibilité. Le *machine learning* peut permettre une estimation plus précise de la masse d'une résonance se désintégrant en paire de τ. Ce sujet est présenté dans le chapitre 5.

Chapitre 5
Reconstruction de la masse d'une résonance grâce au *machine learning*

Sommaire

1 Introduction . . . 158
2 Événements utilisés ou échantillons . . . 159
2.1 Génération avec FastSim . . . 159
2.2 Sélection des événements . . . 160
2.3 Événements obtenus et pondération . . . 161
2.4 Cible et variables d'entrée des modèles . . . 162
3 Concepts communs aux modèles . . . 163
3.1 Fonction de coût . . . 163
3.2 Entraînement, descente de gradient et mini-lots . . . 163
3.3 Sous-entraînement et surentraînement . . . 164
4 Arbres de décision améliorés . . . 165
4.1 Arbres de décision . . . 165
4.2 *Gradient Boosting* et descente de gradient . . . 166
5 Réseaux de neurones profonds . . . 167
5.1 Neurones . . . 167
5.2 Réseaux de neurones . . . 169
5.3 Entraînement . . . 169
6 Optimisation des hyper-paramètres et choix d'un modèle . . . 172
6.1 Variables d'entrée . . . 173
6.2 Type de modèle . . . 175
6.3 Fonction de coût . . . 176
6.4 Algorithme d'optimisation . . . 177
6.5 Autres hyper-paramètres . . . 178
7 Discussions . . . 184
7.1 Effet de l'empilement . . . 184
7.2 Effet de la reconstruction des particules . . . 186
7.3 Effet des faux taus hadroniques . . . 187
7.4 Effet de la séparation des canaux . . . 188
7.5 Effet de la définition de $E_{\mathrm{T}}^{\mathrm{miss}}$ . . . 194
7.6 Effet de l'intervalle de masse . . . 196
7.7 Modèle final . . . 201
8 Utilisation du modèle dans les analyses CMS . . . 201
8.1 Utilisation de m_{ML} comme variable discriminante . . . 201
8.2 Comparaison de m_{ML} à m_{SVFIT} . . . 205
9 Conclusion . . . 210

1 Introduction

L'utilisation de l'intelligence artificielle (IA) s'est grandement développée au cours des dernières années. L'IA regroupe l'ensemble des théories et des techniques développant des programmes informatiques complexes capables de simuler certains traits de l'intelligence humaine tels que le raisonnement et l'apprentissage. L'entreprise Google DeepMind a par exemple développé AlphaGo [248], un programme destiné à jouer au jeu de Go, qui a battu en 2016 le champion du monde de la discipline 4 à 1.

Le *machine learning* (ML) est une branche de l'IA dans laquelle un modèle (algorithme ou programme) s'améliore dans la réalisation d'une tâche par accumulation d'expérience sur des jeux de données d'entraînement, sans pour autant être programmé explicitement pour effectuer cette tâche. Pour y parvenir, les jeux de données d'entraînement comprennent les informations $\{ \vec{x}_i \}$ à donner au modèle ainsi que les « bonnes réponses » $\{ y_{\text{vraie},i} \}$ qu'il doit fournir en sortie. L'objectif est donc d'obtenir un modèle agissant comme une fonction F approximant celle reliant les entrées $\{ \vec{x}_i \}$ aux cibles $\{ y_{\text{vraie},i} \}$. Il peut alors donner une prédiction $y_{\text{préd}}$ sur une nouvelle entrée $\vec{x}$ selon $y_{\text{préd}} = F(\vec{x})$. La tâche du modèle peut être :

une classification lorsque y est discrète, par exemple lorsqu'il s'agit de déterminer si une image représente un chat ou un chien [249] ;

une régression lorsque y est continue, par exemple estimer le prix d'un bien immobilier [250].

Les applications du ML à la physique des particules sont variées et font l'objet de nombreux sujets d'étude [211, 251-253]. Dans les chapitres précédents, le ML est déjà utilisé pour diverses tâches :

— identification des jets issus de quarks b (*b-tagging*) avec DEEPCSV [166, 167, 221] ;
— identification des taus hadroniques avec DEEPTAU [176] ;
— catégorisation des événements comme exposé dans le chapitre 4 [227, 228].

Dans les événements $H \to \tau\tau$ présentés au chapitre 1, et plus généralement lors de tout processus physique $X \to \tau\tau$ où une particule X donne une paire de leptons τ, des neutrinos sont émis lors des désintégrations des τ. Or, ils sont invisibles dans les détecteurs tels que CMS ou ATLAS. Il est donc impossible de déterminer la masse invariante totale du système $\tau\tau$ issu de X. Plusieurs méthodes ont été développées afin de reconstruire la masse du système $\tau\tau$ [254-256]. Dans le cadre des analyses $H \to \tau\tau$, la collaboration CMS utilise SVFIT [229], un estimateur de la masse d'une particule se désintégrant en paire de leptons τ par ajustement d'un profil de vraisemblance.

La reconstruction de la masse d'une particule X, ou résonance, se désintégrant en paire de leptons τ grâce au *machine learning* a été étudiée par BÄRTSCHI & coll. [257] dans le cas où X est un boson de Higgs avec une masse comprise entre 80 et 300 GeV. Ils ont obtenu une résolution de 8,4 % sur la masse du Higgs, contre 17 % avec SVFIT. De plus, le temps de calcul nécessaire à l'obtention de la masse est moindre avec le ML. L'utilisation du ML est donc très prometteuse. Cependant, ces travaux utilisent des événements générés avec une simulation grossière du détecteur CMS basée sur DELPHES [258, 259] et sans empilement, notion introduite dans le chapitre 2.

Les travaux présentés dans ce chapitre vont plus loin. La génération des événements, introduite dans la section 2, utilise FASTSIM [260-263] pour modéliser le détecteur CMS. Bien qu'il ne s'agisse pas de la simulation complète basée sur GEANT4 [184-186], FASTSIM est bien plus proche de la réalité que DELPHES. De plus, l'empilement est pris en compte. Les modèles obtenus sont ainsi directement utilisables dans de réelles analyses, telle que celle présentée dans le chapitre 4.

La section 3 introduit des concepts communs aux deux types de modèles étudiés. Les arbres de décision améliorés et les réseaux de neurones profonds sont ensuite introduits dans les sections 4 et 5 respectivement. La comparaison des modèles obtenus et les raisons du choix de l'un d'entre eux sont présentées section 6. Dans la section 7, divers effets sur les performances du modèle sélectionné sont discutés, en particulier la prise en compte de l'empilement. Enfin, l'utilisation en conditions réelles du modèle issu de ces travaux dans des analyses de physique est présentée dans la section 8.

Une note d'analyse [264] est en cours de rédaction. Les scripts utilisés pour la génération des événements peuvent être consultés [265], tout comme ceux permettant d'entraîner les modèles étudiés [266]. Le modèle issu de ces travaux est disponible [267] pour pouvoir être utilisé dans d'autres analyses.

2 Événements utilisés ou échantillons

L'objectif de cette étude est de reconstruire la masse des particules se désintégrant en paire de leptons τ. Il s'agit donc d'une tâche de régression. Dans l'optique d'une utilisation dans les analyses telles que celle présentée dans le chapitre 4, il a été choisi d'utiliser des événements $\mathcal{H} \to \tau\tau$ où $\mathcal{H}$ est le boson de Higgs du modèle standard h dont la masse est modifiée, à l'instar de ce qu'ont fait BÄRTSCHI & coll. [257]. La cible du modèle est donc la masse $m_{\mathcal{H}}$. Un événement est, dans la terminologie du ML, un « échantillon ».

2.1 Génération avec FASTSIM

Nous avons généré nos propres données simulées [265] afin d'obtenir une distribution continue des valeurs de $m_{\mathcal{H}}$ et suffisamment d'échantillons pour chaque point de masse. Dans le contexte de la collaboration CMS, nous avons utilisé FASTSIM [260-263]. Cet outil permet de procéder à l'ensemble de la simulation des événements introduite chapitre 2, de la génération du processus initial à la reconstruction des objets physiques par le détecteur.

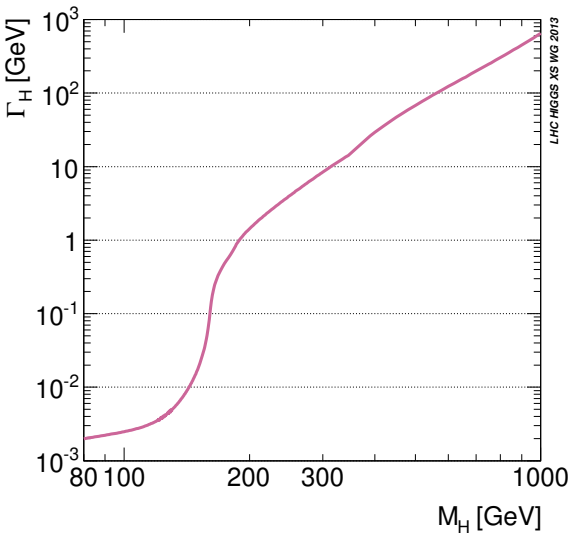
Les données simulées correspondent à des collisions de protons avec une énergie dans le centre de masse de 13 TeV. Les processus physiques sont générés par PYTHIA 8 [179] avec les réglages CUEP8M1 [193, 197]. La production du boson $\mathcal{H}$ se fait par fusion de gluons, il s'agit du mode de production dominant dans le cas du boson de Higgs du modèle standard h. De plus, le rapport de branchement $\mathcal{BR}(\mathcal{H} \to \tau\tau)$ est fixé à 1, c'est-à-dire que $\mathcal{H}$ se désintègre forcément en paires de leptons τ. Tous les événements obtenus sont donc bien du type $\mathcal{H} \to \tau\tau$.

La masse de $\mathcal{H}$ varie de 50 à 800 GeV par pas de 1 GeV. Il est important d'utiliser l'intervalle le plus étendu possible, il correspond à la gamme utile des modèles obtenus. L'effet de l'étendue de cet intervalle est discuté dans la section 7.

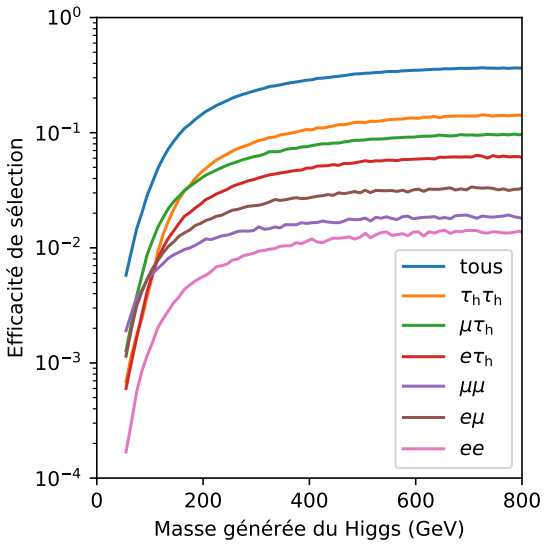
Il n'est pas possible, par la méthode que nous utilisons, de générer des événements avec $m_{\mathcal{H}} \gtrsim 1$ TeV. Cela est dû à la largeur Γ_h de h, représentée figure 5.1a en fonction de m_h. La largeur d'une particule est liée à sa durée de vie τ selon

$$\Gamma = \frac{\hbar}{\tau}. \tag{5.1}$$

Ainsi, plus une particule se désintègre rapidement, plus sa largeur est importante. Le principe de Heisenberg mène alors à une incertitude sur la masse, due à la durée de vie τ, égale à Γ. Or, vers



(a) *Largeur du boson de Higgs du modèle standard [85].*



(b) *Efficacité de sélection des événements pour $m_{\mathcal{H}} \in [\![50, 800]\!]$ GeV dans les différents canaux et pour tous les canaux.*

Figure 5.1 – *Origine des limites haute (gauche) et basse (droite) de l'intervalle de masse utilisé.*

1 TeV, $\Gamma_{\mathcal{H}} \simeq m_{\mathcal{H}}$. C'est pourquoi la génération de tels événements est compromise avec $\mathcal{H}$ défini comme h avec une masse modifiée. Nous générons donc uniquement des événements en-deçà de 800 GeV.

La sélection des événements est présentée dans la section 2.2. Son efficacité est représentée sur la figure 5.1b. Plus de 99 % des événements sont rejetés lorsque $m_{\mathcal{H}} < 50$ GeV. Nous ne considérerons donc pas de masse plus basse.

S'il est possible d'appliquer des poids aux échantillons afin d'équilibrer l'entraînement sur l'ensemble des valeurs de la cible, plus d'événements sont générés à basse masse afin d'obtenir des topologies d'événements variées malgré la faible efficacité de sélection. Ainsi, la quantité d'événements générés pour chaque valeur de $m_{\mathcal{H}}$ est de :

— 60 000 pour 50 GeV $\leqslant m_{\mathcal{H}} <$ 300 GeV ;
— 20 000 pour 300 GeV $\leqslant m_{\mathcal{H}} <$ 500 GeV ;
— 10 000 pour 500 GeV $\leqslant m_{\mathcal{H}} \leqslant$ 800 GeV.

L'empilement est modélisé par superposition du signal $\mathcal{H} \to \tau\tau$ à des événements dits de « biais minimum » [179]. Il s'agit d'événements pouvant contenir des interactions dures, mais n'activant pas de chemin de déclenchement. La quantité d'empilement ajoutée à l'événement $\mathcal{H} \to \tau\tau$ suit le profil d'empilement de l'année 2017. Les conditions des collisions simulées sont ainsi identiques à celles de l'année 2017 au LHC.

2.2 Sélection des événements

2.2.1 Canaux $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ et $e\mu$

Dans le cas des canaux $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ et $e\mu$, la sélection des événements se fait comme exposé dans le chapitre 4 pour l'année 2017. Afin d'obtenir un modèle dont les prédictions auront non seulement un sens dans la région de signal, mais aussi dans les régions de contrôle et de détermination, les coupures sur

— $m_T^{(\mu)}$ dans le canal $\mu\tau_h$;
— $m_T^{(e)}$ dans le canal $e\tau_h$;
— D_ζ dans le canal $e\mu$

ne sont pas appliquées. La construction du dilepton est inchangée. La correspondance des objets du dilepton avec ceux ayant activé le chemin de déclenchement n'est pas vérifiée.

2.2.2 Canal $\mu\mu$

Sélection des muons Tout muon respectant les critères listés ci-après est retenu pour jouer le rôle de L_1 ou L_2 dans le dilepton :

— $p_T{}^\mu > 10$ GeV ;
— $|\eta^\mu| < 2.4$;
— paramètres d'impact $d_z < 0{,}2$ cm et $d_{xy} < 0{,}045$ cm ;
— $I^\mu < 0{,}15\, p_T{}^\mu$;
— passer le point de fonctionnement *medium* du *muonID*.

Sélection du dilepton L'événement est retenu à condition qu'au moins une paire $L_1L_2 = \mu\mu$ puisse être construite avec L_1 et L_2 de charges électriques opposées. Il est de plus requis que L_1 et L_2 soient séparés dans le plan (η, ϕ) tel que $\Delta R > 0{,}3$. Si plus d'une paire possible existe dans l'événement, une seule est retenue selon la logique exposée dans le chapitre 4.

Vetos de leptons supplémentaires Les vetos de leptons supplémentaires doivent être respectés, c'est-à-dire que l'événement ne contient pas :

— de muon supplémentaire tel que $p_T{}^\mu > 10$ GeV, $|\eta^\mu| < 2{,}4$, passant le point de fonctionnement *medium* du *muonID* et d'isolation $I^\mu < 0{,}3\, p_T{}^\mu$;

— d'électron tel que $p_T{}^e > 10\,\text{GeV}$, $|\eta^e| < 2{,}5$, passant le point de fonctionnement à 90 % d'efficacité de l'*electron ID MVA* et d'isolation $I^e < 0{,}3\,p_T{}^e$, l'électron devant passer le veto d'électron de conversion et présenter moins de deux points de passage manquants dans le trajectographe.

2.2.3 Canal *ee*

Sélection des électrons Tout électron respectant les critères listés ci-après est retenu pour jouer le rôle de L_1 ou L_2 dans le dilepton :
— $p_T{}^e > 20\,\text{GeV}$;
— $|\eta^e| < 2.4$;
— paramètres d'impact $d_z < 0{,}2\,\text{cm}$ et $d_{xy} < 0{,}045\,\text{cm}$;
— $I^e < 0{,}1\,p_T{}^e$;
— passer le point de fonctionnement à 90 % d'efficacité de l'*electron ID MVA*.

Sélection du dilepton L'événement est retenu à condition qu'au moins une paire $L_1L_2 = ee$ puisse être construite avec L_1 et L_2 de charges électriques opposées. Il est de plus requis que L_1 et L_2 soient séparés dans le plan (η, ϕ) tel que $\Delta R > 0{,}5$. Si plus d'une paire possible existe dans l'événement, une seule est retenue selon la logique exposée dans le chapitre 4.

Vetos de leptons supplémentaires Les vetos de leptons supplémentaires doivent être respectés, c'est-à-dire que l'événement ne contient pas :
— de muon tel que $p_T{}^\mu > 10\,\text{GeV}$, $|\eta^\mu| < 2{,}4$, passant le point de fonctionnement *medium* du *muonID* et d'isolation $I^\mu < 0{,}3\,p_T{}^\mu$;
— d'électron supplémentaire tel que $p_T{}^e > 10\,\text{GeV}$, $|\eta^e| < 2{,}5$, passant le point de fonctionnement à 90 % d'efficacité de l'*electron ID MVA* et d'isolation $I^e < 0{,}3\,p_T{}^e$, l'électron devant passer le veto d'électron de conversion et présenter moins de deux points de passage manquants dans le trajectographe.

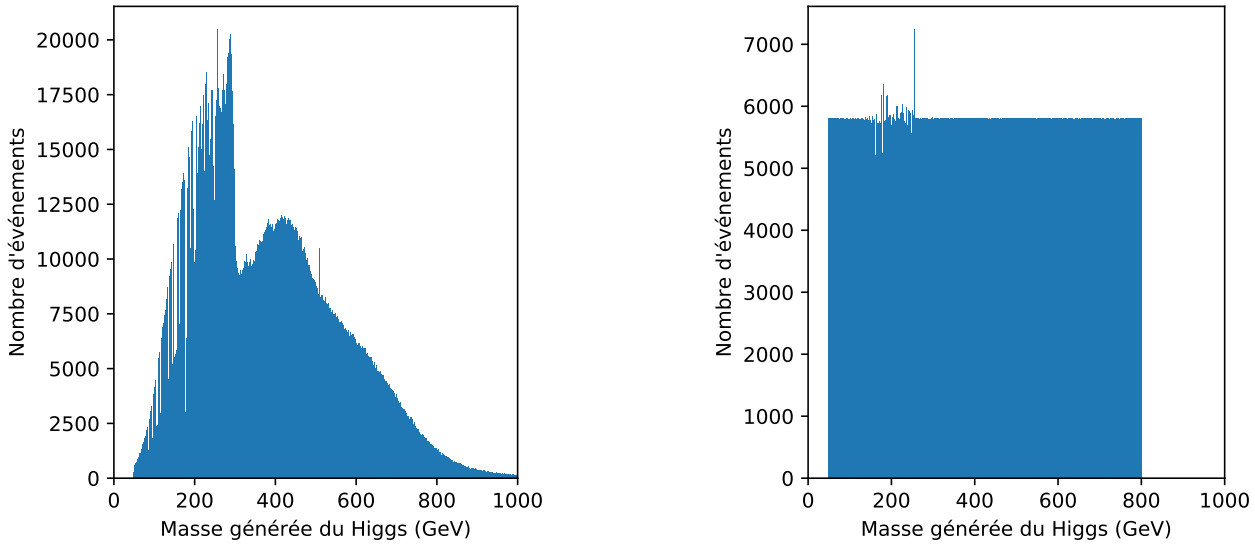
2.2.4 Jets

À l'instar de l'analyse présentée au chapitre 4, les jets sont soumis à la procédure de CHS [192] décrite dans le chapitre 3 et reconstruits par l'algorithme anti-k_T [162] avec un paramètre $R = 0{,}4$. Ces jets doivent également passer les critères d'identification présentés dans le chapitre 2. L'identification des jets issus de quarks b (*b-tagging*) est réalisée par l'algorithme DEEPCSV [167, 221]. Les jets tels que $p_T > 20\,\text{GeV}$ et $|\eta| < 2{,}5$ sont considérés comme issus d'un b si leur score est supérieur à 0,3033. Les jets non identifiés comme issus d'un b ne sont retenus que si $p_T > 30\,\text{GeV}$ et $|\eta| < 4{,}7$.

2.3 Événements obtenus et pondération

Environ 3 millions d'événements ont été sélectionnés selon les critères présentés précédemment parmi plus de 22 millions générés. La distribution de $m_\mathcal{H}$ dans les événements sélectionnés est représentée sur la figure 5.2a. Certains présentent des valeurs de $m_\mathcal{H}$ au-delà de 800 GeV, cet effet est dû à la largeur de cette particule, représentée sur la figure 5.1a en fonction de sa masse. La largeur à 800 GeV est ainsi d'environ 300 GeV. Le réglage $m_\mathcal{H} = 800\,\text{GeV}$ donne donc des événements contenant un boson dont la masse effective se situe entre 500 et 1100 GeV, d'où la queue de la distribution observée à haute masse sur la figure 5.2a. À basse masse en revanche, la largeur est inférieure à 100 MeV, cet effet n'est donc pas présent. La cible du modèle est la masse effective du boson. Les événements retenus dans la suite sont ceux où celle-ci se situe bien entre 50 et 800 GeV, d'où la disparition de la queue à haute masse sur la figure 5.2b.

Ces événements sont de plus séparés en trois groupes selon les proportions suivantes :
— 70 % pour l'entraînement. Ce sont ces échantillons que les modèles pourront exploiter afin d'apprendre à prédire correctement $m_\mathcal{H}$;
— 20 % pour la validation. Ces échantillons permettent de vérifier qu'il n'y a pas de surentraînement, c'est-à-dire que le modèle ne se spécialise pas vis-à-vis du jeu d'entraînement ;



(a) *Distribution brute sur tous les événements.*

(b) *Distribution pondérée pour les événements d'entraînement.*

Figure 5.2 – *Distributions de la masse générée de $\mathcal{H}$.*

— 10 % pour les tests. Ces échantillons ne sont pas utilisés lors des entraînements et permettent donc de tester les modèles sur des données inédites. Sauf contre-indication, les figures sont toutes obtenues avec ce groupe d'échantillons.

La répartition des échantillons dans ces trois groupes est faite de manière aléatoire.

Afin de réaliser un entraînement équitable entre les différentes valeurs de $m_{\mathcal{H}}$, un poids est associé à chaque échantillon de manière à ce que la distribution pondérée de $m_{\mathcal{H}}$ soit plate dans chacun des trois groupes précédemment définis. Cette distribution sur les échantillons utilisés pour l'entraînement des modèles est représentée sur la figure 5.2b.

2.4 Cible et variables d'entrée des modèles

La cible des modèles est la masse de la particule générée $\mathcal{H}$ se désintégrant en paire de leptons τ. Un tel événement est illustré sur la figure 5.3. Les variables d'entrée doivent être des observables accessibles expérimentalement, c'est-à-dire issues de la reconstruction des événements présentée dans le chapitre 2.

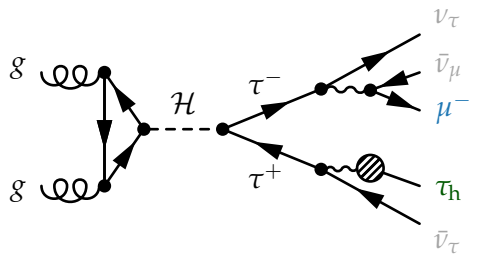


Figure 5.3 – *Diagramme de Feynman des événements d'entraînement des modèles dans le cas du canal $\mu\tau_h$.*

Les variables considérées sont :

— les impulsions de L_1 et L_2, les produits de désintégration visibles des τ, c'est-à-dire le muon et le τ_h dans l'exemple de la figure 5.3 : ${p_T}^{L_1}$, η^{L_1}, ϕ^{L_1}, ${p_T}^{L_2}$, η^{L_2}, ϕ^{L_2} ;

— l'énergie transverse manquante pour rendre compte de la présence des neutrinos : ${E_T}^{\text{miss}}$, $\phi^{{E_T}^{\text{miss}}}$, obtenus par l'algorithme PUPPI [171] ;

— la matrice M de covariance de ${E_T}^{\text{miss}}$, rendant compte de l'incertitude sur la mesure de ${E_T}^{\text{miss}}$: M_{xx}, M_{xy}, M_{yy} ;

— le nombre attendu de neutrinos lié à l'état final identifié N_ν^{reco}, déterminé à partir du canal obtenu par la sélection des événements, c'est-à-dire sans utilisation des informations générées ;
— les masses transverses $m_{\text{T}}(L_1, {E_{\text{T}}}^{\text{miss}})$, $m_{\text{T}}(L_2, {E_{\text{T}}}^{\text{miss}})$ et $m_{\text{T}}(L_1, L_2)$ définies par

$$m_{\text{T}}(A,B) = \sqrt{2\,{p_{\text{T}}}^A\,{p_{\text{T}}}^B\,(1-\cos(\phi^A-\phi^B))}\;; \tag{5.2}$$

— la masse transverse totale $m_{\text{T}}^{\text{tot}}$ définie par

$$m_{\text{T}}^{\text{tot}} = \sqrt{m_{\text{T}}^2(L_1, {E_{\text{T}}}^{\text{miss}}) + m_{\text{T}}^2(L_2, {E_{\text{T}}}^{\text{miss}}) + m_{\text{T}}^2(L_1, L_2)}\;; \tag{5.3}$$

— les impulsions des deux jets principaux (de plus hauts p_{T}) présents dans l'événement : ${p_{\text{T}}}^{\text{jet 1}}$, $\eta^{\text{jet 1}}$, $\phi^{\text{jet 1}}$, ${p_{\text{T}}}^{\text{jet 2}}$, $\eta^{\text{jet 2}}$, $\phi^{\text{jet 2}}$;
— l'Activité Hadronique Additionnelle (AHA), définie par la somme des impulsions des jets autres que les deux principaux : ${p_{\text{T}}}^{\text{AHA}}$, η^{AHA}, ϕ^{AHA} avec

$$\vec{p}^{\,\text{AHA}} = \sum_{\text{jet } i,\, i>2} \vec{p}^{\,\text{jet } i}\;; \tag{5.4}$$

— la quantité de jets utilisés pour déterminer $\vec{p}^{\,\text{AHA}}$, $N_{\text{jets}}^{\text{AHA}}$;
— le nombre de vertex principaux d'empilement, N_{PU}.

Des modèles ont été entraînés sur l'ensemble de ces 27 variables ainsi que sur des sous-ensembles de cette liste.

3 Concepts communs aux modèles

3.1 Fonction de coût

Une fonction de coût compare les prédictions d'un modèle aux valeurs vraies de la cible. Elle doit être différentiable et est définie de manière à être minimale lorsque les prédictions sont égales aux valeurs vraies, c'est-à-dire lorsque le modèle est parfait. Les fonctions de coût les plus répandues sont :

MSE *Mean Squared Error* ou erreur quadratique moyenne,

$$\mathsf{L}_{\text{MSE}}(\{\,y_{\text{vraie},i}\,\},\{\,y_{\text{préd},i}\,\}) = \frac{1}{2N}\sum_{i=1}^{N}\left(y_{\text{préd},i}-y_{\text{vraie},i}\right)^2\;; \tag{5.5}$$

MAE *Mean Absolute Error* ou erreur absolue moyenne,

$$\mathsf{L}_{\text{MAE}}(\{\,y_{\text{vraie},i}\,\},\{\,y_{\text{préd},i}\,\}) = \frac{1}{N}\sum_{i=1}^{N}\left|y_{\text{préd},i}-y_{\text{vraie},i}\right|\;; \tag{5.6}$$

MAPE *Mean Absolute Percentile Error* ou erreur absolue relative moyenne,

$$\mathsf{L}_{\text{MAPE}}(\{\,y_{\text{vraie},i}\,\},\{\,y_{\text{préd},i}\,\}) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_{\text{préd},i}-y_{\text{vraie},i}}{y_{\text{vraie},i}}\right|. \tag{5.7}$$

3.2 Entraînement, descente de gradient et mini-lots

Un modèle peut être vu comme une fonction paramétrique F dont l'application à une entrée $\vec{x}$ donne une prédiction $y_{\text{préd}} = F(\vec{x})$. L'entraînement consiste à régler les paramètres du modèle afin d'obtenir des prédictions fidèles aux valeurs vraies du jeu de données d'entraînement.

La fonction de coût L est minimale lorsque les prédictions du modèle sont parfaites. Il s'agit donc de trouver le minimum de L dans l'espace à D dimensions formé par les $D = N_{\text{params.}}$ paramètres à régler. Cela peut être fait de manière itérative par descente de gradient (GD, *Gradient Descent*) [268].

Il s'agit d'une méthode itérative qui détermine le gradient de L, $\vec{\nabla}(\mathsf{L})$, autour de la « position » du modèle dans l'espace à D dimensions. Chaque paramètre p est alors modifié selon

$$p \to p - \eta \vec{\nabla}\big(\mathsf{L}(\{ y_{\text{vraie},i} \}, \{ y_{\text{préd},i} \})\big) \cdot \vec{e}_p = p - \eta \frac{\partial \mathsf{L}(\{ y_{\text{vraie},i} \}, \{ y_{\text{préd},i} \})}{\partial p} \tag{5.8}$$

avec η le taux d'apprentissage, et $\mathsf{L}(\{ y_{\text{vraie},i} \}, \{ y_{\text{préd},i} \})$ la fonction de coût évaluée sur l'ensemble du jeu de données d'entraînement. Le taux d'apprentissage est généralement pris entre 10^{-5} et 0,5.

Toutefois, l'évaluation du gradient de la fonction de coût sur l'ensemble du jeu de données d'entraînement, contenant éventuellement plusieurs millions d'échantillons comme c'est le cas dans cette thèse, nécessite d'importantes ressources computationnelles (en l'occurrence, de la mémoire vive). Afin de pallier ce problème, la descente de gradient stochastique (SGD, *Stochastic Gradient Descent*) évalue le gradient de L individuellement pour chaque échantillon. Cependant, le SGD amène de fortes fluctuations le long de la descente, ce qui ralenti l'entraînement. De plus, ces mêmes fluctuations une fois proche du minimum mènent à une précision dégradée des prédictions.

Afin d'éviter ce phénomène, le *batch* GD (BGD) [269] évalue le gradient de la fonction de coût par une moyenne sur un « mini-lot » (*mini-batch*), sous-ensemble du jeu de données de taille fixée. Les fluctuations lors de la descente sont moindres que dans le cas de SGD. Cette moyenne introduit un bruit dû à la composition aléatoire des mini-lots qui reste non nul même une fois le minimum de L atteint. Cela permet de s'échapper des minimums locaux, mais dégrade la précision une fois au minimum global.

Une « époque » de l'entraînement correspond à une utilisation de tous les mini-lots, c'est-à-dire de tous les échantillons du jeu de données, pour modifier les paramètres du modèle. Pour ne pas biaiser l'entraînement à cause de l'ordre du jeu de données, il est mélangé aléatoirement à chaque nouvelle époque. La composition des mini-lots est donc également aléatoire. Leur taille est fixée à $2^{11} = 2048$ événements. Une taille de la forme 2^n permet d'optimiser l'utilisation des GPUs (*Graphics Processing Unit*) sur lesquels l'entraînement se fait [175]. Les points de masse générés étant les entiers entre 50 et 800 GeV, soit 750 points de masse, 2048 événements pris au hasard est un compromis entre un petit mini-lot et une bonne probabilité de couvrir une large gamme de masse au sein d'un mini-lot.

3.3 Sous-entraînement et surentraînement

Un modèle doit être suffisamment complexe, c'est-à-dire proposer suffisamment de paramètres réglables à l'entraînement, afin de pouvoir réaliser sa tâche. Dans le cas contraire, ses prédictions ne sont qu'approximatives, voire fausses. Par exemple, en utilisant une droite affine comme modèle, c'est-à-dire avec deux paramètres, il est difficile de prédire correctement une loi polynomiale de degré 2, régie par trois paramètres. Or, le nombre d'itérations d'entraînement nécessaires à l'optimisation d'un modèle augmente avec sa complexité. Avec une quantité limitée d'itérations, le modèle est sous-entraîné. Il est donc nécessaire d'en avoir suffisamment.

Au fur et à mesure des itérations, la valeur de la fonction de coût appliquée au jeu de données d'entraînement diminue. Elle peut donc être un indicateur de l'amélioration des prédictions du modèle d'une itération à une autre. Arrivé à un plateau, le modèle est optimisé et l'entraînement s'arrête.

Cette approche masque toutefois une spécialisation du modèle ou surentraînement. Prédire parfaitement $\{ y_{\text{vraie},i} \}$ sur le jeu de données d'entraînement, ce qui correspond à une fonction de coût nulle, n'est pas équivalent à être optimal sur des données inédites. Or, le but est justement d'utiliser le modèle sur ces dernières. Un autre modèle, moins complexe ou entraîné avec moins d'itérations, peut donc donner de meilleures prédictions sur des données inédites.

Afin d'éviter le surentraînement, il est possible d'utiliser un jeu de données dit de « validation », non utilisé pour régler les paramètres du modèle. L'intérêt du jeu de validation est illustré sur la figure 5.4. Un modèle sous-entraîné ou dont l'entraînement est optimal présente des erreurs similaires dans les deux jeux de données. Dans le cas d'un surentraînement, les erreurs continuent à diminuer sur le jeu d'entraînement, mais pas sur le jeu de validation. Une fonction d'évaluation E, éventuellement égale à la fonction de coût L, permet de quantifier ces erreurs et de mettre fin à l'entraînement avant de surentraîner le modèle. Il s'agit de l'arrêt prématuré. La condition activant cet arrêt est

un « hyper-paramètre » du modèle. Les hyper-paramètres, à ne pas confondre avec les paramètres modifiés lors de l'entraînement, sont fixés par l'utilisateur et propres au modèle.

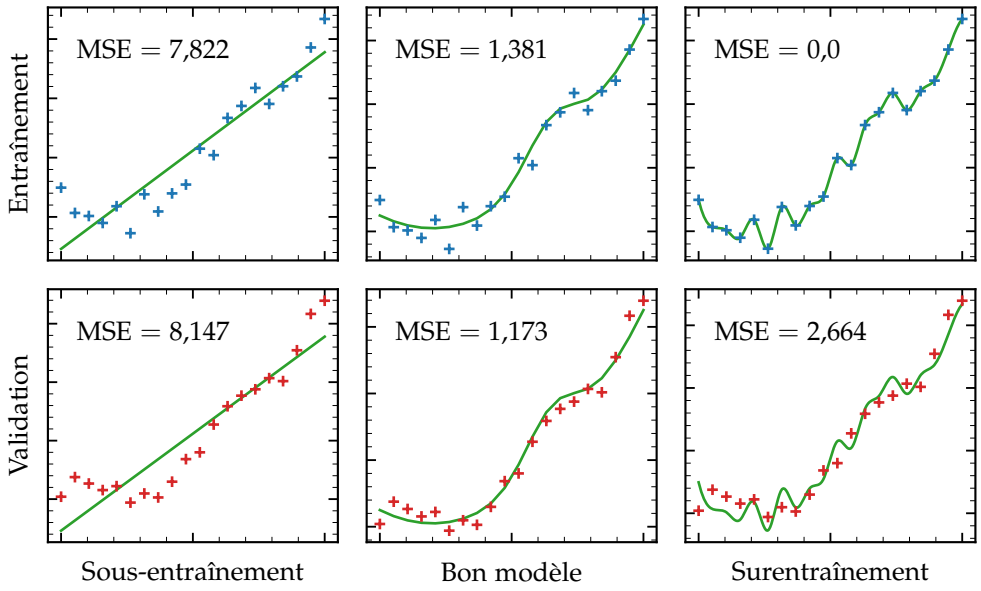


Figure 5.4 – *Illustrations du sous-entraînement et du surentraînement. Un même modèle est trop peu (gauche), suffisamment (milieu) ou trop entraîné (droite). Ses prédictions (ordonnées) en fonction de l'entrée (abscisses) sont tracées en vert. Le jeu de données d'entraînement (de validation) est représenté par des croix bleues (rouges) sur la ligne du haut (bas). Les valeurs de* L_{MSE} *sont également données afin d'illustrer les variations de l'erreur discutées dans le texte.*

4 Arbres de décision améliorés

La librairie XGBOOST [160] (*eXtreme Gradient Boosting*) permet de construire des arbres de décision améliorés. De nombreuses compétitions Kaggle [270] ont été remportées grâce à eux. Ils présentent l'avantage d'être généralement plus rapides à entraîner que les réseaux de neurones présentés section 5, et peuvent fournir des prédictions même si une des variables d'entrée est manquante, ce qui n'est pas le cas des réseaux de neurones.

4.1 Arbres de décision

Un arbre de décision (non amélioré) est une succession de questions dont les réponses mènent à un résultat final, comme illustré sur la figure 5.5. Chaque réponse à une question crée une « branche » menant à une nouvelle question (en bleu) ou à une réponse finale sur une « feuille » (en vert).

De tels arbres peuvent être utilisés avec des variables numériques. Dans ce cas, chaque question consiste en une condition sur l'une des variables, par exemple $p_{\text{T}}{}^{\mu} > 50\,\text{GeV}$. Le choix de la variable ($p_{\text{T}}{}^{\mu}$) et de la coupure correspondante (50 GeV) à utiliser pour former deux nouvelles branches b_1 (condition fausse) et b_2 (condition vraie) se base sur la similarité S. Il s'agit d'une variable quantifiant les écarts d'une liste de valeurs y_i à la moyenne de celles-ci $\langle y \rangle$. Elle est définie comme

$$S = \frac{1}{N}\left(\sum_{i=1}^{N} r_i\right)^2, \quad r_i = y_i - \frac{1}{N}\sum_{j=1}^{N} y_j = y_i - \langle y \rangle \tag{5.9}$$

où N est la taille de la liste de valeurs, r_i le résiduel de y_i et y_i la i^{e} valeur de y.

Le gain G obtenu par la création de deux nouvelles branches b_1 et b_2 s'exprime

$$G = S_{b_1} + S_{b_2} - S_{b_1+b_2} \tag{5.10}$$

avec $S_{b_1+b_2}$ la similarité de la liste non séparée, S_{b_1} (S_{b_2}) la similarité de la liste se retrouvant dans la branche b_1 (b_2). La condition retenue pour former les deux branches est celle présentant le gain

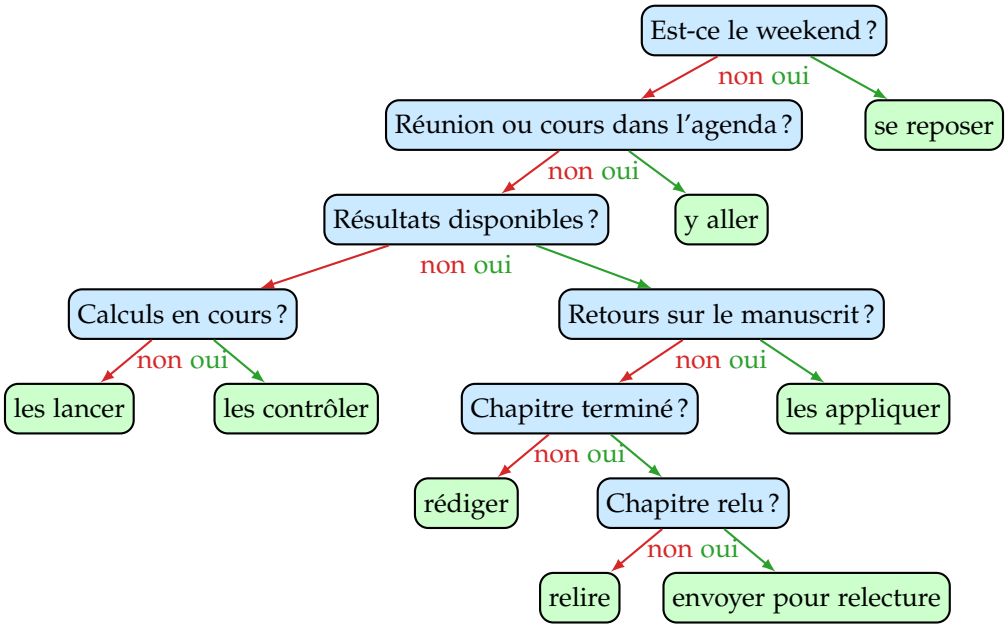


Figure 5.5 – *Exemple d'un arbre de décision utilisable par un doctorant.*

le plus élevé. Cela revient à définir deux sous-listes dans lesquelles les valeurs de y sont proches les unes des autres. Ce processus est alors itéré sur chacune des nouvelles branches, jusqu'à ce que :

— le gain soit inférieur à une valeur γ fixée ;

— la profondeur de l'arbre soit égale à une valeur $N_{\text{max}}^{\text{prof.}}$ fixée ;

— la quantité d'échantillons dans une branche soit inférieure à une valeur $N_{\text{min}}^{\text{échant.}}$ fixée.

Les paramètres γ, $N_{\text{max}}^{\text{prof.}}$ et $N_{\text{min}}^{\text{échant.}}$, choisis par l'utilisateur, sont des hyper-paramètres.

4.2 *Gradient Boosting* et descente de gradient

La technique du *Gradient Boosting* est l'utilisation de modèles simples, ici des arbres de décision, pour obtenir un modèle global plus robuste. La construction se fait de manière itérative.

La première étape consiste à créer un arbre de décision, noté M_0, comme exposé dans la section 4.1. La fonction associée à ce modèle est F_0. Puis à chaque étape $k \geqslant 1$, un arbre de décision M_k est construit avec pour objectif de prédire, pour une entrée $\vec{x}_i$,

$$y_{\text{vraie},i} - F_{k-1}(\vec{x}_i) \tag{5.11}$$

avec $y_{\text{vraie},i}$ la valeur que doit prédire le modèle global et F_{k-1} la fonction du modèle issu de l'étape $k-1$. Le modèle M_k corrige donc l'écart résiduel des prédictions $\{ y_{\text{préd},i} \}$ du modèle F_{k-1} à $\{ y_{\text{vraie},i} \}$. Les prédictions F_k du modèle global s'expriment donc

$$y_{\text{préd},i} = F_k(\vec{x}_i) = F_{k-1}(\vec{x}_i) + \eta\, M_k(\vec{x}_i) = M_0(\vec{x}_i) + \eta \sum_{l=1}^{k} M_l(\vec{x}_i) \tag{5.12}$$

avec η le taux d'apprentissage, inférieur à 1, permettant de corriger progressivement l'écart résiduel. L'itération s'arrête lorsque le nombre maximal d'estimateurs $N_{\text{max}}^{\text{estim.}}$ est atteint. Les grandeurs η et $N_{\text{max}}^{\text{estim.}}$ sont également des hyper-paramètres. Le modèle global obtenu est ici un arbre de décision amélioré.

La dérivée partielle de L_{MSE} par rapport à $y_{\text{préd},i}$ s'exprime

$$\frac{\partial \mathsf{L}_{\text{MSE}}(y_{\text{vraie},i}, y_{\text{préd},i})}{\partial y_{\text{préd},i}} = y_{\text{préd},i} - y_{\text{vraie},i}\,. \tag{5.13}$$

Ainsi, la cible de M_k définie précédemment est

$$y_{\text{vraie},i} - F_{k-1}(\vec{x}_i) = -\frac{\partial \mathsf{L}_{\text{MSE}}(y_{\text{vraie},i}, F_{k-1}(\vec{x}_i))}{\partial F_{k-1}(\vec{x}_i)}\,, \tag{5.14}$$

ce qui revient à appliquer la descente de gradient avec $\mathsf{L} = \mathsf{L}_{\text{MSE}}$. À partir de ce constat, il est possible de généraliser le *Gradient Boosting* en considérant que la cible de M_k est

$$-\frac{\partial \mathsf{L}(y_{\text{vraie},i}, F_{k-1}(\vec{x}_i))}{\partial F_{k-1}(\vec{x}_i)} = -\vec{\nabla}_{F_{k-1}(\vec{x}_i)} \left(\mathsf{L}(y_{\text{vraie},i}, F_{k-1}(\vec{x}_i))\right) \tag{5.15}$$

avec L une fonction de coût quelconque.

Une itération de l'entraînement consiste ainsi en l'ajout d'un estimateur au modèle. Un arrêt prématuré est réalisé lorsque l'erreur quadratique moyenne (L_{MSE}) ne diminue pas sur le jeu de validation pendant 5 itérations.

5 Réseaux de neurones profonds

Les réseaux de neurones (NN, *Neural Networks*) sont un autre type de modèle permettant d'approximer la fonction reliant les entrées $\{\vec{x}_i\}$ aux cibles $\{y_{\text{vraie},i}\}$ [175, 271]. Les librairies KERAS [168] et TENSORFLOW [169] sont utilisées afin de construire et d'entraîner ces modèles.

5.1 Neurones

5.1.1 Principe

Un neurone est une entité ayant un certain nombre d'entrées x_j, $j \in \{1, \ldots, n\}$, auxquelles sont associées des poids w_j, un biais b et une fonction f dite d'« activation », discutée section 5.1.2. Les poids w_j et le biais b sont les paramètres du neurone, la fonction d'activation est un hyper-paramètre. La sortie s du neurone s'exprime comme

$$s = f\left(\sum_{j=1}^{n} w_j x_j + b\right). \tag{5.16}$$

Le fonctionnement d'un neurone est résumé sur la figure 5.6.

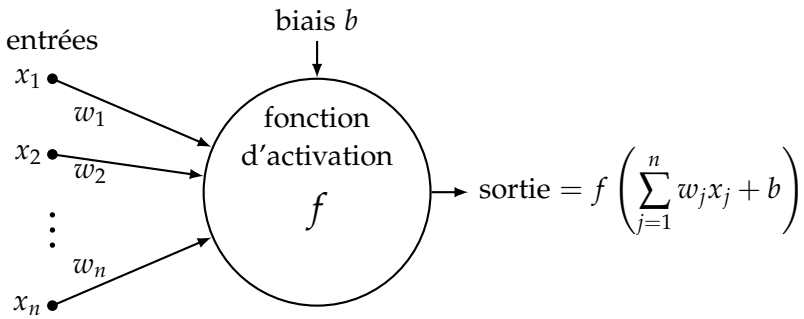


Figure 5.6 – *Structure d'un neurone. Une fonction f dite d'« activation » est appliquée à la somme des entrées x_j pondérées par les poids w_j et du biais b afin d'obtenir la valeur de sortie.*

5.1.2 Fonctions d'activation

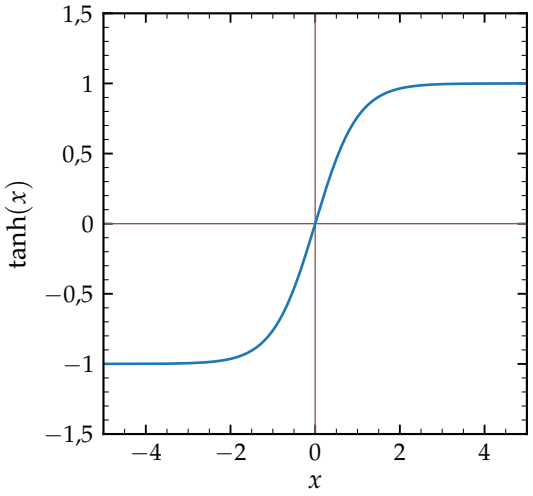
En principe, toute fonction définie sur l'ensemble d'existence de chacune des entrées x_j peut être utilisée comme fonction d'activation. Elles sont ainsi définies sur $\mathbb{R}$. Les plus utilisées sont :

tangente hyperbolique notée tanh, définie par

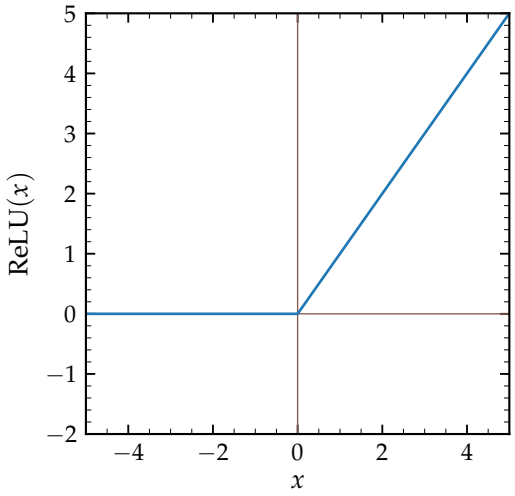
$$\tanh : x \mapsto \frac{e^x - e^{-x}}{e^x + e^{-x}}\ ; \tag{5.17}$$

sigmoïde notée sig, définie par

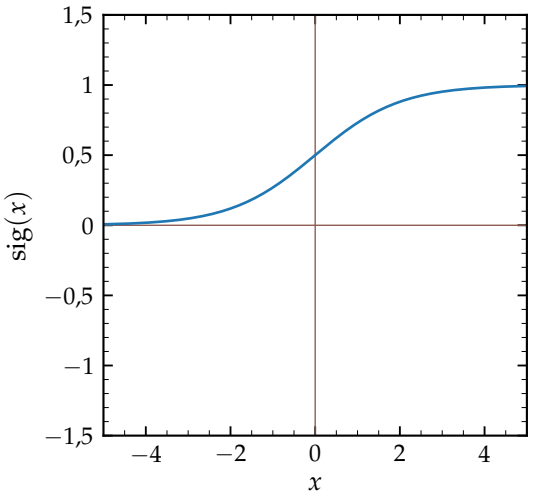
$$\text{sig} : x \mapsto \frac{1}{1 + e^{-x}}\ ; \tag{5.18}$$



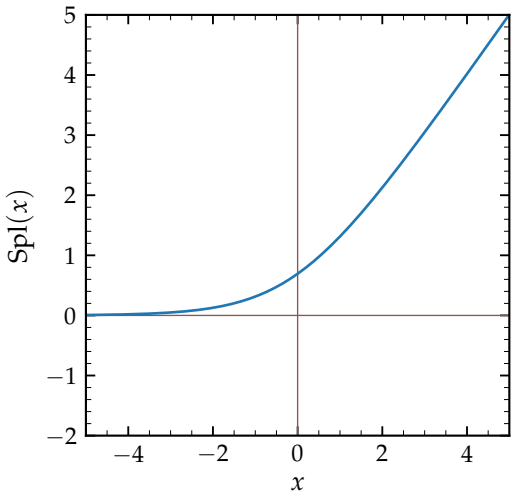
(a) *Tangente hyperbolique.*



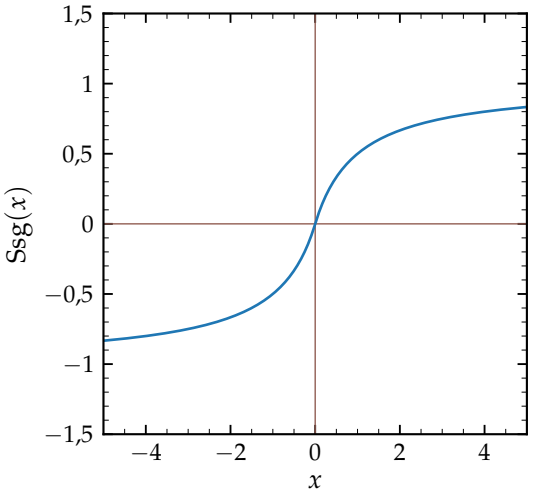
(b) *ReLU.*



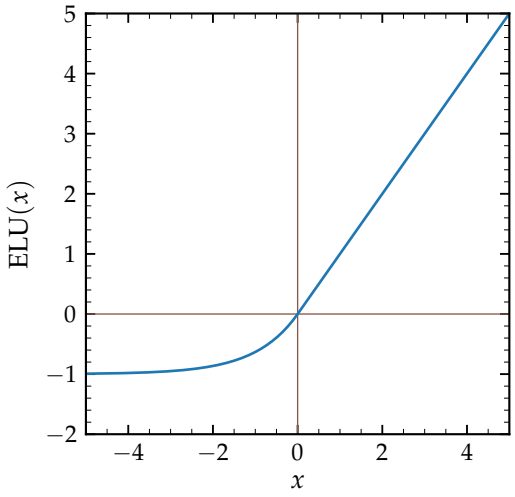
(c) *Sigmoïde.*



(d) *Softplus.*



(e) *Softsign.*



(f) *ELU.*

Figure 5.7 – *Exemples de fonctions d'activation. À gauche, des fonctions à valeurs bornées, généralement utilisées en classification. À droite, des fonctions à valeurs non bornées, utilisables pour des tâches de régression.*

Softsign notée Ssg, définie par

$$\mathrm{Ssg} : x \mapsto \frac{x}{1+|x|} \; ; \tag{5.19}$$

ReLU (*Rectified Linear Unit*), définie par

$$\mathrm{ReLU} : x \mapsto \begin{cases} x\,, & x > 0 \\ 0\,, & x \leqslant 0 \end{cases} \; ; \tag{5.20}$$

Softplus notée Spl, définie par

$$\mathrm{Spl} : x \mapsto \ln(1+\mathrm{e}^{x}) \; ; \tag{5.21}$$

ELU (*Exponential Linear Unit*), définie par

$$\mathrm{ELU} : x \mapsto \begin{cases} x\,, & x > 0 \\ \alpha(\mathrm{e}^{x}-1)\,, & x \leqslant 0 \end{cases}, \quad \alpha = 1 \; ; \tag{5.22}$$

SELU (*Scaled Exponential Linear Unit*), similaire à ELU et définie par

$$\mathrm{SELU} : x \mapsto \lambda \times \begin{cases} x\,, & x > 0 \\ \alpha(\mathrm{e}^{x}-1)\,, & x \leqslant 0 \end{cases}, \quad \alpha \simeq 1{,}67\,, \quad \lambda \simeq 1{,}05 \; ; \tag{5.23}$$

ou encore la fonction linéaire identité $\mathbb{1} : x \mapsto x$. Certaines d'entre elles sont représentées sur la figure 5.7.

5.2 Réseaux de neurones

Un NN est obtenu par l'interconnexion de plusieurs neurones entre eux. Ces connexions peuvent se faire selon diverses architectures [175, 271]. Nous utilisons ici, comme dans les travaux de BÄRTSCHI & coll. [257], une architecture normale profonde à propagation avant complètement connectée (*normal deep feedforward fully-connected*), représentée sur la figure 5.8, c'est-à-dire avec :

— des neurones répartis en couches (normale) ;
— plusieurs couches « cachées », situées entre les couches d'entrée et de sortie (profonde) ;
— toutes les sorties de la couche $k-1$ utilisées comme entrées de chacun des neurones de la couche k (à propagation avant complètement connectée).

Le nombre de neurones par couche cachée est noté $N_{\mathrm{n/c}}$, le nombre de couches cachées N_{cc}. Le NN ayant une structure profonde, il s'agit d'un DNN (*Deep Neural Network*).

La tâche du réseau est une régression vers une seule grandeur, $m_{\mathcal{H}}$, à partir de n variables d'entrée x_j, $j \in \{1, \ldots, n\}$. La couche de sortie est donc composée d'un seul neurone dont la fonction d'activation est l'identité. La couche d'entrée comporte n neurones, chacun se contentant de transmettre la variable d'entrée correspondante. Il s'agit donc d'une couche d'adaptation entre le nombre d'entrées n_{in} et le nombre de neurones dans la couche suivante $N_{\mathrm{n/c}}$. Tous les neurones des couches cachées ont la même fonction d'activation. Plusieurs fonctions d'activation sont testées dans la section 6.

5.3 Entraînement

L'entraînement d'un NN est le réglage des paramètres des neurones du réseau situés sur les couches cachées et la couche de sortie. Il s'agit des poids w_i et du biais b. Pour un DNN avec $n_{\mathrm{in}} = 27$ variables d'entrée, $N_{\mathrm{cc}} = 3$ couches cachées de $N_{\mathrm{n/c}} = 1000$ neurones, le nombre de paramètres est ainsi de

$$\begin{aligned} N_{\mathrm{params.}} &= \underbrace{N_{\mathrm{n/c}} \times (n_{\mathrm{in}}+1)}_{\text{couche cachée 1}} + \underbrace{(N_{\mathrm{cc}}-1) \times N_{\mathrm{n/c}} \times (N_{\mathrm{n/c}}+1)}_{\text{autres couches cachées}} + \underbrace{N_{\mathrm{n/c}}+1}_{\text{couche de sortie}} \\ &= 28\,000 + 2 \times 1\,001\,000 + 1001 = 2\,031\,001\,, \end{aligned} \tag{5.24}$$

soit près de deux millions. Les termes « $+1$ » correspondent aux biais b à ajouter au nombre d'entrées des neurones.

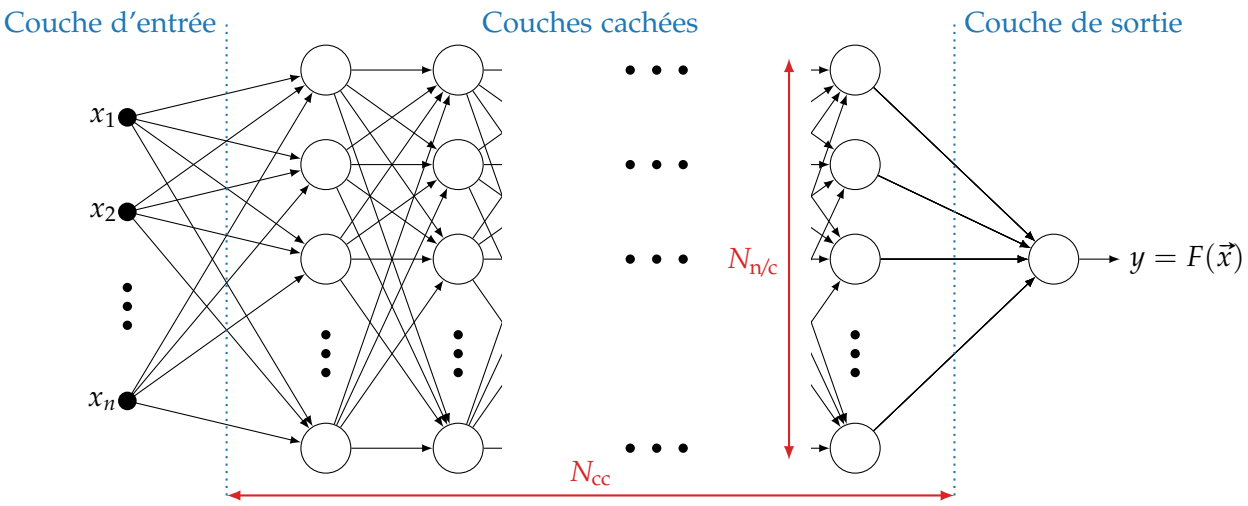


Figure 5.8 – *Structure normale profonde à propagation avant complètement connectée d'un réseau de neurones. Une couche d'entrée comporte autant de neurones que de variables x_i. La couche de sortie en comporte autant que de valeurs à donner, c'est-à-dire une. Les fonctions d'activation de ces deux couches sont linéaires. Entre elles se trouvent N_{cc} couches cachées, chacune contenant $N_{n/c}$ neurones. Diverses fonctions d'activation peuvent être utilisées dans les couches cachées.*

5.3.1 Initialisation des paramètres

Les biais b sont initialement fixés à 0, les poids w_i à une valeur constante donnée ou aléatoirement selon une loi de probabilité. Le mode d'initialisation est un hyper-paramètre du modèle. Lors de ces travaux, nous avons testé les lois normale et uniforme. Dans le cas des DNNs, ces modes d'initialisation peuvent être améliorés par la méthode de GLOROT & BENGIO [272] afin de faciliter l'entraînement. Il s'agit alors des lois « Glorot uniforme » et « Glorot normale », également testées.

5.3.2 Optimisation des paramètres

L'optimisation des paramètres est réalisée en utilisant les mini-lots introduits en section 3.2. Un maximum de 500 époques est autorisé, avec un arrêt prématuré au bout de 20 époques sans diminution de l'erreur absolue moyenne (L_{MAE}) sur les données de validation. Plusieurs algorithmes d'optimisation existent [175], présentés de manière non exhaustive ci-après.

***Batch Gradient Descent* (BGD)** [269] L'algorithme BGD, déjà introduit en section 3.2, applique la méthode de descente de gradient sur les mini-lots. Le bruit dû à la composition aléatoire des mini-lots permet de s'échapper des minimums locaux mais dégrade la précision une fois au minimum global. Pour pallier cet effet, le taux d'apprentissage η peut être diminué à chaque époque. La condition sur les taux d'apprentissage η_k avec k l'époque afin de s'assurer de la convergence du modèle optimisé par BGD est [175]

$$\sum_{k=1}^{\infty} \eta_k = \infty\,, \quad \sum_{k=1}^{\infty} \eta_k^2 < \infty\,. \tag{5.25}$$

La mise à jour des paramètres à la fin d'un mini-lot pendant l'époque k est alors réalisée selon

$$p \to p - \eta_k \langle \vec{\nabla}(\mathsf{L}) \rangle_{\text{mini-lot}} \cdot \vec{e}_p = p - \eta_k \left\langle \frac{\partial \mathsf{L}}{\partial p} \right\rangle_{\text{mini-lot}}\,. \tag{5.26}$$

BGD avec moments [175] Les moments sont une « mémoire » des valeurs du gradient de la fonction de coût des époques précédentes. Ce peut être vu comme une inertie du mouvement du modèle dans l'espace des paramètres, prise en compte à travers une vitesse $\vec{v}$ définie initialement par l'utilisateur

et mise à jour à chaque mini-lot selon

$$\vec{v}[t-1] \to \vec{v}[t] = \alpha\vec{v}[t-1] - \eta_k \langle \vec{\nabla}(\mathsf{L})[t] \rangle_{\text{mini-lot}} \tag{5.27}$$

$$\Rightarrow \vec{v}[t] \cdot \vec{e}_p = v_p[t] = \alpha v_p[t-1] - \eta_k \left\langle \frac{\partial \mathsf{L}}{\partial p}[t] \right\rangle_{\text{mini-lot}} \tag{5.28}$$

avec t l'indice d'itération ou indice temporel de l'entraînement, et $0 \leqslant \alpha < 1$ le paramètre des moments. La mise à jour des paramètres lors de l'itération t se fait alors selon

$$p[t-1] \to p[t] = p[t-1] + v_p[t] = p[t-1] + \alpha v_p[t-1] - \eta_k \left\langle \frac{\partial \mathsf{L}}{\partial p}[t] \right\rangle_{\text{mini-lot}} . \tag{5.29}$$

***Adaptive Gradient* (AdaGrad)** [273] L'algorithme AdaGrad adapte le taux d'apprentissage individuellement pour chaque paramètre p à l'aide d'une variable de mémoire $\vec{r}$. Elle est initialement définie à $\vec{0}$ et est modifiée à chaque mini-lot selon

$$\vec{r} \cdot \vec{e}_p = r_p \to r_p + \left\langle \frac{\partial \mathsf{L}}{\partial p} \right\rangle^2_{\text{mini-lot}} . \tag{5.30}$$

La mise à jour des paramètres se fait alors suivant

$$p \to p - \eta \frac{1}{\sqrt{r_p} + \delta} \left\langle \frac{\partial \mathsf{L}}{\partial p} \right\rangle_{\text{mini-lot}} \tag{5.31}$$

où δ est une variable de régularisation évitant les divisions par zéro. Le taux d'apprentissage effectif pour le paramètre p est ainsi η divisé par la somme quadratique des gradients précédents $\sqrt{r_p}$.

Plus un paramètre modifie la valeur de la fonction de coût, plus sa modification est progressive. Dans l'optique de la recherche d'un minimum, cela revient à descendre une pente lentement et à se mouvoir rapidement dans une direction plane. Cependant, l'accumulation depuis le début de l'entraînement des gradients au carré dans r_p peut mener à une diminution excessive du taux d'apprentissage effectif d'un paramètre.

RMSProp [274] L'algorithme RMSProp consiste en une légère modification de AdaGrad. Une décroissance exponentielle de la mémoire des gradients passés est mise en place en remplaçant (5.30) par

$$r_p \to \rho\, r_p + (1-\rho) \left\langle \frac{\partial \mathsf{L}}{\partial p} \right\rangle^2_{\text{mini-lot}} \tag{5.32}$$

où $0 < \rho < 1$ est le taux de diminution de la mémoire. RMSProp est ainsi une version de AdaGrad dont la mémoire est plus adaptée à la situation locale.

***Adaptive Delta* (AdaDelta)** À l'instar de RMSProp, AdaDelta est une modification de AdaGrad visant à améliorer l'effet de mémoire. La variable r_p est mise à jour par (5.30). Cependant, la valeur précédente de r_p est également utilisée lors de la mise à jour de p. Ainsi, lors de l'itération t,

$$p[t-1] \to p[t] = p[t-1] - \frac{\sqrt{r_p[t-1]} + \delta}{\sqrt{r_p[t]} + \delta} \left\langle \frac{\partial \mathsf{L}}{\partial p}[t] \right\rangle_{\text{mini-lot}} . \tag{5.33}$$

Il n'y a donc pas besoin de définir un taux d'apprentissage initial avec AdaDelta.

***Adaptive Moments* (Adam)** [175, 275] L'algorithme Adam est une combinaison de la méthode des moments et de RMSProp. Il adapte donc le taux d'apprentissage pour chaque paramètre à chaque mini-lot. Pour cela sont définis initialement :

— le pas $\epsilon = 0{,}001$;
— les moments d'ordres 1 et 2, $\vec{v} = \vec{0}$ et $\vec{r} = \vec{0}$;
— les taux de diminution de moments d'ordre 1 et 2, $\rho_1 = 0{,}9$ et $\rho_2 = 0{,}999$;

— le paramètre temporel $t = 0$.

Puis, à chaque mini-lot, les moments sont redéfinis selon

$$\vec{v}\cdot\vec{e}_p = v_p \to \rho_1 v_p + (1-\rho_1)\left\langle\frac{\partial \mathsf{L}}{\partial p}\right\rangle_{\text{mini-lot}}, \quad \vec{r}\cdot\vec{e}_p = r_p \to \rho_2 r_p + (1-\rho_2)\left\langle\frac{\partial \mathsf{L}}{\partial p}\right\rangle^2_{\text{mini-lot}}. \tag{5.34}$$

Le biais d'initialisation des moments est corrigé en appliquant

$$t \to t+1, \quad v_p \to \frac{v_p}{1-\rho_1^t}, \quad r_p \to \frac{r_p}{1-\rho_2^t}. \tag{5.35}$$

Les paramètres du modèle sont alors mis à jour selon

$$p \to p - \epsilon\frac{v_p}{\sqrt{r_p}+\delta} \tag{5.36}$$

où $\delta = 10^{-8}$ permet de stabiliser les calculs en évitant une division par zéro.

6 Optimisation des hyper-paramètres et choix d'un modèle

Le choix d'un modèle et de ses hyper-paramètres est l'objet de cette section. Les hyper-paramètres des XGBs, introduits section 4, sont :

— la profondeur maximale des arbres $N_{\text{max}}^{\text{prof.}}$;
— la quantité d'échantillons minimale dans une branche $N_{\text{min}}^{\text{échant.}}$;
— le nombre d'arbres $N_{\text{max}}^{\text{estim.}}$;
— le gain minimal γ ;
— le taux d'apprentissage η ;
— la fonction de coût L ;
— la liste des variables d'entrée.

Les hyper-paramètres des DNNs, introduits section 5, sont :

— le nombre de couches cachées N_{cc} ;
— le nombre de neurones par couche cachée $N_{\text{n/c}}$;
— la fonction d'activation des neurones des couches cachées ;
— l'algorithme d'optimisation ;
— la fonction de coût L ;
— le mode d'initialisation des poids ;
— la liste des variables d'entrée.

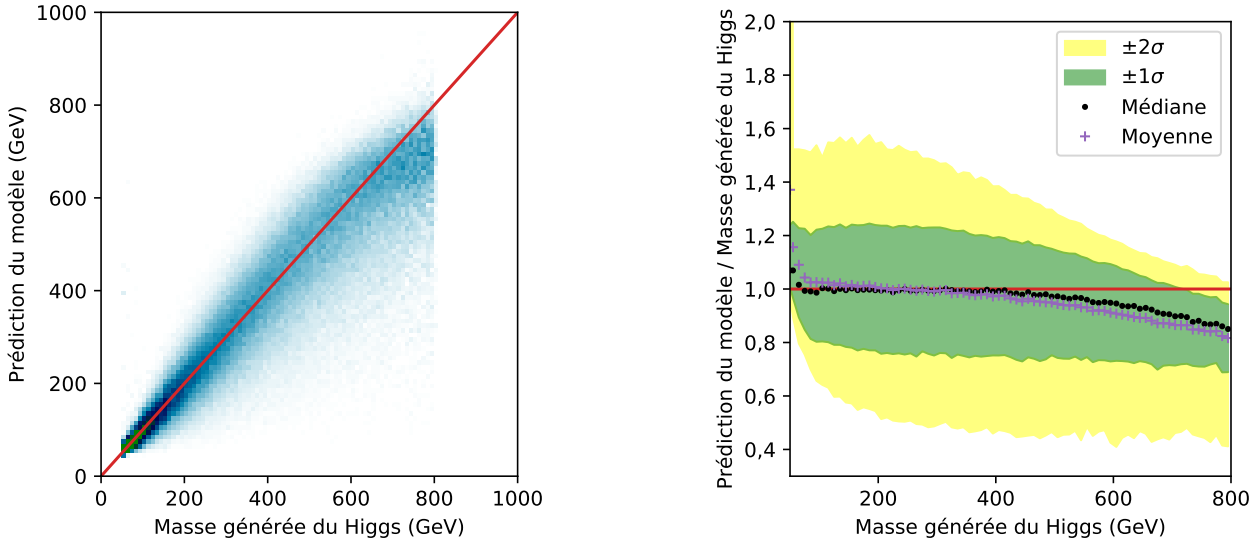
Dans ce chapitre, les modèles ont pour but de prédire la masse générée du boson de Higgs $m_{\mathcal{H}}$. Une représentation graphique possible afin de montrer les performances d'un modèle est de tracer ses prédictions $y_{\text{préd}}$ en fonction de y_{vraie} dans un histogramme à deux dimensions comme sur la figure 5.9a. L'objectif des modèles est alors de se rapprocher autant que possible de la première bissectrice, tracée en rouge. Toutefois la large gamme explorée, de 50 à 800 GeV, rend difficile la visualisation des performances à basse masse. Or, cette région est importante car elle contient les bosons Z et h du modèle standard. La réponse R du modèle, définie comme

$$R = \frac{y_{\text{préd}}}{y_{\text{vraie}}} = \frac{F(\vec{x})}{m_{\mathcal{H}}}, \tag{5.37}$$

permet de ramener l'objectif des modèles à 1 sur toute la gamme de masse. La réponse du même modèle est ainsi représentée sur la figure 5.9b. Pour chaque intervalle de 10 GeV sur $m_{\mathcal{H}}$, la distribution de r est déterminée. La valeur moyenne et la médiane de cette distribution sont données, ainsi que les largeurs à $\pm1\sigma$ et $\pm2\sigma$, correspondant respectivement aux zones contenant 68 et 95 % des valeurs de R les plus proches de la médiane. Ces zones sont déterminées de manière indépendante sur les valeurs inférieures et supérieures à la médiane, elles peuvent donc être asymétriques.

Il est difficile de définir un seul score quantifiant la qualité d'un modèle. Plusieurs métriques sont considérées afin de l'évaluer :

— les valeurs de L_{MSE}, L_{MAE}, L_{MAPE} ;



(a) *Histogramme à deux dimensions de $y_{\text{préd}}$ en fonction de y_{vraie}.*

(b) *Réponse du modèle $y_{\text{préd}}/y_{\text{vraie}}$ en fonction de y_{vraie}.*

Figure 5.9 – *Exemples de graphiques rendant compte des performances des modèles.*

— la largeur de -1σ à $+1\sigma$ de la réponse R du modèle, notée $\Delta_{1\sigma}$, égale à la moyenne sur des intervalles de 10 GeV sur $y_{\text{vraie}} = m_{\mathcal{H}}$ de la distance entre les écarts-types supérieur et inférieur de la distribution de R, c'est-à-dire

$$\Delta_{1\sigma} = \left\langle \left[\sigma_+ \left(\frac{y_{\text{préd}}}{y_{\text{vraie}}} \right) - \sigma_- \left(\frac{y_{\text{préd}}}{y_{\text{vraie}}} \right) \right] \Big|_{y_{\text{vraie}} \in [n, n+1] \times 10\,\text{GeV}} \right\rangle_n . \tag{5.38}$$

Il s'agit donc de la moyenne de la largeur verticale des bandes vertes ($\pm 1\sigma$) sur les graphiques des réponses des modèles comme celui de la figure 5.9b, aussi égale à deux fois la résolution relative du modèle.

Pour toutes ces métriques, l'objectif est d'avoir la plus petite valeur possible. De plus, quatre domaines de masse sont définis :

— basse masse : $m_{\mathcal{H}} < 150\,\text{GeV}$, incluant en particulier les bosons Z et h ;
— moyenne masse : $150\,\text{GeV} \leqslant m_{\mathcal{H}} < 500\,\text{GeV}$;
— haute masse : $m_{\mathcal{H}} \geqslant 500\,\text{GeV}$;
— toute masse : aucune restriction sur $m_{\mathcal{H}}$.

Ils permettent de comparer les performances des modèles sur certaines gammes de masse uniquement. Sauf contre-indication, toute la gamme de masse est considérée.

Face à l'immense quantité de combinaisons différentes d'hyper-paramètres, toutes n'ont pas été testées. Nous avons en revanche entraîné suffisamment de modèles afin d'observer les distributions des différentes métriques d'évaluation pour des groupes de modèles ayant une valeur donnée d'un hyper-paramètre. La comparaison des différentes distributions permet dans un premier temps de voir quelles valeurs d'hyper-paramètres donnent des modèles moins performants et ainsi se rapprocher d'une combinaison optimale, comme l'exposent les sections 6.1 à 6.4. Une fois certains hyper-paramètres fixés, la sélection finale d'un seul modèle est réalisée selon la procédure présentée en section 6.5.

6.1 Variables d'entrée

Utiliser le moins de variables d'entrée possible, si cela ne dégrade pas la qualité de nos modèles, pourrait faciliter leur intégration dans les analyses de CMS. En effet, si toutes les variables d'entrée considérées listées section 2.4 sont généralement déjà exploitées dans les analyses en cours, ce n'est toutefois pas toujours le cas, en particulier pour les variables relatives à l'activité hadronique ad-

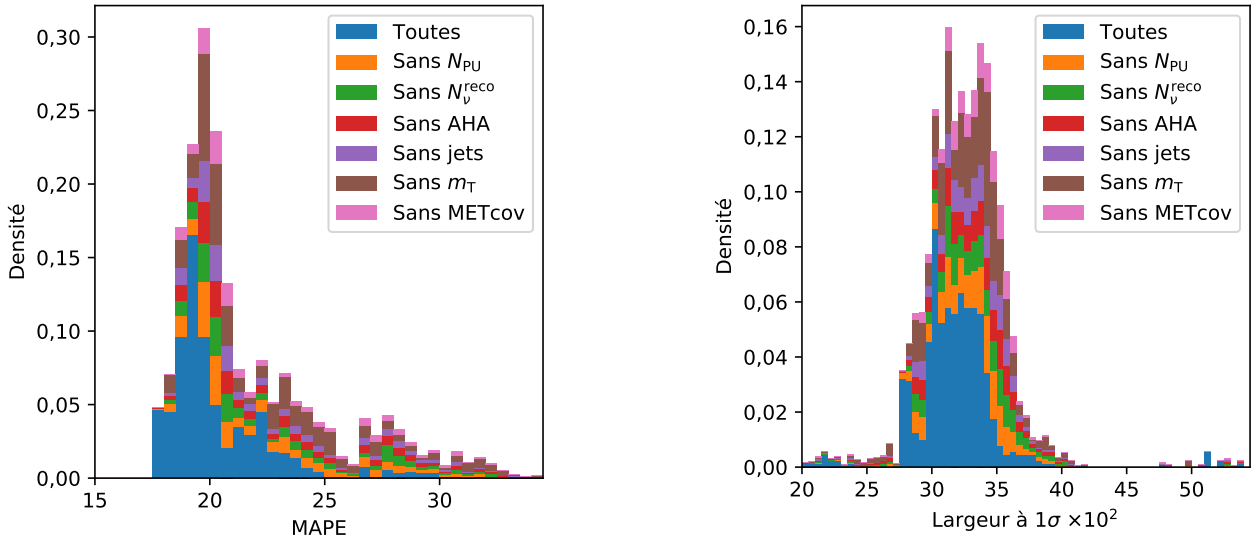
ditionnelle. Il est donc pertinent de déterminer les variables dont les modèles peuvent aisément se passer.

Pour cela, des sous-ensembles des variables d'entrée sont définis par les restrictions suivantes :

— sans N_{PU} : la variable N_{PU} n'est pas utilisée ;
— sans N_ν^{reco} : la variable N_ν^{reco} n'est pas utilisée ;
— sans AHA : les variables d'activité hadronique additionnelle ne sont pas utilisées ;
— sans jets : les variables relatives aux jets (dont AHA) ne sont pas utilisées ;
— sans m_{T} : les masses transverses ne sont pas utilisées ;
— sans METcov : la matrice de covariance de ${E_{\text{T}}}^{\text{miss}}$ n'est pas utilisée.

L'application de plusieurs de ces restrictions simultanément est également testée.

Les performances des modèles entraînés avec les différents ensembles de variables d'entrée sont données figure 5.10 pour les XGBs et figure 5.11 pour les DNNs. Les modèles concernés par plusieurs restrictions sont comptés de manière pondérée dans chaque groupe correspondant à une restriction unique. Par exemple, un modèle soumis à la restriction « sans N_{PU} » et « sans N_ν^{reco} » a un poids de $\frac{1}{2}$ dans chacun de ces deux groupes. Les histogrammes ainsi créés sont superposés. Il est alors possible de voir les contributions de chacune des restrictions aux valeurs obtenues sur la métrique d'évaluation illustrée.



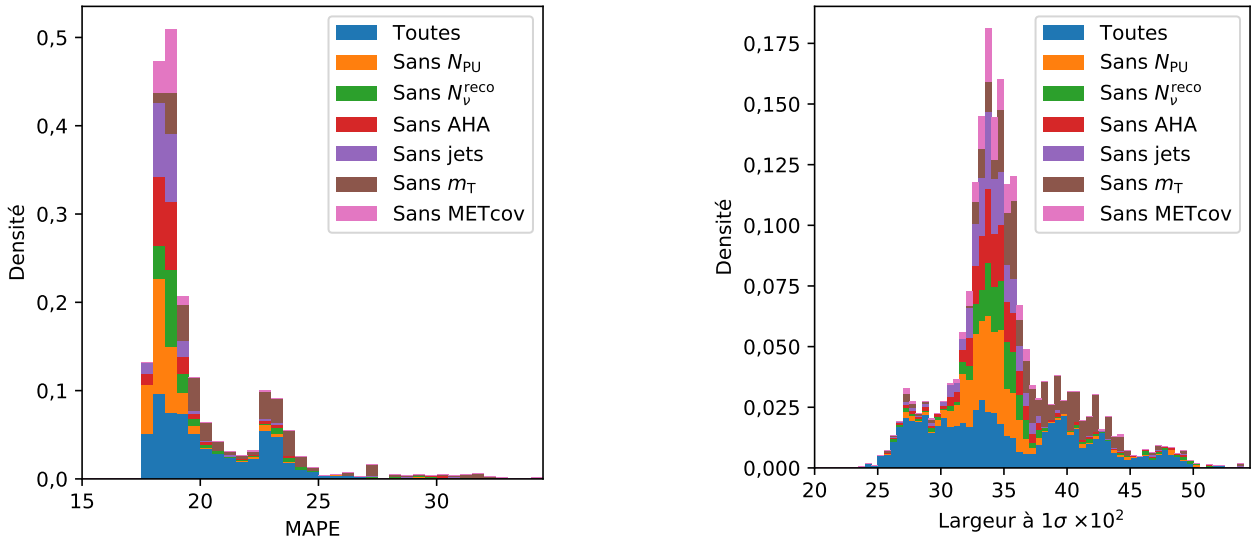
(a) *Évaluation par* L_{MAPE}. **(b)** *Évaluation par* $\Delta_{1\sigma}$.

Figure 5.10 – *Évaluations des XGBs regroupés selon les variables d'entrée.*

Dans le cas des XGBs, l'évaluation des modèles par L_{MAPE}, en figure 5.10a, donne des valeurs situées entre 17 et 35. Le cœur de la distribution, à $\text{L}_{\text{MAPE}} = 19 \pm 2$, est plutôt constitué de modèles utilisant toutes les entrées dans sa partie gauche ($\text{L}_{\text{MAPE}} < 19$) et de modèles utilisant un sous-ensemble d'entrées dans sa partie droite ($19 < \text{L}_{\text{MAPE}} < 22$). De plus, les basses valeurs de L_{MAPE}, en dessous de 18,5, sont presque exclusivement obtenues avec des modèles utilisant toutes les entrées. À l'inverse, la queue à hautes valeurs de la distribution obtenue ($\text{L}_{\text{MAPE}} > 23$) est largement dominée par les contributions des modèles avec un sous-ensemble d'entrées.

La plupart des XGBs ont une largeur $\Delta_{1\sigma}$, en figure 5.10b, située entre 27 et 38. Cependant, les XGBs utilisant toutes les variables d'entrée exhibent une distribution de $\Delta_{1\sigma}$ légèrement décalée vers de plus faibles valeurs.

Dans le cas des DNNs, la distribution de la métrique L_{MAPE}, en figure 5.11a, contient des valeurs situées majoritairement entre 17,5 et 25. Les DNNs n'utilisant pas N_ν^{reco} se situent à $\text{L}_{\text{MAPE}} > 18$. Cette variable permet aux modèles de différencier les canaux hadroniques, semi-leptoniques et leptoniques, dont la séparation est discutée dans la section 7. Ceux n'utilisant pas m_{T} présentent également des valeurs de L_{MAPE} uniquement au-delà de 18. L'utilisation de ces variables permet donc d'obtenir de meilleurs modèles. Elles sont de plus facilement obtenues à partir du dilepton, défini chapitre 4, et de ${E_{\text{T}}}^{\text{miss}}$. Les analyses avec deux leptons τ dans l'état final exploitent déjà ces observables, leur uti-



(a) *Évaluation par* L_{MAPE}. **(b)** *Évaluation par* $\Delta_{1\sigma}$.

Figure 5.11 – *Évaluations des DNNs regroupés selon les variables d'entrée.*

lisation par nos modèles est donc à la fois pertinente, car les scores de L_{MAPE} obtenus sont meilleurs, et sans incidence sur la facilité d'intégration du modèle à l'analyse. Les DNNs avec $L_{MAPE} \lesssim 18$ exploitent presque tous les variables relatives aux jets, à l'AHA et à la matrice de covariance de E_T^{miss}. Ces entrées sont donc vraisemblablement utiles aux DNNs afin de réaliser la régression. Enfin, la restriction sur N_{PU} ne semble pas dégrader les performances des DNNs selon L_{MAPE}.

La distribution de $\Delta_{1\sigma}$, en figure 5.11b, montre que les modèles utilisant toutes les entrées peuvent se répartir en plusieurs groupes, aux alentour des valeurs 0,275, 0,335, 0,395, 0,425 et 0,475. À 0,395 apparaît également un groupe de modèles entraînés sans m_T. À 0,335 se trouvent la majorité des modèles entraînés avec une restriction des entrées. Pour $\Delta_{1\sigma} < 0{,}3$, les modèles sont très majoritairement ceux utilisant l'ensemble des variables proposées.

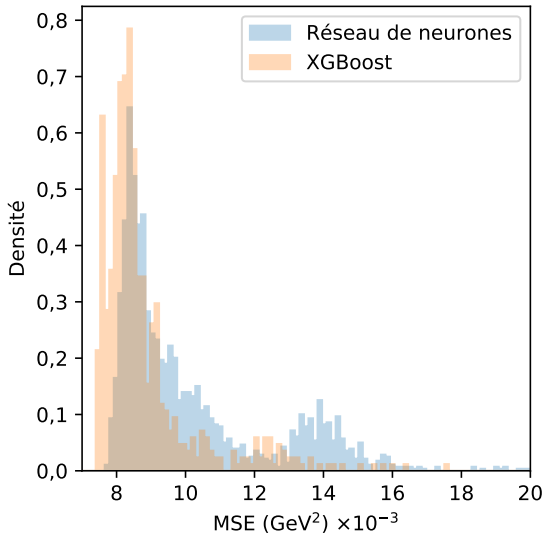
L'utilisation de toutes les variables listées dans la section 2.4 est donc corrélée avec de meilleures performances selon les métriques L_{MAPE} et $\Delta_{1\sigma}$. Par la suite, seuls les modèles utilisant toutes les variables, au nombre de 27, sont considérés.

6.2 Type de modèle

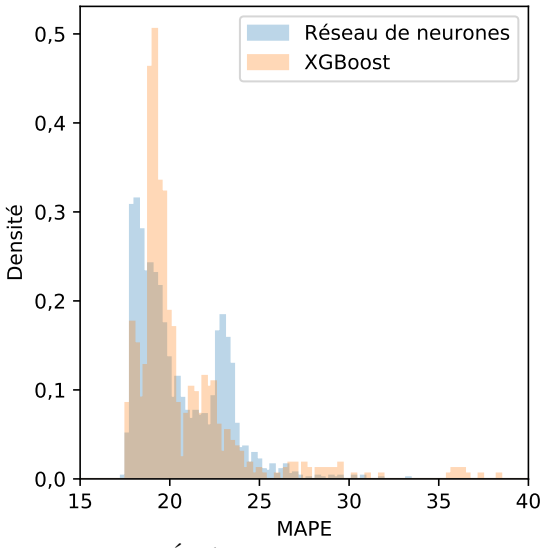
Les figures 5.12 et 5.13 présentent les distributions des scores de L_{MSE}, L_{MAPE} et $\Delta_{1\sigma}$ pour l'ensemble des DNNs et des XGBs utilisant toutes les variables d'entrée.

L'évaluation par L_{MSE}, en figure 5.12a, favorise les XGBs. Le cœur de la distribution de L_{MSE} pour ces modèles est en effet à $8{,}1 \times 10^3\,\text{GeV}^2$ contre $8{,}5 \times 10^3\,\text{GeV}^2$ pour les DNNs. En revanche, l'évaluation par L_{MAPE}, en figure 5.12b, favorise les DNNs avec un groupe de DNNs à $L_{MAPE} = 18$ contre 19 pour les XGBs. Un second groupe de DNNs est présent à $L_{MAPE} = 23$. L'existence de ces deux groupes est due à l'utilisation de plusieurs algorithmes d'optimisation, comme discuté dans la section 6.4.

La résolution des modèles est évaluée par $\Delta_{1\sigma}$ en figure 5.13a pour toute la gamme de masse et en figure 5.13b à basse masse. Sur l'ensemble de la gamme de masse, les XGBs ont un score de $0{,}32 \pm 0{,}04$ et les DNNs se répartissent en plusieurs groupes à environ 0,28, 0,33, 0,40, 0,42 et 0,48. Les XGBs sont ainsi compétitifs d'après cette évaluation. Cependant, les performances des modèles à basse masse, c'est-à-dire pour $m_{\mathcal{H}} < 150\,\text{GeV}$, sont importantes car c'est dans cette gamme de masse que se trouvent les bosons Z et h du modèle standard. En particulier, il faut s'assurer que le signal du Z présente une queue à haute masse, c'est-à-dire dans la région de signal des bosons de Higgs. Dans le cadre des analyses telles que celle présentée au chapitre 4, le Z est en effet un bruit de fond important. Sur la figure 5.13b montrant l'évaluation à basse masse par $\Delta_{1\sigma}$ des modèles, les scores des XGBs se situent majoritairement à $0{,}70 \pm 0{,}15$ alors que ceux des DNNs donnent deux groupes, le premier à $0{,}42 \pm 0{,}05$ et le second entre 0,50 et 1,0. Le premier ensemble de DNNs propose les

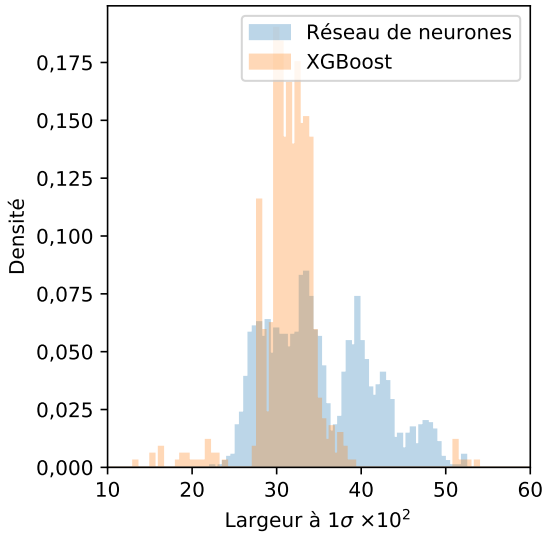


(a) *Évaluation par* $\mathsf{L}_{\mathrm{MSE}}$.

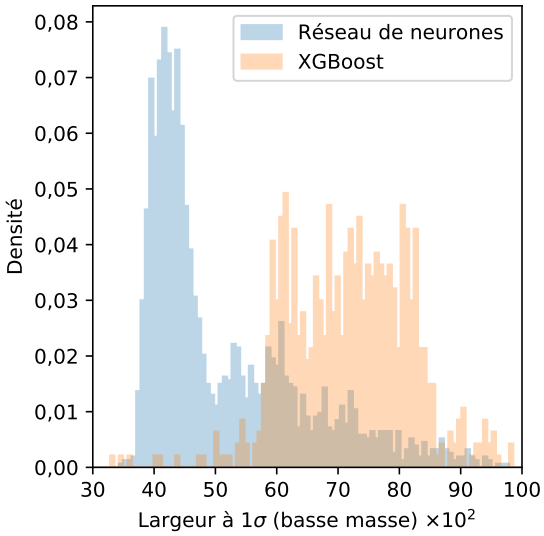


(b) *Évaluation par* $\mathsf{L}_{\mathrm{MAPE}}$.

Figure 5.12 – *Évaluations des XGBs et des DNNs par* $\mathsf{L}_{\mathrm{MSE}}$ *et* $\mathsf{L}_{\mathrm{MAPE}}$.



(a) *Sur toute la gamme de masse.*



(b) *À basse masse.*

Figure 5.13 – *Évaluations des XGBs et des DNNs par* $\Delta_{1\sigma}$.

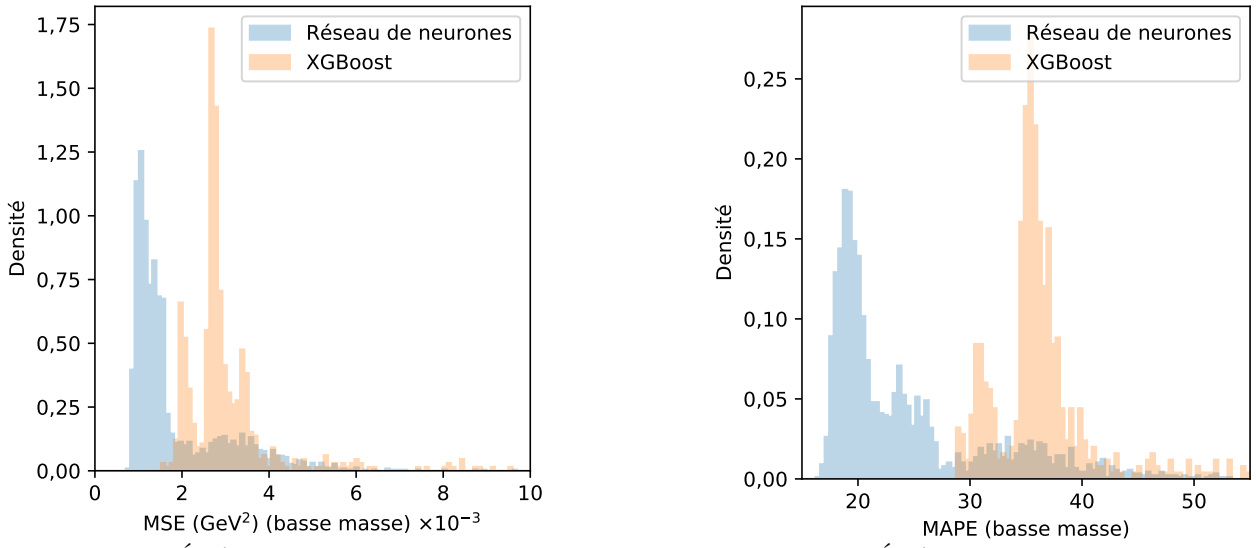
meilleures résolutions sur les masses des particules du modèle standard.

La réévaluation des modèles par $\mathsf{L}_{\mathrm{MSE}}$ et $\mathsf{L}_{\mathrm{MAPE}}$ à basse masse, en figures 5.14a et 5.14b, confirme l'obtention de meilleures performances avec les DNNs. En effet, les DNNs sont les seuls modèles avec $\mathsf{L}_{\mathrm{MSE}} < 1{,}5\times10^{3}\,\mathrm{GeV}^2$ et $\mathsf{L}_{\mathrm{MAPE}} < 28$ à basse masse. Les XGBs ont des scores de $\mathsf{L}_{\mathrm{MSE}}$ et $\mathsf{L}_{\mathrm{MAPE}}$ généralement compris entre $1{,}5\times10^{3}\,\mathrm{GeV}^2$ et $4{,}0\times10^{3}\,\mathrm{GeV}^2$ et entre 28 et 43, respectivement. Dans la suite, seuls les DNNs seront donc considérés.

6.3 Fonction de coût

Les évaluations des DNNs, regroupés d'après la fonction de coût utilisée lors de leurs entraînements, selon les métriques $\mathsf{L}_{\mathrm{MSE}}$, $\mathsf{L}_{\mathrm{MAPE}}$, $\mathsf{L}_{\mathrm{MAE}}$ sur toute la gamme de masse et $\Delta_{1\sigma}$ à basse masse sont représentées sur la figure 5.15.

L'évaluation par $\mathsf{L}_{\mathrm{MSE}}$ est représentée figure 5.15a. Les DNNs entraînés avec $\mathsf{L} = \mathsf{L}_{\mathrm{MSE}}$ y présentent un score compris entre $7{,}8\times10^{3}\,\mathrm{GeV}^2$ et $15\times10^{3}\,\mathrm{GeV}^2$, la majorité d'entre eux se trouvant en dessous de $11\times10^{3}\,\mathrm{GeV}^2$ avec un pic de leur distribution à $8{,}7\times10^{3}\,\mathrm{GeV}^2$. Les DNNs entraînés avec



(a) *Évaluation par* $\mathsf{L}_{\mathrm{MSE}}$. **(b)** *Évaluation par* $\mathsf{L}_{\mathrm{MAPE}}$.

Figure 5.14 – *Évaluations des XGBs et des DNNs par* $\mathsf{L}_{\mathrm{MSE}}$ *et* $\mathsf{L}_{\mathrm{MAPE}}$ *à basse masse.*

$\mathsf{L} = \mathsf{L}_{\mathrm{MAE}}$ se situent majoritairement entre $7{,}7\times10^3\,\mathrm{GeV}^2$ et $10\times10^3\,\mathrm{GeV}^2$, avec un pic de leur distribution à $8{,}3\times10^3\,\mathrm{GeV}^2$. Les DNNs entraînés avec $\mathsf{L} = \mathsf{L}_{\mathrm{MAPE}}$ se répartissent en deux groupes, le premier entre $7{,}9\times10^3\,\mathrm{GeV}^2$ et $10\times10^3\,\mathrm{GeV}^2$, le second entre $13\times10^3\,\mathrm{GeV}^2$ et $16\times10^3\,\mathrm{GeV}^2$. La fonction de coût $\mathsf{L}_{\mathrm{MAE}}$ semble ainsi préférable à $\mathsf{L}_{\mathrm{MSE}}$ lorsque la comparaison se fait sur $\mathsf{L}_{\mathrm{MSE}}$ elle-même. Il est en revanche plus difficile de conclure quant à $\mathsf{L}_{\mathrm{MAPE}}$.

L'évaluation par $\mathsf{L}_{\mathrm{MAPE}}$, figure 5.15b, montre également un avantage de $\mathsf{L}_{\mathrm{MAE}}$ sur $\mathsf{L}_{\mathrm{MSE}}$. En effet, les modèles entraînés avec $\mathsf{L} = \mathsf{L}_{\mathrm{MAE}}$ se situent majoritairement à $\mathsf{L}_{\mathrm{MAPE}} < 21$ alors que ceux entraînés avec $\mathsf{L} = \mathsf{L}_{\mathrm{MSE}}$ sont plutôt dans la région $\mathsf{L}_{\mathrm{MAPE}} > 20$. Les valeurs les plus basses sont obtenues sur les modèles entraînés avec $\mathsf{L}_{\mathrm{MAPE}}$. Or, l'évaluation est basée sur $\mathsf{L}_{\mathrm{MAPE}}$ elle-même, il n'est donc pas équitable de se baser uniquement sur la figure 5.15b pour affirmer que $\mathsf{L}_{\mathrm{MAPE}}$ peut être préférable à $\mathsf{L}_{\mathrm{MAE}}$ ou $\mathsf{L}_{\mathrm{MSE}}$.

La figure 5.15c représente l'évaluation des DNNs par $\mathsf{L}_{\mathrm{MAE}}$. La distribution obtenue avec les DNNs entraînés avec $\mathsf{L} = \mathsf{L}_{\mathrm{MSE}}$ s'étend de 65 GeV à près de 10 GeV avec un pic à 71 GeV. En revanche, de nombreux modèles entraînés avec $\mathsf{L}_{\mathrm{MAE}}$ ou $\mathsf{L}_{\mathrm{MAPE}}$ se situent à 67 ± 4 GeV.

Enfin, sur la figure 5.15d se trouvent les distributions de $\Delta_{1\sigma}$ à basse masse pour ces trois groupes de DNNs. Les modèles utilisant $\mathsf{L}_{\mathrm{MSE}}$ ont tous un score supérieur à 0,5. Ceux entraînés avec $\mathsf{L}_{\mathrm{MAE}}$ se situent entre 0,4 et 0,6. Les modèles basés sur $\mathsf{L}_{\mathrm{MAPE}}$ forment encore deux groupes, le premier entre 0,34 et 0,5, le second entre 0,5 et 0,64. Les fonctions de coût $\mathsf{L}_{\mathrm{MAPE}}$ et $\mathsf{L}_{\mathrm{MAE}}$ permettent donc d'obtenir des modèles avec une meilleure résolution à basse masse que $\mathsf{L}_{\mathrm{MSE}}$.

Les modèles entraînés avec $\mathsf{L} = \mathsf{L}_{\mathrm{MAE}}$ ou $\mathsf{L} = \mathsf{L}_{\mathrm{MAPE}}$ proposent ainsi de meilleurs scores que ceux obtenus avec $\mathsf{L} = \mathsf{L}_{\mathrm{MSE}}$, quelle que soit la métrique d'évaluation utilisée. Lors des évaluations avec $\mathsf{L}_{\mathrm{MSE}}$ ou $\mathsf{L}_{\mathrm{MAE}}$, aucun avantage net n'est visible entre $\mathsf{L} = \mathsf{L}_{\mathrm{MAE}}$ et $\mathsf{L} = \mathsf{L}_{\mathrm{MAPE}}$. En revanche, les métriques $\mathsf{L}_{\mathrm{MAPE}}$ et $\Delta_{1\sigma}$ montrent que certains modèles entraînés avec $\mathsf{L} = \mathsf{L}_{\mathrm{MAPE}}$ donnent de meilleurs résultats. La sélection d'un modèle est donc poursuivie parmi ceux ayant comme fonction de coût $\mathsf{L}_{\mathrm{MAPE}}$.

6.4 Algorithme d'optimisation

Les algorithmes d'optimisation sont présentés dans la section 5.3.2. L'algorithme SGD ne permet pas aux modèles de converger, il est donc exclu de nos investigations. Deux algorithmes sont comparés, AdaDelta et Adam.

Les évaluations des DNNs précédemment sélectionnés, regroupés d'après l'algorithme d'optimisation utilisé lors de leurs entraînements, selon les métriques $\mathsf{L}_{\mathrm{MAPE}}$ sur toute la gamme de masse et $\Delta_{1\sigma}$ à basse masse sont représentées sur la figure 5.16. Les deux groupes observés dans les sections

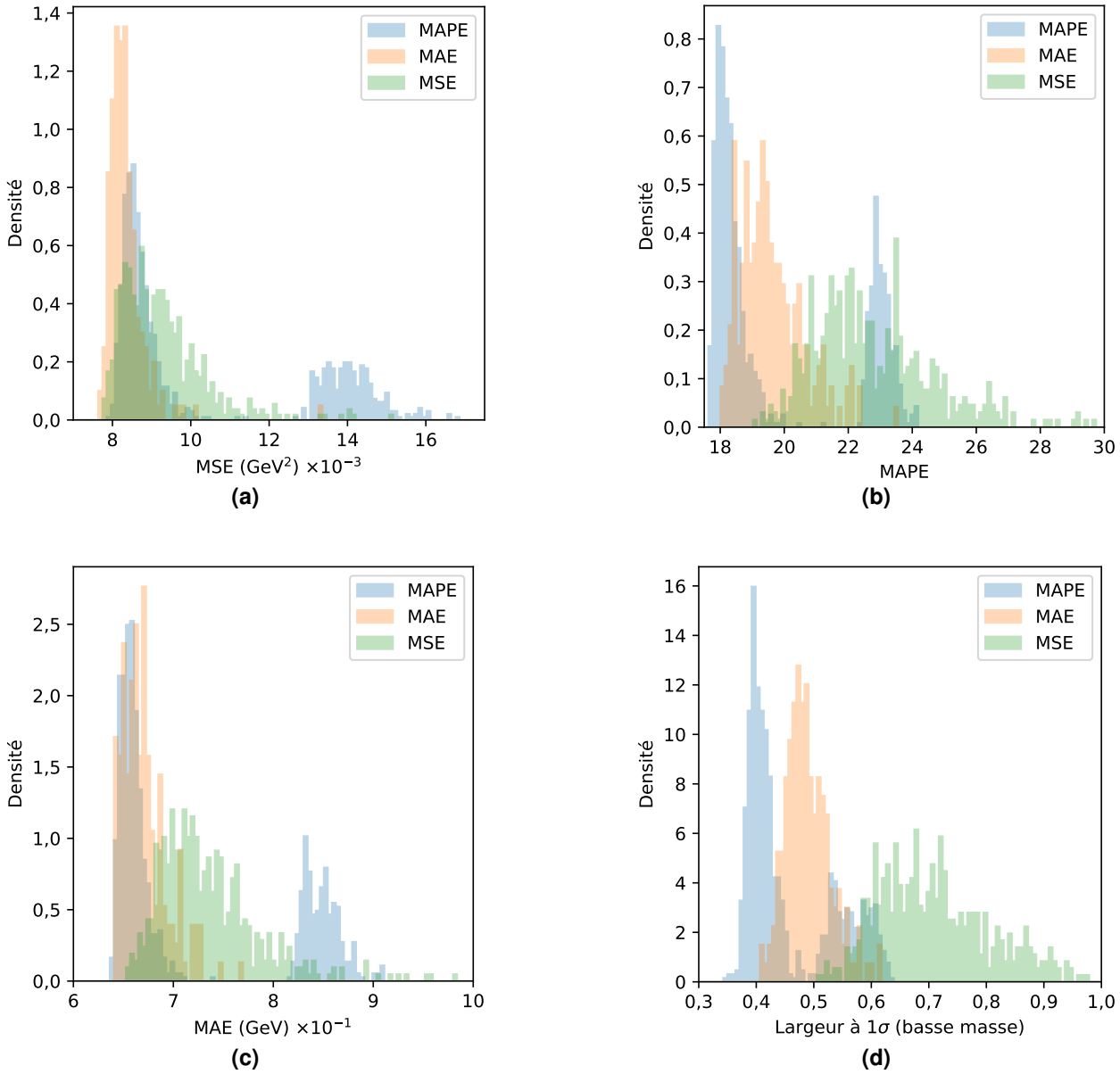


Figure 5.15 – *Évaluations des DNNs regroupés selon la fonction de coût par* $\mathsf{L}_{\mathrm{MSE}}$, $\mathsf{L}_{\mathrm{MAPE}}$, $\mathsf{L}_{\mathrm{MAE}}$ *et* $\Delta_{1\sigma}$.

précédentes sont identifiés comme étant les modèles entraînés respectivement par Adam et AdaDelta. Dans le cadre de nos travaux, lors de la recherche de la combinaison optimale d'hyper-paramètres, nous avons initialement utilisé Adam jusqu'à sélectionner le jeu de variables d'entrée (section 6.1) et la fonction de coût (section 6.3) à utiliser. C'est pourquoi ces deux groupes liés à Adam et AdaDelta n'apparaissent que dans certaines sélections de modèles.

Sur la figure 5.16a, les modèles optimisés par Adam présentent un score de $\mathsf{L}_{\mathrm{MAPE}}$ entre 17,5 et 20 alors que ceux optimisés par AdaDelta se situent entre 22,2 et 24,3. L'optimisation par Adam semble donc meilleure que celle par AdaDelta. L'évaluation à basse masse par $\Delta_{1\sigma}$ sur la figure 5.16b confirme cette observation. Les modèles optimisés par Adam se situent en effet entre 0,35 et 0,48, ceux optimisés par AdaDelta entre 0,49 et 0,64. L'algorithme d'optimisation Adam donne donc de meilleurs modèles qu'AdaDelta avec les hyper-paramètres choisis précédemment.

6.5 Autres hyper-paramètres

Les hyper-paramètres restant à fixer et leurs valeurs testées sont :

— le nombre de couches cachées N_{cc}, 2 à 5 ;

— le nombre de neurones par couche cachée $N_{\mathrm{n/c}}$, 200 à 2000 par pas de 100 ;

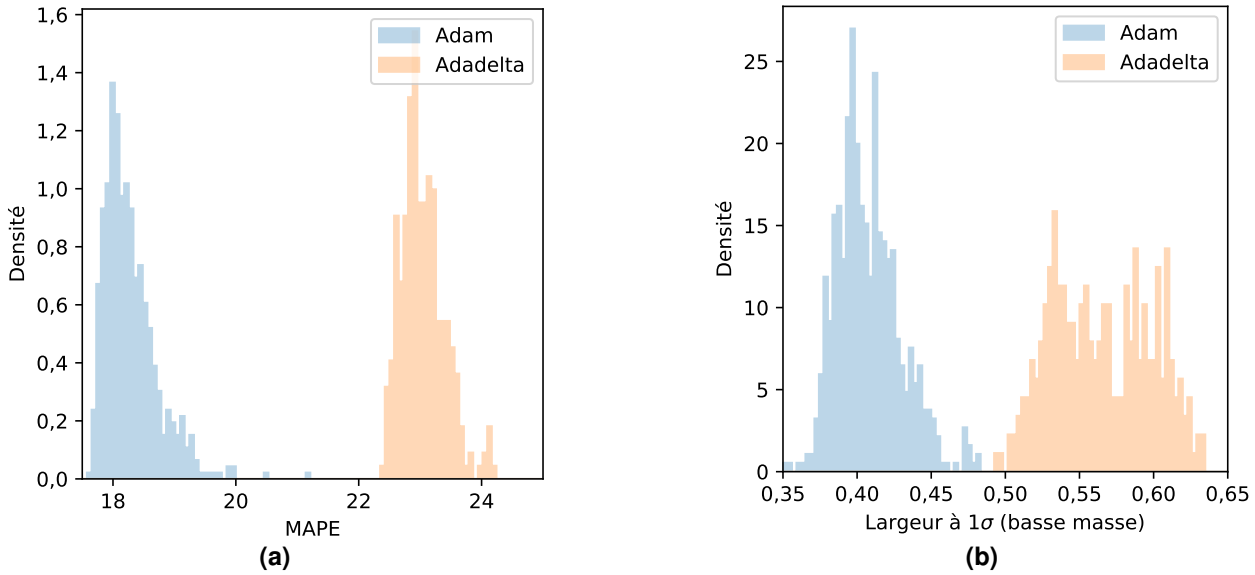


Figure 5.16 – *Évaluations des DNNs regroupés selon l'algorithme d'optimisation par* L_{MAPE} *et* $\Delta_{1\sigma}$.

— le mode d'initialisation des poids (WI), uniforme (u), normale (n), Glorot uniforme (gu), Glorot normale (gn) ;
— la fonction d'activation (FA) des neurones des couches cachées, ReLU, SELU, ELU, Softplus.

Les évaluations à basse, moyenne et haute masse des DNNs utilisant les 27 variables d'entrée et entraînés par Adam avec $\mathsf{L} = \mathsf{L}_{\text{MAPE}}$, regroupés par N_{cc}, $N_{\text{n/c}}$, mode d'initialisation des poids et fonction d'activation sont données sur les figures 5.17, 5.18, 5.19 et 5.20 respectivement. Les distributions des scores pour les métriques d'évaluation y étant présentées sont semblables pour tous les groupes de modèles formés par une valeur fixée d'un de ces hyper-paramètres. La méthode employée jusqu'ici ne permet donc pas de conclure sur le choix d'une valeur pour un hyper-paramètre.

Nous avons alors choisi de sélectionner un modèle parmi ceux restant à ce stade à l'aide d'une procédure itérative rejetant tout modèle avec un score élevé selon une des métriques d'évaluation jusqu'à ce qu'il n'en reste que 10 au plus. Cette procédure est la suivante :

1. Déterminer la valeur maximale $x_{\text{max}}^{\text{métrique } m}$, sur l'ensemble des modèles sélectionnés, de chacune des métriques d'évaluation m utilisées. La valeur maximale autorisée $x_{\text{OK}}^{\text{métrique } m}$ pour la métrique m est initialement fixée à $x_{\text{max}}^{\text{métrique } m}$;
2. Fixer la valeur maximale autorisée à 99 % de sa valeur actuelle pour chacune des métriques m ;
3. Rejeter tout modèle dont une des métriques donne un score supérieur à $x_{\text{OK}}^{\text{métrique } m}$;
4. Reprendre à l'étape 2 si plus de 10 modèles sont encore sélectionnés.

Les modèles ainsi sélectionnés, au nombre de 7, sont listés dans le tableau 5.1 sans ordre particulier. Leurs réponses dont données sur les figures 5.21 et 5.22.

La même procédure appliquée à tous les modèles entraînés mène à une liste ne contenant que des DNNs, utilisant majoritairement les 27 variables d'entrée, tous entraînés par Adam avec comme fonction de coût L_{MAPE}, ce qui confirme que nos choix d'hyper-paramètres précédents sont pertinents.

Aucun modèle avec $N_{\text{cc}} \in \{2, 5\}$ n'est sélectionné. Pour 4 modèles, $N_{\text{cc}} = 3$. Le nombre de neurones par couche cachée est de 1000 pour 4 modèles sur 7, dont 3 sur les 4 avec $N_{\text{cc}} = 3$. Le WI le plus représenté est Glorot uniforme (5/7). Les FA sont disparates, chacune apparaissant une ou deux fois dans la sélection.

Chacun de ces modèles présente une réponse proche de 1 entre 70 et 400 GeV avec une résolution relative de l'ordre de 22 % à basse masse et 10 % à haute masse. Le modèle F conserve une réponse proche de 1 jusqu'à environ 500 GeV, cependant sa résolution à basse masse est légèrement dégradée par rapport aux autres modèles. Le modèle B présente l'avantage d'avoir des hyper-paramètres « consensus », c'est-à-dire que chacune des valeurs de ses hyper-paramètres correspond à la valeur la plus représentée dans la sélection. C'est à partir de ce modèle que nous avons choisi de continuer

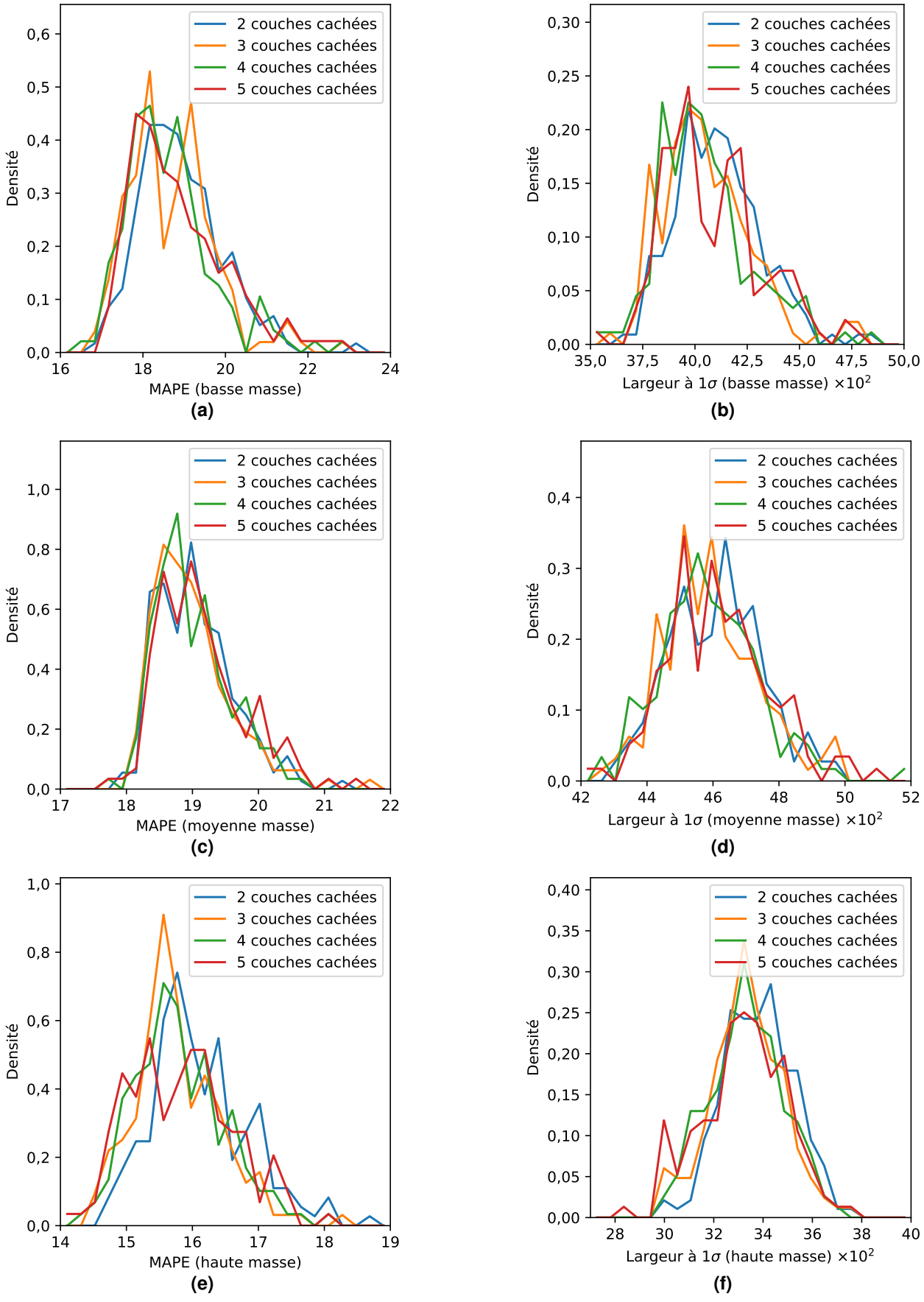


Figure 5.17 – *Évaluations des DNNs regroupés selon N_{cc} par* $\mathrm{L}_{\mathrm{MAPE}}$ *et* $\Delta_{1\sigma}$.

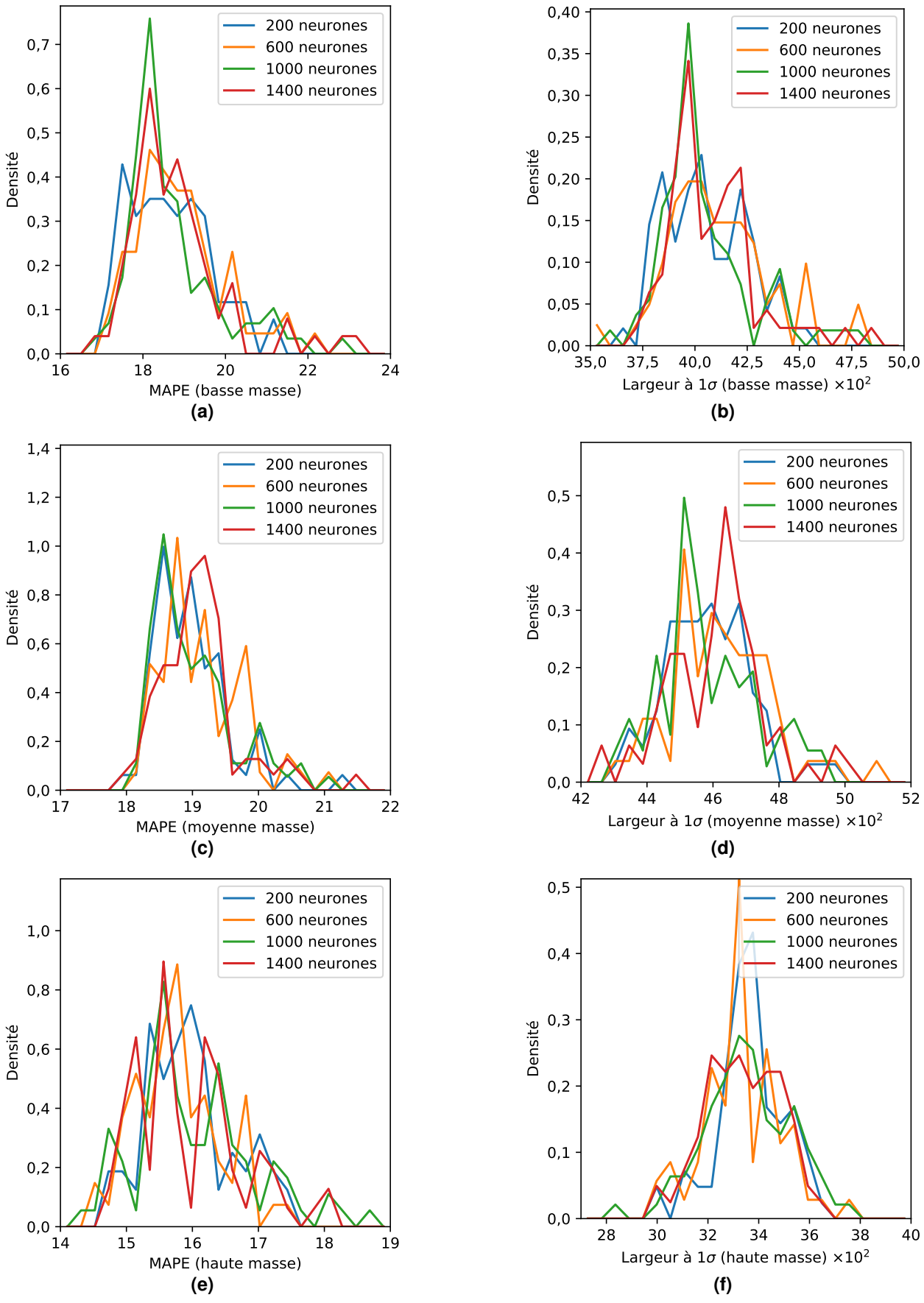


Figure 5.18 – *Évaluations des DNNs regroupés selon N_{nt} par* $\mathrm{L}_{\mathrm{MAPE}}$ *et* $\Delta_{1\sigma}$.

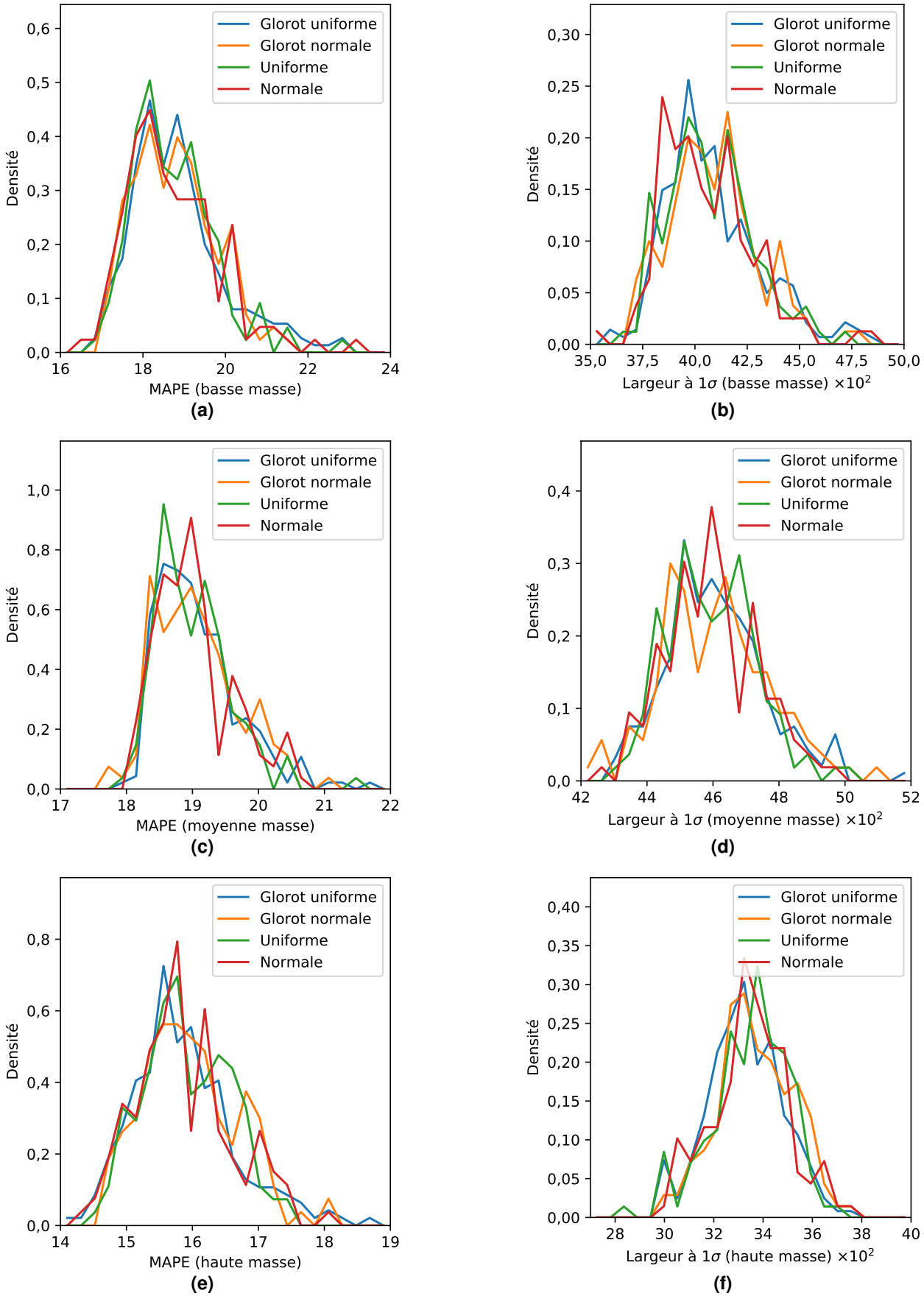


Figure 5.19 – *Évaluations des DNNs regroupés selon le mode d'initialisation des poids par* $\mathrm{L}_{\mathrm{MAPE}}$ *et* $\Delta_{1\sigma}$.

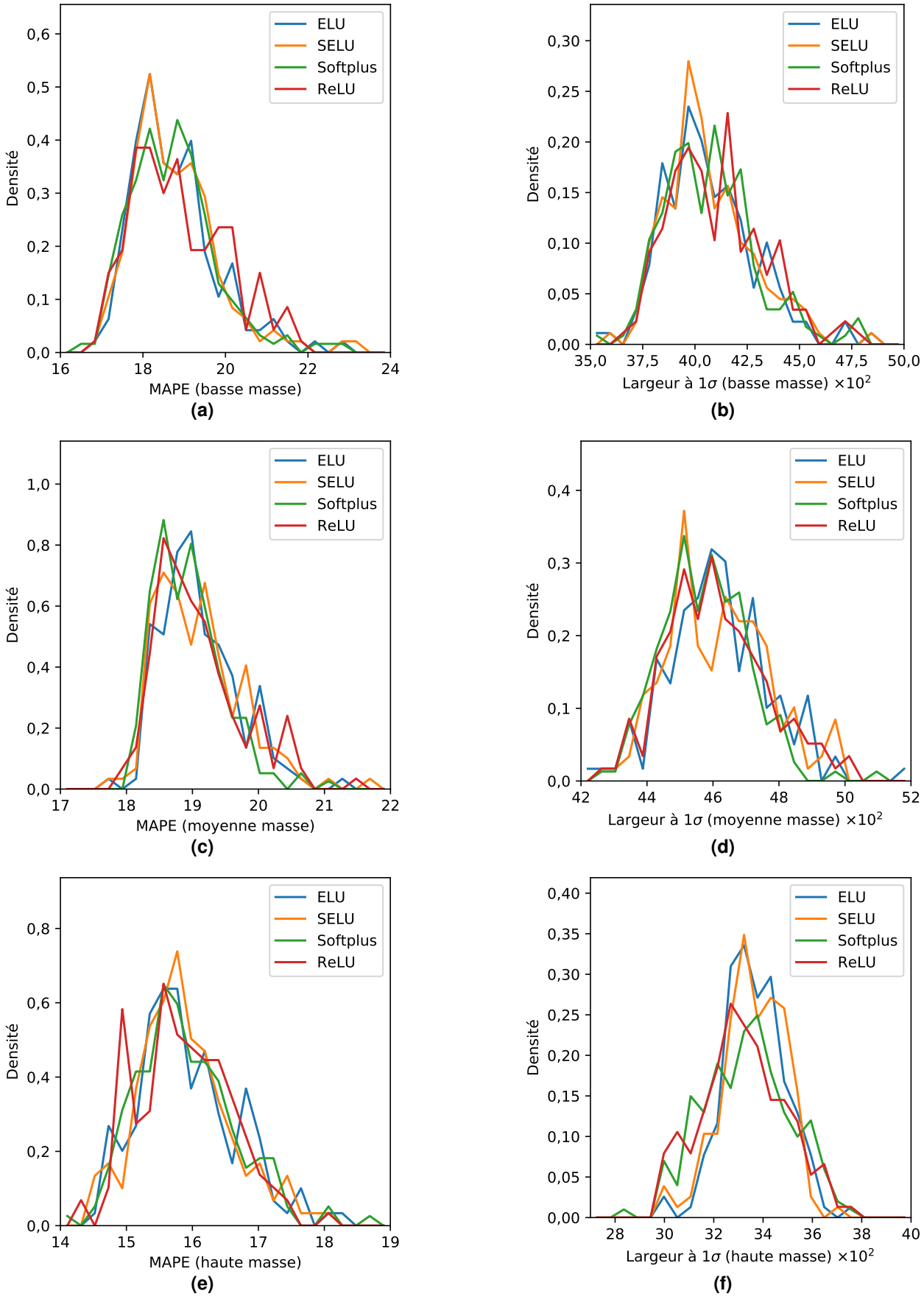


Figure 5.20 – *Évaluations des DNNs regroupés selon la fonction d'activation par* $\mathrm{L}_{\mathrm{MAPE}}$ *et* $\Delta_{1\sigma}$.

Modèle	N_{cc}	$N_{n/c}$	WI	FA
A	3	1000	gu	ELU
B	3	1000	gu	Softplus
C	3	1000	n	SELU
D	3	1400	gu	ReLU
E	4	200	gn	ReLU
F	4	1000	gu	ELU
G	4	1400	gu	Softplus

Tableau 5.1 – *Liste des 7 modèles sélectionnés.*

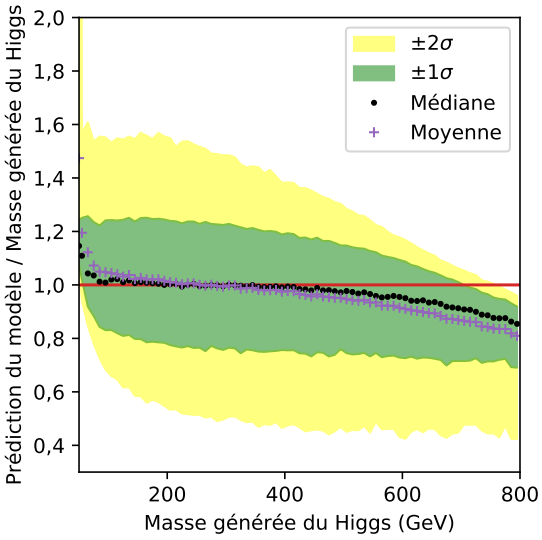


Figure 5.21 – *Réponse du modèle A.*

notre étude. Les hyper-paramètres sélectionnés sont donnés dans le tableau 5.2 avec une comparaison à ceux utilisés par Bärtschi & coll. [257].

Hyper-paramètre	Notre DNN	DNN de Bärtschi & coll. [257]
Nombre de couches cachées N_{cc}	3	4
Neurones par couche cachée $N_{n/c}$	1000	200
Fonction d'activation	Softplus	ReLU
Algorithme d'optimisation	Adam	Adam
Fonction de coût	L_{MAPE}	L_{MSE}
Initialisation des poids	« Glorot Uniforme » [272]	?
Nombre d'entrées	27 (voir section 2.4)	17 (voir [257])

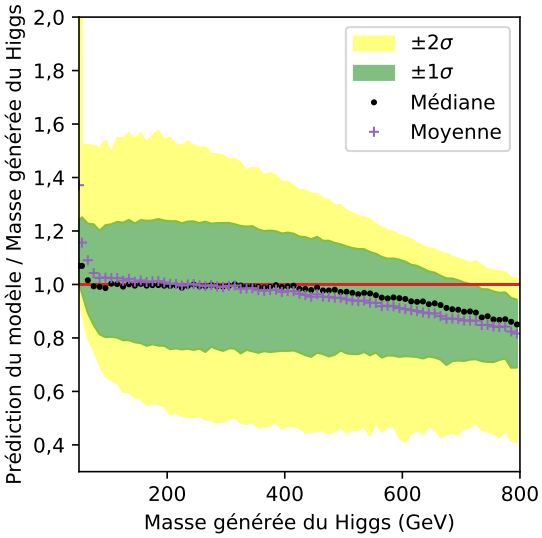
Tableau 5.2 – *Comparaison de nos hyper-paramètres à ceux de Bärtschi & coll. Le mode d'initialisation des poids utilisé par Bärtschi & coll. n'est pas donné dans leur article.*

7 Discussions

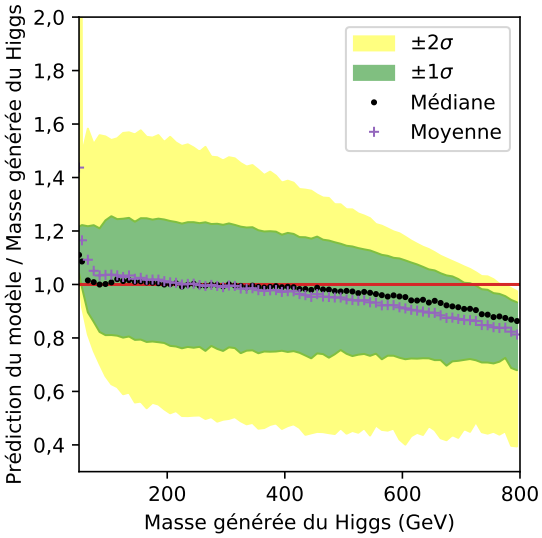
Lors de cette thèse, en parallèle de l'exploration des hyper-paramètres présentée en section 6, les effets de différents facteurs ont été étudiés. À des fins de cohérence dans la comparaison des effets, nous utilisons ici le modèle B sélectionné précédemment comme référence.

7.1 Effet de l'empilement

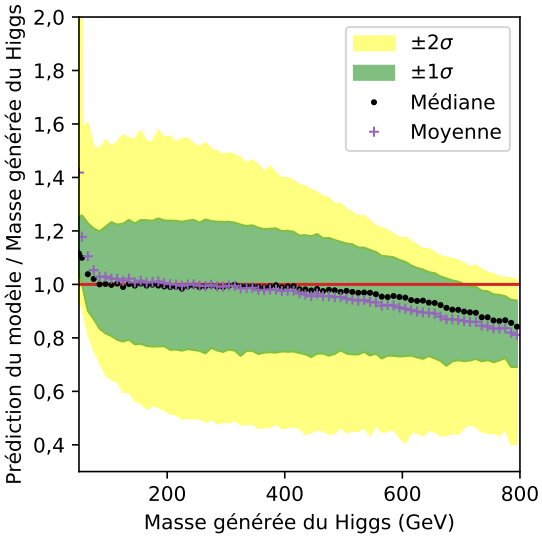
Dans les travaux de Bärtschi & coll. [257], l'empilement (PU, *Pile-Up*) n'est pas considéré. Nous avons donc souhaité déterminer son effet sur les prédictions de notre modèle. Pour cela, les mêmes événements que ceux décrits en section 2 ont été générés sans empilement. Un DNN, noté B^{0PU}, est entraîné sur ces événements sans empilement. Les hyper-paramètres de B^{0PU} sont identiques à ceux de B, à l'exception des variables d'entrée auxquelles N_{PU} est retiré, car $N_{PU} = 0$ pour tous les événements sans empilement. La réponse de B^{0PU} sur ces événements est représentée sur la figure 5.23a. Sa moyenne est de $1{,}00 \pm 0{,}05$ pour $m_{\mathcal{H}}$ entre 70 GeV et 600 GeV avec une résolution relative de l'ordre de 22 % à basse masse et 10 % à haute masse. Ces performances sont donc similaires à celles de B sur les événements avec empilement.



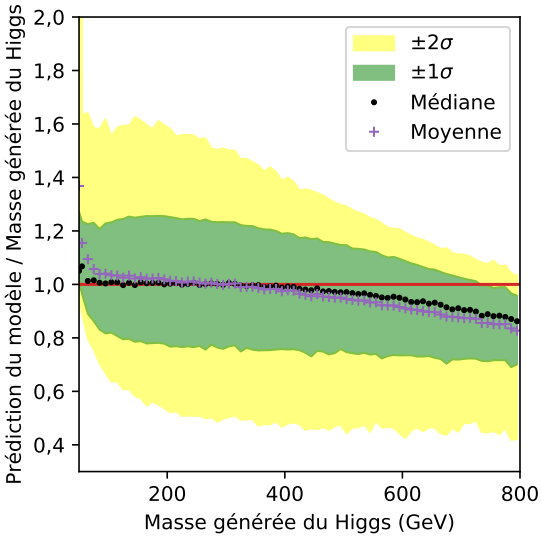
(a) *Réponse du modèle B.*



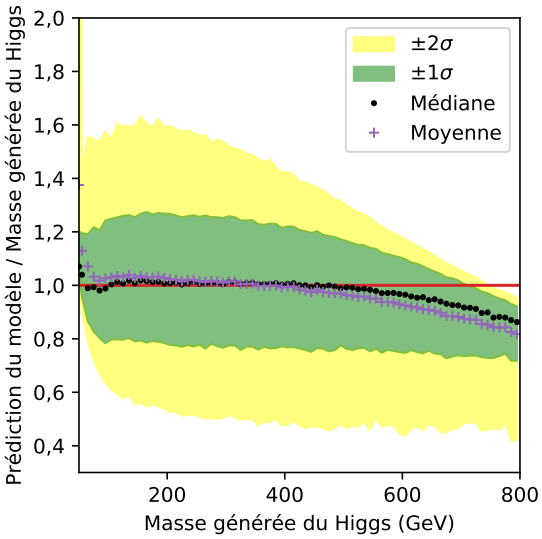
(b) *Réponse du modèle C.*



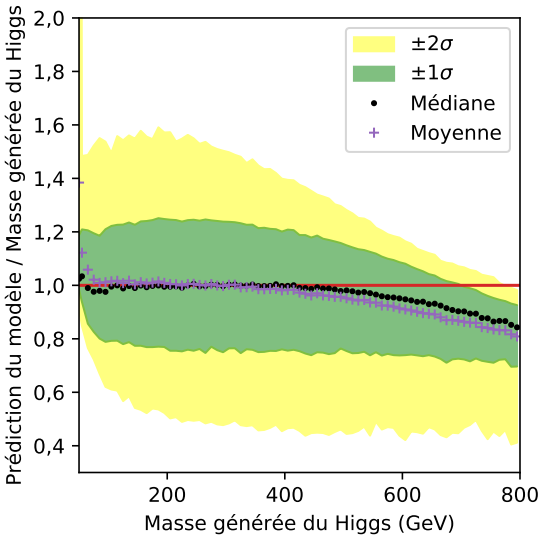
(c) *Réponse du modèle D.*



(d) *Réponse du modèle E.*

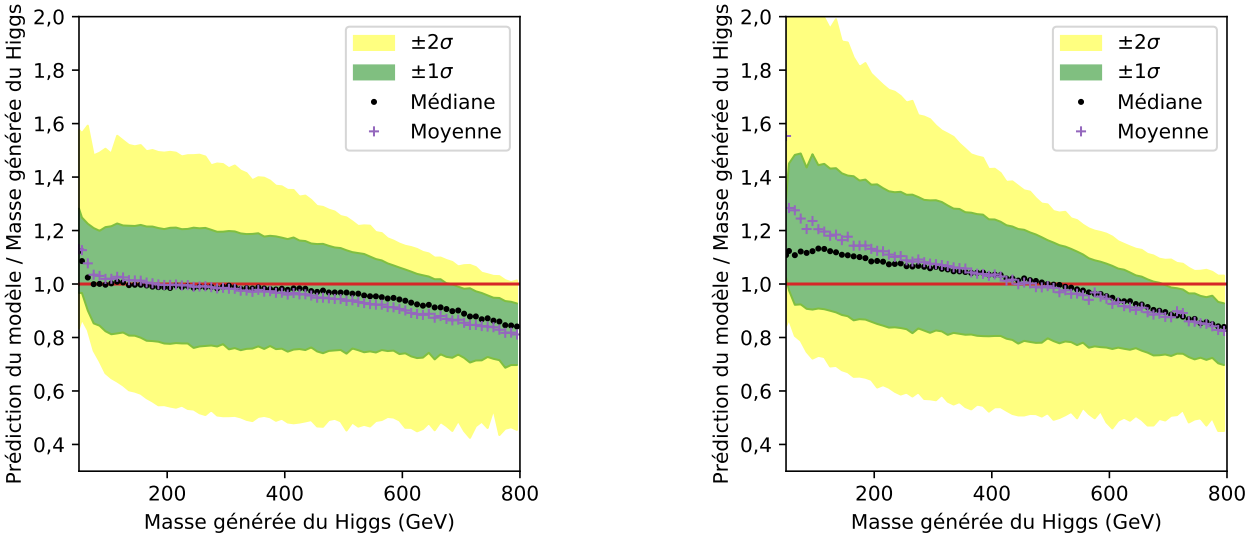


(e) *Réponse du modèle F.*



(f) *Réponse du modèle G.*

Figure 5.22 – *Réponse des modèles B à G.*



(a) *Sans empilement.* **(b)** *Avec empilement.*

Figure 5.23 – *Réponses du modèle B^{0PU} sur les événements sans et avec empilement.*

Cependant, la réponse de B^{0PU} est dégradée sur des événements contenant de l'empilement, figure 5.23b. La réponse médiane se situe en effet à 1,13 à $m_{\mathcal{H}} = 100\,\text{GeV}$ et diminue à 0,83 à $m_{\mathcal{H}} = 800\,\text{GeV}$ avec empilement contre 1,00 et 0,83 sans empilement. La résolution relative à basse masse est de l'ordre de 25 % avec empilement.

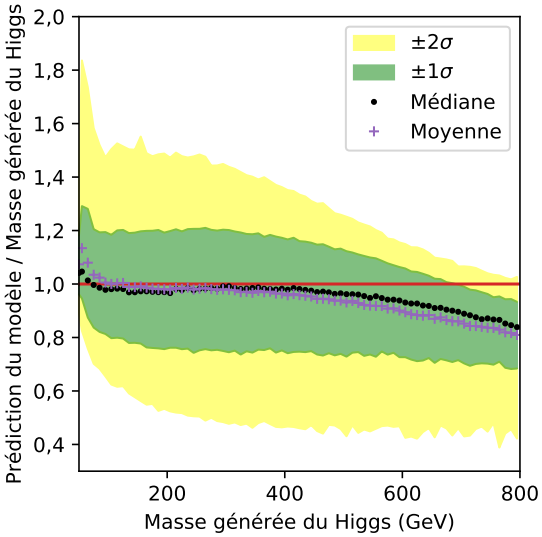
L'empilement peut être considéré comme un bruit blanc, dont l'énergie moyenne est reliée à N_{PU}. L'énergie portée par L_1 et L_2, les éléments visibles de la désintégration de $\mathcal{H}$, est en revanche reliée à $m_{\mathcal{H}}$. À haute masse, elle est grande par rapport à celle de l'empilement. Les prédictions de B^{0PU} ne sont alors pas perturbées, menant à des performances similaires à celles du cas sans empilement. Lorsque $m_{\mathcal{H}}$ diminue, l'énergie disponible pour L_1 et L_2 est moindre et le bruit d'empilement devient significatif. Le modèle B^{0PU} n'est pas entraîné pour traiter ce bruit et ses prédictions sont perturbées. Il est donc primordial d'inclure l'empilement dans l'entraînement dans l'optique d'une utilisation de nos modèles dans les analyses de CMS.

La réponse du modèle B dans le cas d'événements sans empilement, figure 5.24a, peut être comparée à celle sur les événements en contenant, figure 5.24b (identique à 5.22a). Le profil d'empilement utilisé pour générer les événements d'entraînement est celui de l'année 2017. Il apparaît que le modèle B est peu sensible au retrait de l'empilement, les réponses étant similaires sur les figures 5.24a et 5.24b. L'utilisation de B sur des événements dont le profil d'empilement est légèrement différent de celui de l'année 2017, comme c'est le cas pour les autres années du Run II (2016, 2018), est ainsi directement envisageable.

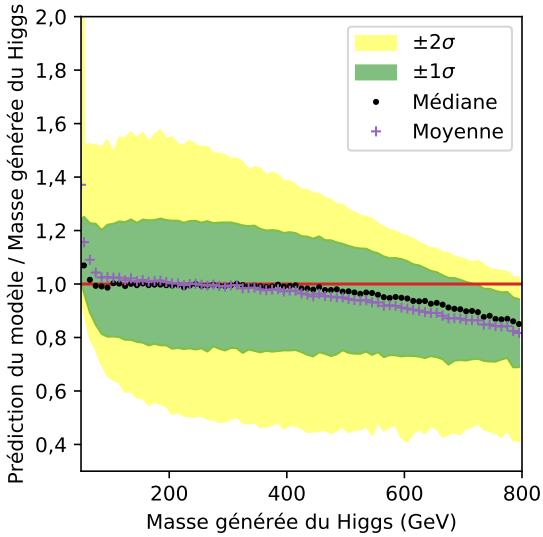
7.2 Effet de la reconstruction des particules

La reconstruction des particules est présentée dans le chapitre 2. Son effet peut être caractérisé par l'étude du modèle B^{gen}, ayant les mêmes hyper-paramètres que B, mais entraîné en utilisant les objets générés au lieu de ceux reconstruits pour L_1, L_2 et ${E_{\text{T}}}^{\text{miss}}$, c'est-à-dire pour les trois objets physiques liés à la désintégration des leptons τ (deux parties visibles L_1 et L_2 et ${E_{\text{T}}}^{\text{miss}}$ pour les neutrinos). En particulier, les valeurs de $\vec{p}_{\text{T}}{}^{L_1}$, $\vec{p}_{\text{T}}{}^{L_2}$ et $\vec{E}_{\text{T}}^{\text{miss}}$ correspondent exactement à la réalité. Il s'agit donc du cas dans lequel les objets physiques issus de $\mathcal{H}$ sont parfaitement reconstruits. Toutes les autres variables, y compris la matrice de covariance de ${E_{\text{T}}}^{\text{miss}}$, restent celles obtenues avec les objets reconstruits.

La figure 5.25 montre les réponses du modèle B^{gen} sur les événements avec reconstruction parfaite et réelle. Dans le cas d'une reconstruction parfaite, la réponse médiane de B^{gen} est de l'ordre de $1{,}01 \pm 0{,}02$ de 70 à 800 GeV. La résolution relative est quant à elle de l'ordre de 3 %, soit près de sept fois mieux que B.

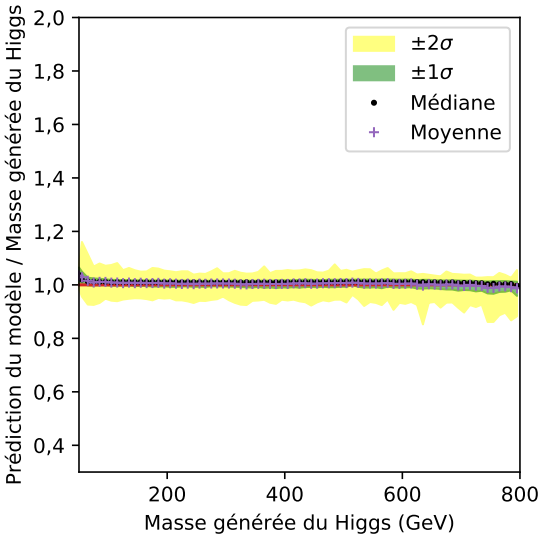


(a) *Sans empilement.*

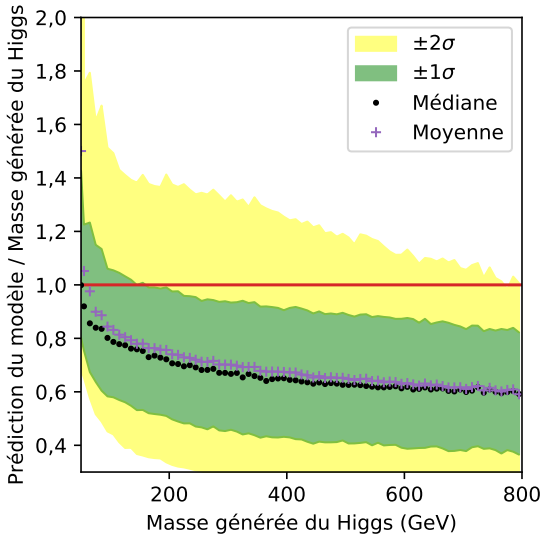


(b) *Avec empilement.*

Figure 5.24 – *Réponses du modèle B sur les événements sans et avec empilement.*



(a) *Reconstruction parfaite.*



(b) *Reconstruction réelle.*

Figure 5.25 – *Réponses du modèle B^{gen} dans le cas d'une reconstruction des particules parfaite et réelle.*

Les DNNs sont donc en mesure de comprendre la physique des événements $\mathcal{H} \to \tau\tau$ afin d'estimer $m_{\mathcal{H}}$ à partir des objets physiques générés correspondant aux objets effectivement reconstruits par le détecteur. Cependant, comme le montre la figure 5.25b, l'utilisation de B^{gen} sur les variables reconstruites, effectivement accessibles expérimentalement, ne permet pas d'obtenir $m_{\mathcal{H}}$. En effet, la réponse moyenne de B^{gen} avec ces variables est inférieure à 1 et de l'ordre de 0,7 à haute masse. De plus, la résolution relative est de l'ordre de 40 %. Une des tâches des DNNs est donc de corriger cet effet de reconstruction.

7.3 Effet des faux taus hadroniques

La phénoménologie des événements contenant une paire de leptons τ est décrite dans le chapitre 1. Ces leptons peuvent se désintégrer hadroniquement en tau hadronique (τ_h) ou leptoniquement en électron (e) ou en muon (μ). Il existe ainsi six canaux différents dans les événements avec une paire de leptons τ, pouvant être répartis en trois groupes :

— complètement hadronique : $\tau_h\tau_h$, avec deux τ_h ;

— semi-leptoniques : $\mu\tau_h$ et $e\tau_h$, ou simplement $\ell\tau_h$, avec un τ_h ;
— leptoniques : $\mu\mu$, $e\mu$ et ee, ou simplement $\ell\ell$, sans τ_h.

Les faux taus hadroniques (faux τ_h) sont des objets physiques tels que des électrons, des muons et surtout des jets identifiés à tort comme des τ_h. Ils représentent près de 70 % des événements dans le canal $\tau_h\tau_h$, 38 % dans le canal $\mu\tau_h$ et 68 % dans le canal $e\tau_h$. Les faux τ_h sont particulièrement difficiles à modéliser dans les simulations [224, 225].

L'identification des τ_h est réalisée dans nos travaux à l'aide de l'algorithme DEEPTAU [176], qui présente un faible taux de mauvaise identification des τ_h, inférieur à 1 %. Cependant, une autre méthode d'identification des τ_h, basée sur un arbre de décision (BDT), peut être utilisée et présente un taux de mauvaise identification de jets en tant que τ_h pouvant atteindre 4 % [173]. Une sélection plus riche en faux τ_h est ainsi obtenue.

Les réponses du modèle B sur chacun des trois groupes de canaux (hadronique, semi-leptoniques et leptoniques) sont représentées figure 5.26 pour les deux ensembles de sélection des τ_h. Quel que soit le groupe d'état final, les réponses pour $m_{\mathcal{H}} > 600\,\text{GeV}$ ne sont pas affectées par la sélection des τ_h. En effet, pour de hautes valeurs de $m_{\mathcal{H}}$, les τ_h ont des impulsions suffisamment élevées pour être correctement sélectionnés par la séquence d'analyse. À basse masse en revanche, les impulsions des faux τ_h sont compétitives vis-à-vis de celles des vrais τ_h. La séquence d'analyse forme alors des dileptons contenant des faux τ_h. La réponse du modèle s'en retrouve modifiée, jusqu'à 20 % de plus pour des masses entre 100 GeV et 600 GeV. L'effet le plus important se situe à très basse masse où la résolution est fortement dégradée.

La figure 5.27 montre la différence $y_{\text{préd}} - y_{\text{vraie}}$ entre les prédictions du modèle B et la valeur vraie de $m_{\mathcal{H}}$ pour des valeurs de $m_{\mathcal{H}}$ entre 50 et 200 GeV sur chacun des trois groupes de canaux et pour les deux ensembles de sélection des τ_h. Dans les canaux leptoniques ($\ell\ell$), figures 5.27e et 5.27f, l'effet de la sélection des τ_h est moindre que dans les autres canaux. Il n'y a en effet aucun τ_h dans le dilepton, seule la sélection des événements est modifiée. Un objet physique identifié comme un τ_h par le BDT et non par DEEPTAU peut en effet faire basculer l'événement d'un canal à l'autre, si le τ_h identifié par le BDT permet de construire un dilepton.

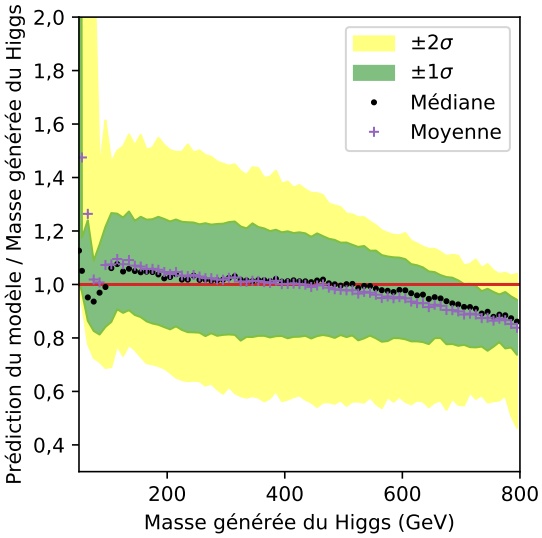
Dans le cas des canaux semi-leptoniques ($\ell\tau_h$), la différence entre $y_{\text{préd}}$ de B et y_{vraie} à basse masse est en moyenne inférieure à 10 GeV pour une sélection des τ_h par DEEPTAU, figure 5.27c. La résolution relative est quant à elle inférieure à 25 %. Lorsque les τ_h sont identifiés par le BDT, figure 5.27d, le modèle surestime $m_{\mathcal{H}}$ de 25 GeV en moyenne pour $70\,\text{GeV} < m_{\mathcal{H}} < 200\,\text{GeV}$ et de près de 40 GeV à $m_{\mathcal{H}} = 50\,\text{GeV}$. La résolution relative est de l'ordre de 25 % au-dessus de 70 GeV, moins bonne qu'avec DEEPTAU, et augmente drastiquement pour des masses plus basses, ce qui n'est pas le cas avec DEEPTAU. Il s'agit donc de la contribution des faux τ_h.

Dans le canal $\tau_h\tau_h$, figures 5.27a et 5.27b, un effet similaire existe. La résolution relative est toujours de l'ordre de 22 % au-delà de 100 GeV, mais la présence des faux τ_h mène à une surestimation moyenne de 30 GeV pour $m_{\mathcal{H}} > 110\,\text{GeV}$ et pouvant aller jusqu'à 100 GeV pour $m_{\mathcal{H}} \simeq 50\,\text{GeV}$, soit une erreur de 200 %. La dégradation de la résolution à très basse masse s'étend jusqu'à 100 GeV, au lieu de 70 GeV pour les canaux $\ell\tau_h$. L'effet des faux τ_h est donc plus important que dans les autres canaux, ce qui s'explique facilement par la présence de deux τ_h. Pour $m_{\mathcal{H}} = 50\,\text{GeV}$, la résolution de B sur les événements avec DEEPTAU est également mauvaise. La sélection des τ_h se fait avec $p_T > 40\,\text{GeV}$, ce qui est difficile à obtenir pour $m_{\mathcal{H}} = 50\,\text{GeV}$. Ces événements sont donc peu nombreux et vraisemblablement très contaminés par les faux τ_h.

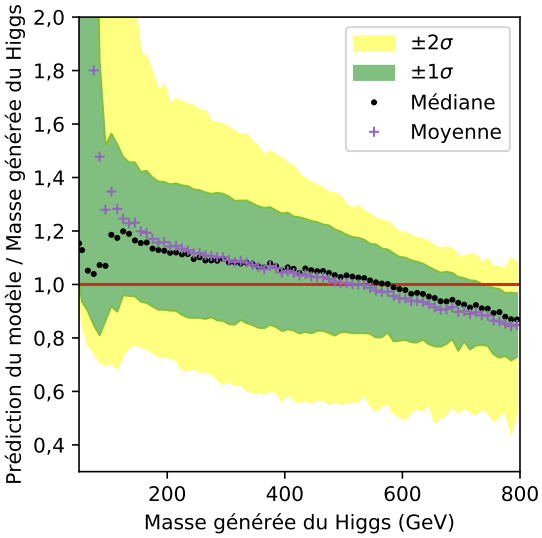
Les faux τ_h introduisent donc un biais important sur une large gamme de masse et en particulier dans la région des bosons Z ($m_Z = 91{,}2\,\text{GeV}$) et h ($m_h = 125{,}1\,\text{GeV}$). L'inclusion des faux τ_h dans l'entraînement est non triviale, car la masse à prédire n'est pas définie, les faux τ_h n'étant pas des objets physiques provenant de $\mathcal{H}$.

7.4 Effet de la séparation des canaux

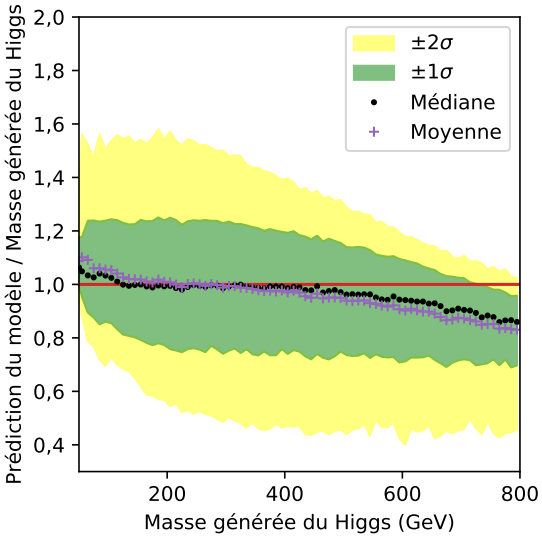
Les modèles construits sont entraînés et testés sur l'ensemble des événements, sans sélection sur le canal. Or, il est possible d'entraîner un DNN par canal afin de le spécialiser à la phénoménologie associée et obtenir, potentiellement, de meilleures estimations de $m_{\mathcal{H}}$.



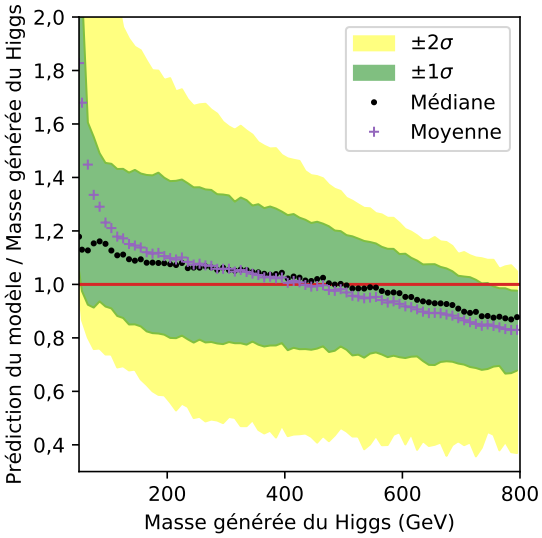
(a) *Canal* $\tau_h\tau_h$, DEEPTAU.



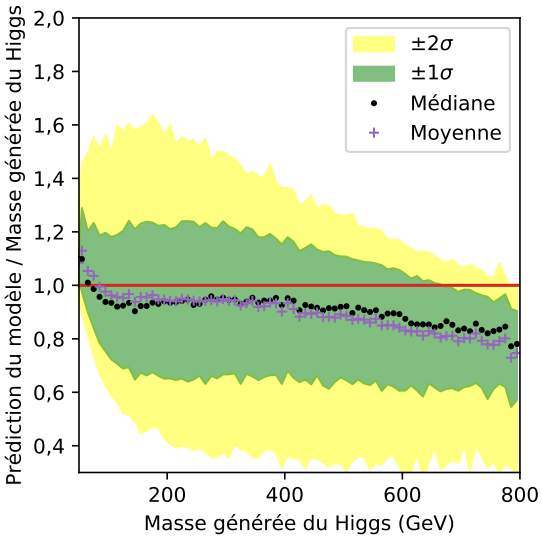
(b) *Canal* $\tau_h\tau_h$, *BDT.*



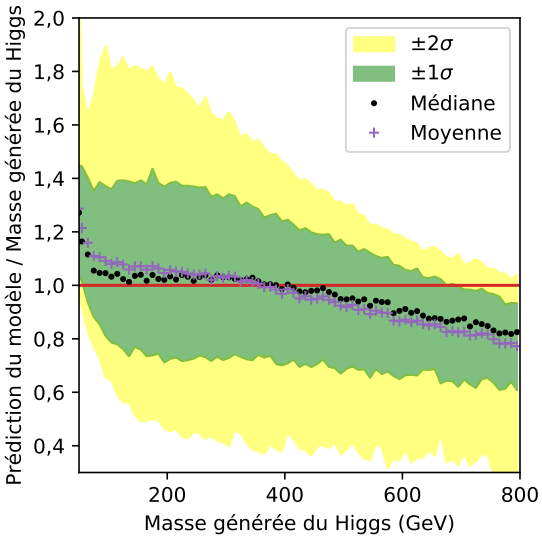
(c) *Canaux* $\ell\tau_h$, DEEPTAU.



(d) *Canaux* $\ell\tau_h$, *BDT.*

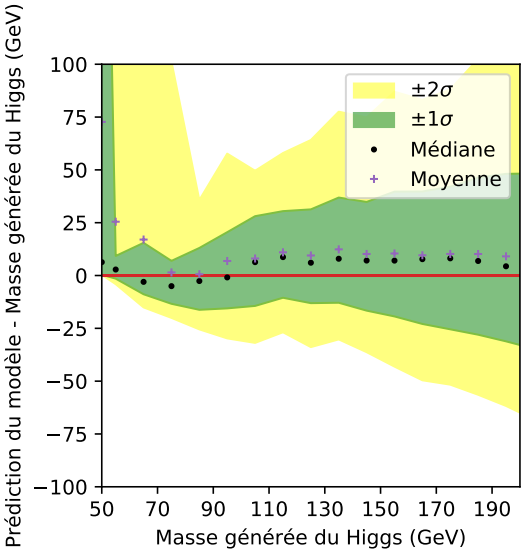


(e) *Canaux* $\ell\ell$, DEEPTAU.

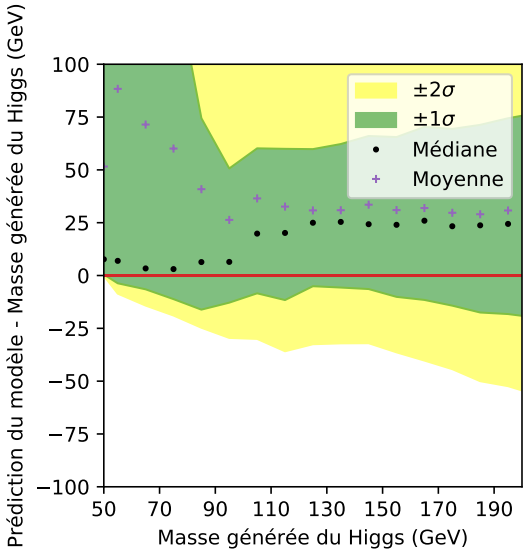


(f) *Canaux* $\ell\ell$, *BDT.*

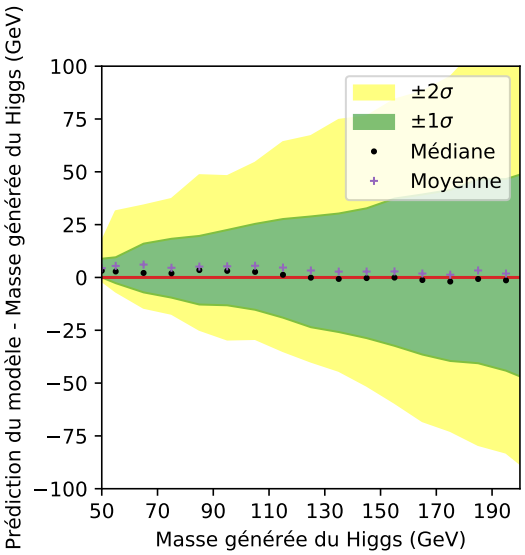
Figure 5.26 – *Réponses du modèle B sur les différents types de canaux avec une quantité variable de faux* τ_h.



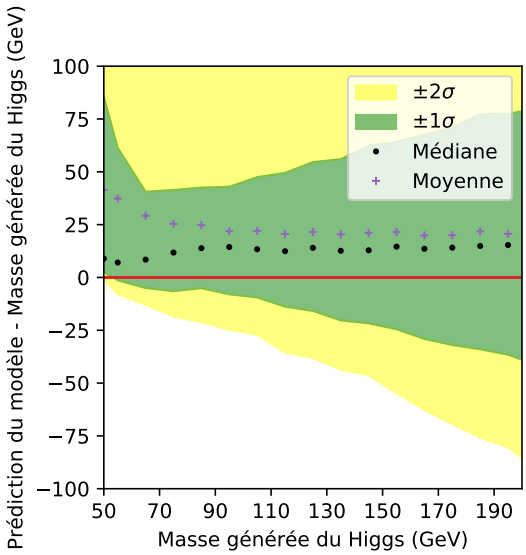
(a) *Canal* $\tau_h\tau_h$, DeepTau.



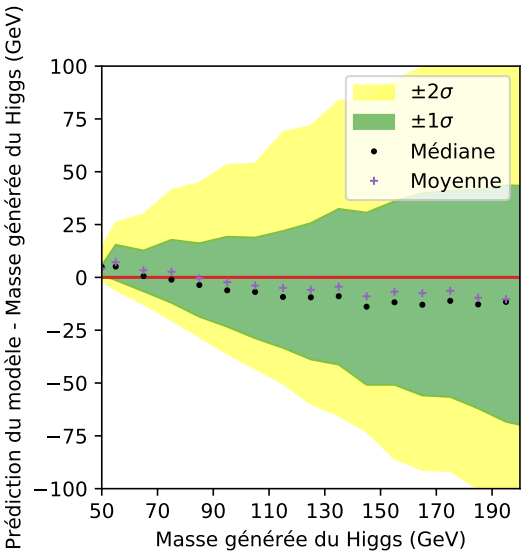
(b) *Canal* $\tau_h\tau_h$, *BDT.*



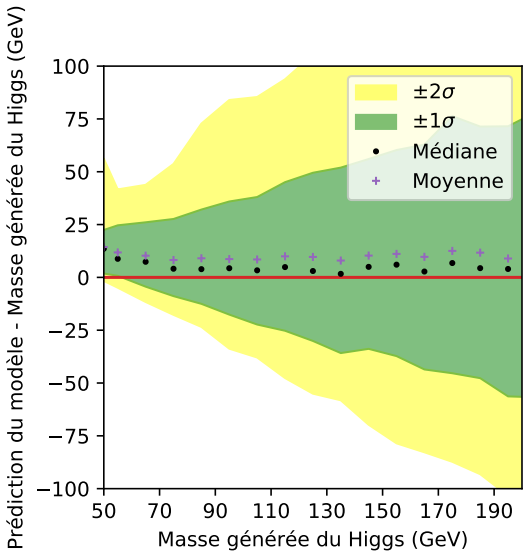
(c) *Canaux* $\ell\tau_h$, DeepTau.



(d) *Canaux* $\ell\tau_h$, *BDT.*



(e) *Canaux* $\ell\ell$, DeepTau.



(f) *Canaux* $\ell\ell$, *BDT.*

Figure 5.27 – *Écarts à basse masse du modèle B sur les différents types de canaux avec une quantité variable de faux* τ_h.

Les modèles notés B^{x} possèdent les mêmes hyper-paramètres que B mais sont entraînés uniquement sur les événements du canal x.

7.4.1 Séparation en six canaux

Les figures 5.28 et 5.29 donnent les réponses des modèles $\mathrm{B}^{\tau_\mathrm{h}\tau_\mathrm{h}}$, $\mathrm{B}^{\mu\tau_\mathrm{h}}$, $\mathrm{B}^{e\tau_\mathrm{h}}$ et $\mathrm{B}^{\mu\mu}$, $\mathrm{B}^{e\mu}$, B^{ee} testés sur leurs canaux respectifs, comparées à celles de B sur les mêmes canaux.

Dans le canal $\tau_\mathrm{h}\tau_\mathrm{h}$, la réponse de $\mathrm{B}^{\tau_\mathrm{h}\tau_\mathrm{h}}$, figure 5.28a, est plus proche de 1 que celle de B, figure 5.28b, entre 200 et 450 GeV. Au-delà, elle est jusqu'à 3 % plus basse. Elle est en revanche jusqu'à 10 % plus haute à basse masse. Le comportement des modèles est toutefois similaire dans cette région : une baisse locale de la réponse est observable pour $m_{\mathcal{H}} \simeq 80\,\mathrm{GeV}$. La coupure sur l'impulsion transverse des τ_h étant de 40 GeV pour chacun des deux τ_h, il s'agit probablement d'une transition entre les événements avec une majorité de vrais τ_h ($m_{\mathcal{H}} > 80\,\mathrm{GeV}$) et ceux avec une contamination importante par les faux τ_h ($m_{\mathcal{H}} < 80\,\mathrm{GeV}$). Le modèle $\mathrm{B}^{\tau_\mathrm{h}\tau_\mathrm{h}}$ est donc difficilement entraîné dans cette région de masse. Pour $\mathrm{B}^{\tau_\mathrm{h}\tau_\mathrm{h}}$ et B, la résolution relative sur le canal $\tau_\mathrm{h}\tau_\mathrm{h}$ est de 20 %.

Le modèle $\mathrm{B}^{\mu\tau_\mathrm{h}}$, figure 5.28c, possède une réponse équivalente à celle de B sur les mêmes événements, figure 5.28d, pour des masses inférieures à 400 GeV. À haute masse, la réponse moyenne du modèle B est toutefois plus proche de 1. Le même constat peut être fait dans le cas du canal $e\tau_\mathrm{h}$, figures 5.28e et 5.28f.

Dans le cas des canaux leptoniques, figure 5.29, l'utilisation de B plutôt que $\mathrm{B}^{\mu\mu}$, $\mathrm{B}^{e\mu}$ ou B^{ee} selon le canal permet d'améliorer la résolution relative sur $m_{\mathcal{H}}$ dont les valeurs sont données dans le tableau 5.3. Les valeurs des réponses moyennes sont peu modifiées par rapport aux valeurs des résolutions.

Canal x	Modèle B^{x}		Modèle B	
	min	max	min	max
$\mu\mu$	20	50	10	40
$e\mu$	20	40	20	30
ee	20	50	10	30

Tableau 5.3 – *Résolutions relatives minimales et maximales sur des intervalles de* 10 GeV *pour* $\mathrm{B}^{\mu\mu}$, $\mathrm{B}^{e\mu}$ *ou* B^{ee} *et B.*

Il semble ainsi préférable d'utiliser un seul modèle global plutôt qu'un modèle par canal. Cet effet peut être dû à la statistique plus faible à disposition lors de l'entraînement des DNNs séparément pour chaque canal. Un compromis peut être obtenu en séparant non pas par canal, mais par groupe de canaux.

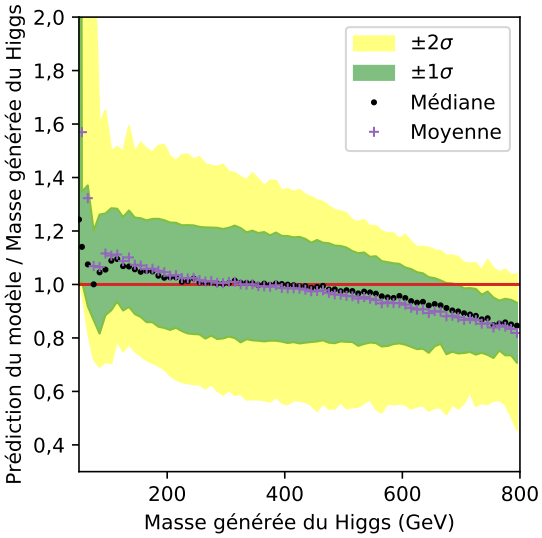
7.4.2 Séparation en trois groupes

En dehors de toute considération de reconstruction des particules, la phénoménologie des canaux d'un même groupe est sensiblement la même. Au lieu de séparer les six canaux ($\tau_\mathrm{h}\tau_\mathrm{h}$, $\mu\tau_\mathrm{h}$, $e\tau_\mathrm{h}$, $\mu\mu$, $e\mu$, ee), il est possible de former trois groupes ($\tau_\mathrm{h}\tau_\mathrm{h}$, $\ell\tau_\mathrm{h}$, $\ell\ell$), dans lesquels les quantités de τ_h et de neutrinos issus des désintégrations des leptons τ sont constantes. Cette nouvelle séparation permet ainsi d'avoir accès à de plus grandes quantités d'événements lors des entraînements, +100 % pour les canaux semi-leptoniques et +100 à +300 % pour les canaux leptoniques.

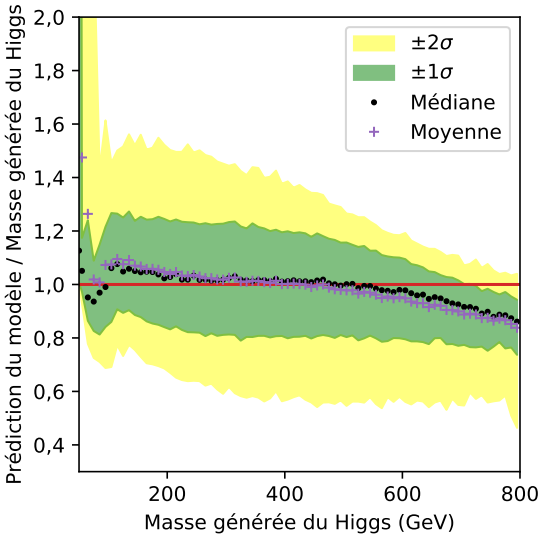
Le canal $\tau_\mathrm{h}\tau_\mathrm{h}$, seul de son groupe, est ainsi déjà traité dans la section précédente.

La figure 5.30 compare le modèle $\mathrm{B}^{\ell\tau_\mathrm{h}}$, entraîné sur les canaux semi-leptoniques, à B utilisé sur ces mêmes événements. Pour des masses supérieures à 300 GeV, les deux modèles sont équivalents en termes de réponse et de résolution relative. En revanche, à basse masse, le modèle B a une réponse moyenne de 1,10 contre 1,18 pour $\mathrm{B}^{\ell\tau_\mathrm{h}}$.

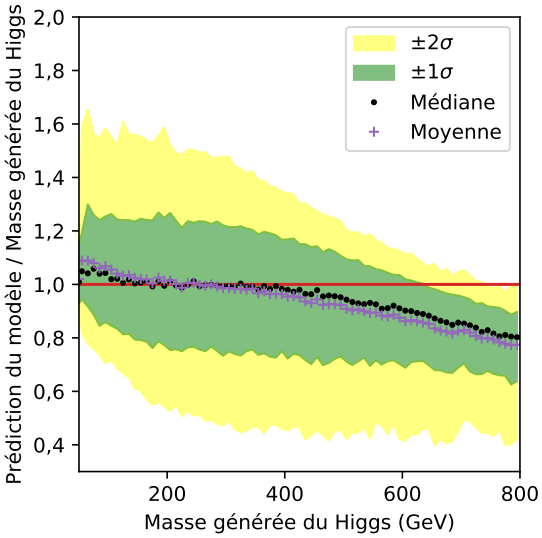
La figure 5.31 compare le modèle $\mathrm{B}^{\ell\ell}$, entraîné sur les canaux leptoniques, à B utilisé sur ces mêmes événements. Les réponses de ces deux modèles sont équivalentes. Entre 100 et 200 GeV, la réponse moyenne de $\mathrm{B}^{\ell\ell}$ est légèrement plus proche de 1 que celle de B. Cette différence est toutefois négligeable face à la résolution relative de ces modèles, de l'ordre de 30 % dans cette région.



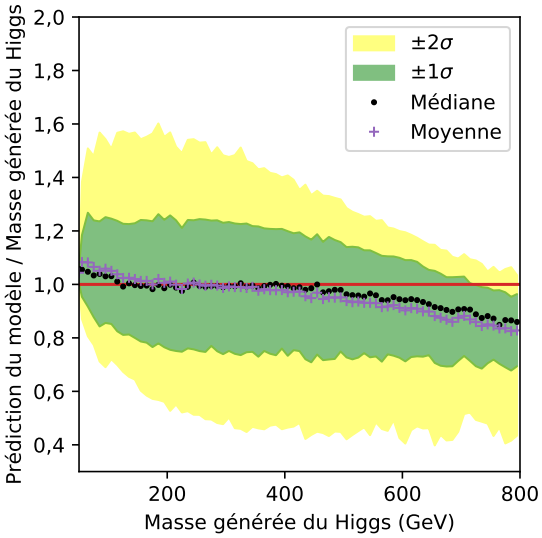
(a) *Modèle $B^{\tau_h\tau_h}$ testé sur $\tau_h\tau_h$.*



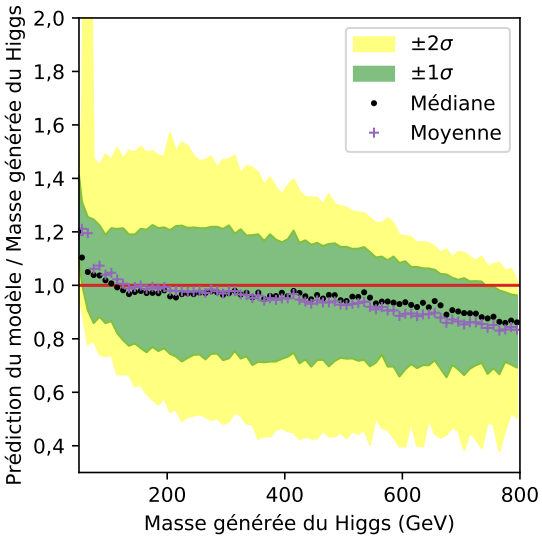
(b) *Modèle B testé sur $\tau_h\tau_h$.*



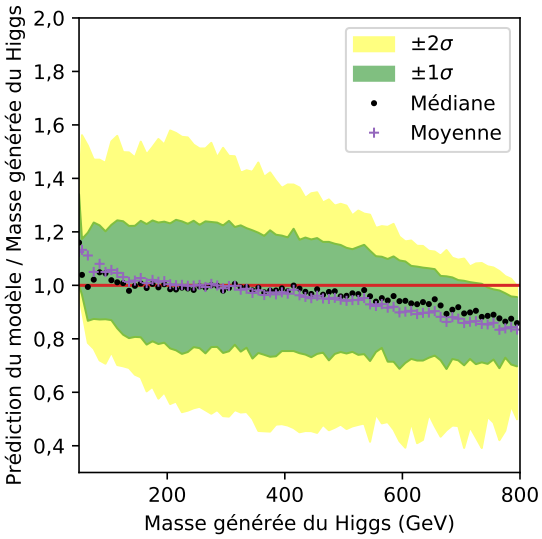
(c) *Modèle $B^{\mu\tau_h}$ testé sur $\mu\tau_h$.*



(d) *Modèle B testé sur $\mu\tau_h$.*

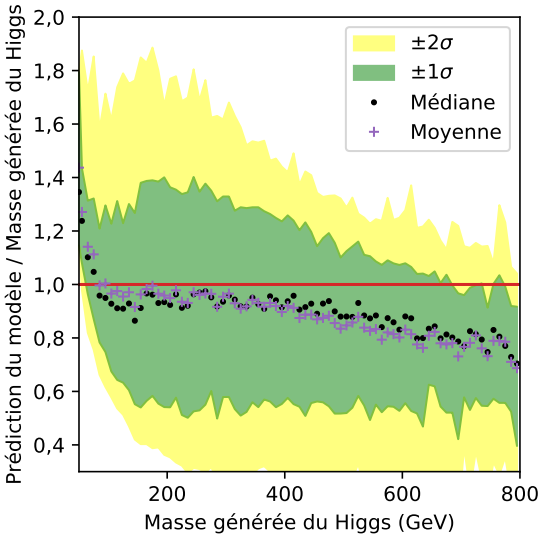


(e) *Modèle $B^{e\tau_h}$ testé sur $e\tau_h$.*

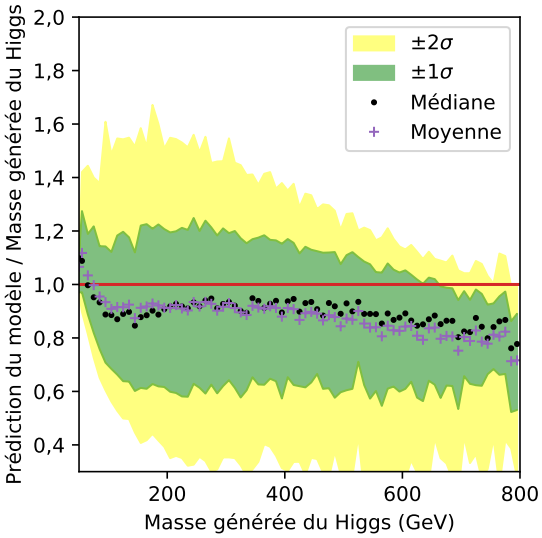


(f) *Modèle B testé sur $e\tau_h$.*

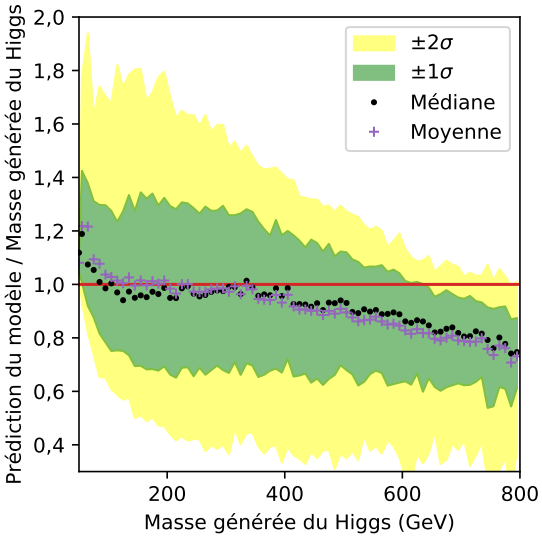
Figure 5.28 – *Comparaison des modèles entraînés par canal ($\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$) au modèle B.*



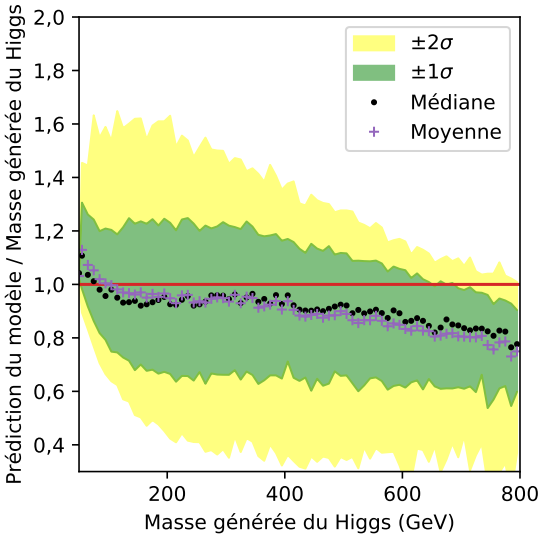
(a) *Modèle $B^{\mu\mu}$ testé sur $\mu\mu$.*



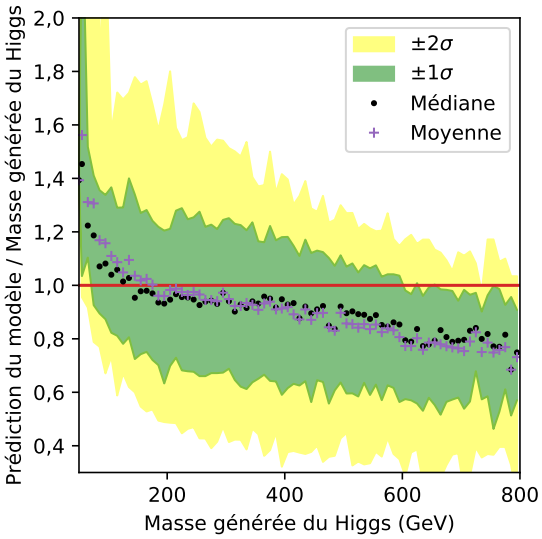
(b) *Modèle B testé sur $\mu\mu$.*



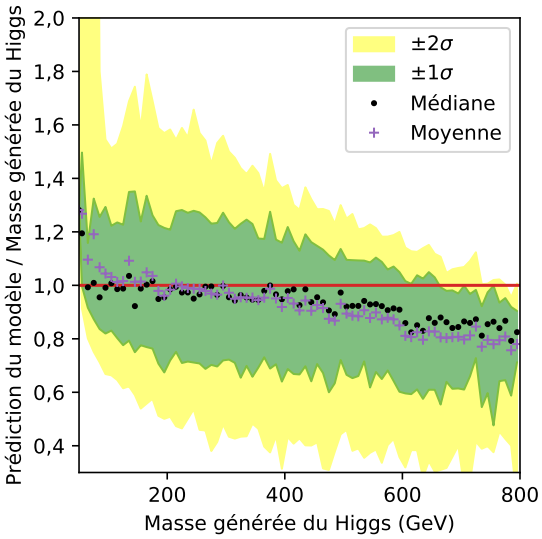
(c) *Modèle $B^{e\mu}$ testé sur $e\mu$.*



(d) *Modèle B testé sur $e\mu$.*

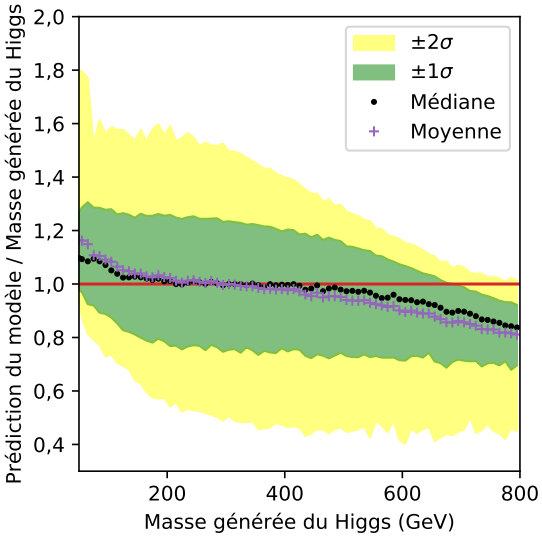


(e) *Modèle B^{ee} testé sur ee.*

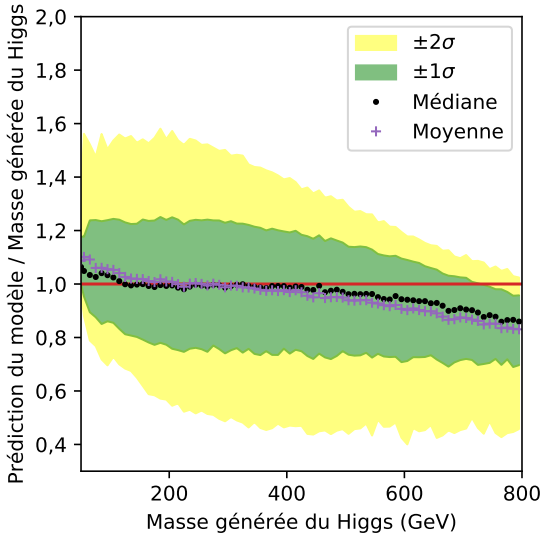


(f) *Modèle B testé sur ee.*

Figure 5.29 – *Comparaison des modèles entraînés par canal ($\mu\mu$, $e\mu$, ee) au modèle B.*

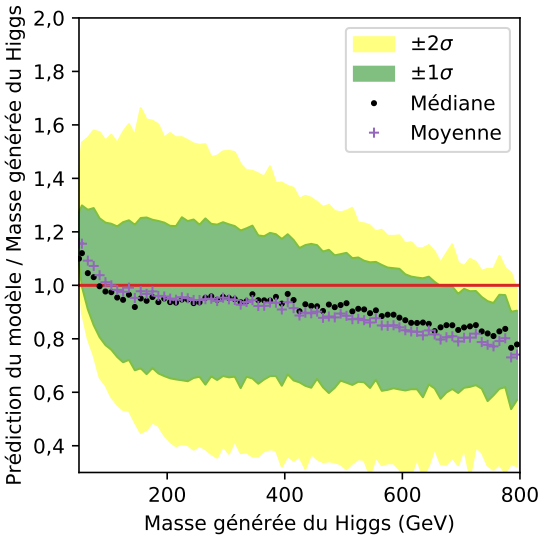


(a) *Modèle $B^{\ell\tau_h}$ testé sur $\ell\tau_h$.*

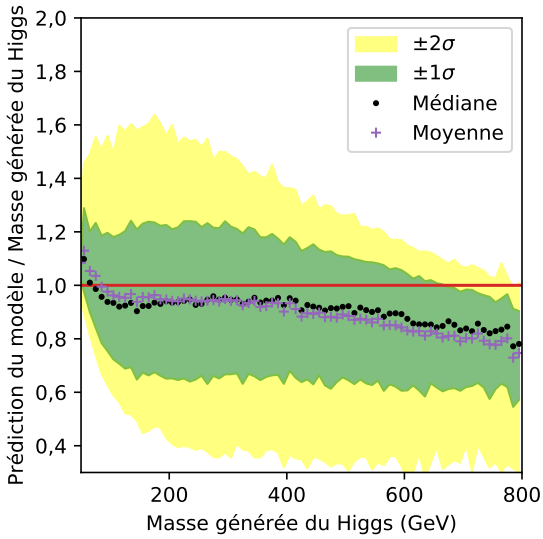


(b) *Modèle B testé sur $\ell\tau_h$.*

Figure 5.30 – *Comparaison de $B^{\ell\tau_h}$ à B.*



(a) *Modèle $B^{\ell\ell}$ testé sur $\ell\ell$.*



(b) *Modèle B testé sur $\ell\ell$.*

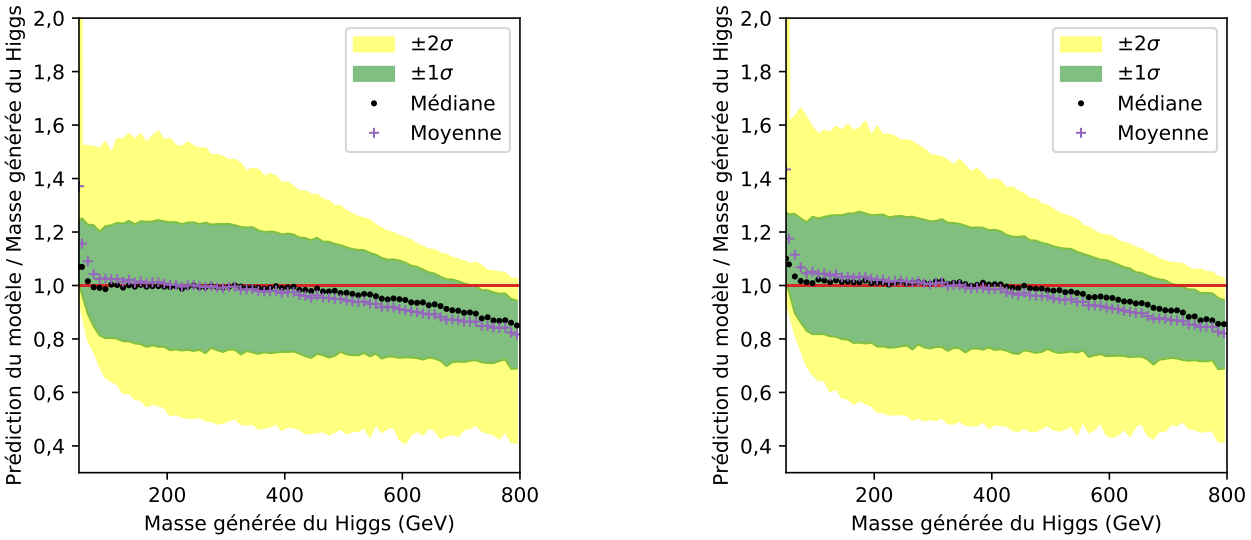
Figure 5.31 – *Comparaison de $B^{\ell\ell}$ à B.*

Utiliser un modèle par groupe de canaux ne permet donc pas non plus d'améliorer les estimations obtenues par rapport au modèle de référence entraîné sur l'ensemble des canaux. Ce modèle, B, a en variable d'entrée le nombre attendu de neutrinos dans l'état final, N_ν^{reco}, directement relié au groupe du canal. En effet, $N_\nu^{\text{reco}} = 2$ pour le canal hadronique, 3 pour les semi-leptoniques et 4 pour les leptoniques. Or, comme vu dans la section 6.1, tout modèle privé de cette information a des performances dégradées. Le modèle B identifie donc vraisemblablement correctement le groupe de canal grâce à N_ν^{reco}.

7.5 Effet de la définition de E_T^{miss}

Dans l'analyse présentée au chapitre 4, E_T^{miss} est déterminée par l'algorithme PUPPI [171]. Il s'agit de la « PUPPIMET », avec laquelle les modèles ont donc été entraînés. Toutefois, une autre estimation de E_T^{miss} existe, obtenue directement à partir de l'algorithme de PF, la « PFMET ». Ces deux définitions de E_T^{miss} sont introduites dans le chapitre 2.

Certaines analyses utilisent PFMET plutôt que PUPPIMET. Nous avons donc souhaité vérifier la portabilité du modèle B entraîné avec PUPPIMET à une utilisation avec PFMET. La figure 5.32 montre les réponses de B sur les mêmes événements, lorsque les variables reliées à E_T^{miss} sont obtenues à partir de PUPPIMET (figure 5.32a) ou PFMET (figure 5.32b).

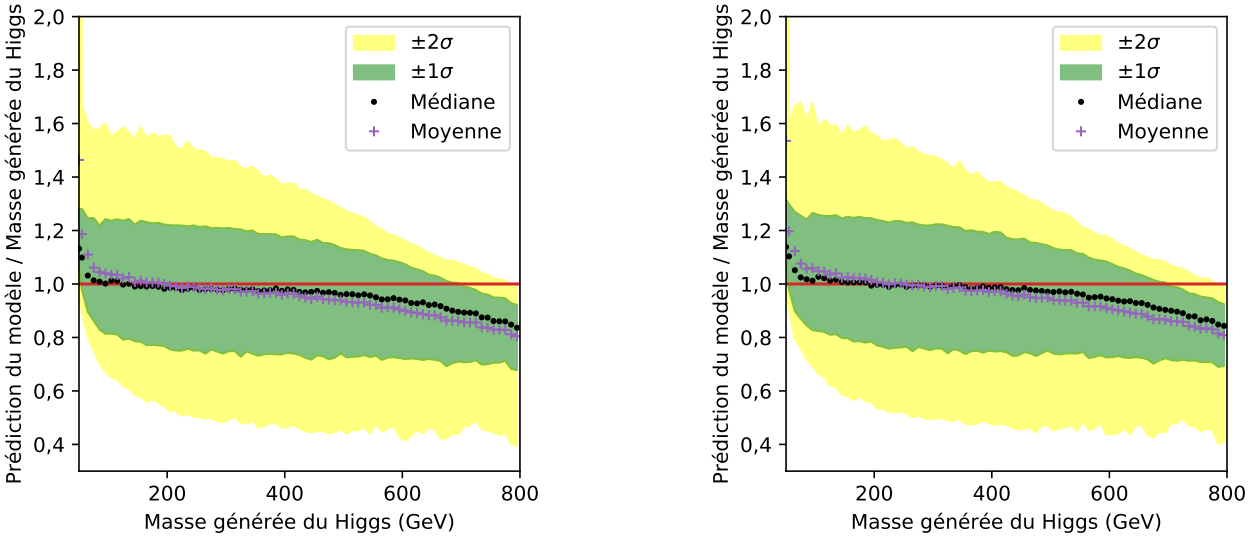


(a) *Modèle B testé avec* PUPPIMET. **(b)** *Modèle B testé avec PFMET.*

Figure 5.32 – *Réponses du modèle B avec* PUPPIMET *ou PFMET.*

L'utilisation de PFMET augmente la réponse de B. Cette augmentation est cependant inférieure à 3 %. La résolution relative est inchangée. Il est donc tout à fait possible d'utiliser le modèle B avec PFMET, bien qu'il soit entraîné avec PUPPIMET.

La possibilité d'obtenir de meilleures prédictions en utilisant PFMET à l'aide du modèle B^{PF} entraîné directement avec cette définition de E_T^{miss} a également été étudiée. La figure 5.33 présente les réponses obtenues.



(a) *Modèle* B^{PF} *testé avec* PUPPIMET. **(b)** *Modèle* B^{PF} *testé avec PFMET.*

Figure 5.33 – *Réponses du modèle* B^{PF} *avec* PUPPIMET *ou PFMET.*

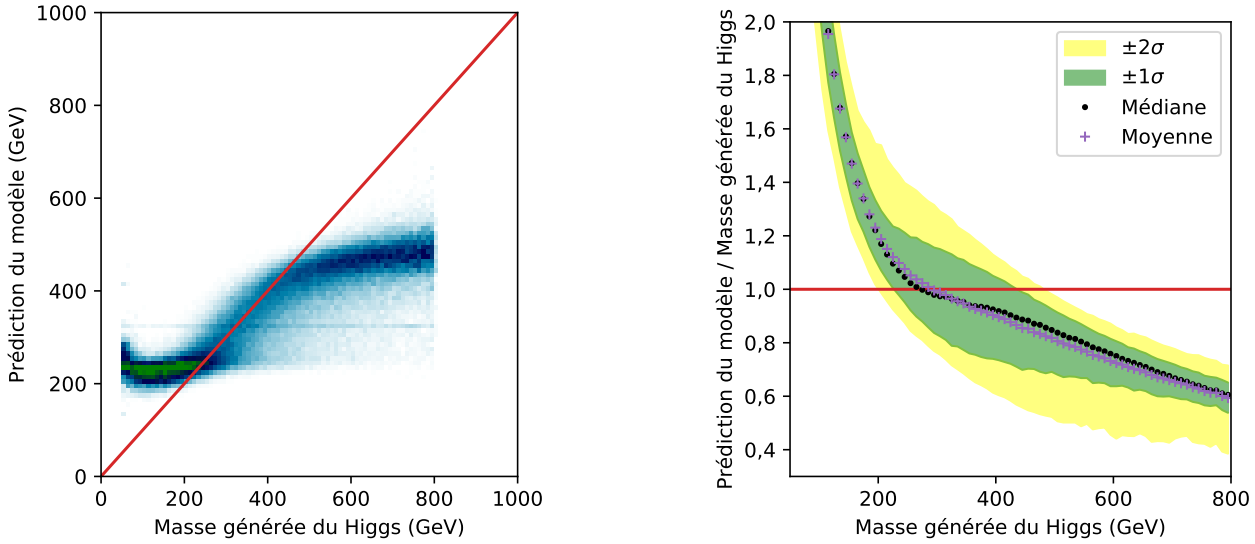
Le même effet de transition entre PFMET et PUPPIMET qu'avec B est observable avec B^{PF}. Cependant, la réponse moyenne de B avec PUPPIMET (figure 5.32a) est égale à $1{,}00 \pm 0{,}05$ de 80 GeV à 425 GeV, alors que ce n'est le cas que de 200 GeV à 425 GeV pour B^{PF} avec PFMET (figure 5.33b). De

plus, la réponse moyenne de B avec PFMET (figure 5.32b) est plus proche de 1 pour $m_{\mathcal{H}} \simeq 100\,\text{GeV}$ que celle de B^{PF} avec PFMET (figure 5.33b). Pour les analyses utilisant PFMET, le modèle B basé sur PUPPIMET est donc recommandé plutôt que B^{PF}.

7.6 Effet de l'intervalle de masse

7.6.1 Étendue de l'intervalle

L'intervalle de masse exploré lors de l'entraînement s'étend de 50 à 800 GeV, ce qui correspond aux valeurs de la masse du boson de Higgs modifié $\mathcal{H}$ utilisé lors de la génération des événements présentée dans la section 2. Cet intervalle est également le domaine de validité du modèle. La figure 5.34 montre les performances du modèle $\text{B}^{\text{200-500}}$, équivalent au modèle B mais entraîné uniquement entre 200 et 500 GeV.



(a) *Prédictions en fonction des valeurs vraies.* **(b)** *Réponse.*

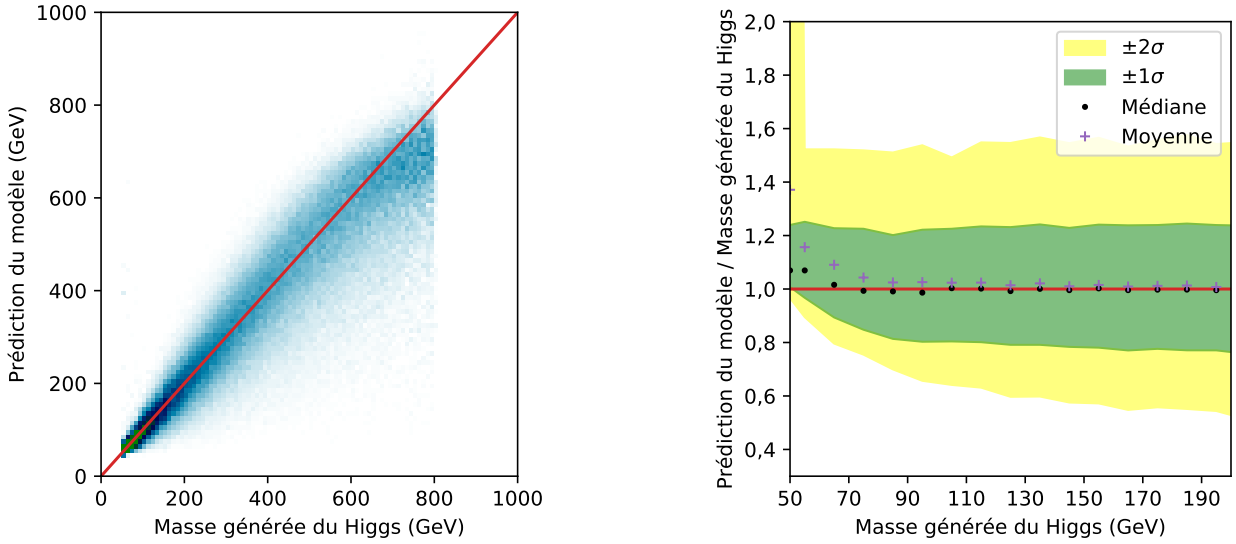
Figure 5.34 – *Performances du modèle* $B^{200\text{-}500}$.

L'histogramme à deux dimensions des prédictions de $\text{B}^{\text{200-500}}$ en fonction de la vraie valeur de $m_{\mathcal{H}}$, figure 5.34a, montre que les prédictions du modèle sont contenues dans l'intervalle d'entraînement, à quelques exceptions près. Dans l'intervalle d'entraînement, les prédictions sont cohérentes avec $m_{\mathcal{H}}$, la population de l'histogramme est proche de la première bissectrice ($y_{\text{préd}} = y_{\text{vraie}}$) en rouge. Les événements avec $m_{\mathcal{H}} < 200\,\text{GeV}$ sont prédits vers 230 GeV et ceux avec $m_{\mathcal{H}} > 500\,\text{GeV}$ sont prédits vers 480 GeV. Un modèle ne peut donc pas être utilisé afin de prédire des masses en dehors de son intervalle d'entraînement.

Afin d'obtenir un modèle pertinent dans l'optique d'une utilisation dans les analyses de CMS, il est donc important d'utiliser un intervalle contenant la gamme de masse des particules du modèle standard se désintégrant en paire de leptons τ, en particulier des bosons Z et h à 91,2 et 125,1 GeV respectivement. Pour une recherche de bosons de Higgs supplémentaires de haute masse, la limite supérieure de l'intervalle d'entraînement doit être la plus haute possible.

7.6.2 Effet de bord

Origine de l'effet de bord L'intervalle de masse utilisé pour l'entraînement du modèle B s'étend de 50 à 800 GeV. Comme discuté dans la section 2, il ne nous est pas possible de l'étendre avec la méthode utilisée pour générer les événements. Or, il existe un effet de bord sur les prédictions du modèle B lié à cet intervalle. La figure 5.35a montre l'histogramme à deux dimensions des prédictions de B en fonction de la vraie valeur de $m_{\mathcal{H}}$. Pour $m_{\mathcal{H}} > 600\,\text{GeV}$, les prédictions de B saturent progressivement en dessous de 800 GeV. De même, la réponse moyenne de B à basse masse, figure 5.35b, est de



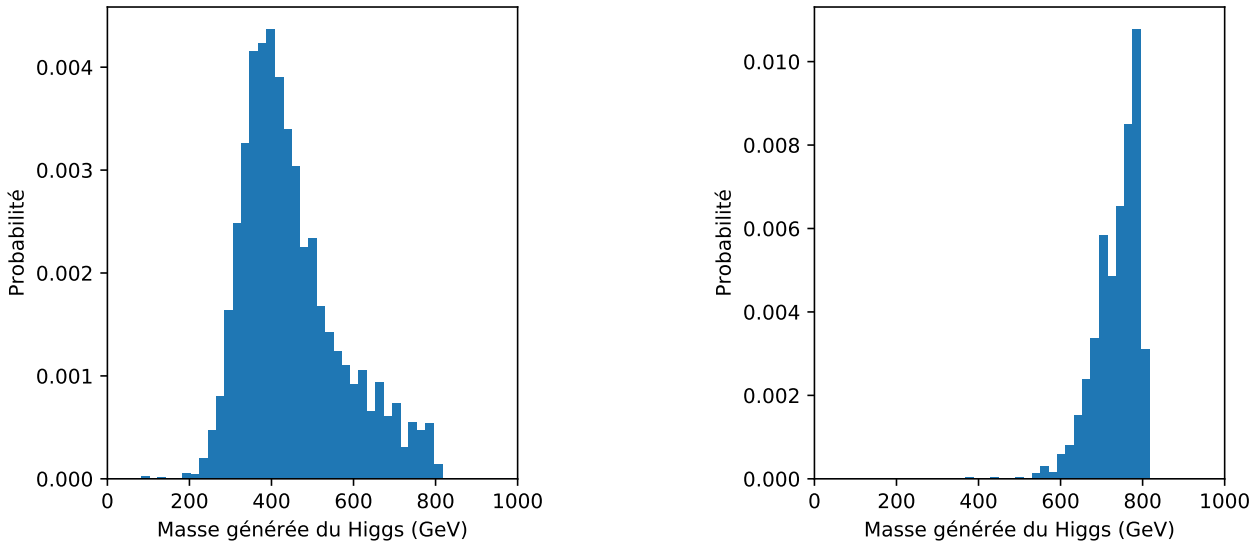
(a) *Prédictions en fonction des valeurs vraies.*

(b) *Réponse du modèle B à basse masse.*

Figure 5.35 – *Performances du modèle B.*

$1{,}01 \pm 0{,}01$ pour $80\,\text{GeV} < m_{\mathcal{H}} < 200\,\text{GeV}$. En dessous de 80 GeV, la réponse moyenne augmente et la limite de l'écart-type inférieur (limite basse de la bande verte $\pm 1\sigma$) passe de 0,8 à 1,0.

Ainsi, $m_{\mathcal{H}}$ est surestimée à basse masse et sous-estimée à haute masse. Il s'agit de l'effet de bord de l'intervalle de masse. L'interprétation de cet effet est la suivante. Chaque ensemble d'événements prédits à une valeur de $y_{\text{préd}}$ donnée est une même famille du point de vue du DNN. En termes de classification en un nombre infini de catégories au lieu de régression, cela revient à dire qu'une famille est donc une catégorie identifiée. Sur la figure 5.36 sont représentées les distributions de $y_{\text{vraie}} = m_{\mathcal{H}}$ pour des valeurs de la masse prédite par le modèle $y_{\text{préd}}$ de 400 (figure 5.36a) et 750 GeV (figure 5.36b) à ± 5 GeV. Il s'agit donc de tranches horizontales de l'histogramme de la figure 5.35a, ce qui correspond à une famille d'événements selon le DNN.



(a) *Distribution de y_{vraie} pour $y_{\text{préd}} = 400 \pm 5\,\text{GeV}$.*

(b) *Distribution de y_{vraie} pour $y_{\text{préd}} = 750 \pm 5\,\text{GeV}$.*

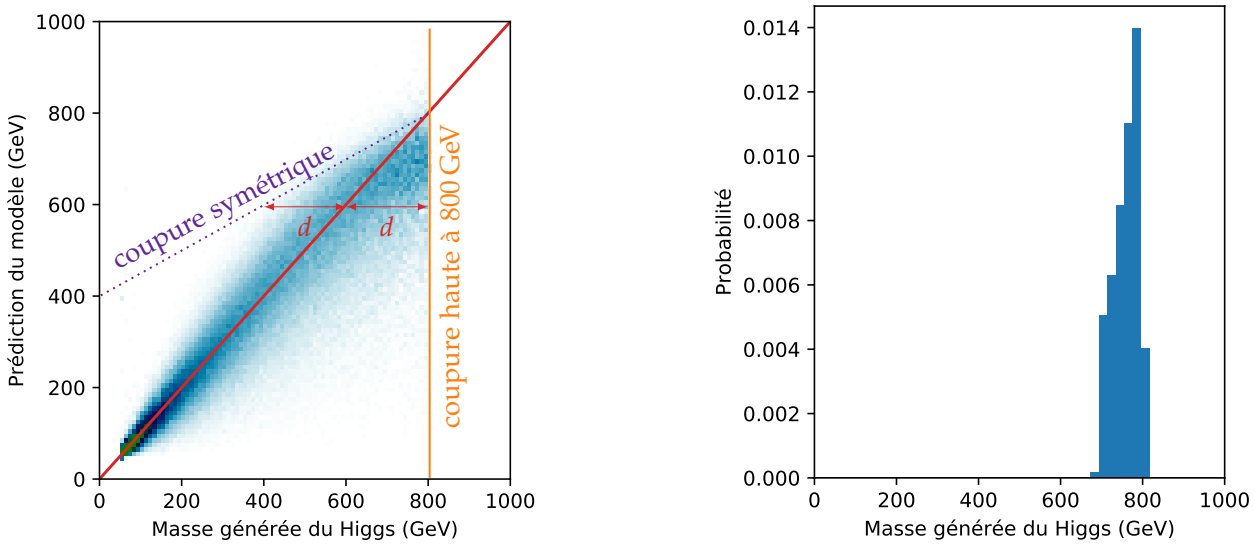
Figure 5.36 – *Distributions de y_{vraie} à $y_{\text{préd}}$ fixée.*

Loin des bords de l'intervalle d'entraînement, figure 5.36a, la distribution de y_{vraie} pour une famille est complète ; les queues de la distribution sont présentes des deux côtés de la valeur centrale. Lors de l'entraînement, le DNN « apprend » à prédire pour les événements de cette famille la valeur de $y_{\text{préd}}$ minimisant la fonction de coût sur cette distribution. Nous obtenons dans ce cas une valeur

proche de 400 GeV, ce qui est correct.

Au niveau du bord à haute masse, figure 5.36b, l'absence d'événements au-delà de 800 GeV donne une distribution tronquée de y_{vraie} pour une famille donnée. Seule l'extrémité à basse masse de la queue de la distribution est présente. Par conséquent, le DNN ne connaît que les basses masses de cette famille, il est ainsi tout à fait cohérent qu'il en sous-estime la masse. La minimisation de la fonction de coût mène donc à des prédictions biaisées. Si les événements au-delà de 800 GeV étaient présents dans cette famille, la distribution serait plus étendue vers les hautes masses, et la minimisation de la fonction de coût mènerait donc le DNN à prédire une masse plus élevée. L'effet est inversé pour le bord à basse masse, d'où la surestimation.

Principe de la correction La minimisation de la fonction de coût est donc réalisée correctement, mais sur des familles d'événements tronquées. Afin de contrer cet effet, l'idée retenue est de reformer des familles équilibrées en les tronquant de manière symétrique par rapport à la valeur de y_{vraie} devant leur correspondre. La figure 5.37a illustre le principe de cette coupure symétrique.



(a) *Principe de la coupure symétrique à haute masse.*

(b) *Distribution de y_{vraie} pour $y_{\text{préd}} = 750 \pm 5$ GeV avec coupure symétrique.*

Figure 5.37 – *Mise en place de la coupure symétrique.*

La troncature des familles due à l'extrémité de l'intervalle à 800 GeV est symétrisée par rapport au centre de valeur y_{vraie} devant correspondre à chaque famille. Ainsi, tout événement d'une famille (tranche horizontale) situé à une distance d de la valeur centrale à prédire (sur la diagonale rouge) plus grande que la distance de cette valeur centrale à l'extrémité de l'intervalle (trait orange vertical) est rejeté de l'entraînement. La distribution de la famille de la figure 5.36b symétrisée par troncature selon cette méthode est présentée sur la figure 5.37b. La valeur de la fonction de coût sur cette distribution n'est plus minimale pour la valeur précédemment prédite par le modèle, elle l'est en revanche pour une valeur prédite plus élevée. Le modèle actuel sous-estimant $m_{\mathcal{H}}$ à haute masse, cette correction doit donc permettre d'obtenir un nouveau modèle avec une réponse plus proche de 1.

Modification de la fonction de coût Une coupure symétrique est également mise en place à basse masse en suivant le même principe. Cinq zones peuvent alors être définies dans le plan $(y_{\text{vraie}}, y_{\text{préd}})$ afin d'appliquer les coupures symétriques aux deux extrémités de l'intervalle de masse :

Zone 1 $y_{\text{vraie}} < 50$ GeV : absence d'événements due à la masse minimale de l'entraînement ;

Zone 2 $y_{\text{vraie}} > 800$ GeV : absence d'événements due à la masse maximale de l'entraînement ;

Zone 3 $|y_{\text{préd}} - y_{\text{vraie}}| > |800\,\text{GeV} - y_{\text{préd}}|$: zone d'exclusion à haute masse ;

Zone 4 $|y_{\text{préd}} - y_{\text{vraie}}| > |50\,\text{GeV} - y_{\text{préd}}|$: zone d'exclusion à basse masse ;

Zone 5 zone centrale.

Ces zones sont illustrées sur la figure 5.38. Dans la zone centrale, la fonction de coût est utilisée de manière classique, sans changement. Dans les zones 3 et 4 d'exclusion en revanche, pour ne pas prendre en compte les événements afin de symétriser les bords de l'intervalle d'entraînement au sein d'une famille, la fonction de coût est rendue égale à zéro. En d'autres termes, nous procédons au changement $\mathsf{L} = \mathsf{L}_{\text{MAPE}} \to \mathsf{L}'$ avec

$$\mathsf{L}' = \mathsf{L}_{\text{MAPE}} \times \begin{cases} 0 & \text{si } (y_{\text{vraie}}, y_{\text{préd}}) \in \text{zones 3 ou 4} \\ 1 & \text{sinon} \end{cases} . \tag{5.39}$$

Cette fonction de coût ne respecte pas la condition

$$\arg\min_{y_{\text{préd}}} \big(\mathsf{L}(y_{\text{vraie}}, y_{\text{préd}})\big) = \{\, y_{\text{vraie}} \,\} , \tag{5.40}$$

c'est-à-dire que le minimum de L soit atteint lorsque $y_{\text{préd}} = y_{\text{vraie}}$. Des problèmes de convergence lors de l'entraînement peuvent alors survenir. C'est effectivement ce que nous avons pu observer lors de la mise en place de cette fonction de coût avec la condition d'exclusion de la zone 4. Nous avons alors choisi de multiplier la valeur de la fonction de coût par 0,1 dans la zone 4 au lieu de 0. De plus, nous avons observé que la multiplication de L_{MAPE} par la racine de y_{vraie}, conjointement avec les conditions d'exclusion, permettait d'améliorer encore la réponse du modèle. La fonction de coût ainsi utilisée est $\mathsf{L}_{\text{MA}\sqrt{\text{PE}}\times\text{b}}$, définie par

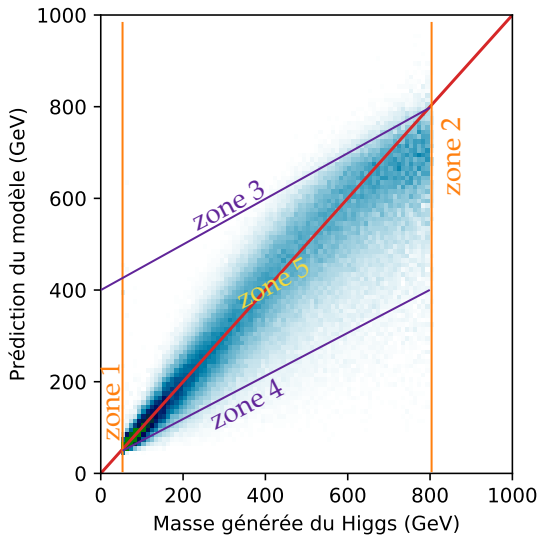


Figure 5.38 – *Zones considérées pour l'entraînement.*

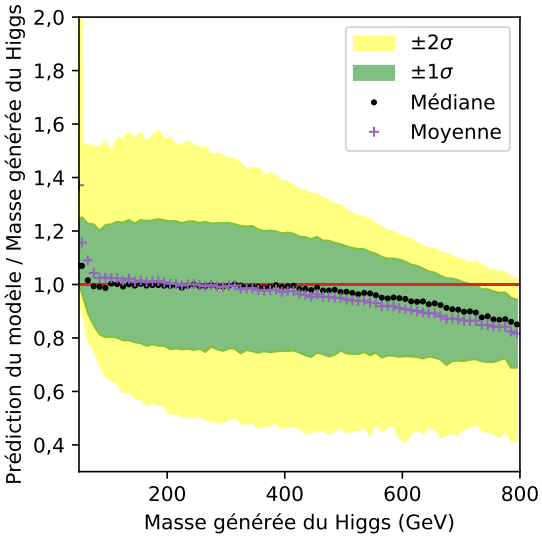
$$\mathsf{L}_{\text{MA}\sqrt{\text{PE}}\times\text{b}}(y_{\text{vraie}}, y_{\text{préd}}) = \mathsf{L}_{\text{MA}\sqrt{\text{PE}}}(y_{\text{vraie}}, y_{\text{préd}}) \times \begin{cases} 0 & \text{si } (y_{\text{vraie}}, y_{\text{préd}}) \in \text{zone 3} \\ 0{,}1 & \text{si } (y_{\text{vraie}}, y_{\text{préd}}) \in \text{zone 4} \\ 1 & \text{sinon} \end{cases} \tag{5.41}$$

avec

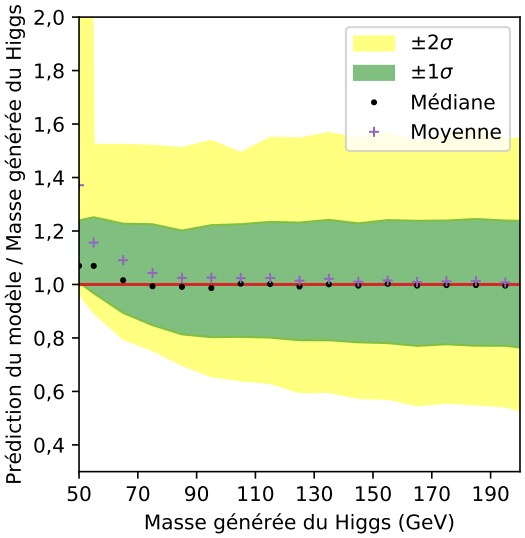
$$\begin{aligned} \mathsf{L}_{\text{MA}\sqrt{\text{PE}}}(y_{\text{vraie}}, y_{\text{préd}}) &= \mathsf{L}_{\text{MAPE}}(y_{\text{vraie}}, y_{\text{préd}}) \times \sqrt{y_{\text{vraie}}} = \left| \frac{y_{\text{préd}} - y_{\text{vraie}}}{y_{\text{vraie}}} \right| \times \sqrt{y_{\text{vraie}}} \\ \Leftrightarrow \mathsf{L}_{\text{MA}\sqrt{\text{PE}}}(y_{\text{vraie}}, y_{\text{préd}}) &= \left| \frac{y_{\text{préd}} - y_{\text{vraie}}}{\sqrt{y_{\text{vraie}}}} \right| . \end{aligned} \tag{5.42}$$

Nouveau modèle obtenu Le nouveau modèle obtenu, noté B', est comparé à B sur la figure 5.39. Pour des masses inférieures à 70 GeV, la réponse médiane de B', en figure 5.39d, est égale à 1 alors que celle de B, en figure 5.39b, est sujette à l'effet de bord. À haute masse, la réponse de B' en figure 5.39c est également plus proche de 1 que celle de B, en figure 5.39a. L'utilisation de $\mathsf{L}_{\text{MA}\sqrt{\text{PE}}}$ comme fonction de coût permet donc de supprimer l'effet de bord à basse masse et de le réduire à haute masse.

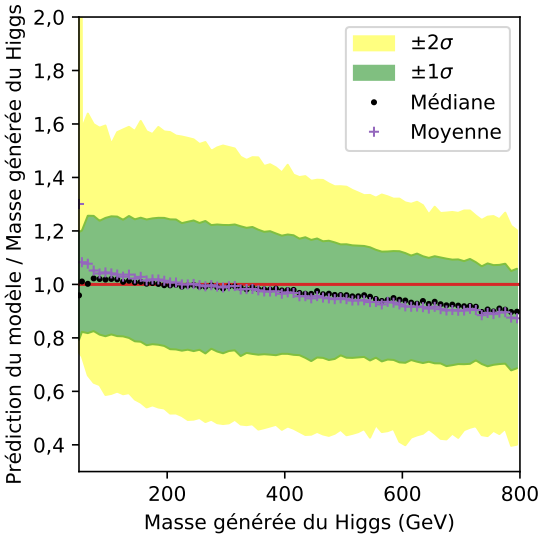
Exploitation de la queue à haute masse des événements générés Lors de la génération des événements à haute masse, la largeur du boson de Higgs permet d'obtenir des événements avec $m_{\mathcal{H}}$ supérieure à 800 GeV, comme discuté en section 2. Jusqu'ici, nous ne considérions que les événements tels que $50\,\text{GeV} \leqslant m_{\mathcal{H}} \leqslant 800\,\text{GeV}$. L'inclusion de la queue à haute masse des événements générés permet d'étendre artificiellement l'intervalle d'entraînement jusqu'à 1 TeV. Le biais éventuel dû à la faible quantité d'événements au-delà de 800 GeV est évité grâce à la pondération présentée dans la section 2. De même, les définitions des cinq zones utilisées pour déterminer la fonction de coût sont adaptées à la nouvelle valeur maximale de $m_{\mathcal{H}}$ fixée à 1 TeV. Le modèle B" ainsi obtenu est comparé à B' sur la figure 5.39. La réponse à basse masse de B", figure 5.39f, est semblable à celle de B', figure 5.39d, ce qui est attendu. En revanche, la réponse moyenne de B" à haute masse, figure 5.39e, est de $1{,}00 \pm 0{,}04$ contre $0{,}93 \pm 0{,}07$ pour B', figure 5.39c. Enfin, la résolution relative de B' est de



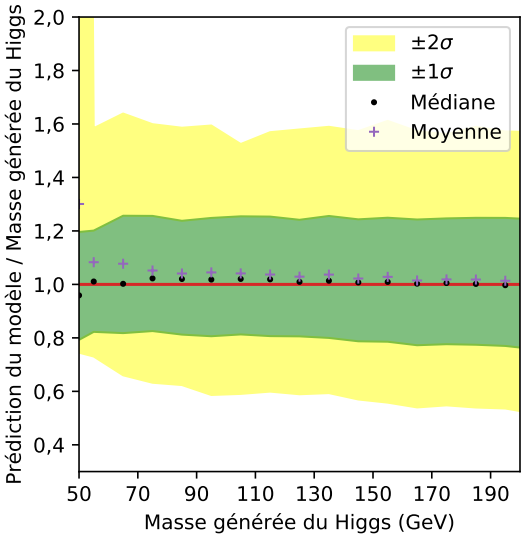
(a) *Réponse du modèle B.*



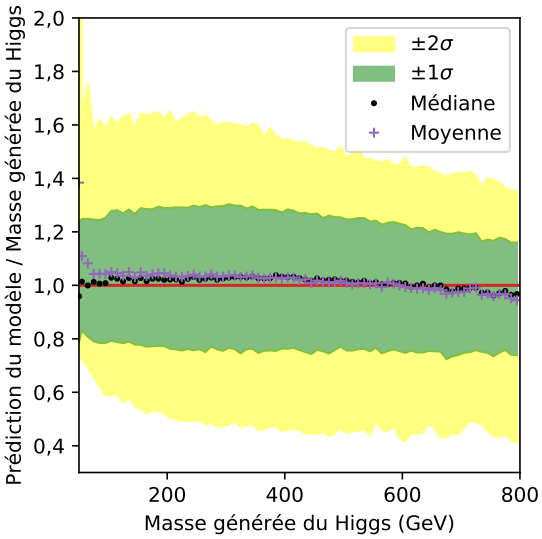
(b) *Réponse du modèle B à basse masse.*



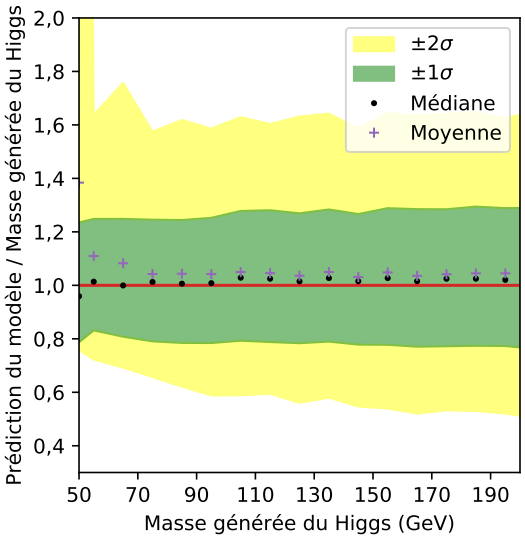
(c) *Réponse du modèle B′.*



(d) *Réponse du modèle B′ à basse masse.*



(e) *Réponse du modèle B″.*



(f) *Réponse du modèle B″ à basse masse.*

Figure 5.39 – *Comparaison des modèles B, B′ et B″.*

22 %, celle de B" de 25 %. Malgré une légère dégradation de la résolution, l'utilisation de la fonction de coût modifiée $L_{MA\sqrt{PE}\times b}$ et des événements entre 800 GeV et 1 TeV obtenus grâce à la largeur de $\mathcal{H}$ permet de ramener la réponse moyenne du modèle à des valeurs de $1{,}00 \pm 0{,}05$ pour des valeurs de $m_{\mathcal{H}}$ allant de 80 à 800 GeV.

7.7 Modèle final

Le modèle B" que nous avons construit est donc entraîné sur des événements $\mathcal{H} \to \tau\tau$ où $\mathcal{H}$ est le boson de Higgs du modèle standard h avec une masse modifiée entre 50 et 800 GeV, avec addition d'empilement selon le profil de l'année 2017, dont la sélection est réalisée selon la procédure décrite section 2. En particulier, l'algorithme DEEPTAU [176] est utilisé pour l'identification des taus hadroniques. La largeur de $\mathcal{H}$ permet d'exploiter des événements où sa masse effective est supérieure à 800 GeV, jusqu'à 1 TeV. La gamme d'utilisation de notre modèle est toutefois considérée comme allant de 50 à 800 GeV uniquement. Les hyper-paramètres de B" sont ceux de B à l'exception de la fonction de coût :

— 3 couches cachées ;
— 1000 neurones par couche cachée ;
— fonction d'activation Softplus, $x \mapsto \ln(1 + e^x)$;
— algorithme d'optimisation Adam, présenté en section 5.3.2 ;
— fonction de coût $L_{MA\sqrt{PE}\times b}$, définie équation (5.41) ;
— initialisation des poids selon le mode « Glorot Uniforme » [272] ;
— 27 variables d'entrée données en section 2.4.

L'utilisation de B" dans les analyses de CMS est abordé dans la section suivante.

8 Utilisation du modèle dans les analyses CMS

Dans cette section, l'utilisation du modèle B" introduit dans la section 7 sur les événements de l'analyse présentée dans le chapitre 4 est discutée. Ce modèle, issu du ML, est conçu pour prédire la masse de la particule se désintégrant en paire de leptons τ. Ces prédictions sont notées m_{ML}.

8.1 Utilisation de m_{ML} comme variable discriminante

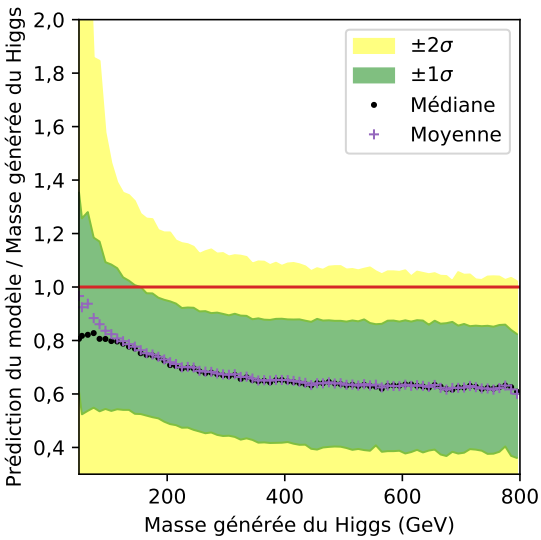
8.1.1 Performances de m_{ML} et de $m_{\text{T}}^{\text{tot}}$ en tant qu'estimateurs de la masse

Dans l'analyse présentée au chapitre 4, la variable discriminante utilisée est $m_{\text{T}}^{\text{tot}}$. Il s'agit d'une approximation de la masse invariante du système dans l'état final, restreinte au plan transverse car l'ensemble des neutrinos est remplacé par $E_{\text{T}}{}^{\text{miss}}$. Nous avons souhaité étudier l'utilisation de m_{ML} comme variable discriminante.

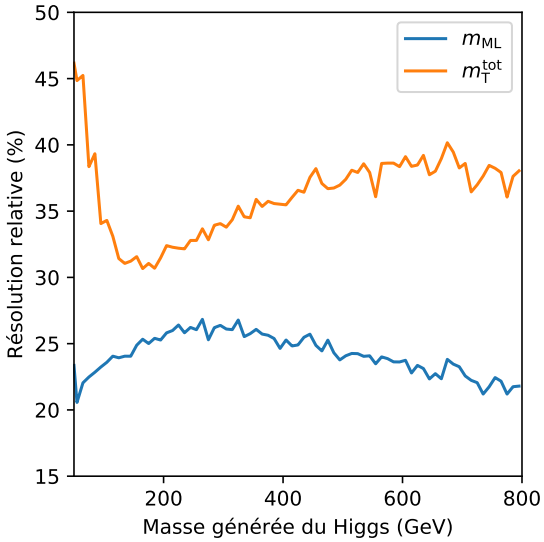
La réponse de $m_{\text{T}}^{\text{tot}}$ est représentée en fonction de $m_{\mathcal{H}}$ sur la figure 5.40a. Sa médiane est comprise entre 0,83 et 0,60 sur toute la gamme de masse, les valeurs de $m_{\text{T}}^{\text{tot}}$ ne sont donc absolument pas concordantes avec la masse $m_{\mathcal{H}}$ du boson initial. La réponse médiane du modèle B" en revanche, figure 5.39e, est de $1{,}00 \pm 0{,}05$. Les valeurs de m_{ML} sont donc plus fidèles à $m_{\mathcal{H}}$ que celles de $m_{\text{T}}^{\text{tot}}$.

Cependant, pour une variable discriminante, l'enjeu est d'avantage de séparer signal et bruit de fond que de présenter une réponse de 1. Pour cela, la résolution relative est importante. Une mauvaise résolution étale les distributions des différents signaux qui se superposent, rendant difficile leur séparation.

Les résolutions relatives de $m_{\text{T}}^{\text{tot}}$ et m_{ML} sont représentées figure 5.40b. Dans le cas de $m_{\text{T}}^{\text{tot}}$, la résolution relative est de 45 % à 50 GeV. Elle diminue à 31 % à 150 GeV, puis augmente jusqu'à 500 GeV où elle est d'environ 37 %. Pour m_{ML} en revanche, elle est de 21 % à basse masse, augmente progressivement jusqu'à 26 % à 270 GeV puis redescend à 22 % à haute masse. La résolution relative de m_{ML} est systématiquement meilleure que celle de $m_{\text{T}}^{\text{tot}}$. L'utilisation de m_{ML} au lieu de $m_{\text{T}}^{\text{tot}}$ en tant que variable discriminante pour l'analyse du chapitre 4 est donc particulièrement intéressante.



(a) *Réponse de* $m_{\mathrm{T}}^{\mathrm{tot}}$.



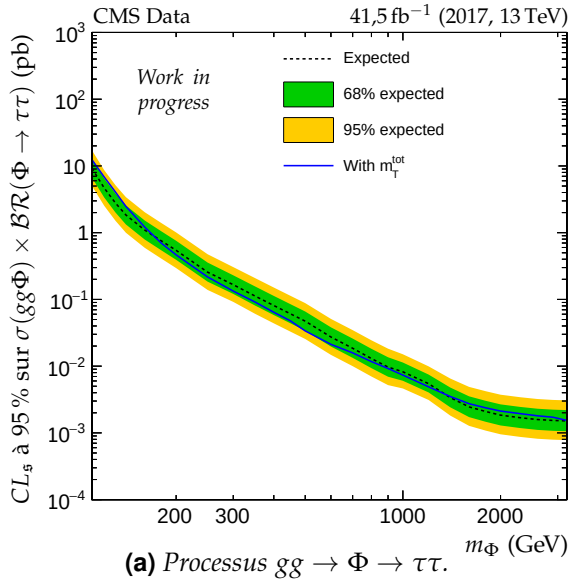
(b) *Résolutions relatives de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML}.

Figure 5.40 – *Réponse de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et comparaison de sa résolution relative à celle de* m_{ML}.

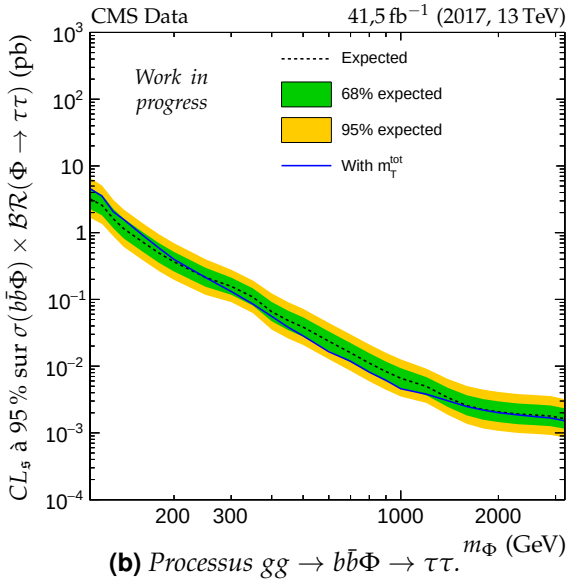
8.1.2 Limites d'exclusion avec m_{ML}

L'analyse indépendante du modèle introduite au chapitre 4 a été réalisée sur l'année 2017 avec $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} comme variables discriminantes. Les incertitudes systématiques de normalisation et de forme, introduites au chapitre 4, ont été recalculées avec m_{ML} afin de réaliser le traitement statistique des données. Les résultats sont donnés en figure 5.41 pour les processus $gg \to \Phi \to \tau\tau$ et $gg \to b\bar{b}\Phi \to \tau\tau$. Il s'agit des limites « attendues », obtenues en considérant que l'estimation du bruit de fond correspond exactement aux observations.

Les limites d'exclusion avec m_{ML} sont équivalentes à celles données par $m_{\mathrm{T}}^{\mathrm{tot}}$ dans la limite de l'incertitude à $\pm 1\sigma$ sur toute la gamme de masse étudiée. À basse masse, $m_{\Phi} < 180\,\mathrm{GeV}$ pour $gg\Phi$ en figure 5.41a et $m_{\Phi} < 250\,\mathrm{GeV}$ pour $b\bar{b}\Phi$ en figure 5.41b, m_{ML} est un peu plus performante que $m_{\mathrm{T}}^{\mathrm{tot}}$. Pour des masses plus élevées et jusqu'à 1 TeV, c'est en revanche $m_{\mathrm{T}}^{\mathrm{tot}}$ qui donne des résultats légèrement meilleurs. Au-delà, les deux variables ont les mêmes performances. Un bon estimateur de la masse d'une résonance n'est donc pas forcément une bonne variable discriminante.



(a) *Processus* $gg \to \Phi \to \tau\tau$.

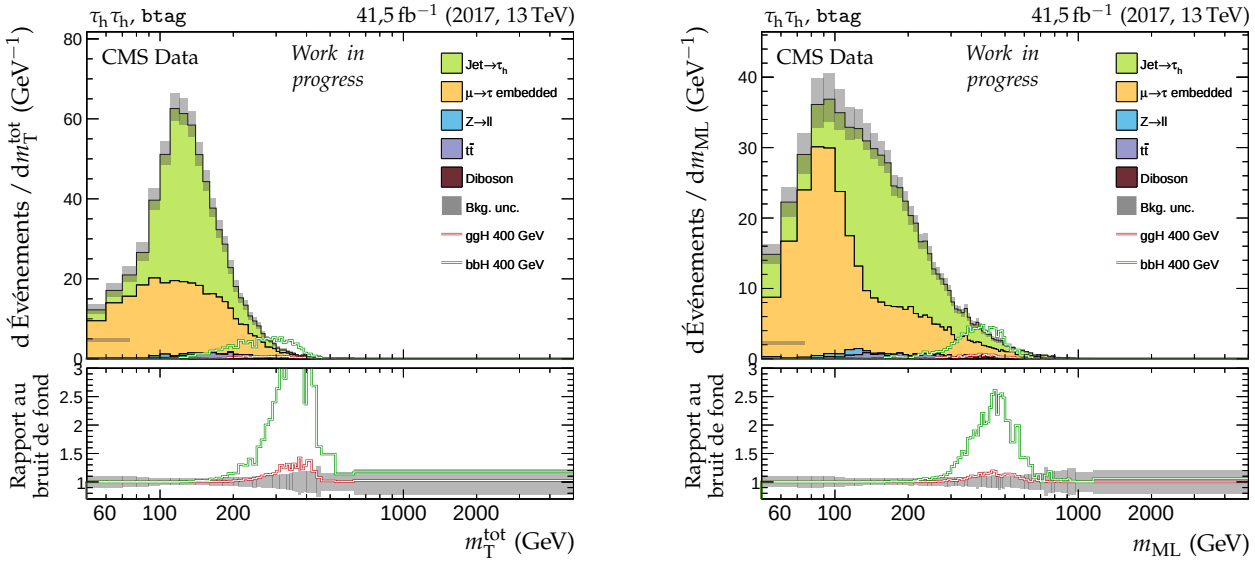


(b) *Processus* $gg \to b\bar{b}\Phi \to \tau\tau$.

Figure 5.41 – *Limites d'exclusion attendues avec* m_{ML} *en tant que variable discriminante* (Expected) *obtenues avec l'année 2017. Les limites obtenues avec* $m_{\mathrm{T}}^{\mathrm{tot}}$ *sont également données* (With $m_{\mathrm{T}}^{\mathrm{tot}}$).

8.1.3 Distributions de m_{ML} et de $m_{\mathrm{T}}^{\mathrm{tot}}$

Afin d'interpréter ces résultats, il est possible de se référer aux distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML}. La figure 5.42 les présente pour la catégorie `btag` du canal $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$. D'autres distributions sont disponibles dans l'annexe H.



(a) *Distribution de* $m_{\mathrm{T}}^{\mathrm{tot}}$.

(b) *Distribution de* m_{ML}.

Figure 5.42 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2017 dans la catégorie* `btag` *du canal* $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$. *Un signal correspondant à un boson de Higgs supplémentaire de masse* 400 GeV *est ajouté à titre d'illustration.*

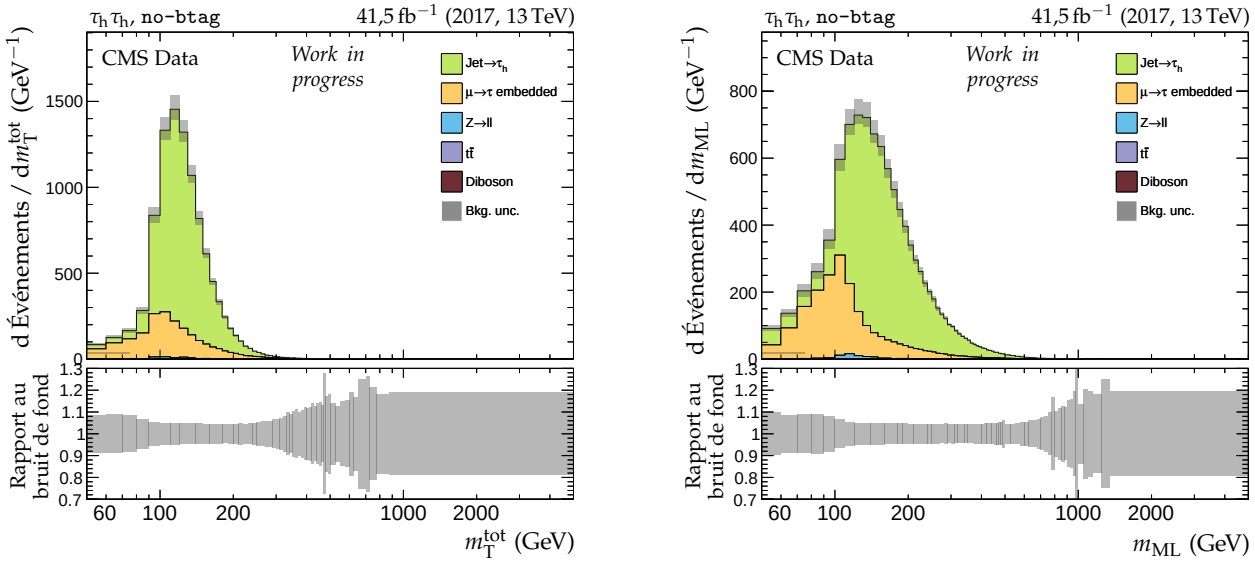
À haute masse, au-delà de 1 TeV, l'effet de la statistique est dominant à cause de la faible quantité d'événements présents. Les deux variables donnent alors des résultats équivalents. Pour les masses intermédiaires en revanche, l'avantage de $m_{\mathrm{T}}^{\mathrm{tot}}$ sur m_{ML} provient de la présence des faux τ_{h}, c'est-à-dire des événements dans lesquels au moins un jet est identifié à tort comme un τ_{h} et utilisé pour former le dilepton. De tels jets ne proviennent pas de la désintégration d'un boson de Higgs en paire de leptons τ. S'il est possible de définir une masse pour un dilepton formé avec un ou deux faux τ_{h}, la valeur obtenue n'a pas de sens physique, ces jets n'étant pas issus de la même désintégration.

Comme cela est visible sur la figure 5.42, le bruit de fond dû aux faux τ_{h} (Jet $\rightarrow \tau_{\mathrm{h}}$) forme une queue de distribution à haute masse plus importante avec m_{ML} qu'avec $m_{\mathrm{T}}^{\mathrm{tot}}$. Or, cette queue se retrouve dans la région de signal. Par exemple, selon $m_{\mathrm{T}}^{\mathrm{tot}}$, le signal d'un boson de Higgs supplémentaire de masse 400 GeV se trouve entre 120 et 500 GeV avec un maximum de 5 GeV^{-1} à 300 GeV. À cet endroit, le nombre d'événements de bruit de fond pour $m_{\mathrm{T}}^{\mathrm{tot}}$ est d'environ 3 GeV^{-1}. Selon m_{ML} en revanche, le même signal se trouve entre 200 et 650 GeV avec un maximum de 6 GeV^{-1} à 400 GeV. À cet endroit, le nombre d'événements de bruit de fond pour m_{ML} est d'environ 5 GeV^{-1}.

Bien que les valeurs données par m_{ML} soient plus proches de la valeur vraie de la masse du boson et que le signal présente une valeur maximale plus élevée qu'avec $m_{\mathrm{T}}^{\mathrm{tot}}$, la queue à haute masse due aux faux τ_{h} dégrade le rapport du signal au bruit de fond. Ce dernier est en effet plus faible avec m_{ML} qu'avec $m_{\mathrm{T}}^{\mathrm{tot}}$.

Un effet intéressant est par contre la meilleure séparation par m_{ML} que par $m_{\mathrm{T}}^{\mathrm{tot}}$ des événements des données encapsulées ($\mu \rightarrow \tau$ embedded), introduites dans le chapitre 4, vis-à-vis des faux τ_{h}. Les données encapsulées décrivent principalement le boson Z et une fraction d'événements $t\bar{t}$. Dans le cas de $m_{\mathrm{T}}^{\mathrm{tot}}$, leur distribution est largement étalée jusqu'à 300 GeV, avec un maximum d'événements de 20 GeV^{-1} à 95 GeV. La distribution des valeurs de m_{ML} sur ces mêmes événements exhibe deux composantes. La première forme un pic avec un maximum d'événements de 30 GeV^{-1} à 90 GeV, il s'agit du signal du boson Z. La seconde se situe à plus haute masse, jusqu'à 600 GeV, avec un maximum d'événements de 7 GeV^{-1} à environ 200 GeV. Ce groupe peut correspondre à la composante $t\bar{t}$ des données encapsulées. Pour le vérifier, la figure 5.43 montre les distributions de $m_{\mathrm{T}}^{\mathrm{tot}}$ et m_{ML} pour

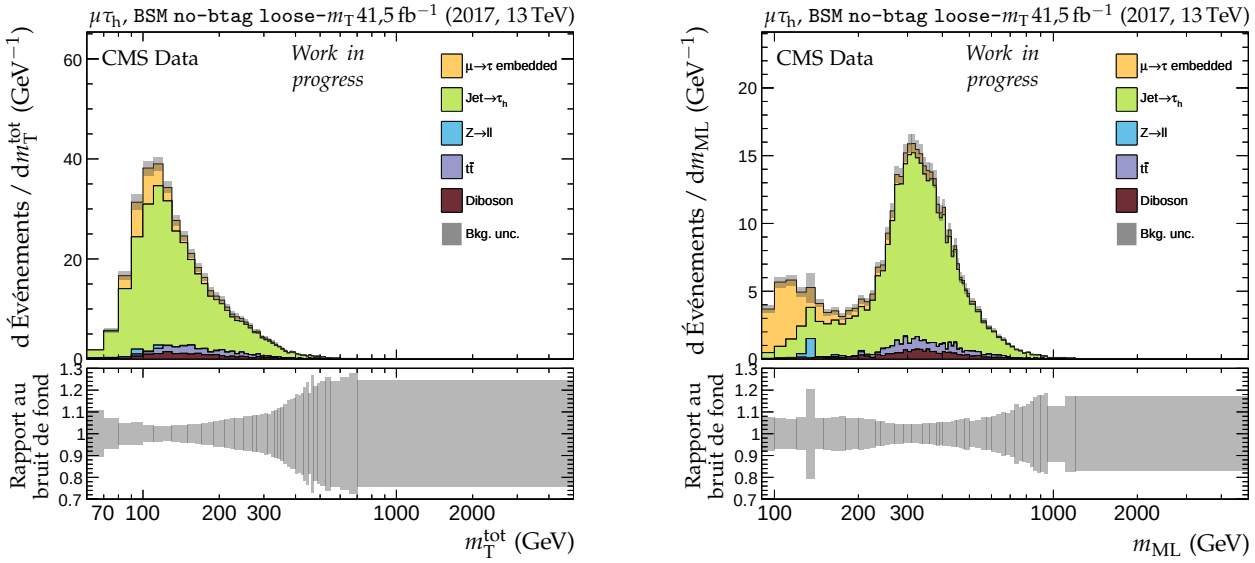
la catégorie `no-btag` du canal $\tau_h\tau_h$, où l'exclusion des événements $t\bar{t}$ est réalisée par le rejet des jets issus de quarks b. La distribution de m_{ML} dans les données encapsulées ne montre pas deux parties comme dans la catégories `btag`, ce qui confirme cette hypothèse. Notre modèle est donc en mesure de séparer les composantes du boson Z et $t\bar{t}$ des données encapsulées du canal $\tau_h\tau_h$, bien qu'il n'ait pas été entraîné dans ce but, alors que m_T^{tot} ne le permet pas.



(a) *Distribution de* m_T^{tot}. **(b)** *Distribution de* m_{ML}.

Figure 5.43 – *Distributions de* m_T^{tot} *et* m_{ML} *en 2017 dans la catégorie* `no-btag` *du canal* $\tau_h\tau_h$.

Notre modèle permet de plus de mieux séparer les événements du boson Z des faux τ_h. Par exemple, les distributions de m_T^{tot} et m_{ML} sont données dans la catégorie `BSM no-btag loose-`m_T pour l'année 2017 en figure 5.44. La composante du boson Z apparaît nettement avec m_{ML} aux alentours de 100 GeV, alors qu'elle est confondue avec les faux τ_h d'après m_T^{tot}. Or, la sélection des événements pour les catégories `BSM`, introduite au chapitre 4, contient en particulier la coupure $m_{SVFIT} \geqslant 250$ GeV, avec m_{SVFIT} la masse prédite par l'algorithme SVFIT [229]. Notre modèle identifie donc des événements correspondant au boson Z, de masse 91,2 GeV, alors que SVFIT les estime au-delà de 250 GeV.



(a) *Distribution de* m_T^{tot}. **(b)** *Distribution de* m_{ML}.

Figure 5.44 – *Distributions de* m_T^{tot} *et* m_{ML} *en 2017 dans la catégorie* `BSM no-btag loose-`m_T *du canal* $\mu\tau_h$.

8.2 Comparaison de m_{ML} à m_{SVFIT}

Au sein de la collaboration CMS, l'algorithme SVFIT [229] a le même objectif que notre modèle B", estimer la masse d'une particule se désintégrant en paire de leptons τ, bien que la méthode employée soit différente. Cependant, nous avons observé précédemment que certains événements estimés au-delà de 250 GeV par SVFIT semblent correspondre au boson Z de masse 91,2 GeV selon notre modèle. Nous avons donc souhaité comparer les valeurs de m_{ML} prédites par B" à celles de m_{SVFIT} fournies par SVFIT.

8.2.1 Distributions inclusives

La figure 5.45 montre les distributions inclusives, c'est-à-dire sur les événements de toutes les catégories confondues, de m_{SVFIT} et m_{ML} en 2017 dans les canaux $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $e\tau_{\mathrm{h}}$ et $e\mu$. Ces distributions sont deux à deux très similaires, ce qui est attendu car m_{SVFIT} comme m_{ML} sont deux estimateurs de la même grandeur physique. Les faux τ_{h} sont toutefois prédits à des masses légèrement plus élevées par B" que par SVFIT. Il en résulte des queues de distribution à haute masse plus importantes avec m_{ML}, effet déjà observé lors de la comparaison de m_{ML} à $m_{\mathrm{T}}^{\mathrm{tot}}$. Ces comparaisons sur l'ensemble des topologies d'événements d'un canal ne permet pas de comparer B" et SVFIT plus en détail. C'est pourquoi des comparaisons sur des sous-ensembles d'événements sont réalisées dans les sections suivantes.

8.2.2 Distributions des événements de haute masse

D'après la figure 5.44, notre modèle identifie des événements correspondant au boson Z parmi ceux que SVFIT estime au-delà de 250 GeV. Il y a donc un désaccord entre les deux estimateurs. Toutefois, SVFIT pourrait aussi trouver un signal à basse masse parmi les événements estimés à haute masse par B". Pour le vérifier, les distributions de m_{ML} sur les événements tels que $m_{\mathrm{SVFIT}} > 250$ GeV sont représentées en figure 5.46 pour les canaux $\mu\tau_{\mathrm{h}}$, $e\tau_{\mathrm{h}}$ et $e\mu$ en 2017, en regard des distributions de m_{SVFIT} sur les événements tels que $m_{\mathrm{ML}} > 250$ GeV.

Les distributions de m_{ML} confirment que B" parvient à trouver des événements correspondant au boson Z alors que SVFIT les estime au-delà de 250 GeV. Les distributions de m_{SVFIT} en revanche ne font apparaître aucun signal en particulier parmi les événements estimés au-delà de 250 GeV par B". Des investigations plus poussées peuvent être menées en sélectionnant des topologies particulières d'événements.

8.2.3 Distributions par topologies

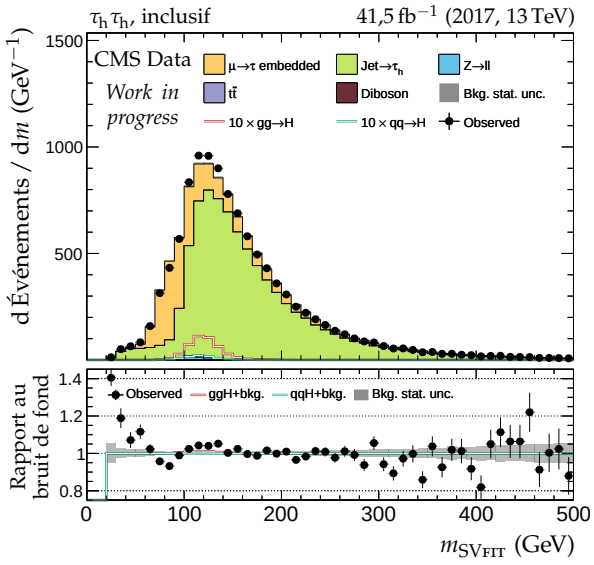
Les distributions de m_{SVFIT} et m_{ML} sont comparées dans le cas de topologies spécifiques d'événements. Ces topologies correspondent à certaines catégories de l'analyse du boson de Higgs du modèle standard dans sa désintégration en paire de leptons τ n'utilisant pas de réseau de neurones pour obtenir une variable discriminante [36, 230].

La figure 5.47 correspond aux événements du canal $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ sans jets et tels que la distance ΔR entre les deux éléments du dilepton dans le plan (η, ϕ) soit inférieure à 3,2. Avec m_{SVFIT}, le signal du boson Z est situé au-dessus de 100 GeV et est complètement mélangé avec les faux τ_{h}. En revanche, le passage à m_{ML} a pour effet de décaler les faux τ_{h} d'une dizaine de GeV vers les hautes valeurs alors que le signal du Z est en partie estimé à plus basse masse.

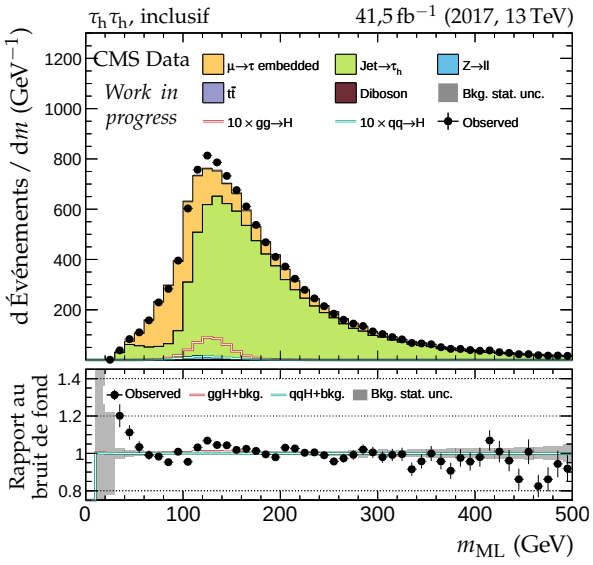
Cet effet est encore plus visible dans le cas des événements vérifiant l'une des deux conditions suivantes :

— moins de deux jets et tels que $\Delta R \geqslant 3{,}2$;
— au moins deux jets et tels que $\Delta R \geqslant 2{,}5$;

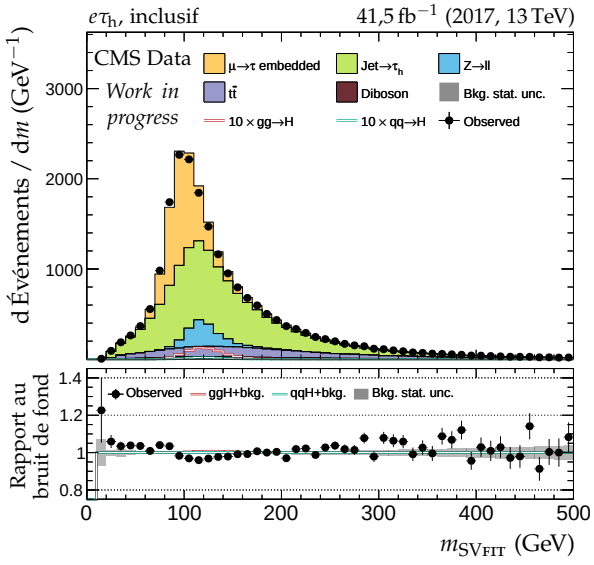
dont les distributions de m_{SVFIT} et m_{ML} sont données en figure 5.48. Dans ce cas, la distribution de m_{SVFIT} sur les données encapsulées est complètement confondue avec celle des faux τ_{h}, tandis que les valeurs de m_{ML} font apparaître un signal aux alentours de 100 GeV correspondant au boson Z, sa masse étant donc surestimée de 10 GeV environ.



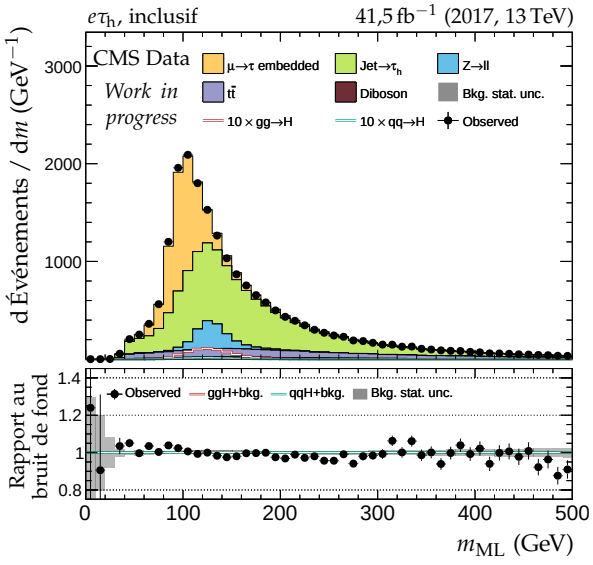
(a) m_{SVFIT}, *canal* $\tau_h\tau_h$.



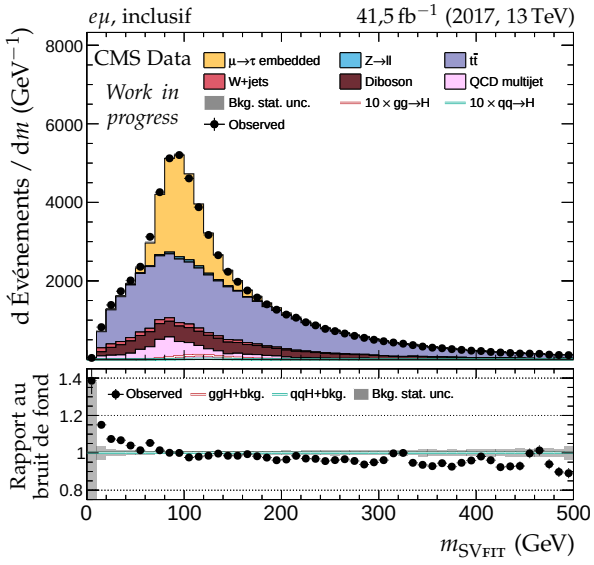
(b) m_{ML}, *canal* $\tau_h\tau_h$.



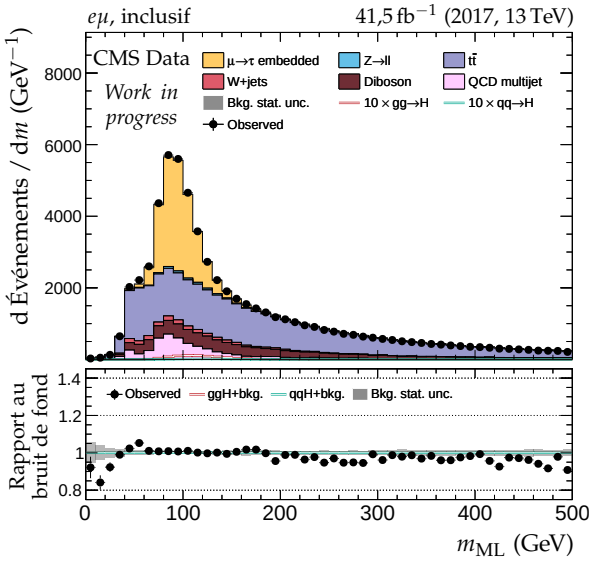
(c) m_{SVFIT}, *canal* $e\tau_h$.



(d) m_{ML}, *canal* $e\tau_h$.

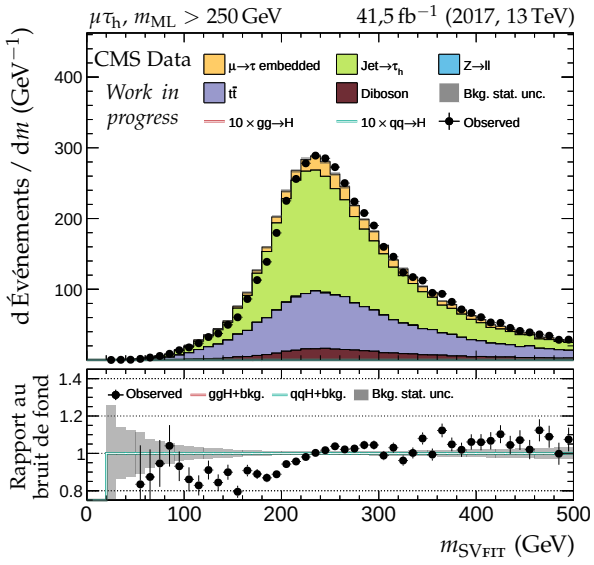


(e) m_{SVFIT}, *canal* $e\mu$.

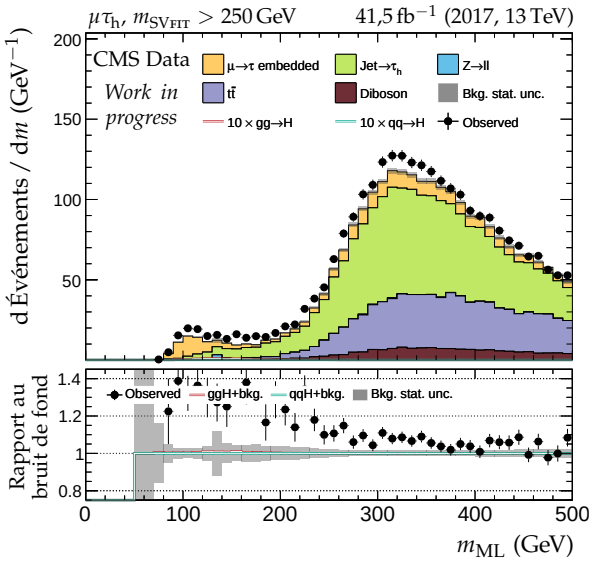


(f) m_{ML}, *canal* $e\mu$.

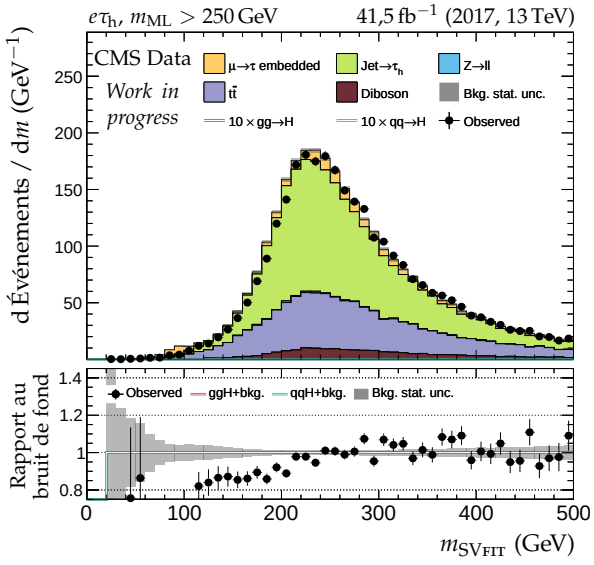
Figure 5.45 – *Distributions de* m_{SVFIT} *et* m_{ML} *en 2017 dans les canaux* $\tau_h\tau_h$, $e\tau_h$ *et* $e\mu$.



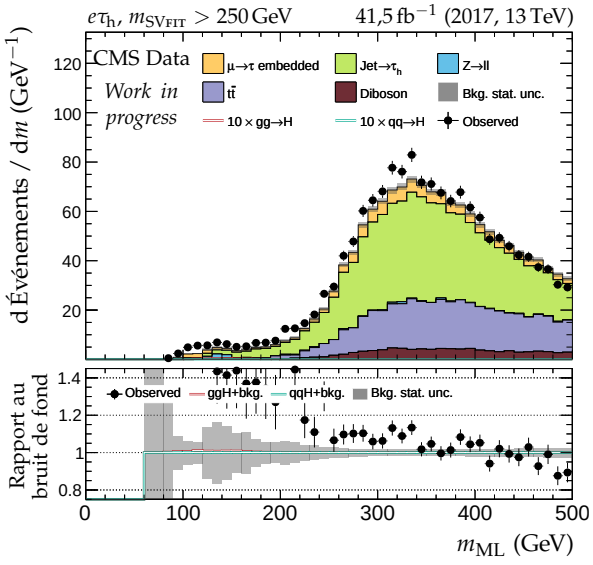
(a) m_{SVFIT} *pour* $m_{\text{ML}} > 250\,\text{GeV}$, *canal* $\mu\tau_{\text{h}}$.



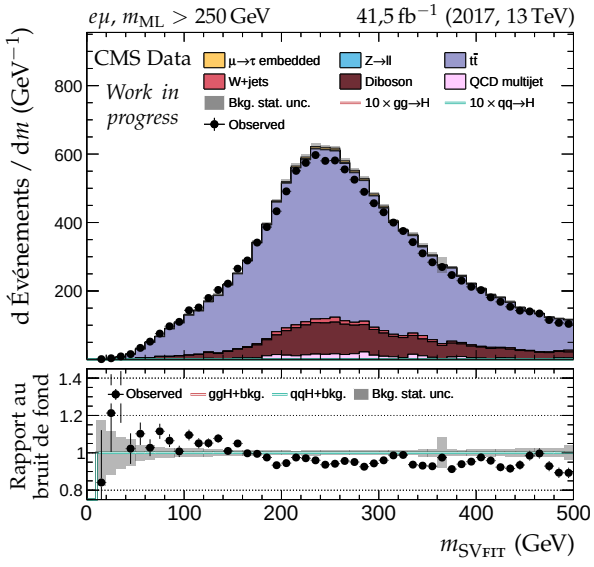
(b) m_{ML} *pour* $m_{\text{SVFIT}} > 250\,\text{GeV}$, *canal* $\mu\tau_{\text{h}}$.



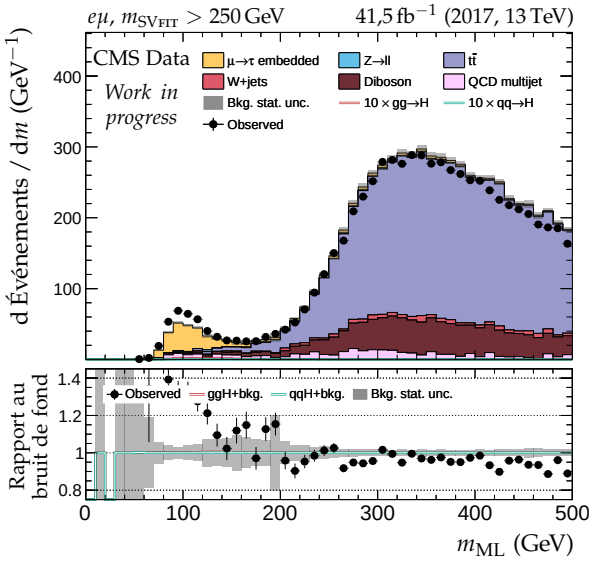
(c) m_{SVFIT} *pour* $m_{\text{ML}} > 250\,\text{GeV}$, *canal* $e\tau_{\text{h}}$.



(d) m_{ML} *pour* $m_{\text{SVFIT}} > 250\,\text{GeV}$, *canal* $e\tau_{\text{h}}$.

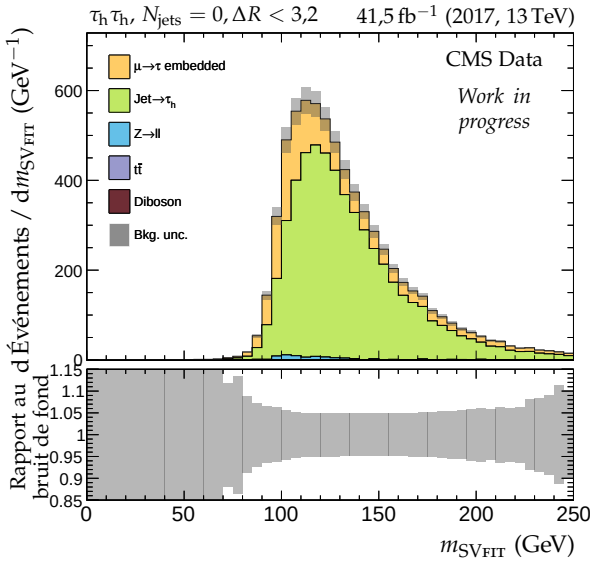


(e) m_{SVFIT} *pour* $m_{\text{ML}} > 250\,\text{GeV}$, *canal* $e\mu$.

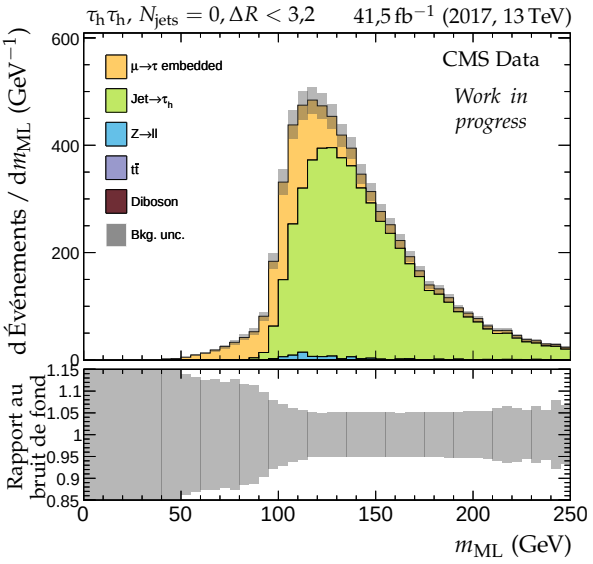


(f) m_{ML} *pour* $m_{\text{SVFIT}} > 250\,\text{GeV}$, *canal* $e\mu$.

Figure 5.46 – *Distributions de* m_{SVFIT} *et* m_{ML} *en 2017 sur les événements de haute masse des canaux* $\mu\tau_{\text{h}}$, $e\tau_{\text{h}}$ *et* $e\mu$.

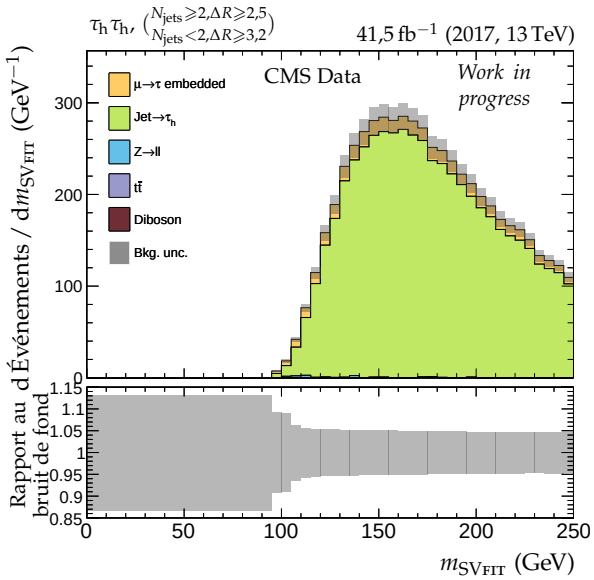


(a) m_{SVFIT}, *canal* $\tau_h\tau_h$.

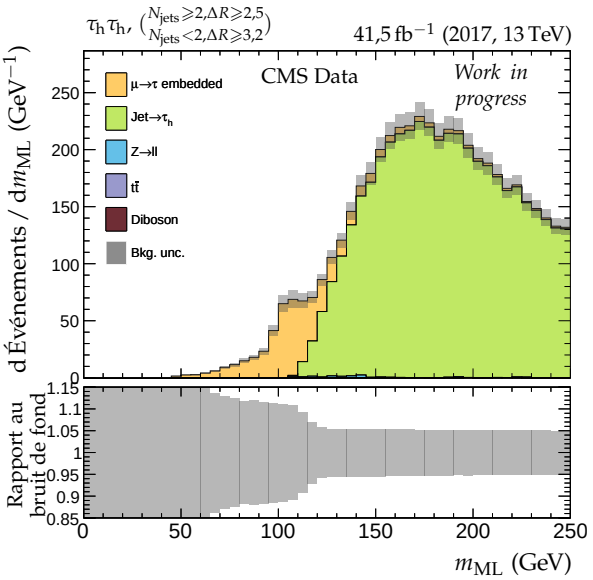


(b) m_{ML}, *canal* $\tau_h\tau_h$.

Figure 5.47 – *Distributions de* m_{SVFIT} *et* m_{ML} *en 2017 dans le canal* $\tau_h\tau_h$ *avec les événements sans jets et tels que la distance* ΔR *entre les deux éléments du dilepton dans le plan* (η, ϕ) *soit inférieure à* 3,2.



(a) m_{SVFIT}, *canal* $\tau_h\tau_h$.



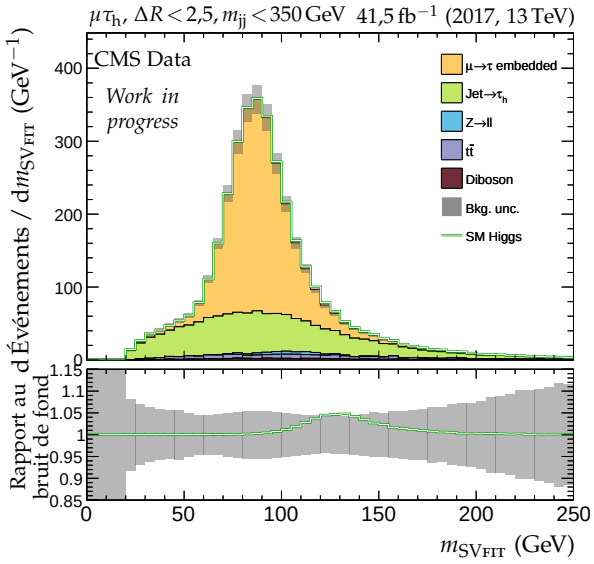
(b) m_{ML}, *canal* $\tau_h\tau_h$.

Figure 5.48 – *Distributions de* m_{SVFIT} *et* m_{ML} *en 2017 dans le canal* $\tau_h\tau_h$ *avec les événements contenant moins de deux jets et tels que la distance* ΔR *entre les deux éléments du dilepton dans le plan* (η, ϕ) *soit supérieure à* 3,2 *ou contenant au moins deux jets et* $\Delta R \geqslant 2{,}5$.

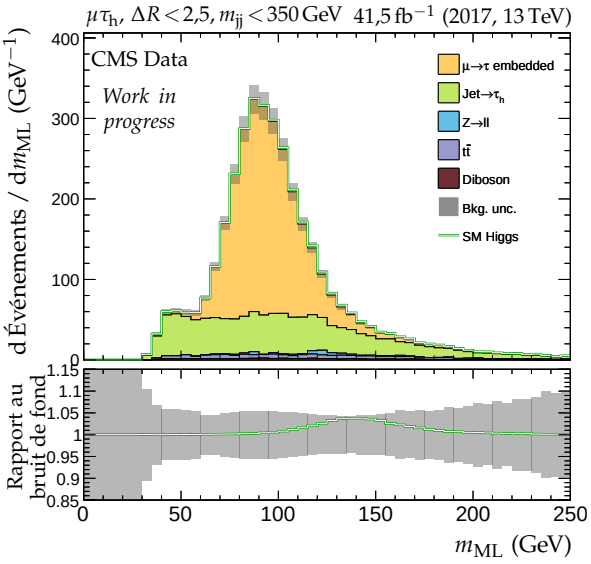
La sensibilité de B" au boson de Higgs du modèle standard h est comparable avec celle de SVFIT. Par exemple, avec les événements du canal $\mu\tau_h$ contenant au moins deux jets avec la masse invariante des deux jets principaux m_{jj} inférieure à 350 GeV et tels que la distance ΔR entre les deux éléments du dilepton dans le plan (η, ϕ) soit inférieure à 2,5, les distributions de m_{SVFIT} et m_{ML} en figure 5.49 montrent un rapport au bruit de fond similaire pour le signal associé à h.

Lorsque m_{jj} est supérieure à 1 TeV en revanche, en figure 5.50, la distribution du signal de h est plus étalée avec m_{ML} qu'avec m_{SVFIT}. Il en résulte une sensibilité diminuée.

Notre modèle propose donc des estimations de la masse des particules se désintégrant en paires de leptons τ comparables à SVFIT, actuellement utilisé au sein de la collaboration CMS. Selon les topologies des événements considérés, la description du boson de Higgs h est similaire. Celle du boson Z est parfois meilleure avec notre modèle.

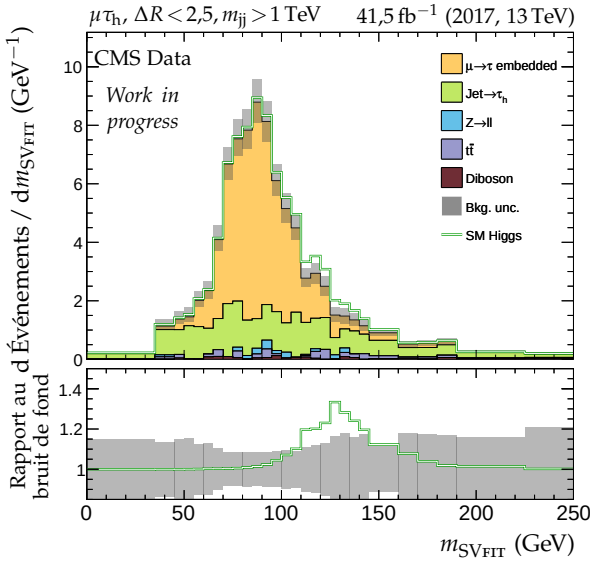


(a) m_{SVFIT}, *canal* $\mu\tau_{\text{h}}$.

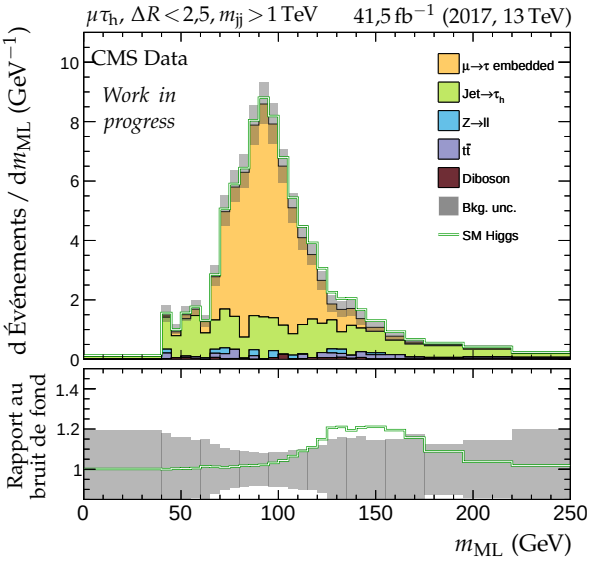


(b) m_{ML}, *canal* $\mu\tau_{\text{h}}$.

Figure 5.49 – *Distributions de* m_{SVFIT} *et* m_{ML} *en 2017 dans le canal* $\mu\tau_{\text{h}}$ *avec les événements contenant au moins deux jets avec la masse invariante des deux jets principaux* m_{jj} *inférieure à* 350 GeV *et tels que la distance* ΔR *entre les deux éléments du dilepton dans le plan* (η, ϕ) *soit inférieure à* 2,5.



(a) m_{SVFIT}, *canal* $\mu\tau_{\text{h}}$.



(b) m_{ML}, *canal* $\mu\tau_{\text{h}}$.

Figure 5.50 – *Distributions de* m_{SVFIT} *et* m_{ML} *en 2017 dans le canal* $\mu\tau_{\text{h}}$ *avec les événements contenant au moins deux jets avec la masse invariante des deux jets principaux* m_{jj} *supérieure à* 1 TeV *et tels que la distance* ΔR *entre les deux éléments du dilepton dans le plan* (η, ϕ) *soit inférieure à* 2,5.

8.2.4 Vitesses d'exécution

L'algorithme de SVFIT est connu pour être relativement long en termes de temps d'exécution. Cela provient de sa démarche qui consiste à réaliser un ajustement de la valeur de m_{SVFIT} aux observables de chaque événement [229]. Dans le cas d'un DNN, l'ajustement en fonction des variables d'entrées est réalisé une fois pour toutes lors de l'entraînement du modèle. Le DNN B" a été entraîné en 15 minutes environ. Puis, l'obtention des prédictions m_{ML} de B" consiste à appliquer la fonction F de B" aux variables d'entrée. La figure 5.51 compare les temps d'exécution moyens par événement de deux scripts, l'un en C++ pour obtenir les valeurs de m_{SVFIT}, l'autre en Python pour obtenir celles de m_{ML}.

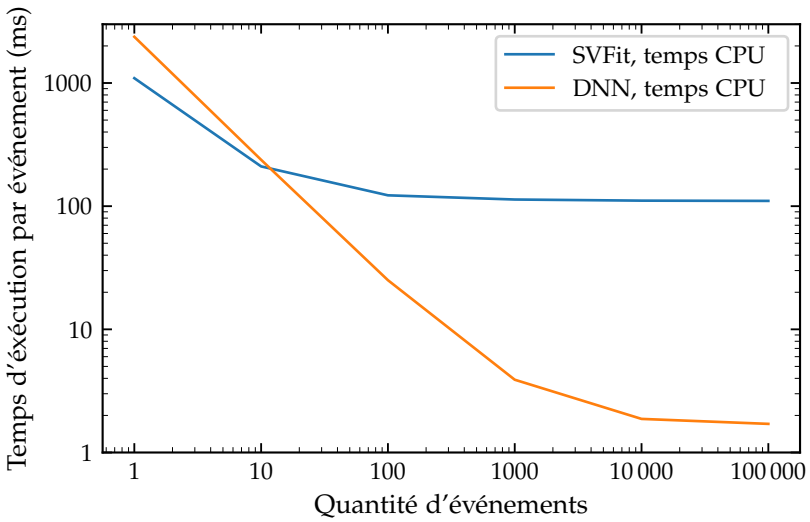


Figure 5.51 – *Temps nécessaire à l'obtention des prédictions de masse avec* SVFIT *et avec le DNN B".*

Pour 10 événements ou moins à traiter, le script de SVFIT est plus rapide que celui du DNN. Ceci est dû au langage utilisé ainsi qu'au temps de chargement du DNN. Cependant, une quantité si faible d'événements à traiter est rarement rencontrée dans les analyses du Run II à CMS dans lesquelles les événements se comptent au moins en millions. Sur la figure 5.51, le temps moyen par événement pour chacun des deux scripts stagne à partir de 100 événements avec SVFIT et 10 000 avec le DNN. Le script de SVFIT nécessite alors 110 ms par événement contre seulement 1,71 ms avec le DNN.

Les prédictions de notre modèle sont donc obtenues 60 fois plus rapidement que celles de SVFIT. Dans le contexte des futurs *runs* du LHC et surtout du HL-LHC, l'augmentation de la luminosité et donc du nombre d'événements à traiter rend l'utilisation d'un DNN à la place de SVFIT pertinente dans l'optique de minimiser les temps d'exécution des analyses.

9 Conclusion

Le *machine learning* (ML) est une branche de l'intelligence artificielle permettant d'obtenir des modèles pouvant réaliser des classifications ou des régressions. De tels modèles sont déjà exploités en physique des particules afin de réaliser diverses tâches, comme l'identification des jets issus de quarks b par exemple.

Nous avons étudié la possibilité de prédire la masse d'une résonance se désintégrant en paire de leptons τ grâce au ML. En effet, la phénoménologie de ces événements ne permet pas d'obtenir la masse invariante totale du système dans l'état final. Les travaux réalisés par BÄRTSCHI & coll. sur ce sujet donnent des résultats prometteurs, mais l'empilement n'y est pas pris en compte et la modélisation du détecteur CMS approximée. À partir d'événements que nous avons nous-mêmes simulés à l'aide de FASTSIM, l'empilement a été pris en compte et le détecteur modélisé plus précisément.

Nous avons construit et entraîné des arbres de décision améliorés à l'aide de la librairie XGBOOST et des réseaux de neurones profonds à l'aide des librairies KERAS et TENSORFLOW. Le principe et l'entraînement de ces types de modèle ont été présentés. De nombreuses combinaisons d'hyper-paramètres, propriétés des modèles régissant par exemple leur structure interne, ont été étudiées et comparées. Il en ressort que certaines variables d'entrée sont des informations pertinentes afin d'estimer plus fidèlement la masse de la résonance. Les réseaux de neurones proposent de meilleures performances que les arbres de décision améliorés d'après les métriques d'évaluation que nous avons utilisées, en particulier pour les valeurs de masse de la résonance correspondant aux bosons Z et h du modèle standard. Les performances des réseaux de neurones dépendent également fortement de l'algorithme d'optimisation utilisé lors de l'entraînement.

Nous avons alors déterminé une combinaison performante d'hyper-paramètres correspondant au modèle B. Divers effets sur ses prédictions ont été étudiés. Ainsi, l'empilement doit être pris en

compte lors de l'entraînement afin d'obtenir un modèle pertinent pour les analyses.

Un effet majeur sur la précision des prédictions est lié à la reconstruction des particules. Dans le cas d'une reconstruction parfaite des particules, c'est-à-dire en utilisant les objets générés correspondants au lieu des objets reconstruits, la résolution relative sur la masse de la résonance est de 3 % contre 20 à 25 % sinon. De plus, les faux τ_h perturbent les prédictions des modèles à basse masse, en particulier dans la région des bosons Z et h. Cependant, l'entraînement de modèles spécifiques aux différents canaux ou aux différentes phénoménologies de canaux n'apporte pas de gain en termes de précision des prédictions. L'utilisation de la PFMET au lieu de la PUPPIMET a un effet négligeable sur les prédictions du modèle face à sa résolution.

La gamme de masse explorée lors de l'entraînement définit la zone utile du modèle, ses prédictions étant en bonne approximation restreintes à cet intervalle. Cependant, il n'est pas possible d'étendre cet intervalle à l'infini et des effets de bord apparaissent sur les prédictions du modèle. Nous avons modifié la fonction de coût afin de rejeter dynamiquement certains événements de l'entraînement pour réduire cet effet de bord avec succès. L'exploitation de la queue à hautes valeurs de la distribution de la masse de la résonance, objectif des prédictions des modèles, a permis d'améliorer encore les prédictions moyennes obtenues avec le modèle B''. Ce modèle permet de reconstruire avec succès la masse d'une résonance se désintégrant en paire de leptons τ entre 50 GeV et 800 GeV avec une précision de 20 à 25 %.

La prédiction du modèle B'', m_{ML}, a été utilisée en tant que variable discriminante à la place de $m_{\mathrm{T}}^{\mathrm{tot}}$ pour obtenir les limites d'exclusion indépendantes du modèle de l'analyse présentée dans le chapitre 4 sur l'année 2017. Malgré des valeurs plus proches de la vraie masse de la résonance ainsi qu'une meilleure résolution que $m_{\mathrm{T}}^{\mathrm{tot}}$, m_{ML} ne permet pas de repousser les limites d'exclusion obtenues. Ceci est dû aux processus physiques tels que les faux τ_h ne correspondant pas à une résonance se désintégrant en paire de leptons τ mais passant les critères de sélection des événements appliqués. L'utilisation de m_{ML} en tant que variable discriminante pour la recherche de bosons de Higgs supplémentaires de haute masse n'est donc pas pertinente. D'autres analyses peuvent en revanche bénéficier de ce projet. Son utilisation dans un autre but que d'obtenir une variable discriminante est également envisageable, par exemple pour la sélection des événements.

Notre modèle permet en effet de mieux séparer les événements $Z \to \tau\tau$ des faux τ_h que $m_{\mathrm{T}}^{\mathrm{tot}}$. De plus, une comparaison des valeurs de m_{ML} à celles de m_{SVFIT}, obtenues par l'algorithme SVFIT déjà utilisé par la collaboration CMS, a permis de mettre en lumière une meilleure description du boson Z par notre modèle. En effet, certains événements $Z \to \tau\tau$ sont prédits au-delà de 250 GeV par SVFIT, alors que la masse du Z est de 91,2 GeV. Parmi ces événements, notre modèle en prédit aux alentours de 100 GeV, ce qui est plus proche de la valeur vraie. L'effet inverse, c'est-à-dire des événements tels que $m_{\mathrm{ML}} > 250\,\mathrm{GeV}$ alors que $m_{\mathrm{SVFIT}} \simeq 100\,\mathrm{GeV}$ pour le boson Z, n'est pas observé. La sensibilité au boson de Higgs du modèle standard h est similaire avec m_{SVFIT} ou m_{ML} dans certaines topologies d'événements. Pour d'autres, le signal de h est plus étendu avec m_{ML} qu'avec m_{SVFIT}, donnant une sensibilité moindre. L'utilisation de processus physiques plus variés pour entraîner les modèles pourrait améliorer leurs prédictions sur de telles topologies. Une résolution similaire voire meilleure que celle de SVFIT est donc envisageable avec des réseaux de neurones.

Le temps nécessaire pour obtenir les prédictions de masse est 60 fois plus court avec notre modèle qu'avec SVFIT. Les futures analyses de la collaboration CMS seront basées sur de plus grandes quantité d'événements, l'utilisation des réseaux de neurones au lieu de SVFIT présente donc un intérêt certain afin de minimiser leur coût computationnel.

Le modèle B'' développé au cours de ma thèse [266] peut être récupéré [267] et utilisé dans d'autres analyses. Le groupe en charge de l'analyse des événements avec une paire de leptons τ dans le cadre du NMSSM (*Next to MSSM*), modèle contenant sept bosons de Higgs contre cinq dans le MSSM introduit au chapitre 1, a déjà manifesté un intérêt pour notre modèle. De même, l'utilisation de B'' pour l'analyse des événements $hh \to b\bar{b}\tau\tau$, c'est-à-dire avec deux bosons de Higgs dont l'un se désintègre en paire de quarks b et l'autre en paire de leptons τ, est déjà étudiée. La topologie de ces événements, différente de ceux contenant uniquement $h \to \tau\tau$, permet de tester notre modèle dans des situations inédites. Les résultats obtenus à cette occasion pourront permettre d'améliorer le modèle et d'étendre son domaine d'application.

Conclusion et perspectives

L'extension supersymétrique minimale du modèle standard (MSSM) introduite au chapitre 1 est une des théories visant à combler des lacunes du modèle standard (SM), par exemple l'absence de la matière noire. Dans le MSSM, quatre bosons de Higgs sont présents en plus de celui prédit par le SM (h), deux chargés ($H^{\pm}$) et deux neutres (H, A). La phénoménologie de ces derniers dans les collisionneurs tels que le LHC favorise le canal de désintégration en paire de leptons τ ($H/A \to \tau\tau$) pour la recherche d'un signal indiquant leur présence, qui serait une preuve de l'existence d'une nouvelle physique au-delà du SM (BSM).

Les travaux réalisés lors de cette thèse exploitent les données expérimentales récoltées par le détecteur CMS lors des collisions de protons du Run II du LHC à une énergie dans le centre de masse de 13 TeV, correspondant à une luminosité intégrée de 137 fb^{-1}. Le détecteur CMS a été présenté dans le chapitre 2. Les particules issues des collisions y laissent des signaux lors de leur passage, utilisés pour reconstruire les particules correspondantes. Les propriétés de ces dernières permettent d'étudier les processus physiques ayant eu lieu lors de la collision initiale.

La signature expérimentale d'un parton issu de la collision, c'est-à-dire un quark ou un gluon, est un jet. Il s'agit d'un objet physique complexe composé de plusieurs particules. De nombreuses analyses de la collaboration CMS exploitent les jets et les incertitudes liées à leur énergie sont généralement parmi les plus importantes. La bonne caractérisation de ces objets physiques est donc cruciale. Dans cette thèse, les événements contenant un photon et un jet (γ + jet) sont utilisés afin de calibrer les jets en énergie. Les résultats obtenus lors de cette thèse pour les années 2018 et 2017-UL, présentés dans le chapitre 3, sont exploités dans la calibration officielle de la collaboration.

La recherche de bosons de Higgs supplémentaires de haute masse se désintégrant en paire de τ a été présentée dans le chapitre 4. L'estimation des bruits de fond dus aux processus du SM est majoritairement obtenue à partir des données observées elles-mêmes. Plusieurs catégories d'événements sont définies afin d'augmenter la sensibilité de l'analyse à H et A. Aucune déviation significative par rapport aux prédictions du SM n'est observée dans les distributions des variables discriminantes des différentes catégories. Des limites hautes ont alors été déterminées par la méthode CL_s sur la section efficace de production de H et A, multipliée par leur rapport de branchement à la désintégration en paire de leptons τ. Une région d'exclusion du MSSM en faveur du SM, déterminée dans le plan $(m_A, \tan\beta)$ pour le scénario M_h^{125}, permet d'exclure $m_A < 600\,\text{GeV}$. Dans le cas du scénario $M_{H_1}^{125}(\text{CPV})$, une violation de CP par les bosons de Higgs est possible. Les états propres de masse sont H_1, H_2 et H_3 au lieu de h, H et A, c'est pourquoi la région d'exclusion est donnée dans le plan $(m_{H^{\pm}}, \tan\beta)$ et H_1 doit jouer le rôle du boson découvert en 2012. Les valeurs de $m_{H^{\pm}}$ inférieures à 400 GeV sont exclues.

Les neutrinos issus des désintégrations des leptons τ sont invisibles dans les détecteurs du LHC. L'incertitude inhérente due à l'impossibilité de les mesurer dégrade la précision et les performances des analyses. Les travaux présentés dans le chapitre 5 ont permis d'obtenir un réseau de neurones profond (DNN) capable de reconstruire la masse d'une particule se désintégrant en paire de leptons τ entre 50 GeV et 800 GeV avec une précision de 20 à 25 %. Une avancée majeure par rapport aux précédentes études est l'inclusion de l'empilement dans l'entraînement du DNN qui peut alors être directement exploité dans les analyses de CMS.

Malgré une meilleure résolution sur le signal de H et A que $m_{\text{T}}^{\text{tot}}$, la variable discriminante actuellement utilisée dans l'analyse du chapitre 4, ce modèle ne permet pas d'étendre les limites d'exclusion obtenues. Cela est dû aux contributions des fausses paires de leptons τ, dont l'estimation de la masse par le DNN se retrouve dans la région de signal du MSSM. Un bon estimateur de masse n'est donc

pas forcément une bonne variable discriminante. Cependant, ce DNN permet de mieux séparer le signal du boson Z de celui des processus $t\bar{t}$ que $m_{\mathrm{T}}^{\mathrm{tot}}$. Il pourrait donc déjà bénéficier à d'autres analyses. De plus, une comparaison de ses estimations de masse à celles de l'algorithme SVFIT utilisé par la collaboration CMS montre une meilleure description du boson Z et une sensibilité au boson de Higgs du modèle standard équivalente dans certaines topologies d'événements, tout en étant 60 fois plus rapide à fournir ces estimations. Avec l'augmentation considérable de la quantité d'événements à traiter dans les futures analyses de physique des particules, notamment durant le Run III du LHC et au HL-LHC, la rapidité du DNN par rapport à SVFIT est un atout majeur.

L'entraînement du DNN est basé sur des événements dans lesquels un boson de Higgs $\mathcal{H}$ de masse variable est produit par fusion de gluons. L'exploitation d'autres modes de production pourrait améliorer les performances du DNN sur les topologies correspondantes. Il est de plus envisageable d'inclure des événements correspondant expressément au boson Z et aux modes de production associés, en plus de ceux pour lesquels $m_{\mathcal{H}} \simeq m_Z$, pour augmenter encore le pouvoir discriminant du DNN déjà plus important que celui de $m_{\mathrm{T}}^{\mathrm{tot}}$ dans cette région de l'espace des phases. La prise en compte des fausses paires de leptons τ dans l'entraînement n'est pas triviale car il n'est pas possible de définir une valeur physique de masse à prédire dans ce cas.

L'utilisation du DNN issu des travaux de cette thèse est déjà envisagée dans deux analyses de la collaboration CMS. La première est l'analyse des événements avec une paire de leptons τ dans le cadre du NMSSM (*Next to MSSM*), théorie plus complexe que le MSSM dans laquelle existent sept bosons de Higgs au lieu de cinq. La topologie des événements y est la même que dans l'analyse menée dans cette thèse, l'implémentation de notre modèle est donc directe. La seconde porte sur les événements avec deux bosons de Higgs du SM (h), l'un se désintégrant en paire de τ et l'autre en paire de quarks b ($hh \to b\bar{b}\tau\tau$). L'utilisation du DNN y est plus complexe à cause de la présence des deux jets de quarks b, situation inédite par rapport aux travaux menés dans cette thèse. La prise en compte de ces jets dans les variables d'entrée du réseau de neurones n'est pas triviale et les études actuelles menées par le groupe en charge de cette analyse tendent à montrer une dégradation des performances par rapport au simple cas $\mathcal{H} \to \tau\tau$ pour lequel il est conçu. Comme exprimé précédemment, inclure des événements $hh \to b\bar{b}\tau\tau$ dans l'entraînement du DNN pourrait remédier à cette dégradation. Une piste prometteuse est d'étudier l'effet du boson de Higgs se désintégrant en $b\bar{b}$ sur le recul de celui donnant la paire de τ.

Dans l'optique de l'analyse « résonnante » des événements $hh \to b\bar{b}\tau\tau$, c'est-à-dire lorsque les deux bosons h sont eux-mêmes issus d'une seule particule initiale, il est envisageable de créer un autre DNN spécialement conçu pour prédire directement la masse de celle-ci. Par rapport au DNN utilisé pour les événements $\mathcal{H} \to \tau\tau$, la liste des variables d'entrée pourrait être complétée par des observables liées aux quarks b et aux jets résultants.

L'utilisation d'un DNN, ou plus généralement du *machine learning*, afin d'estimer la masse d'une résonance se désintégrant en paire de leptons τ est ainsi très prometteuse. Les analyses futures, en particulier lors du Run III dès 2022, pourraient bénéficier d'une meilleure résolution et d'une vitesse accrue par rapport aux outils actuels. Ces avancées pourraient mener à des mesures encore plus précises des paramètres du SM, à mieux contraindre les modèles BSM tels que le MSSM et le NMSSM voire à l'observation d'une nouvelle physique. De tels résultats permettraient aux théoriciens de formuler de nouveaux modèles plus complets et plus à même de décrire l'Univers que l'actuel SM, qui seront eux-mêmes expérimentalement testés, perpétuant ainsi la marche scientifique ayant déjà mené avec succès à de nombreuses découvertes.

Annexe A
Systèmes d'unités

Le système d'unité utilisé dans cette thèse est le système dit « naturel » (UN), différent du système international (SI) [276] plus largement utilisé. À chaque grandeur physique correspond une dimension, basée sur les sept dimensions et unités fondamentales du SI :

— le temps T, en seconde ;
— la longueur L, en mètre ;
— la masse M, en kilogramme ;
— le courant électrique I, en ampère ;
— la température Θ, en kelvin ;
— la quantité de matière N, en mole ;
— l'intensité lumineuse J, en candela.

La définition des sept unités fondamentales repose sur autant de constantes physiques dont la valeur est fixée par convention. Ainsi, la seconde est déterminée par la fréquence de la transition hyperfine du césium $\Delta\nu_{\mathrm{Cs}} = 9\,192\,631\,770\,\mathrm{Hz}$. Le mètre peut alors être défini à partir de la vitesse de la lumière dans le vide $c = 299\,792\,458\,\mathrm{m{\cdot}s^{-1}}$. Puis, le kilogramme est obtenu grâce à ces deux premières unités et la constante de Planck $h = 6{,}626\,070\,15 \times 10^{-34}\,\mathrm{J{\cdot}s}$. Les grandeurs physiques usuellement rencontrées, exprimées avec ces unités, ont généralement des valeurs facilement manipulables. Par exemple, une personne peut mesurer 1,7 m et peser 70 kg.

Cependant, en physique des particules, ces unités ne sont pas adaptées aux échelles des grandeurs physiques exploitées. C'est pourquoi un autre système d'unités, dites naturelles, (UN) est utilisé dans lequel $\hbar = c = 1$. Les dimensions fondamentales de masse, longueur et temps sont remplacées par les dimensions de $\hbar$, c et d'énergie en GeV, avec $\hbar = \frac{h}{2\pi}$ et $1\,\mathrm{GeV} = 10^9\,\mathrm{eV} = 10^9 \times e \times 1\,\mathrm{V}$, e étant la charge électrique élémentaire. Ce changement de dimensions fondamentales permet toujours de décrire toutes les grandeurs physiques, comme un changement de système de coordonnées permet toujours de décrire toutes les positions possibles.

Dans le UN, une masse s'exprime ainsi en $\mathrm{GeV}{\cdot}c^{-2}$, une longueur en $\hbar{\cdot}c{\cdot}\mathrm{GeV}^{-1}$, une durée en $\hbar{\cdot}\mathrm{GeV}^{-1}$, etc. Par abus de langage, comme $\hbar = c = 1$, les facteurs $\hbar$ et c sont omis. Une masse s'exprime alors en GeV, ce qui devrait pourtant correspondre à une énergie. Par exemple, l'électron possède une masse de $9{,}11 \times 10^{-31}\,\mathrm{kg}$, soit 511 keV. Le tableau A.1 résume le passage entre (M, L, T), ($\hbar$, c, GeV) et ($\hbar = c = 1$, GeV).

Grandeur	Dimensions : (M, L, T), ($\hbar$, c, GeV), ($\hbar = c = 1$, GeV)			Conversion SI $\rightarrow$ UN
Masse	M	$\mathrm{GeV}{\cdot}c^{-2}$	GeV	$1\,\mathrm{kg} = 5{,}61 \times 10^{26}\,\mathrm{GeV}$
Longueur	L	$\hbar{\cdot}c{\cdot}\mathrm{GeV}^{-1}$	GeV^{-1}	$1\,\mathrm{m} = 5{,}07 \times 10^{15}\,\mathrm{GeV}^{-1}$
Durée	T	$\hbar{\cdot}\mathrm{GeV}^{-1}$	GeV^{-1}	$1\,\mathrm{s} = 1{,}52 \times 10^{24}\,\mathrm{GeV}^{-1}$
Énergie	$\mathsf{ML}^2\mathsf{T}^{-2}$	GeV	GeV	$1\,\mathrm{J} = \frac{10^{-3}}{1{,}6 \times 10^{-19}}\,\mathrm{GeV}$
Impulsion	MLT^{-1}	$\mathrm{GeV}{\cdot}c^{-1}$	GeV	$1\,\mathrm{kg{\cdot}m{\cdot}s^{-1}} = 1{,}87 \times 10^{18}\,\mathrm{GeV}$

Tableau A.1 – *Équivalence entre les systèmes d'unités international et naturel pour quelques grandeurs physiques.*

Annexe B
Notions mathématiques

Convention de sommation d'Einstein Il s'agit d'un raccourci de notation allégeant les équations. Ainsi, pour un espace de Minkowski de métrique η,

$$a_\mu b^\mu = a^\mu b_\mu = a^\mu \eta_{\mu\nu} b^\nu = a_\mu \eta^{\mu\nu} b_\nu = a^0 b^0 - \sum_{i=1}^{3} a^i b^i \tag{B.1}$$

car la métrique de Minkowski est ici

$$\eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{B.2}$$

Matrices de Pauli

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{B.3}$$

Matrices de Dirac ou matrices gamma

$$\gamma^0 = \begin{pmatrix} \mathbb{1}_{2\times2} & 0_{2\times2} \\ 0_{2\times2} & -\mathbb{1}_{2\times2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} , \quad \gamma^1 = \begin{pmatrix} 0_{2\times2} & \sigma_1 \\ -\sigma_1 & 0_{2\times2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} ,$$
$$\gamma^2 = \begin{pmatrix} 0_{2\times2} & \sigma_2 \\ -\sigma_2 & 0_{2\times2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix} , \quad \gamma^3 = \begin{pmatrix} 0_{2\times2} & \sigma_3 \\ -\sigma_3 & 0_{2\times2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} . \tag{B.4}$$

Projecteur chiral

$$\gamma^5 = \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} . \tag{B.5}$$

Matrices de Gell-Mann

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \quad \lambda_2 = \begin{pmatrix} 0 & -\mathrm{i} & 0 \\ \mathrm{i} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} ,$$
$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} , \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -\mathrm{i} \\ 0 & 0 & 0 \\ \mathrm{i} & 0 & 0 \end{pmatrix} , \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} ,$$
$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{pmatrix} , \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} . \tag{B.6}$$

Annexe C
Diagrammes de Feynman

Les diagrammes de Feynman sont des représentations graphiques des équations mathématiques, issues de la théorie quantique des champs, décrivant les interactions entre particules. Grâce à leur aspect visuel simple, ils permettent une approche intuitive de ces équations souvent complexes. Ces diagrammes ont été inventés par Richard Feynman à la fin des années 40 afin de réaliser des calculs de diffusion de particules [277]. Un des tout premiers diagrammes de Feynman présents dans la littérature, sur la figure C.1, représente ainsi la diffusion entre deux électrons. Le temps s'y écoule de bas en haut ; l'état initial est donc en bas du diagramme et l'état final en haut. Dans les diagrammes de Feynman du reste de cette thèse, le temps s'écoule de gauche à droite. Ainsi, l'état initial se trouve à gauche et l'état final à droite.

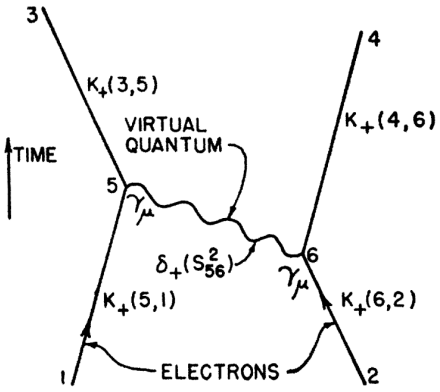


Figure C.1 – *Diagramme de Feynman de la diffusion électron-électron présenté dans la référence [277].*

Les diagrammes de Feynman ne représentent pas pour autant la réalité physique des processus étudiés. En effet, seuls les états initiaux et finaux sont réels, les parties internes de chaque diagramme correspondant aux manières de passer d'un état à l'autre. Il faut prendre en compte plusieurs diagrammes de Feynman, parfois de plus en plus complexes, c'est-à-dire d'ordres supérieurs, afin d'obtenir une description fidèle du processus étudié.

Ainsi, pour un processus de type $e^+e^- \rightarrow e^+e^-$, c'est-à-dire un état initial contenant un anti-électron d'impulsion $\vec{p}_1$ et un électron d'impulsion $\vec{p}_2$ et un état final contenant un anti-électron d'impulsion $\vec{p}_3$ et un électron d'impulsion $\vec{p}_4$, l'étude complète du phénomène physique (membre de gauche sur la figure C.2) doit prendre en compte la diffusion (premier terme du membre de droite), l'annihilation et réapparition de la paire e^+e^- par interaction électromagnétique (deuxième terme) ou faible (troisième terme) mais aussi des processus plus complexes (quatrième et cinquième termes).

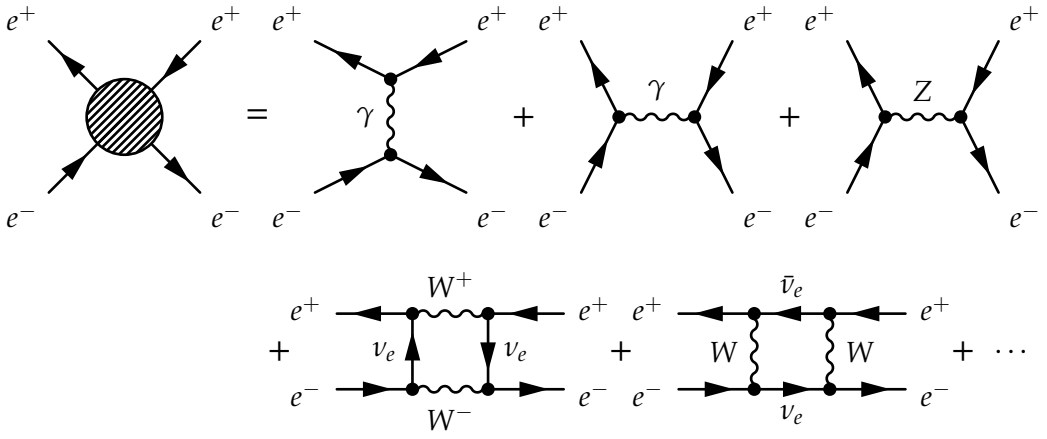


Figure C.2 – *Prise en compte de plusieurs diagrammes de Feynman lors de l'étude d'un processus physique donné.*

Selon le type de particule se propageant (fermion, photon, boson de Higgs ...), différents types de lignes sont utilisés dans les diagrammes de Feynman. Plusieurs conventions existent, en particulier pour les bosons. La convention utilisée dans cette thèse est illustrée sur la figure C.3.

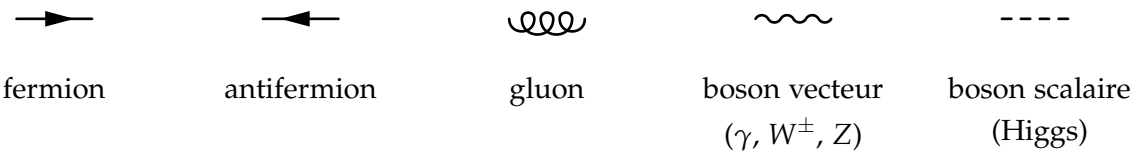


Figure C.3 – *Types de particules et lignes correspondantes dans les diagrammes de Feynman.*

Les interactions entre les particules ont lieu, dans les diagrammes de Feynman, au niveau des connexions entre les lignes correspondant à ces particules. Ces connexions sont les « vertex » du diagramme. Dans le cadre de la théorie quantique des champs, les interactions entre les particules sont possibles grâce aux termes du lagrangien faisant intervenir les champs correspondants à ces particules. Par exemple, dans le cas de l'interaction électromagnétique abordée dans le chapitre 1, un des termes du lagrangien est

$$\bar{\psi}\gamma^{\mu}eQA_{\mu}\psi = \bar{\psi}_a\ [\gamma^{\mu}]_{ab}\, eQ\ A_{\mu}\ \psi_b\,, \qquad \text{(C.1)}$$

faisant intervenir le champ ψ des fermions, $\bar{\psi}$ pour les antifermions ainsi que le champ A_{μ} des photons. Ce terme correspond directement aux vertex comme celui de la figure C.4, sur laquelle les rôles des différents facteurs présents dans ce terme du lagrangien sont illustrés.

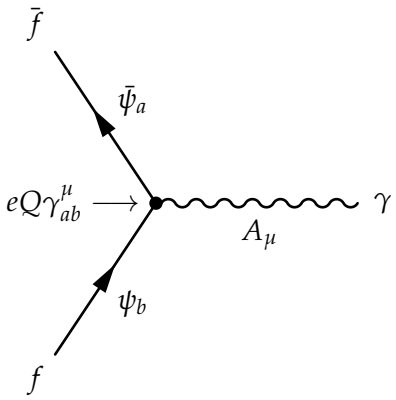


Figure C.4 – *Diagramme de Feynman issu du terme du lagrangien du modèle standard de l'équation (C.1). Un fermion f et un antifermion $\bar{f}$ sont présents dans l'état initial et sont décrits par un champ fermionique ψ et son adjoint $\bar{\psi}$. Le champ ψ interagit par ses composantes a et b avec le champ vectoriel A_{μ} au vertex, avec un couplage donné par $eQ\gamma^{\mu}_{ab}$. Il en résulte un photon γ, décrit par ce champ A_{μ}.*

La probabilité de passer d'un état initial $|i\rangle$ à un état final $|f\rangle$ s'exprime

$$P(|i\rangle \to |f\rangle) = \left|\mathcal{S}_{fi}\right|^2, \quad \mathcal{S}_{fi} = \langle f|\,\mathcal{S}\,|i\rangle\ . \qquad \text{(C.2)}$$

où $\mathcal{S}$ est la matrice de diffusion ou « matrice S » pour *scattering*. Pour déterminer $\mathcal{S}_{fi}$, il faut sommer les contributions des différents modes de passage de $|i\rangle$ à $|f\rangle$. Les termes de cette somme correspondent aux diagrammes de Feynman permettant le passage de $|i\rangle$ à $|f\rangle$, comme sur la figure C.2. Chacun de ces termes consiste en un produit de différents facteurs. Certains d'entre eux assurent la conservation de l'énergie et de l'impulsion. De plus, chaque ligne et vertex du diagramme de Feynman correspondant apporte un facteur supplémentaire, dont l'expression exacte est donnée par les règles de Feynman.

Ces diagrammes sont ainsi un outil incontournable en physique des particules. Des diagrammes analogues à ceux de Feynman sont utilisés en physique nucléaire et en physique de la matière condensée, le formalisme étant proche de celui de la physique des particules.

Annexe D
Jeux de données – γ + jets

Cette annexe présente les jeux de données utilisés pour les analyses du chapitre 3.

1 Jeux de données pour l'analyse 2018 des événements γ + jets

L'analyse est basée sur les données à $\sqrt{s} = 13$ TeV collectées en 2018 par l'expérience CMS, correspondant à une luminosité intégrée de 59,7 fb^{-1}. Les jeux de données utilisés ainsi que leurs gammes de *runs* et luminosités intégrées sont donnés dans le tableau D.1. Seuls les événements certifiés par la collaboration CMS sont considérés. Cette sélection est renseignée dans le fichier `JSON`

`Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt`

se trouvant dans `Collisions18/13TeV/ReReco/`. Les jeux de données simulées utilisés ainsi que leurs sections efficaces sont donnés dans le tableau D.2.

2 Jeux de données pour l'analyse 2017UL des événements γ + jets

L'analyse est basée sur les données à $\sqrt{s} = 13$ TeV collectées en 2017 par l'expérience CMS, correspondant à une luminosité intégrée de 41,6 fb^{-1}. Les jeux de données utilisés ainsi que leurs gammes de *runs* et luminosités intégrées sont donnés dans le tableau D.3. Seuls les événements certifiés par la collaboration CMS sont considérés. Cette sélection est renseignée dans le fichier `JSON`

`Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON.txt`

se trouvant dans `Collisions17/13TeV/Legacy_2017/`. Avant que ce fichier ne soit disponible et pour des études préliminaires, le fichier de certification utilisé était celui de l'analyse 2017 « non UL »,

`Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON.txt`

se trouvant dans `Collisions17/13TeV/ReReco/`. Les jeux de données simulées utilisés ainsi que leurs sections efficaces sont donnés dans le tableau D.4.

Jeu de données	Gamme de *run*	$\mathcal{L}$ (fb^{-1})
/EGamma/Run2018A-17Sep2018-v2/MINIAOD	315 252 – 316 995	13,98
/EGamma/Run2018B-17Sep2018-v1/MINIAOD	317 080 – 319 310	7,064
/EGamma/Run2018C-17Sep2018-v1/MINIAOD	319 337 – 320 065	6,899
/EGamma/Run2018D-PromptReco-v2/MINIAOD	320 673 – 325 175	31,75

Tableau D.1 – *Jeux de données utilisés pour l'analyse 2018 des événements γ + jets.*

Jeu de données simulées	σ (pb)
/GJet_Pt-15To6000_TuneCP5-Flat_13TeV_pythia8[1]	283 000,0

[1] /RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM

Tableau D.2 – *Jeux de données simulées utilisés pour l'analyse 2018 des événements γ + jets.*

Jeu de données	Gamme de *run*	$\mathcal{L}$ (fb^{-1})
/SinglePhoton/Run2017B-09Aug2019_UL2017-v1/MINIAOD	297 046 – 299 329	4,823
/SinglePhoton/Run2017C-09Aug2019_UL2017-v1/MINIAOD	299 368 – 302 029	9,664
/SinglePhoton/Run2017D-09Aug2019_UL2017-v1/MINIAOD	302 030 – 303 434	4,252
/SinglePhoton/Run2017E-09Aug2019_UL2017-v1/MINIAOD	303 824 – 304 797	9,278
/SinglePhoton/Run2017F-09Aug2019_UL2017-v1/MINIAOD	305 040 – 306 462	13,54

Tableau D.3 – *Jeux de données utilisés pour l'analyse 2017UL des événements γ + jets.*

Jeu de données simulées	σ (pb)
/GJets_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8[1]	18 700,0
/GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8[2]	8640,0
/GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8[1]	2185,0
/GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8[1]	259,9
/GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8[1]	85,31

[1] /RunIISummer19UL17MiniAOD-106X_mc2017_realistic_v6-v1/MINIAODSIM

[2] /RunIISummer19UL17MiniAOD-4cores5k_106X_mc2017_realistic_v6-v1/MINIAODSIM

Tableau D.4 – *Jeux de données simulées utilisés pour l'analyse 2017UL des événements γ + jets.*

Annexe E
Jeux de données – $H \to \tau\tau$

L'analyse du chapitre 4 est basée sur les données à $\sqrt{s} = 13\,\text{TeV}$ collectées en 2016, 2017 et 2018 par l'expérience CMS, correspondant à une luminosité intégrée de $35{,}9 + 41{,}5 + 59{,}7\,\text{fb}^{-1}$. Seuls les événements certifiés par la collaboration CMS sont considérés. Cette sélection est renseignée dans les fichiers `JSON` du tableau E.1. Les jeux de données utilisés pour chacun des états finaux considérés, ainsi que leurs gammes de *runs* et luminosités intégrées respectives, sont donnés dans les tableaux E.2, E.3 et E.4.

La modélisation du boson de Higgs du modèle standard est obtenue avec les jeux de données simulées correspondent aux modes de production du boson de Higgs ggh, VBF, VH (Wh, Zh et $ggZh$) et $t\bar{t}h$. Les listes de ces jeux de données simulées utilisés pour les trois années analysées sont données dans les tableaux E.5, E.7 et E.9.

La modélisation des bosons de Higgs neutres additionnels du MSSM, c'est-à-dire $\Phi \to \tau\tau$ avec $\Phi = H, A$, est obtenue avec les jeux de données $gg \to \Phi \to \tau\tau$ et $gg \to bb\Phi \to \tau\tau$ simulé avec POWHEG [180] et PYTHIA [178, 179] pour l'hadronisation. Les listes de ces jeux de données simulées utilisés pour les trois années analysées sont données dans les tableaux E.6, E.8 et E.10.

Les jeux de données simulées utilisés afin de modéliser les bruits de fond sont listés dans les tableaux E.11, E.12 et E.13. Les différents processus sont regroupés comme suit :

- $Z \to \tau\tau$, $Z \to \ell\ell$:
  - $Z \to LL$,
  - $Z + 1$ jet,
  - $Z + 2$ jets,
  - $Z + 3$ jets,
  - $Z + 4$ jets,
  - EWK $Z \to LL$,
  - EWK $Z \to \nu\nu$;
- $t\bar{t}$;
- $W +$ jets :
  - $W +$ jets,
  - $W + 1$ jet,
  - $W + 2$ jets,
  - $W + 3$ jets,
  - $W + 4$ jets,
  - EWK W^-,
  - EWK W^+,
  - $W\gamma$;
- Diboson :
  - *Single top*,
  - VVTo2L2Nu,
  - WZTo2L2Q,
  - WZTo3LNu,
  - ZZTo2L2Q,
  - ZZTo4L.

Les jeux de données encapsulées (*embedded*) sont listés dans les tableaux E.14, E.15 et E.16. Ces jeux de données sont utilisés dans une estimation du bruit de fond contenant des paires de leptons τ à partir des données elles-mêmes.

Année	Fichier de certification `JSON`
2016	`Cert_271036-284044_13TeV_ReReco_07Aug2017_Collisions16_JSON.txt`
2017	`Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON_v1.txt`
2018	`Cert_314472-325175_13TeV_17SeptEarlyReReco 2018ABC_PromptEraD_Collisions18_JSON.txt`

Tableau E.1 – *Fichiers de certification `JSON`.*

Canal	Jeu de données	Gamme de *run*	$\mathcal{L}$ (fb^{-1})
$\tau_h\tau_h$	`/Tau/Run2016B-17Jul2018_ver2-v1/MINIAOD`	272 007 – 275 376	5,788
$\tau_h\tau_h$	`/Tau/Run2016C-17Jul2018-v1/MINIAOD`	275 657 – 276 283	2,573
$\tau_h\tau_h$	`/Tau/Run2016D-17Jul2018-v1/MINIAOD`	276 315 – 276 811	4,248
$\tau_h\tau_h$	`/Tau/Run2016E-17Jul2018-v1/MINIAOD`	276 831 – 277 420	4,009
$\tau_h\tau_h$	`/Tau/Run2016F-17Jul2018-v1/MINIAOD`	277 772 – 278 808	3,102
$\tau_h\tau_h$	`/Tau/Run2016G-17Jul2018-v1/MINIAOD`	278 820 – 280 385	7,540
$\tau_h\tau_h$	`/Tau/Run2016H-17Jul2018-v1/MINIAOD`	280 919 – 284 044	8,606
$\mu\tau_h$	`/SingleMuon/Run2016B-17Jul2018_ver2-v1/MINIAOD`	272 007 – 275 376	5,788
$\mu\tau_h$	`/SingleMuon/Run2016C-17Jul2018-v1/MINIAOD`	275 657 – 276 283	2,573
$\mu\tau_h$	`/SingleMuon/Run2016D-17Jul2018-v1/MINIAOD`	276 315 – 276 811	4,248
$\mu\tau_h$	`/SingleMuon/Run2016E-17Jul2018-v1/MINIAOD`	276 831 – 277 420	4,009
$\mu\tau_h$	`/SingleMuon/Run2016F-17Jul2018-v1/MINIAOD`	277 772 – 278 808	3,102
$\mu\tau_h$	`/SingleMuon/Run2016G-17Jul2018-v1/MINIAOD`	278 820 – 280 385	7,540
$\mu\tau_h$	`/SingleMuon/Run2016H-17Jul2018-v1/MINIAOD`	280 919 – 284 044	8,606
$e\tau_h$	`/SingleElectron/Run2016B-17Jul2018_ver2-v1/MINIAOD`	272 007 – 275 376	5,788
$e\tau_h$	`/SingleElectron/Run2016C-17Jul2018-v1/MINIAOD`	275 657 – 276 283	2,573
$e\tau_h$	`/SingleElectron/Run2016D-17Jul2018-v1/MINIAOD`	276 315 – 276 811	4,248
$e\tau_h$	`/SingleElectron/Run2016E-17Jul2018-v1/MINIAOD`	276 831 – 277 420	4,009
$e\tau_h$	`/SingleElectron/Run2016F-17Jul2018-v1/MINIAOD`	277 772 – 278 808	3,102
$e\tau_h$	`/SingleElectron/Run2016G-17Jul2018-v1/MINIAOD`	278 820 – 280 385	7,540
$e\tau_h$	`/SingleElectron/Run2016H-17Jul2018-v1/MINIAOD`	280 919 – 284 044	8,606
$e\mu$	`/MuonEG/Run2016B-17Jul2018_ver2-v1/MINIAOD`	272 007 – 275 376	5,788
$e\mu$	`/MuonEG/Run2016C-17Jul2018-v1/MINIAOD`	275 657 – 276 283	2,573
$e\mu$	`/MuonEG/Run2016D-17Jul2018-v1/MINIAOD`	276 315 – 276 811	4,248
$e\mu$	`/MuonEG/Run2016E-17Jul2018-v1/MINIAOD`	276 831 – 277 420	4,009
$e\mu$	`/MuonEG/Run2016F-17Jul2018-v1/MINIAOD`	277 772 – 278 808	3,102
$e\mu$	`/MuonEG/Run2016G-17Jul2018-v1/MINIAOD`	278 820 – 280 385	7,540
$e\mu$	`/MuonEG/Run2016H-17Jul2018-v1/MINIAOD`	280 919 – 284 044	8,606

Tableau E.2 – *Jeux de données utilisés en 2016.*

Canal	Jeu de données	Gamme de *run*	$\mathcal{L}$ (fb^{-1})
$\tau_h\tau_h$	/Tau/Run2017B-31Mar2018-v1/MINIAOD	297 046 – 299 329	4,823
$\tau_h\tau_h$	/Tau/Run2017C-31Mar2018-v1/MINIAOD	299 368 – 302 029	9,664
$\tau_h\tau_h$	/Tau/Run2017D-31Mar2018-v1/MINIAOD	302 030 – 303 434	4,252
$\tau_h\tau_h$	/Tau/Run2017E-31Mar2018-v1/MINIAOD	303 824 – 304 797	9,278
$\tau_h\tau_h$	/Tau/Run2017F-31Mar2018-v1/MINIAOD	305 040 – 306 462	13,54
$\mu\tau_h$	/SingleMuon/Run2017B-31Mar2018-v1/MINIAOD	297 046 – 299 329	4,823
$\mu\tau_h$	/SingleMuon/Run2017C-31Mar2018-v1/MINIAOD	299 368 – 302 029	9,664
$\mu\tau_h$	/SingleMuon/Run2017D-31Mar2018-v1/MINIAOD	302 030 – 303 434	4,252
$\mu\tau_h$	/SingleMuon/Run2017E-31Mar2018-v1/MINIAOD	303 824 – 304 797	9,278
$\mu\tau_h$	/SingleMuon/Run2017F-31Mar2018-v1/MINIAOD	305 040 – 306 462	13,54
$e\tau_h$	/SingleElectron/Run2017B-31Mar2018-v1/MINIAOD	297 046 – 299 329	4,823
$e\tau_h$	/SingleElectron/Run2017C-31Mar2018-v1/MINIAOD	299 368 – 302 029	9,664
$e\tau_h$	/SingleElectron/Run2017D-31Mar2018-v1/MINIAOD	302 030 – 303 434	4,252
$e\tau_h$	/SingleElectron/Run2017E-31Mar2018-v1/MINIAOD	303 824 – 304 797	9,278
$e\tau_h$	/SingleElectron/Run2017F-31Mar2018-v1/MINIAOD	305 040 – 306 462	13,54
$e\mu$	/MuonEG/Run2017B-31Mar2018-v1/MINIAOD	297 046 – 299 329	4,823
$e\mu$	/MuonEG/Run2017C-31Mar2018-v1/MINIAOD	299 368 – 302 029	9,664
$e\mu$	/MuonEG/Run2017D-31Mar2018-v1/MINIAOD	302 030 – 303 434	4,252
$e\mu$	/MuonEG/Run2017E-31Mar2018-v1/MINIAOD	303 824 – 304 797	9,278
$e\mu$	/MuonEG/Run2017F-31Mar2018-v1/MINIAOD	305 040 – 306 462	13,54

Tableau E.3 – *Jeux de données utilisés en 2017.*

Canal	Jeu de données	Gamme de *run*	$\mathcal{L}$ (fb^{-1})
$\tau_h\tau_h$	/Tau/Run2018A-17Sep2018-v1/MINIAOD	315 252 – 316 995	13,98
$\tau_h\tau_h$	/Tau/Run2018B-17Sep2018-v1/MINIAOD	317 080 – 319 310	7,064
$\tau_h\tau_h$	/Tau/Run2018C-17Sep2018-v1/MINIAOD	319 337 – 320 065	6,899
$\tau_h\tau_h$	/Tau/Run2018D-PromptReco-v2/MINIAOD	320 673 – 325 175	31,75
$\mu\tau_h$	/SingleMuon/Run2018A-17Sep2018-v2/MINIAOD	315 252 – 316 995	13,98
$\mu\tau_h$	/SingleMuon/Run2018B-17Sep2018-v1/MINIAOD	317 080 – 319 310	7,064
$\mu\tau_h$	/SingleMuon/Run2018C-17Sep2018-v1/MINIAOD	319 337 – 320 065	6,899
$\mu\tau_h$	/SingleMuon/Run2018D-22Jan2019-v2/MINIAOD	320 673 – 325 175	31,75
$e\tau_h$	/EGamma/Run2018A-17Sep2018-v2/MINIAOD	315 252 – 316 995	13,98
$e\tau_h$	/EGamma/Run2018B-17Sep2018-v1/MINIAOD	317 080 – 319 310	7,064
$e\tau_h$	/EGamma/Run2018C-17Sep2018-v1/MINIAOD	319 337 – 320 065	6,899
$e\tau_h$	/EGamma/Run2018D-22Jan2019-v2/MINIAOD	320 673 – 325 175	31,75
$e\mu$	/MuonEG/Run2018A-17Sep2018-v1/MINIAOD	315 252 – 316 995	13,98
$e\mu$	/MuonEG/Run2018B-17Sep2018-v1/MINIAOD	317 080 – 319 310	7,064
$e\mu$	/MuonEG/Run2018C-17Sep2018-v1/MINIAOD	319 337 – 320 065	6,899
$e\mu$	/MuonEG/Run2018D-PromptReco-v2/MINIAOD	320 673 – 325 175	31,75

Tableau E.4 – *Jeux de données utilisés en 2018.*

Processus	Jeu de données simulées	$\sigma \times \mathcal{BR}$ (pb)
$ggh \to \tau\tau$	/GluGluHToTauTau_M125_13TeV$^{\dagger 1,2,3}$	3,00 (N3LO)
VBF $h \to \tau\tau$	/VBFHToTauTau_M125_13TeV$^{\dagger 1,2,3}$	0,237 (NNLO)
$W^+h \to \tau\tau$	/WplusHToTauTau_M125_13TeV$^{\dagger 1}$	0,0527 (NNLO)
$W^-h \to \tau\tau$	/WminusHToTauTau_M125_13TeV$^{\dagger 1}$	0,0334 (NNLO)
$Zh \to \tau\tau$	/ZHToTauTau_M125_13TeV$^{\dagger 1}$	0,0477 (NNLO)
$ggZh \to qq\tau\tau$	/ggZH_HToTauTau_ZToQQ_M125_13TeV$^{\dagger 1}$	0,0054 (NNLO)
$ggZh \to \nu\nu\tau\tau$	/ggZH_HToTauTau_ZToNuNu_M125_13TeV$^{\dagger 1}$	0,0015 (NNLO)
$ggZh \to LL\tau\tau$	/ggZH_HToTauTau_ZToLL_M125_13TeV$^{\dagger 1}$	0,0008 (NNLO)
$t\bar{t}h \to \tau\tau$	/ttHJetToTT_M125_13TeV$^{\|\| 4}$	0,0318 (NLO)
$ggh \to WW$	/GluGluHToWWTo2L2Nu_M125_13TeV$^{\ddagger 1}$	1,09 (N3LO)
VBF $h \to WW$	/VBFHToWWTo2L2Nu_M125_13TeV$^{\ddagger 1}$	0,0850 (NNLO)
$W^+h \to WW$	/HWplusJ_HToWW_M125_13TeV$^{\dagger 1}$	0,18 (NLO)
$W^-h \to WW$	/HWminusJ_HToWW_M125_13TeV$^{\dagger 1}$	0,114 (NLO)
$Zh \to WW$	/HZJ_HToWW_M125_13TeV$^{\dagger 1}$	0,163 (NLO)
$ggZh \to WW$	/GluGluZH_HToWW_M125_13TeV$^{\dagger 1}$	0,0262 (NLO)

[1] /RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v*/MINIAODSIM

[2] /RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3_ext1-v*/MINIAODSIM

[3] /RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3_ext2-v*/MINIAODSIM

[4] /RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3_ext4-v*/MINIAODSIM

[†] _powheg_pythia8 [||] _amcatnloFXFX_madspin_pythia8

[‡] _powheg_JHUgenv628_pythia8

Tableau E.5 – *Jeux de données simulées modélisant le boson de Higgs du modèle standard en 2016.*

Processus	Jeu de données simulées
$gg \to \Phi \to \tau\tau$	/SUSYGluGluToHToTauTau_M-*_TuneCUETP8M1_13TeV-powheg-pythia8[1]
$gg \to bb\Phi \to \tau\tau$	/SUSYGluGluToBBHToTauTau_M-*_TuneCUETP8M1_13TeV-powheg-pythia8[1]

[1] /RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v*/MINIAODSIM

Tableau E.6 – *Jeux de données simulées modélisant les bosons de Higgs du MSSM en 2016.*

Processus	Jeu de données simulées	$\sigma \times \mathcal{BR}$ (pb)
$ggh \to \tau\tau$	/GluGluHToTauTau_M125_13TeV[†1,2]	3,00 (N3LO)
VBF $h \to \tau\tau$	/VBFHToTauTau_M125_13TeV[†2]	0,237 (NNLO)
$W^+h \to \tau\tau$	/WplusHToTauTau_M125_13TeV[†3]	0,0527 (NNLO)
$W^-h \to \tau\tau$	/WminusHToTauTau_M125_13TeV[†3]	0,0334 (NNLO)
$Zh \to \tau\tau$	/ZHToTauTau_M125_13TeV[†3]	0,0477 (NNLO)
$ggZh \to qq\tau\tau$	/ggZH_HToTauTau_ZToQQ_M125_13TeV[†3]	0,0054 (NNLO)
$ggZh \to \nu\nu\tau\tau$	/ggZH_HToTauTau_ZToNuNu_M125_13TeV[†3]	0,0015 (NNLO)
$ggZh \to LL\tau\tau$	/ggZH_HToTauTau_ZToLL_M125_13TeV[†3]	0,0008 (NNLO)
$t\bar{t}h \to \tau\tau$	/ttHToTauTau_M125_TuneCP5_13TeV[†2]	0,0318 (NLO)
$ggh \to WW$	/GluGluHToWWTo2L2Nu_M125_13TeV[§3]	1,09 (N3LO)
VBF $h \to WW$	/VBFHToWWTo2L2Nu_M125_13TeV[§3]	0,0850 (NNLO)
$W^+h \to WW$	/HWplusJ_HToWW_M125_13TeV[‡3]	0,18 (NLO)
$W^-h \to WW$	/HWminusJ_HToWW_M125_13TeV[‡3]	0,114 (NLO)
$Zh \to WW$	/HZJ_HToWW_M125_13TeV[‖3]	0,163 (NLO)
$ggZh \to WW$	/GluGluZH_HToWW_M125_13TeV[‡3]	0,0262 (NLO)

[1] /RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14_ext1-v*/MINIAODSIM

[2] /RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v*/MINIAODSIM

[3] /RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v*/MINIAODSIM

[†] _powheg_pythia8

[‡] _powheg_pythia8_TuneCP5

[§] _powheg2_JHUGenV714_pythia8

[‖] _powheg_JHUgenv714_pythia8_TuneCP5

Tableau E.7 – *Jeux de données simulées modélisant le boson de Higgs du modèle standard en 2017.*

Processus	Jeu de données simulées
$gg \to \Phi \to \tau\tau$	/SUSYGluGluToHToTauTau_M-*_TuneCP5_13TeV-powheg-pythia8[1]
$gg \to bb\Phi \to \tau\tau$	/SUSYGluGluToBBHToTauTau_M-*_TuneCP5_13TeV-powheg-pythia8[1]

[1] /RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM

Tableau E.8 – *Jeux de données simulées modélisant les bosons de Higgs du MSSM en 2017.*

Processus	Jeu de données simulées	$\sigma \times \mathcal{BR}$ (pb)
$ggh \to \tau\tau$	`/GluGluHToTauTau_M125_13TeV`$^{\dagger 1}$	3,00 (N3LO)
VBF $h \to \tau\tau$	`/VBFHToTauTau_M125_13TeV`$^{\dagger 2}$	0,237 (NNLO)
$W^+h \to \tau\tau$	`/WplusHToTauTau_M125_13TeV`$^{\dagger 1}$	0,0527 (NNLO)
$W^-h \to \tau\tau$	`/WminusHToTauTau_M125_13TeV`$^{\dagger 1}$	0,0334 (NNLO)
$Zh \to \tau\tau$	`/ZHToTauTau_M125_13TeV`$^{\dagger 1}$	0,0477 (NNLO)
$ggZh \to qq\tau\tau$	`/ggZH_HToTauTau_ZToQQ_M125_13TeV`$^{\dagger 1}$	0,0054 (NNLO)
$ggZh \to \nu\nu\tau\tau$	`/ggZH_HToTauTau_ZToNuNu_M125_13TeV`$^{\dagger 1}$	0,0015 (NNLO)
$ggZh \to LL\tau\tau$	`/ggZH_HToTauTau_ZToLL_M125_13TeV`$^{\dagger 1}$	0,0008 (NNLO)
$t\bar{t}h \to \tau\tau$	`/ttHToTauTau_M125_TuneCP5_13TeV`$^{\dagger 1}$	0,0318 (NLO)
$ggh \to WW$	`/GluGluHToWWTo2L2Nu_M125_13TeV`$^{\ddagger 1}$	1,09 (N3LO)
VBF $h \to WW$	`/VBFHToWWTo2L2Nu_M125_13TeV`$^{\ddagger 1}$	0,0850 (NNLO)
$W^+h \to WW$	`/HWplusJ_HToWW_M125_13TeV`$^{\S 1}$	0,18 (NLO)
$W^-h \to WW$	`/HWminusJ_HToWW_M125_13TeV`$^{\S 1}$	0,114 (NLO)
$Zh \to WW$	`/HZJ_HToWW_M125_13TeV`$^{\S 1}$	0,163 (NLO)
$ggZh \to WW$	`/GluGluZH_HToWW_M125_13TeV`$^{\| 1}$	0,0262 (NLO)

[1] `/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v*/MINIAODSIM`

[2] `/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15_ext1-v*/MINIAODSIM`

† `_powheg_pythia8`

‡ `_powheg2_JHUGenV714_pythia8`

§ `_powheg_jhugen714_pythia8_TuneCP5`

|| `_powheg_pythia8_TuneCP5_PSweights`

Tableau E.9 – *Jeux de données simulées modélisant le boson de Higgs du modèle standard en 2018.*

Processus	Jeu de données simulées
$gg \to \Phi \to \tau\tau$	`/SUSYGluGluToHToTauTau_M-*_TuneCP5_13TeV-powheg-pythia8`[1]
$gg \to bb\Phi \to \tau\tau$	`/SUSYGluGluToBBHToTauTau_M-*_TuneCP5_13TeV-powheg-pythia8`[1]

[1] `/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v*/MINIAODSIM`

Tableau E.10 – *Jeux de données simulées modélisant les bosons de Higgs du MSSM en 2018.*

Processus	Jeu de données simulées	σ (pb)
$Z \to LL$ (basse masse)	/DYJetsToLL_M-10to50[†1]	21 658,0 (NLO)
$Z \to LL$	/DYJetsToLL_M-50[†2,3]	6077,22 (NNLO)
$Z + 1$ jet	/DY1JetsToLL_M-50[†1]	1253,1[*]
$Z + 2$ jets	/DY2JetsToLL_M-50[†1]	409,4[*]
$Z + 3$ jets	/DY3JetsToLL_M-50[†1]	124,8[*]
$Z + 4$ jets	/DY4JetsToLL_M-50[†1]	67,33[*]
EWK $Z \to LL$	/EWKZ2Jets_ZToLL_M-50[‡2,3]	4,321 (LO)
EWK $Z \to \nu\nu$	/EWKZ2Jets_ZToNuNu[‡2,3]	10,66 (LO)
$t\bar{t}$	/TTTo2L2Nu[§1]	88,29
$t\bar{t}$	/TTToHadronic[§1]	377,96
$t\bar{t}$	/TTToSemiLeptonic[§1]	365,35
VVTo2L2Nu	/VVTo2L2Nu_13TeV[‖1,2]	13,84
WZTo2L2Q	/WZTo2L2Q_13TeV[‖1]	5,52
WZTo3LNu	/WZTo3LNu_TuneCUETP8M1_13TeV[¶1]	4,43
ZZTo2L2Q	/ZZTo2L2Q_13TeV[‖1]	3,38
ZZTo4L	/ZZTo4L_13TeV[¶2]	1,26
Single top	/ST_t-channel_antitop_4f_inclusiveDecays[◇1]	80,95
Single top	/ST_t-channel_top_4f_inclusiveDecays[◇1]	136,02
Single top	/ST_tW_antitop_5f_inclusiveDecays[□2]	35,85
Single top	/ST_tW_top_5f_inclusiveDecays[□2]	35,85
$W +$ jets	/WJetsToLNu[†1,3]	61 526,7 (NNLO)
$W + 1$ jet	/W1JetsToLNu[†1]	11 805,6[*]
$W + 2$ jets	/W2JetsToLNu[†2]	3891,0[*]
$W + 3$ jets	/W3JetsToLNu[†2]	1153,2[*]
$W + 4$ jets	/W4JetsToLNu[†2,3]	60,67[*]
EWK W^-	/EWKWMinus2Jets_WToLNu_M-50[‡2,3]	23,24 (LO)
EWK W^+	/EWKWPlus2Jets_WToLNu_M-50[‡2,3]	29,59 (LO)

[1] /RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v*/MINIAODSIM

[2] /RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3_ext1-v*/MINIAODSIM

[3] /RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3_ext2-v*/MINIAODSIM

[†] _TuneCUETP8M1_13TeV-madgraphMLM-pythia8

[‡] _13TeV-madgraph-pythia8

[§] _TuneCP5_PSweights_13TeV-powheg-pythia8

[‖] _amcatnloFXFX_madspin_pythia8

[¶] -amcatnloFXFX-pythia8

[◇] _13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1

[□] _13TeV-powheg-pythia8_TuneCUETP8M1

[*] Déterminée à partir de la section efficace du jeu inclusif.

Tableau E.11 – *Jeux de données simulées modélisant le bruit de fond en 2016.*

Processus	Jeu de données simulées	σ (pb)
$Z \to LL$ (basse masse)	`/DYJetsToLL_M-10to50`[†2]	21 658,0 (NLO)
$Z \to LL$	`/DYJetsToLL_M-50`[†3,4]	6077,22 (NNLO)
$Z + 1$ jet	`/DY1JetsToLL_M-50`[†5,8]	977,1[*]
$Z + 2$ jets	`/DY2JetsToLL_M-50`[†1,6]	347,3[*]
$Z + 3$ jets	`/DY3JetsToLL_M-50`[†1,2]	126,1[*]
$Z + 4$ jets	`/DY4JetsToLL_M-50`[†7]	71,67[*]
EWK $Z \to LL$	`/EWKZ2Jets_ZToLL_M-50`[‡5]	4,321 (LO)
EWK $Z \to \nu\nu$	`/EWKZ2Jets_ZToNuNu`[‡5]	10,66 (LO)
$t\bar{t}$	`/TTTo2L2Nu`[§5]	88,29
$t\bar{t}$	`/TTToHadronic`[§5]	377,96
$t\bar{t}$	`/TTToSemiLeptonic`[§5]	365,35
VVTo2L2Nu	`/VVTo2L2Nu_13TeV`[||1]	13,84
WZTo2L2Q	`/WZTo2L2Q_13TeV`[||1]	5,52
WZTo3LNu	`/WZTo3LNu`[¶5]	4,43
ZZTo2L2Q	`/ZZTo2L2Q_13TeV`[||1]	3,38
ZZTo4L	`/ZZTo4L`[¶1]	1,26
Single top	`/ST_t-channel_antitop_4f_inclusiveDecays`[◇1]	80,95
Single top	`/ST_t-channel_top_4f_inclusiveDecays`[◇5]	136,02
Single top	`/ST_tW_antitop_5f_inclusiveDecays`[§1]	35,85
Single top	`/ST_tW_top_5f_inclusiveDecays`[§1]	35,85
$W +$ jets	`/WJetsToLNu`[†1,2]	61 526,7 (NNLO)
$W + 1$ jet	`/W1JetsToLNu`[†1]	9370,5[*]
$W + 2$ jets	`/W2JetsToLNu`[†1]	3170,9[*]
$W + 3$ jets	`/W3JetsToLNu`[†1]	1132,5[*]
$W + 4$ jets	`/W4JetsToLNu`[†8]	631,5[*]
EWK W^-	`/EWKWMinus2Jets_WToLNu_M-50`[‡5]	23,24 (LO)
EWK W^+	`/EWKWPlus2Jets_WToLNu_M-50`[‡5]	29,59 (LO)
$W\gamma$ (canal $e\mu$)	`/WGToLNuG`[†1]	464,4 (LO)

[1] `/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v*/MINIAODSIM`

[2] `/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14_ext1-v*/MINIAODSIM`

[3] `/RunIIFall17MiniAODv2-PU2017RECOSIMstep_12Apr2018_94X_mc2017_realistic_v14-v*/MINIAODSIM`

[4] `/RunIIFall17MiniAODv2-PU2017RECOSIMstep_12Apr2018_94X_mc2017_realistic_v14_ext1-v*/MINIAODSIM`

[5] `/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v*/MINIAODSIM`

[6] `/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14_ext1-v*/MINIAODSIM`

[7] `/RunIIFall17MiniAODv2-PU2017_12Apr2018_v2_94X_mc2017_realistic_v14-v*/MINIAODSIM`

[8] `/RunIIFall17MiniAODv2-PU2017_12Apr2018_v3_94X_mc2017_realistic_v14_ext1-v*/MINIAODSIM`

[†] `_TuneCP5_13TeV-madgraphMLM-pythia8`

[‡] `_TuneCP5_13TeV-madgraph-pythia8`

[§] `_TuneCP5_13TeV-powheg-pythia8`

[||] `_amcatnloFXFX_madspin_pythia8`

[¶] `_TuneCP5_13TeV-amcatnloFXFX-pythia8`

[◇] `_TuneCP5_13TeV-powhegV2-madspin-pythia8`

[*] Déterminée à partir de la section efficace du jeu inclusif.

Tableau E.12 – *Jeux de données simulées modélisant le bruit de fond en 2017.*

Processus	Jeu de données simulées	σ (pb)
$Z \to LL$ (basse masse)	/DYJetsToLL_M-10to50[†1]	21 658,0 (NLO)
$Z \to LL$	/DYJetsToLL_M-50[†1]	6077,22 (NNLO)
$Z + 1$ jet	/DY1JetsToLL_M-50[†1]	1007,6[*]
$Z + 2$ jets	/DY2JetsToLL_M-50[†1]	344,3[*]
$Z + 3$ jets	/DY3JetsToLL_M-50[†1]	125,3[*]
$Z + 4$ jets	/DY4JetsToLL_M-50[†1]	71,20[*]
EWK $Z \to LL$	/EWKZ2Jets_ZToLL_M-50[‡1]	4,321 (LO)
EWK $Z \to \nu\nu$	/EWKZ2Jets_ZToNuNu[‡1]	10,66 (LO)
$t\bar{t}$	/TTTo2L2Nu[§1]	88,29
$t\bar{t}$	/TTToHadronic[§1]	377,96
$t\bar{t}$	/TTToSemiLeptonic[§1]	365,35
VVTo2L2Nu	/VVTo2L2Nu[||1]	13,84
WZTo2L2Q	/WZTo2L2Q[||1]	5,52
WZTo3LNu	/WZTo3LNu[¶2]	4,43
ZZTo2L2Q	/ZZTo2L2Q[||1]	3,38
ZZTo4L	/ZZTo4L[¶1]	1,26
Single top	/ST_t-channel_antitop_4f_inclusiveDecays[◇1]	80,95
Single top	/ST_t-channel_top_4f_inclusiveDecays[◇1]	136,02
Single top	/ST_tW_antitop_5f_inclusiveDecays[§2]	35,85
Single top	/ST_tW_top_5f_inclusiveDecays[§2]	35,85
$W +$ jets	/WJetsToLNu[†1]	61 526,7 (NNLO)
$W + 1$ jet	/W1JetsToLNu[†1]	9328,1[*]
$W + 2$ jets	/W2JetsToLNu[†1]	3181,5[*]
$W + 3$ jets	/W3JetsToLNu[†1]	1116,2[*]
$W + 4$ jets	/W4JetsToLNu[†1]	629,3[*]
EWK W^-	/EWKWMinus2Jets_WToLNu_M-50[□1]	23,24 (LO)
EWK W^+	/EWKWPlus2Jets_WToLNu_M-50[□1]	29,59 (LO)

[1] /RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v*/MINIAODSIM

[2] /RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15_ext1-v*/MINIAODSIM

[†] _TuneCP5_13TeV-madgraphMLM-pythia8

[‡] _TuneCP5_PSweights_13TeV-madgraph-pythia8

[§] _TuneCP5_13TeV-powheg-pythia8

[||] _13TeV_amcatnloFXFX_madspin_pythia8

[¶] _TuneCP5_13TeV-amcatnloFXFX-pythia8

[◇] _TuneCP5_13TeV-powheg-madspin-pythia8

[□] _TuneCP5_13TeV-madgraph-pythia8

[*] Déterminée à partir de la section efficace du jeu inclusif.

Tableau E.13 – *Jeux de données simulées modélisant le bruit de fond en 2018.*

Canal	Jeu de données	Canal	Jeu de données
$\tau_h\tau_h$	/EmbeddingRun2016B/TauTau*	$e\mu$	/EmbeddingRun2016B/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2016C/TauTau*	$e\mu$	/EmbeddingRun2016C/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2016D/TauTau*	$e\mu$	/EmbeddingRun2016D/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2016E/TauTau*	$e\mu$	/EmbeddingRun2016E/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2016F/TauTau*	$e\mu$	/EmbeddingRun2016F/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2016G/TauTau*	$e\mu$	/EmbeddingRun2016G/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2016H/TauTau*	$e\mu$	/EmbeddingRun2016H/ElMu*
$\mu\tau_h$	/EmbeddingRun2016B/MuTau*	$e\tau_h$	/EmbeddingRun2016B/ElTau*
$\mu\tau_h$	/EmbeddingRun2016C/MuTau*	$e\tau_h$	/EmbeddingRun2016C/ElTau*
$\mu\tau_h$	/EmbeddingRun2016D/MuTau*	$e\tau_h$	/EmbeddingRun2016D/ElTau*
$\mu\tau_h$	/EmbeddingRun2016E/MuTau*	$e\tau_h$	/EmbeddingRun2016E/ElTau*
$\mu\tau_h$	/EmbeddingRun2016F/MuTau*	$e\tau_h$	/EmbeddingRun2016F/ElTau*
$\mu\tau_h$	/EmbeddingRun2016G/MuTau*	$e\tau_h$	/EmbeddingRun2016G/ElTau*
$\mu\tau_h$	/EmbeddingRun2016H/MuTau*	$e\tau_h$	/EmbeddingRun2016H/ElTau*

* FinalState-inputDoubleMu_94X_Legacy_miniAOD-v5/USER

Tableau E.14 – *Jeux de données encapsulées en 2016.*

Canal	Jeu de données	Canal	Jeu de données
$\tau_h\tau_h$	/EmbeddingRun2017B/TauTau*	$e\mu$	/EmbeddingRun2017B/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2017C/TauTau*	$e\mu$	/EmbeddingRun2017C/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2017D/TauTau*	$e\mu$	/EmbeddingRun2017D/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2017E/TauTau*	$e\mu$	/EmbeddingRun2017E/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2017F/TauTau*	$e\mu$	/EmbeddingRun2017F/ElMu*
$\mu\tau_h$	/EmbeddingRun2017B/MuTau*	$e\tau_h$	/EmbeddingRun2017B/ElTau*
$\mu\tau_h$	/EmbeddingRun2017C/MuTau*	$e\tau_h$	/EmbeddingRun2017C/ElTau*
$\mu\tau_h$	/EmbeddingRun2017D/MuTau*	$e\tau_h$	/EmbeddingRun2017D/ElTau*
$\mu\tau_h$	/EmbeddingRun2017E/MuTau*	$e\tau_h$	/EmbeddingRun2017E/ElTau*
$\mu\tau_h$	/EmbeddingRun2017F/MuTau*	$e\tau_h$	/EmbeddingRun2017F/ElTau*

* FinalState-inputDoubleMu_94X_miniAOD-v2/USER

Tableau E.15 – *Jeux de données encapsulées en 2017.*

Canal	Jeu de données	Canal	Jeu de données
$\tau_h\tau_h$	/EmbeddingRun2018A/TauTau*	$e\mu$	/EmbeddingRun2018A/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2018B/TauTau*	$e\mu$	/EmbeddingRun2018B/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2018C/TauTau*	$e\mu$	/EmbeddingRun2018C/ElMu*
$\tau_h\tau_h$	/EmbeddingRun2018D/TauTau*	$e\mu$	/EmbeddingRun2018D/ElMu*
$\mu\tau_h$	/EmbeddingRun2018A/MuTau*	$e\tau_h$	/EmbeddingRun2018A/ElTau*
$\mu\tau_h$	/EmbeddingRun2018B/MuTau*	$e\tau_h$	/EmbeddingRun2018B/ElTau*
$\mu\tau_h$	/EmbeddingRun2018C/MuTau*	$e\tau_h$	/EmbeddingRun2018C/ElTau*
$\mu\tau_h$	/EmbeddingRun2018D/MuTau*	$e\tau_h$	/EmbeddingRun2018D/ElTau*

* FinalState-inputDoubleMu_102X_miniAOD-v1/USER

Tableau E.16 – *Jeux de données encapsulées en 2018.*

Annexe F
Chemins de déclenchement – $H \rightarrow \tau\tau$

Pour l'analyse détaillée dans le chapitre 4, l'enregistrement des données à CMS est activé selon les chemins de déclenchements (HLT *paths*) listés dans les tableaux F.2 à F.13, pour les années 2016, 2017 et 2018 et pour les canaux $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ et $e\mu$ selon la répartition donnée dans le tableau F.1.

Année	$\tau_h\tau_h$	$\mu\tau_h$	$e\tau_h$	$e\mu$
2016	F.2	F.3	F.4	F.5
2017	F.6	F.7	F.8	F.9
2018	F.10	F.11	F.12	F.13

Tableau F.1 – *Tableaux contenant les informations des HLT* paths *pour chaque année et canal de l'analyse.*

La fréquence des collisions ainsi que la variété des objets à reconstruire rendent impossible la reconstruction de chaque trace du trajectographe en temps réel [145]. Afin d'estimer les objets physiques en présence, une « graine » (*L1 Seed*) est générée à partir de quelques informations directement issues du détecteur. Elle définit une estimation initiale de la trajectoire d'une particule d'un type donné ainsi que l'incertitude sur celle-ci. Pour chaque HLT *path*, une liste de *L1 Seed* utilisées est indiquée.

De plus, dans l'analyse détaillée dans le chapitre 4, il est requis que les objets d'intérêts correspondent aux objets ayant déclenché le HLT *path*. Des filtres permettent de déterminer l'objet ayant déclenché le HLT *path* à comparer avec l'objet d'intérêt et sont également renseignés.

HLT *path*	*L1 Seed*	Filtre à tau vérifié
HLT_VLooseIsoPFTau 120_Trk50_eta2p1_v	L1_SingleTau100er	hltPFTau120TrackPt50LooseAbsOrRelVLooseIso
HLT_VLooseIsoPFTau 140_Trk50_eta2p1_v	L1_SingleTau100er	hltPFTau140TrackPt50LooseAbsOrRelVLooseIso
HLT_DoubleMedium Iso_PFTau35_Trk1 _eta2p1_Reg_v	L1_DoubleIsoTau*er, * in 26, 27, 28, 30, 32, 39	hltDoublePFTau35TrackPt1MediumIsolationDz02Reg
HLT_DoubleMedium CombinedIsoPFTau 35Trk1_eta2p1_Reg_v	L1_DoubleIsoTau*er, * in 26, 27, 28, 30, 32, 39	hltDoublePFTau35TrackPt1MediumCombinedIsolationDz02Reg

Tableau F.2 – *Chemins de déclenchement utilisés en 2016 pour le canal* $\tau_h\tau_h$.

HLT *path*	*L1 Seed*	Filtre à muon vérifié	Filtre à tau vérifié
HLT_IsoMu22_v	L1_SingleMu20	hltL3crIsoL1sMu20L1f0L2f10QL3 f22QL3trkIsoFiltered0p09	-
HLT_IsoMu22 _eta2p1_v	L1_SingleMu20er	hltL3crIsoL1sSingleMu20erL1f0L2 f10QL3f22QL3trkIsoFiltered0p09	-
HLT_VLooseIsoPFTau 120_Trk50_eta2p1_v	L1_SingleTau100er	-	hltPFTau120TrackPt50LooseAbs OrRelVLooseIso
HLT_VLooseIsoPFTau 140_Trk50_eta2p1_v	L1_SingleTau100er	-	hltPFTau140TrackPt50LooseAbs OrRelVLooseIso
HLT_IsoMu19_eta2p1 _LooseIsoPFTau20_v	L1_Mu18er_Tau20er	hltL3crIsoL1sMu18erTauJet20erL1 f0L2f10QL3f19QL3trkIsoFiltered 0p09hltOverlapFilterIsoMu19 LooseIsoPFTau20	hltPFTau20TrackLooseIsoAgainst MuonhltOverlapFilterIsoMu19 LooseIsoPFTau20
HLT_IsoMu19_eta2p1 _LooseIsoPFTau20 _SingleL1_v	L1_SingleMu18er **ou** L1_SingleMu20er	hltL3crIsoL1sSingleMu18erIor SingleMu20erL1f0L2f10QL3f19 QL3trkIsoFiltered0p09hltOverlap FilterSingleIsoMu19LooseIsoPFTau20	hltPFTau20TrackLooseIsoAgainst MuonhltOverlapFilterSingleIso Mu19LooseIsoPFTau20

Tableau F.3 – *Chemins de déclenchement utilisés en 2016 pour le canal* $\mu\tau_h$.

HLT *path*	*L1 Seed*	Filtre à électron vérifié	Filtre à tau vérifié
HLT_Ele25_eta2p1 _WPTight_Gsf_v	L1_SingleEG40 **ou** SingleIsoEG24er **ou** L1_SingleIsoEG22er	hltEle25erWPTightGsf TrackIsoFilter	-
HLT_VLooseIsoPF Tau120_Trk50_eta2p1_v	L1_SingleTau100er	-	hltPFTau120TrackPt50Loose AbsOrRelVLooseIso
HLT_VLooseIsoPF Tau140_Trk50_eta2p1_v	L1_SingleTau100er	-	hltPFTau140TrackPt50Loose AbsOrRelVLooseIso
		Run $<$ 276215 and MC	
HLT_Ele24_eta2p1 _WPLoose_Gsf_Loose IsoPFTau20_SingleL1_v	L1_SingleEG40 **ou** SingleIsoEG24er **ou** L1_SingleIsoEG22er	hltEle24WPLooseL1SingleIsoEG 22erGsfTrackIsoFilterhltOverlap FilterSingleIsoEle24WPLoose GsfLooseIsoPFTau20	hltPFTau20TrackLooseIsohlt OverlapFilterSingleIsoEle24WP LooseGsfLooseIsoPFTau20
		276215 $\leqslant$ Run $<$ 278270	
HLT_Ele24_eta2p1 _WPLoose_Gsf_Loose IsoPFTau20_v	L1_IsoEG22er_Tau 20er_dEta_Min0p2	hltEle24WPLooseL1IsoEG22erTau 20erGsfTrackIsoFilterhltOverlap FilterIsoEle24WPLooseGsfLoose IsoPFTau20	hltPFTau20TrackLooseIsohlt OverlapFilterIsoEle24WPLoose GsfLooseIsoPFTau20
		278270 $\leqslant$ Run	
HLT_Ele24_Eta2p1 _WPLoose_Gsf_Loose IsoPFTau30_v	L1_IsoEG22er_Iso Tau26er_dEta _Min0p2	hltEle24WPLooseL1IsoEG22erIso Tau26erGsfTrackIsoFilterhlt OverlapFilterIsoEle24WPLoose GsfLooseIsoPFTau30	hltPFTau30TrackLooseIsohlt OverlapFilterIsoEle24WPLoose GsfLooseIsoPFTau30

Tableau F.4 – *Chemins de déclenchement utilisés en 2016 pour le canal* $e\tau_h$.

HLT *path*	*L1 Seed*	Filtre à muon vérifié	Filtre à électron vérifié
		Runs B-F and MC	
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v	L1_Mu12_EG10	hltMu23TrkIsoVVLEle12CaloIdLTrackIdLIsoVLMuonlegL3IsoFiltered23	hltMu23TrkIsoVVLEle12CaloIdLTrackIdLIsoVLElectronlegTrackIsoFilter
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v	L1_Mu5_EG15	hltMu8TrkIsoVVLEle23CaloIdLTrackIdLIsoVLMuonlegL3IsoFiltered8	hltMu8TrkIsoVVLEle23CaloIdLTrackIdLIsoVLElectronlegTrackIsoFilter
		Runs G-H	
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v	L1_Mu12_EG10	hltMu23TrkIsoVVLEle12CaloIdLTrackIdLIsoVLMuonlegL3IsoFiltered23 **et** hltMu23TrkIsoVVLEle12CaloIdLTrackIdLIsoVLDZFilter	hltMu23TrkIsoVVLEle12CaloIdLTrackIdLIsoVLElectronlegTrackIsoFilter **et** hltMu23TrkIsoVVLEle12CaloIdLTrackIdLIsoVLDZFilter
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v	L1_Mu5_EG15	hltMu8TrkIsoVVLEle23CaloIdLTrackIdLIsoVLMuonlegL3IsoFiltered8 **et** hltMu8TrkIsoVVLEle23CaloIdLTrackIdLIsoVLDZFilter	hltMu8TrkIsoVVLEle23CaloIdLTrackIdLIsoVLElectronlegTrackIsoFilter **et** hltMu8TrkIsoVVLEle23CaloIdLTrackIdLIsoVLDZFilter

Tableau F.5 – *Chemins de déclenchement utilisés en 2016 pour le canal $e\mu$.*

HLT *path*	*L1 Seed*	Filtre à tau vérifié
HLT_MediumChargedIsoPFTau180HighPtRelaxedIso_Trk50_eta2p1_v	L1_SingleTau80to140er	hltPFTau180TrackPt50LooseAbsOrRelMediumHighPtRelaxedIsoIso **et** hltSelectedPFTau180MediumChargedIsolationL1HLTMatched
HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg_v	L1_DoubleIsoTau*er2p1	hltDoublePFTau40TrackPt1MediumChargedIsolationAndTightOOSCPhotonsDz02Reg
HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg_v	L1_DoubleIsoTau*er2p1	hltDoublePFTau40TrackPt1TightChargedIsolationDz02Reg
HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg_v	L1_DoubleIsoTau*er2p1	hltDoublePFTau35TrackPt1TightChargedIsolationAndTightOOSCPhotonsDz02Reg

* in 28, 30, 32, 33, 34, 35, 36, 38, 70

Tableau F.6 – *Chemins de déclenchement utilisés en 2017 pour le canal $\tau_h\tau_h$.*

HLT *path*	*L1 Seed*	Filtre à muon vérifié	Filtre à tau vérifié
HLT_IsoMu24_v	L1_SingleMu22	hltL3crIsoL1sSingleMu22L1f0L2f10QL3f24QL3trkIsoFiltered0p07	-
HLT_IsoMu27_v	L1_SingleMu22 **ou** L1_SingleMu25	hltL3crIsoL1sMu22Or25L1f0L2f10QL3f27QL3trkIsoFiltered0p07	-
HLT_MediumChargedIsoPFTau180HighPtRelaxedIso_Trk50_eta2p1_v	L1_SingleTau80to140er	-	hltPFTau180TrackPt50LooseAbsOrRelMediumHighPtRelaxedIsoIso **et** hltSelectedPFTau180MediumChargedIsolationL1HLTMatched
HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau27_eta2p1_CrossL1_v	L1_Mu18er2p1_Tau24er2p1	hltL3crIsoL1sMu18erTau24erIorMu20erTau24erL1f0L2f10QL3f20QL3trkIsoFiltered0p07hltOverlapFilterIsoMu20LooseChargedIsoPFTau27L1Seeded	hltSelectedPFTau27LooseChargedIsolationAgainstMuonL1HLTMatchedhltOverlapFilterIsoMu20LooseChargedIsoPFTau27L1Seeded

Tableau F.7 – *Chemins de déclenchement utilisés en 2017 pour le canal $\mu\tau_h$.*

HLT *path*	*L1 Seed*	Filtre à électron vérifié	Filtre à tau vérifié
HLT_Ele27_ WPTight_Gsf_v	L1_SingleEGXX **ou** L1_SingleEGXXer2p1 **ou** L1_SingleIsoEGXX **ou** L1_SingleIsoEGXXer2p1	hltEle27WPTightGsf TrackIsoFilter	-
HLT_Ele32_ WPTight_Gsf_v	L1_SingleEGXX **ou** L1_SingleEGXXer2p1 **ou** L1_SingleIsoEGXX **ou** L1_SingleIsoEGXXer2p1	hltEle32WPTightGsf TrackIsoFilter	-
HLT_Ele35_ WPTight_Gsf_v	L1_SingleEGXX **ou** L1_SingleEGXXer2p1 **ou** L1_SingleIsoEGXX **ou** L1_SingleIsoEGXXer2p1	hltEle35noerWPTightGsf TrackIsoFilter	-
HLT_Medium ChargedIsoPFTau 180HighPtRelaxed Iso_Trk50_eta2p1_v	L1_SingleTau80to140er	-	hltPFTau180TrackPt50LooseAbs OrRelMediumHighPtRelaxed IsoIso **et** hltSelectedPFTau180 MediumChargedIsolationL1 HLTMatched
HLT_Ele24_ eta2p1_WPTight _Gsf_LooseCharged IsoPFTau30 _eta2p1_CrossL1_v	L1_LooseIsoEG22er2p1 _IsoTau26er2p1 _dR_Min0p3	hltEle24erWPTightGsfTrackIso FilterForTauhltOverlapFilterIso Ele24WPTightGsfLooseIso PFTau30	hltSelectedPFTau30Loose ChargedIsolationL1HLTMatched hltOverlapFilterIsoEle24WPTight GsfLooseIsoPFTau30

Tableau F.8 – *Chemins de déclenchement utilisés en 2017 pour le canal* $e\tau_h$.

HLT *path*	*L1 Seed*	Filtre à muon vérifié	Filtre à électron vérifié
HLT_Mu23_Trk IsoVVL_Ele12_ CaloIdL_Track IdL_IsoVL_DZ_v	L1_Mu23EG10 **ou** L1_Mu20EG17	hltMu23TrkIsoVVLEle12CaloIdL TrackIdLIsoVLMuonlegL3 IsoFiltered23	hltMu23TrkIsoVVLEle12CaloIdL TrackIdLIsoVLElectronleg TrackIsoFilter
HLT_Mu8_Trk IsoVVL_Ele23_ CaloIdL_Track IdL_IsoVL_DZ_v	L1_Mu5EG23 **ou** L1_Mu5IsoEG20 **ou** L1_Mu7EG23 **ou** L1_Mu7IsoEG20 **ou** L1_MuIso7EG23	hltL3fL1sMu7EG23f0Filtered8 **ou** hltMu8TrkIsoVVLEle23CaloIdLTrack IdLIsoVLMuonlegL3IsoFiltered8	hltMu8TrkIsoVVLEle23CaloIdL TrackIdLIsoVLElectronlegTrack IsoFilter

Tableau F.9 – *Chemins de déclenchement utilisés en 2017 pour le canal* $e\mu$.

HLT *path*	*L1 Seed*	Filtre à tau vérifié
HLT_MediumCharged IsoPFTau180HighPt RelaxedIso_Trk50_eta2p1_v	L1_SingleTau80to140er	hltPFTau180TrackPt50LooseAbsOrRelMediumHighPtRelaxedIsoIso **et** hltSelectedPFTau180MediumChargedIsolationL1HLTMatched
HLT_DoubleMedium ChargedIsoPFTau40 _Trk1_TightID_eta2p1_Reg_v	L1_DoubleIsoTau*er2p1	hltDoublePFTau40TrackPt1MediumChargedIsolationAnd TightOOSCPhotonsDz02Reg
HLT_DoubleTight ChargedIsoPFTau40 _Trk1_eta2p1_Reg_v	L1_DoubleIsoTau*er2p1	hltDoublePFTau40TrackPt1TightChargedIsolationDz02Reg
HLT_DoubleTight ChargedIsoPFTau35 _Trk1_TightID_eta2p1_Reg_v	L1_DoubleIsoTau*er2p1	hltDoublePFTau35TrackPt1TightChargedIsolationAnd TightOOSCPhotonsDz02Reg
HLT_DoubleMedium ChargedIsoPFTauHPS35_Trk1_eta2p1_Reg_v	L1_DoubleIsoTau*er2p1	hltHpsDoublePFTau35TrackPt1MediumChargedIsolationDz02Reg

* in 28, 30, 32, 33, 34, 35, 36, 38, 70

Tableau F.10 – *Chemins de déclenchement utilisés en 2018 pour le canal* $\tau_h\tau_h$.

HLT *path*	*L1 Seed*	Filtre à muon vérifié	Filtre à tau vérifié
HLT_IsoMu24_v	L1_SingleMu22	hltL3crIsoL1sSingleMu22L1f0L2 f10QL3f24QL3trkIsoFiltered0p07	-
HLT_IsoMu27_v	L1_SingleMu22 **ou** L1_SingleMu25	hltL3crIsoL1sMu22Or25L1f0L2 f10QL3f27QL3trkIsoFiltered0p07	-
HLT_Medium ChargedIsoPFTau 180HighPtRelaxed Iso_Trk50_eta2p1_v	L1_SingleTau 80to140er	-	hltPFTau180TrackPt50LooseAbsOr RelMediumHighPtRelaxedIsoIso **et** hltSelectedPFTau180Medium ChargedIsolationL1HLTMatched
HLT_IsoMu20 _eta2p1_Loose ChargedIsoPFTau27 _eta2p1_CrossL1_v	L1_Mu18er2p1 _Tau24er2p1	hltL3crIsoL1sMu18erTau24erIorMu 20erTau24erL1f0L2f10QL3f20QL3trk IsoFiltered0p07hltOverlapFilterIso Mu20LooseChargedIsoPFTau27L1 Seeded	hltSelectedPFTau27LooseCharged IsolationAgainstMuonL1HLT MatchedhltOverlapFilterIsoMu20 LooseChargedIsoPFTau27L1Seeded

Tableau F.11 – *Chemins de déclenchement utilisés en 2018 pour le canal $\mu\tau_h$.*

HLT *path*	*L1 Seed*	Filtre à électron vérifié	Filtre à tau vérifié
HLT_Ele32_ WPTight_Gsf_v	L1_SingleEGXX **ou** L1_SingleEGXXer2p1 **ou** L1_SingleIsoEGXX **ou** L1_SingleIsoEGXXer2p1	hltEle32WPTightGsf TrackIsoFilter	-
HLT_Ele35_ WPTight_Gsf_v	L1_SingleEGXX **ou** L1_SingleEGXXer2p1 **ou** L1_SingleIsoEGXX **ou** L1_SingleIsoEGXXer2p1	hltEle35noerWPTightGsf TrackIsoFilter	-
HLT_Medium ChargedIsoPFTau 180HighPtRelaxed Iso_Trk50_eta2p1_v	L1_SingleTau80to140er	-	hltPFTau180TrackPt50LooseAbs OrRelMediumHighPtRelaxed IsoIso **et** hltSelectedPFTau180 MediumChargedIsolationL1 HLTMatched
HLT_Ele24_ eta2p1_WPTight _Gsf_LooseCharged IsoPFTau30 _eta2p1_CrossL1_v	L1_LooseIsoEG22er2p1 _IsoTau26er2p1 _dR_Min0p3	hltEle24erWPTightGsfTrackIso FilterForTauhltOverlapFilterIso Ele24WPTightGsfLooseIso PFTau30	hltSelectedPFTau30Loose ChargedIsolationL1HLTMatched hltOverlapFilterIsoEle24WPTight GsfLooseIsoPFTau30
HLT_Ele24_ eta2p1_WPTight _Gsf_LooseCharged IsoPFTauHPS30 _eta2p1_CrossL1_v	L1_LooseIsoEG22er2p1 _IsoTau26er2p1 _dR_Min0p3	hltEle24erWPTightGsfTrackIso FilterForTauhltOverlapFilterIso Ele24WPTightGsfLooseIso PFTau30	hltSelectedPFTau30Loose ChargedIsolationL1HLTMatched hltOverlapFilterIsoEle24WPTight GsfLooseIsoPFTau30

Tableau F.12 – *Chemins de déclenchement utilisés en 2018 pour le canal $e\tau_h$.*

HLT *path*	*L1 Seed*	Filtre à muon vérifié	Filtre à électron vérifié
HLT_Mu23_Trk IsoVVL_Ele12_ CaloIdL_Track IdL_IsoVL_DZ_v	L1_Mu23EG10 **ou** L1_Mu20EG17	hltMu23TrkIsoVVLEle12CaloIdL TrackIdLIsoVLMuonlegL3 IsoFiltered23	hltMu23TrkIsoVVLEle12CaloIdL TrackIdLIsoVLElectronleg TrackIsoFilter
HLT_Mu8_Trk IsoVVL_Ele23_ CaloIdL_Track IdL_IsoVL_DZ_v	L1_Mu5EG23 **ou** L1_Mu5IsoEG20 **ou** L1_Mu7EG23 **ou** L1_Mu7IsoEG20 **ou** L1_MuIso7EG23	hltL3fL1sMu7EG23f0Filtered8 **ou** hltMu8TrkIsoVVLEle23CaloIdLTrack IdLIsoVLMuonlegL3IsoFiltered8	hltMu8TrkIsoVVLEle23CaloIdL TrackIdLIsoVLElectronlegTrack IsoFilter

Tableau F.13 – *Chemins de déclenchement utilisés en 2018 pour le canal $e\mu$.*

Annexe G
Distributions de contrôle – $H \to \tau\tau$

Cette annexe présente des distributions de contrôle avant ajustement des paramètres de nuisance sur les événements utilisés dans l'analyse des événements $H \to \tau\tau$ présentée dans le chapitre 4. La sélection est « inclusive », les événements sont ceux sélectionnés par la définition de la région de signal, sans coupure sur m_{T}^{ℓ} (canaux $\mu\tau_{\mathrm{h}}$, $e\tau_{\mathrm{h}}$) ni D_{ζ} (canal $e\mu$).

Pour chacune des trois années de prise de données (2016, 2017, 2018) et chacun des quatre canaux ($\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, $\mu\tau_{\mathrm{h}}$, $e\tau_{\mathrm{h}}$, $e\mu$), les distributions de plusieurs variables sont données.

Dans chacun des graphiques, les données observées (points noirs) sont comparées à la modélisation des bruits de fond (histogrammes remplis en couleur et empilés). Les bandes grisées correspondent à l'incertitude statistique totale sur le bruit de fond. Le rapport au bruit de fond est donné dans la partie inférieure des graphiques.

Ces distributions montrent un bon accord entre données observées et estimations des bruits de fond, à l'exception :

— des pseudo-rapidités des τ_{h}, $\eta(\tau_{\mathrm{h}})$;
— du nombre de vertex d'empilement N_{PU}.

Pseudo-rapidités des τ_{h} L'écart observé pourrait être réduit en utilisant des facteurs de faux dépendants de $\eta(\tau_{\mathrm{h}})$, ce qui n'est pas le cas actuellement. D'autres variables dépendant directement de la pseudo-rapidité d'un τ_{h} telles que la distance ΔR entre les deux éléments du dilepton montrent également des écarts entre données observées et estimation des bruits de fond, potentiellement dus à ceux sur $\eta(\tau_{\mathrm{h}})$.

L'écart sur $\eta(\tau_{\mathrm{h}})$ pourrait avoir un effet sur $m_{\mathrm{T}}^{\mathrm{tot}}$, cette dernière étant fonction d'impulsions transverses et d'angles azimutaux, car la projection des impulsions dans le plan transverse,

$$p_{\mathrm{T}} = \frac{p}{\cosh \eta}\,, \tag{G.1}$$

fait intervenir $\eta(\tau_{\mathrm{h}})$. Cependant, les modélisations des impulsions transverses et des angles azimutaux sont satisfaisantes. Les valeurs de $m_{\mathrm{T}}^{\mathrm{tot}}$ ne sont donc pas affectées par la mauvaise modélisation de $\eta(\tau_{\mathrm{h}})$.

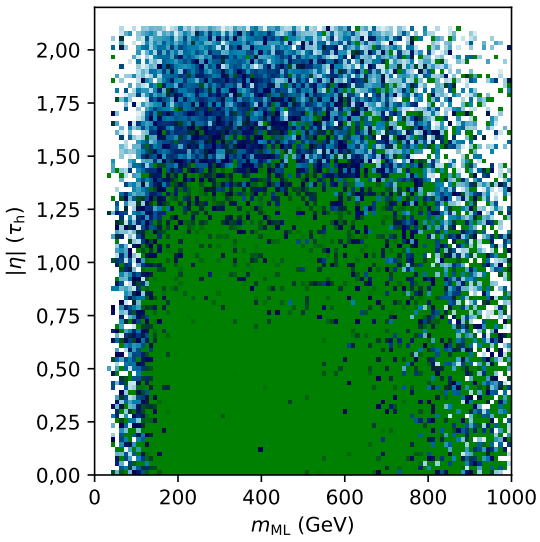


Figure G.1 – *Distribution de $|\eta|(\tau_{\mathrm{h}})$ en fonction de m_{ML}.*

Le réseau de neurones B" utilisé afin d'obtenir m_{ML}, présenté au chapitre 5, exploite les valeurs de $\eta(\tau_{\mathrm{h}})$. La figure G.1 montre la distribution de $|\eta|(\tau_{\mathrm{h}})$ en fonction de m_{ML}. Bien que les hautes valeurs de $|\eta|(\tau_{\mathrm{h}})$ soient moins représentées à haute m_{ML}, ce qui peut s'expliquer par l'effet de la projection mentionnée précédemment, aucune corrélation n'apparaît entre $|\eta|(\tau_{\mathrm{h}})$ et m_{ML}. Les valeurs de m_{ML} ne semblent donc pas affectées par la mauvaise modélisation de $\eta(\tau_{\mathrm{h}})$.

Nombre de vertex d'empilement L'effet de la mauvaise modélisation de N_{PU} constatée dans les distributions de contrôle devrait être marginal sur $m_{\mathrm{T}}^{\mathrm{tot}}$, car cette observable ne dépend pas de N_{PU}. Les autres sources d'incertitudes sur $m_{\mathrm{T}}^{\mathrm{tot}}$ permettent alors de couvrir cet effet.

Les valeurs de m_{ML}, c'est-à-dire des prédictions du réseau de neurones présenté au chapitre 5, dépendent de N_{PU}. Les bons accords entre observations et estimations des bruits de fond obtenus sur les distributions de m_{ML} montrent que cette observable semble peu affectée par la mauvaise modélisation de N_{PU}. La figure G.2 confirme cette hypothèse. Elle représente l'évolution de la réponse du réseau de neurones, à basse, moyenne et haute masse ainsi que sur toute la gamme de masse, ces ensembles étant définis au chapitre 5, lorsque les valeurs de N_{PU} sont modifiées selon

$$N_{\mathrm{PU}} \to N_{\mathrm{PU}} + \Delta N_{\mathrm{PU}} \tag{G.2}$$

avec ΔN_{PU} compris entre -5 et $+5$. Les modifications induites sur la réponse du réseau de neurones sont inférieures à 1 % alors que sa résolution relative est de l'ordre de 20 %.

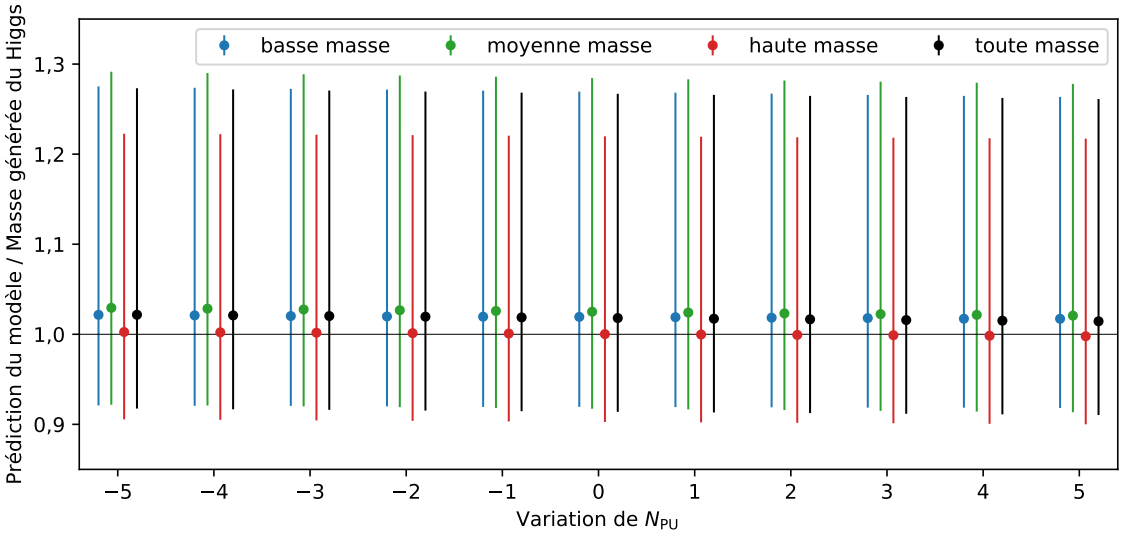


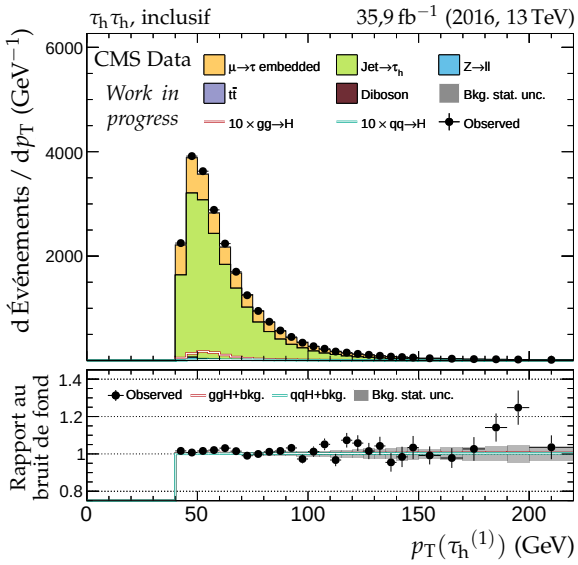
Figure G.2 – *Variations de la réponse du réseau de neurones en fonction de ΔN_{PU} à basse, moyenne et haute masse ainsi que sur toute la gamme de masse. La résolution à $\pm 1\sigma$ est renseignée par les barres verticales.*

Ainsi, l'effet limité de N_{PU} sur les variables discriminantes utilisées permet de conserver des résultats finaux pertinents bien que l'écart sur N_{PU} entre observations et estimations des bruits de fond puisse être de l'ordre de 40 %. Plusieurs effets peuvent causer cet écart sur N_{PU} :

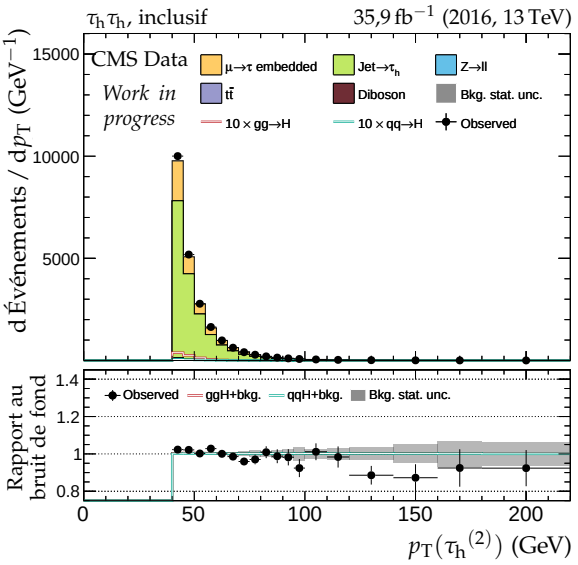
Modélisation des HLT *paths* dans les données encapsulées Les données encapsulées, introduites dans le chapitre 4, sont des hybrides entre données réelles et simulées. Les HLT *paths* activés proviennent de la partie simulée uniquement, c'est-à-dire d'un événement vide à l'exception des leptons τ qui remplacent les muons. Or, l'empilement provient de la partie réelle des données encapsulées et l'acceptation des HLT *paths* en dépend. Actuellement, des facteurs correctifs sont appliqués en fonction des propriétés cinématiques des τ mais ils ne dépendent pas de N_{PU}. Ainsi, les distributions de N_{PU} peuvent être biaisées dans les données encapsulées.

Dépendance en N_{PU} de l'identification des τ_{h} L'identification d'un τ_{h} dépend de son isolation, elle-même sensible à l'empilement. Différents points de fonctionnement d'identification des τ_{h} sont utilisés dans le cadre des facteurs de faux, introduits au chapitre 4. Ces derniers ne sont pas déterminés en fonction de N_{PU}, ce qui peut introduire un biais.

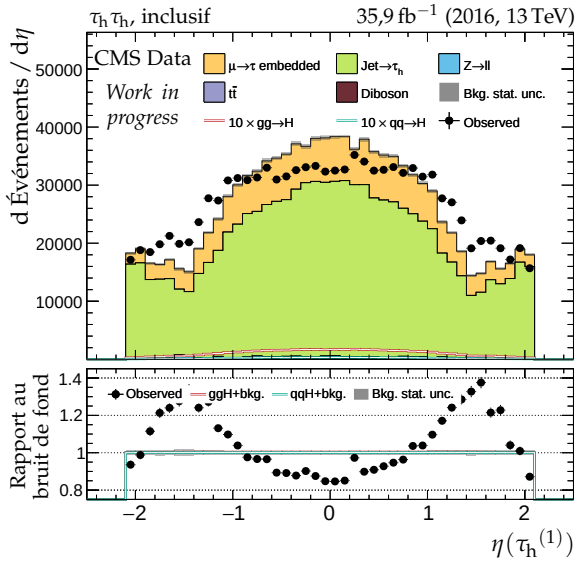
Modélisation de N_{PU} dans les données simulées Un désaccord sur N_{PU} est bien attendu entre données réelles et simulées et ces dernières sont pondérées afin de le corriger, comme exposé au chapitre 2. Toutefois, cette pondération ne permet pas d'obtenir un accord parfait.



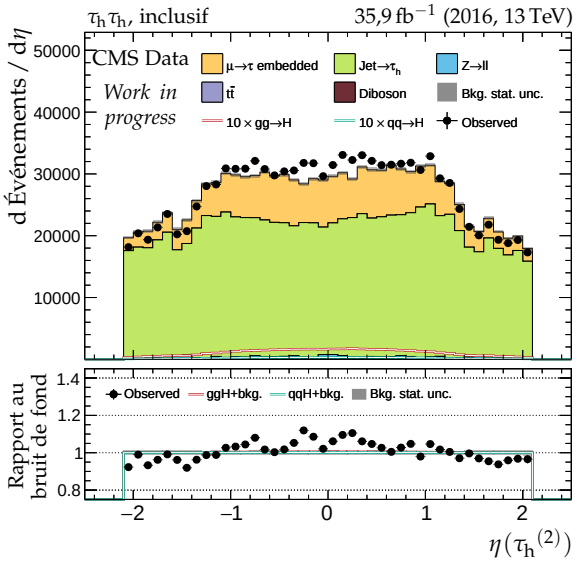
(a) *Impulsion transverse du lepton 1 ($\tau_h^{(1)}$).*



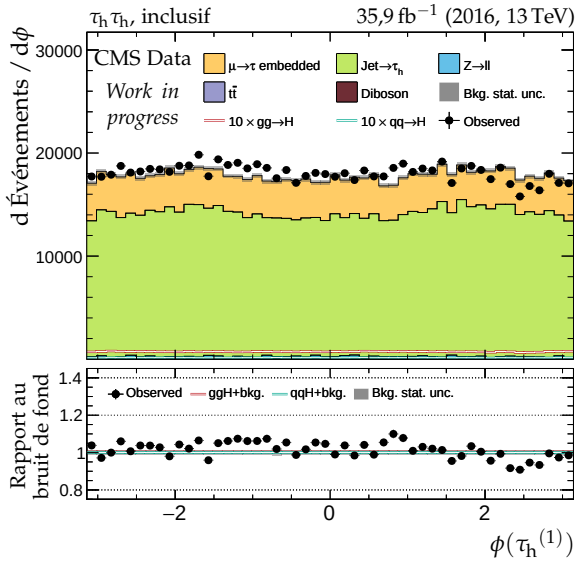
(b) *Impulsion transverse du lepton 2 ($\tau_h^{(2)}$).*



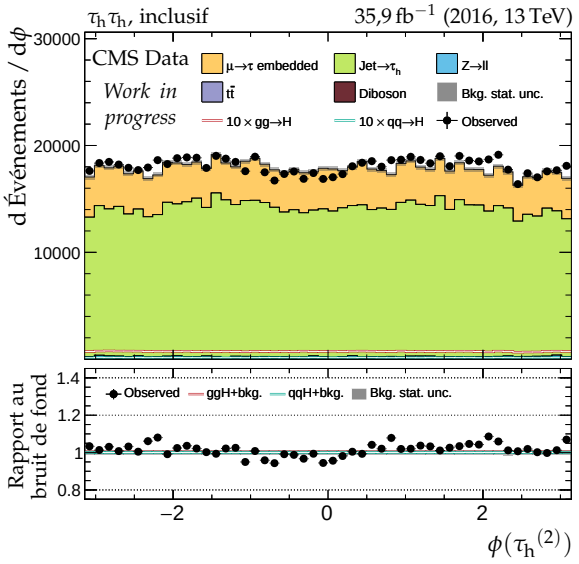
(c) *Pseudo-rapidité du lepton 1 ($\tau_h^{(1)}$).*



(d) *Pseudo-rapidité du lepton 2 ($\tau_h^{(2)}$).*

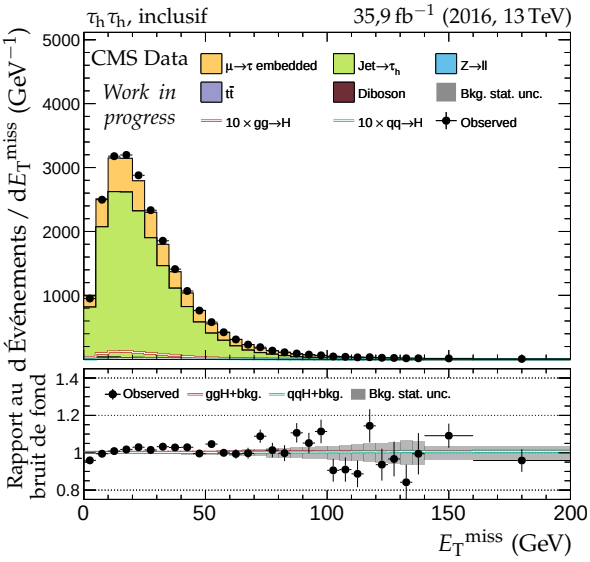


(e) *Angle azimutal du lepton 1 ($\tau_h^{(1)}$).*

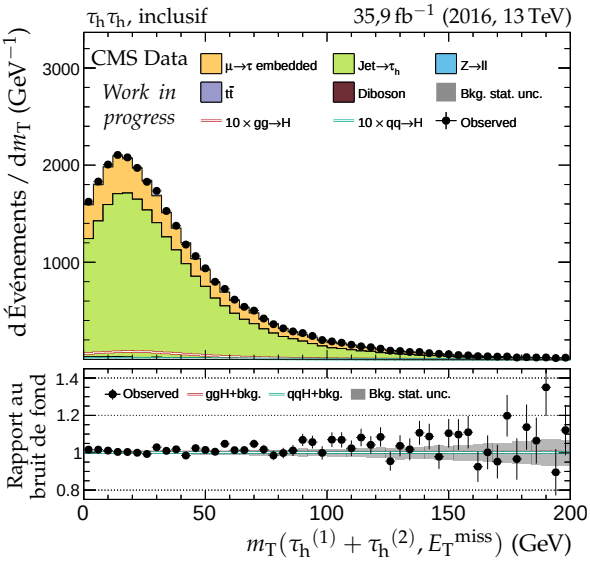


(f) *Angle azimutal du lepton 2 ($\tau_h^{(2)}$).*

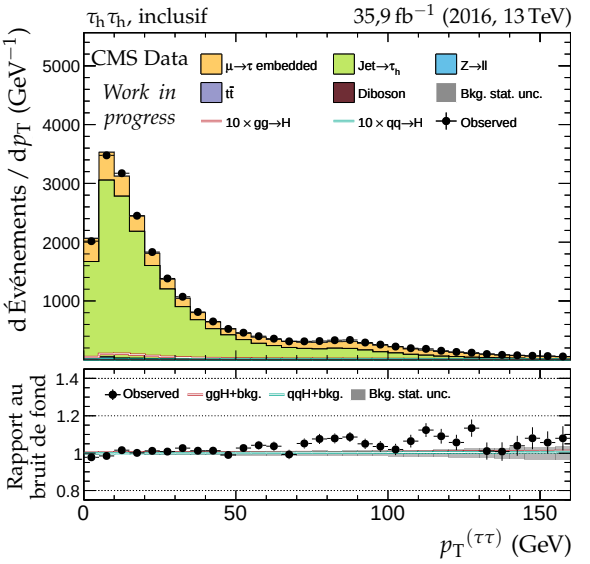
Figure G.3 – *Canal $\tau_h\tau_h$, 2016 : cinématique des leptons ($\tau_h^{(1)}$, $\tau_h^{(2)}$).*



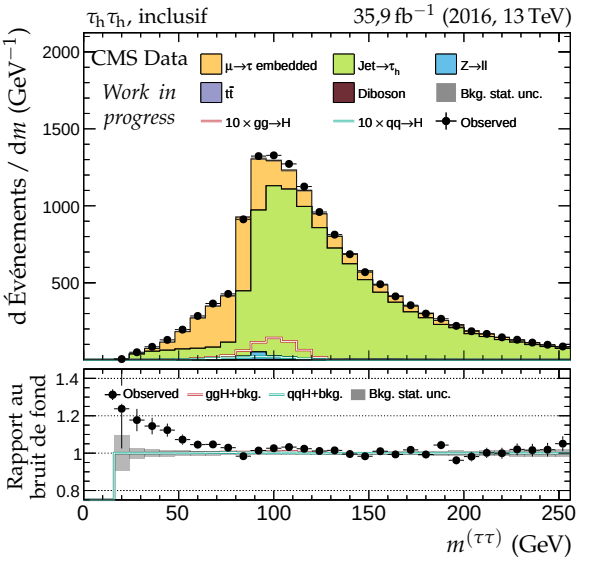
(a) *Énergie transverse manquante.*



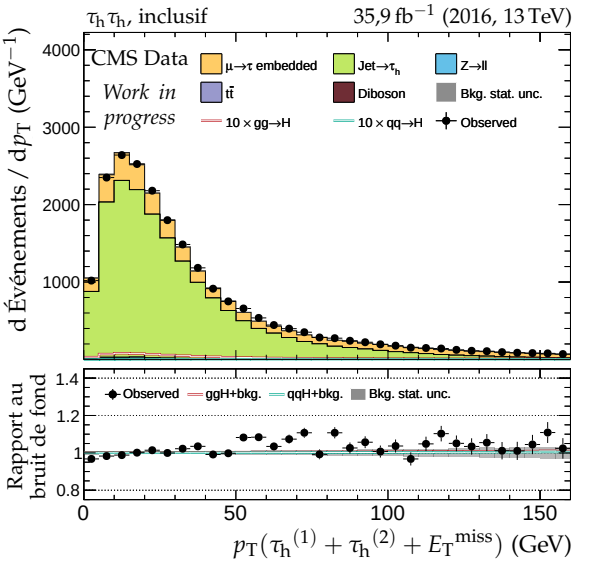
(b) *Masse transverse du dilepton.*



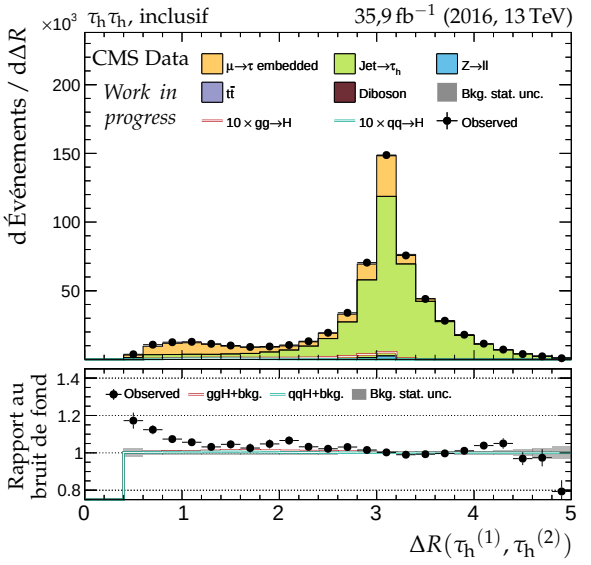
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

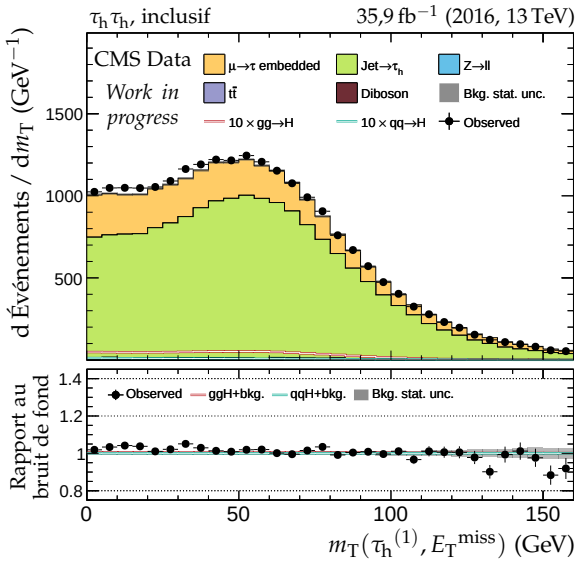


(e) *Impulsion transverse du système di-τ.*

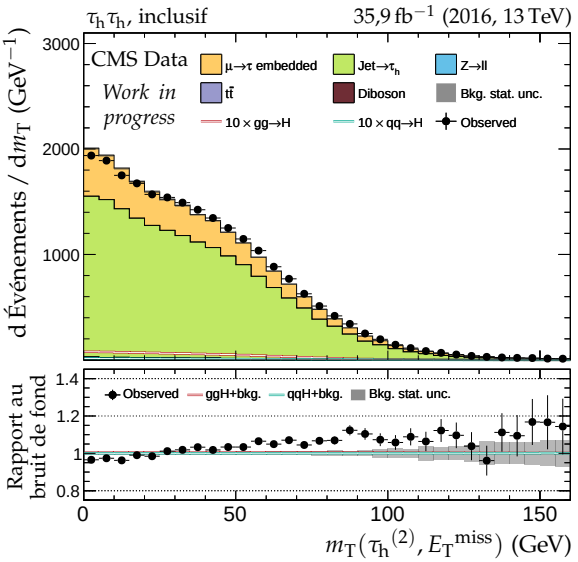


(f) *Distance ΔR entre les leptons ($\tau_h^{(1)}$, $\tau_h^{(2)}$).*

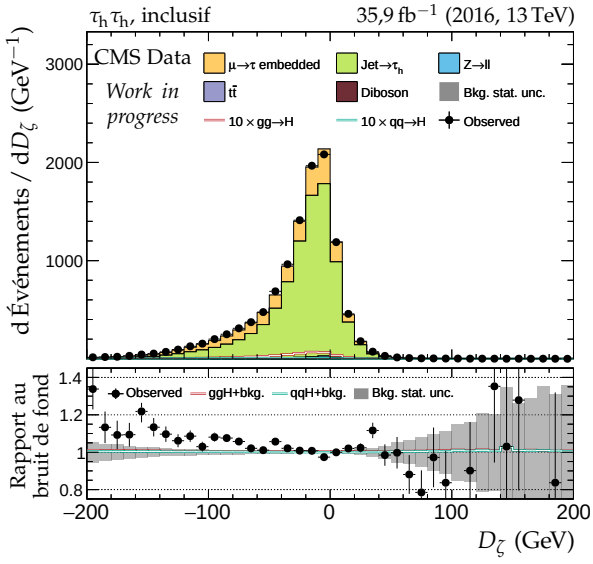
Figure G.4 – *Canal $\tau_h\tau_h$, 2016 : dilepton et énergie transverse manquante.*



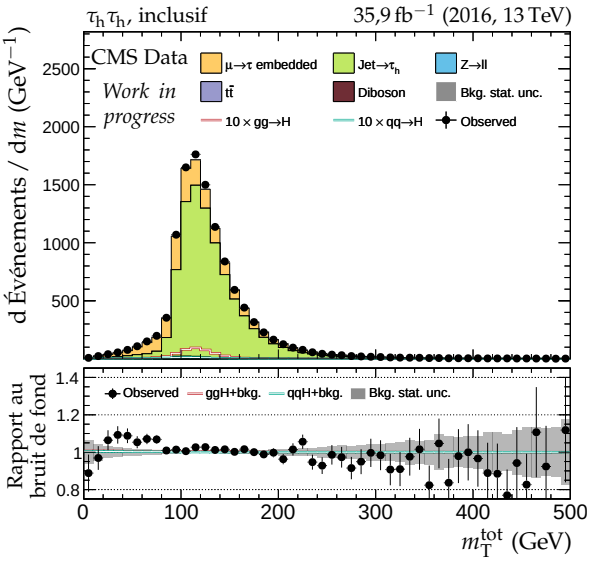
(a) *Masse transverse du lepton 1 ($\tau_h{}^{(1)}$).*



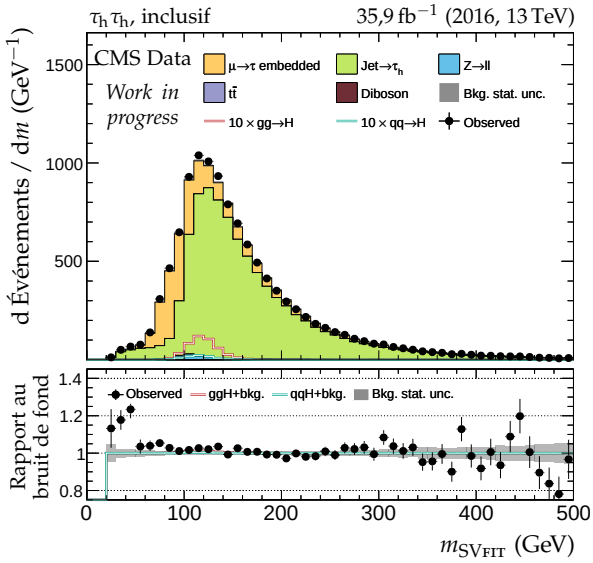
(b) *Masse transverse du lepton 2 ($\tau_h{}^{(2)}$).*



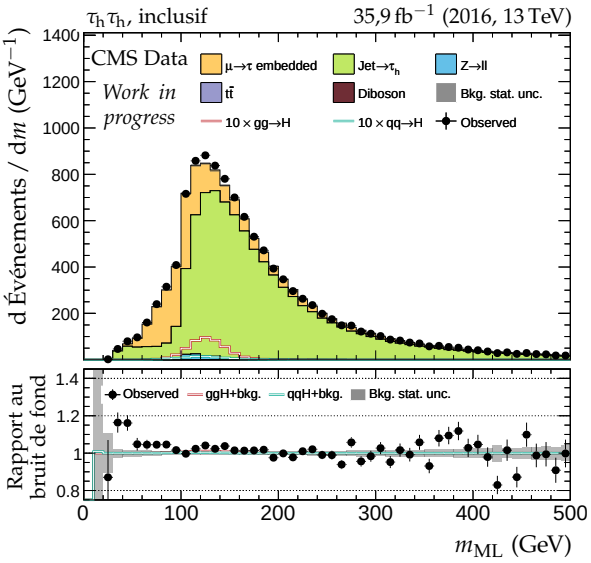
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

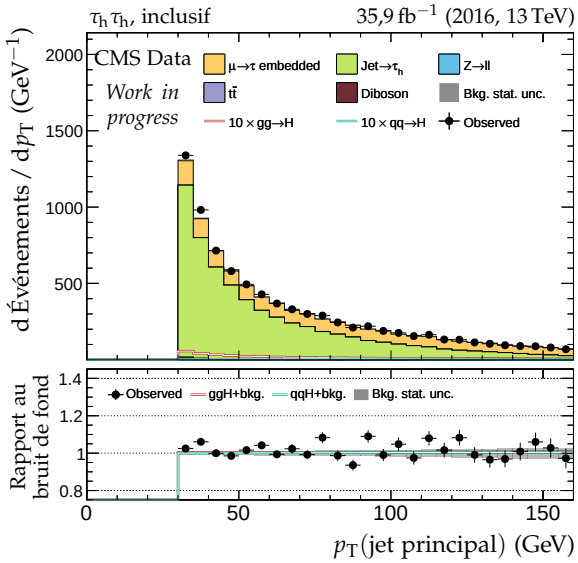


(e) *Masse du système di-τ d'après* SVFIT.

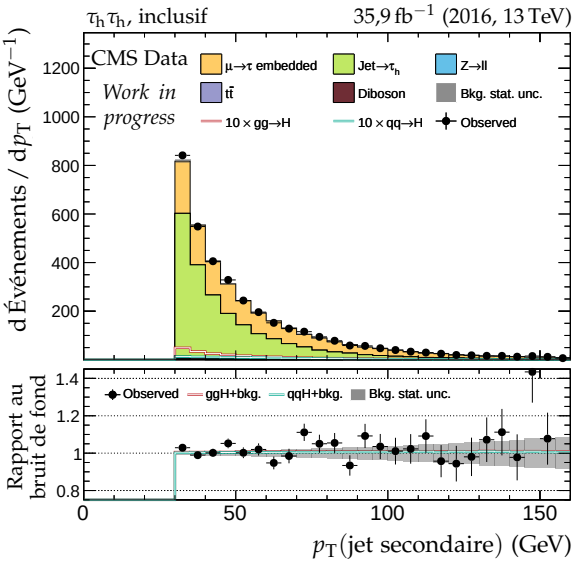


(f) *Masse du système di-τ d'après le ML.*

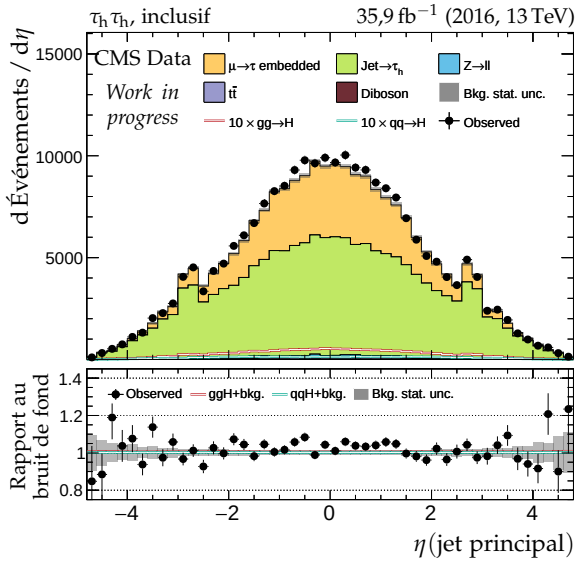
Figure G.5 – *Canal $\tau_h\tau_h$, 2016 : masses transverses, D_ζ et masses.*



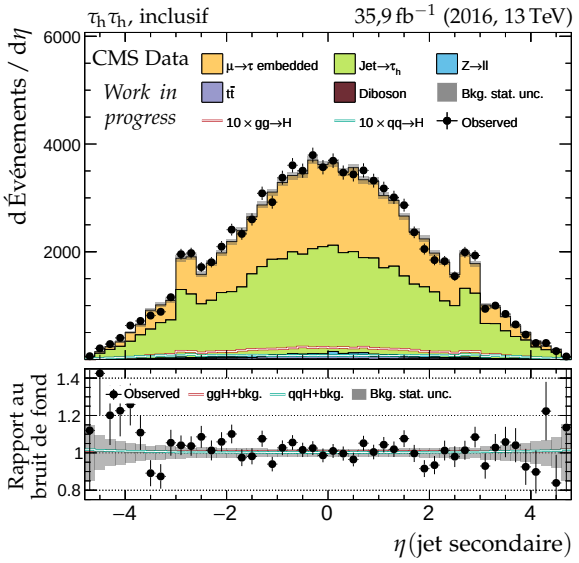
(a) *Impulsion transverse du jet principal.*



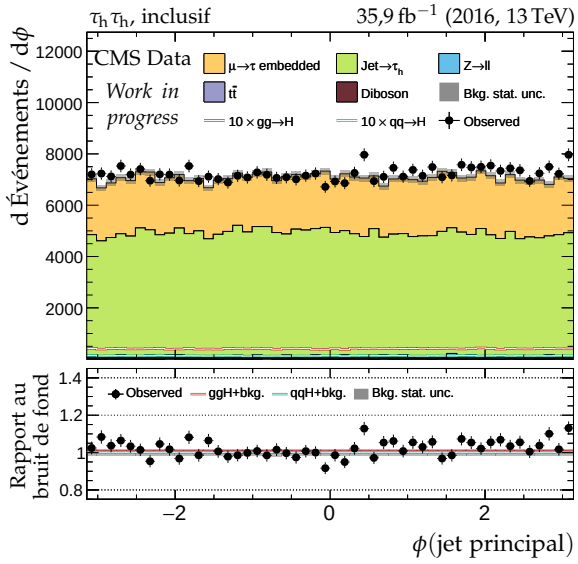
(b) *Impulsion transverse du jet secondaire.*



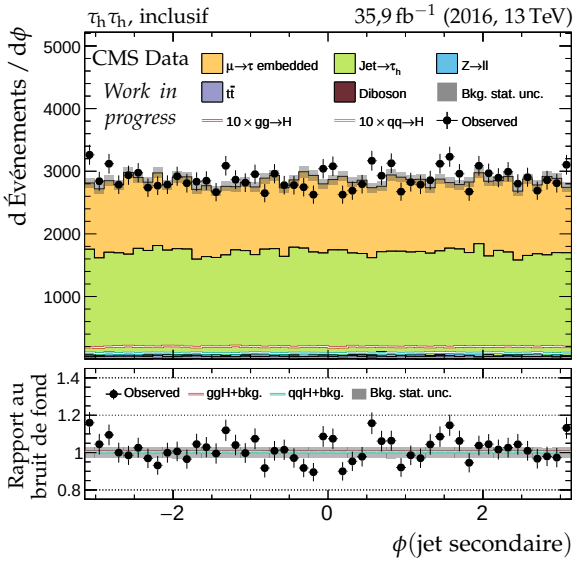
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

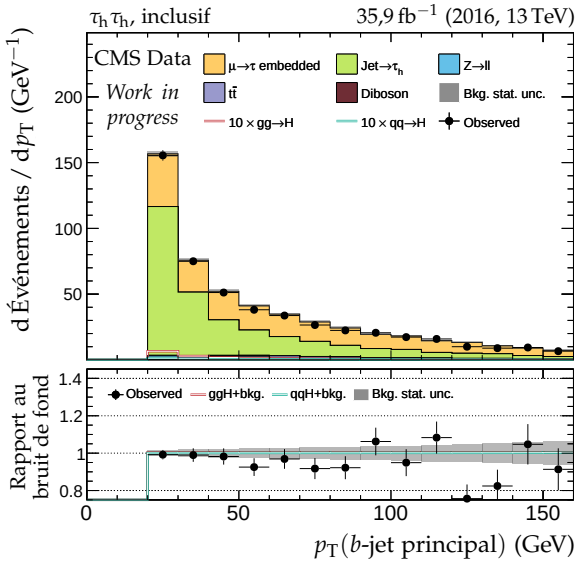


(e) *Angle azimutal du jet principal.*

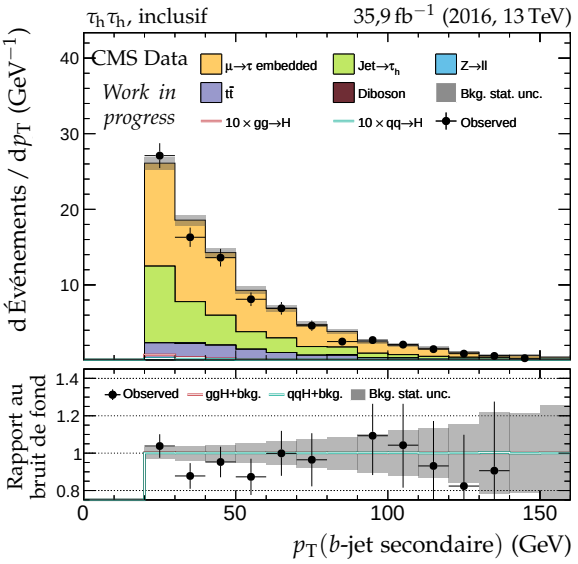


(f) *Angle azimutal du jet secondaire.*

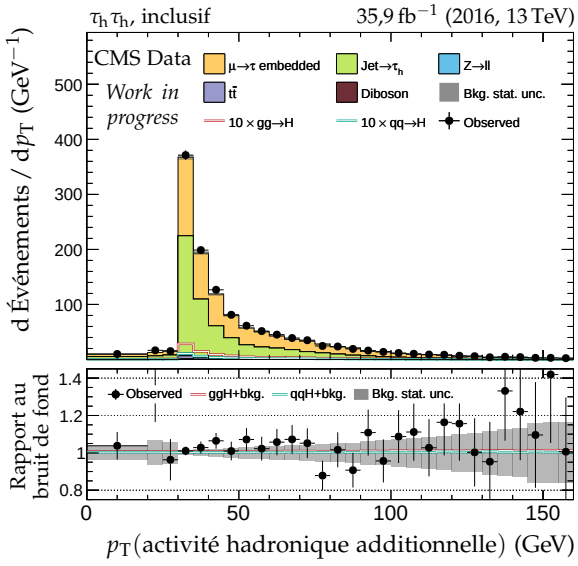
Figure G.6 – *Canal $\tau_h\tau_h$, 2016 : cinématique des deux jets principaux.*



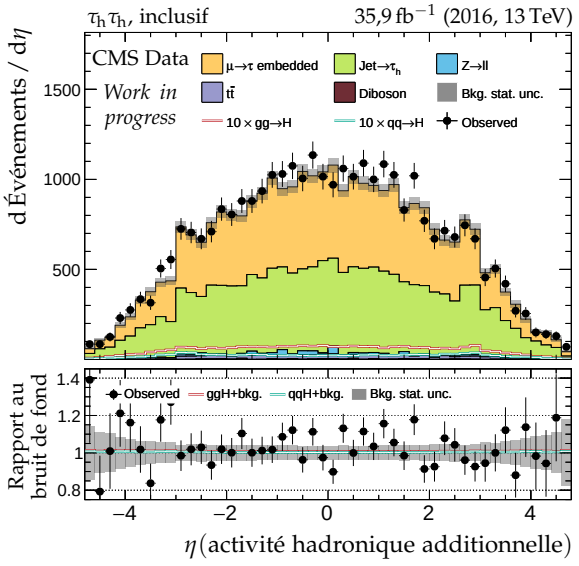
(a) *Impulsion transverse du b-jet principal.*



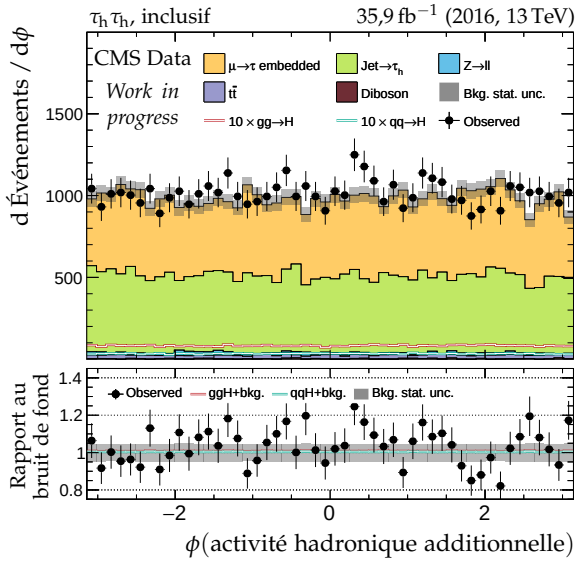
(b) *Impulsion transverse du b-jet secondaire.*



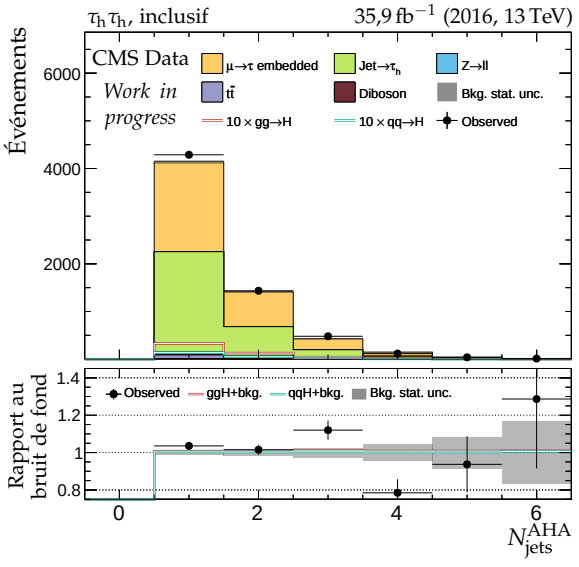
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

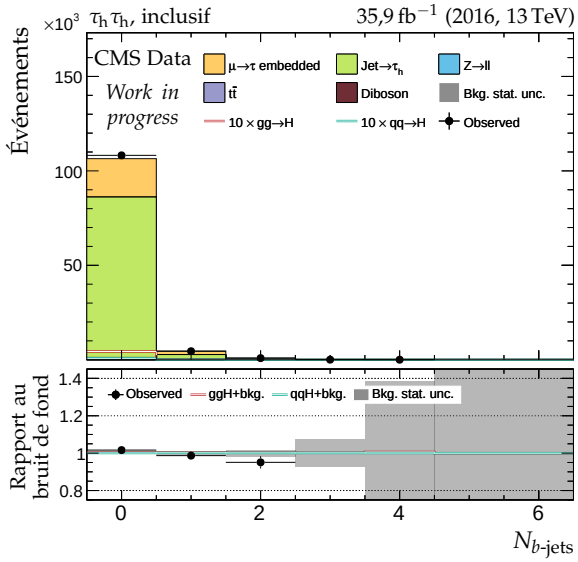


(e) *Angle azimutal de l'AHA.*

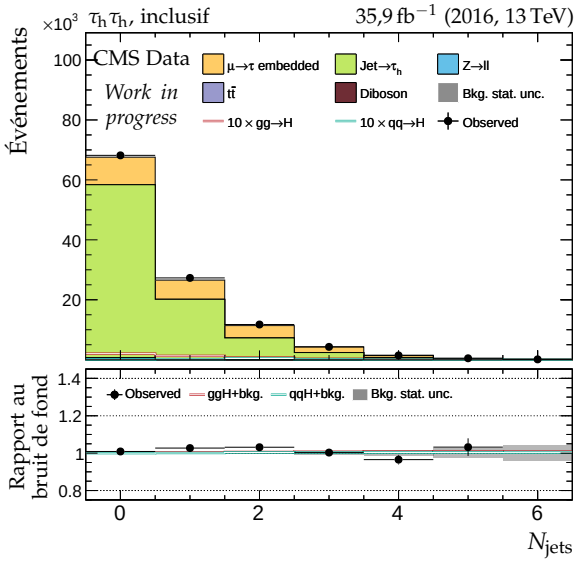


(f) *Nombre de jets dans l'AHA.*

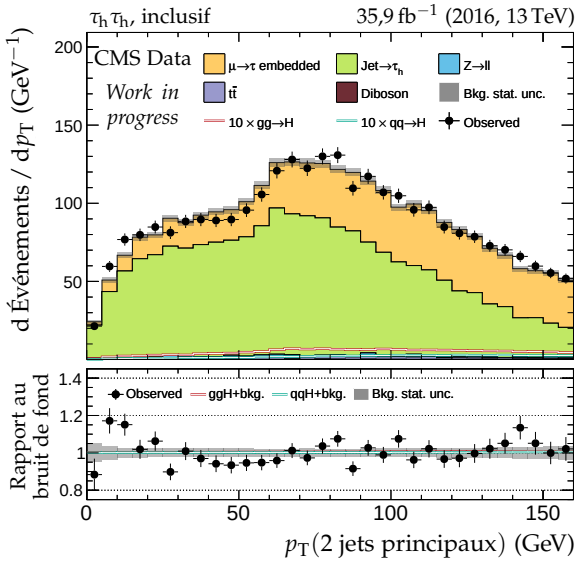
Figure G.7 – *Canal $\tau_h\tau_h$, 2016 : b-jets et activité hadronique additionnelle.*



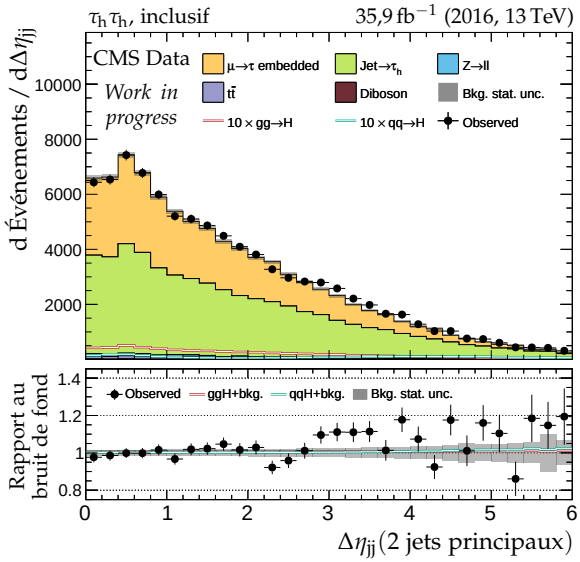
(a) *Nombre de b-jets.*



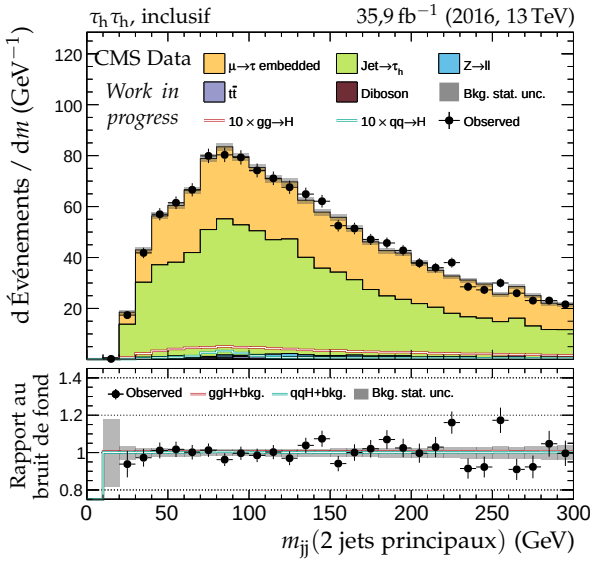
(b) *Nombre de jets.*



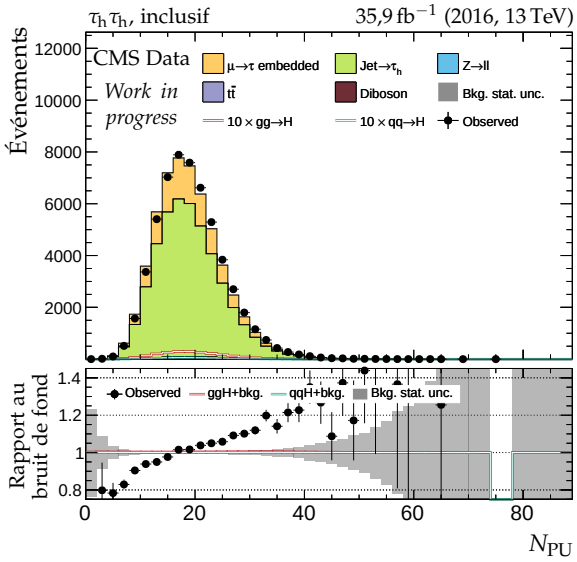
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

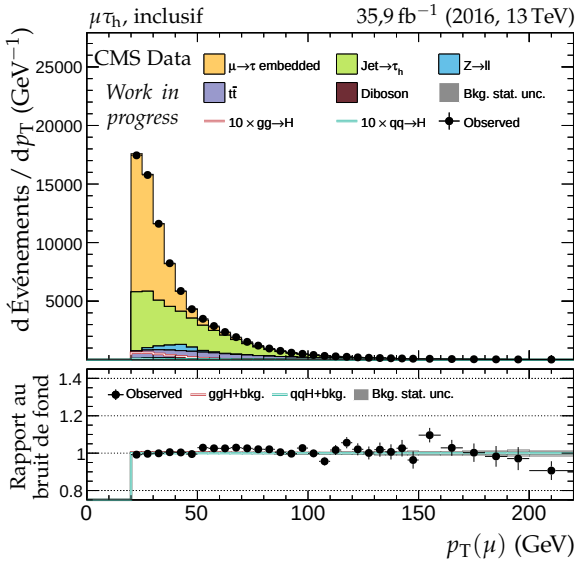


(e) *Masse invariante du système des deux jets.*

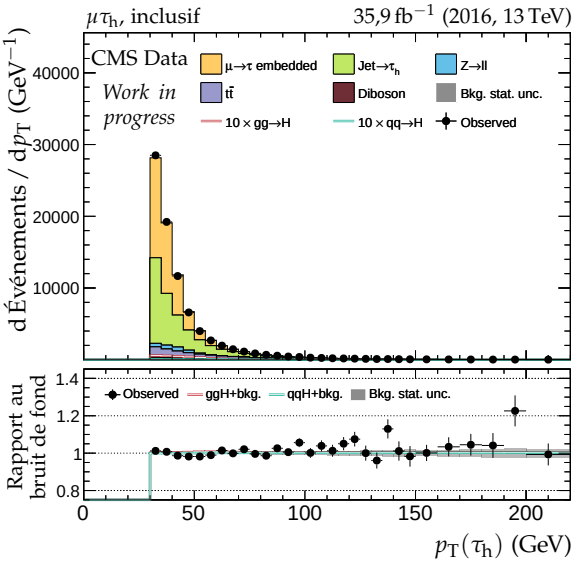


(f) *Nombre de vertex d'empilement.*

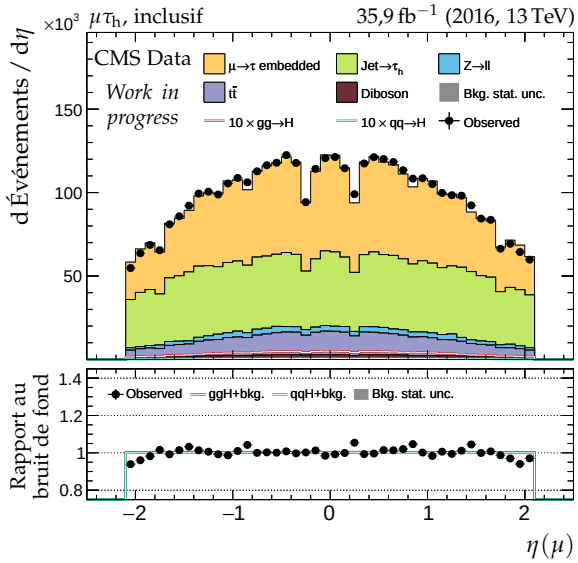
Figure G.8 – *Canal $\tau_h\tau_h$, 2016 : nombre de jets, système des deux jets principaux et empilement.*



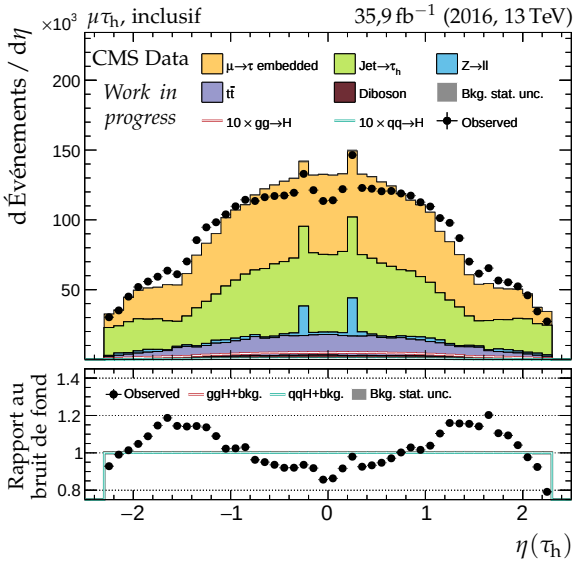
(a) *Impulsion transverse du lepton 1 (μ).*



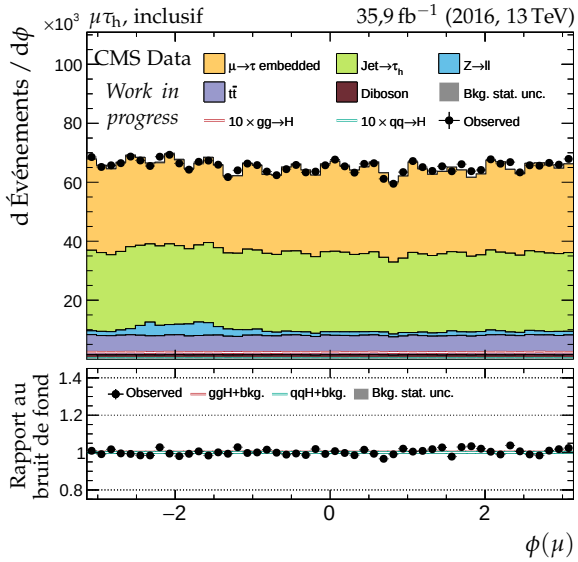
(b) *Impulsion transverse du lepton 2 (τ_h).*



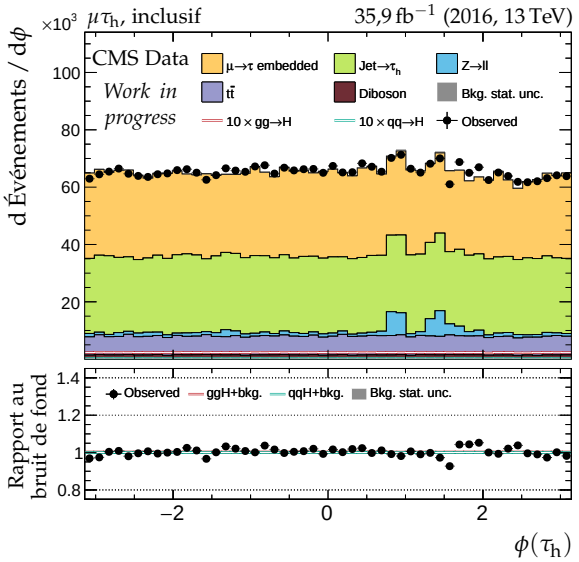
(c) *Pseudo-rapidité du lepton 1 (μ).*



(d) *Pseudo-rapidité du lepton 2 (τ_h).*

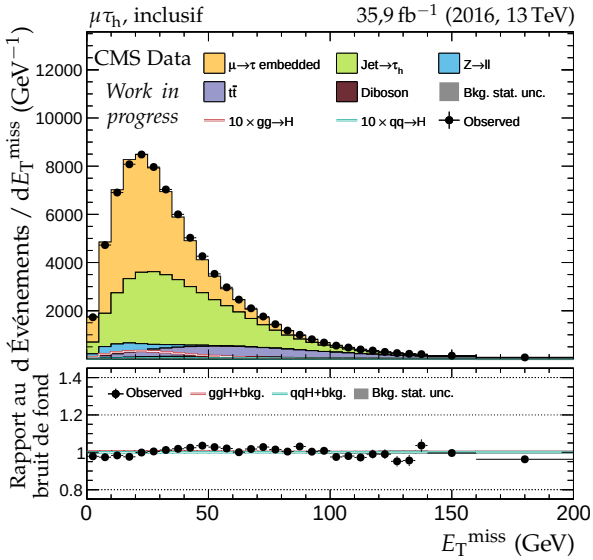


(e) *Angle azimutal du lepton 1 (μ).*

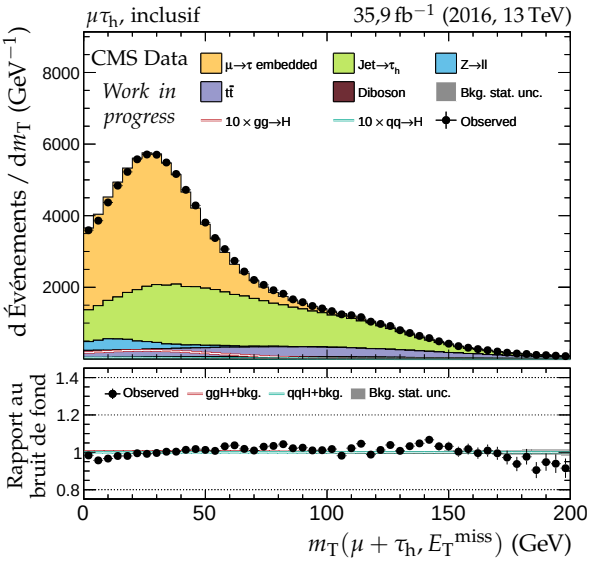


(f) *Angle azimutal du lepton 2 (τ_h).*

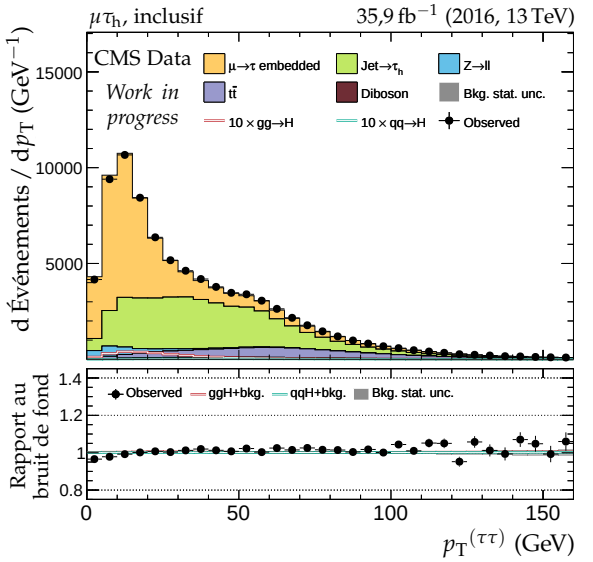
Figure G.9 – *Canal $\mu\tau_h$, 2016 : cinématique des leptons (μ, τ_h).*



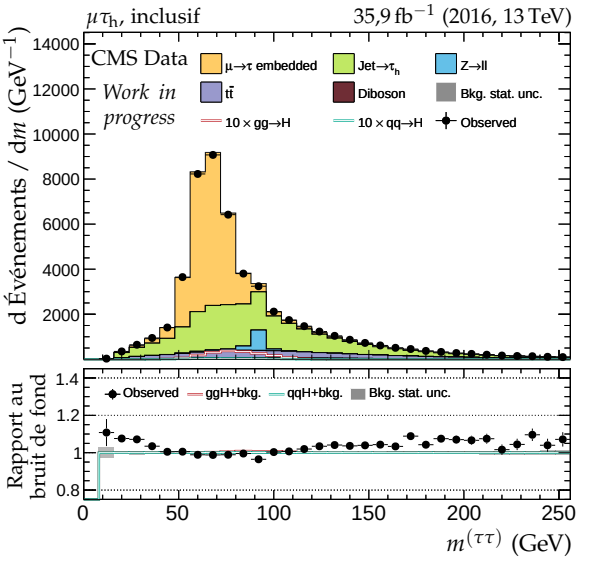
(a) *Énergie transverse manquante.*



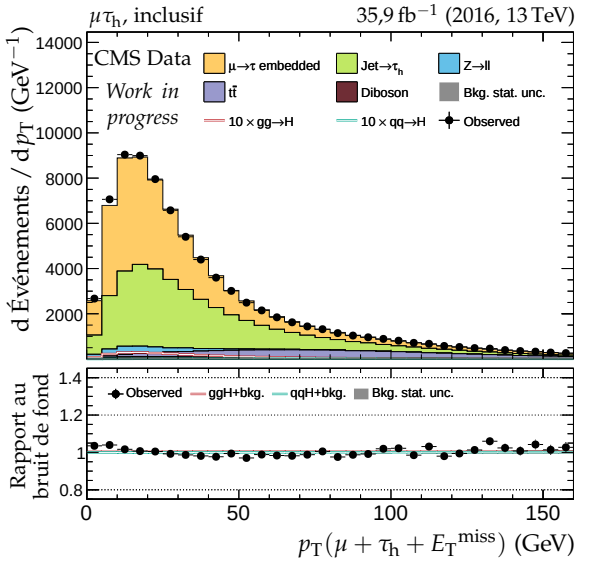
(b) *Masse transverse du dilepton.*



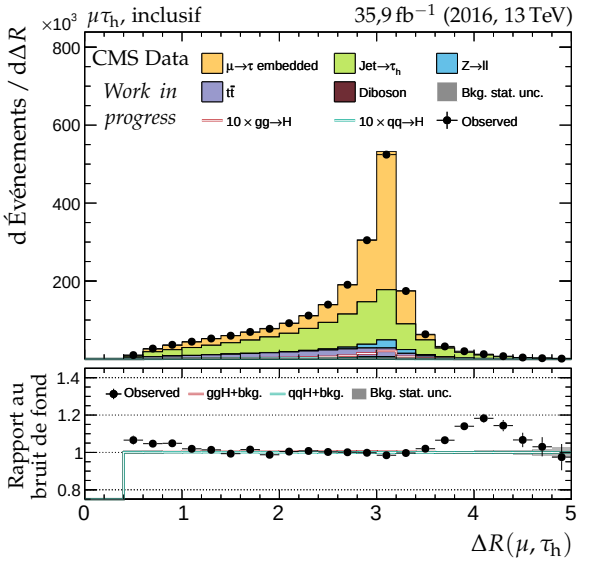
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

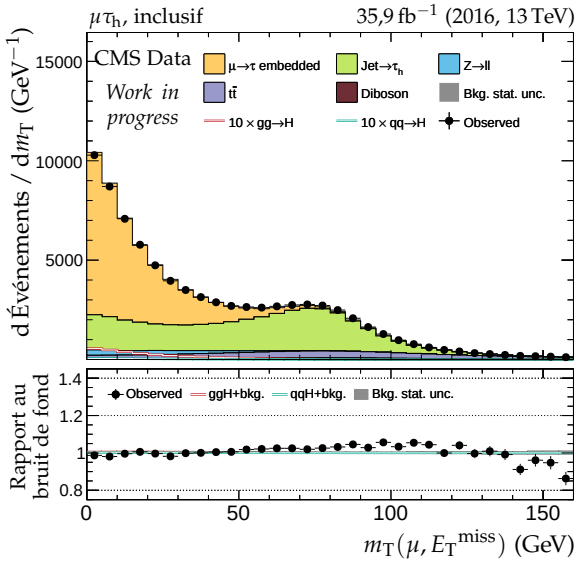


(e) *Impulsion transverse du système di-τ.*

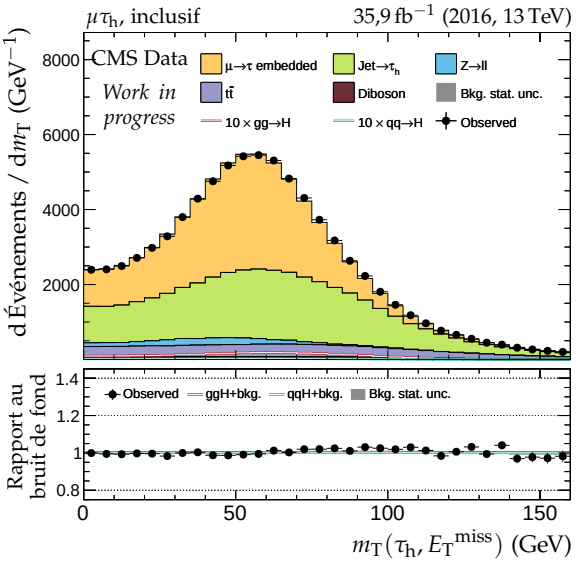


(f) *Distance ΔR entre les leptons (μ, τ_h).*

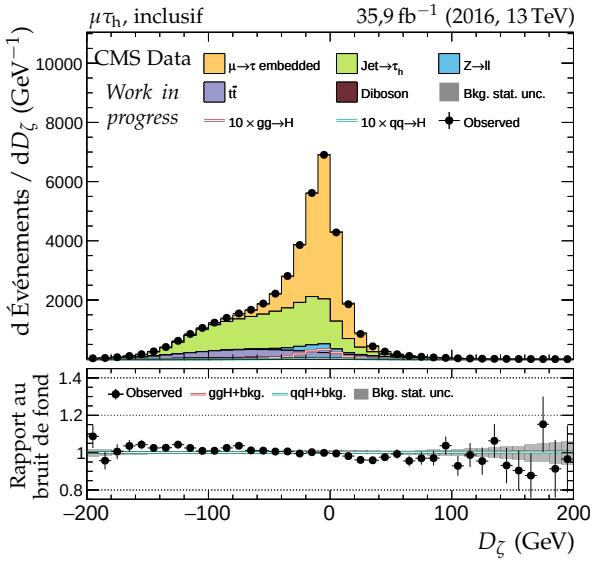
Figure G.10 – *Canal $\mu\tau_h$, 2016 : dilepton et énergie transverse manquante.*



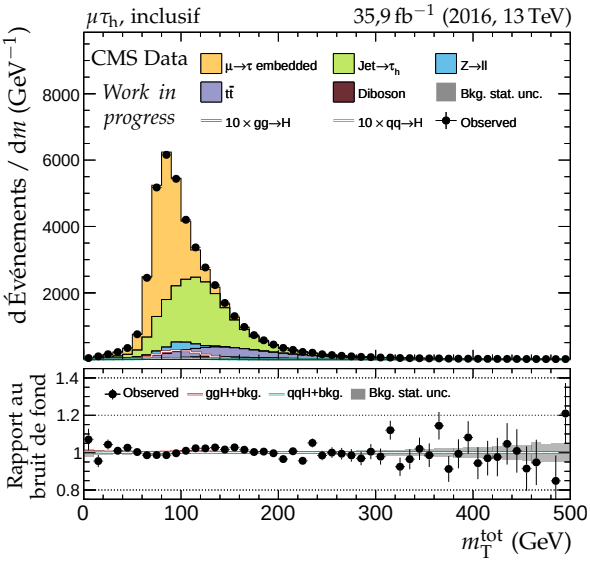
(a) *Masse transverse du lepton 1 (μ).*



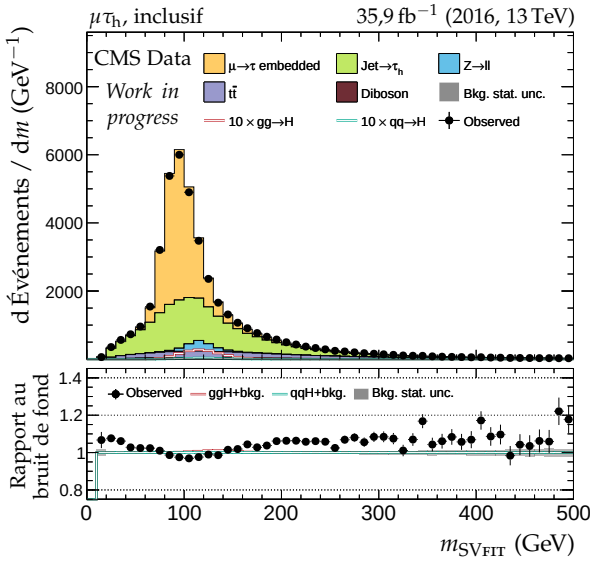
(b) *Masse transverse du lepton 2 (τ_h).*



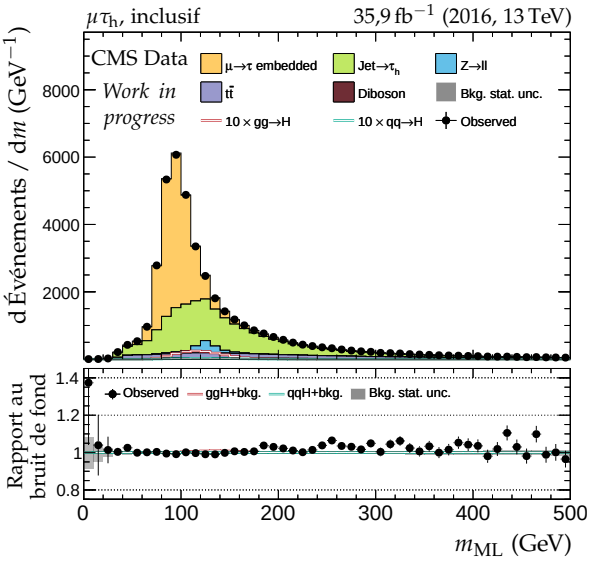
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

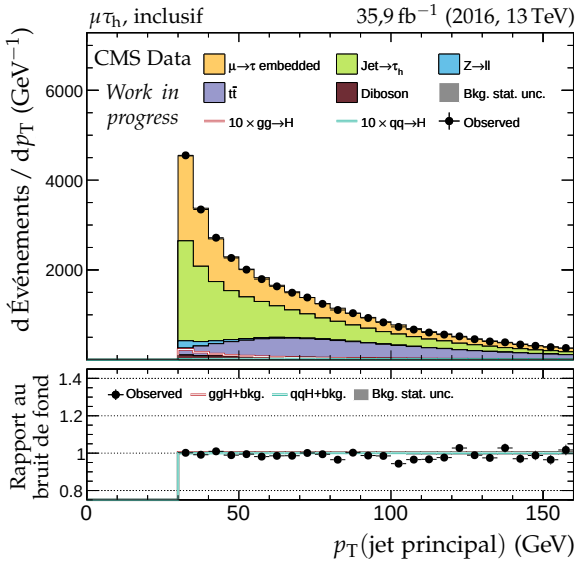


(e) *Masse du système di-τ d'après* SVFIT.

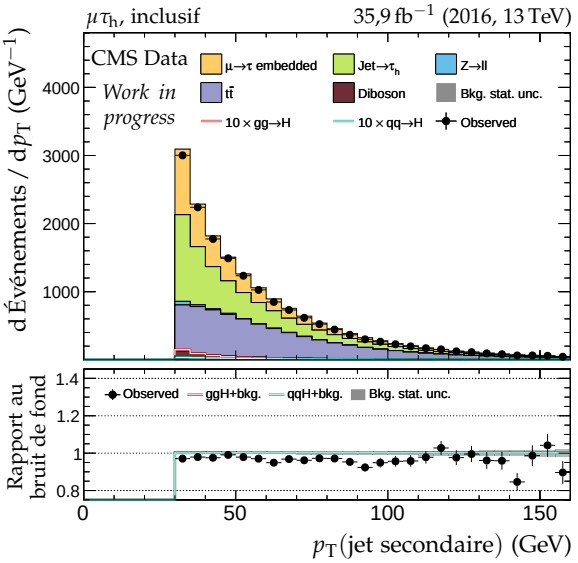


(f) *Masse du système di-τ d'après le ML.*

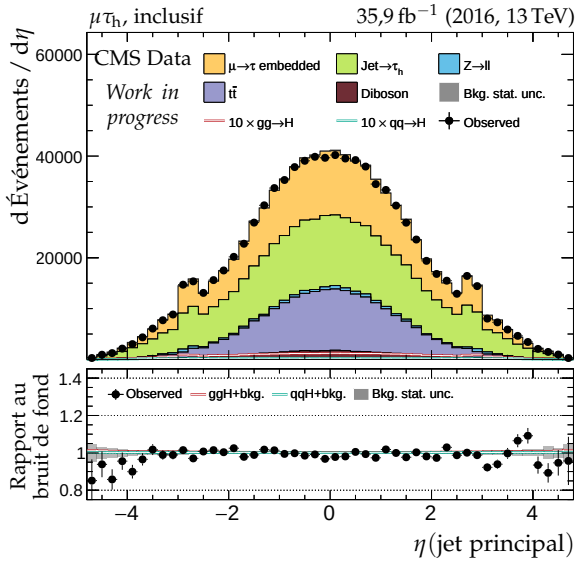
Figure G.11 – *Canal $\mu\tau_h$, 2016 : masses transverses, D_ζ et masses.*



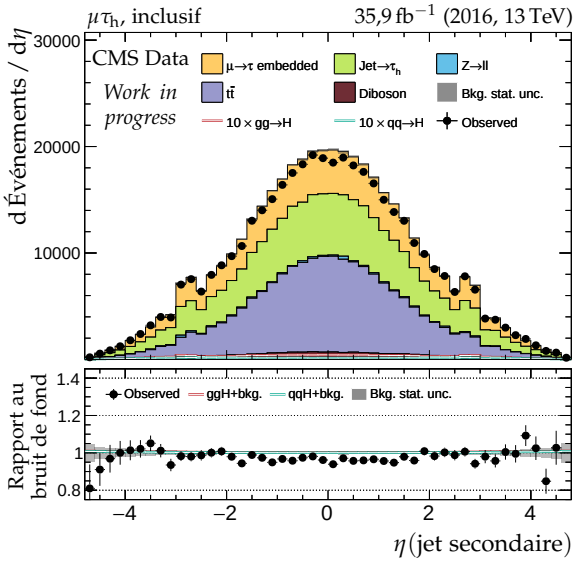
(a) *Impulsion transverse du jet principal.*



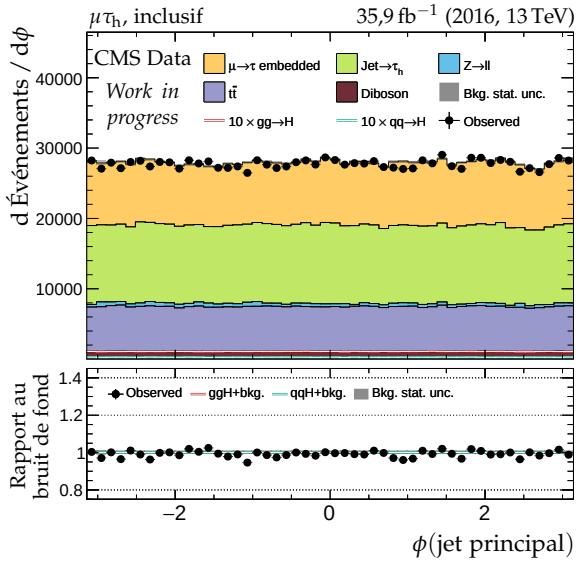
(b) *Impulsion transverse du jet secondaire.*



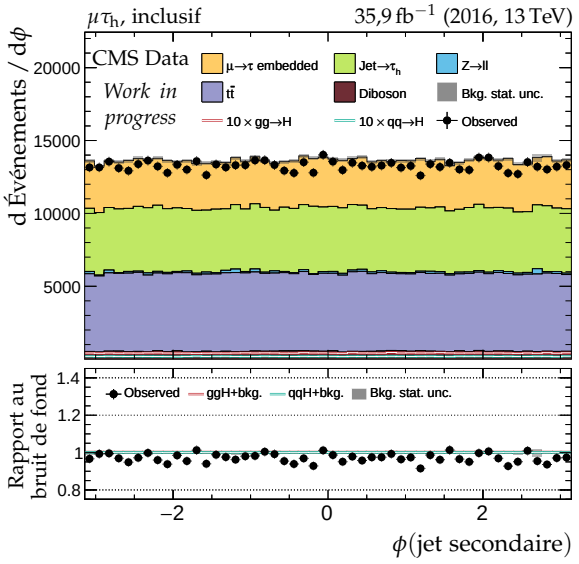
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

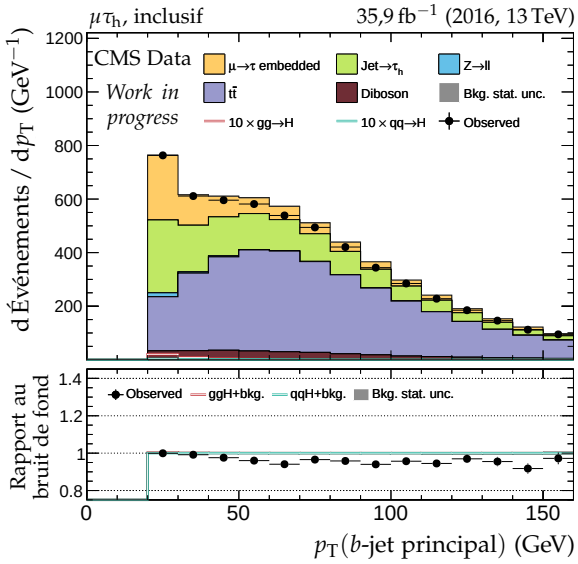


(e) *Angle azimutal du jet principal.*

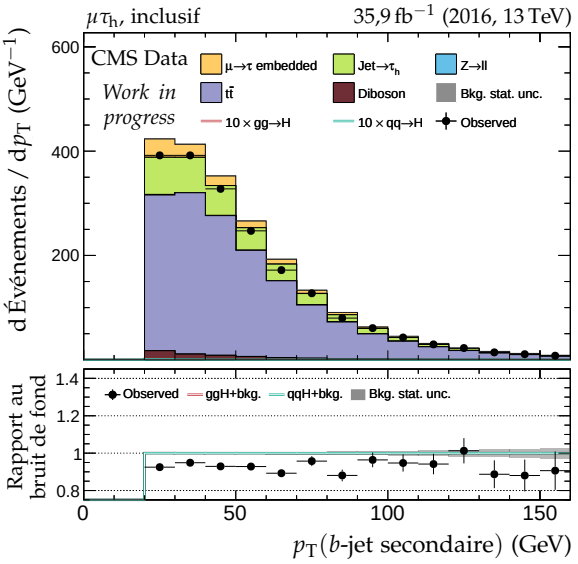


(f) *Angle azimutal du jet secondaire.*

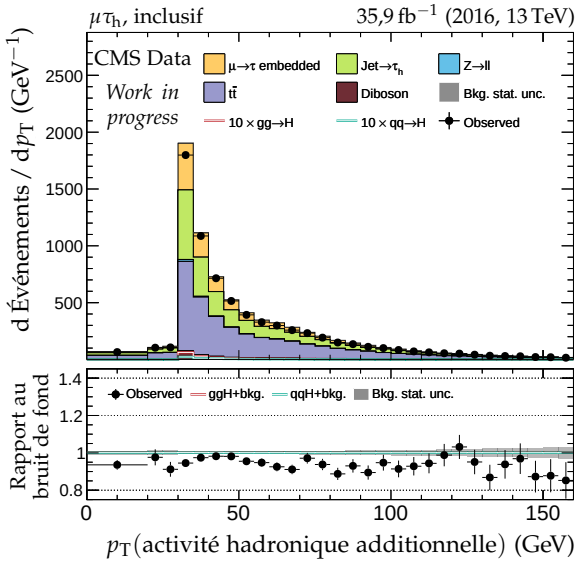
Figure G.12 – *Canal $\mu\tau_h$, 2016 : cinématique des deux jets principaux.*



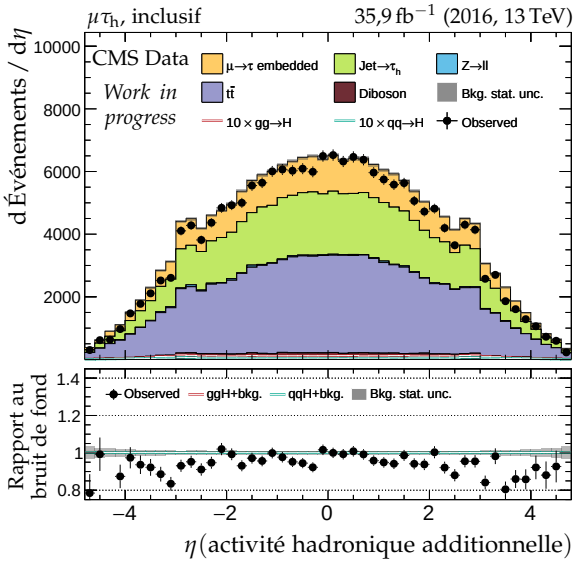
(a) *Impulsion transverse du b-jet principal.*



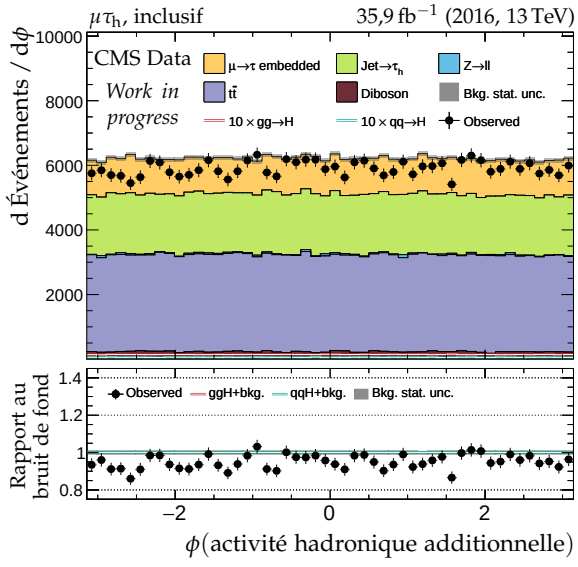
(b) *Impulsion transverse du b-jet secondaire.*



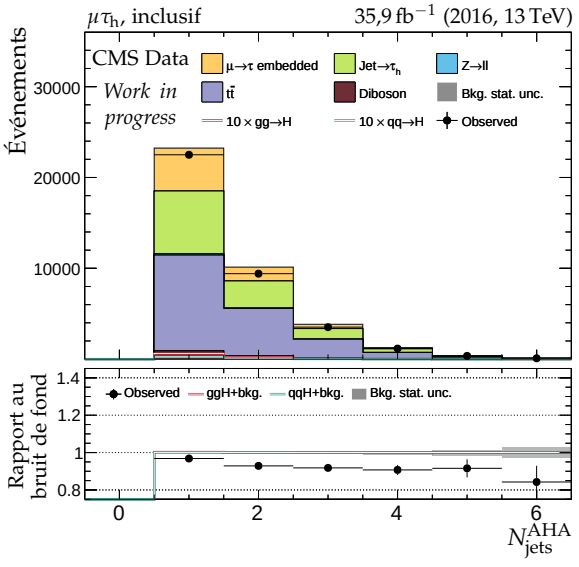
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

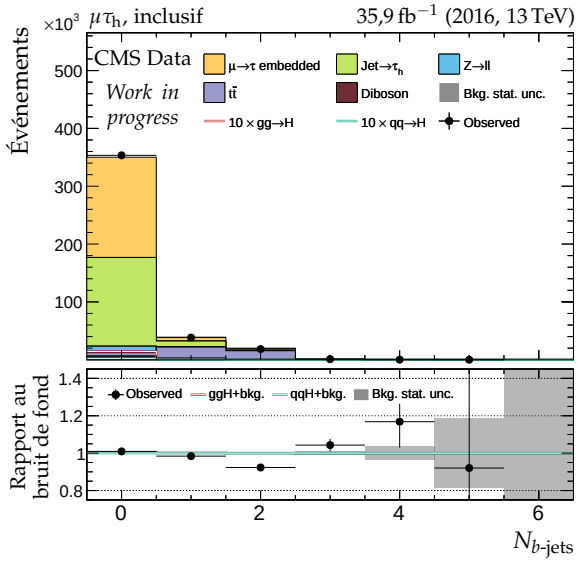


(e) *Angle azimutal de l'AHA.*

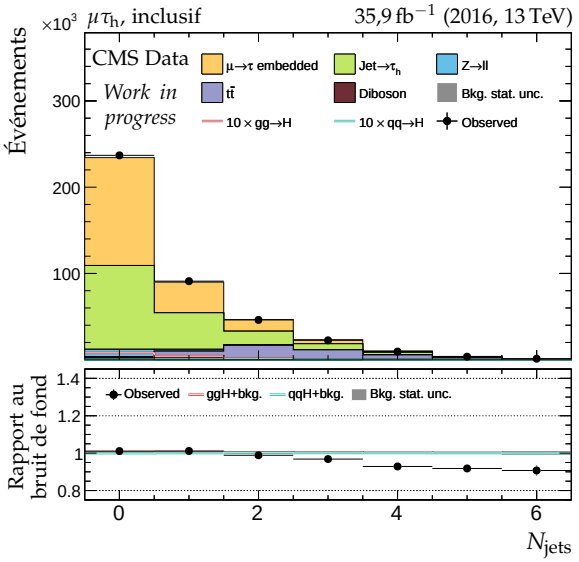


(f) *Nombre de jets dans l'AHA.*

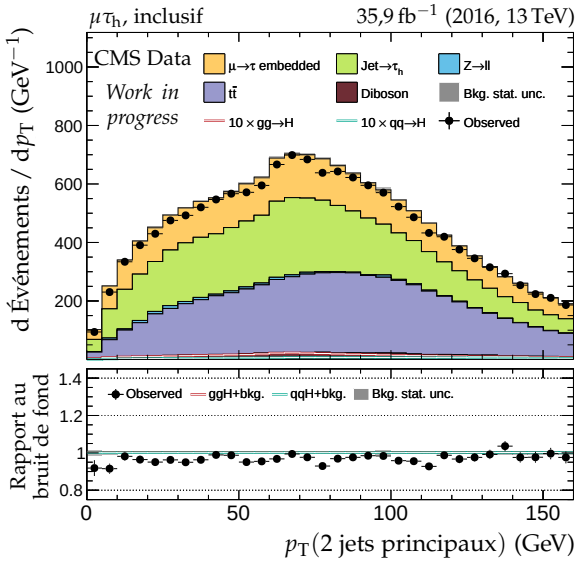
Figure G.13 – *Canal $\mu\tau_h$, 2016 : b-jets et activité hadronique additionnelle.*



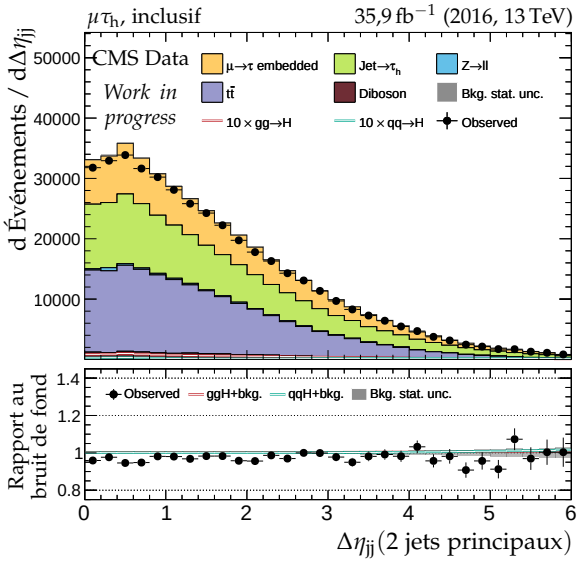
(a) *Nombre de b-jets.*



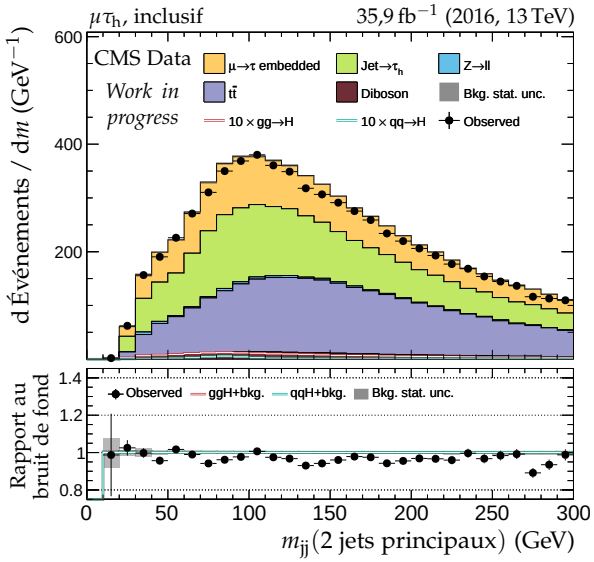
(b) *Nombre de jets.*



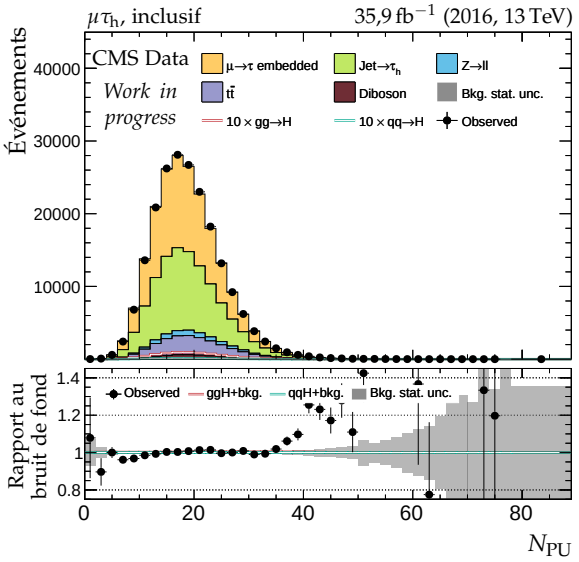
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

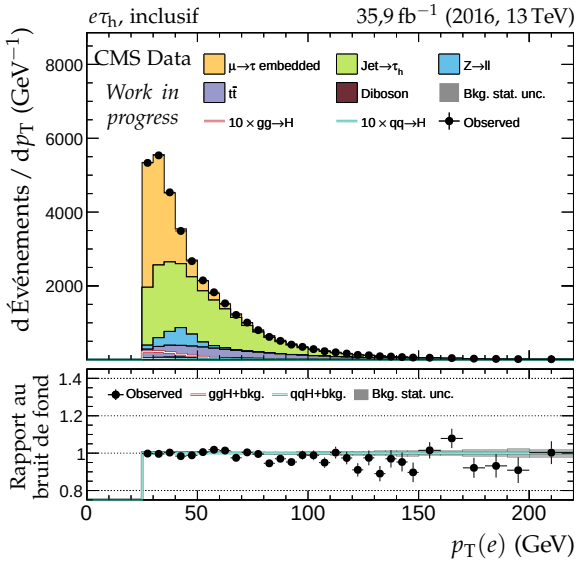


(e) *Masse invariante du système des deux jets.*

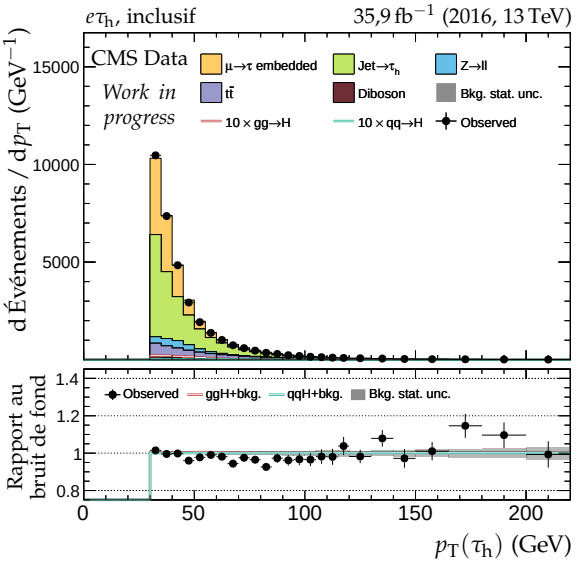


(f) *Nombre de vertex d'empilement.*

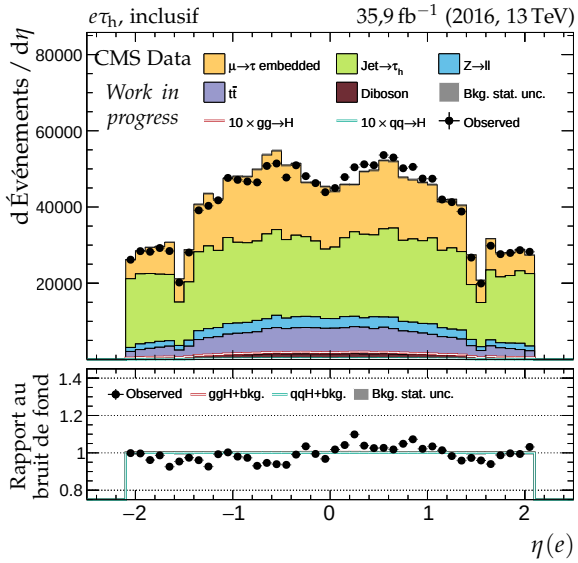
Figure G.14 – *Canal $\mu\tau_h$, 2016 : nombre de jets, système des deux jets principaux et empilement.*



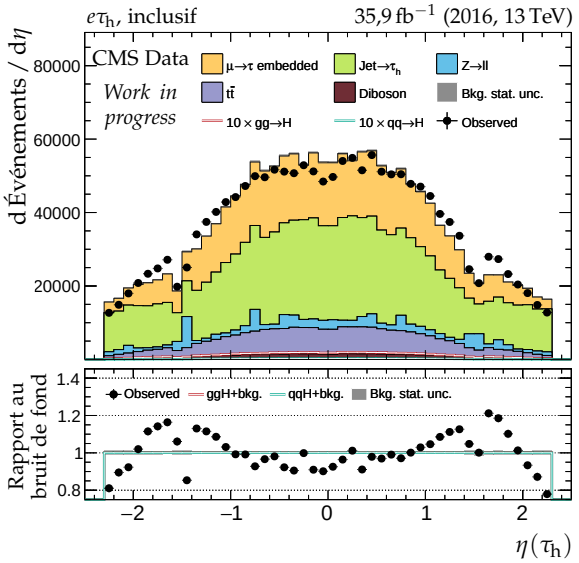
(a) *Impulsion transverse du lepton 1 (e).*



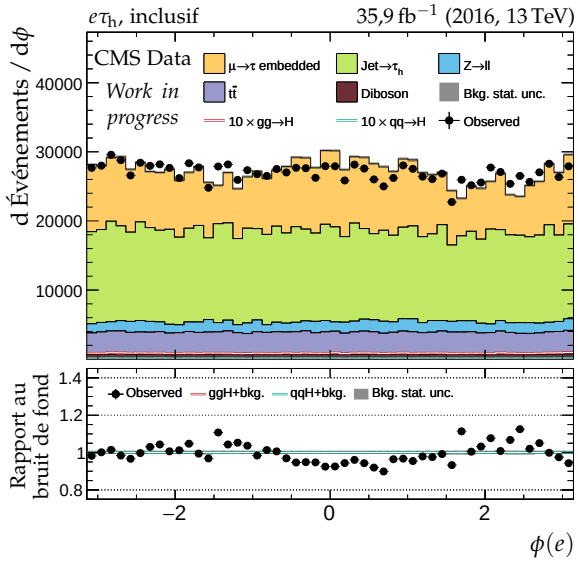
(b) *Impulsion transverse du lepton 2 (τ_h).*



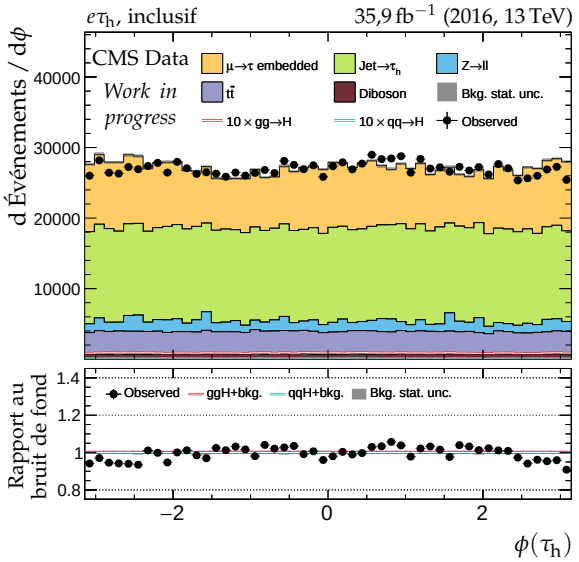
(c) *Pseudo-rapidité du lepton 1 (e).*



(d) *Pseudo-rapidité du lepton 2 (τ_h).*

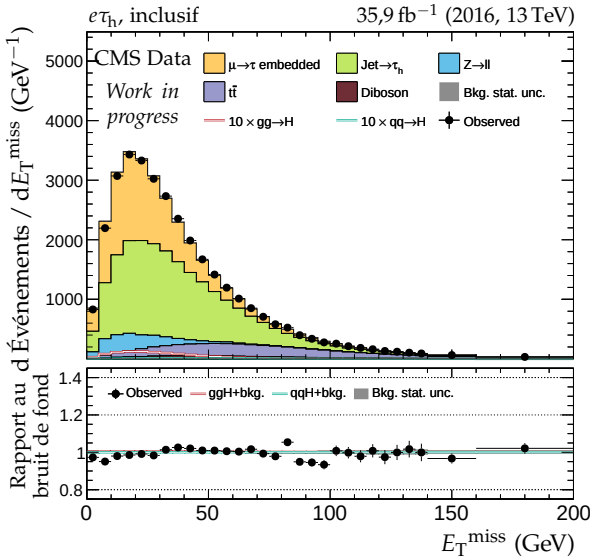


(e) *Angle azimutal du lepton 1 (e).*

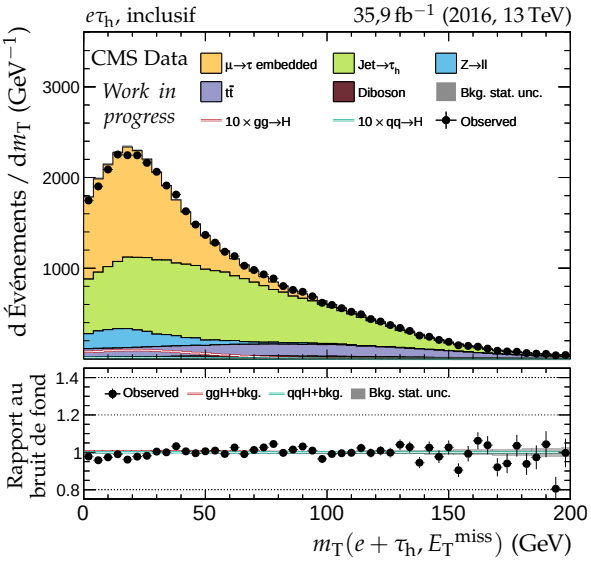


(f) *Angle azimutal du lepton 2 (τ_h).*

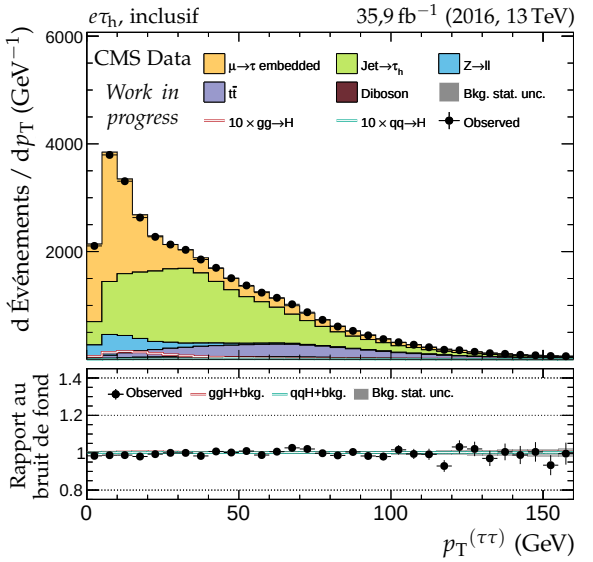
Figure G.15 – *Canal $e\tau_h$, 2016 : cinématique des leptons (e, τ_h).*



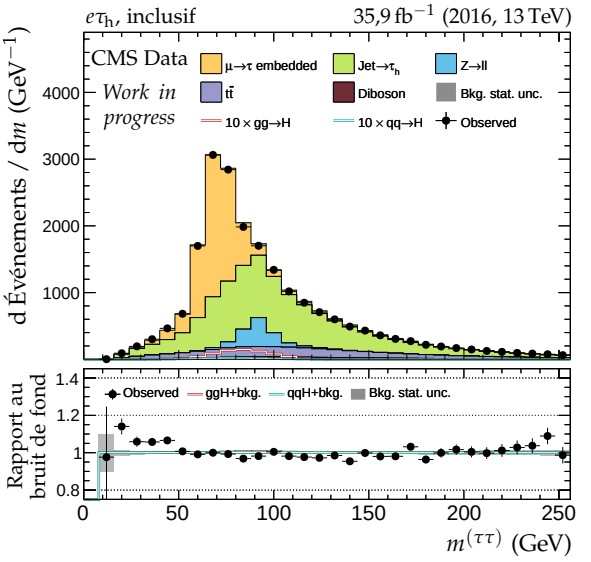
(a) *Énergie transverse manquante.*



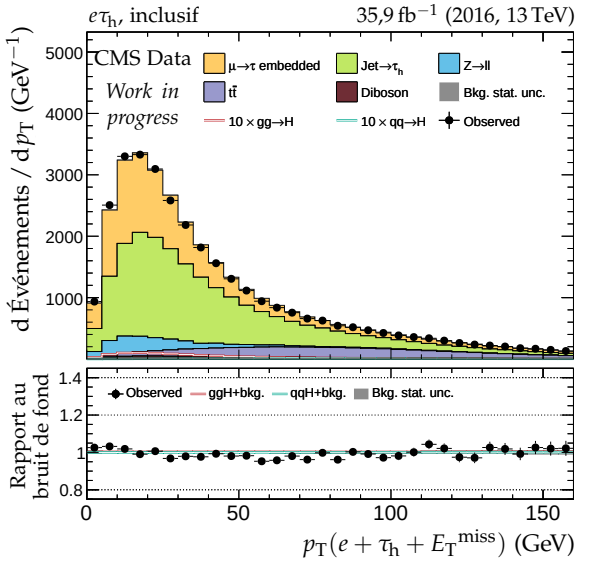
(b) *Masse transverse du dilepton.*



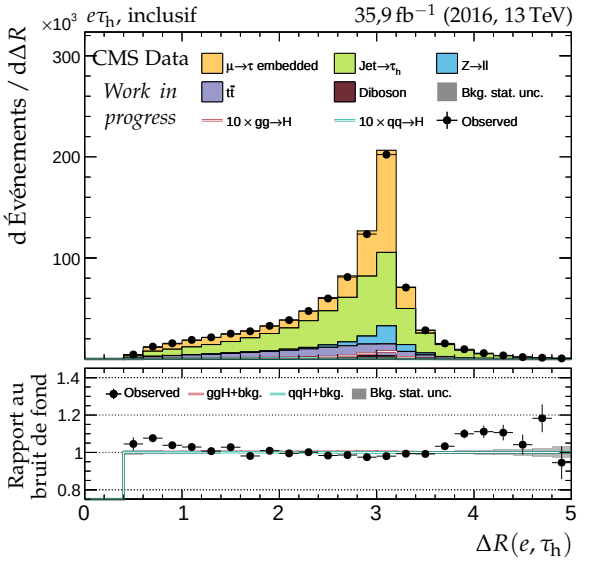
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

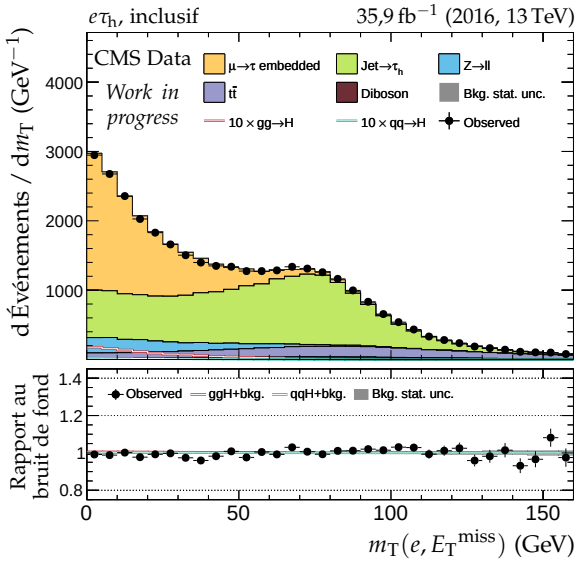


(e) *Impulsion transverse du système di-τ.*

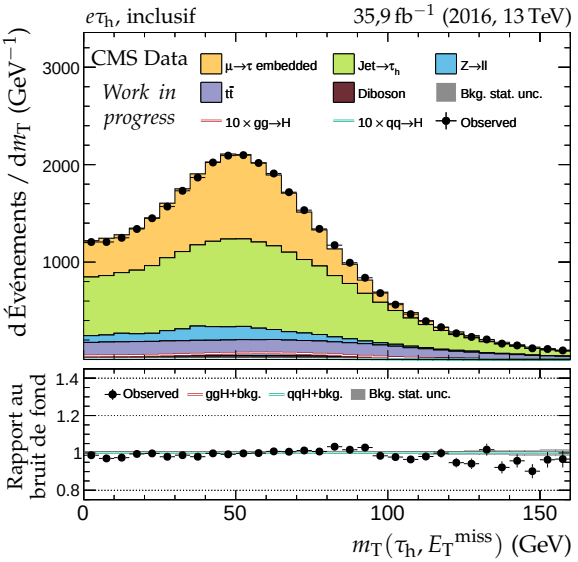


(f) *Distance ΔR entre les leptons (e, τ_h).*

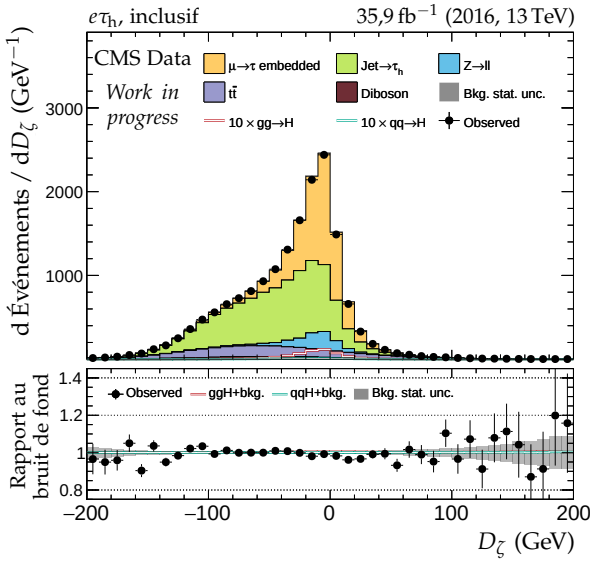
Figure G.16 – *Canal $e\tau_h$, 2016 : dilepton et énergie transverse manquante.*



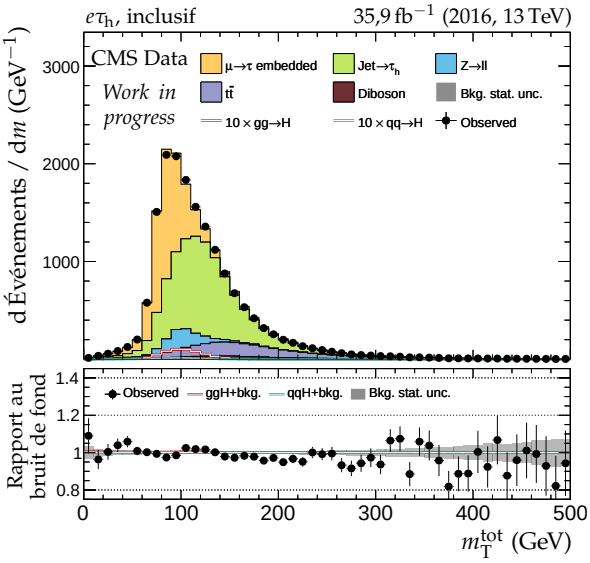
(a) *Masse transverse du lepton 1 (e).*



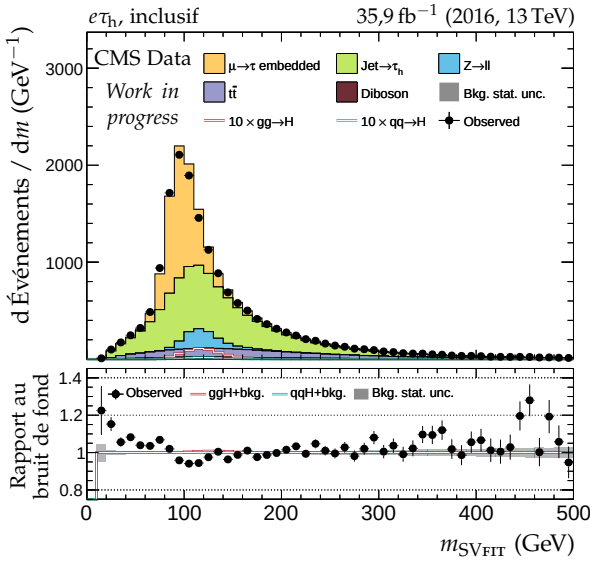
(b) *Masse transverse du lepton 2 (τ_h).*



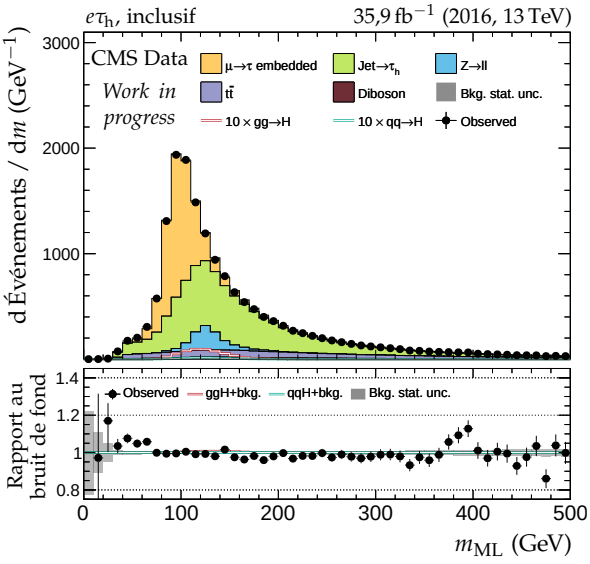
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

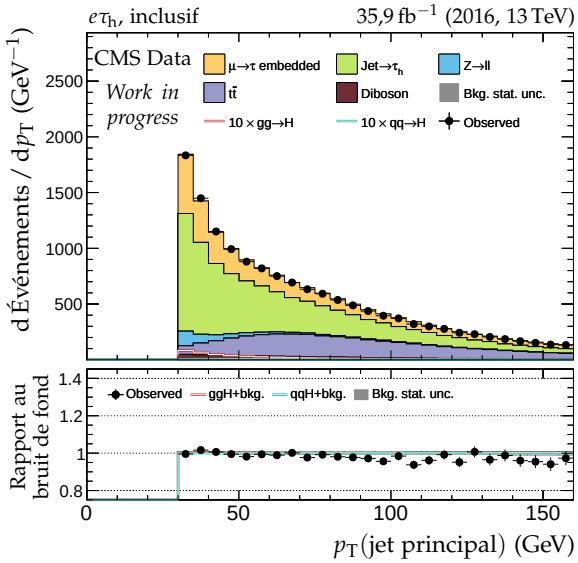


(e) *Masse du système di-τ d'après* SVFIT.

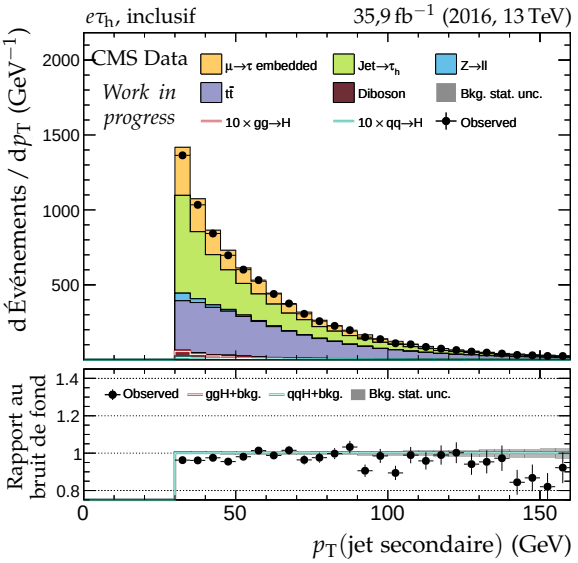


(f) *Masse du système di-τ d'après le ML.*

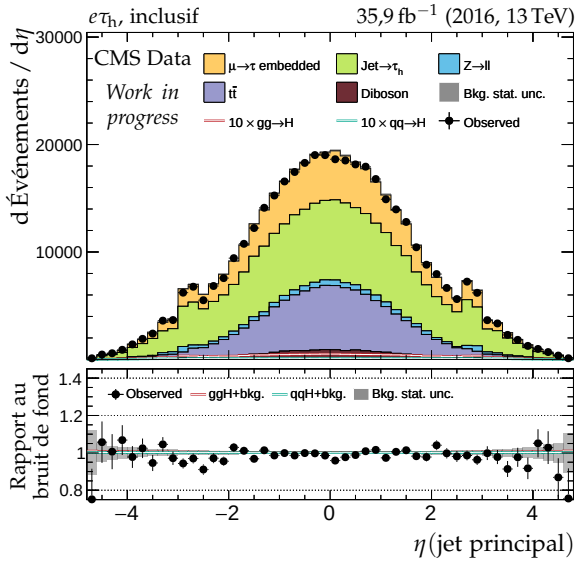
Figure G.17 – *Canal $e\tau_h$, 2016 : masses transverses, D_ζ et masses.*



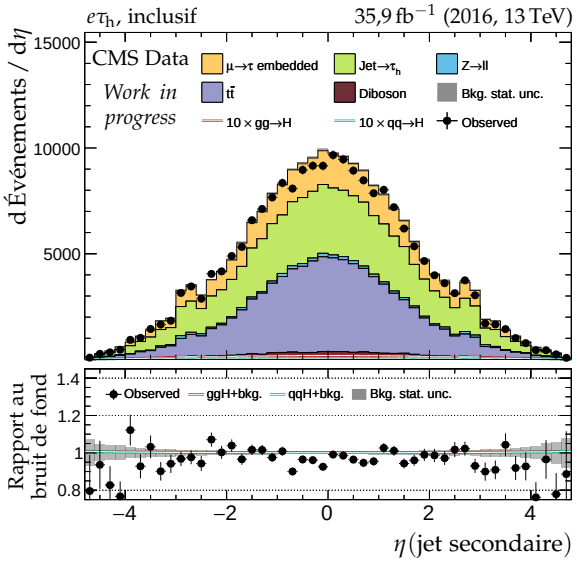
(a) *Impulsion transverse du jet principal.*



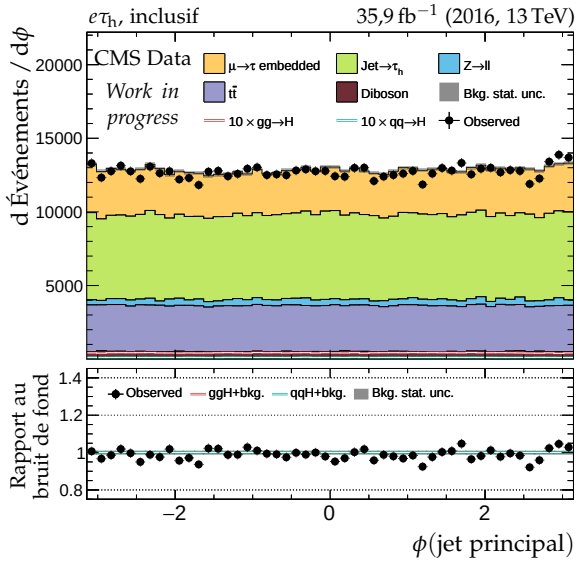
(b) *Impulsion transverse du jet secondaire.*



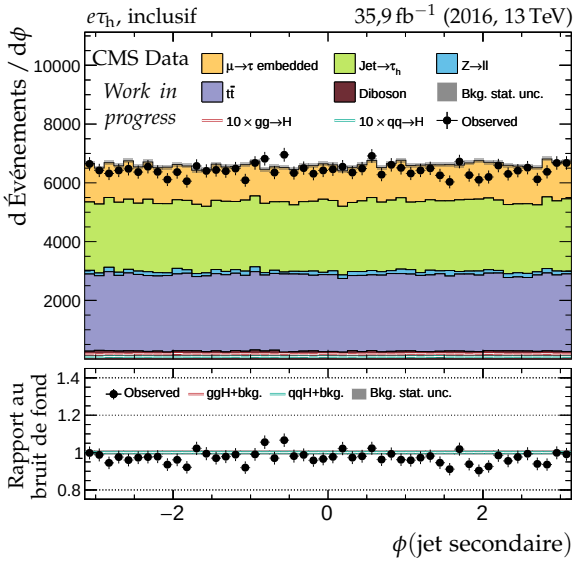
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

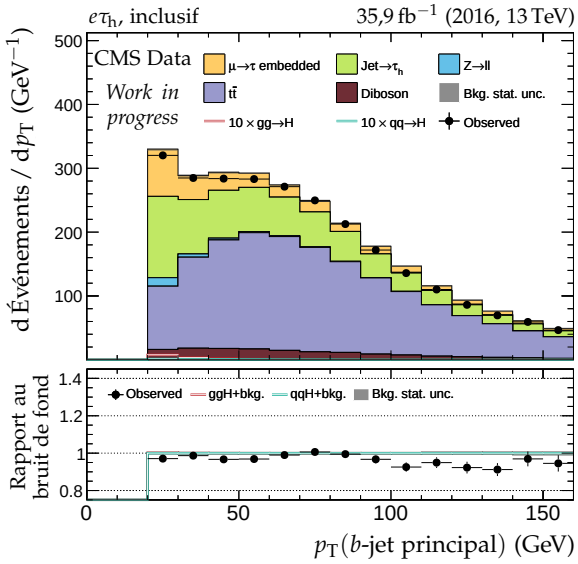


(e) *Angle azimutal du jet principal.*

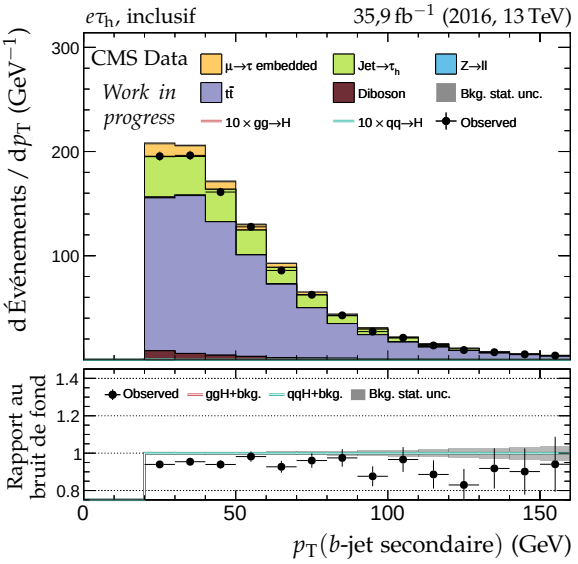


(f) *Angle azimutal du jet secondaire.*

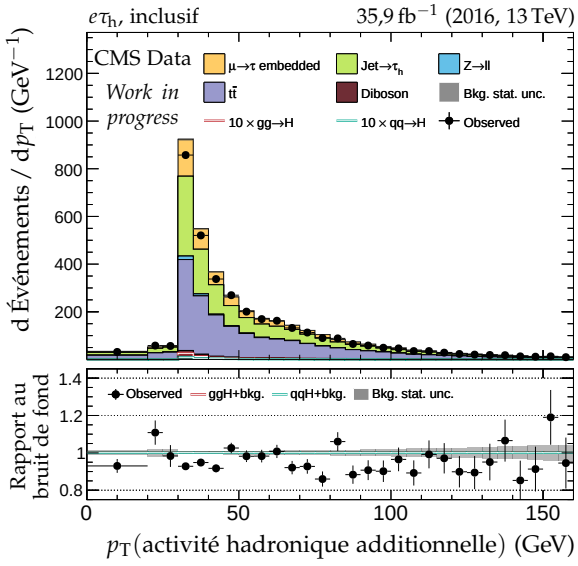
Figure G.18 – *Canal $e\tau_h$, 2016 : cinématique des deux jets principaux.*



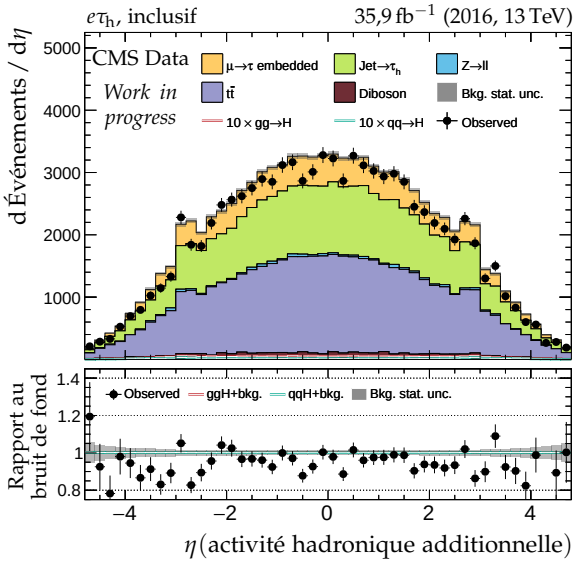
(a) *Impulsion transverse du b-jet principal.*



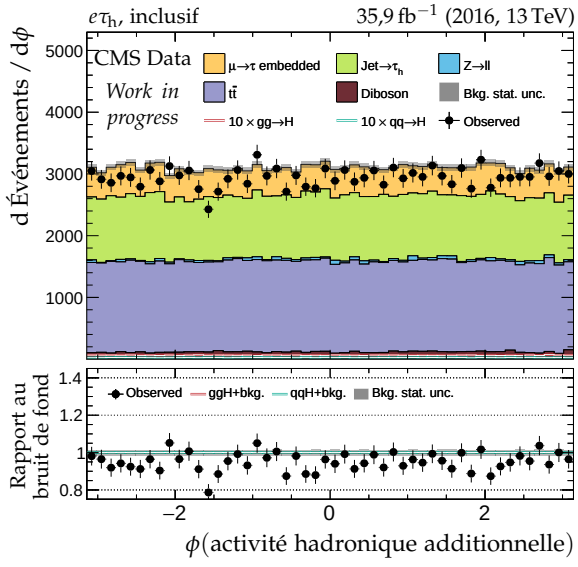
(b) *Impulsion transverse du b-jet secondaire.*



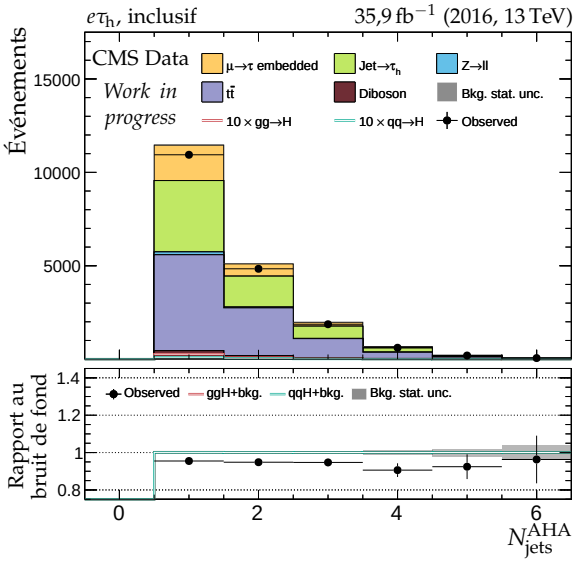
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

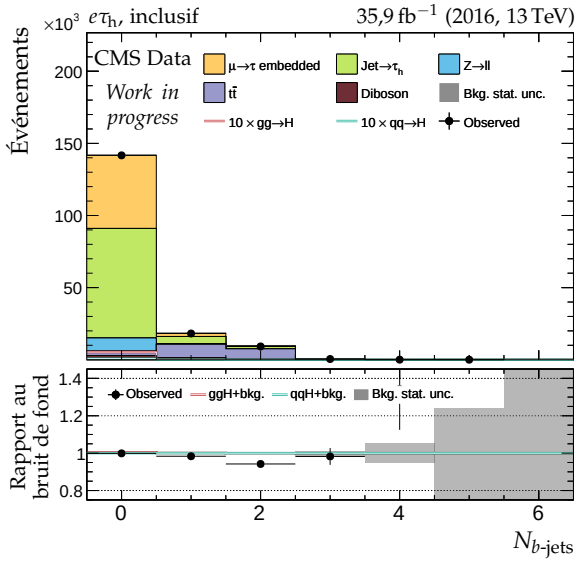


(e) *Angle azimutal de l'AHA.*

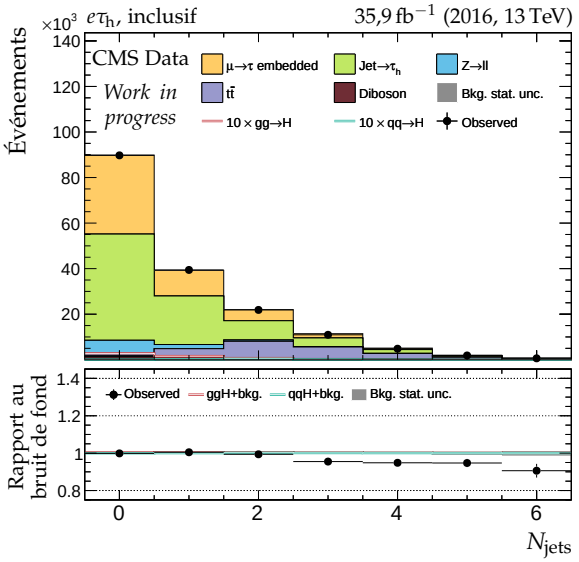


(f) *Nombre de jets dans l'AHA.*

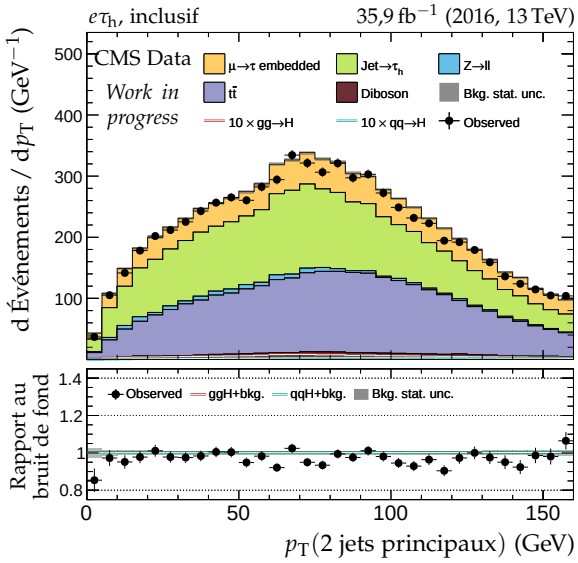
Figure G.19 – *Canal eτh, 2016 : b-jets et activité hadronique additionnelle.*



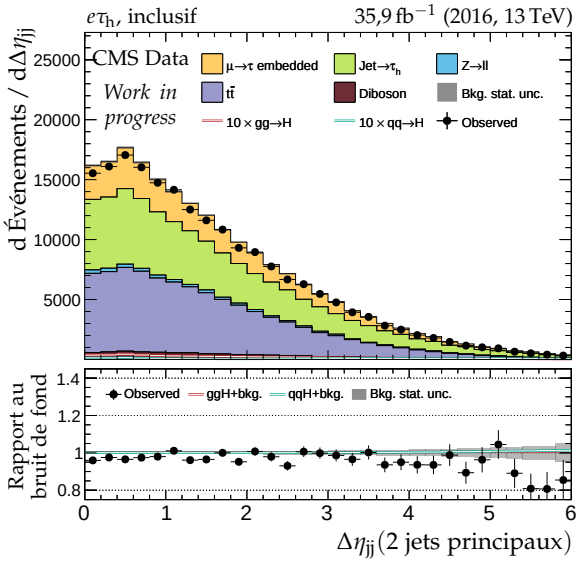
(a) *Nombre de b-jets.*



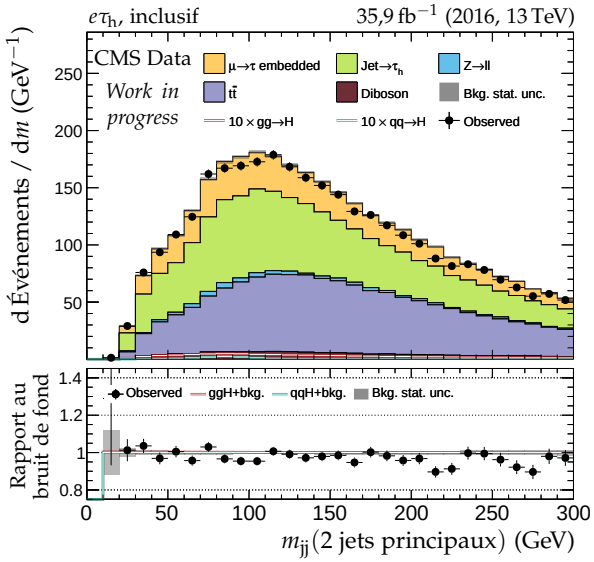
(b) *Nombre de jets.*



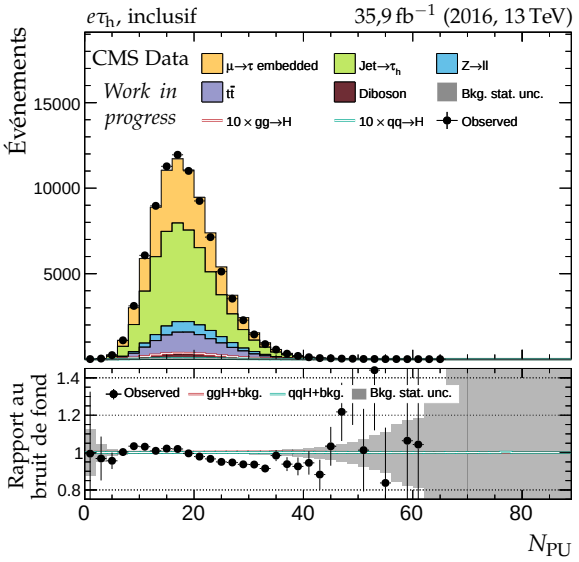
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

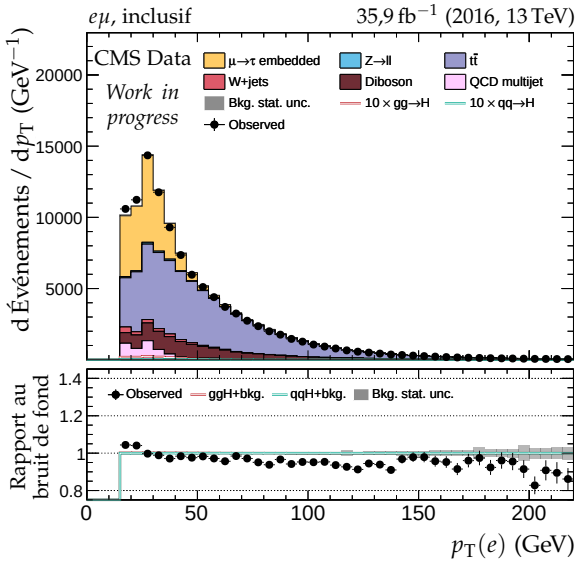


(e) *Masse invariante du système des deux jets.*

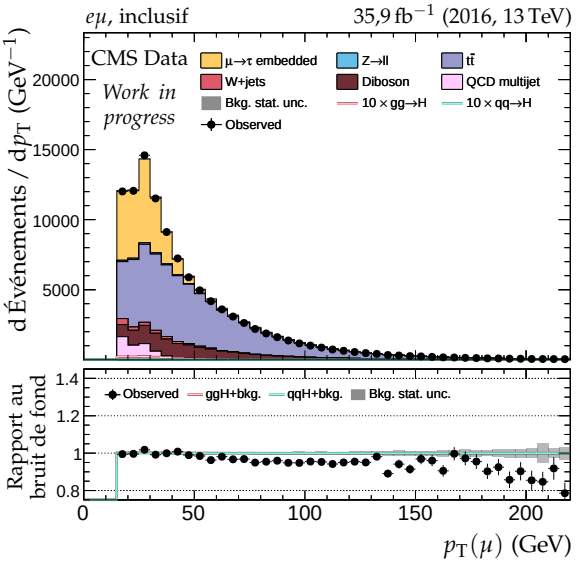


(f) *Nombre de vertex d'empilement.*

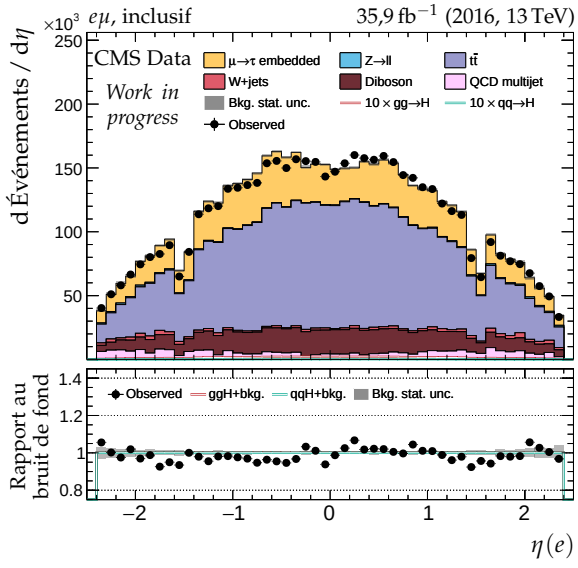
Figure G.20 – *Canal $e\tau_h$, 2016 : nombre de jets, système des deux jets principaux et empilement.*



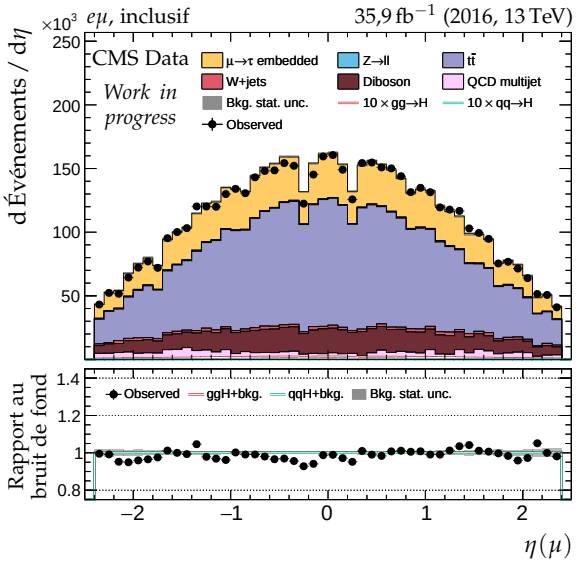
(a) *Impulsion transverse du lepton 1 (e).*



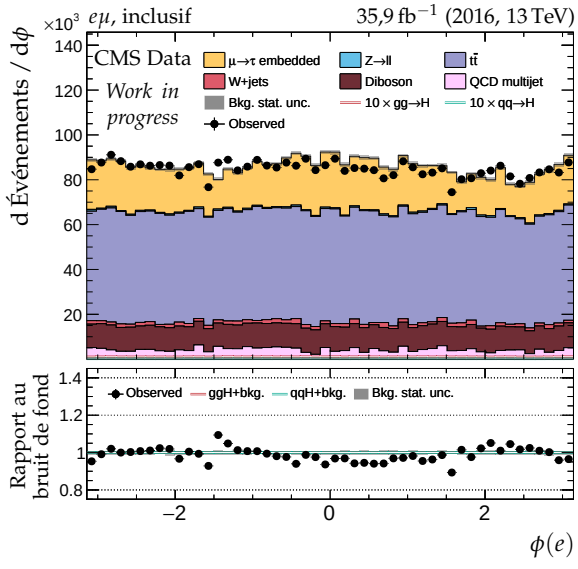
(b) *Impulsion transverse du lepton 2 (μ).*



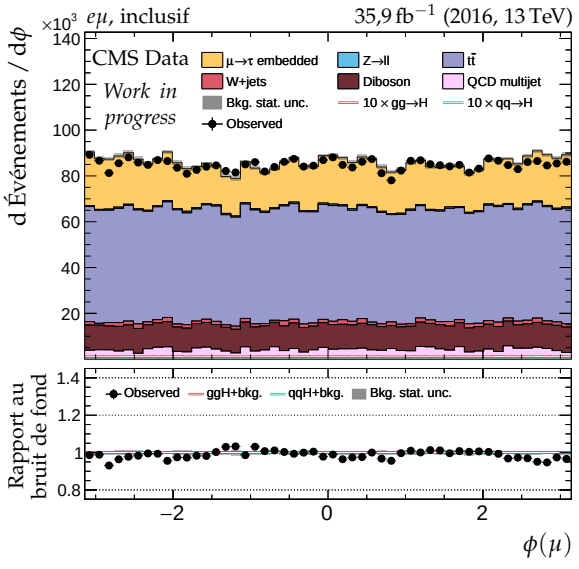
(c) *Pseudo-rapidité du lepton 1 (e).*



(d) *Pseudo-rapidité du lepton 2 (μ).*

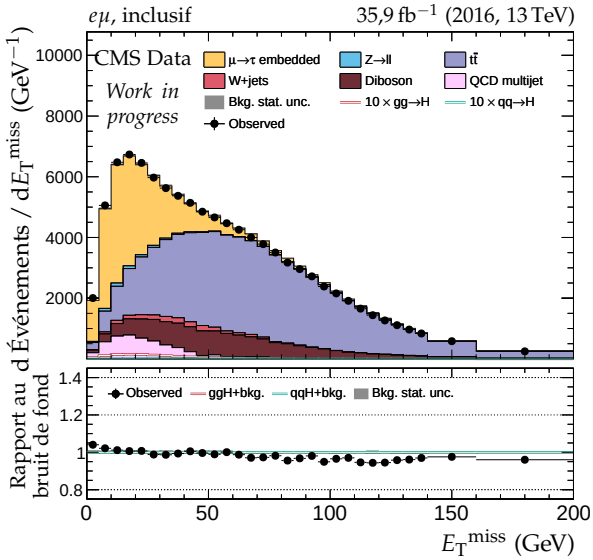


(e) *Angle azimutal du lepton 1 (e).*

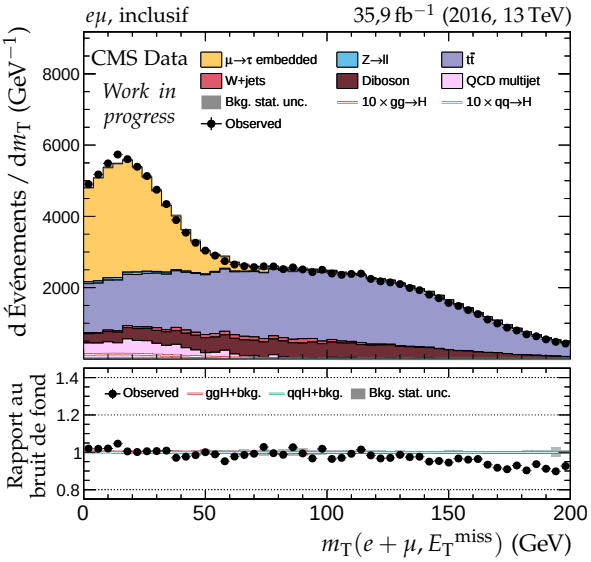


(f) *Angle azimutal du lepton 2 (μ).*

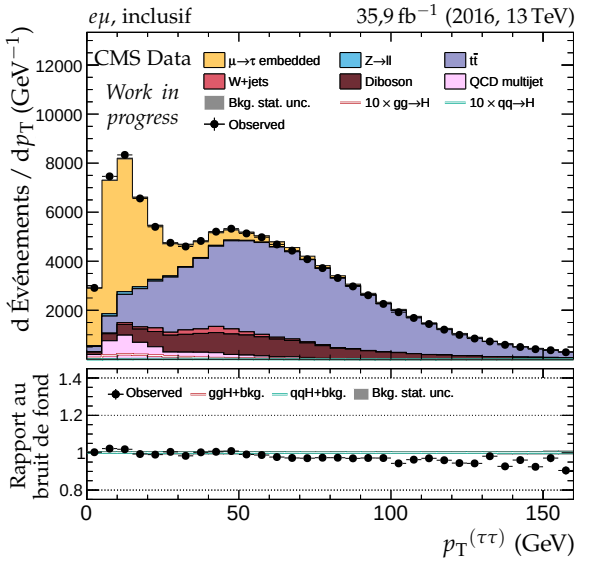
Figure G.21 – *Canal eμ, 2016 : cinématique des leptons (e, μ).*



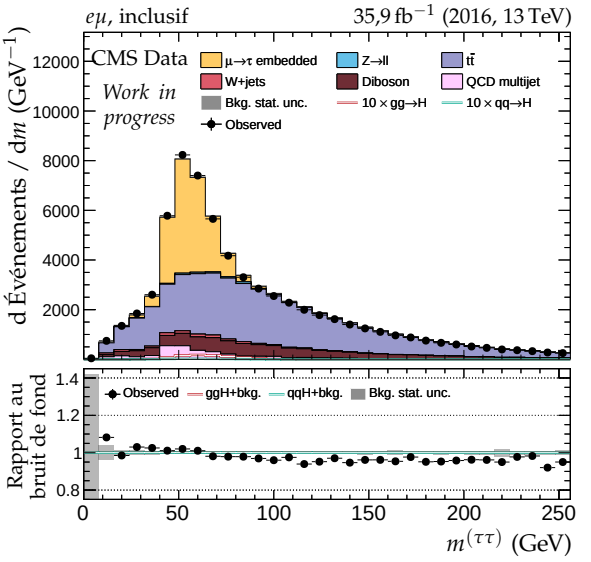
(a) *Énergie transverse manquante.*



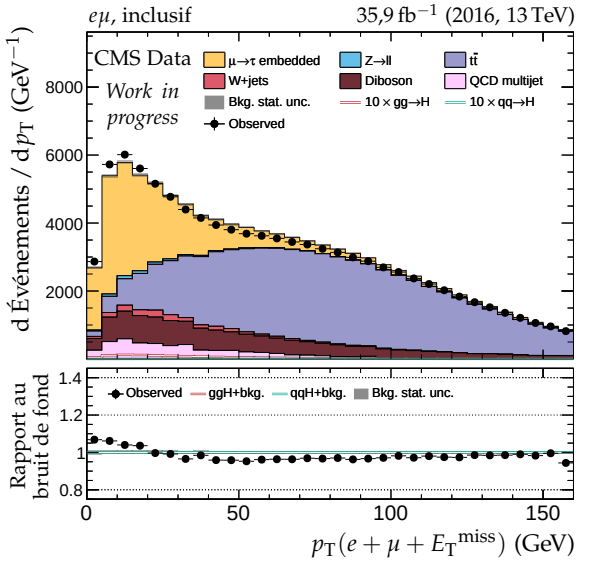
(b) *Masse transverse du dilepton.*



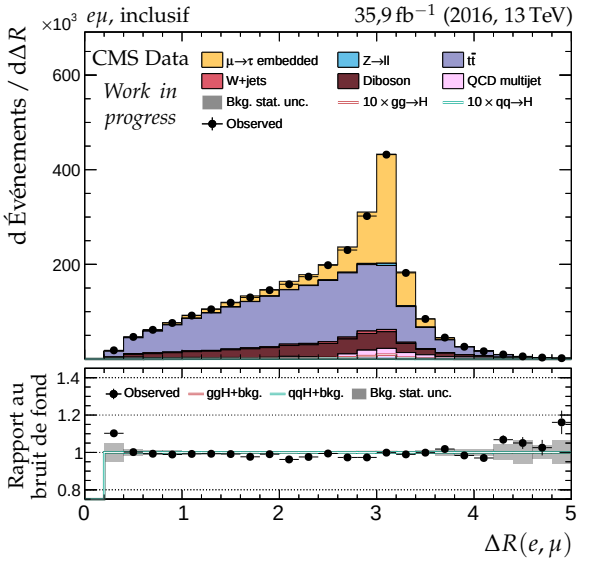
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

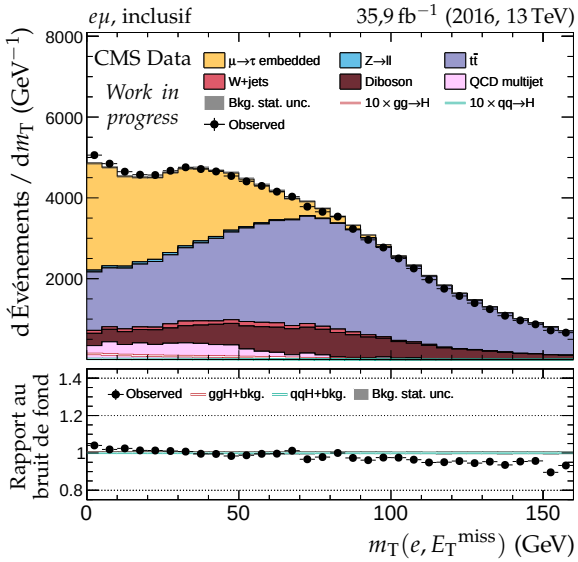


(e) *Impulsion transverse du système di-τ.*

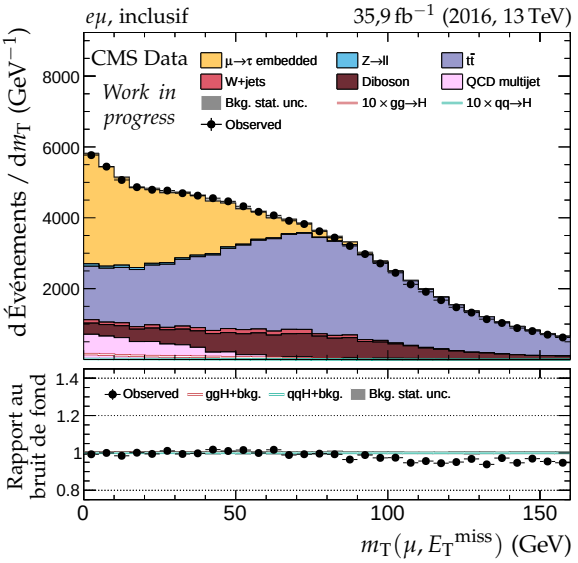


(f) *Distance ΔR entre les leptons (e, μ).*

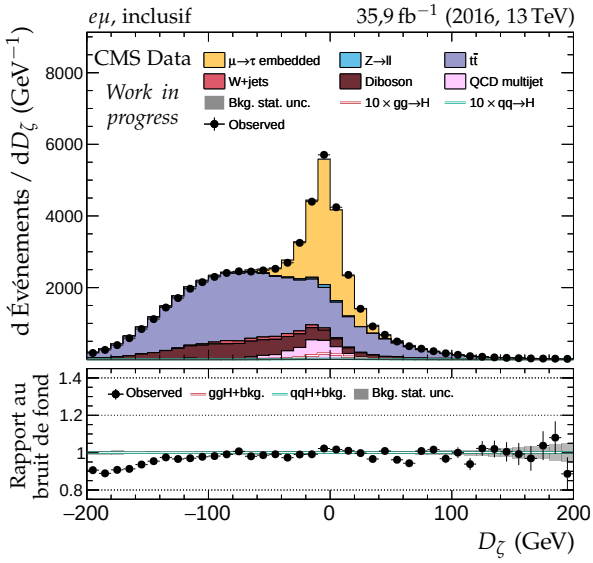
Figure G.22 – *Canal eμ, 2016 : dilepton et énergie transverse manquante.*



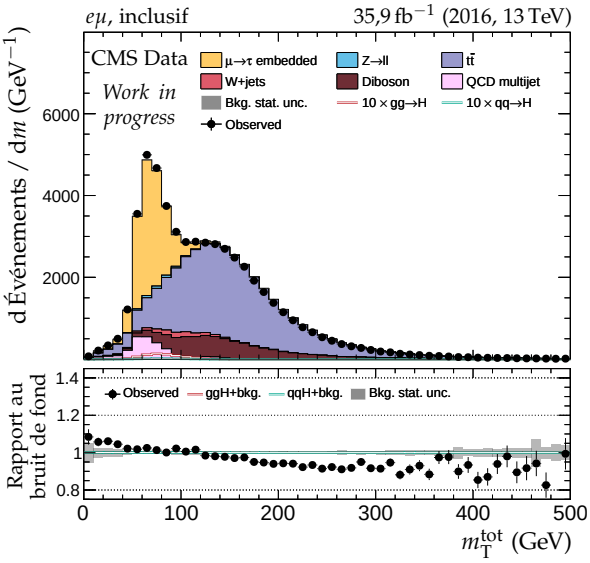
(a) *Masse transverse du lepton 1 (e).*



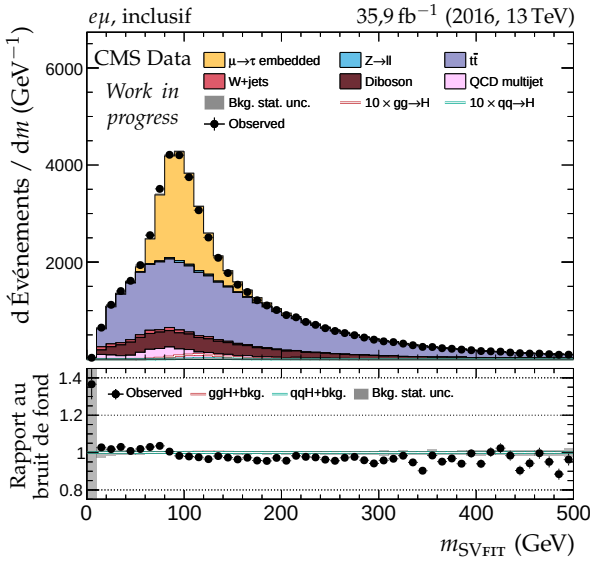
(b) *Masse transverse du lepton 2 (μ).*



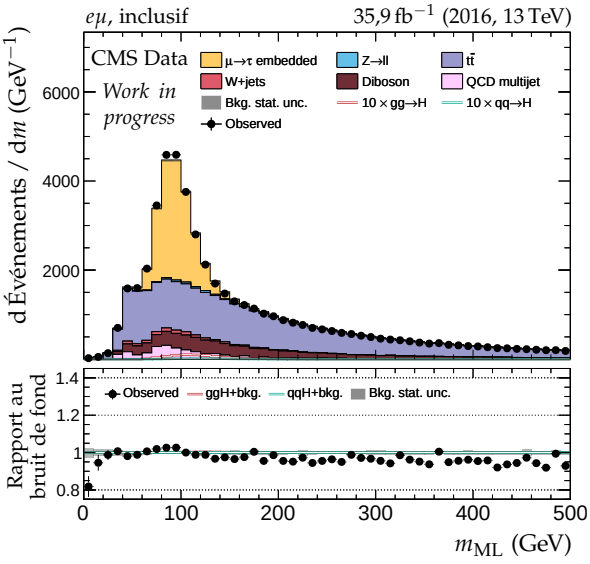
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

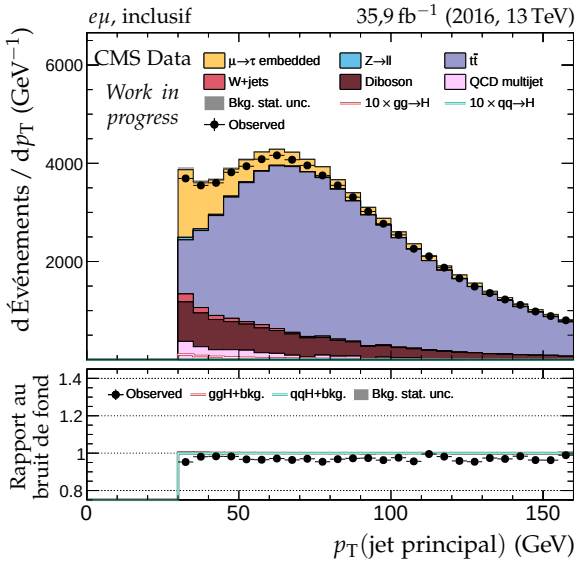


(e) *Masse du système di-τ d'après* SVFIT.

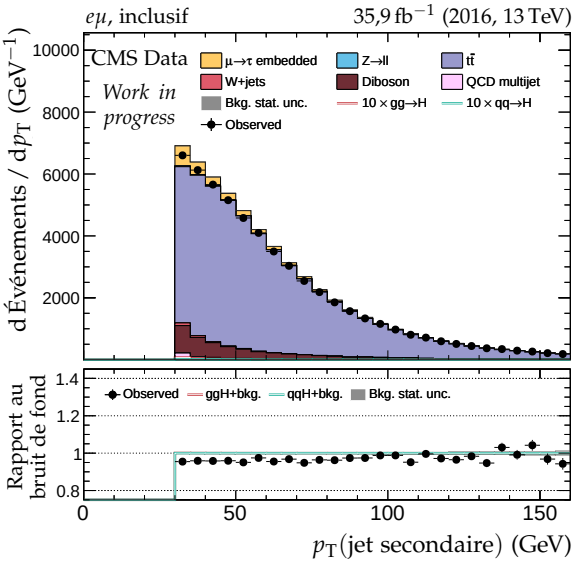


(f) *Masse du système di-τ d'après le ML.*

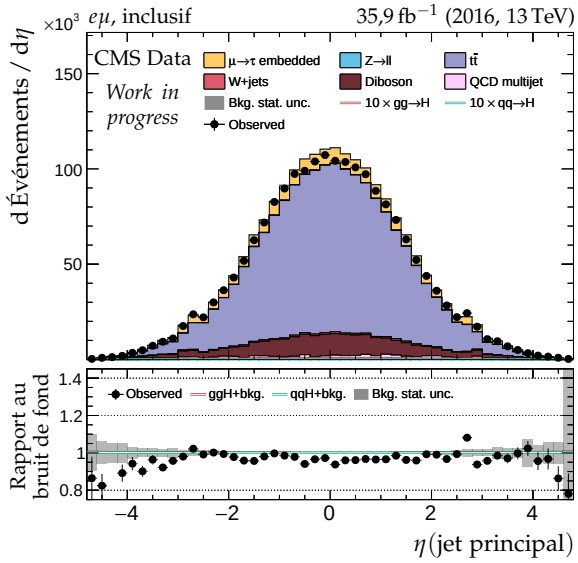
Figure G.23 – *Canal eμ, 2016 : masses transverses, D_ζ et masses.*



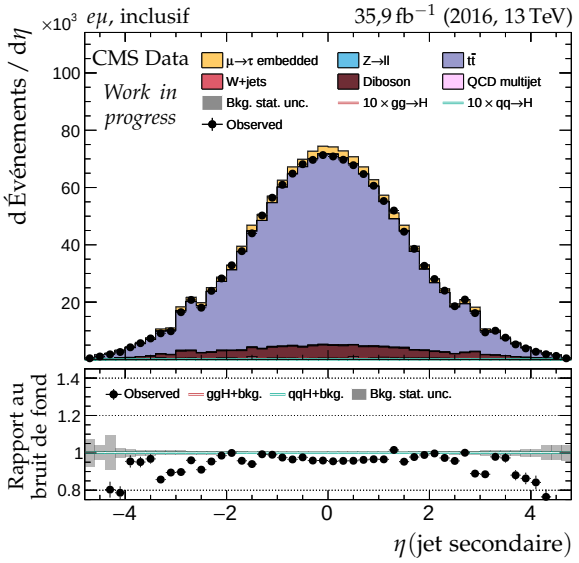
(a) *Impulsion transverse du jet principal.*



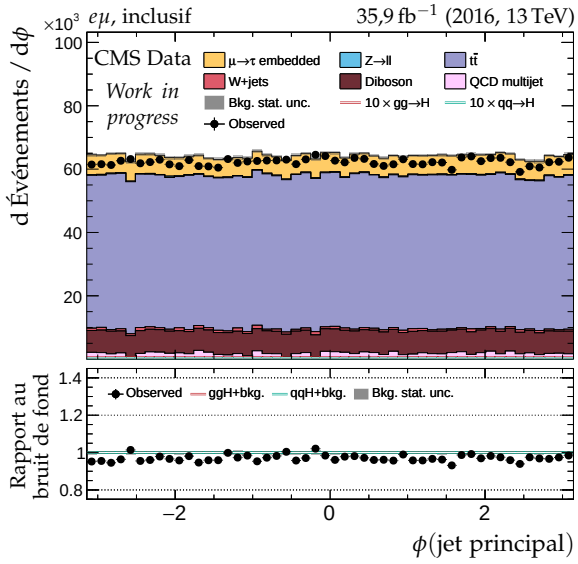
(b) *Impulsion transverse du jet secondaire.*



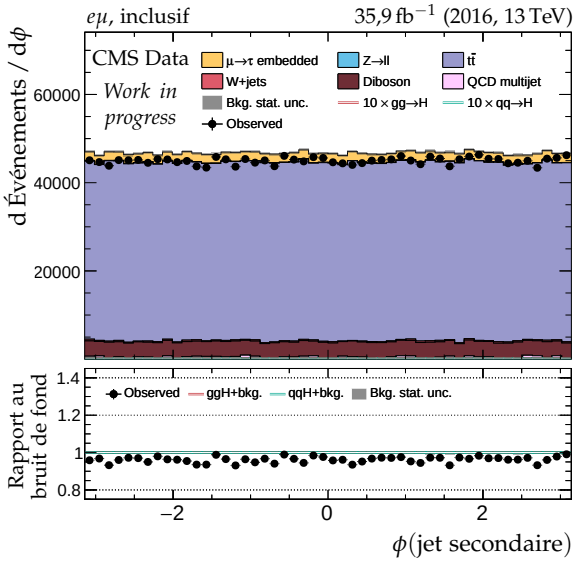
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

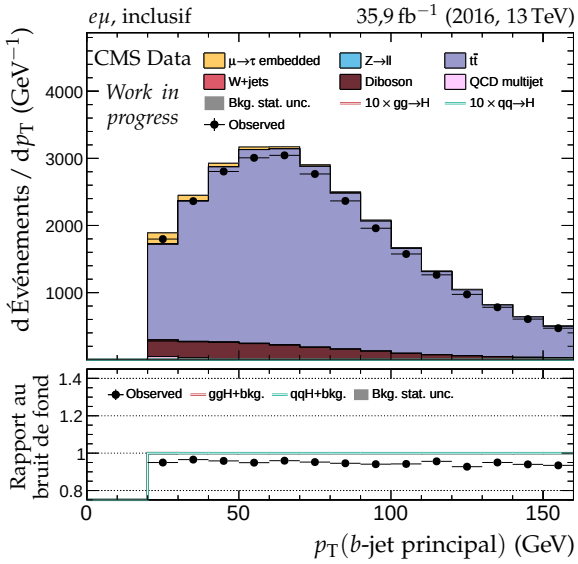


(e) *Angle azimutal du jet principal.*

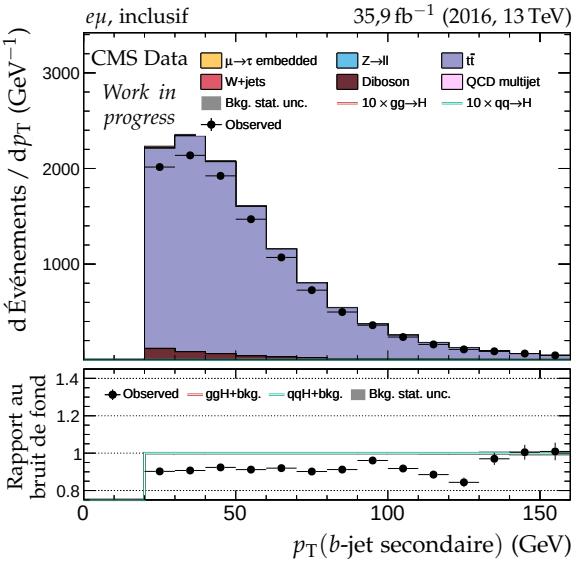


(f) *Angle azimutal du jet secondaire.*

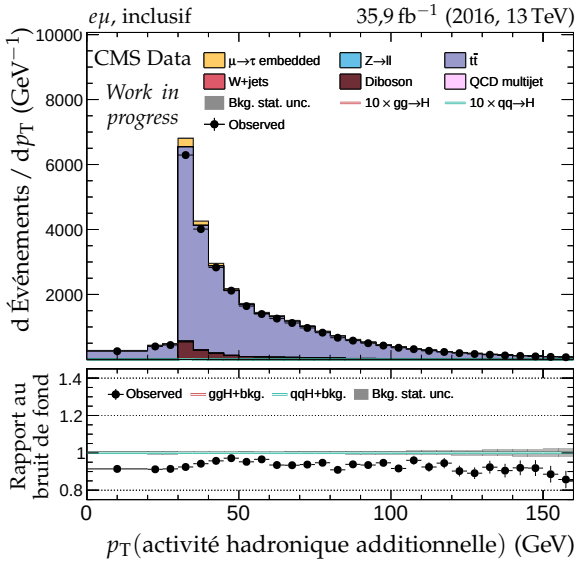
Figure G.24 – *Canal eμ, 2016 : cinématique des deux jets principaux.*



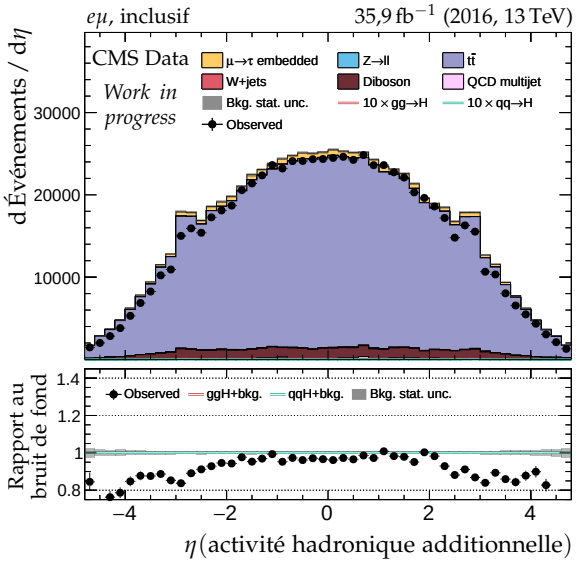
(a) *Impulsion transverse du b-jet principal.*



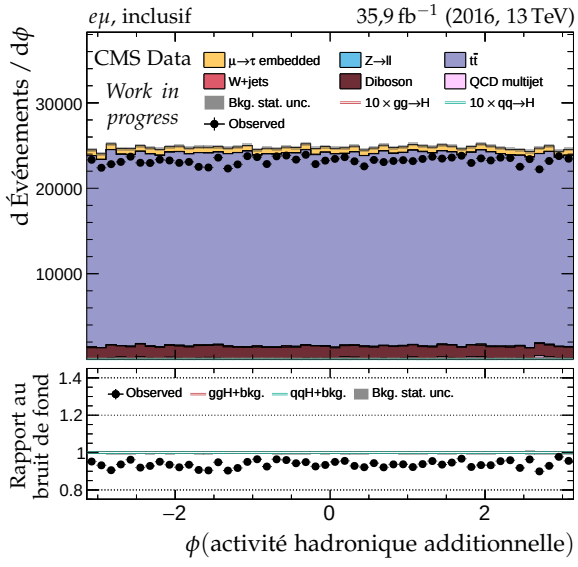
(b) *Impulsion transverse du b-jet secondaire.*



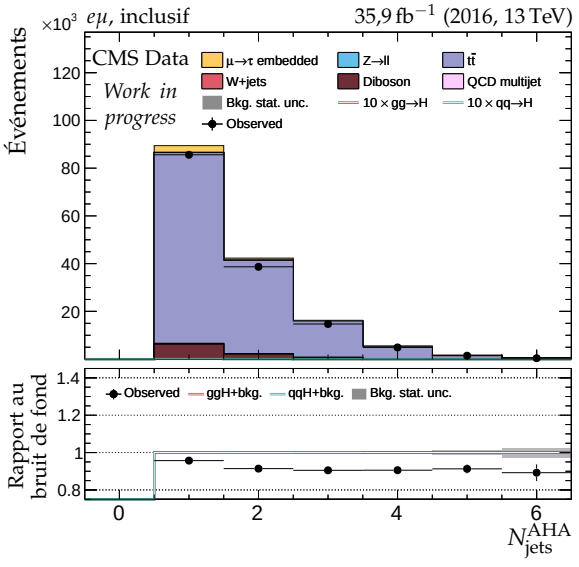
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

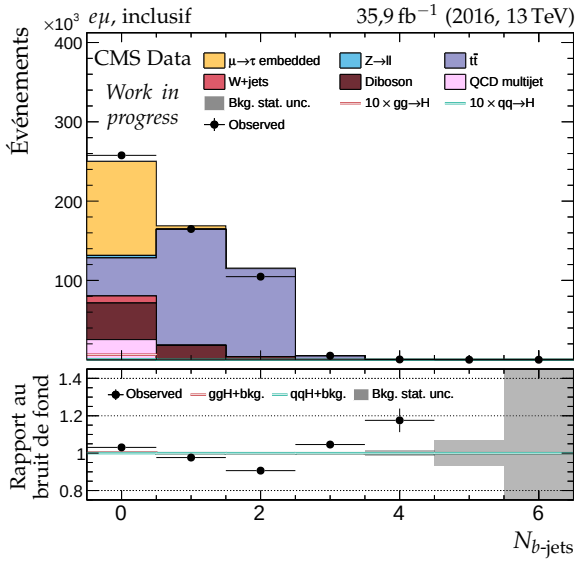


(e) *Angle azimutal de l'AHA.*

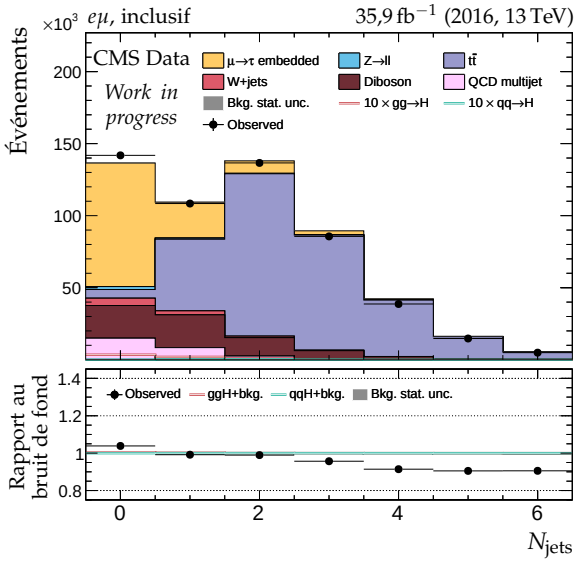


(f) *Nombre de jets dans l'AHA.*

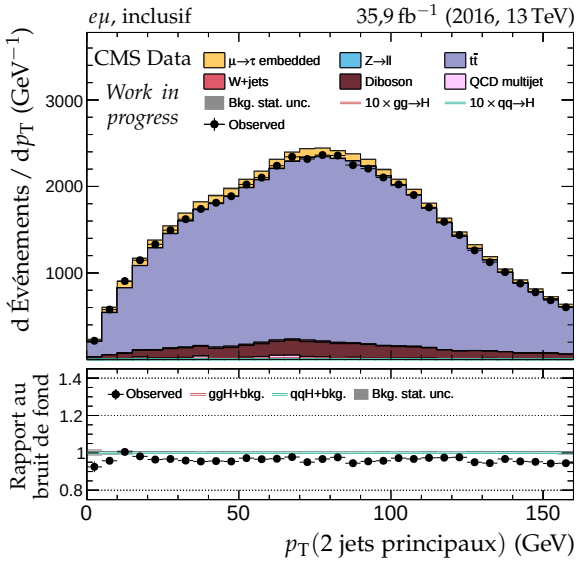
Figure G.25 – *Canal eμ, 2016 : b-jets et activité hadronique additionnelle.*



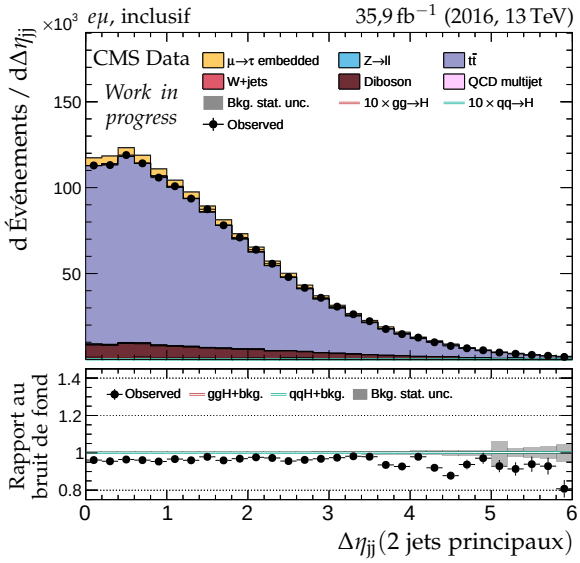
(a) *Nombre de b-jets.*



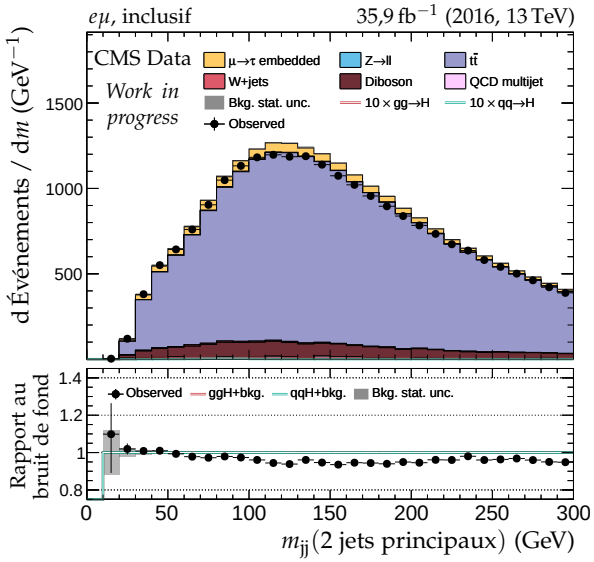
(b) *Nombre de jets.*



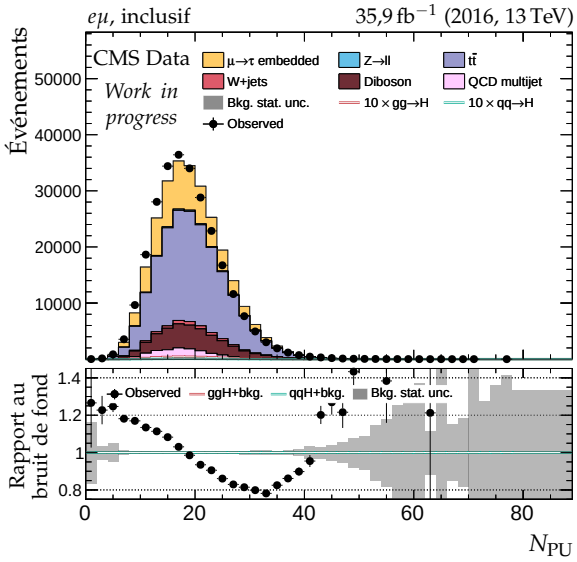
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

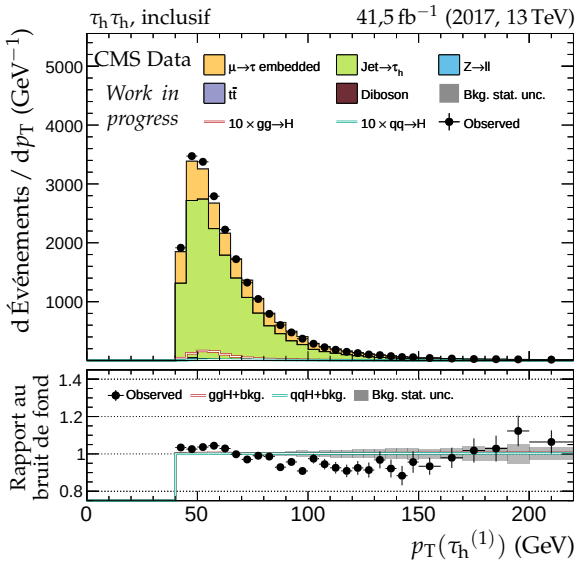


(e) *Masse invariante du système des deux jets.*

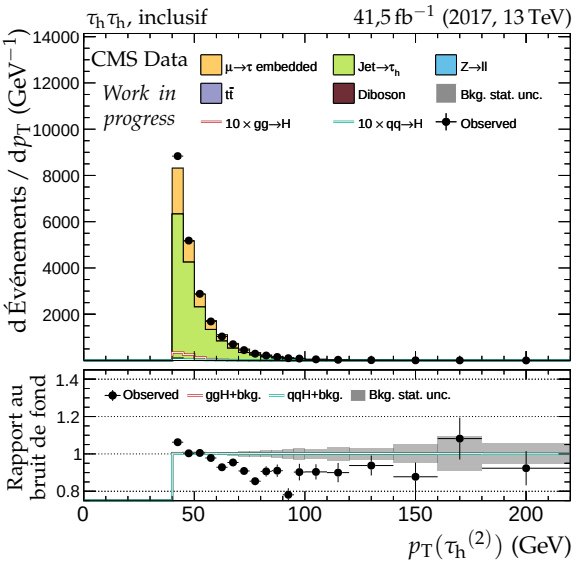


(f) *Nombre de vertex d'empilement.*

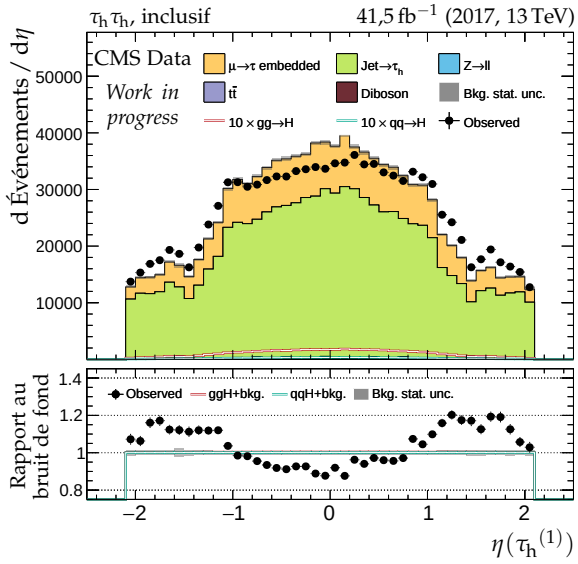
Figure G.26 – *Canal eμ, 2016 : nombre de jets, système des deux jets principaux et empilement.*



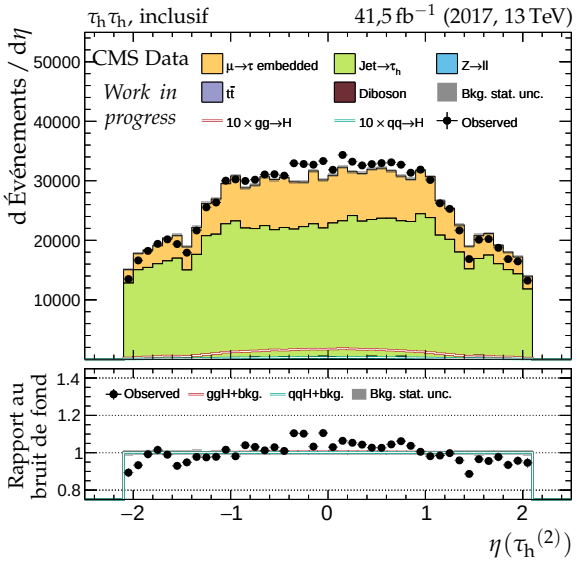
(a) *Impulsion transverse du lepton 1 ($\tau_h{}^{(1)}$).*



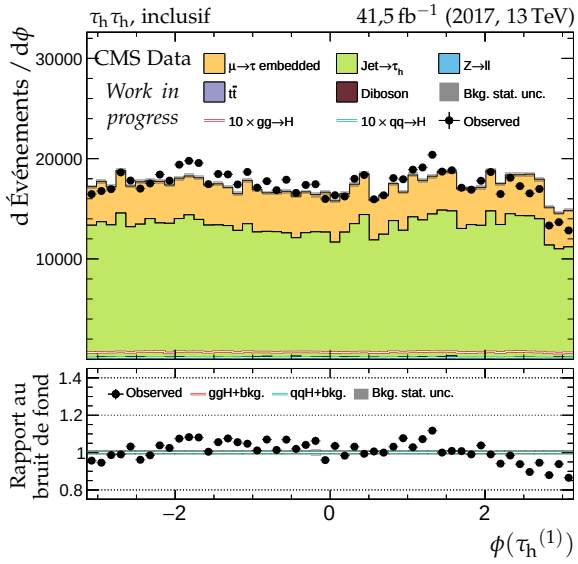
(b) *Impulsion transverse du lepton 2 ($\tau_h{}^{(2)}$).*



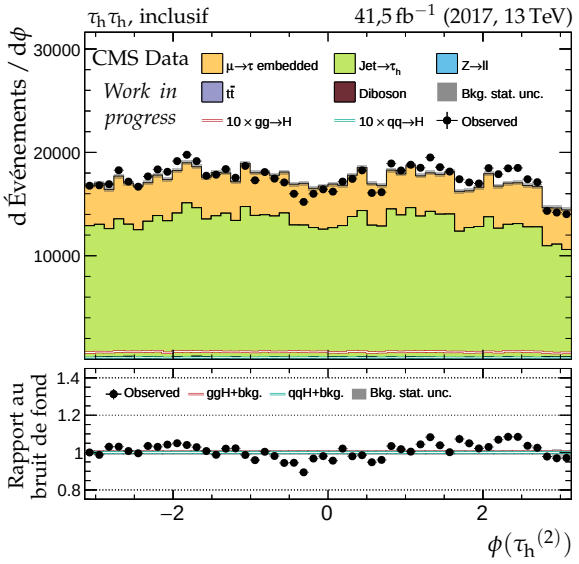
(c) *Pseudo-rapidité du lepton 1 ($\tau_h{}^{(1)}$).*



(d) *Pseudo-rapidité du lepton 2 ($\tau_h{}^{(2)}$).*

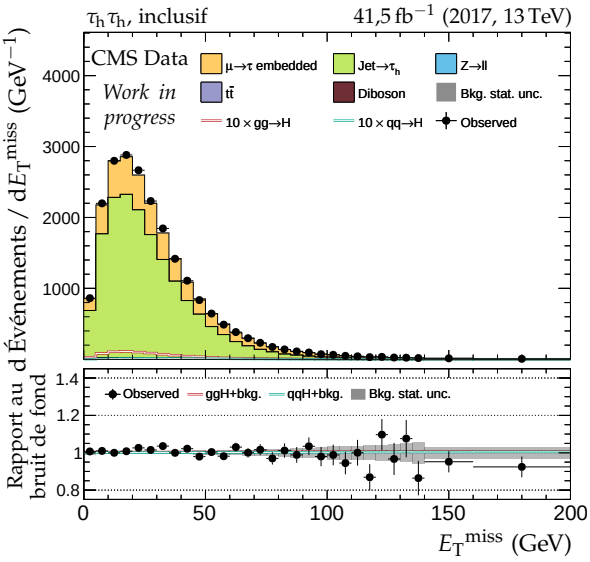


(e) *Angle azimutal du lepton 1 ($\tau_h{}^{(1)}$).*

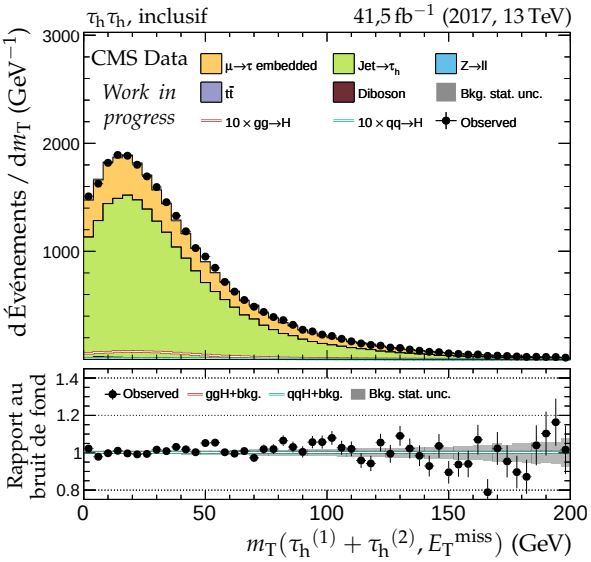


(f) *Angle azimutal du lepton 2 ($\tau_h{}^{(2)}$).*

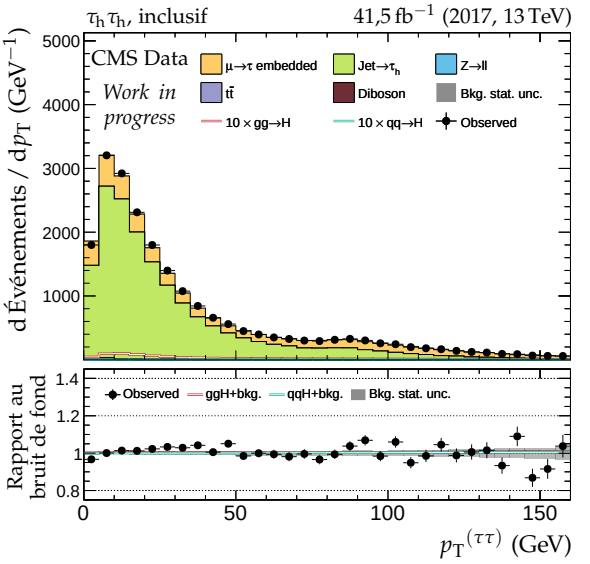
Figure G.27 – *Canal $\tau_h\tau_h$, 2017 : cinématique des leptons ($\tau_h{}^{(1)}$, $\tau_h{}^{(2)}$).*



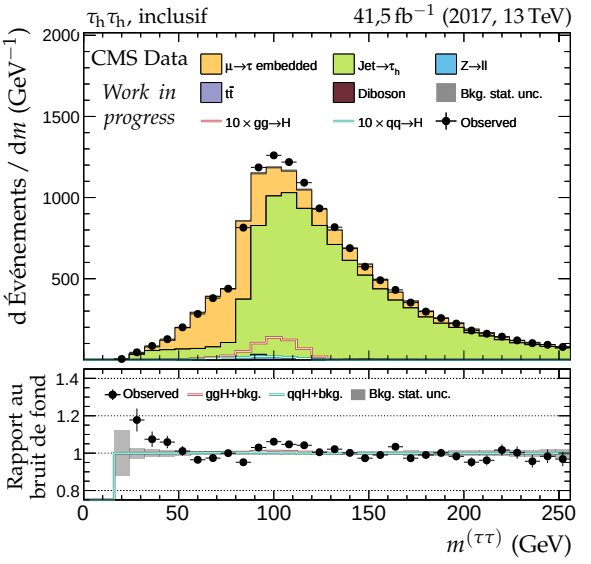
(a) *Énergie transverse manquante.*



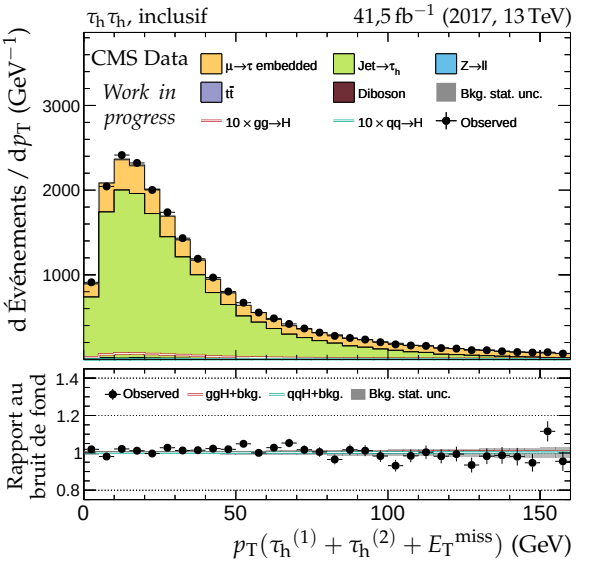
(b) *Masse transverse du dilepton.*



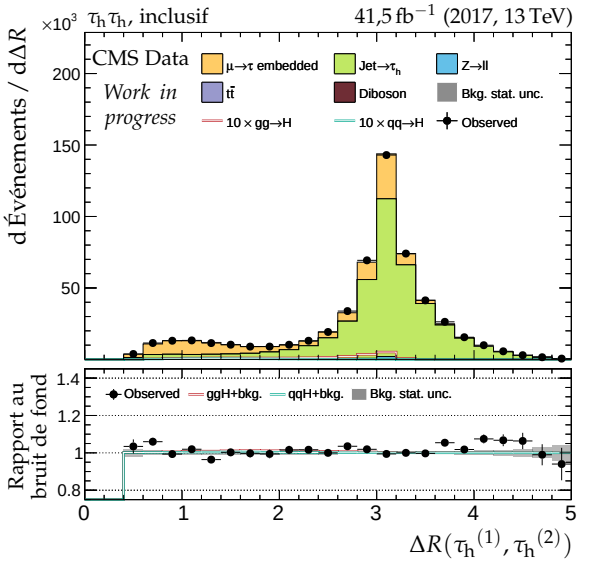
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

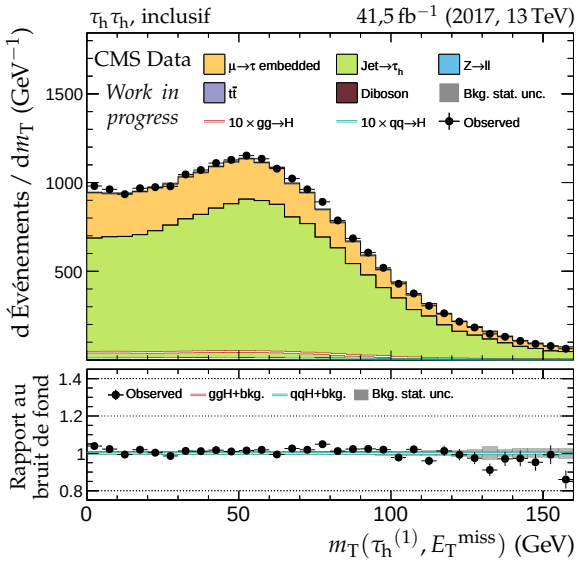


(e) *Impulsion transverse du système di-τ.*

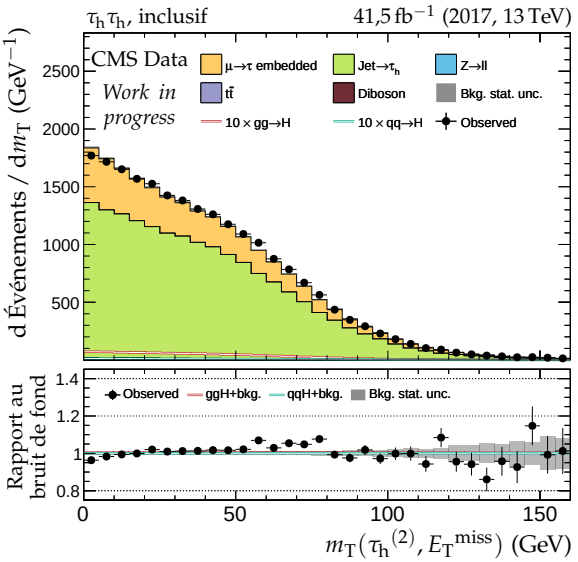


(f) *Distance ΔR entre les leptons ($\tau_h^{(1)}$, $\tau_h^{(2)}$).*

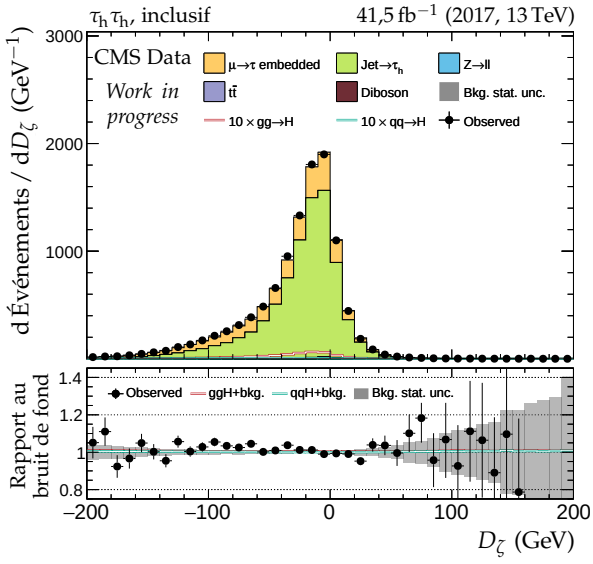
Figure G.28 – *Canal $\tau_h\tau_h$, 2017 : dilepton et énergie transverse manquante.*



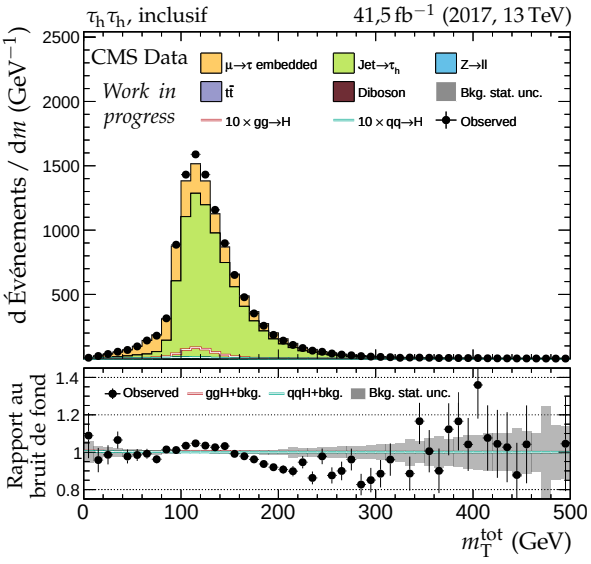
(a) *Masse transverse du lepton 1 ($\tau_h^{(1)}$).*



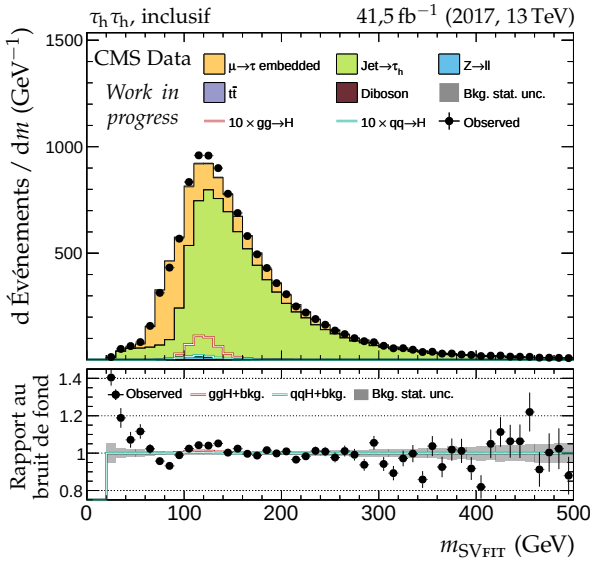
(b) *Masse transverse du lepton 2 ($\tau_h^{(2)}$).*



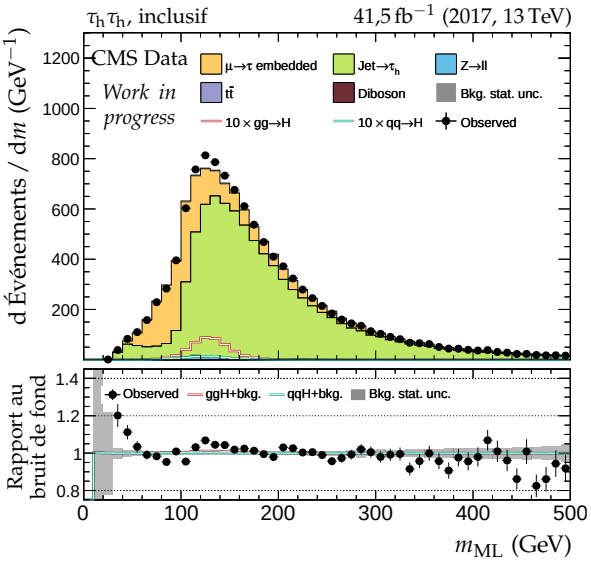
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

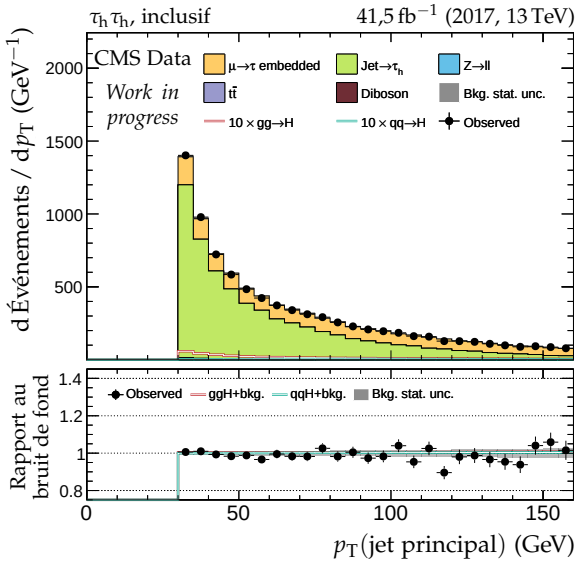


(e) *Masse du système di-τ d'après* SVFIT.

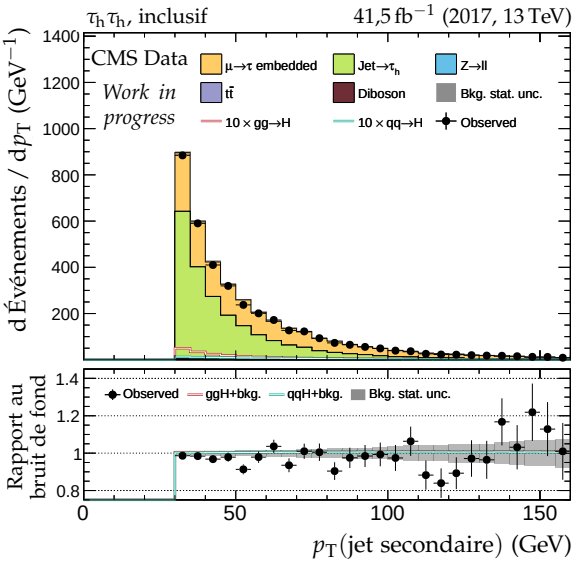


(f) *Masse du système di-τ d'après le ML.*

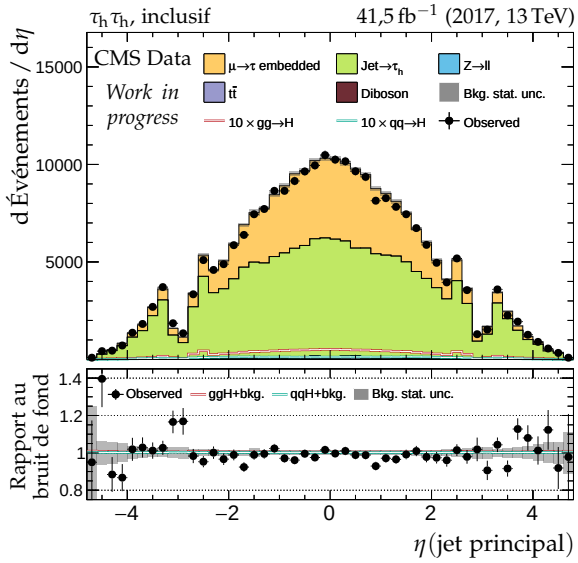
Figure G.29 – *Canal $\tau_h\tau_h$, 2017 : masses transverses, D_ζ et masses.*



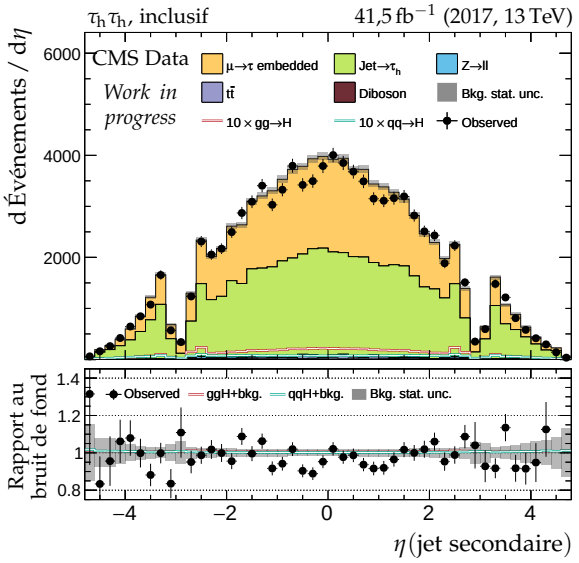
(a) *Impulsion transverse du jet principal.*



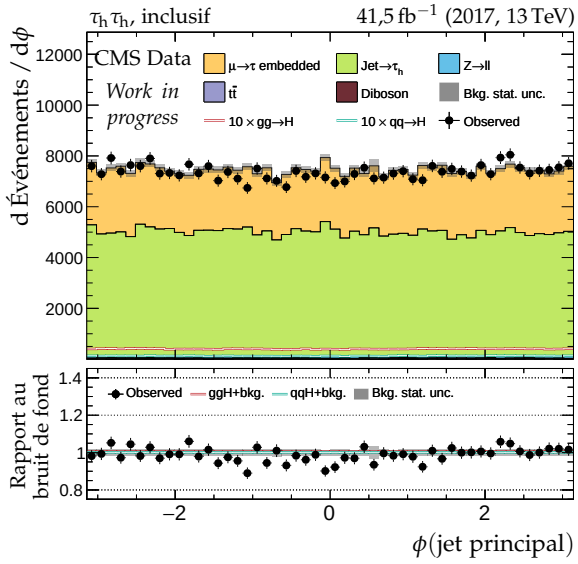
(b) *Impulsion transverse du jet secondaire.*



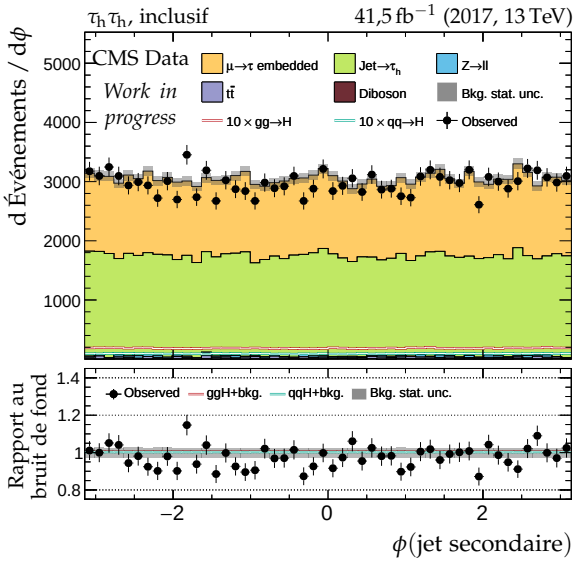
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

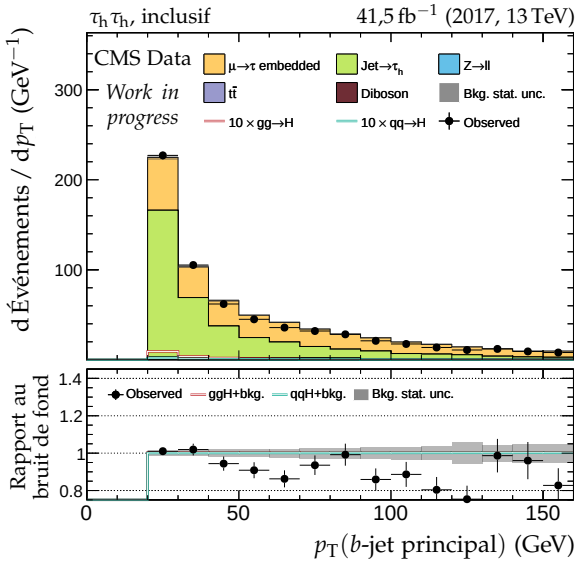


(e) *Angle azimutal du jet principal.*

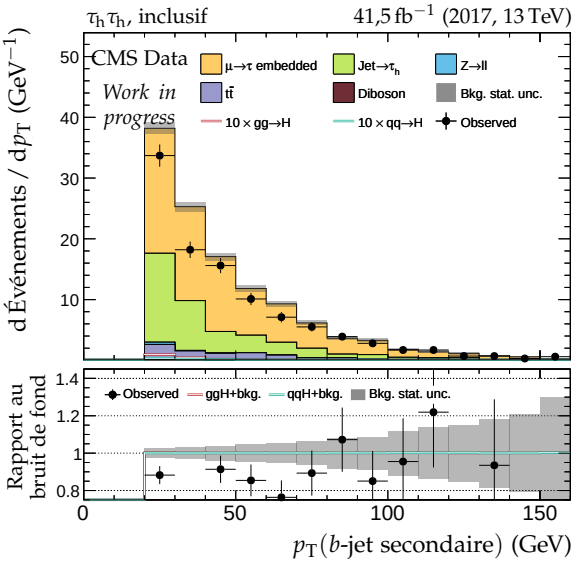


(f) *Angle azimutal du jet secondaire.*

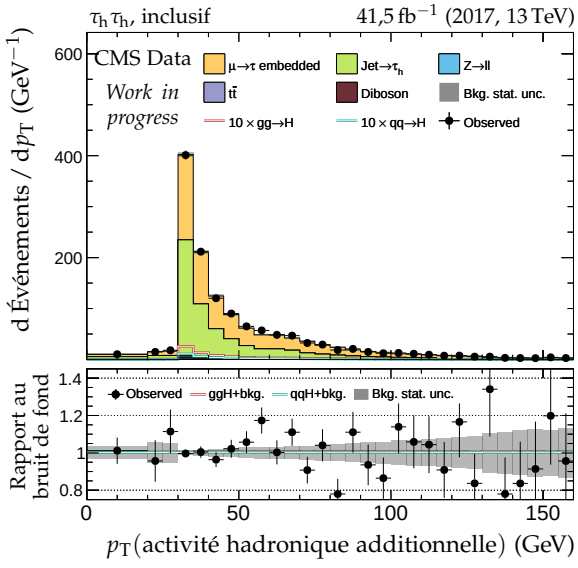
Figure G.30 – *Canal $\tau_h\tau_h$, 2017 : cinématique des deux jets principaux.*



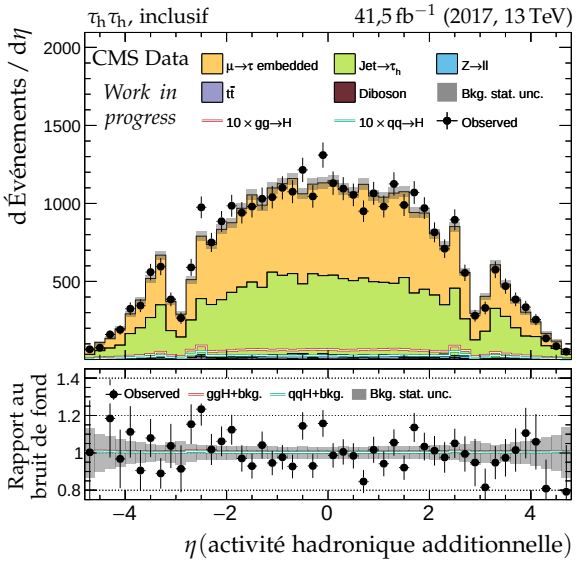
(a) *Impulsion transverse du b-jet principal.*



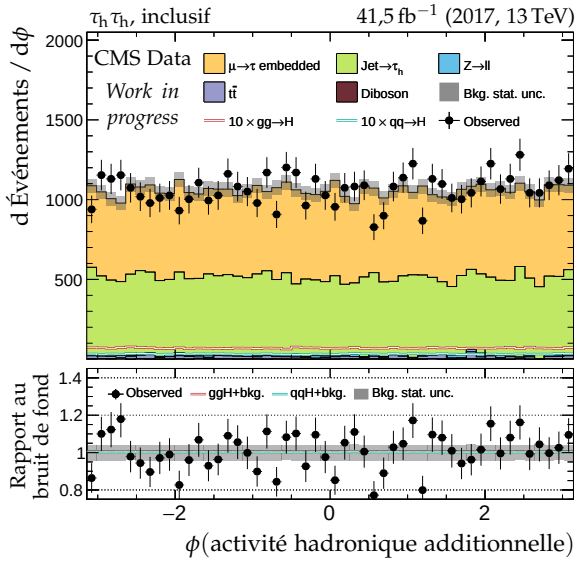
(b) *Impulsion transverse du b-jet secondaire.*



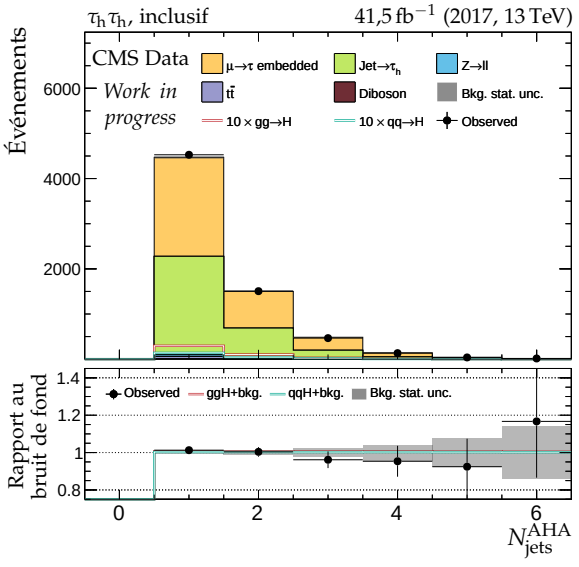
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

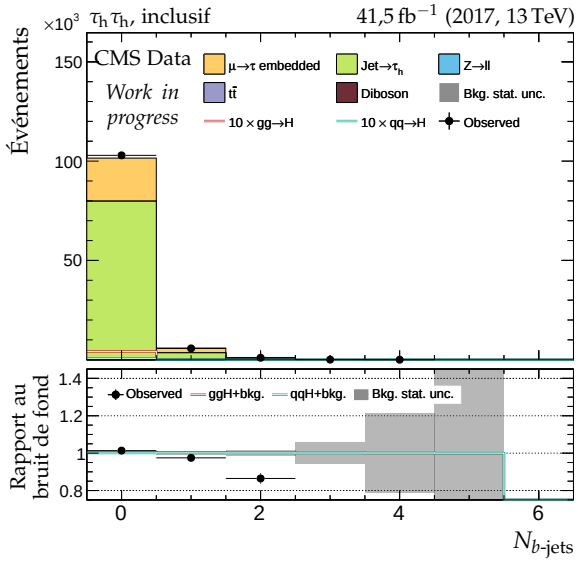


(e) *Angle azimutal de l'AHA.*

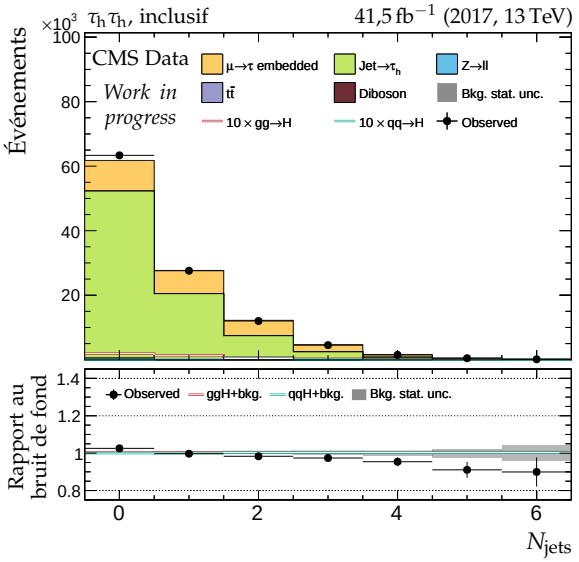


(f) *Nombre de jets dans l'AHA.*

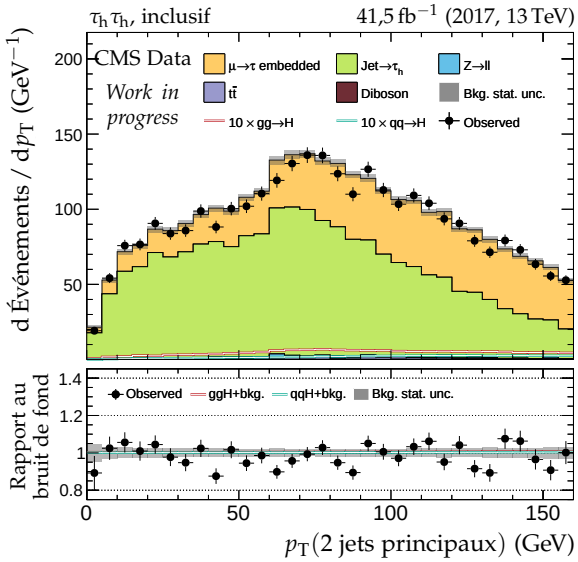
Figure G.31 – *Canal $\tau_h\tau_h$, 2017 : b-jets et activité hadronique additionnelle.*



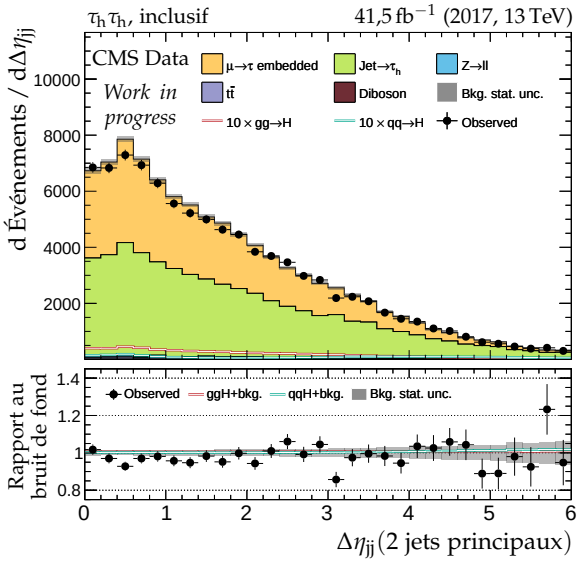
(a) *Nombre de b-jets.*



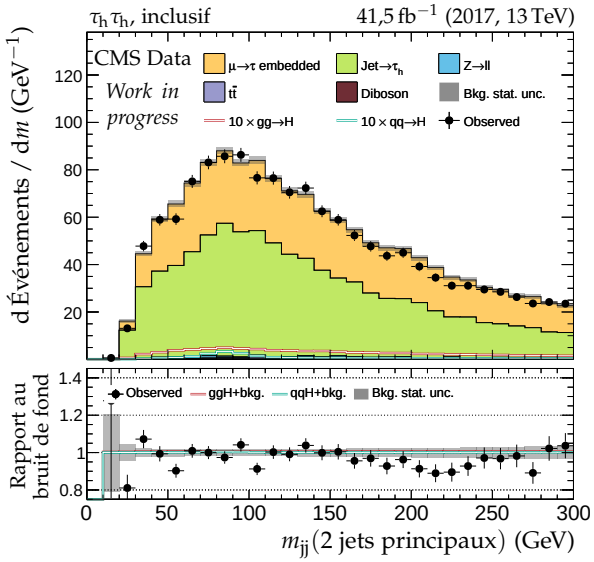
(b) *Nombre de jets.*



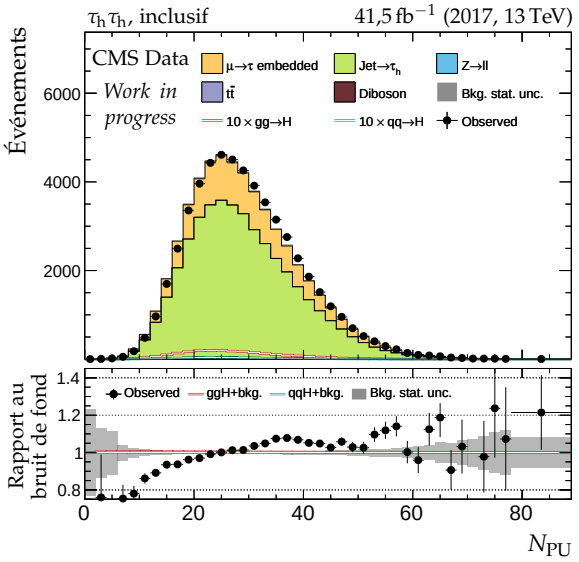
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

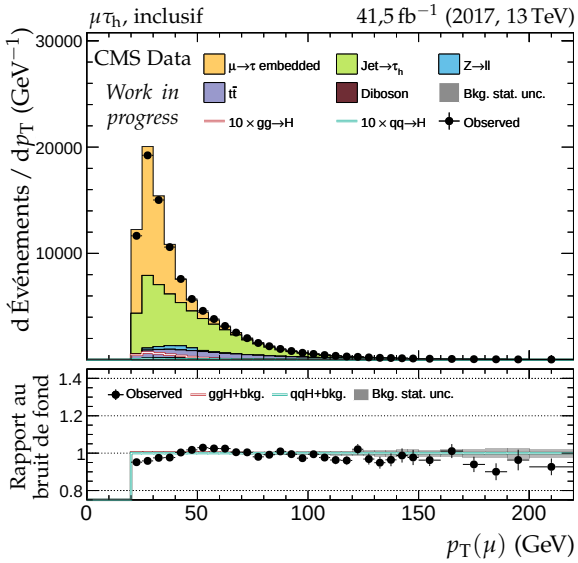


(e) *Masse invariante du système des deux jets.*

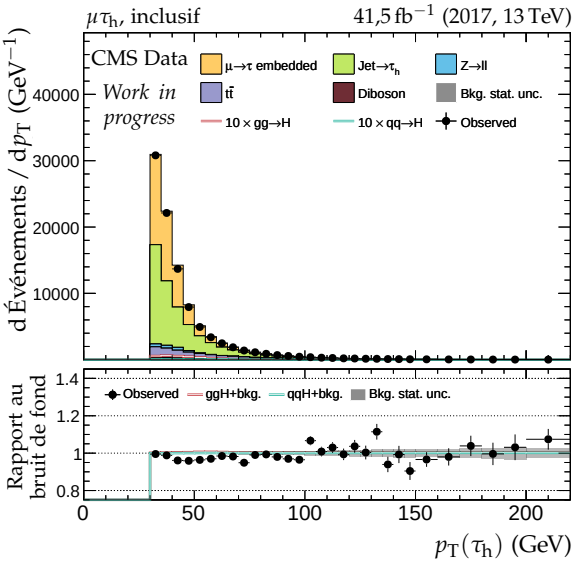


(f) *Nombre de vertex d'empilement.*

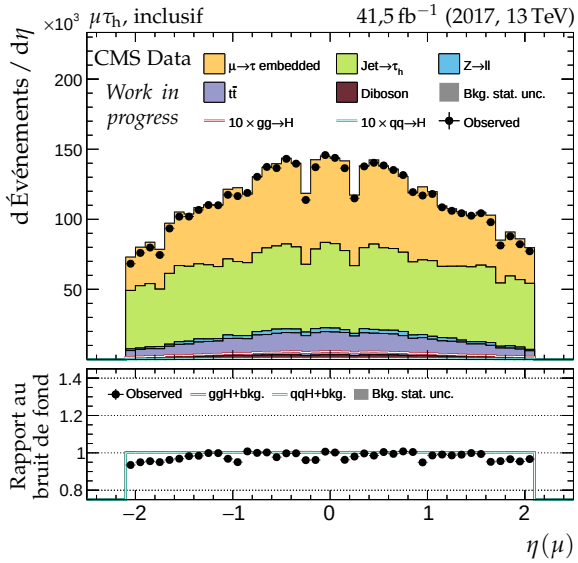
Figure G.32 – *Canal $\tau_h\tau_h$, 2017 : nombre de jets, système des deux jets principaux et empilement.*



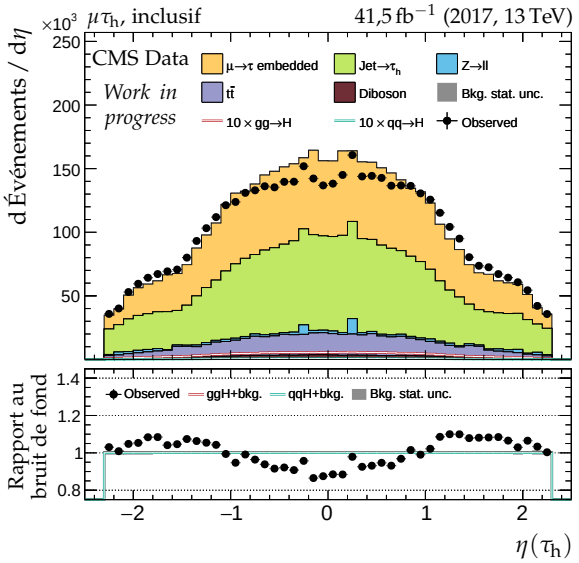
(a) *Impulsion transverse du lepton 1 (μ).*



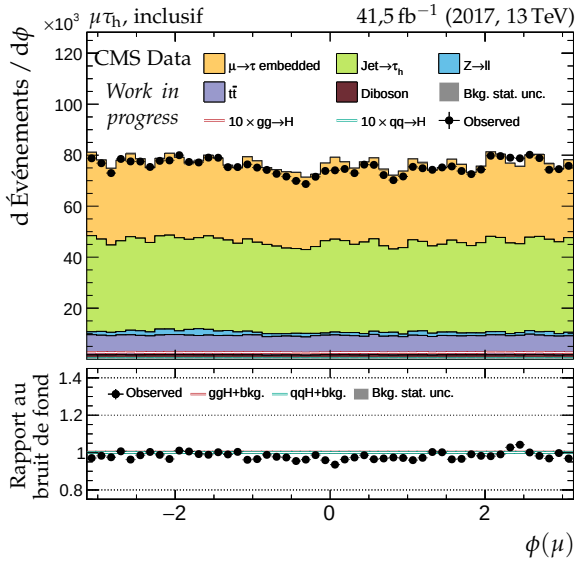
(b) *Impulsion transverse du lepton 2 (τ_h).*



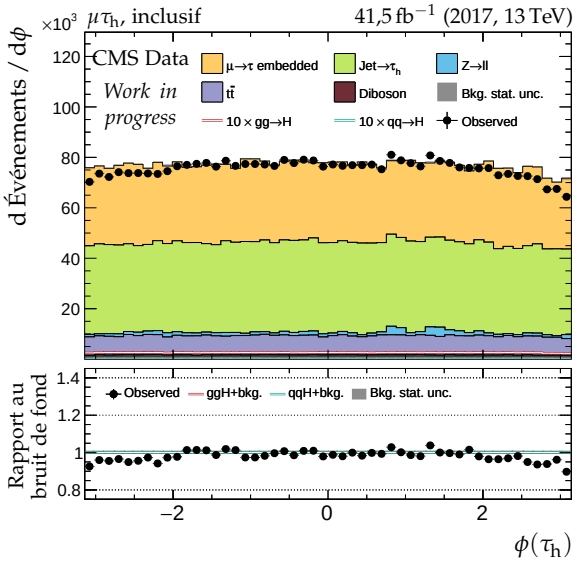
(c) *Pseudo-rapidité du lepton 1 (μ).*



(d) *Pseudo-rapidité du lepton 2 (τ_h).*

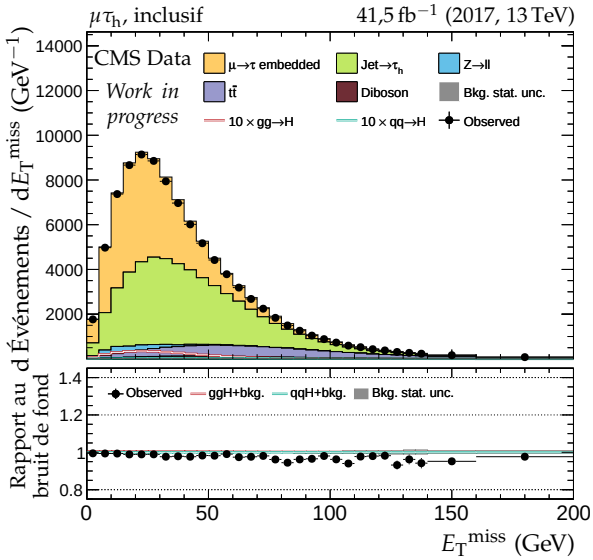


(e) *Angle azimutal du lepton 1 (μ).*

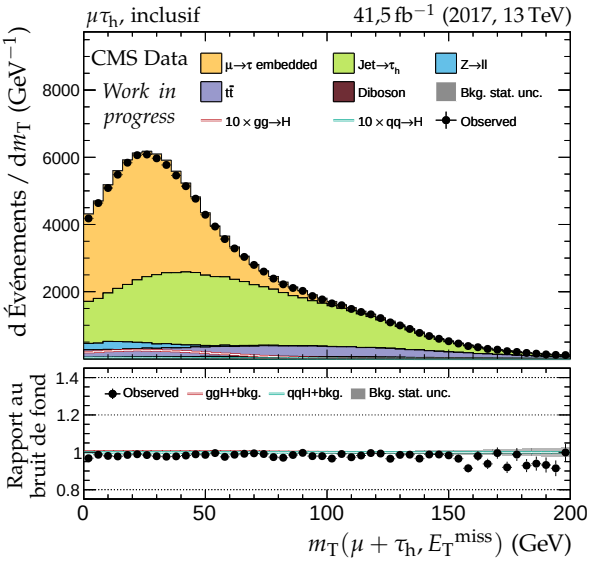


(f) *Angle azimutal du lepton 2 (τ_h).*

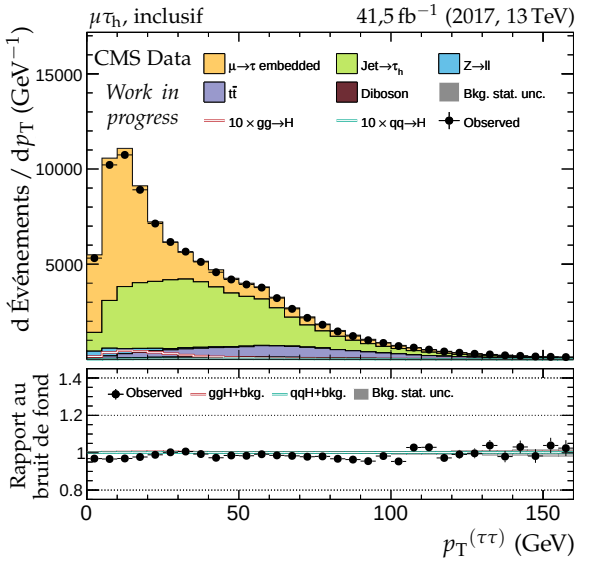
Figure G.33 – *Canal $\mu\tau_h$, 2017 : cinématique des leptons (μ, τ_h).*



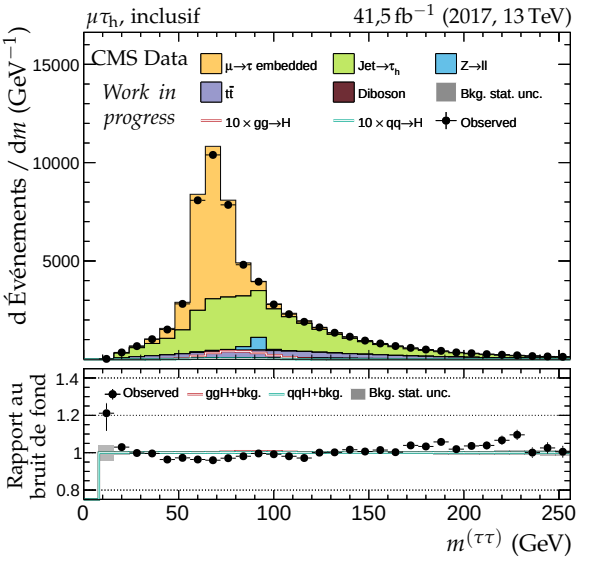
(a) *Énergie transverse manquante.*



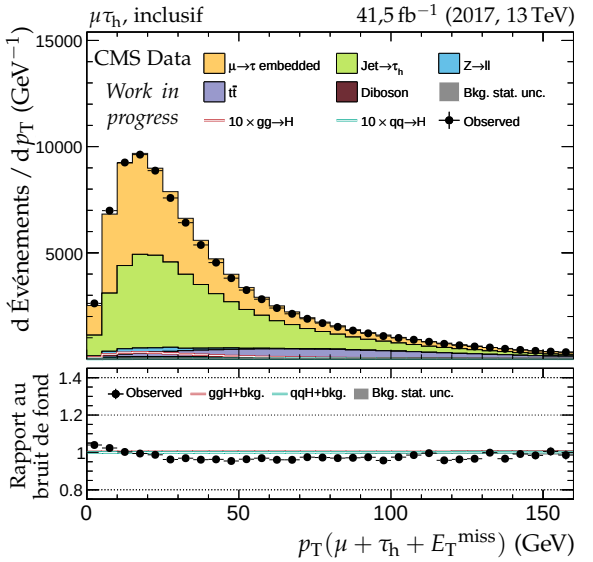
(b) *Masse transverse du dilepton.*



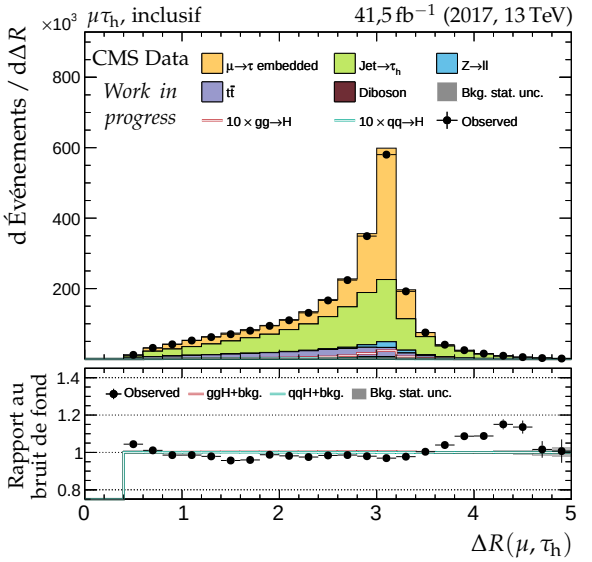
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

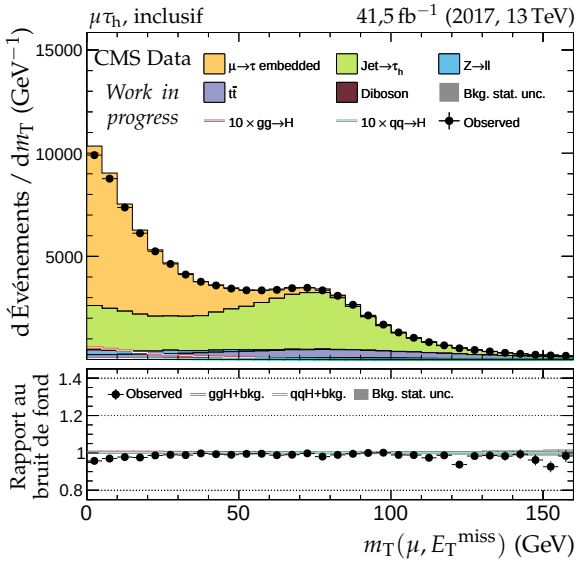


(e) *Impulsion transverse du système di-τ.*

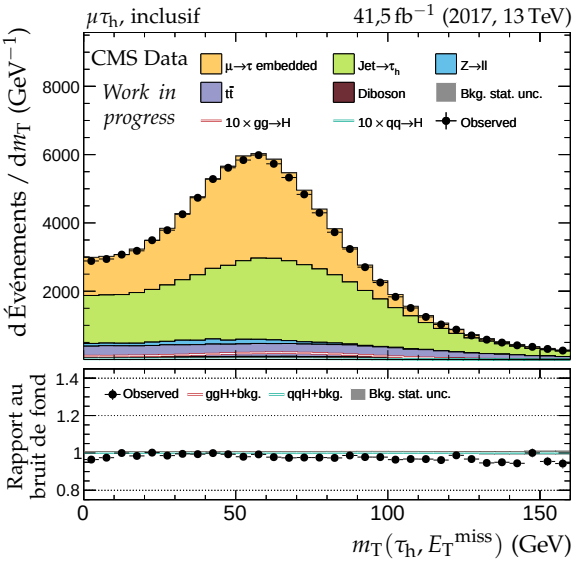


(f) *Distance ΔR entre les leptons (μ, τ_h).*

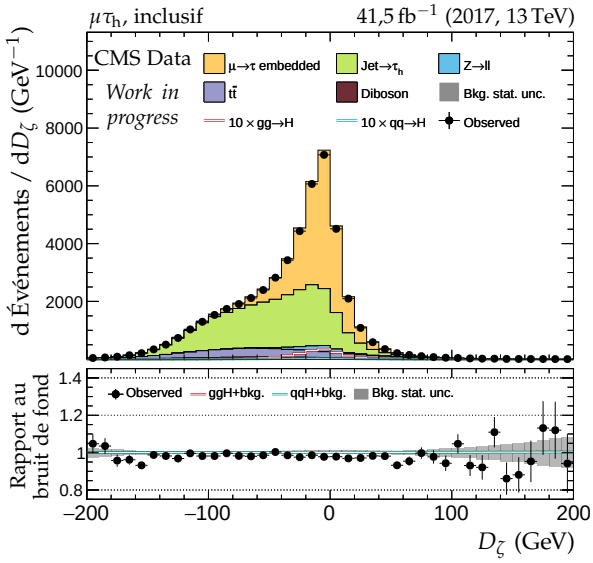
Figure G.34 – *Canal $\mu\tau_h$, 2017 : dilepton et énergie transverse manquante.*



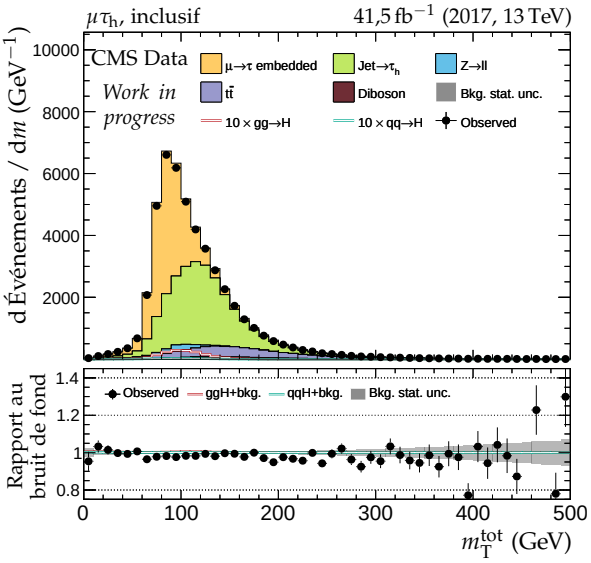
(a) *Masse transverse du lepton 1 (μ).*



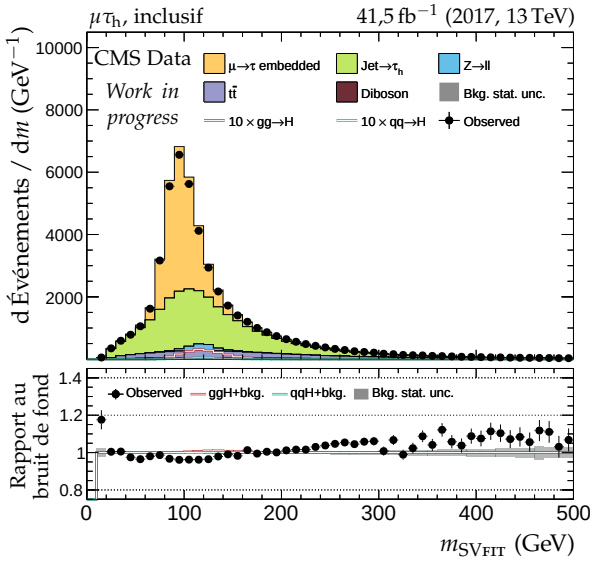
(b) *Masse transverse du lepton 2 (τ_h).*



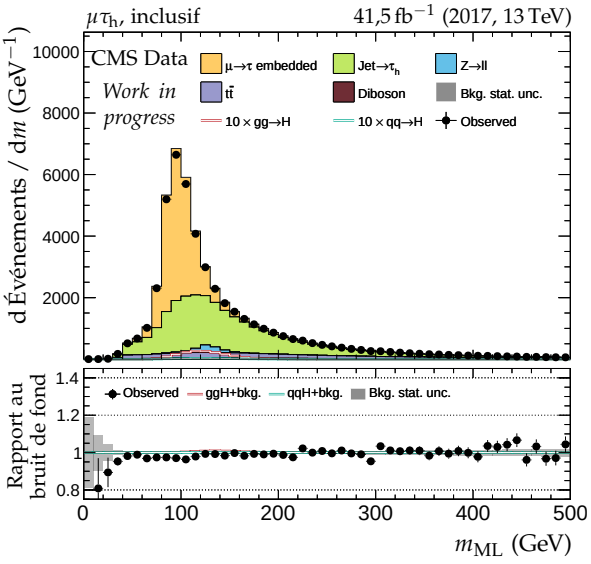
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

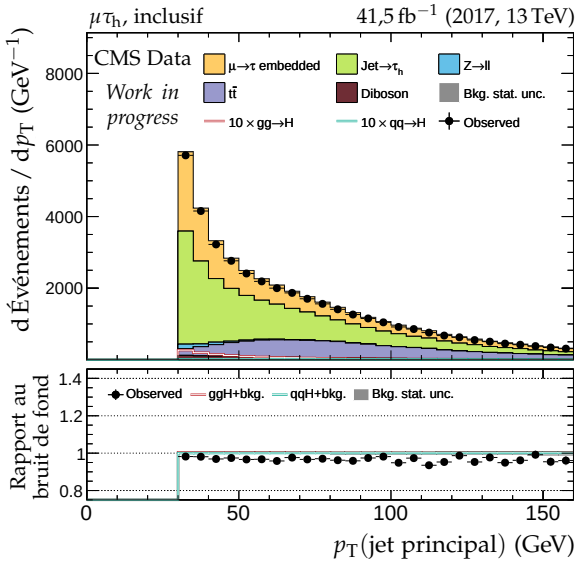


(e) *Masse du système di-τ d'après* SVFIT.

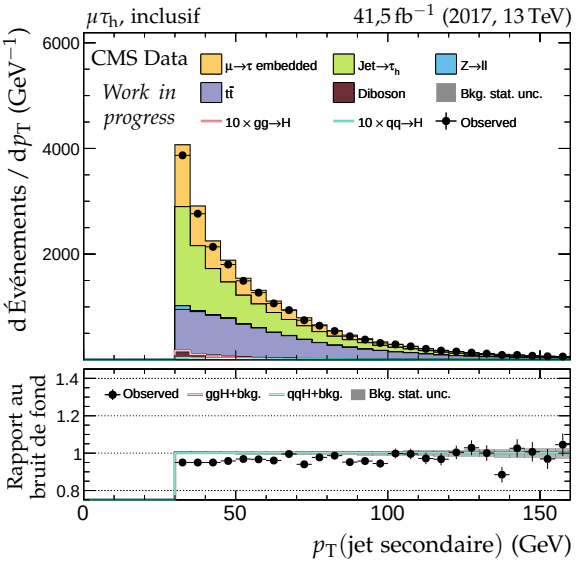


(f) *Masse du système di-τ d'après le ML.*

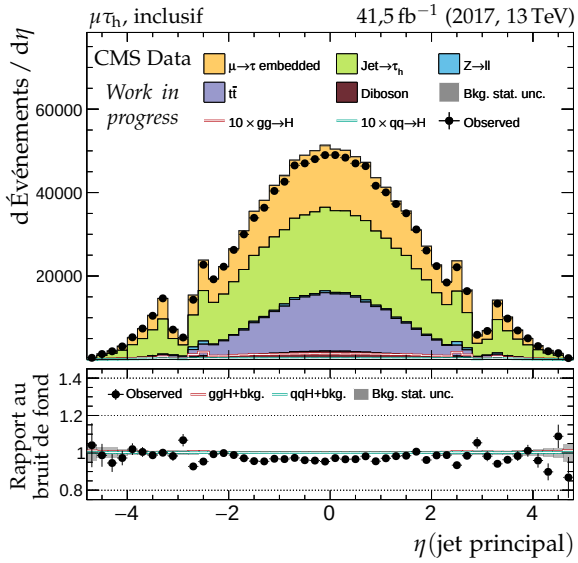
Figure G.35 – *Canal $\mu\tau_h$, 2017 : masses transverses, D_ζ et masses.*



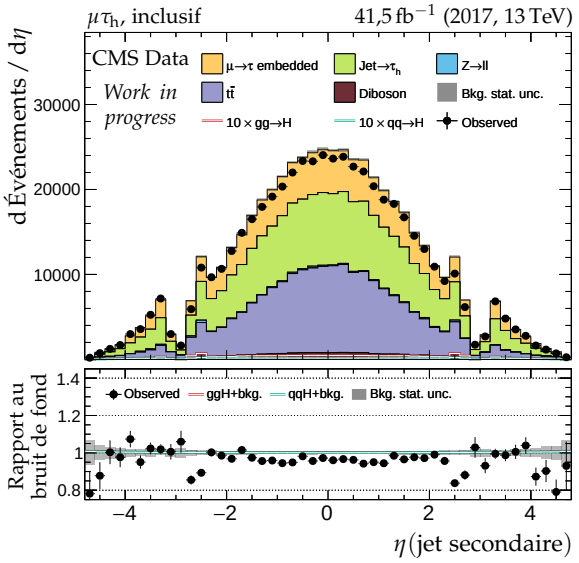
(a) *Impulsion transverse du jet principal.*



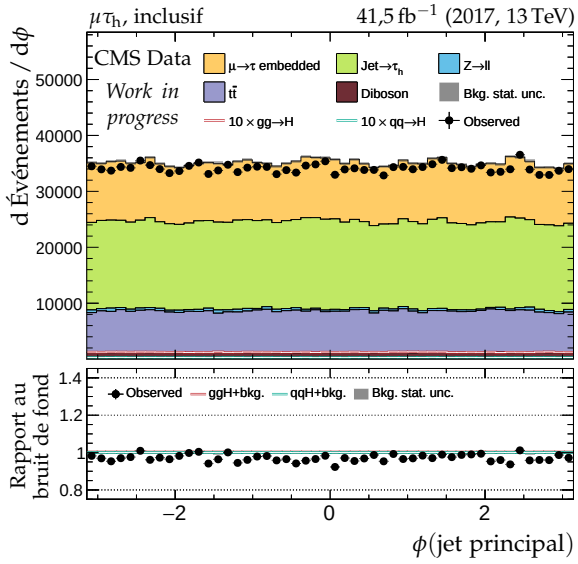
(b) *Impulsion transverse du jet secondaire.*



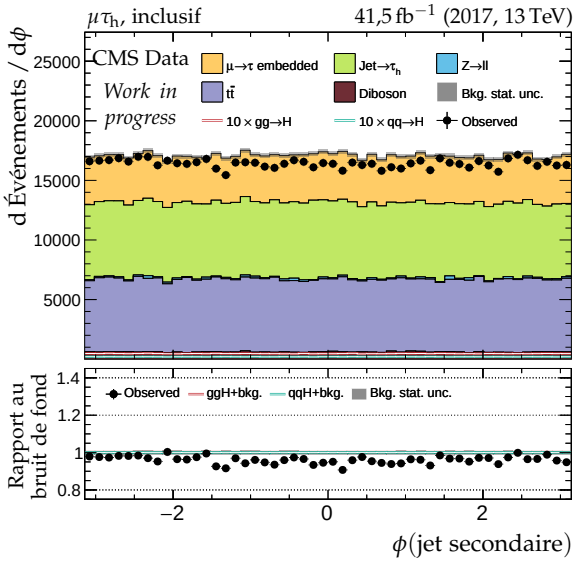
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

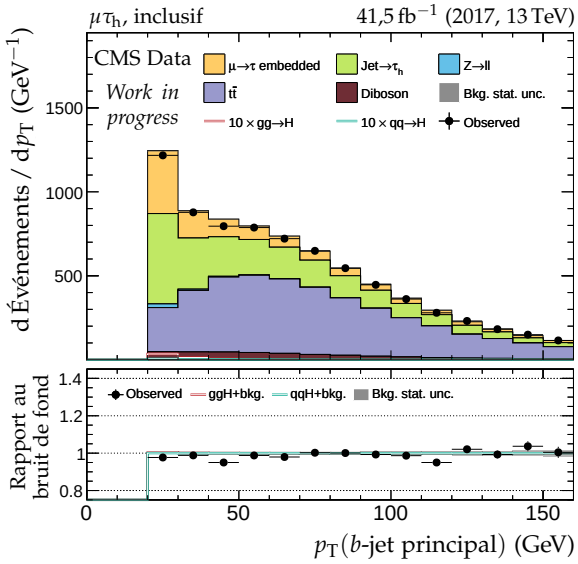


(e) *Angle azimutal du jet principal.*

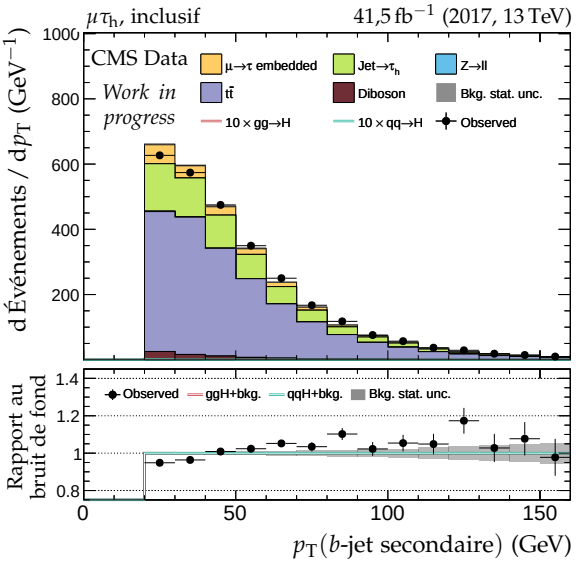


(f) *Angle azimutal du jet secondaire.*

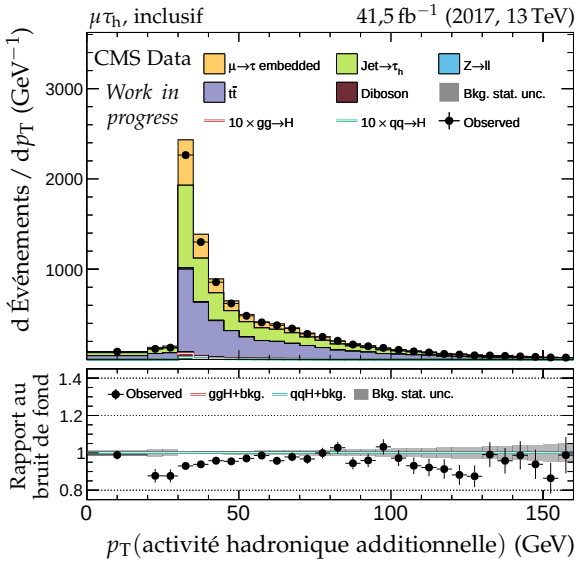
Figure G.36 – *Canal $\mu\tau_h$, 2017 : cinématique des deux jets principaux.*



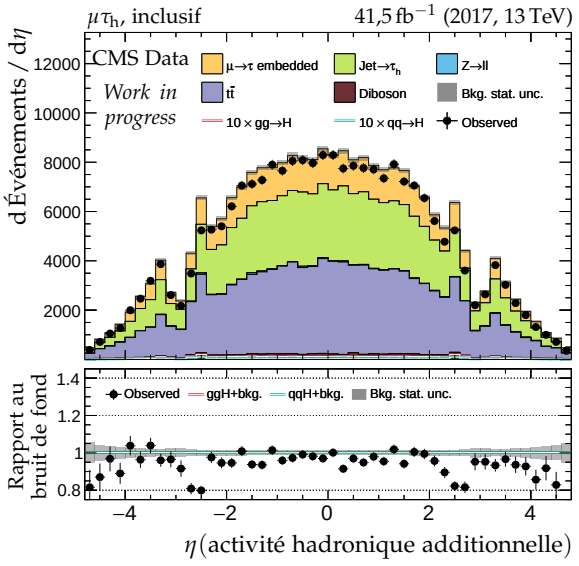
(a) *Impulsion transverse du b-jet principal.*



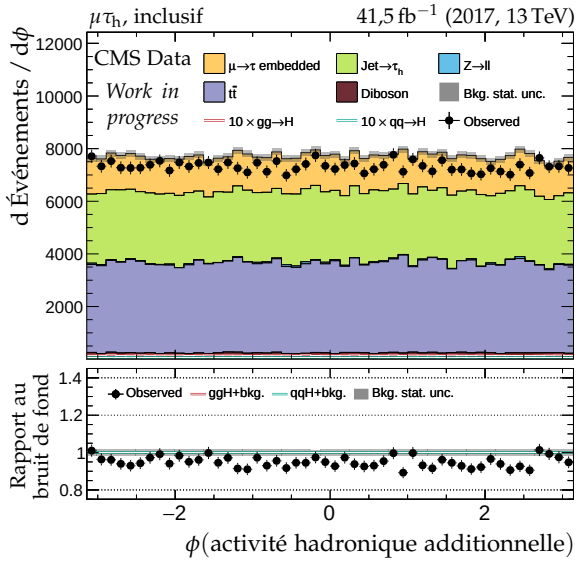
(b) *Impulsion transverse du b-jet secondaire.*



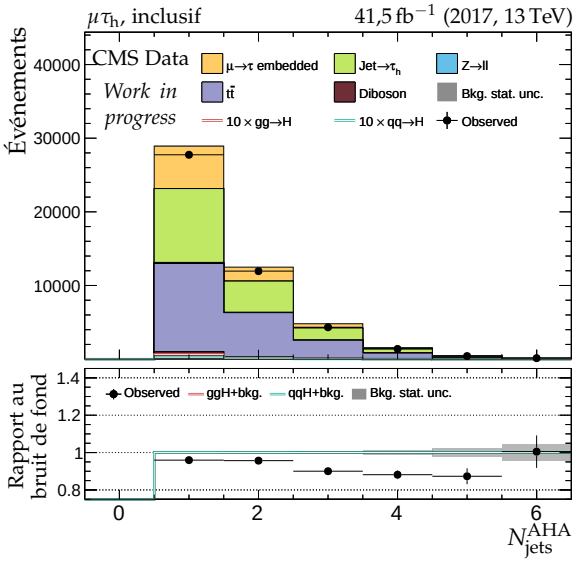
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

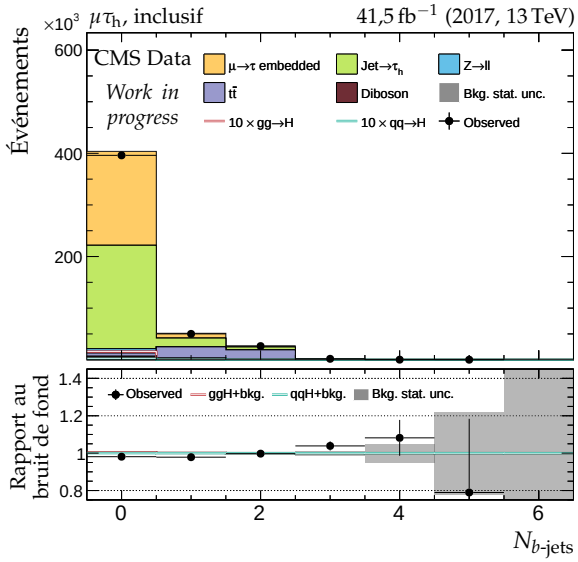


(e) *Angle azimutal de l'AHA.*

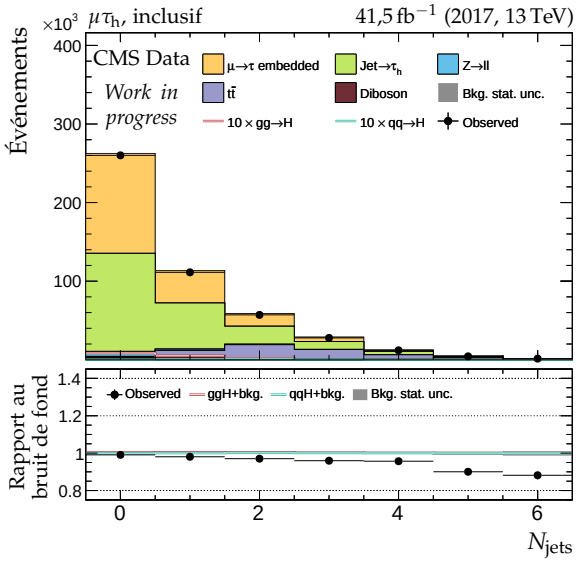


(f) *Nombre de jets dans l'AHA.*

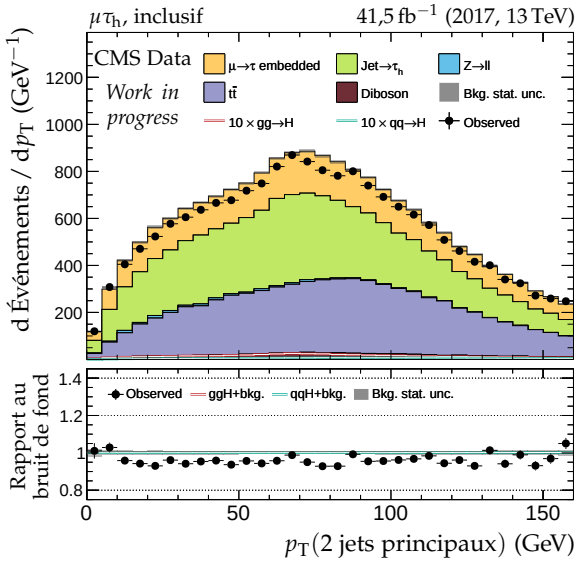
Figure G.37 – *Canal $\mu\tau_h$, 2017 : b-jets et activité hadronique additionnelle.*



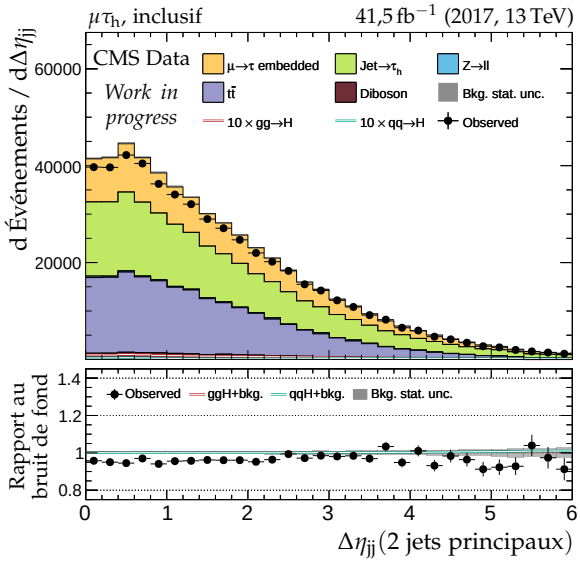
(a) *Nombre de b-jets.*



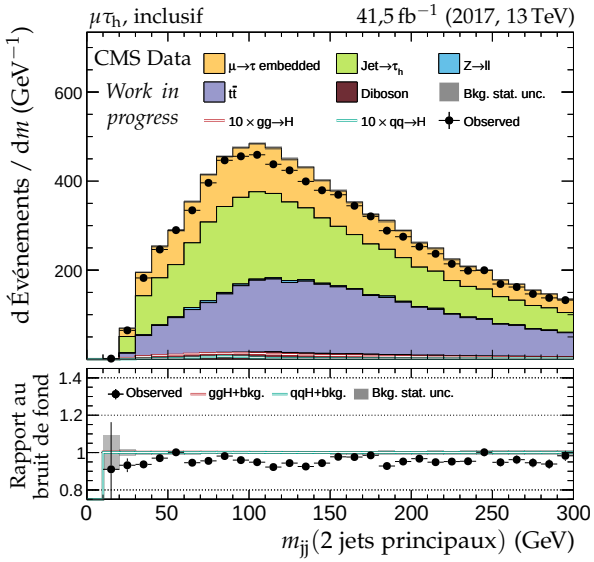
(b) *Nombre de jets.*



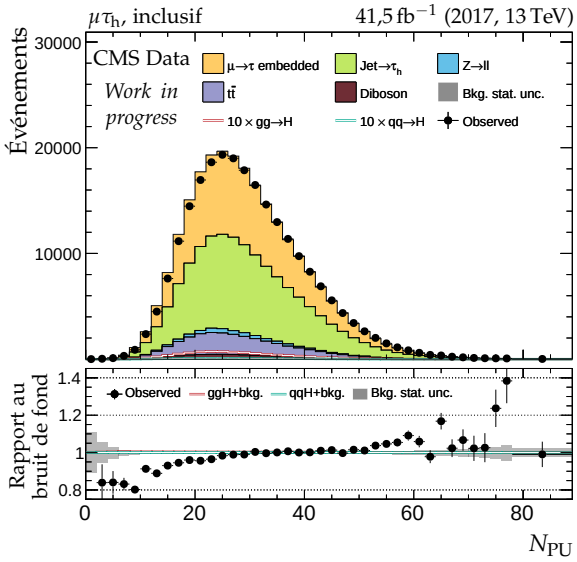
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

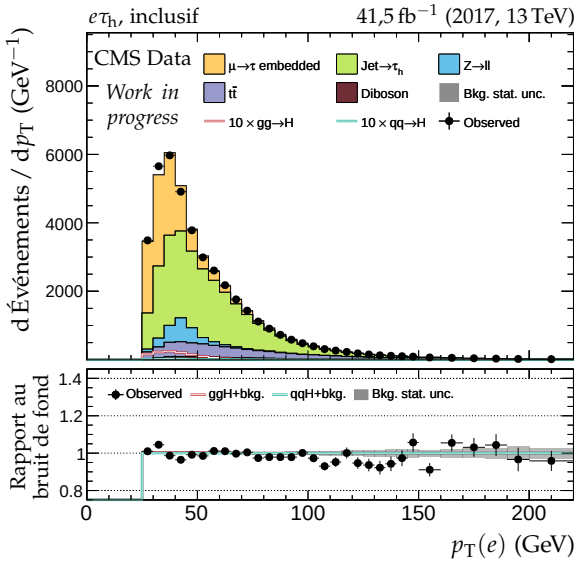


(e) *Masse invariante du système des deux jets.*

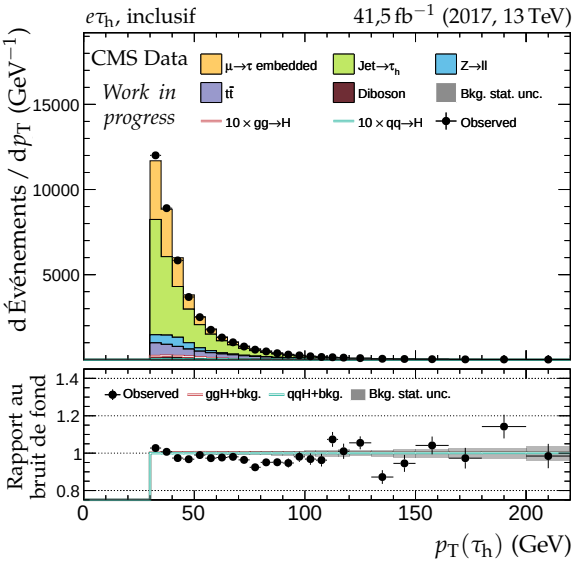


(f) *Nombre de vertex d'empilement.*

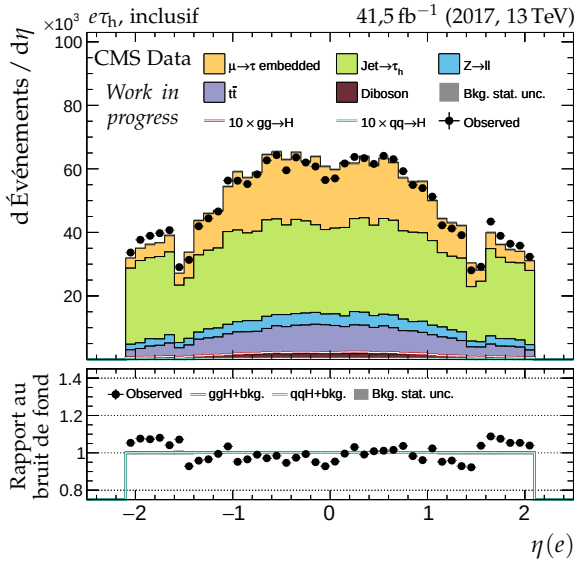
Figure G.38 – *Canal $\mu\tau_h$, 2017 : nombre de jets, système des deux jets principaux et empilement.*



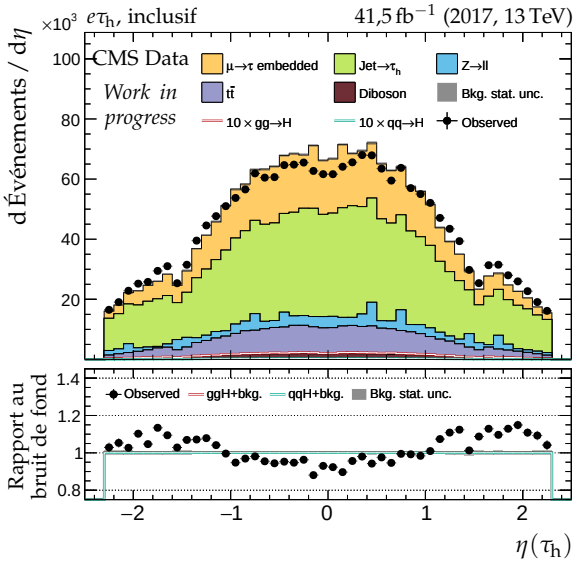
(a) *Impulsion transverse du lepton 1 (e).*



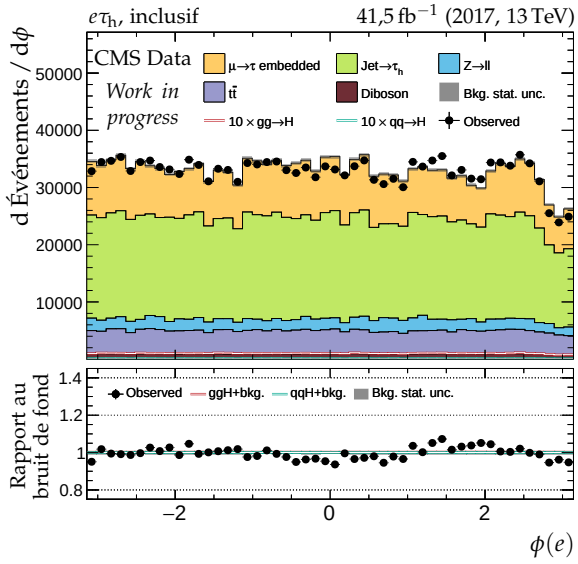
(b) *Impulsion transverse du lepton 2 (τ_h).*



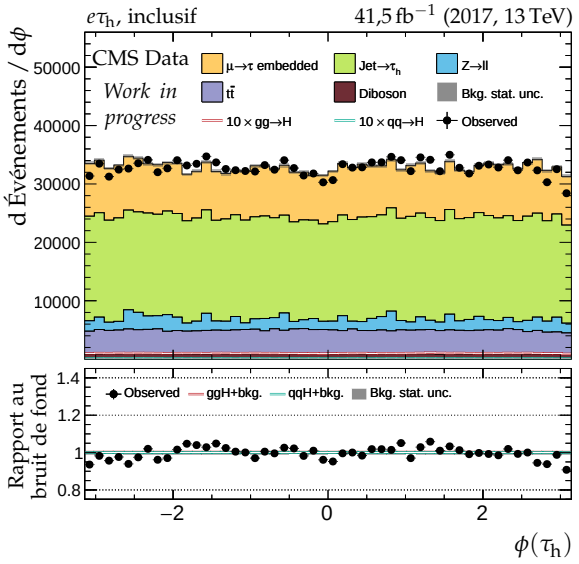
(c) *Pseudo-rapidité du lepton 1 (e).*



(d) *Pseudo-rapidité du lepton 2 (τ_h).*

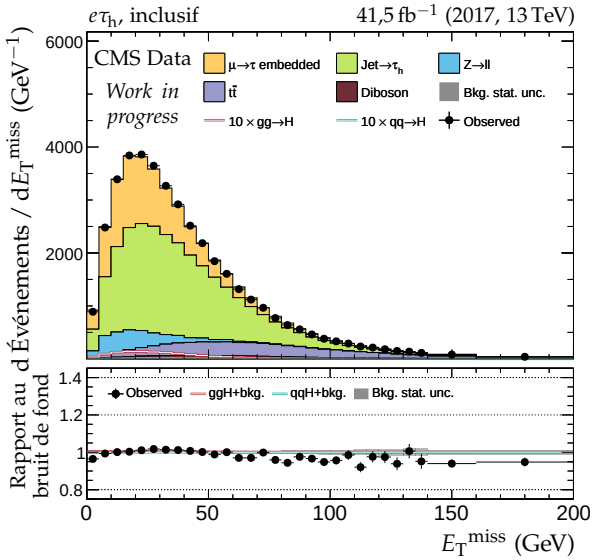


(e) *Angle azimutal du lepton 1 (e).*

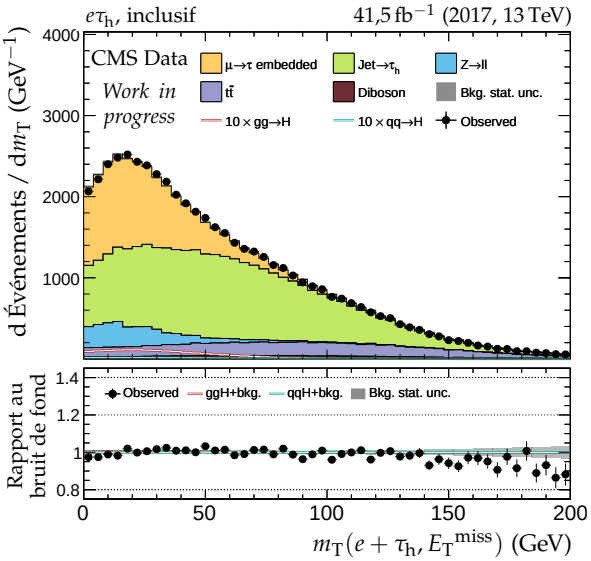


(f) *Angle azimutal du lepton 2 (τ_h).*

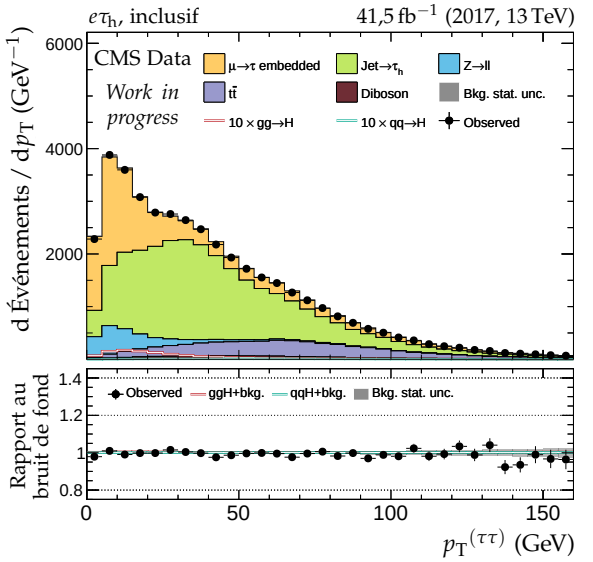
Figure G.39 – *Canal $e\tau_h$, 2017 : cinématique des leptons (e, τ_h).*



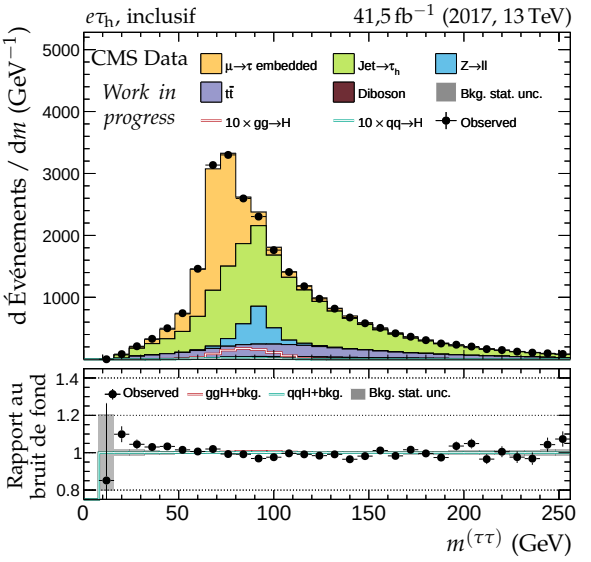
(a) *Énergie transverse manquante.*



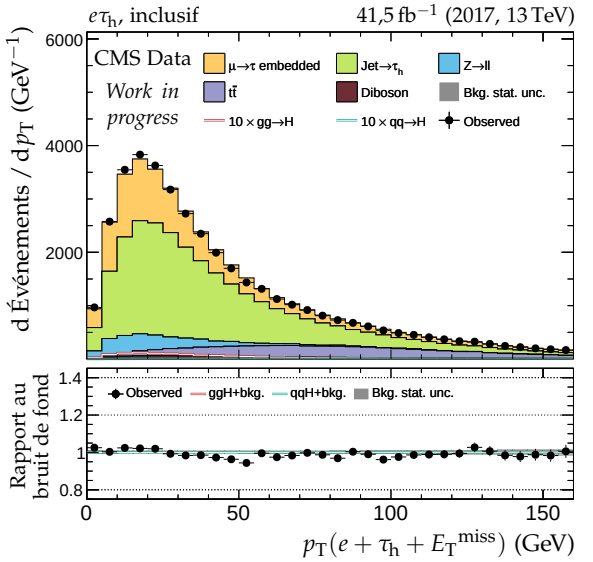
(b) *Masse transverse du dilepton.*



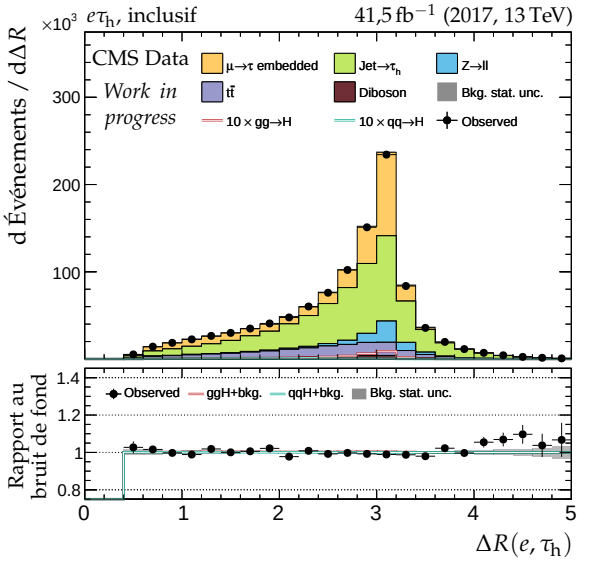
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

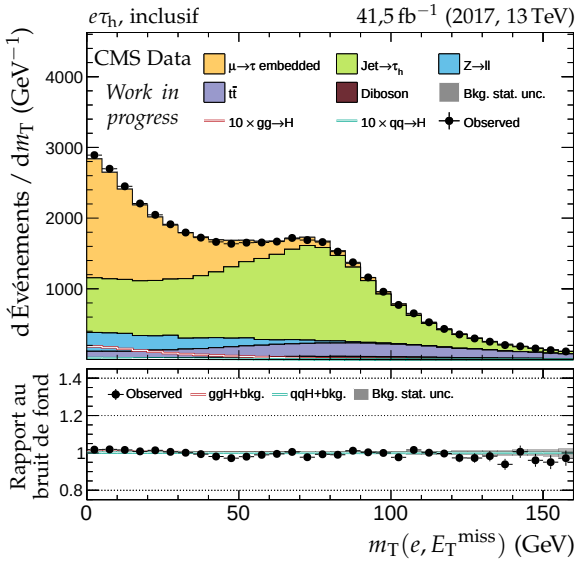


(e) *Impulsion transverse du système di-τ.*

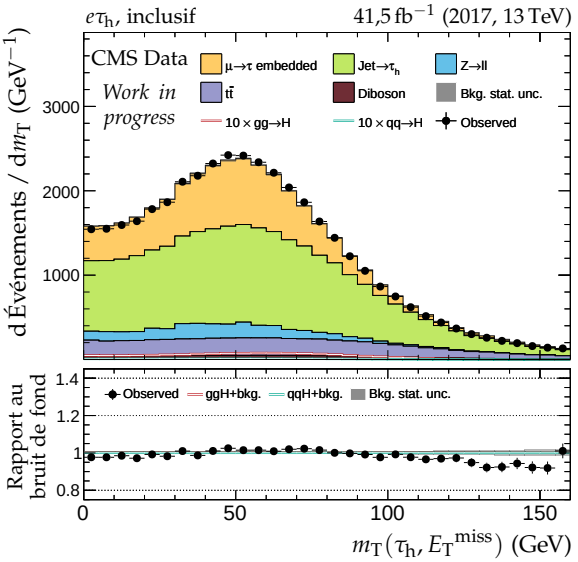


(f) *Distance ΔR entre les leptons (e, τ_h).*

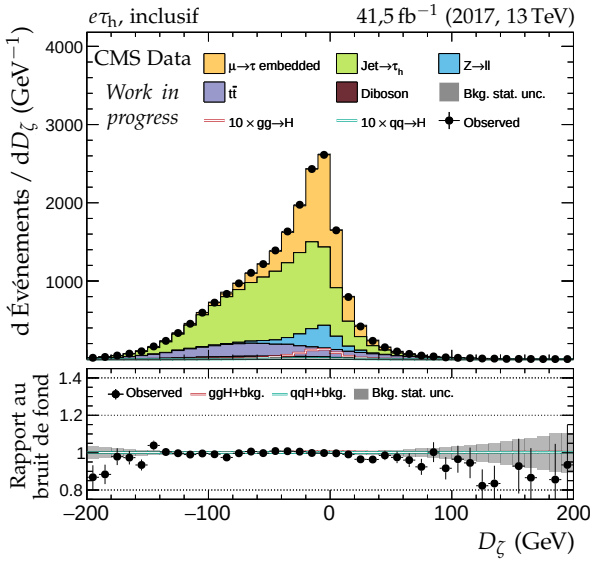
Figure G.40 – *Canal $e\tau_h$, 2017 : dilepton et énergie transverse manquante.*



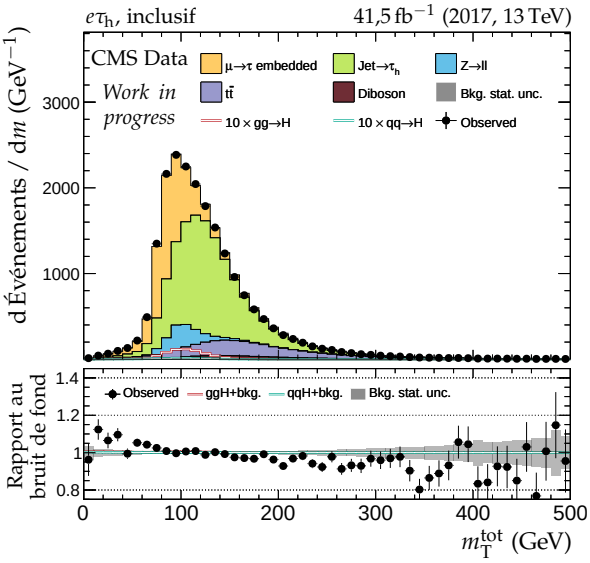
(a) *Masse transverse du lepton 1 (e).*



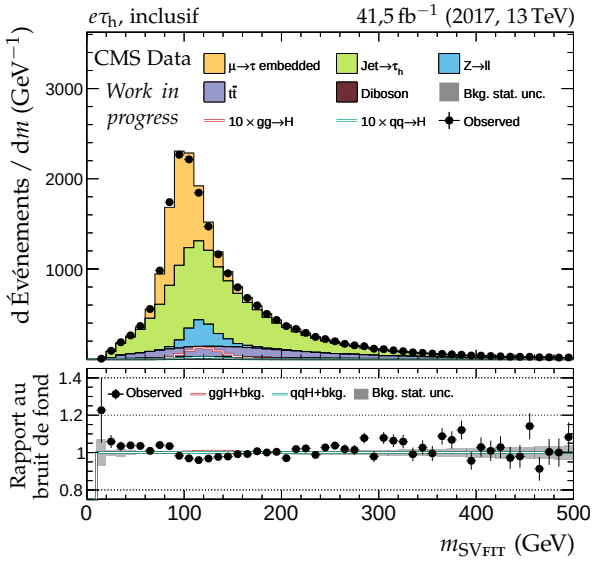
(b) *Masse transverse du lepton 2 (τ_h).*



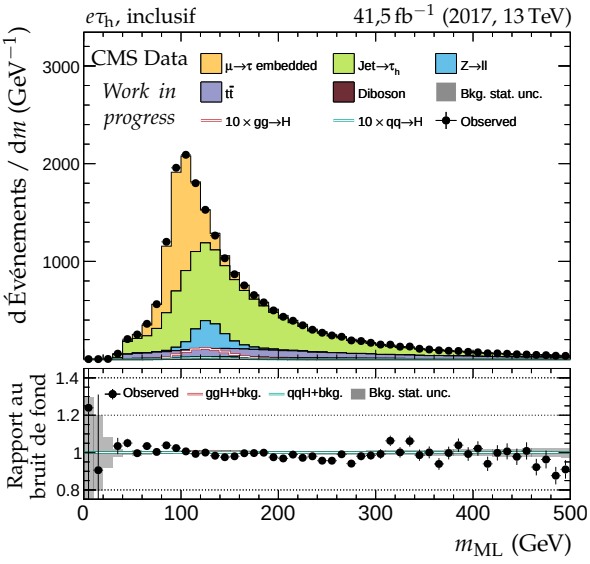
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

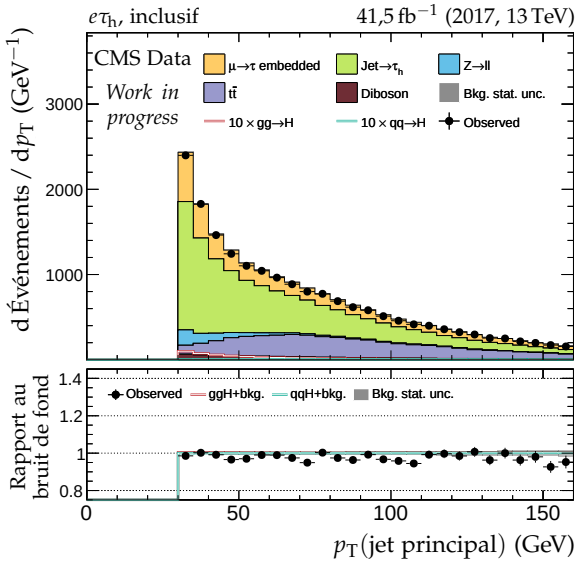


(e) *Masse du système di-τ d'après* SVFIT.

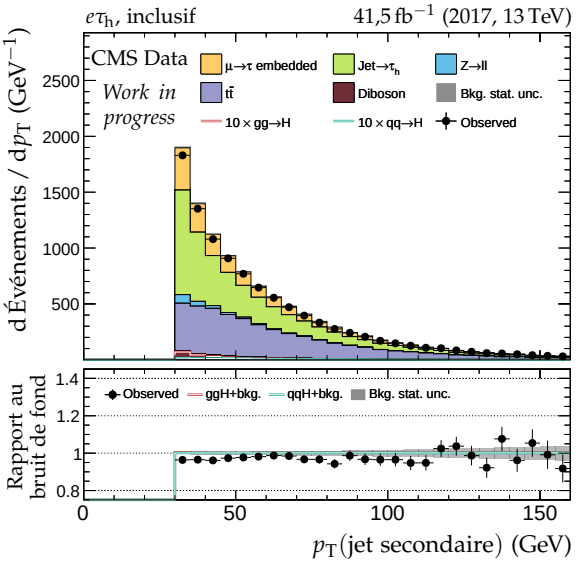


(f) *Masse du système di-τ d'après le ML.*

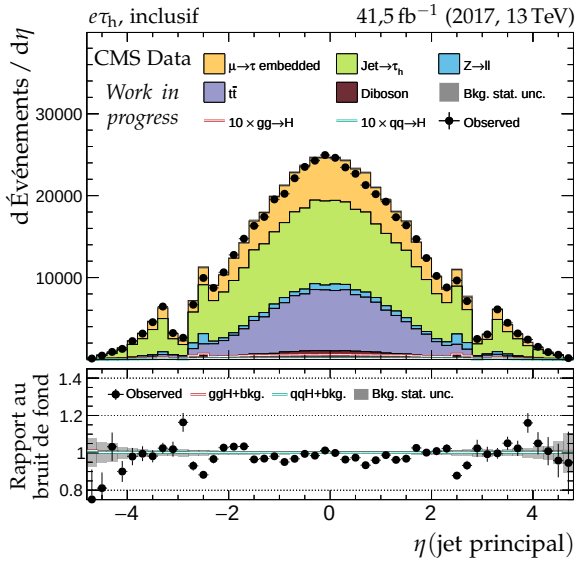
Figure G.41 – *Canal $e\tau_h$, 2017 : masses transverses, D_ζ et masses.*



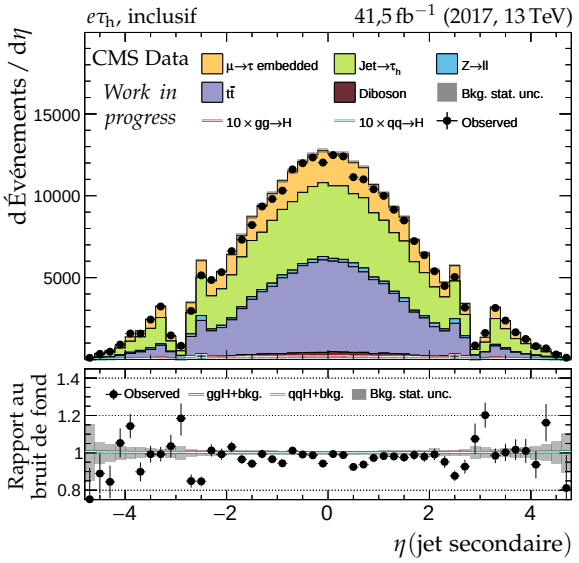
(a) *Impulsion transverse du jet principal.*



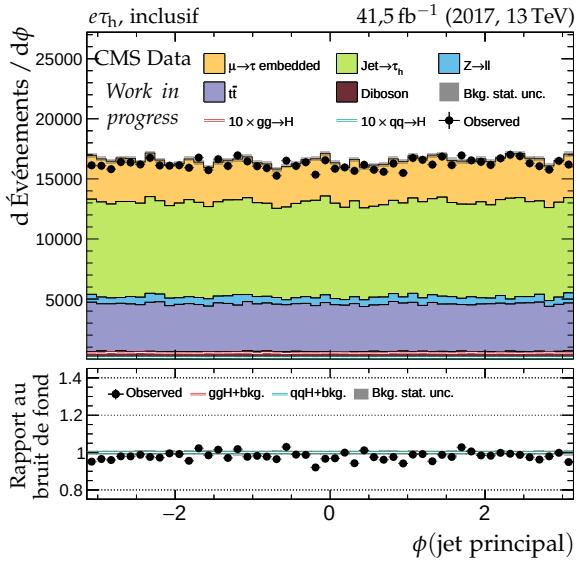
(b) *Impulsion transverse du jet secondaire.*



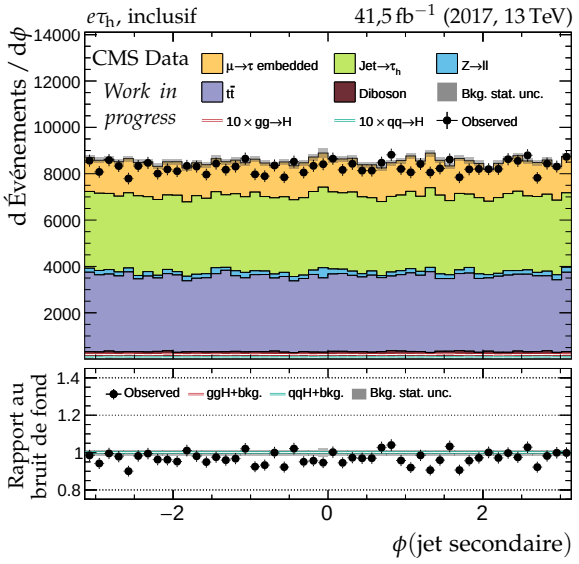
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

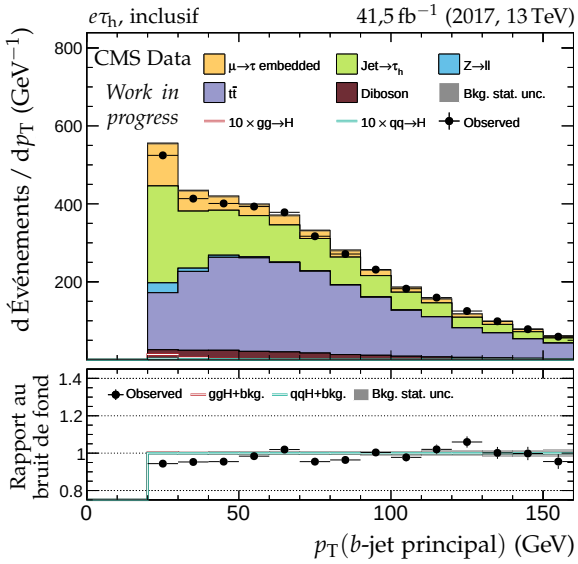


(e) *Angle azimutal du jet principal.*

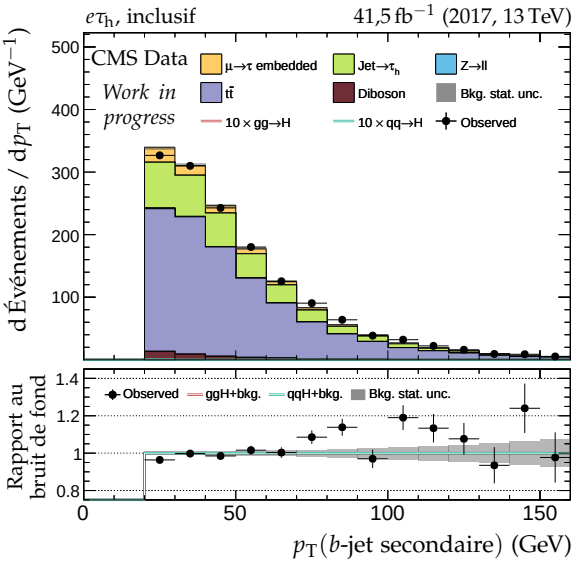


(f) *Angle azimutal du jet secondaire.*

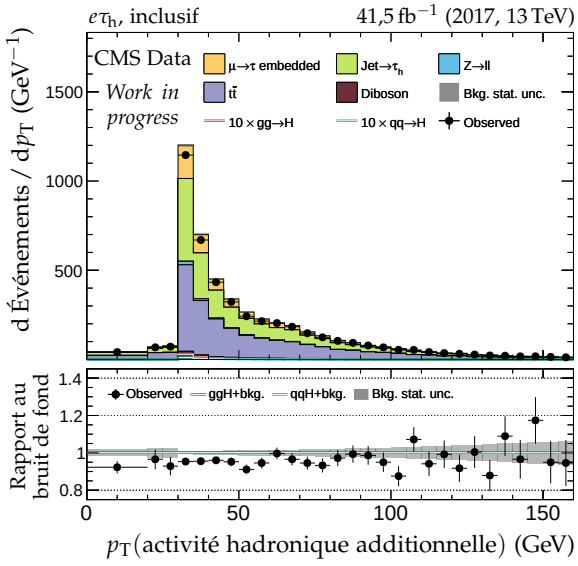
Figure G.42 – *Canal $e\tau_h$, 2017 : cinématique des deux jets principaux.*



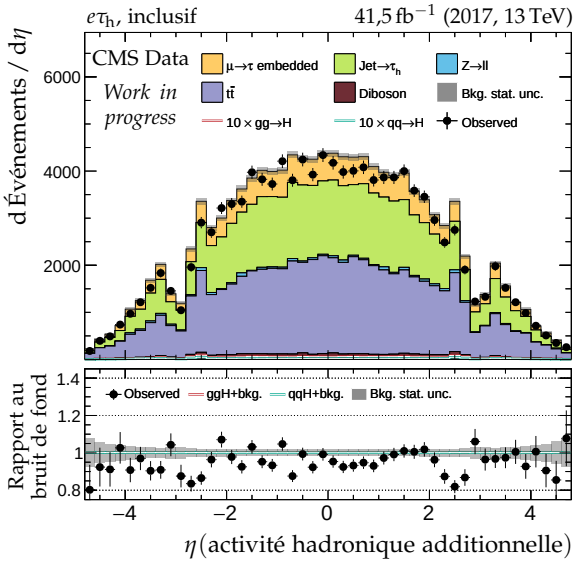
(a) *Impulsion transverse du b-jet principal.*



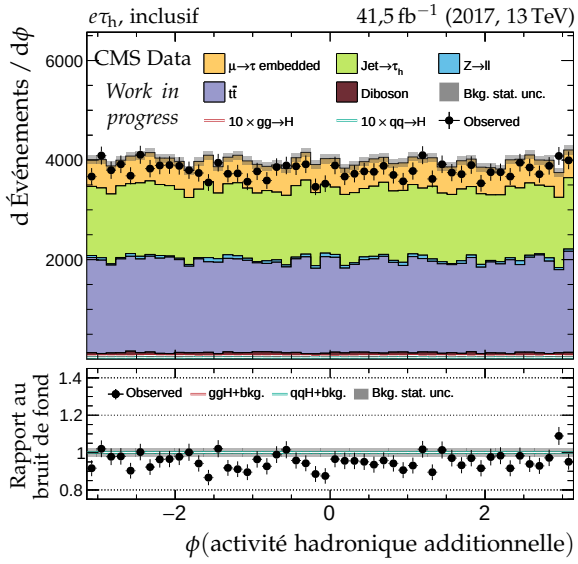
(b) *Impulsion transverse du b-jet secondaire.*



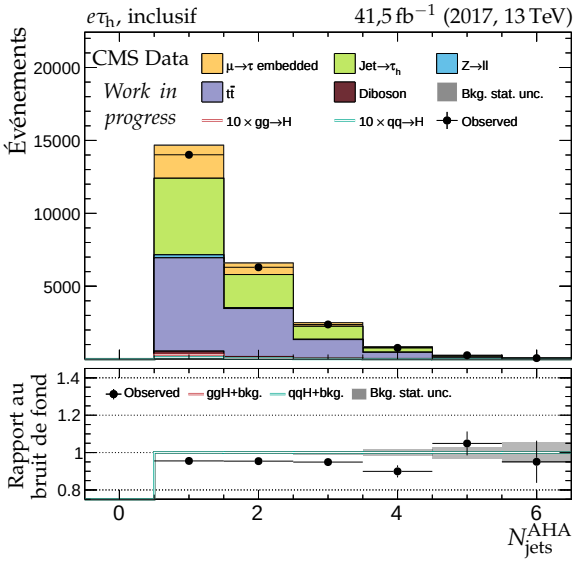
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

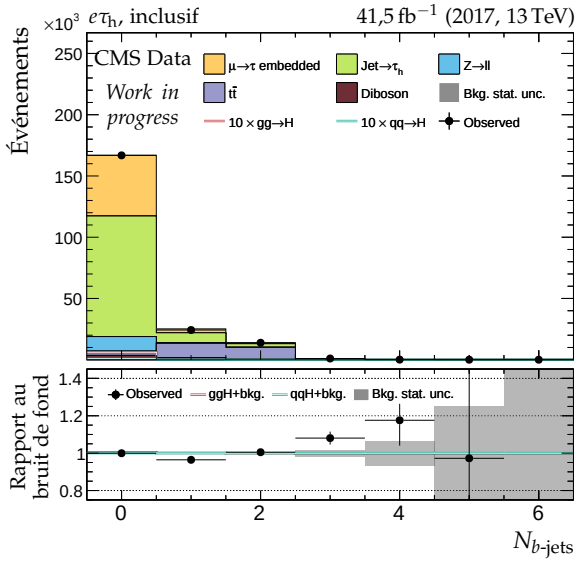


(e) *Angle azimutal de l'AHA.*

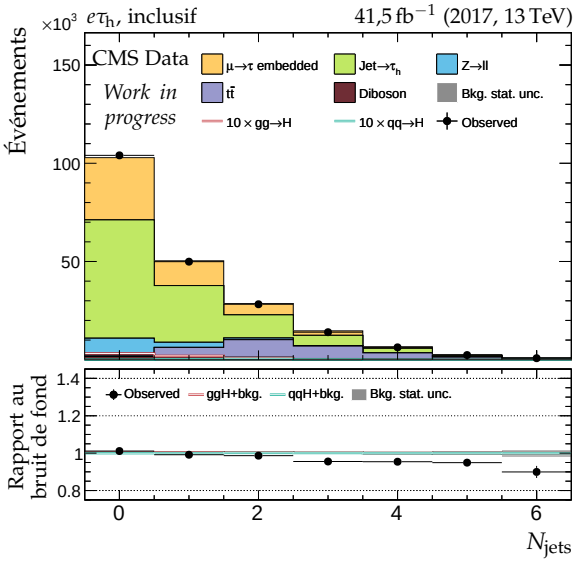


(f) *Nombre de jets dans l'AHA.*

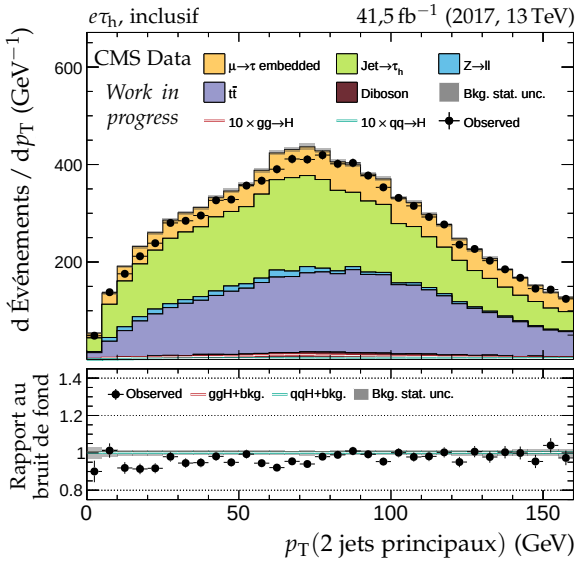
Figure G.43 – *Canal eτh, 2017 : b-jets et activité hadronique additionnelle.*



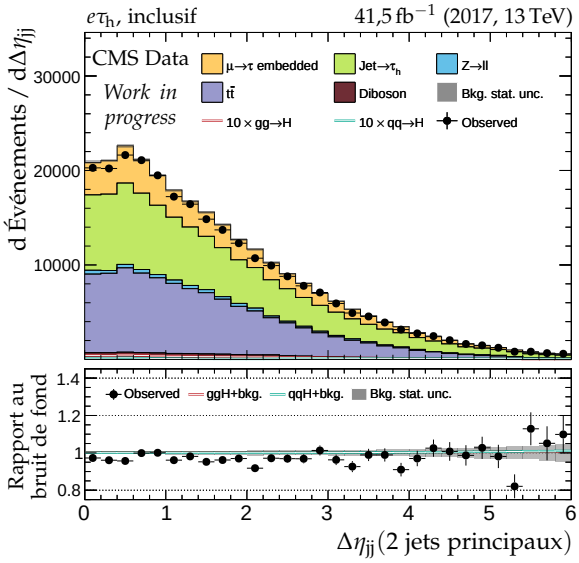
(a) *Nombre de b-jets.*



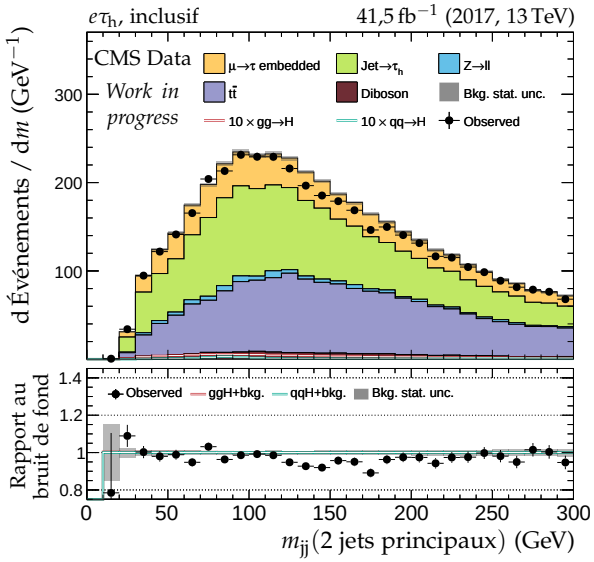
(b) *Nombre de jets.*



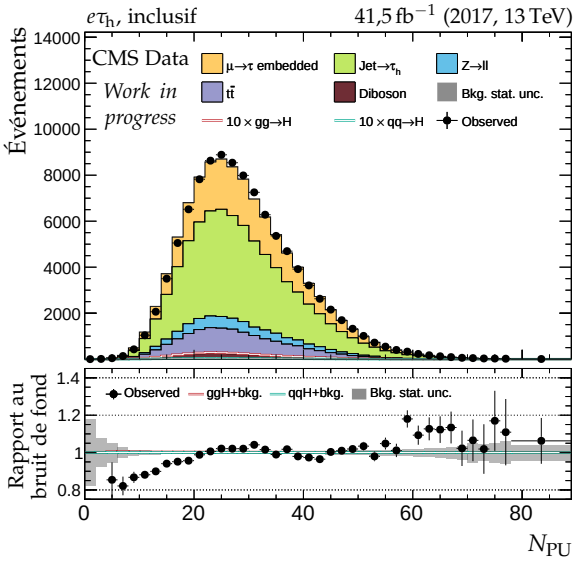
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

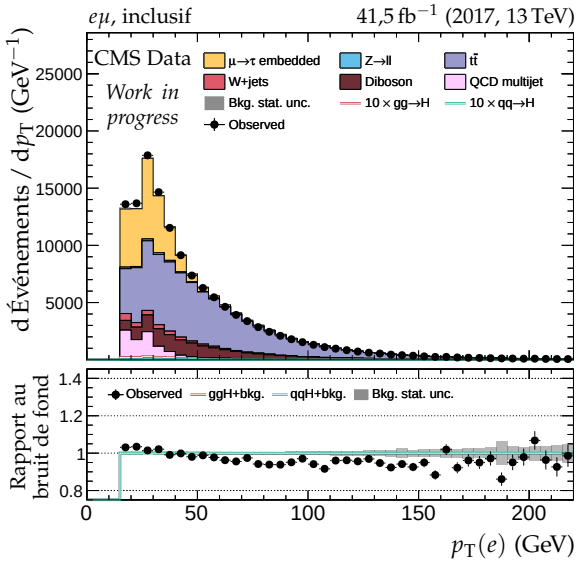


(e) *Masse invariante du système des deux jets.*

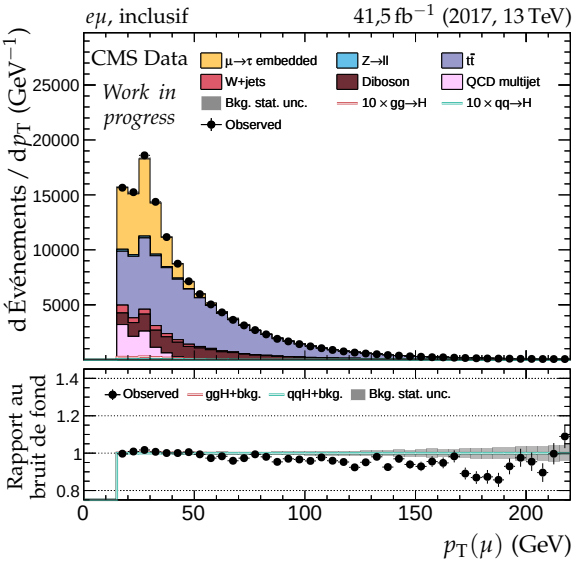


(f) *Nombre de vertex d'empilement.*

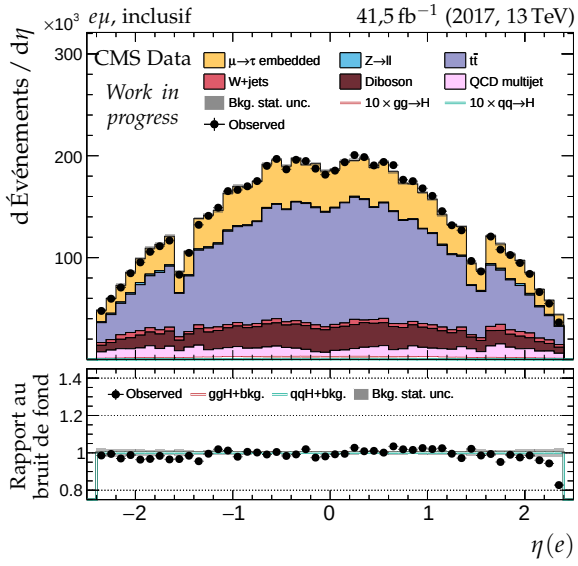
Figure G.44 – *Canal $e\tau_h$, 2017 : nombre de jets, système des deux jets principaux et empilement.*



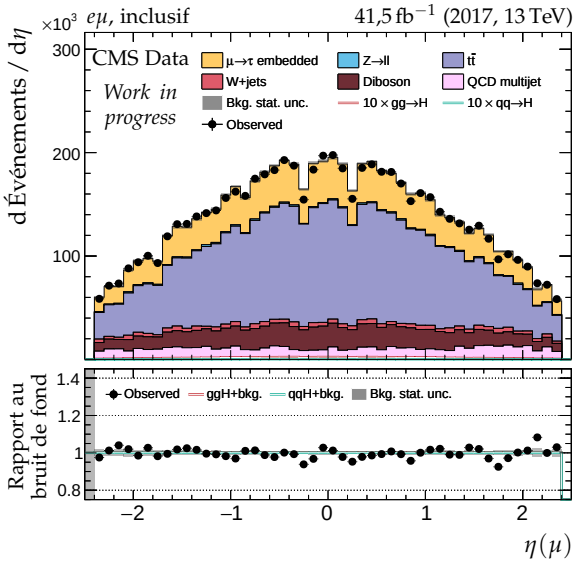
(a) *Impulsion transverse du lepton 1 (e).*



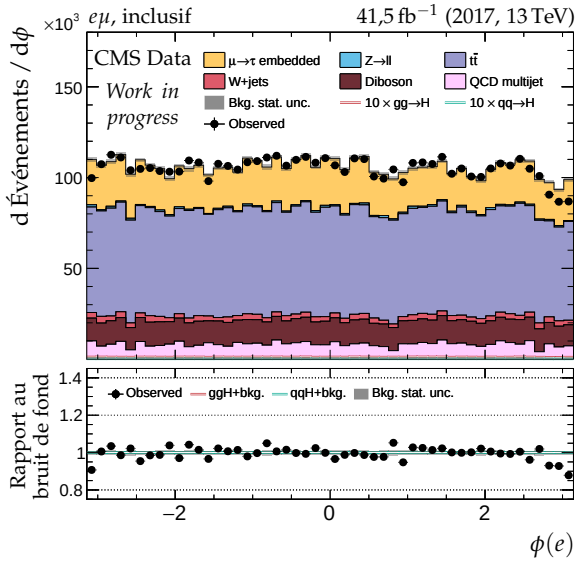
(b) *Impulsion transverse du lepton 2 (μ).*



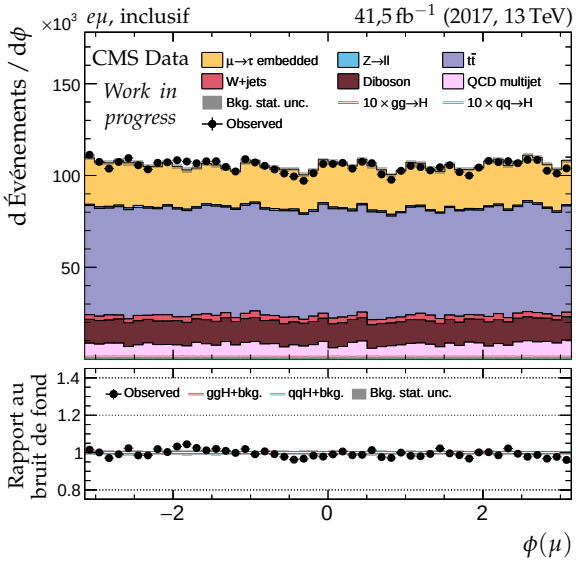
(c) *Pseudo-rapidité du lepton 1 (e).*



(d) *Pseudo-rapidité du lepton 2 (μ).*

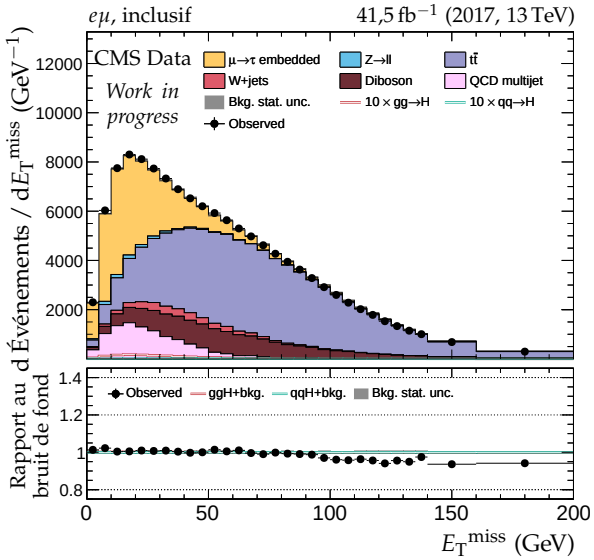


(e) *Angle azimutal du lepton 1 (e).*

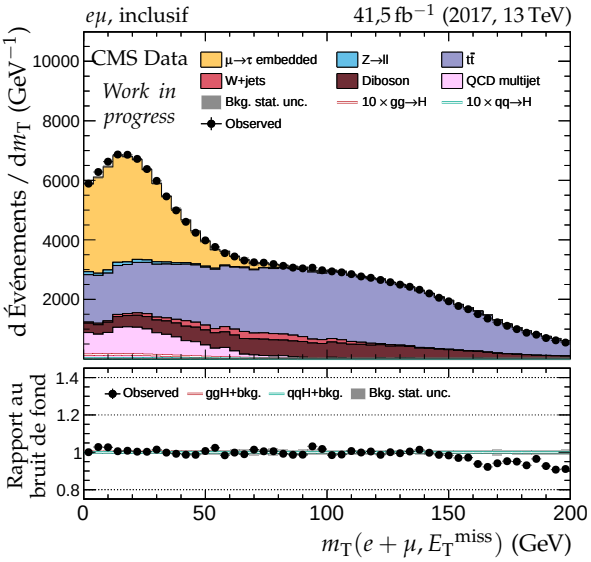


(f) *Angle azimutal du lepton 2 (μ).*

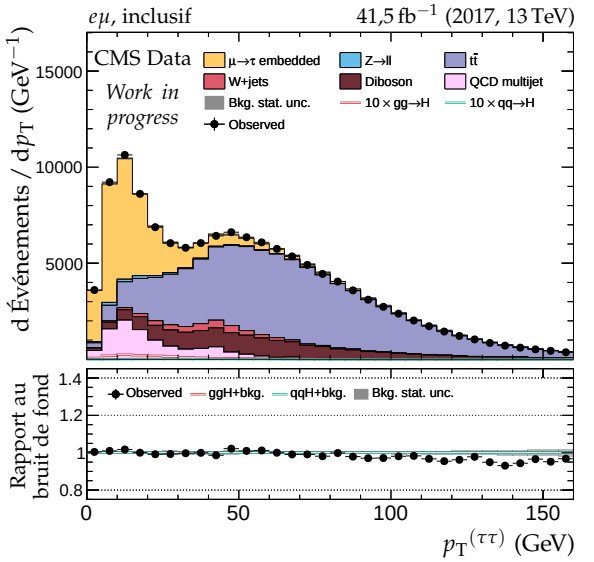
Figure G.45 – *Canal eμ, 2017 : cinématique des leptons (e, μ).*



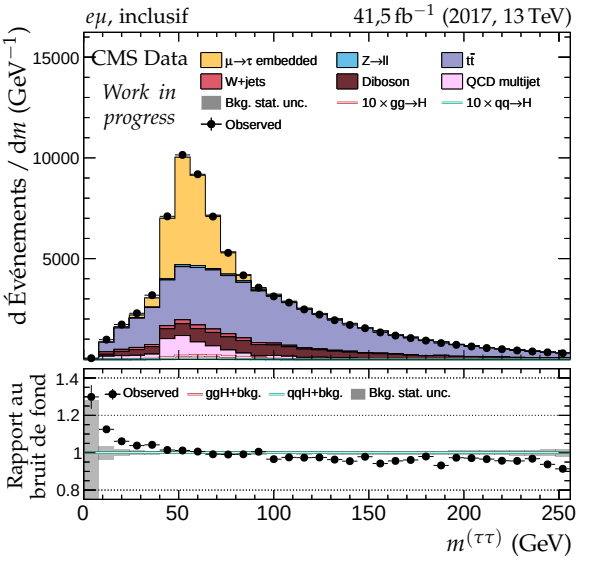
(a) *Énergie transverse manquante.*



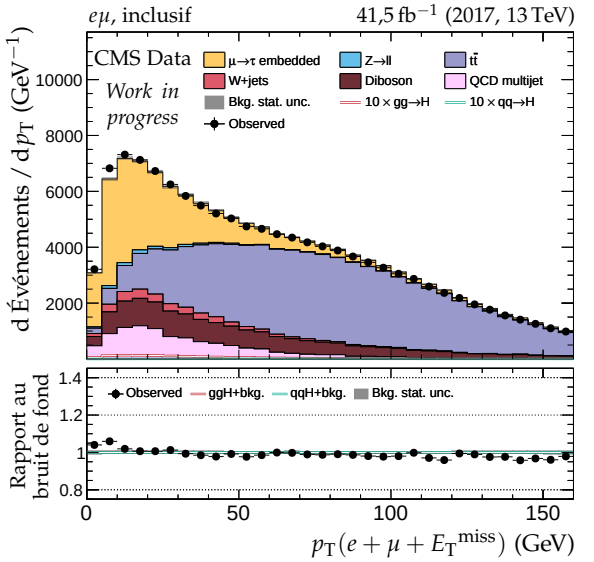
(b) *Masse transverse du dilepton.*



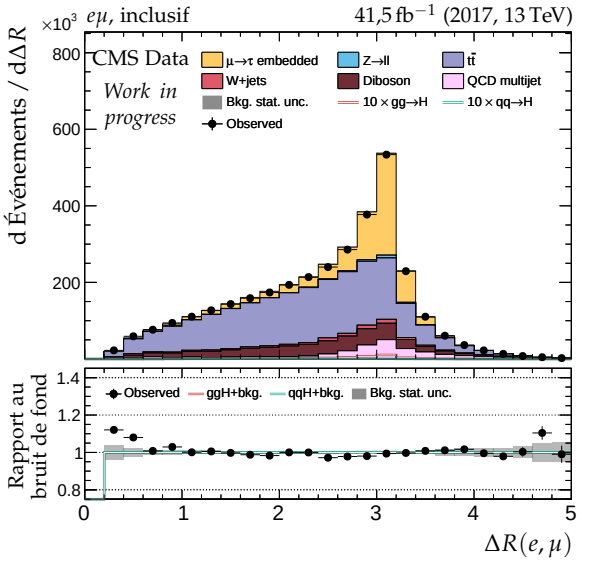
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

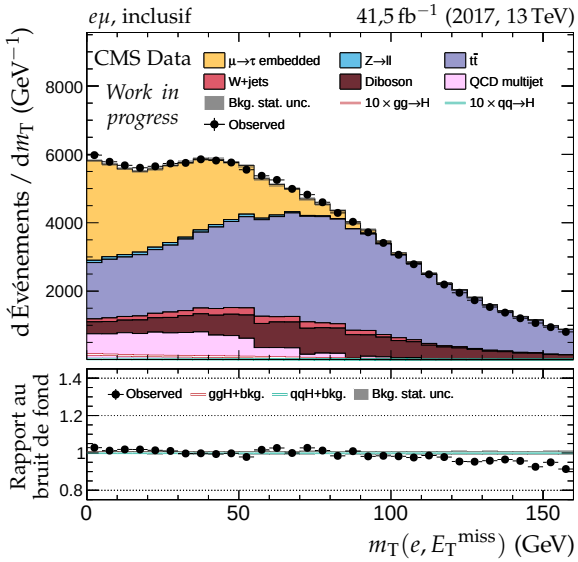


(e) *Impulsion transverse du système di-τ.*

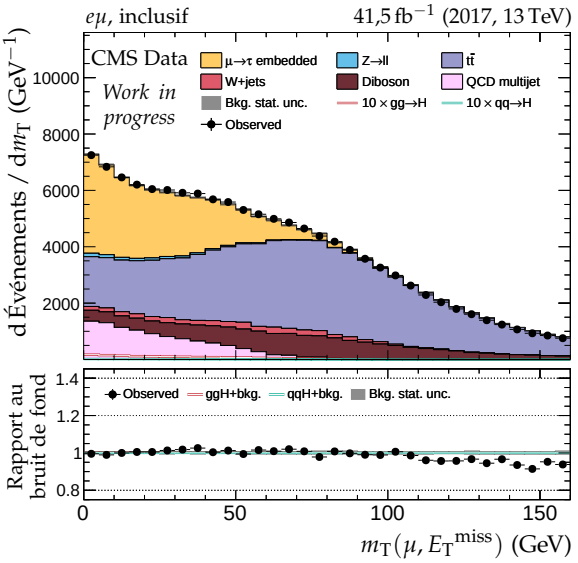


(f) *Distance ΔR entre les leptons (e, μ).*

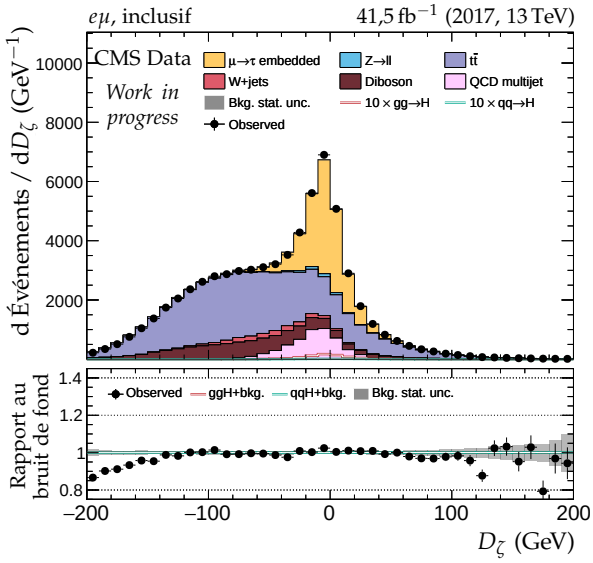
Figure G.46 – *Canal eμ, 2017 : dilepton et énergie transverse manquante.*



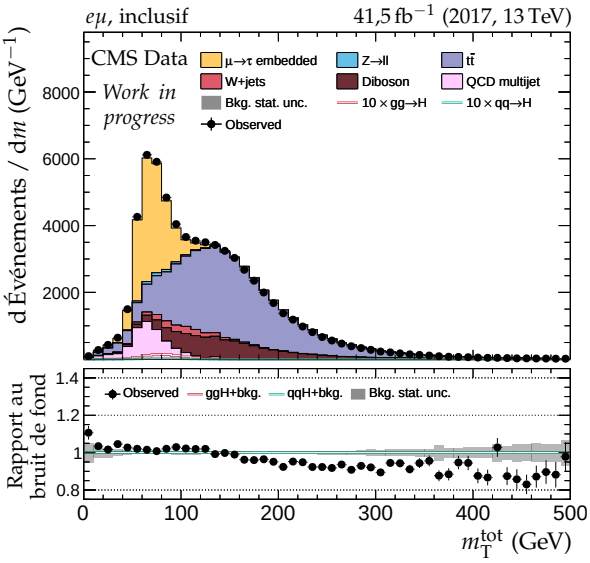
(a) *Masse transverse du lepton 1 (e).*



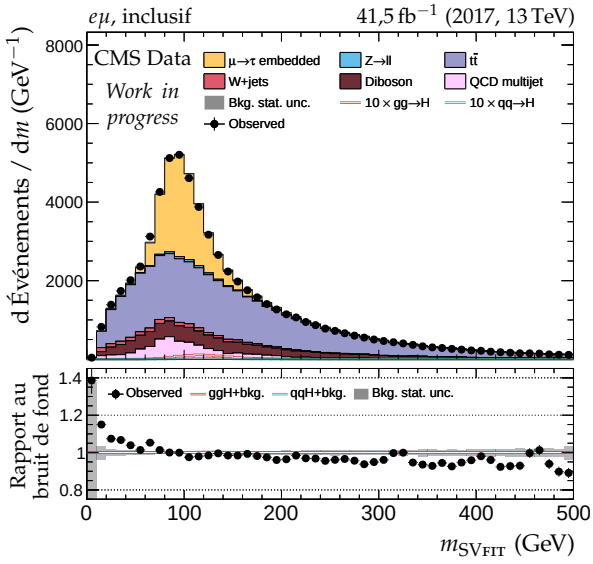
(b) *Masse transverse du lepton 2 (μ).*



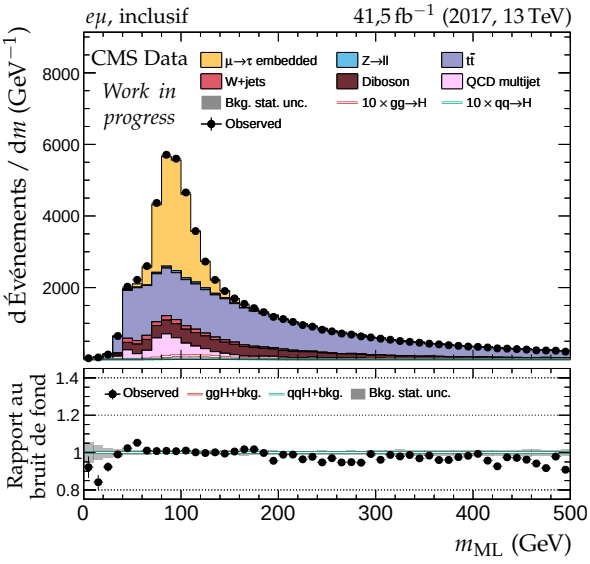
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

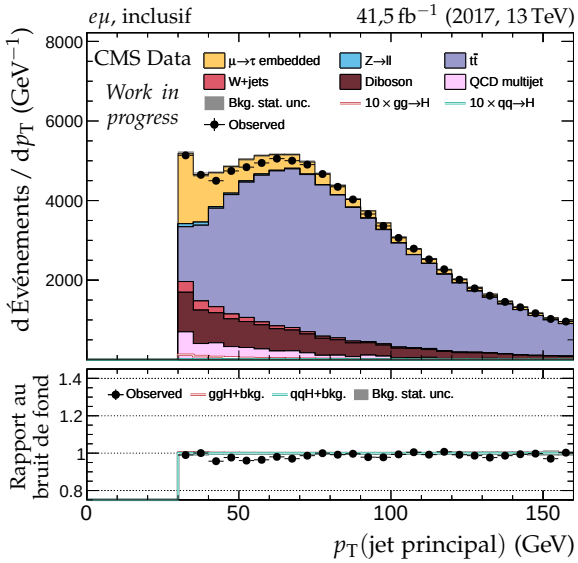


(e) *Masse du système di-τ d'après* SVFIT.

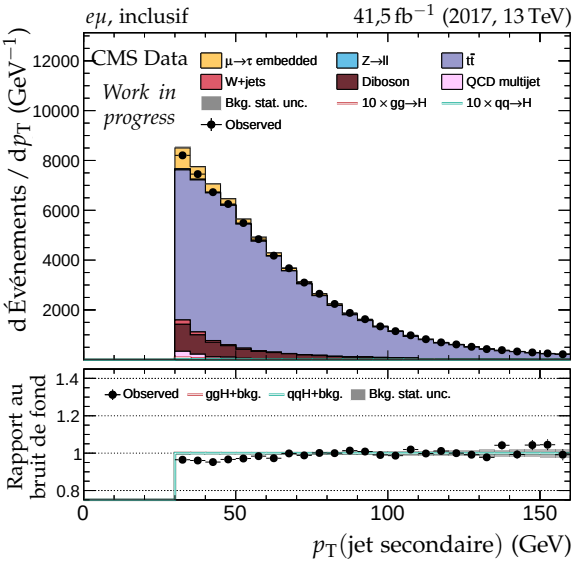


(f) *Masse du système di-τ d'après le ML.*

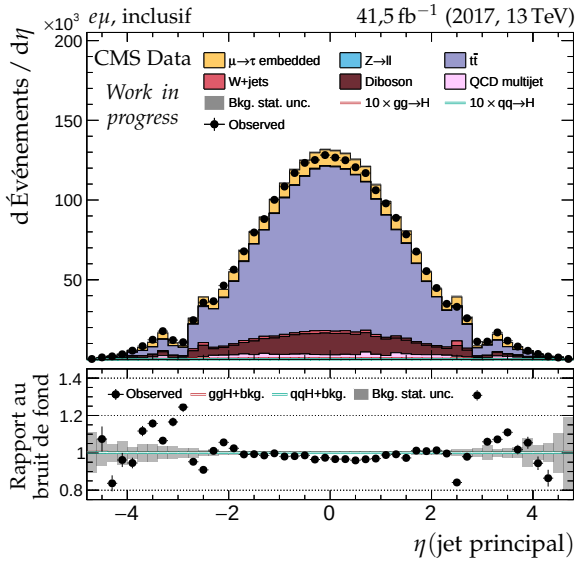
Figure G.47 – *Canal eμ, 2017 : masses transverses, D_ζ et masses.*



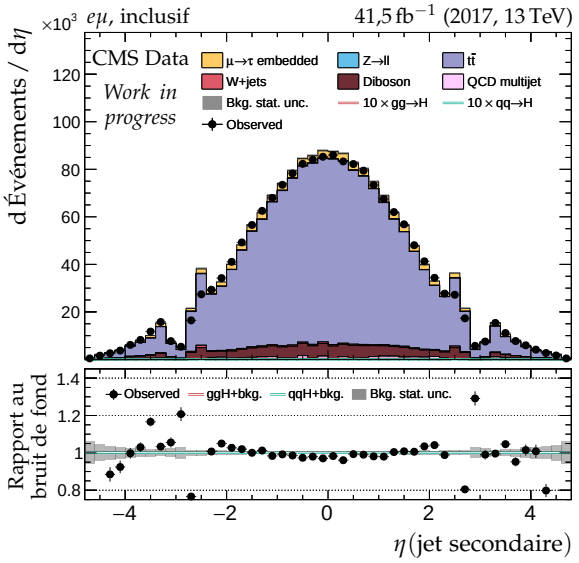
(a) *Impulsion transverse du jet principal.*



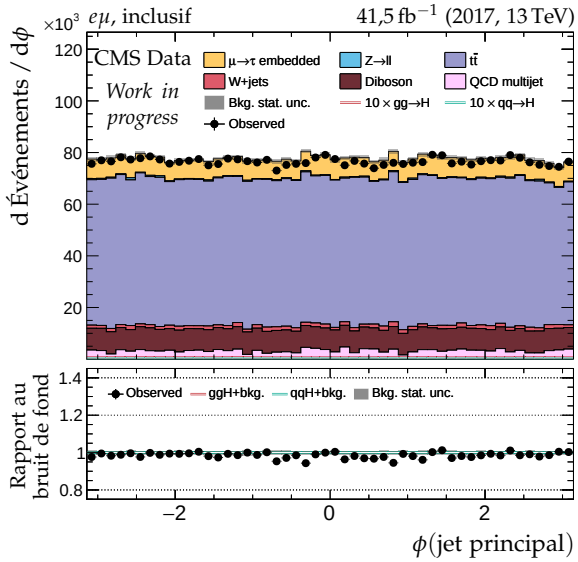
(b) *Impulsion transverse du jet secondaire.*



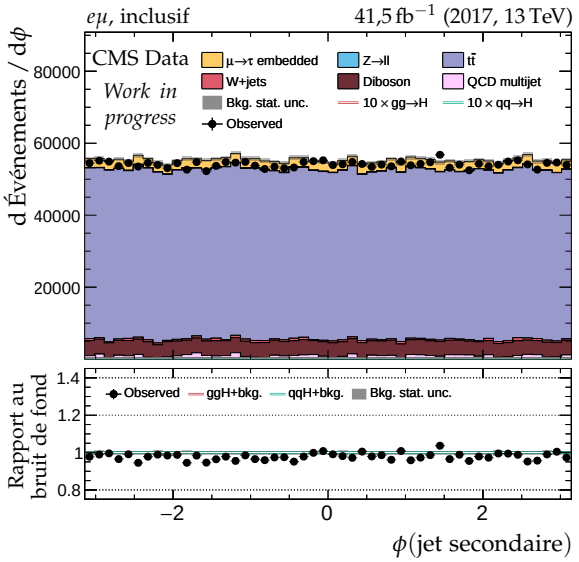
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

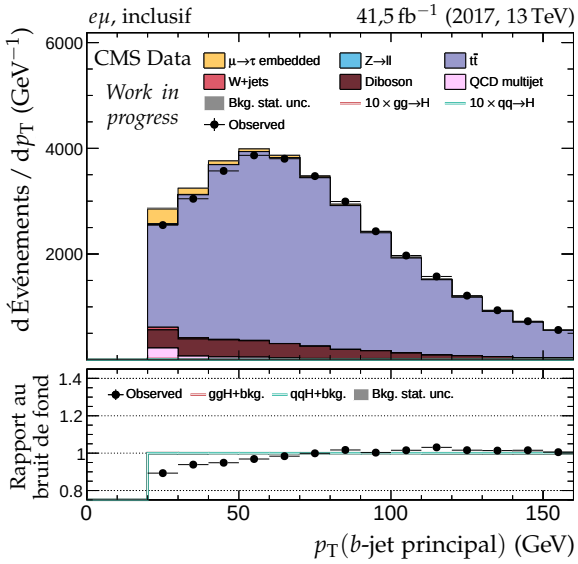


(e) *Angle azimutal du jet principal.*

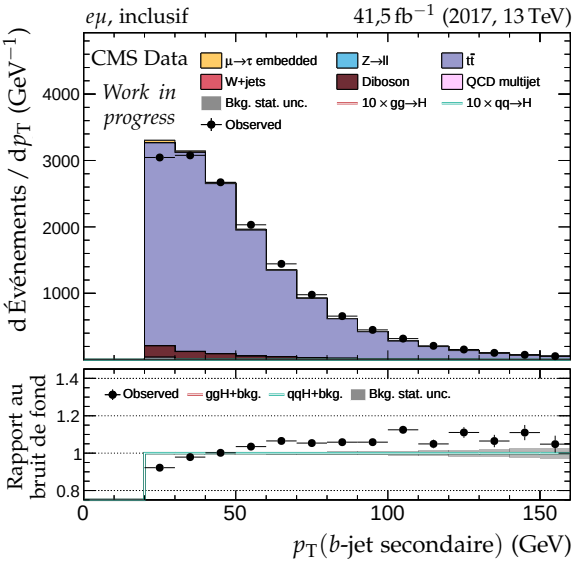


(f) *Angle azimutal du jet secondaire.*

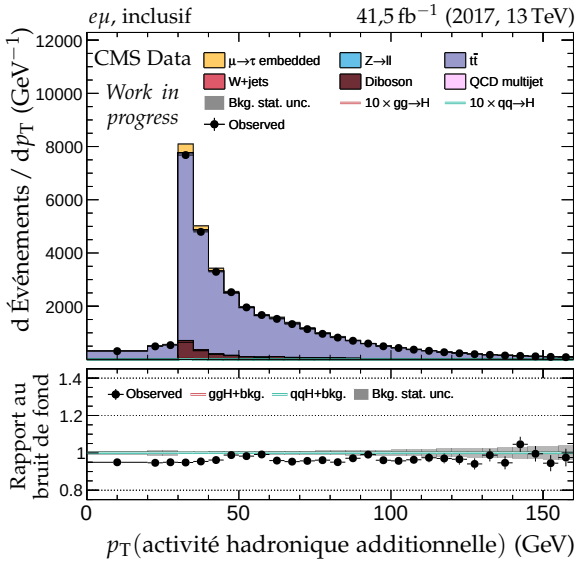
Figure G.48 – *Canal eμ, 2017 : cinématique des deux jets principaux.*



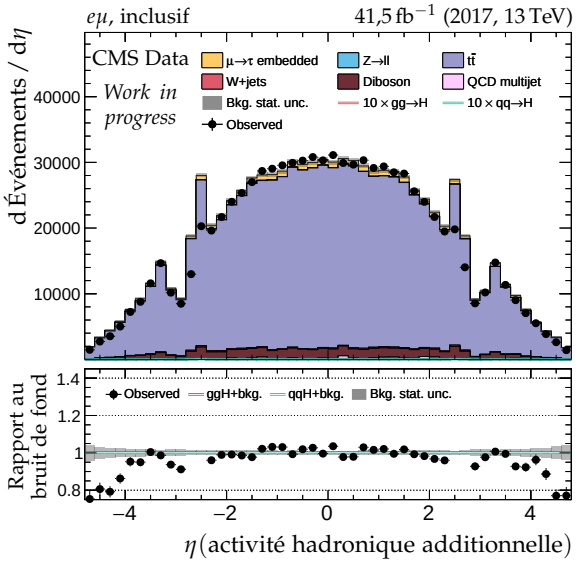
(a) *Impulsion transverse du b-jet principal.*



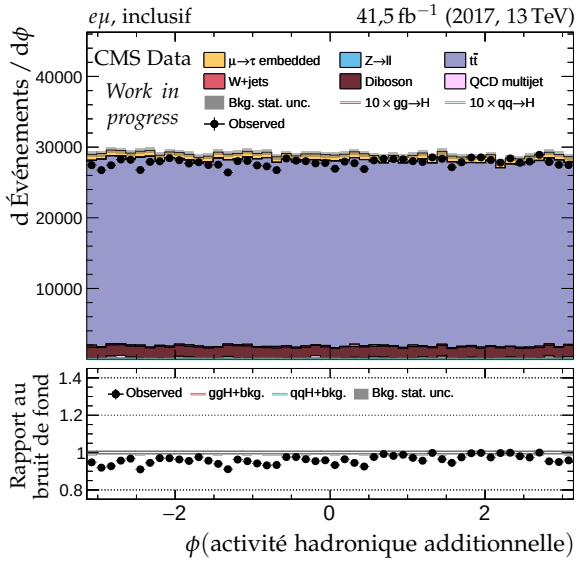
(b) *Impulsion transverse du b-jet secondaire.*



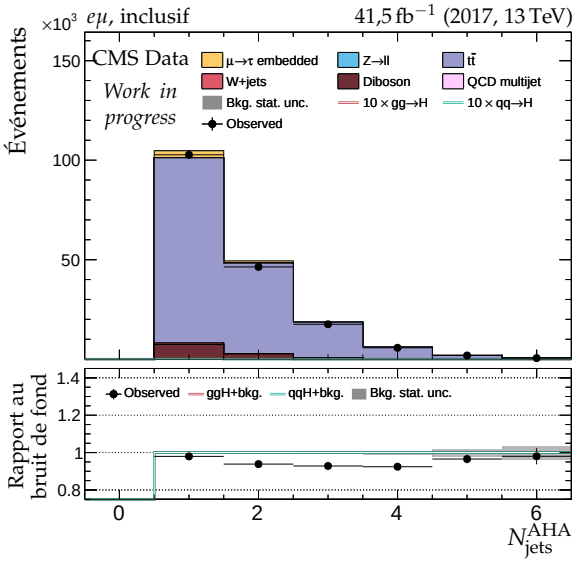
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

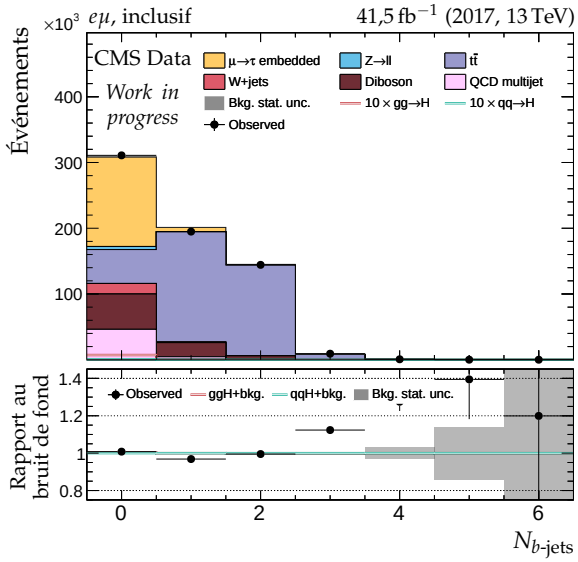


(e) *Angle azimutal de l'AHA.*

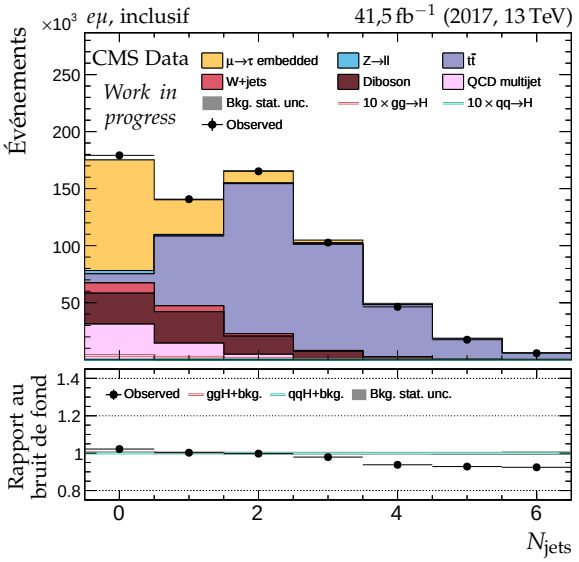


(f) *Nombre de jets dans l'AHA.*

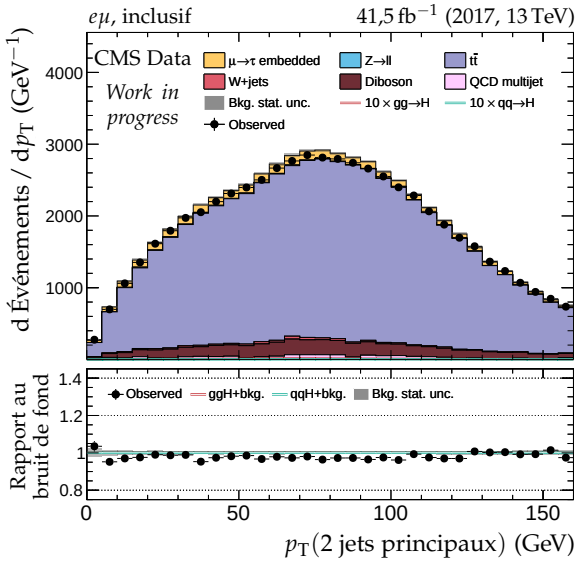
Figure G.49 – *Canal eμ, 2017 : b-jets et activité hadronique additionnelle.*



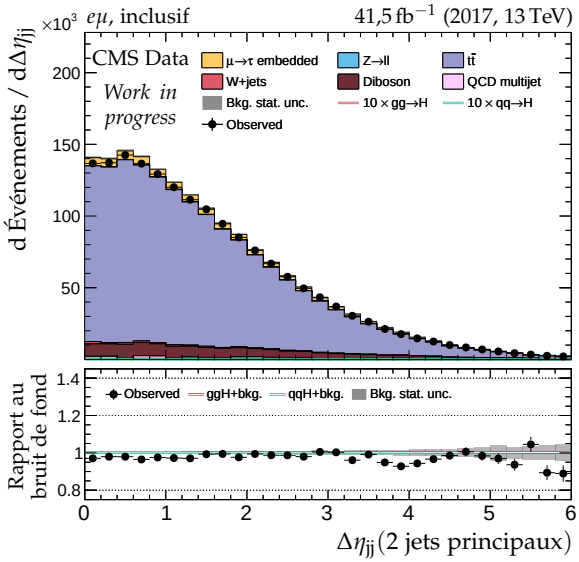
(a) *Nombre de b-jets.*



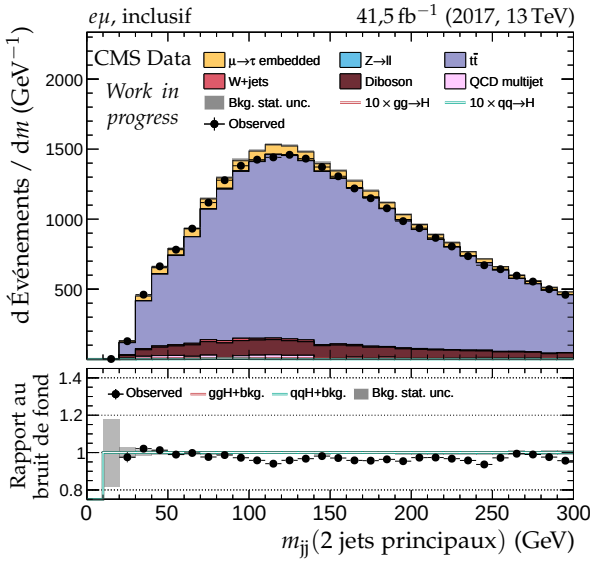
(b) *Nombre de jets.*



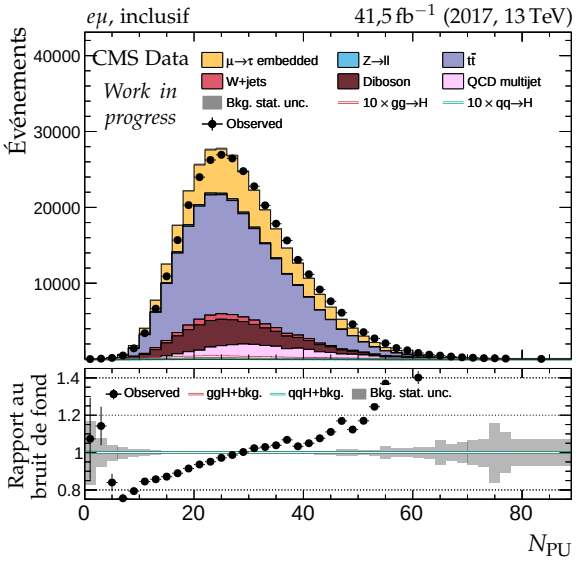
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

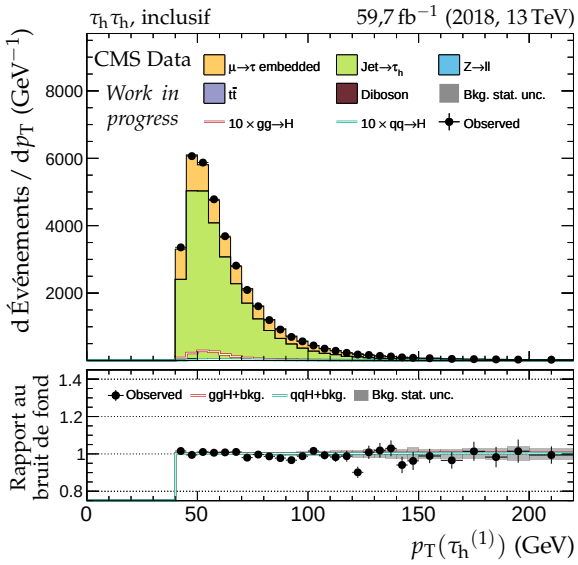


(e) *Masse invariante du système des deux jets.*

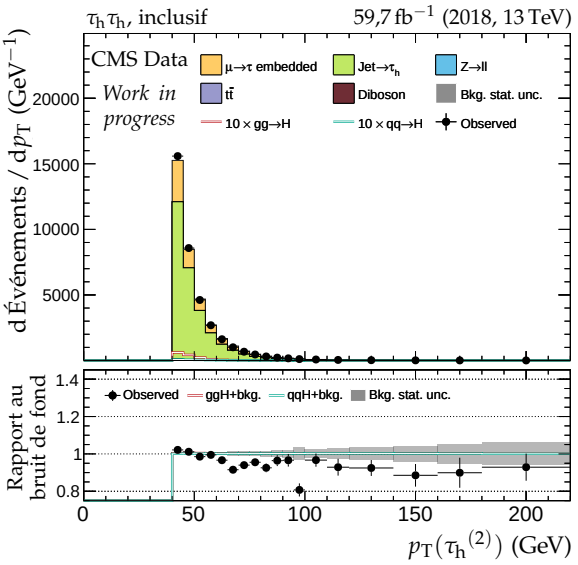


(f) *Nombre de vertex d'empilement.*

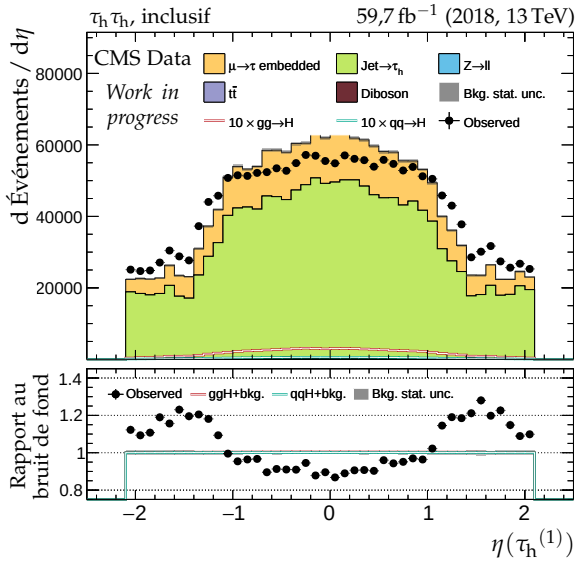
Figure G.50 – *Canal eμ, 2017 : nombre de jets, système des deux jets principaux et empilement.*



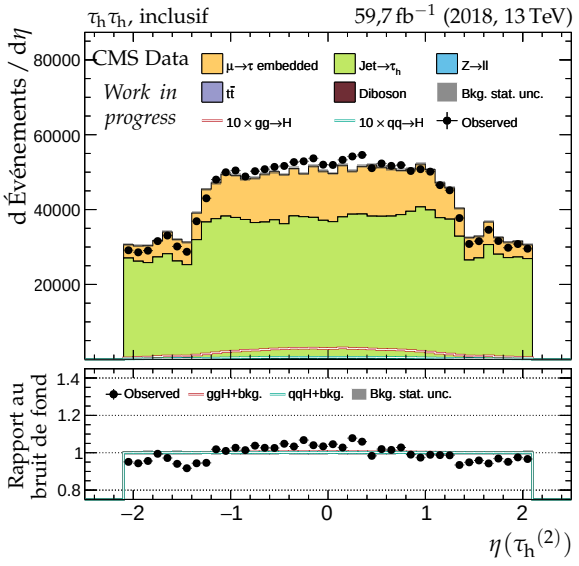
(a) *Impulsion transverse du lepton 1 ($\tau_h^{(1)}$).*



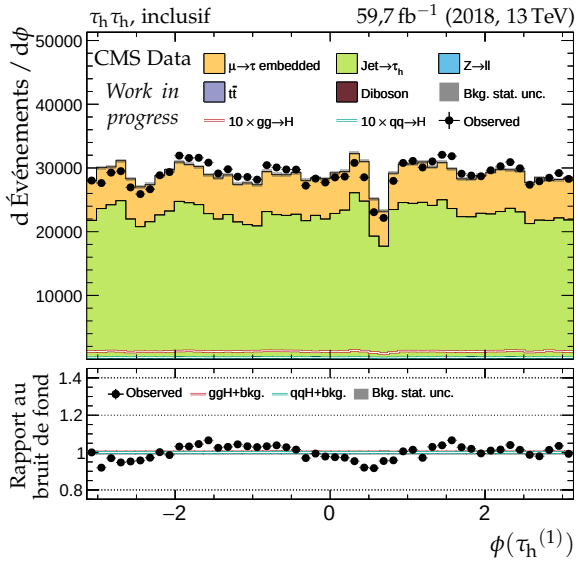
(b) *Impulsion transverse du lepton 2 ($\tau_h^{(2)}$).*



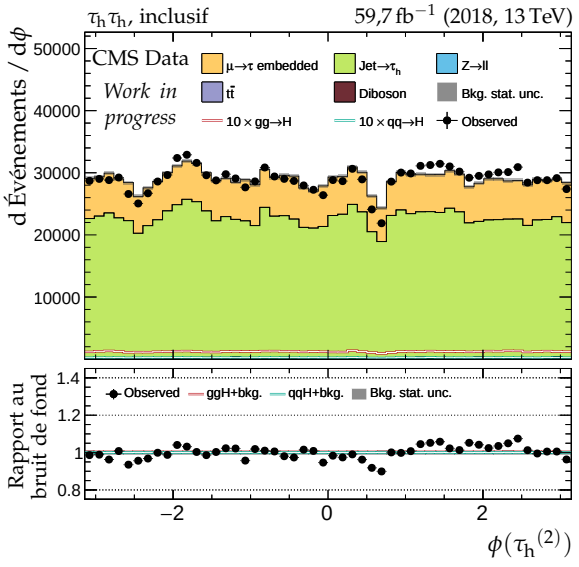
(c) *Pseudo-rapidité du lepton 1 ($\tau_h^{(1)}$).*



(d) *Pseudo-rapidité du lepton 2 ($\tau_h^{(2)}$).*

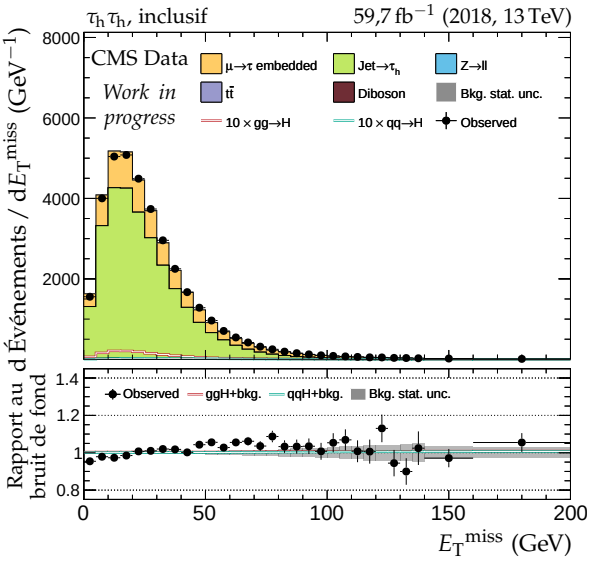


(e) *Angle azimutal du lepton 1 ($\tau_h^{(1)}$).*

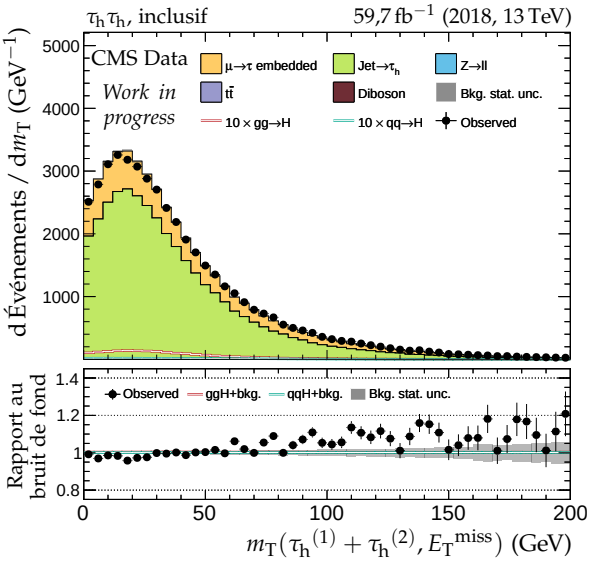


(f) *Angle azimutal du lepton 2 ($\tau_h^{(2)}$).*

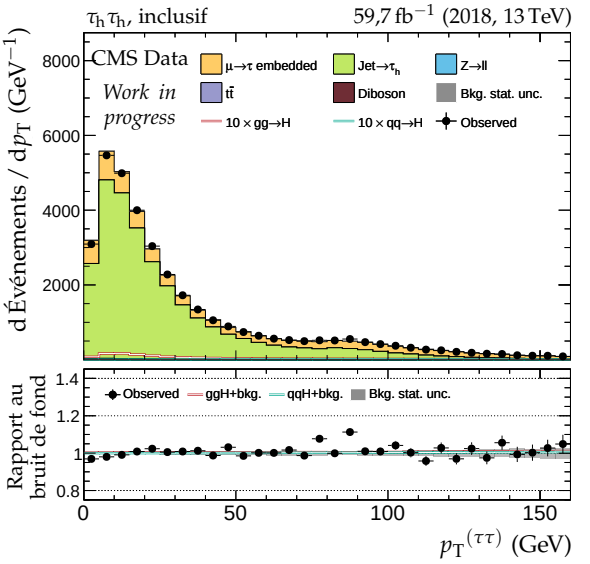
Figure G.51 – *Canal $\tau_h\tau_h$, 2018 : cinématique des leptons ($\tau_h^{(1)}$, $\tau_h^{(2)}$).*



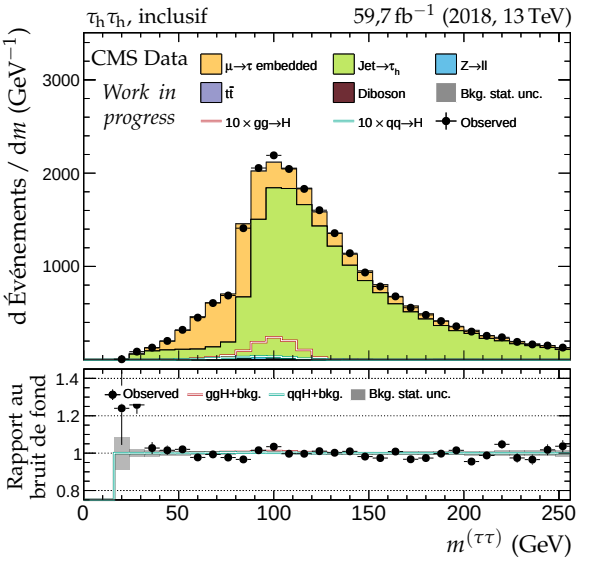
(a) *Énergie transverse manquante.*



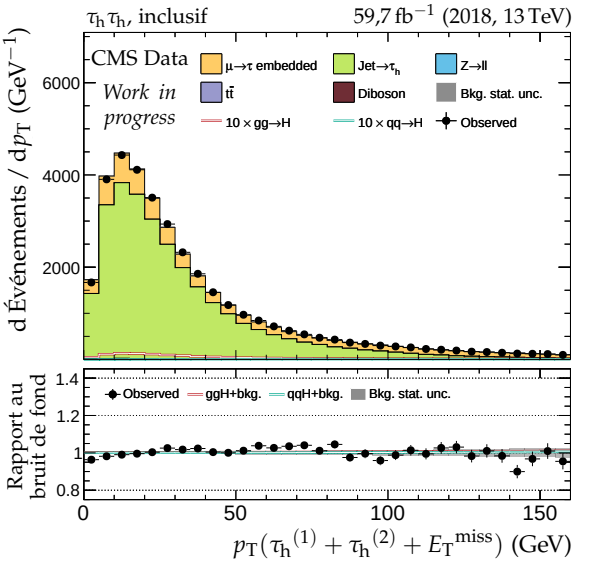
(b) *Masse transverse du dilepton.*



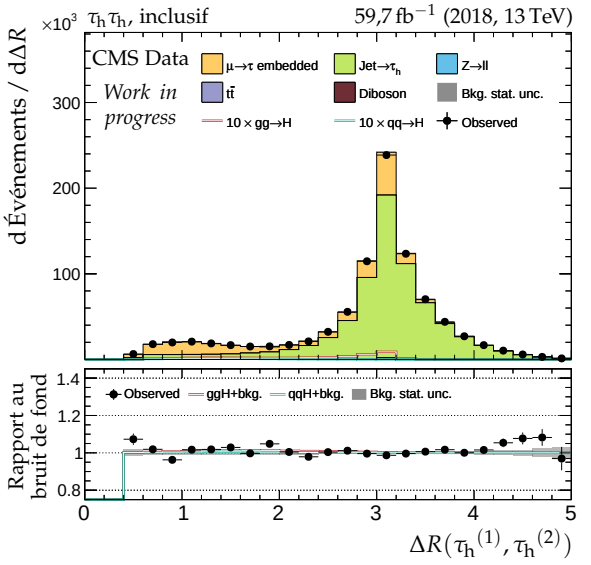
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

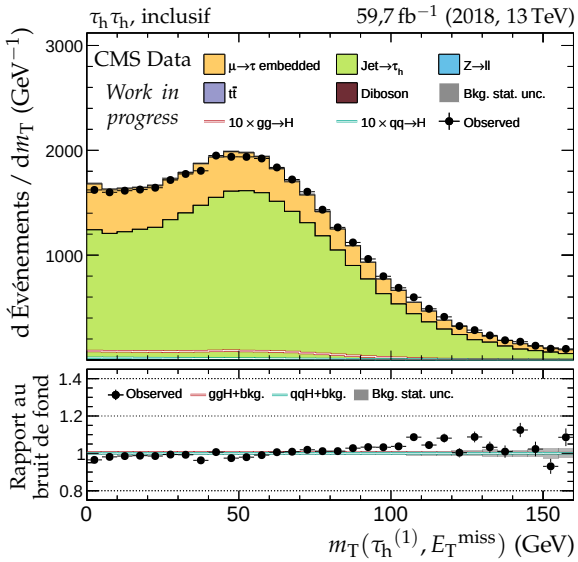


(e) *Impulsion transverse du système di-τ.*

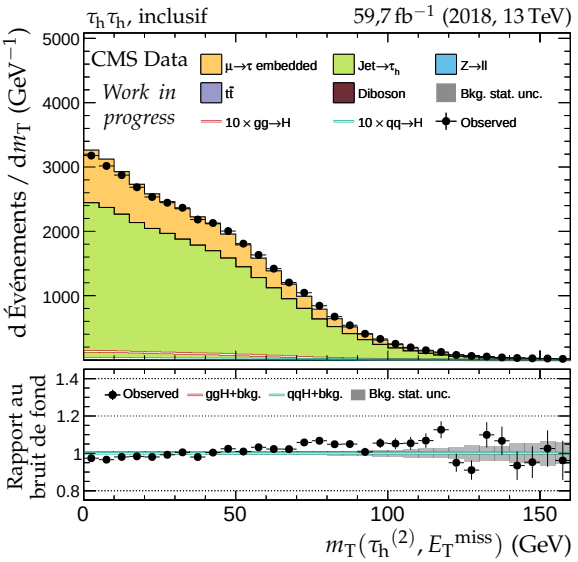


(f) *Distance ΔR entre les leptons ($\tau_h^{(1)}$, $\tau_h^{(2)}$).*

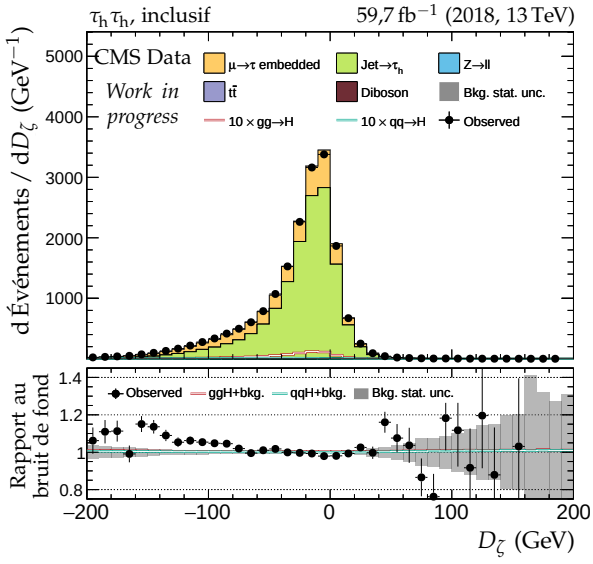
Figure G.52 – *Canal $\tau_h\tau_h$, 2018 : dilepton et énergie transverse manquante.*



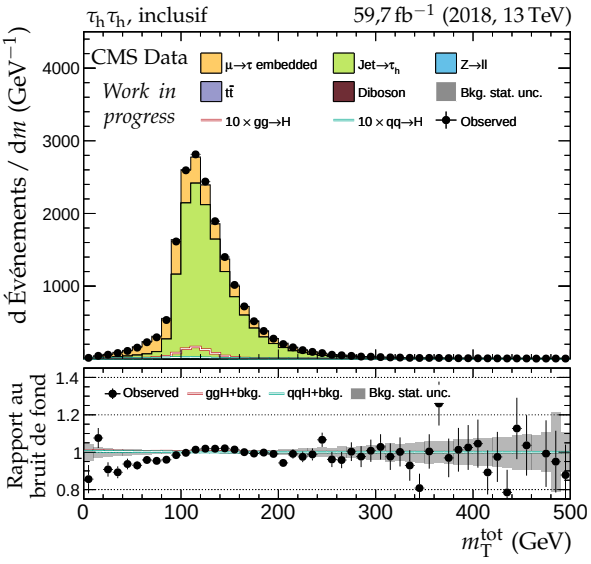
(a) *Masse transverse du lepton 1 ($\tau_h^{(1)}$).*



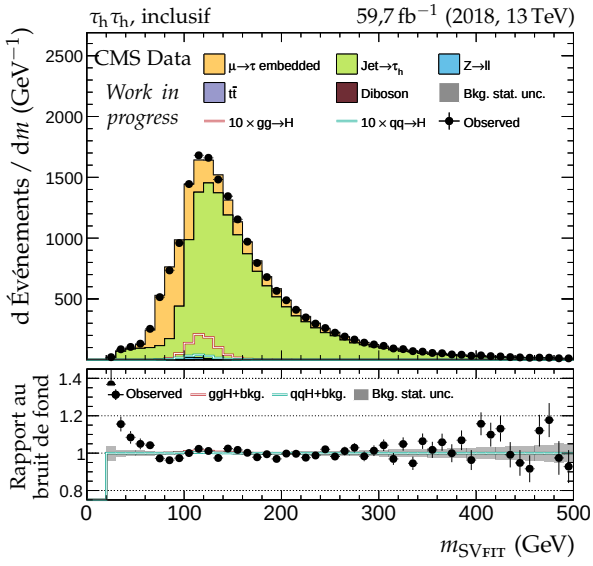
(b) *Masse transverse du lepton 2 ($\tau_h^{(2)}$).*



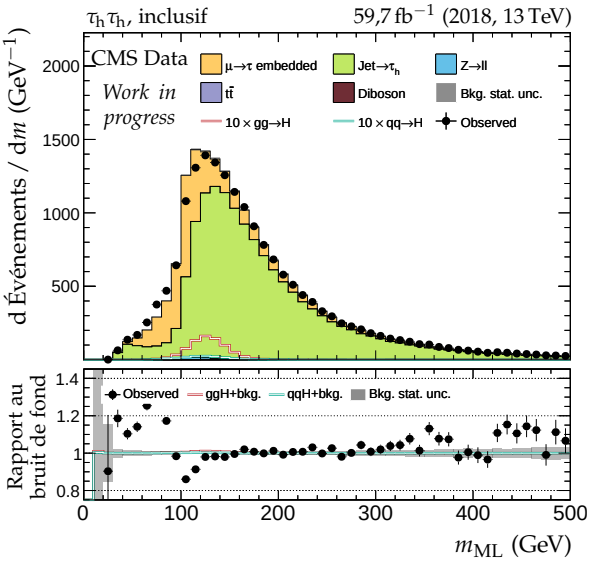
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

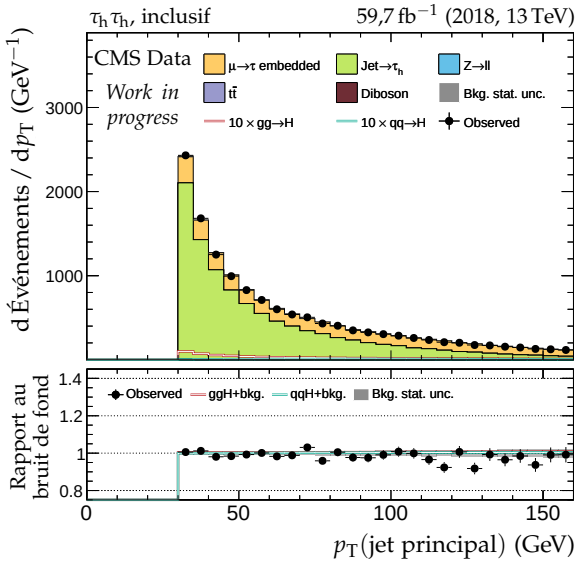


(e) *Masse du système di-τ d'après* SVFIT.

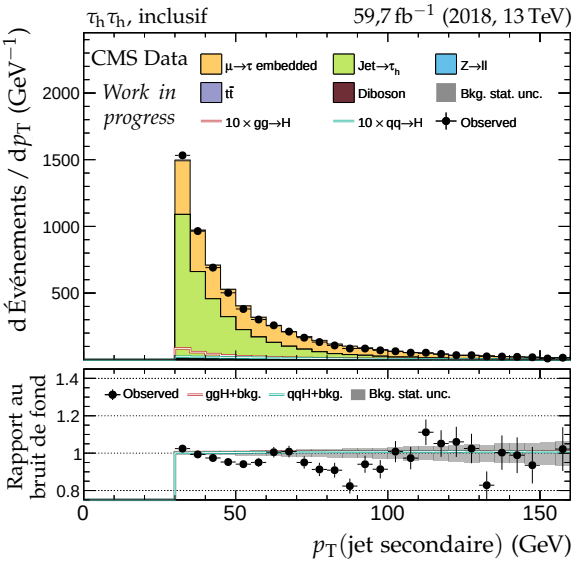


(f) *Masse du système di-τ d'après le ML.*

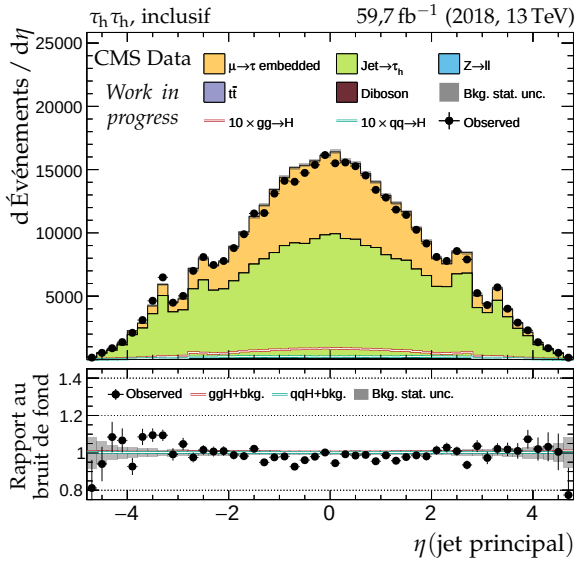
Figure G.53 – *Canal $\tau_h\tau_h$, 2018 : masses transverses, D_ζ et masses.*



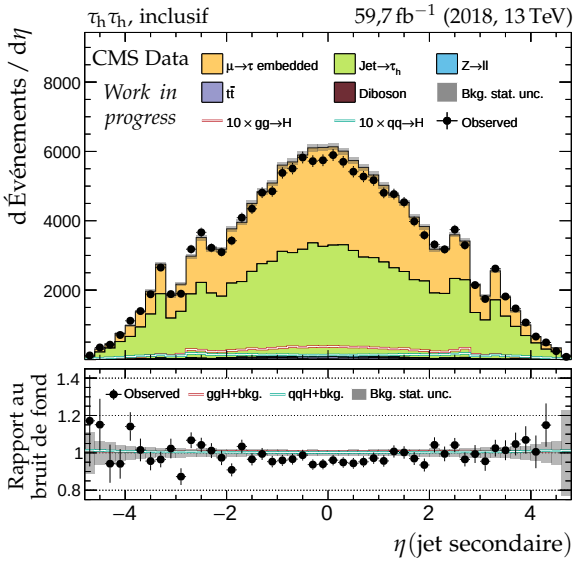
(a) *Impulsion transverse du jet principal.*



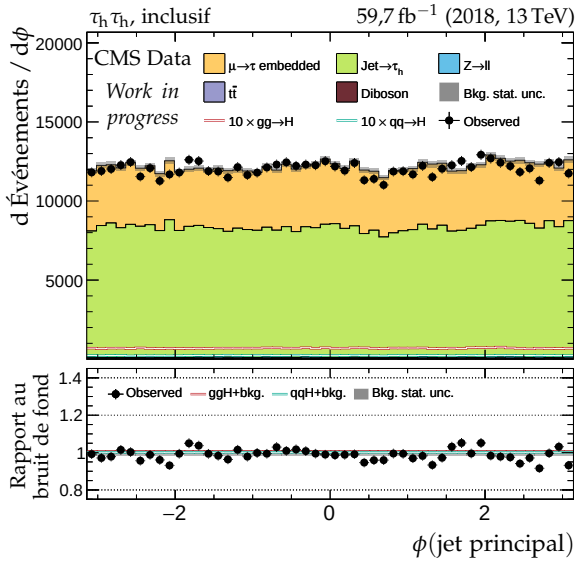
(b) *Impulsion transverse du jet secondaire.*



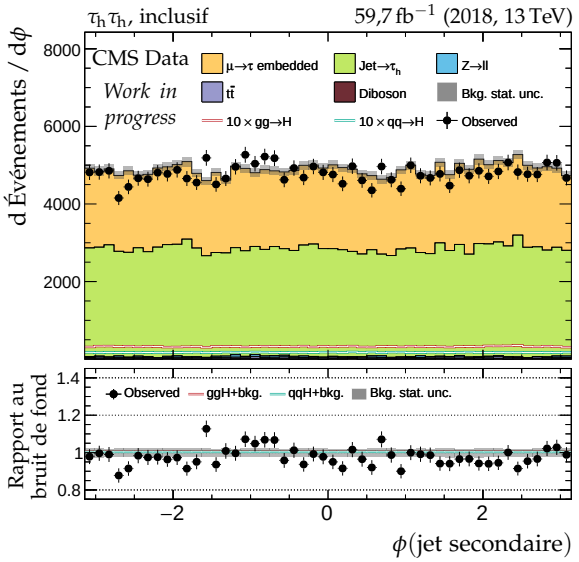
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

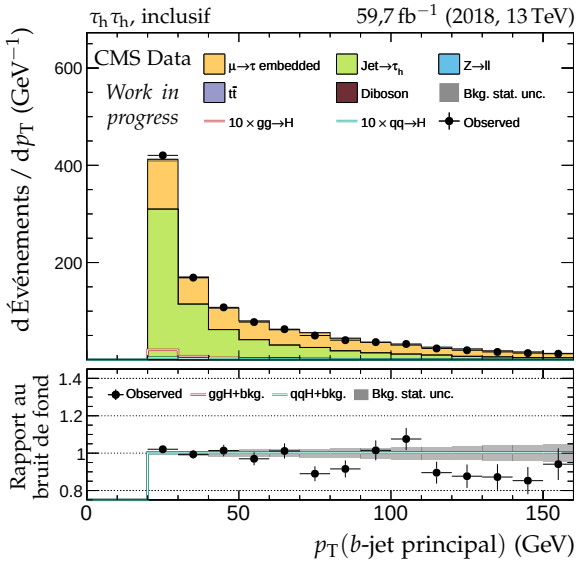


(e) *Angle azimutal du jet principal.*

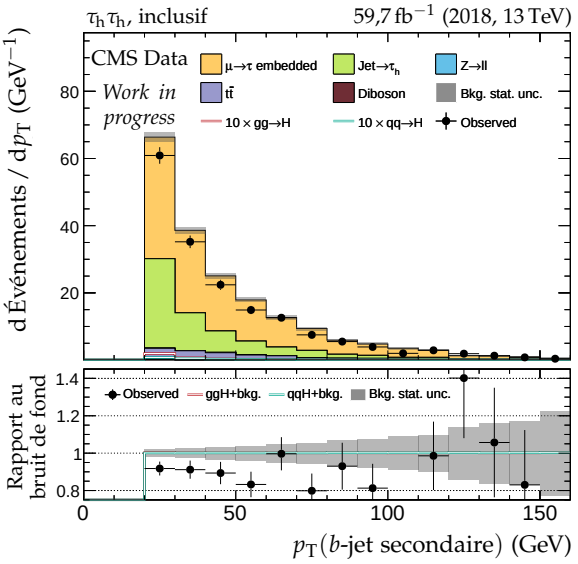


(f) *Angle azimutal du jet secondaire.*

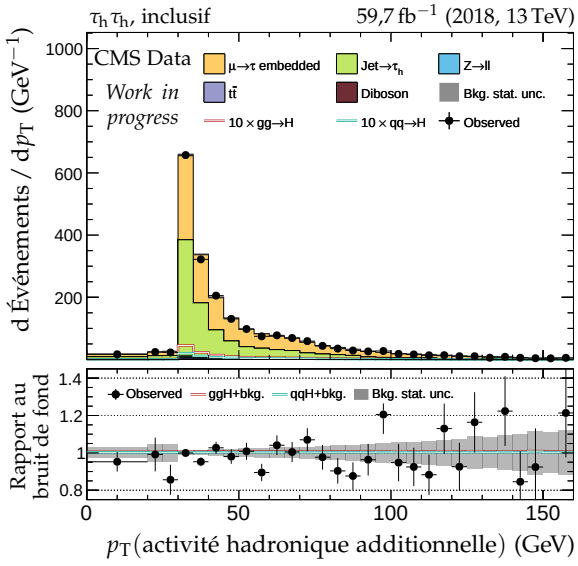
Figure G.54 – *Canal $\tau_h\tau_h$, 2018 : cinématique des deux jets principaux.*



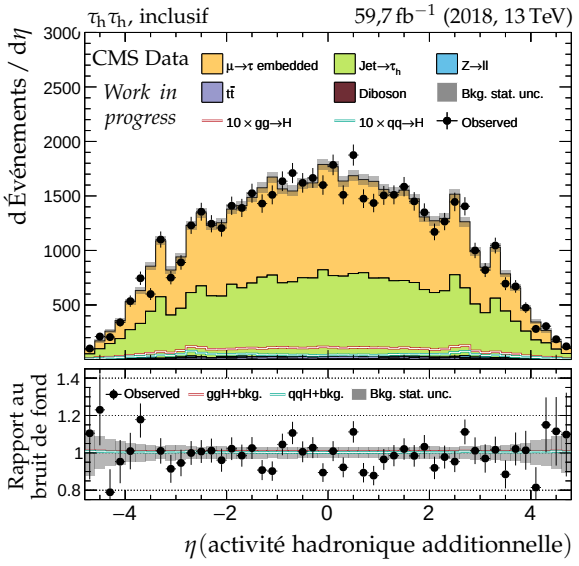
(a) *Impulsion transverse du b-jet principal.*



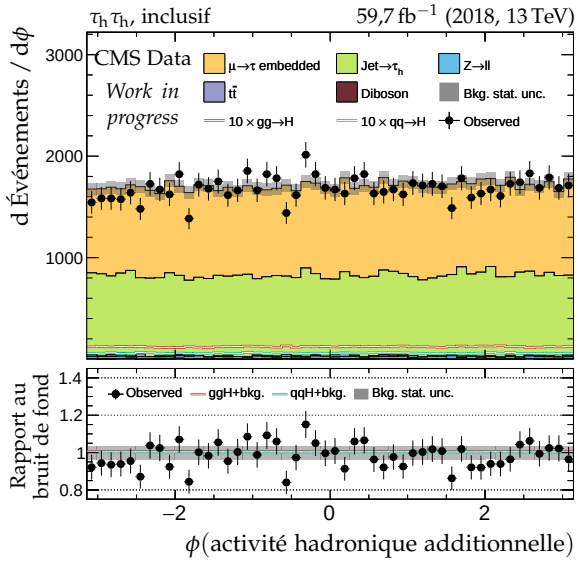
(b) *Impulsion transverse du b-jet secondaire.*



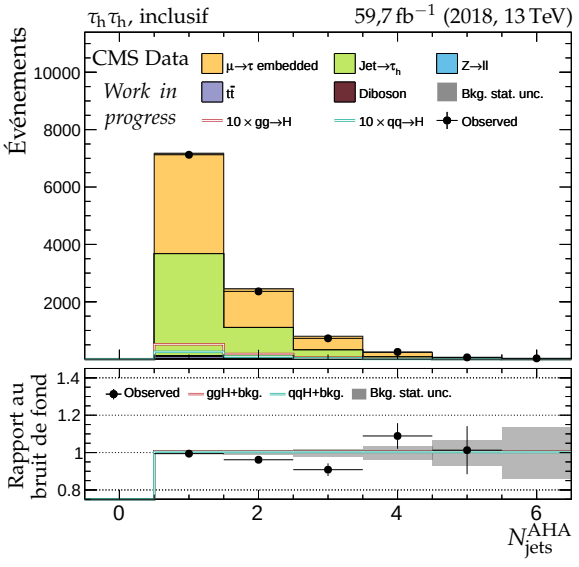
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

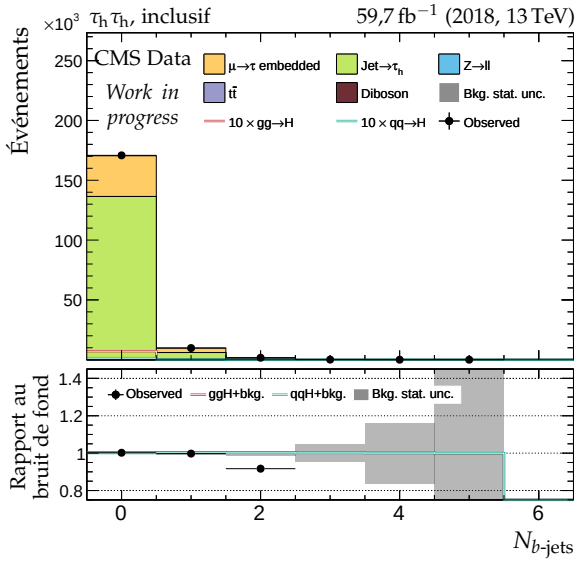


(e) *Angle azimutal de l'AHA.*

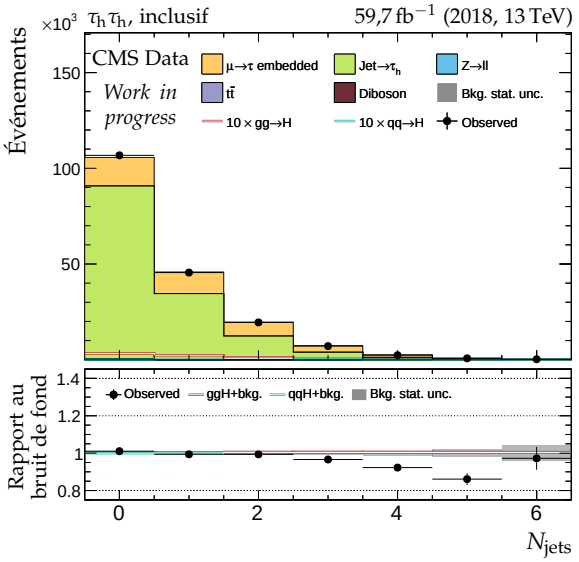


(f) *Nombre de jets dans l'AHA.*

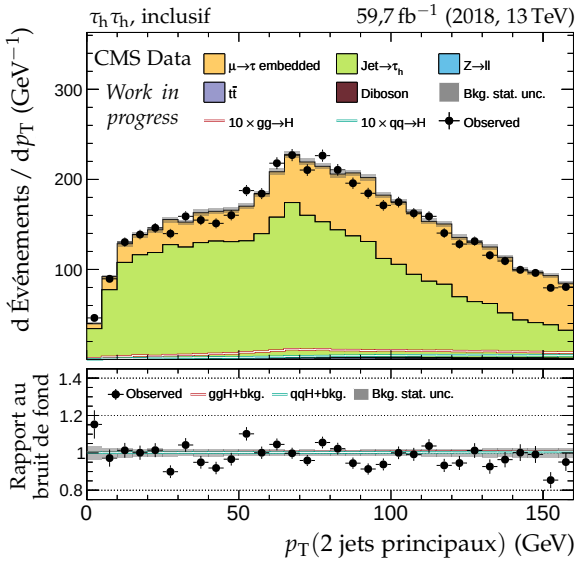
Figure G.55 – *Canal $\tau_h\tau_h$, 2018 : b-jets et activité hadronique additionnelle.*



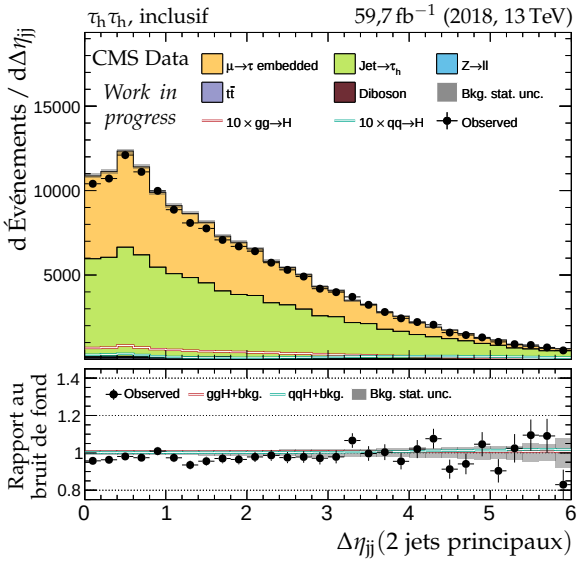
(a) *Nombre de b-jets.*



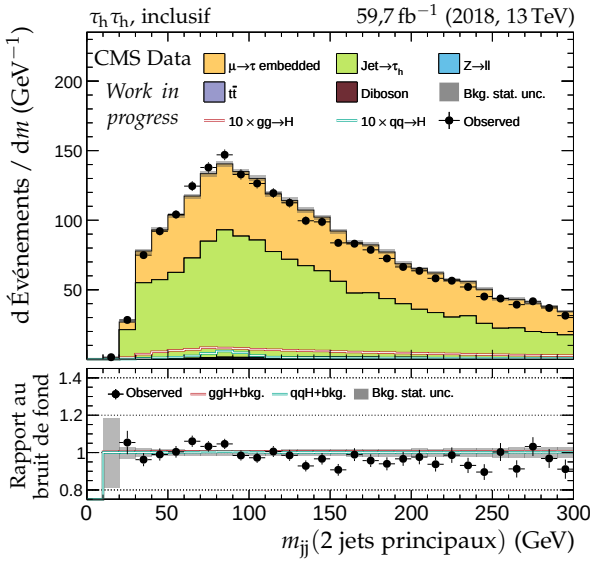
(b) *Nombre de jets.*



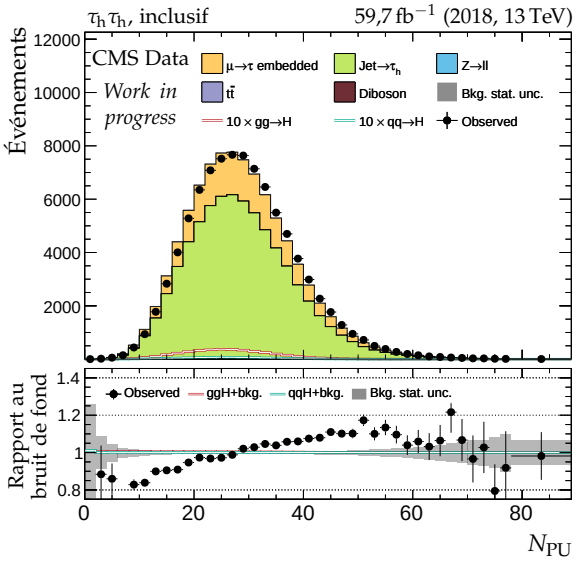
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

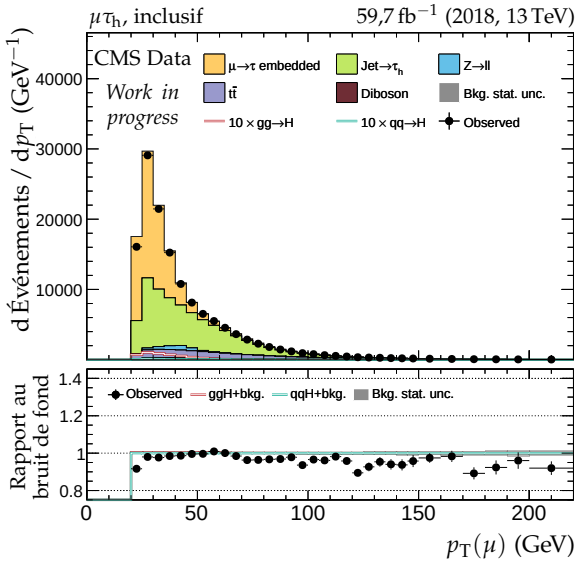


(e) *Masse invariante du système des deux jets.*

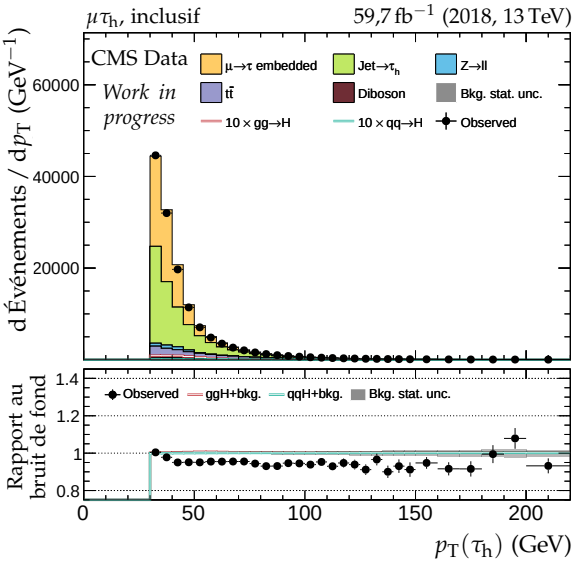


(f) *Nombre de vertex d'empilement.*

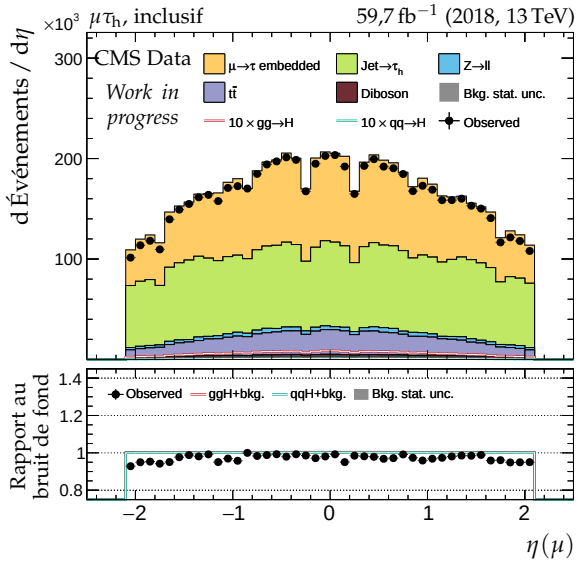
Figure G.56 – *Canal $\tau_h\tau_h$, 2018 : nombre de jets, système des deux jets principaux et empilement.*



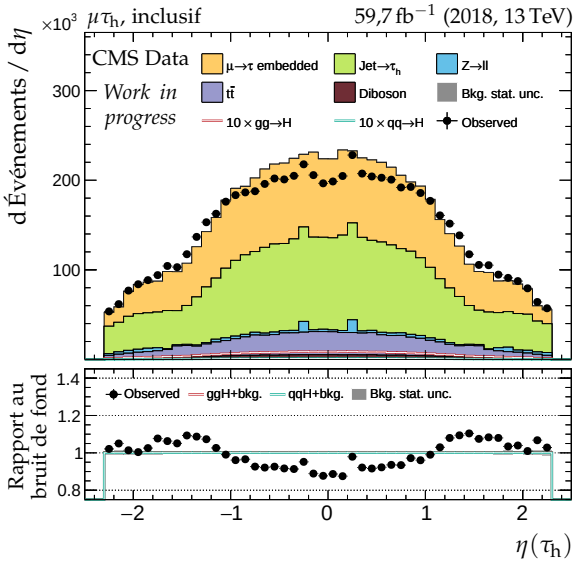
(a) *Impulsion transverse du lepton 1 (μ).*



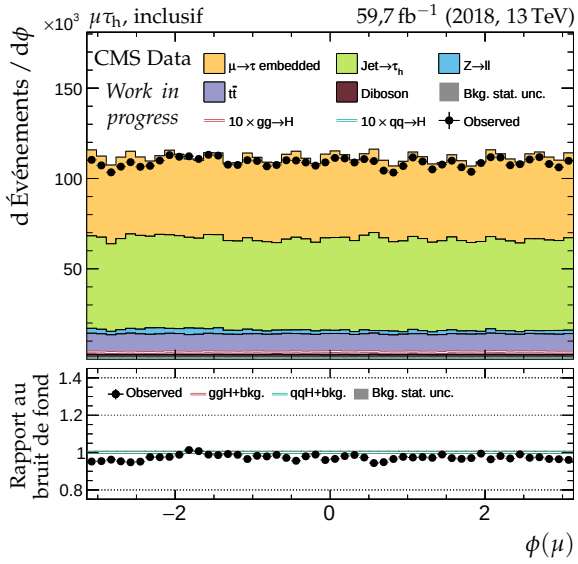
(b) *Impulsion transverse du lepton 2 (τ_h).*



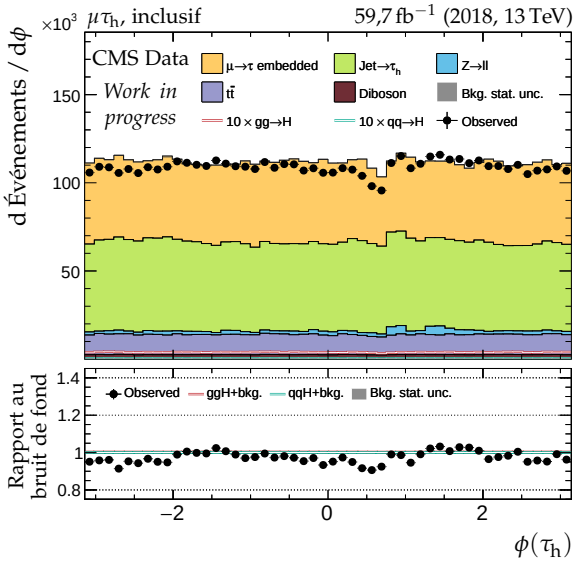
(c) *Pseudo-rapidité du lepton 1 (μ).*



(d) *Pseudo-rapidité du lepton 2 (τ_h).*

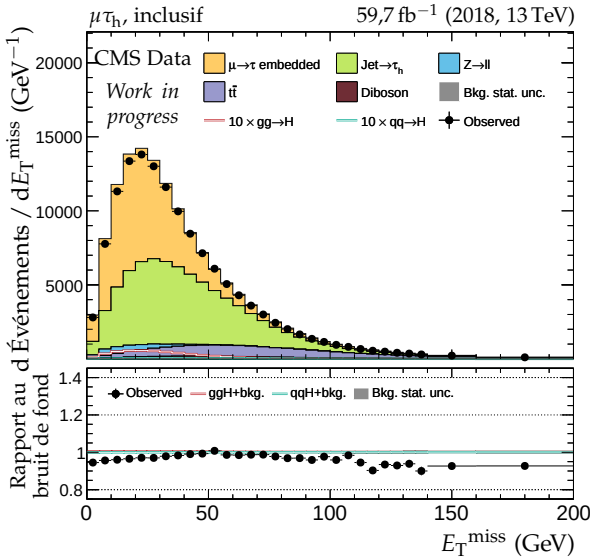


(e) *Angle azimutal du lepton 1 (μ).*

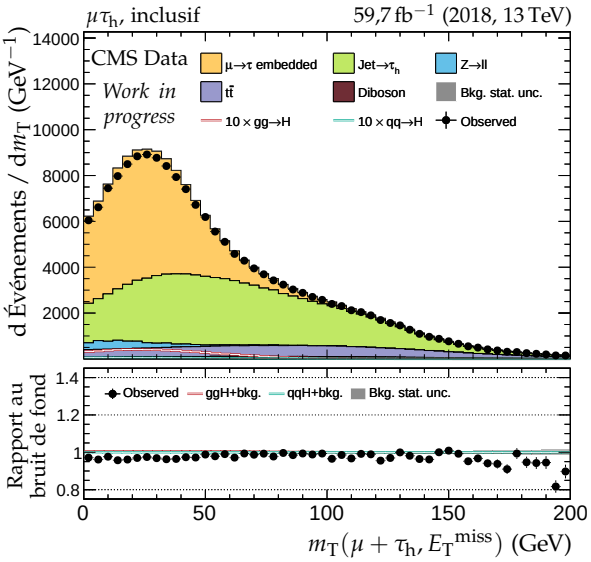


(f) *Angle azimutal du lepton 2 (τ_h).*

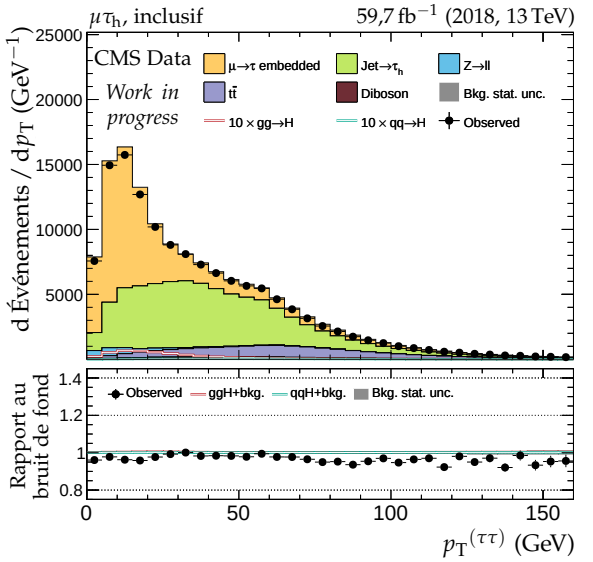
Figure G.57 – *Canal $\mu\tau_h$, 2018 : cinématique des leptons (μ, τ_h).*



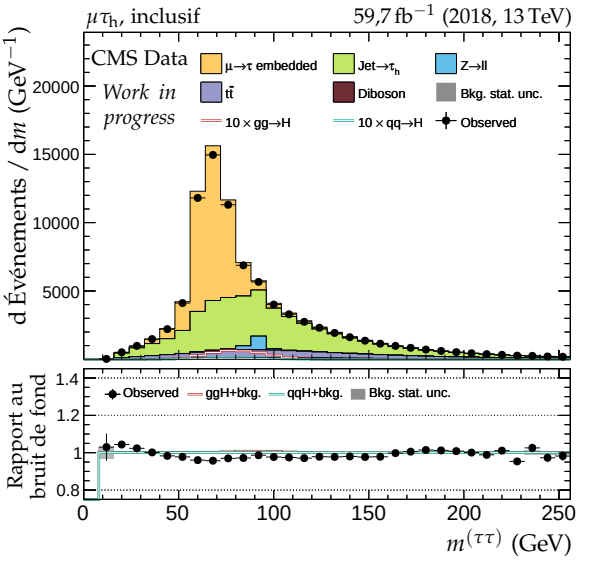
(a) *Énergie transverse manquante.*



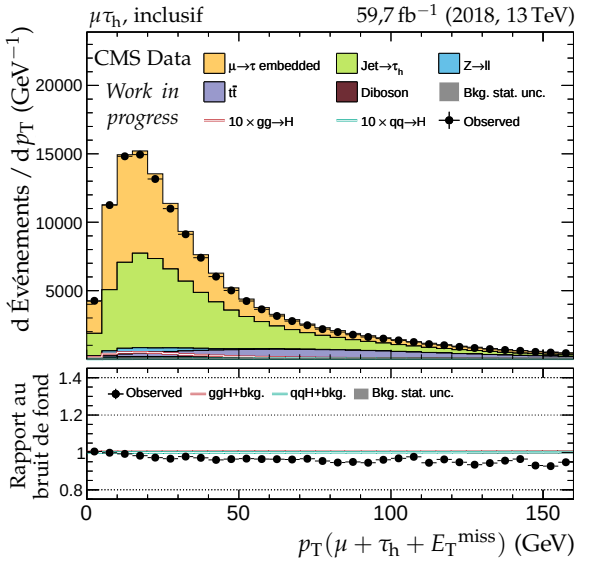
(b) *Masse transverse du dilepton.*



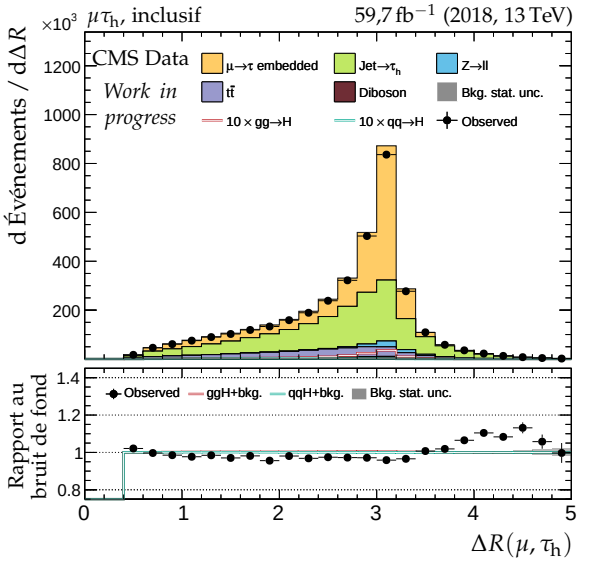
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

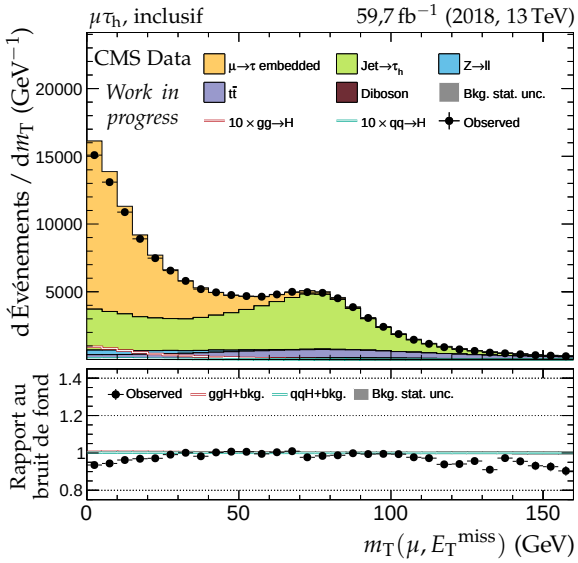


(e) *Impulsion transverse du système di-τ.*

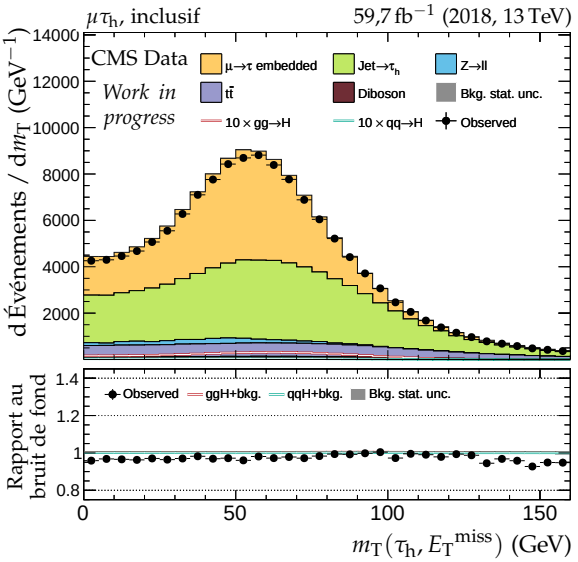


(f) *Distance ΔR entre les leptons (μ, τ_h).*

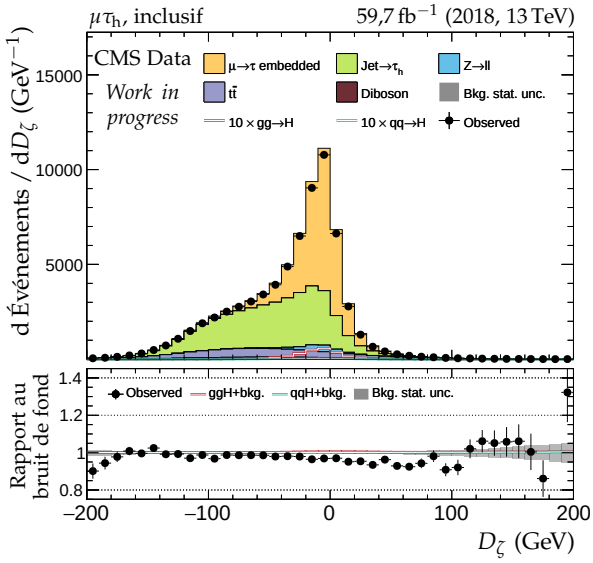
Figure G.58 – *Canal $\mu\tau_h$, 2018 : dilepton et énergie transverse manquante.*



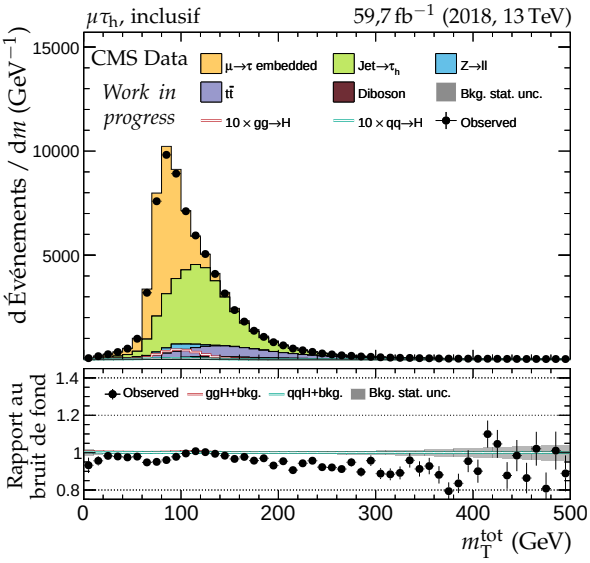
(a) *Masse transverse du lepton 1 (μ).*



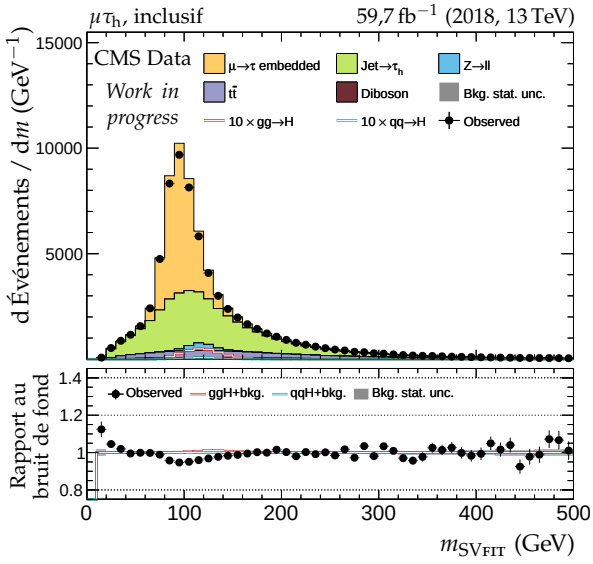
(b) *Masse transverse du lepton 2 (τ_h).*



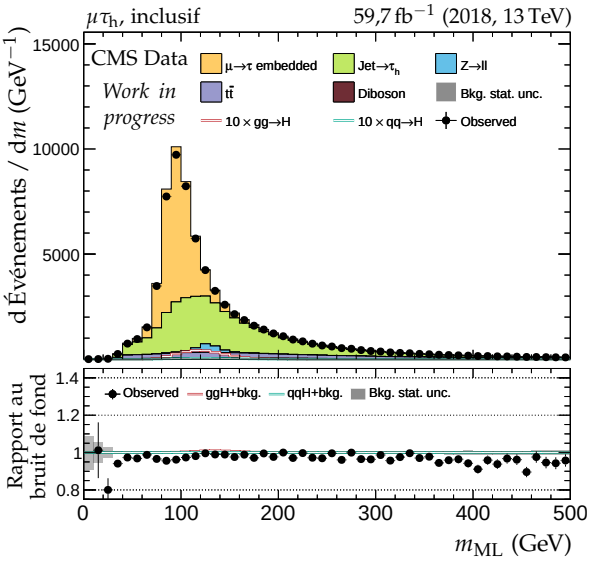
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

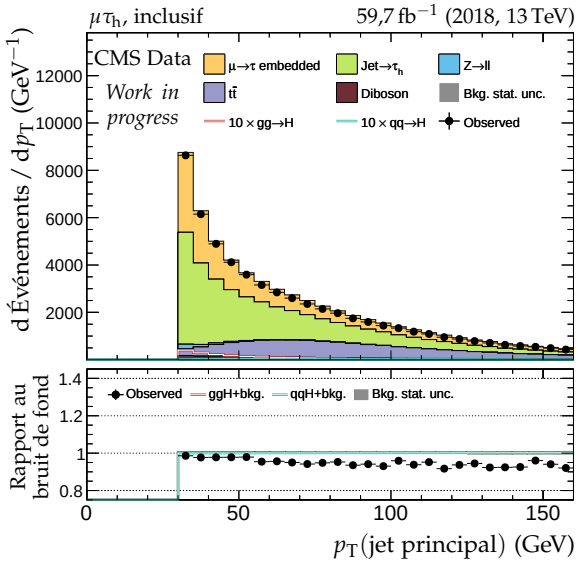


(e) *Masse du système di-τ d'après* SVFIT.

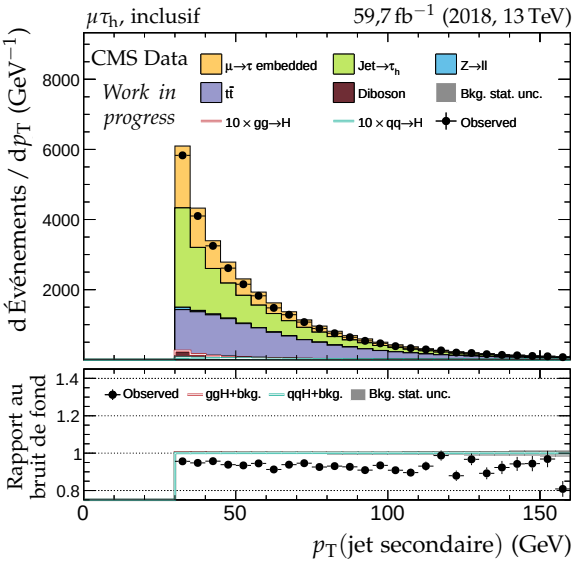


(f) *Masse du système di-τ d'après le ML.*

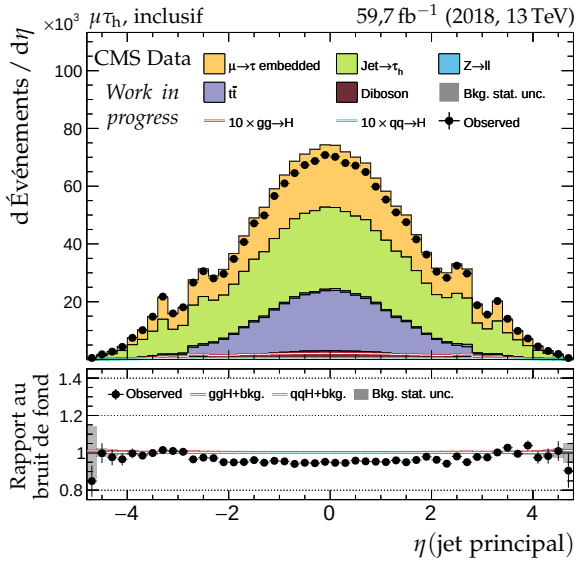
Figure G.59 – *Canal $\mu\tau_h$, 2018 : masses transverses, D_ζ et masses.*



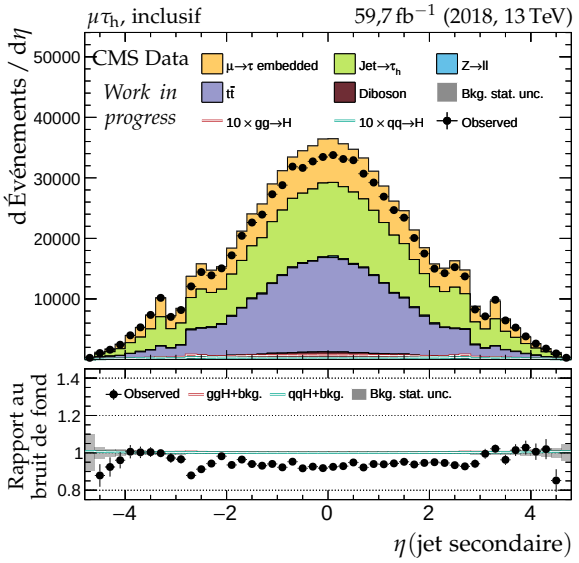
(a) *Impulsion transverse du jet principal.*



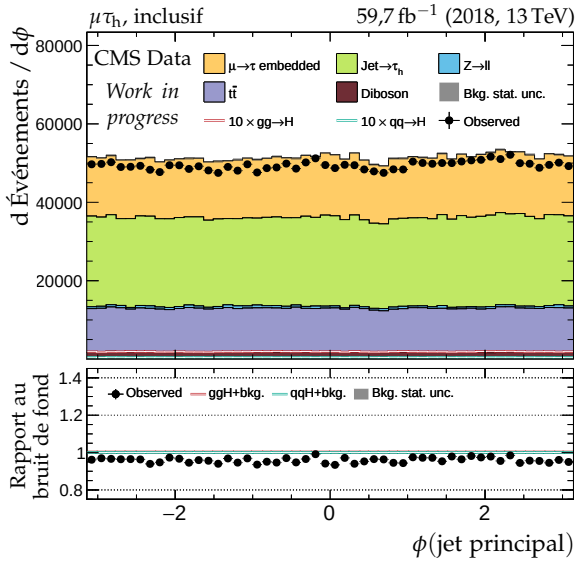
(b) *Impulsion transverse du jet secondaire.*



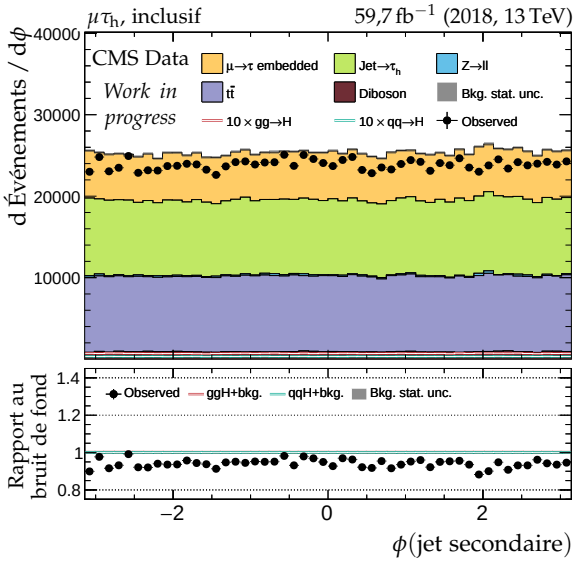
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

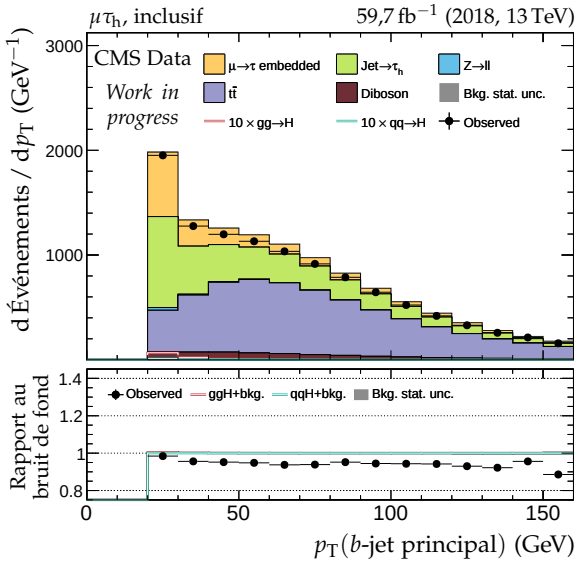


(e) *Angle azimutal du jet principal.*

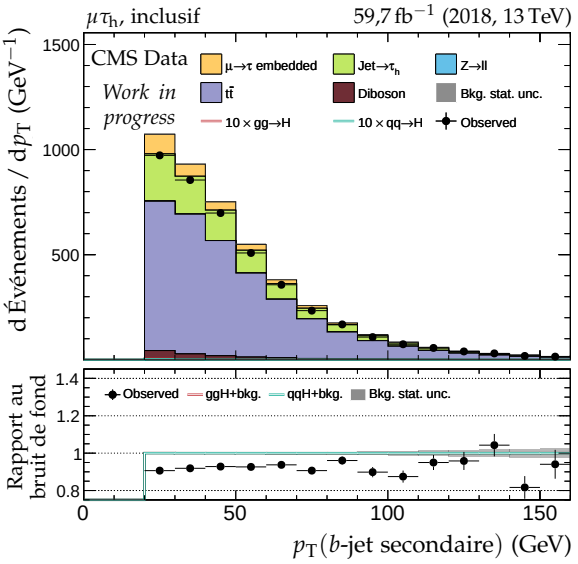


(f) *Angle azimutal du jet secondaire.*

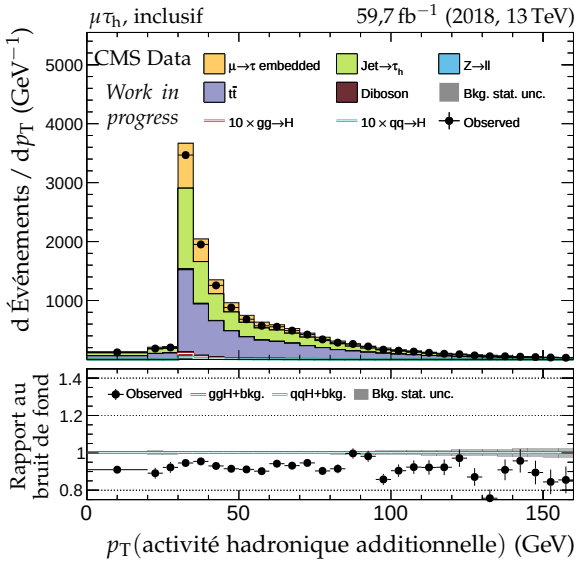
Figure G.60 – *Canal $\mu\tau_h$, 2018 : cinématique des deux jets principaux.*



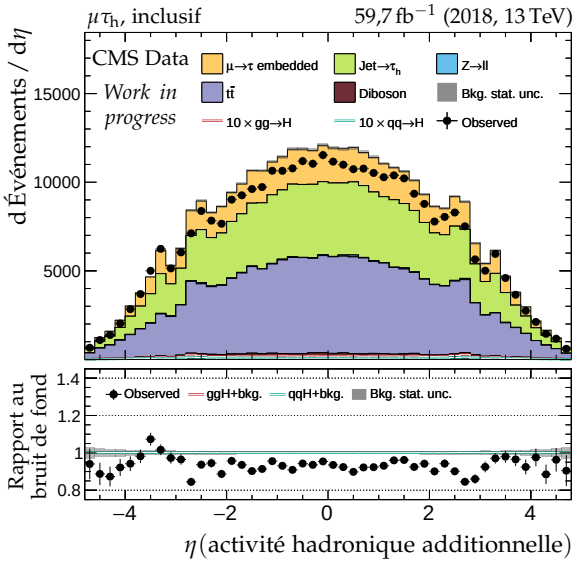
(a) *Impulsion transverse du b-jet principal.*



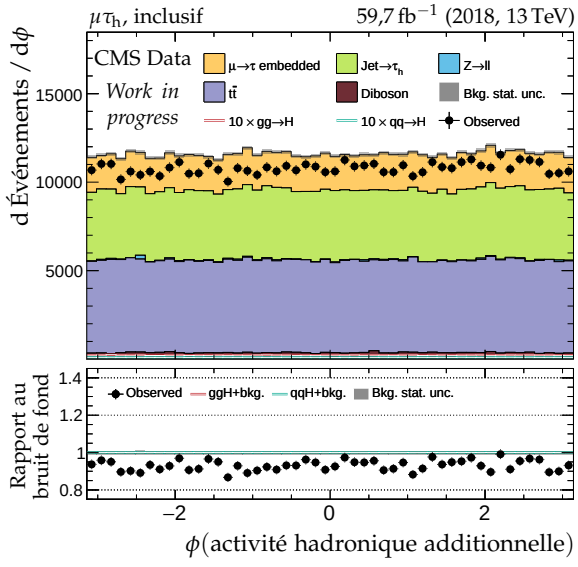
(b) *Impulsion transverse du b-jet secondaire.*



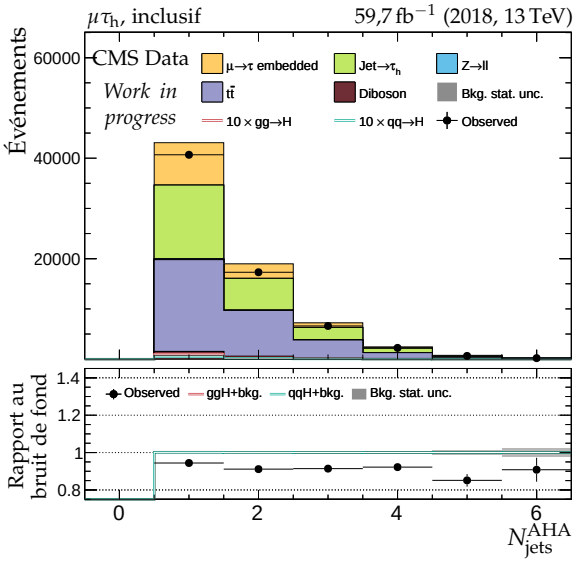
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

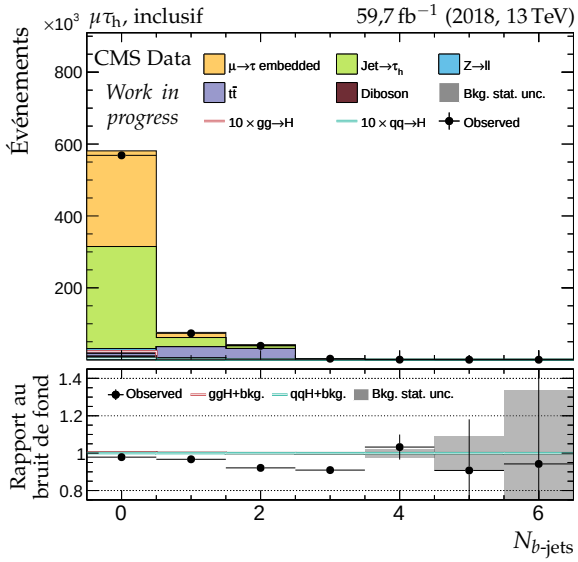


(e) *Angle azimutal de l'AHA.*

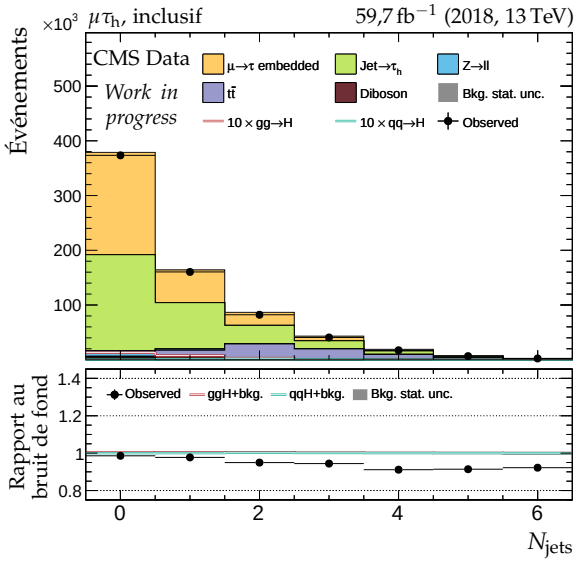


(f) *Nombre de jets dans l'AHA.*

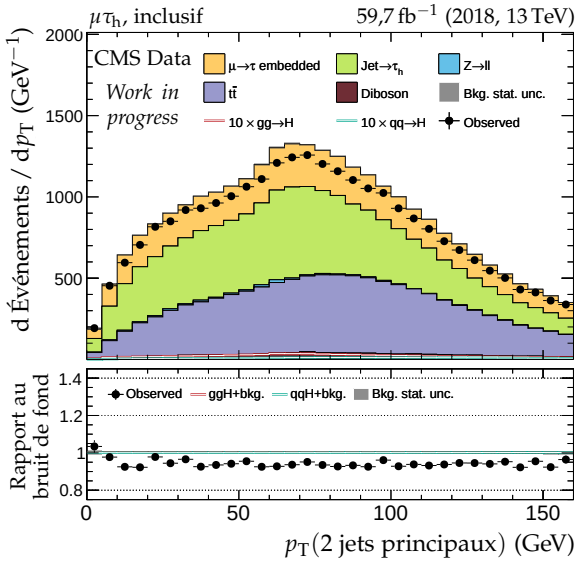
Figure G.61 – *Canal $\mu\tau_h$, 2018 : b-jets et activité hadronique additionnelle.*



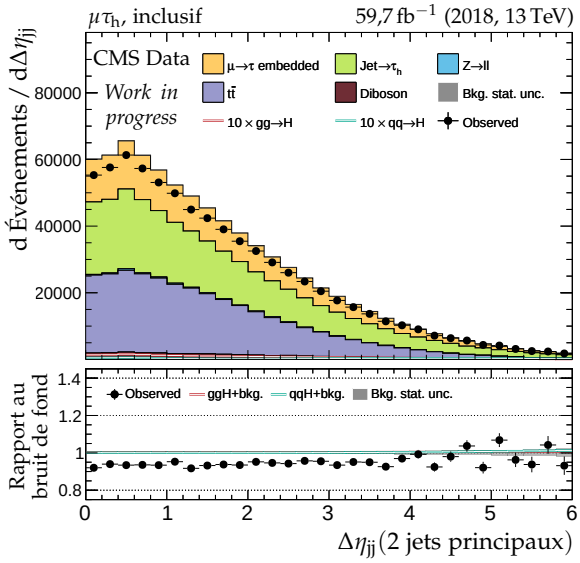
(a) *Nombre de b-jets.*



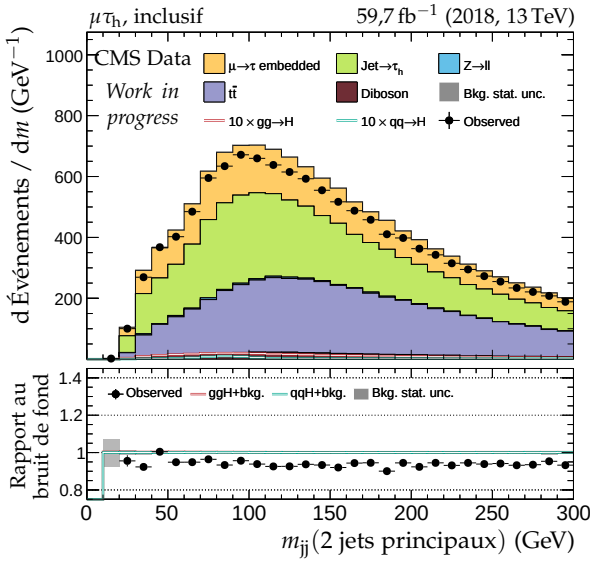
(b) *Nombre de jets.*



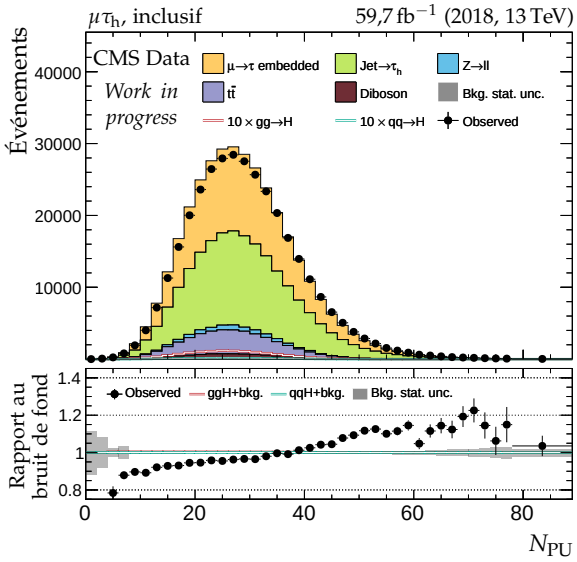
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

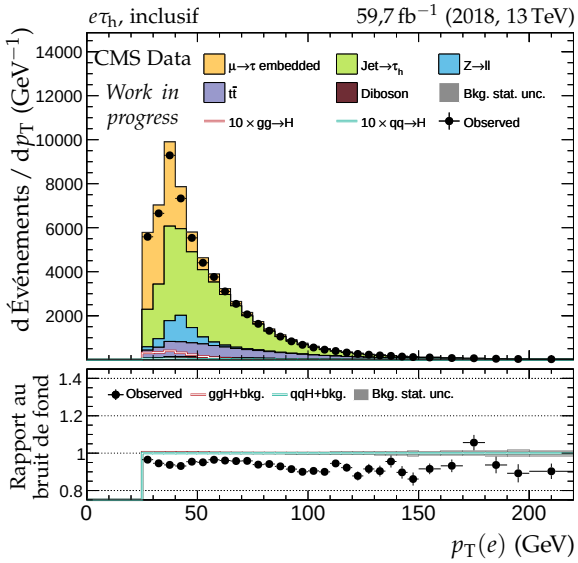


(e) *Masse invariante du système des deux jets.*

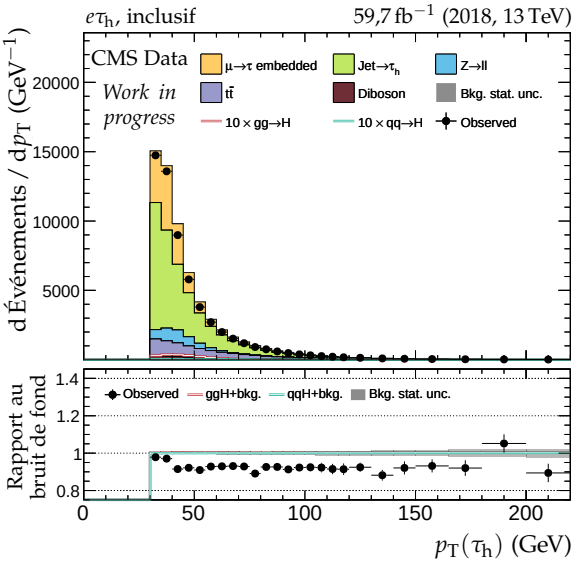


(f) *Nombre de vertex d'empilement.*

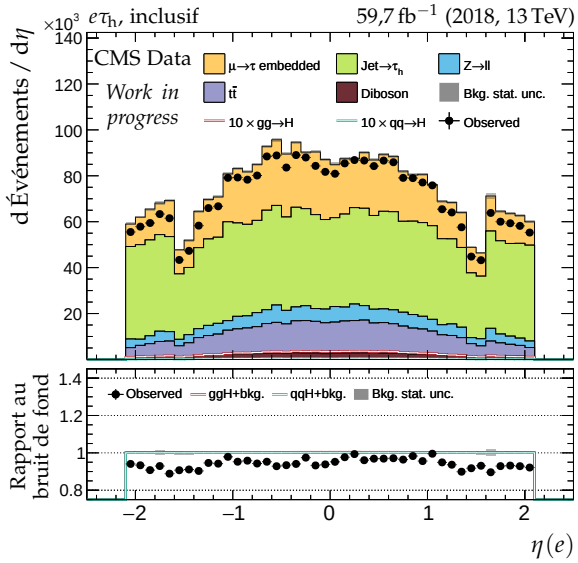
Figure G.62 – *Canal $\mu\tau_h$, 2018 : nombre de jets, système des deux jets principaux et empilement.*



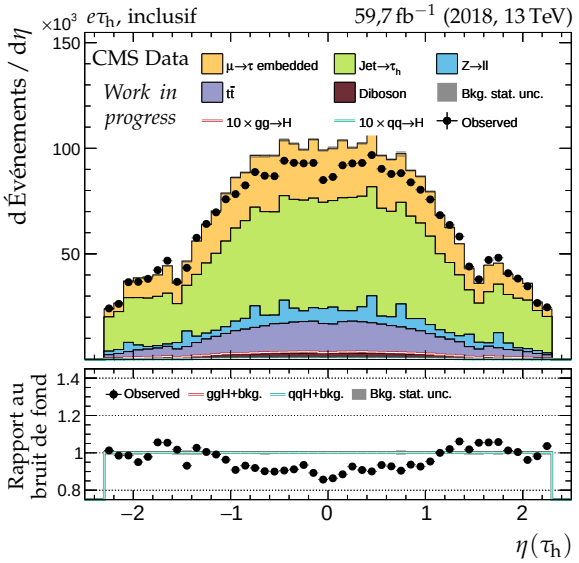
(a) *Impulsion transverse du lepton 1 (e).*



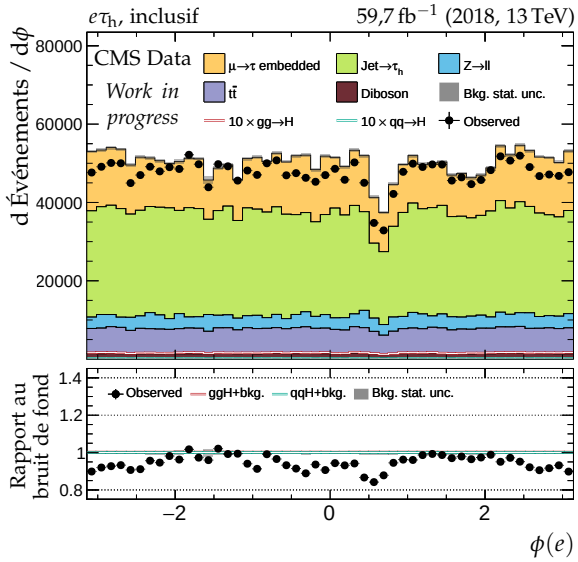
(b) *Impulsion transverse du lepton 2 (τ_h).*



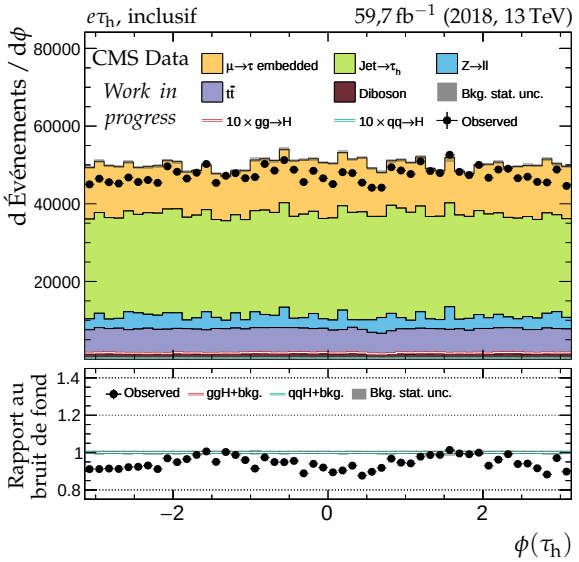
(c) *Pseudo-rapidité du lepton 1 (e).*



(d) *Pseudo-rapidité du lepton 2 (τ_h).*

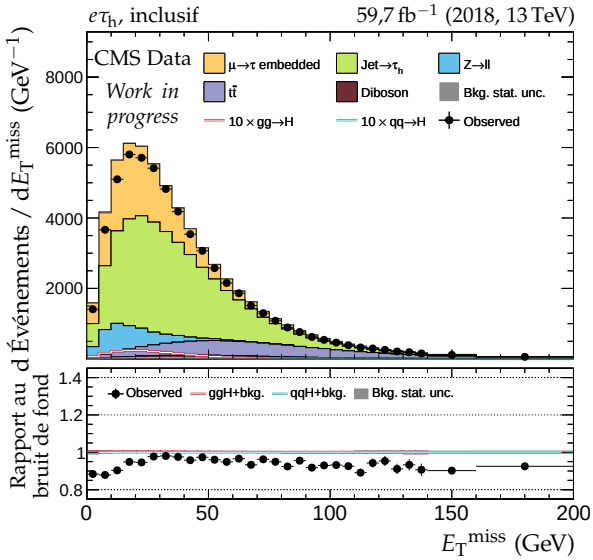


(e) *Angle azimutal du lepton 1 (e).*

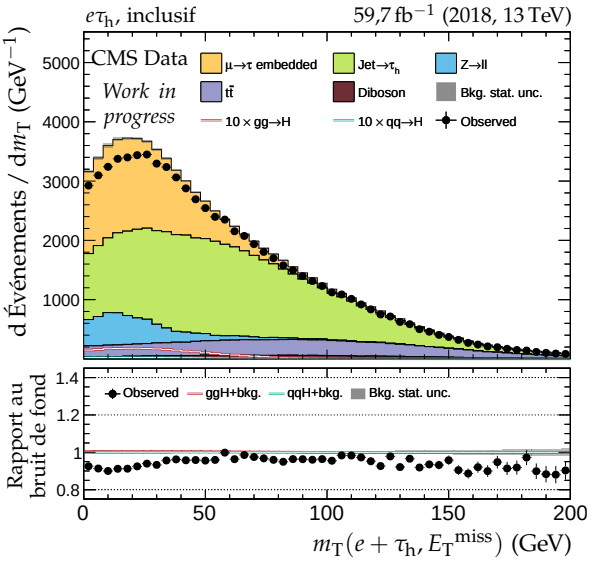


(f) *Angle azimutal du lepton 2 (τ_h).*

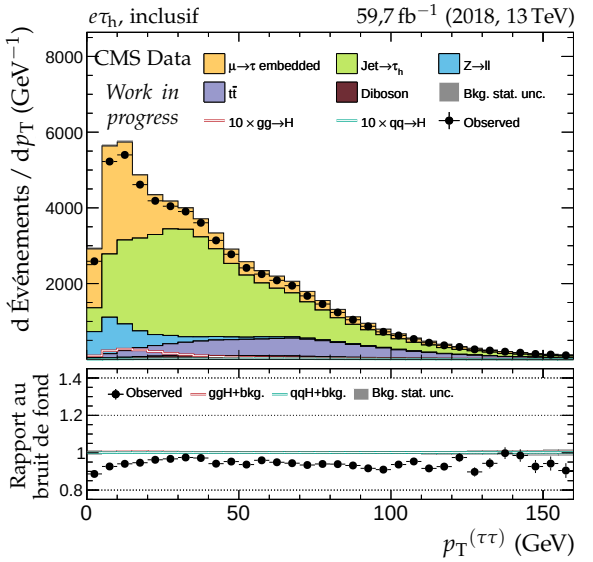
Figure G.63 – *Canal $e\tau_h$, 2018 : cinématique des leptons (e, τ_h).*



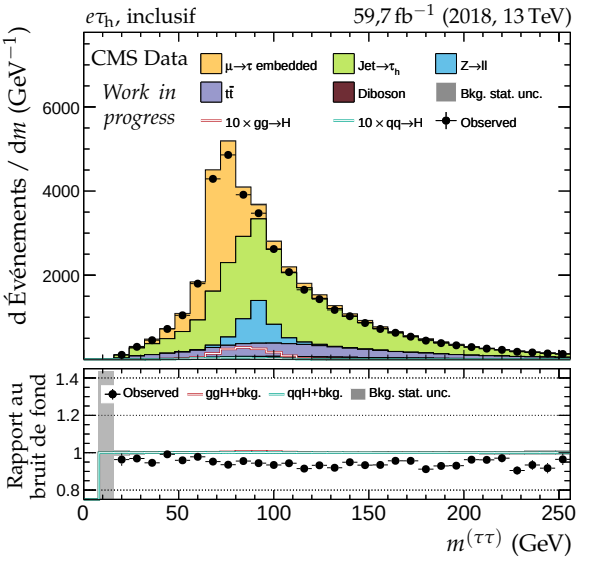
(a) *Énergie transverse manquante.*



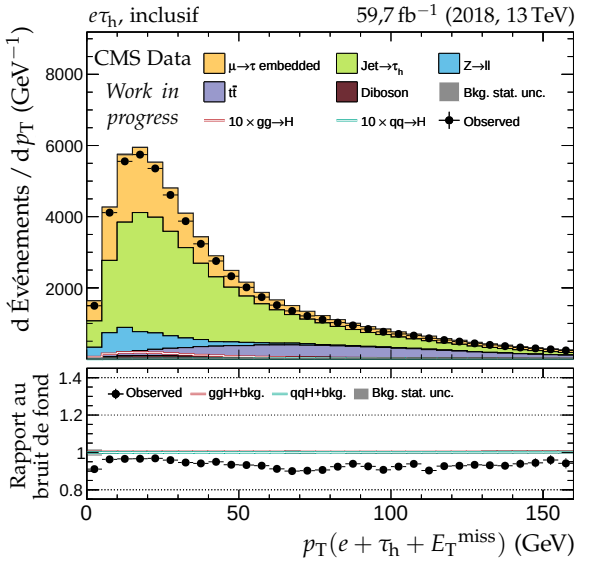
(b) *Masse transverse du dilepton.*



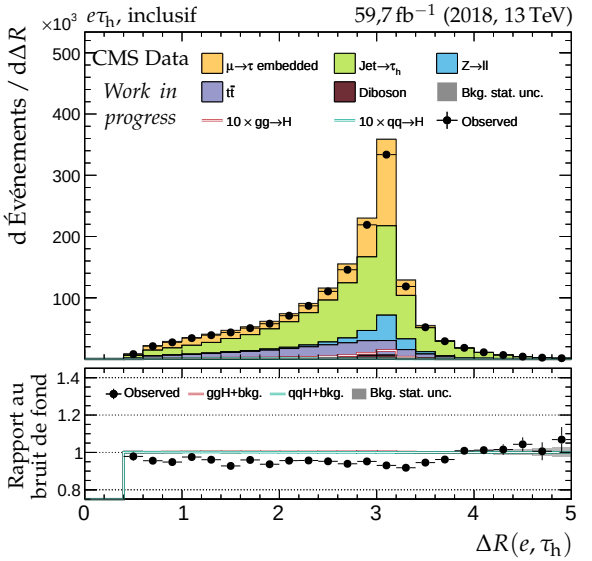
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

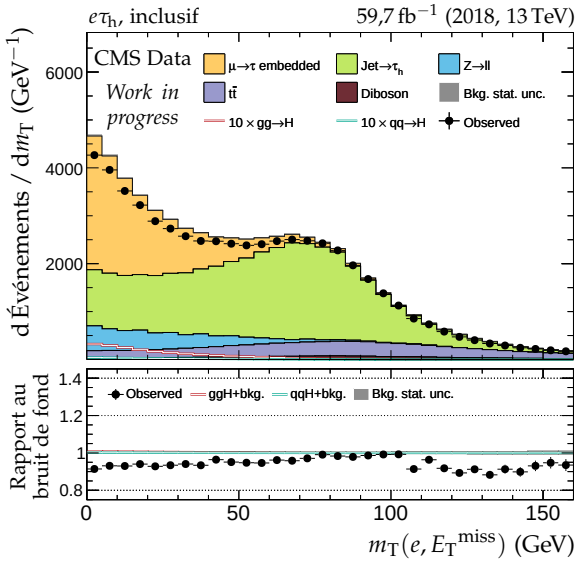


(e) *Impulsion transverse du système di-τ.*

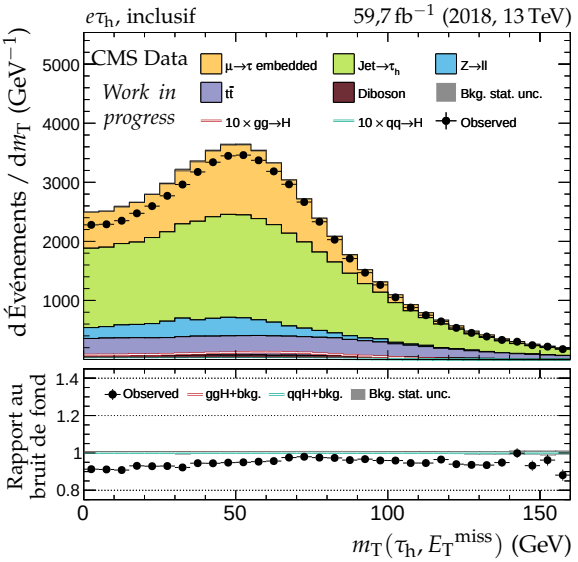


(f) *Distance ΔR entre les leptons (e, τ_h).*

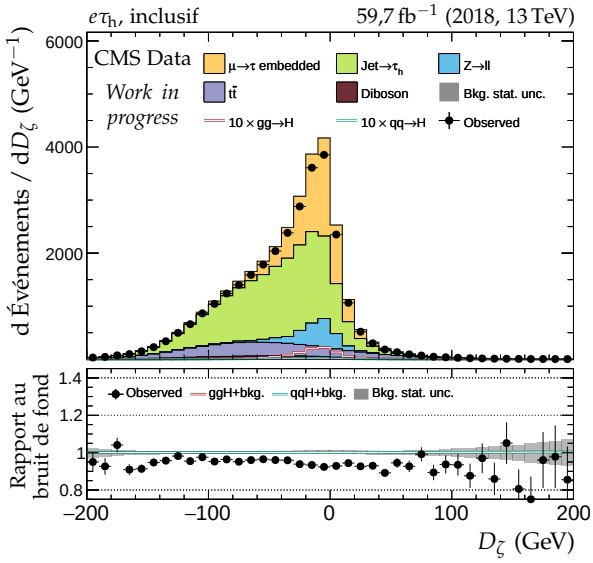
Figure G.64 – *Canal $e\tau_h$, 2018 : dilepton et énergie transverse manquante.*



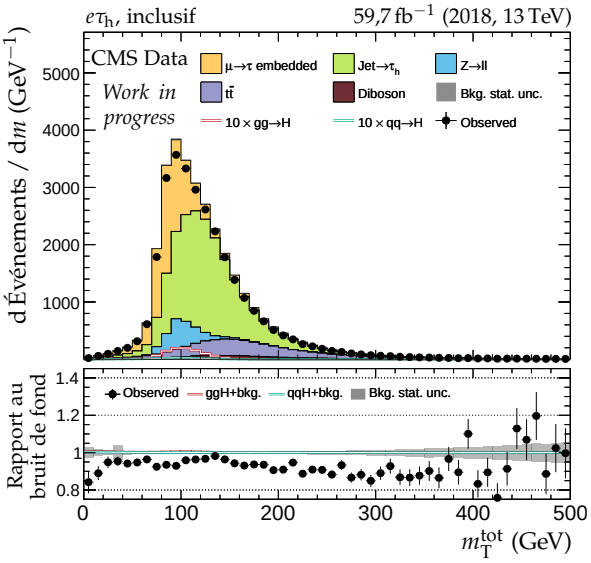
(a) *Masse transverse du lepton 1 (e).*



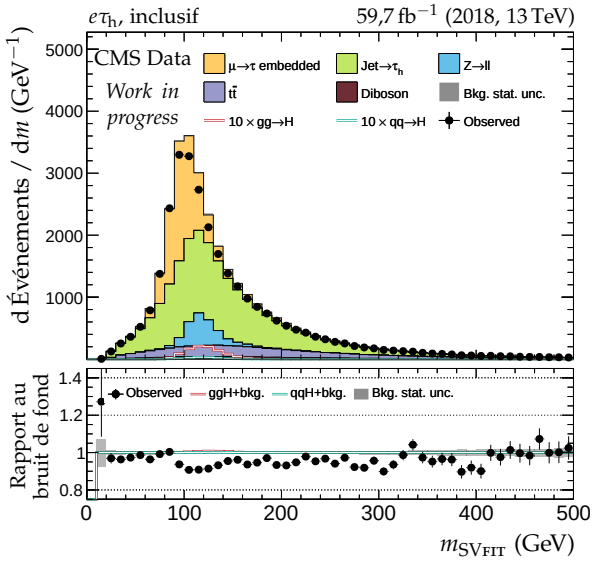
(b) *Masse transverse du lepton 2 (τ_h).*



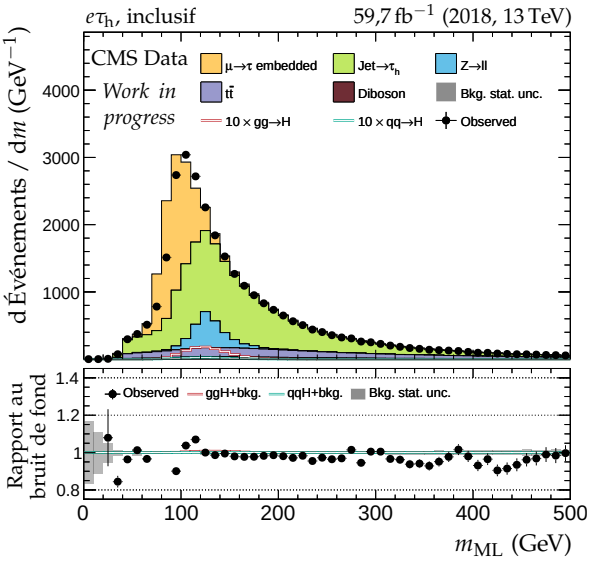
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

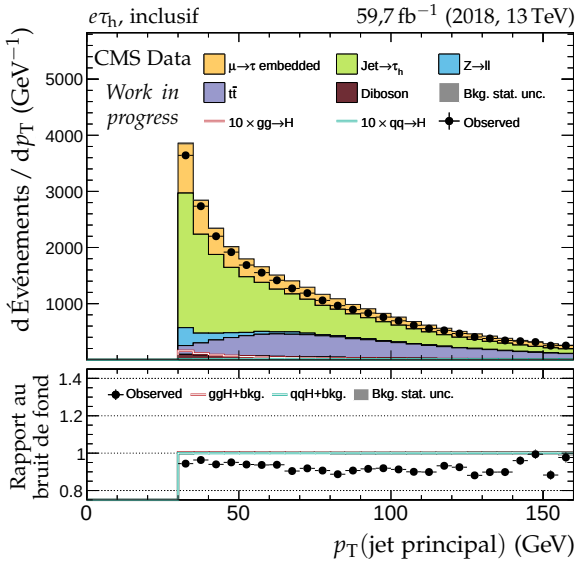


(e) *Masse du système di-τ d'après* SVFIT.

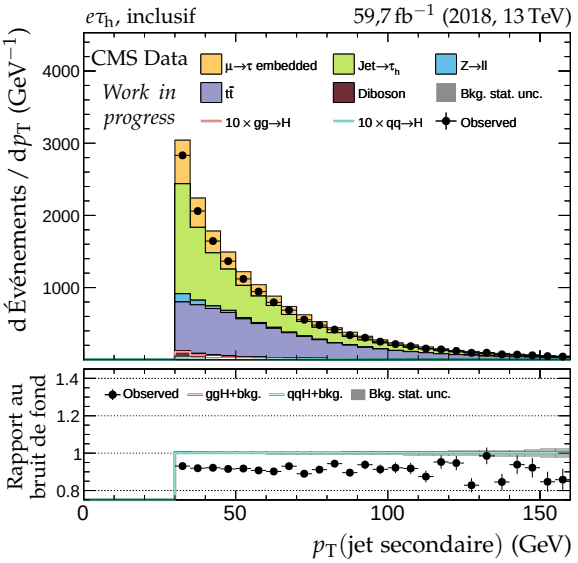


(f) *Masse du système di-τ d'après le ML.*

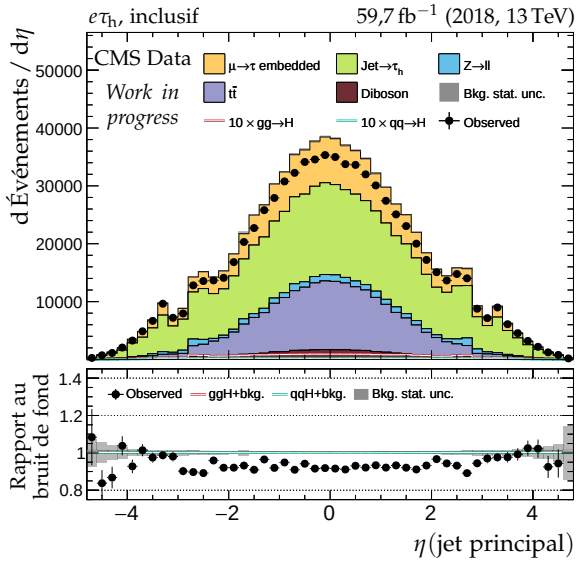
Figure G.65 – *Canal $e\tau_h$, 2018 : masses transverses, D_ζ et masses.*



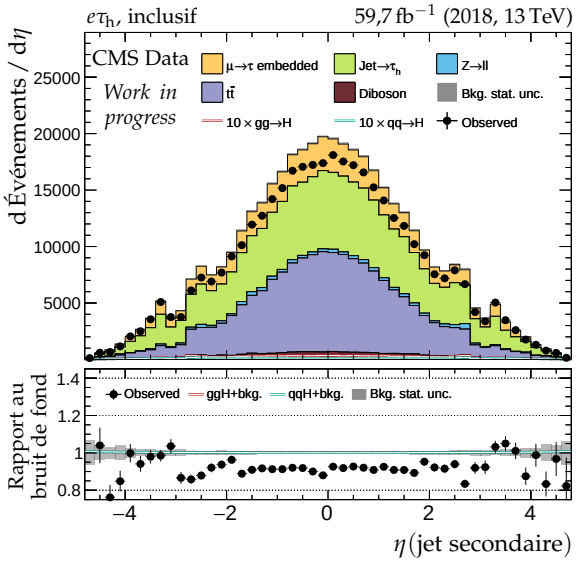
(a) *Impulsion transverse du jet principal.*



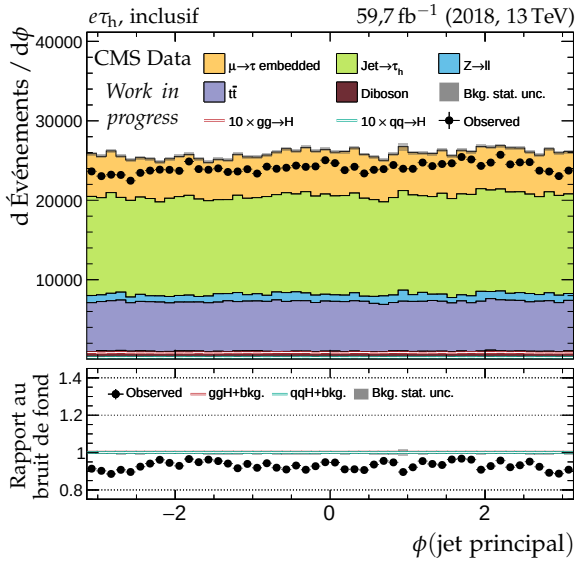
(b) *Impulsion transverse du jet secondaire.*



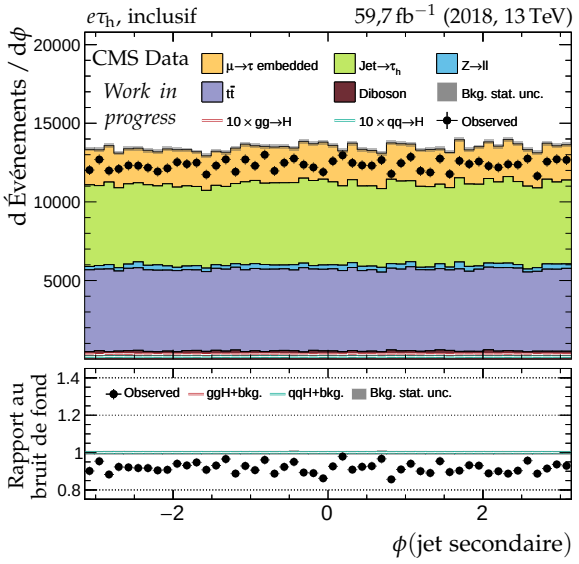
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

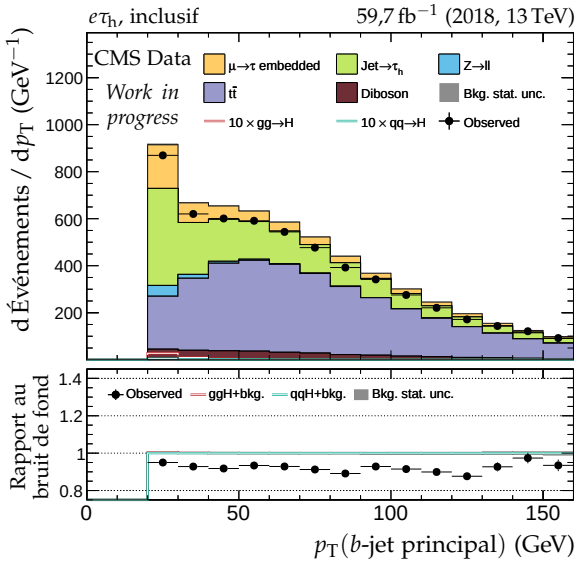


(e) *Angle azimutal du jet principal.*

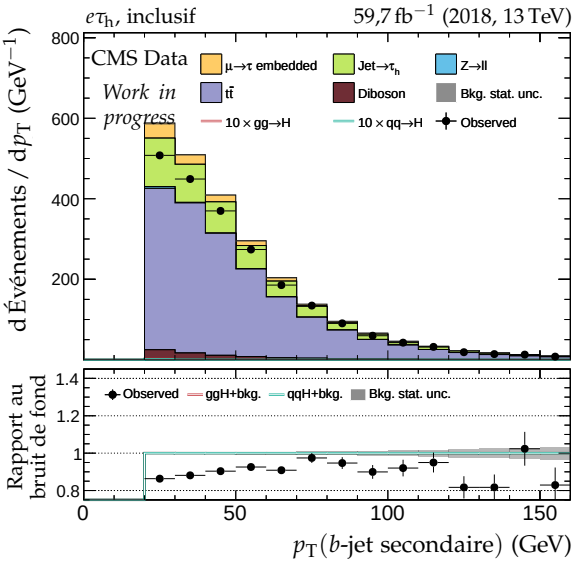


(f) *Angle azimutal du jet secondaire.*

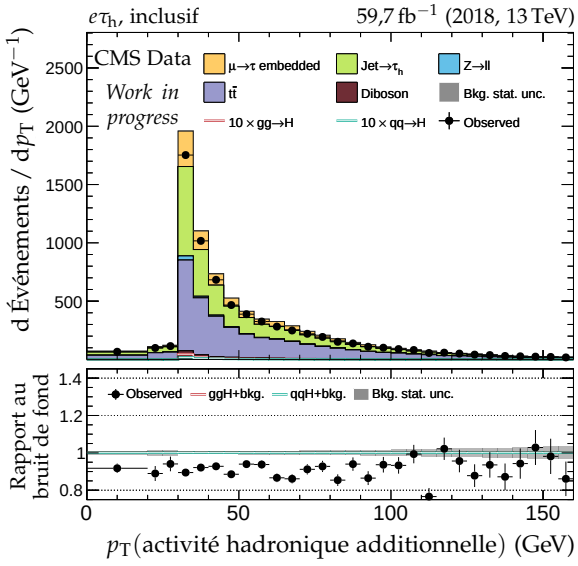
Figure G.66 – *Canal $e\tau_h$, 2018 : cinématique des deux jets principaux.*



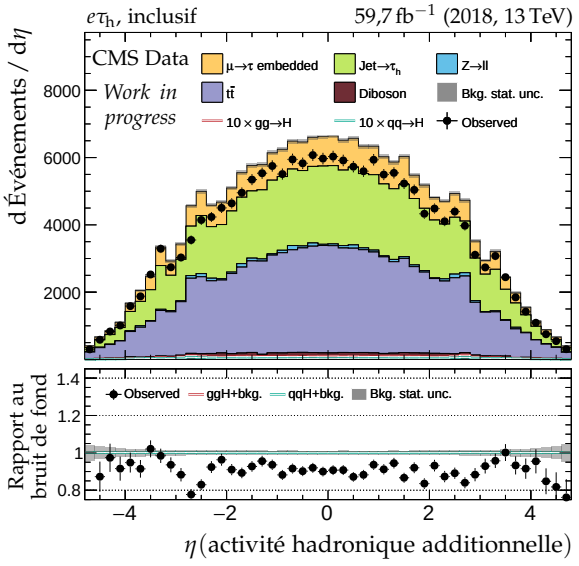
(a) *Impulsion transverse du b-jet principal.*



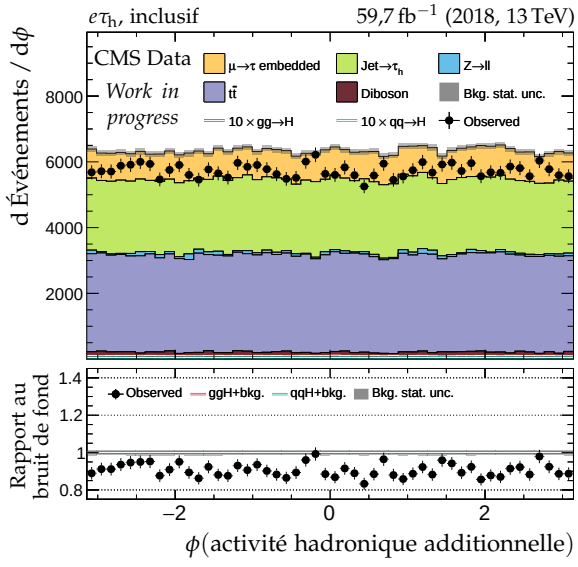
(b) *Impulsion transverse du b-jet secondaire.*



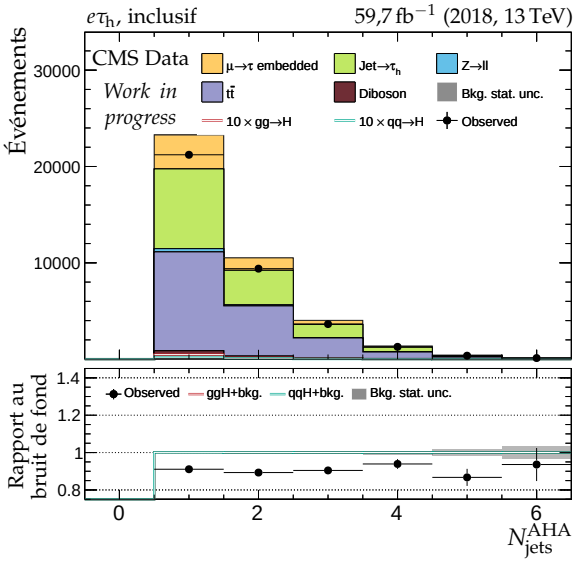
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

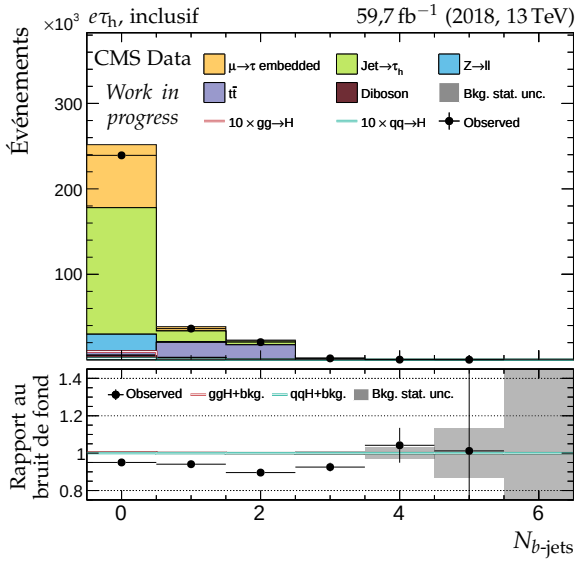


(e) *Angle azimutal de l'AHA.*

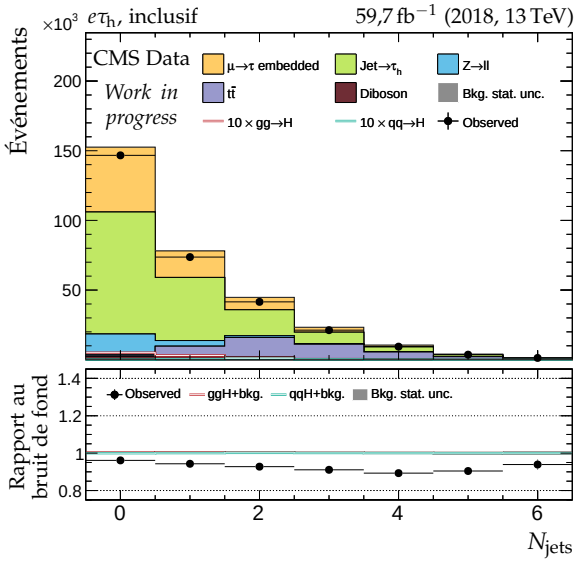


(f) *Nombre de jets dans l'AHA.*

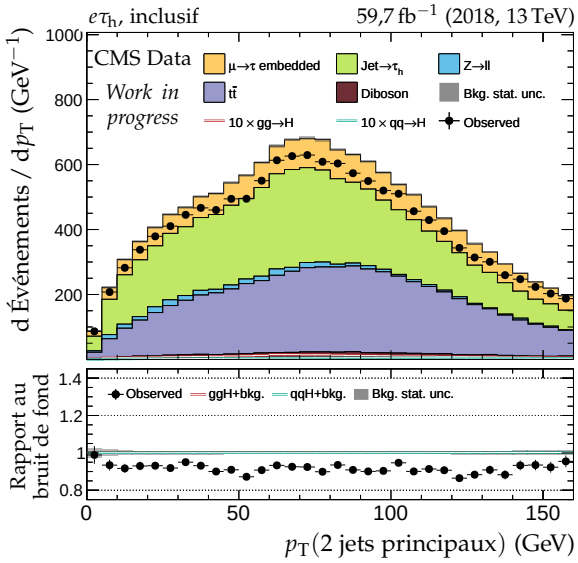
Figure G.67 – *Canal eτh, 2018 : b-jets et activité hadronique additionnelle.*



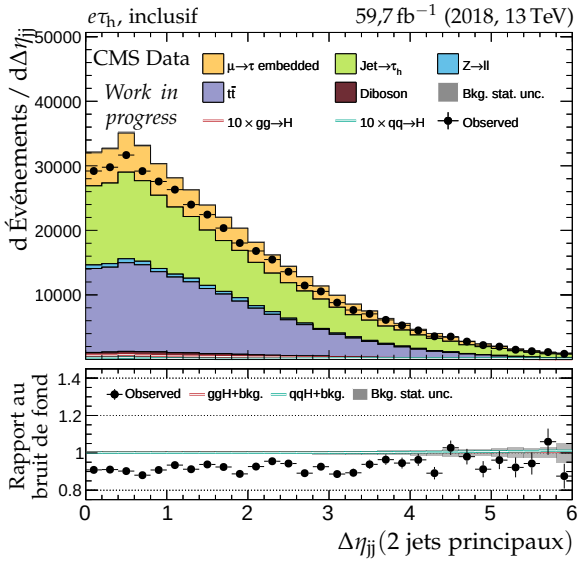
(a) *Nombre de b-jets.*



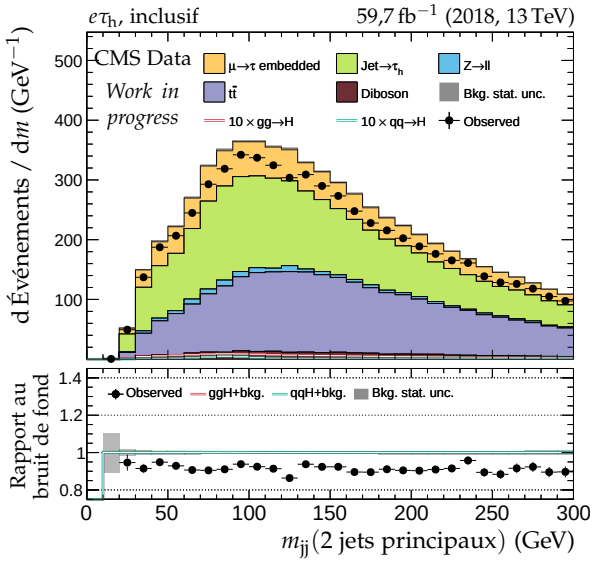
(b) *Nombre de jets.*



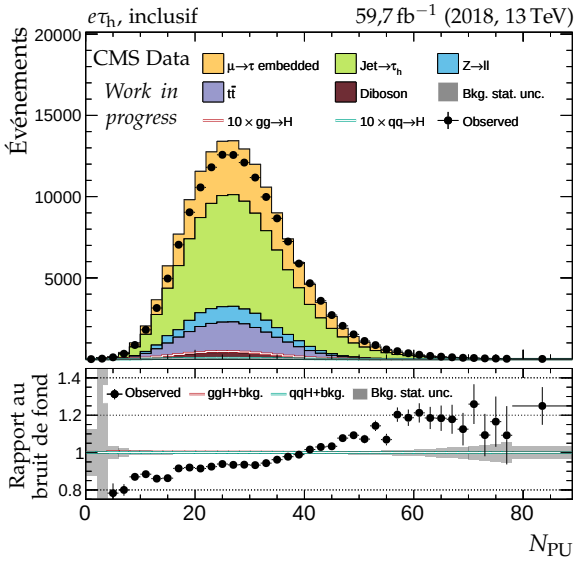
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*

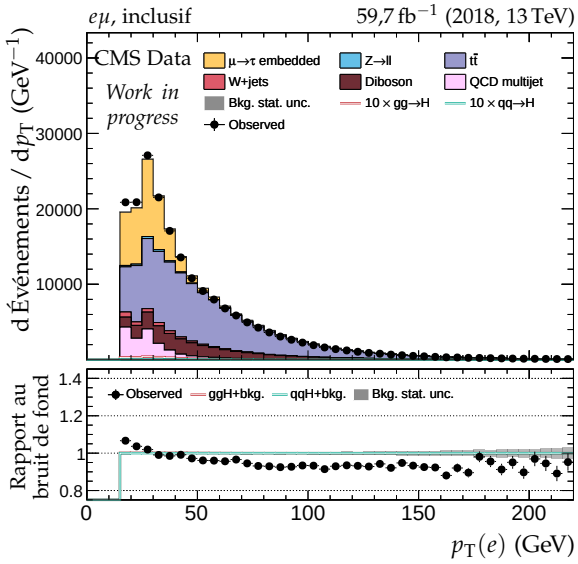


(e) *Masse invariante du système des deux jets.*

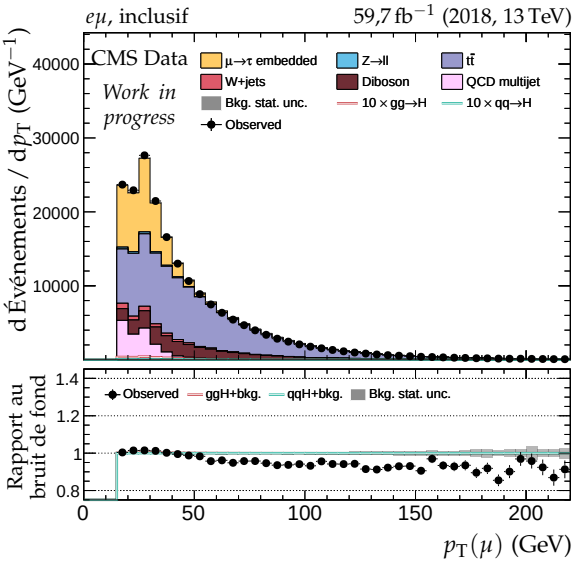


(f) *Nombre de vertex d'empilement.*

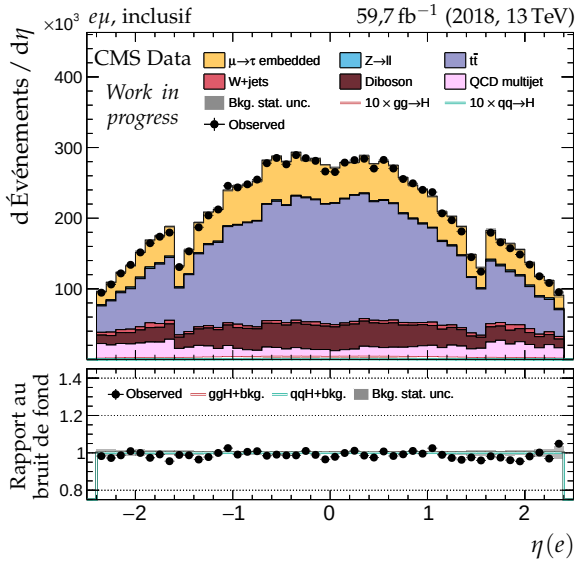
Figure G.68 – *Canal $e\tau_h$, 2018 : nombre de jets, système des deux jets principaux et empilement.*



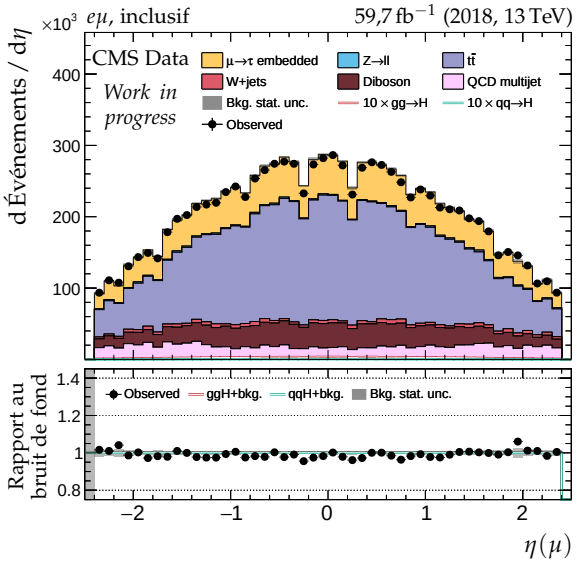
(a) *Impulsion transverse du lepton 1 (e).*



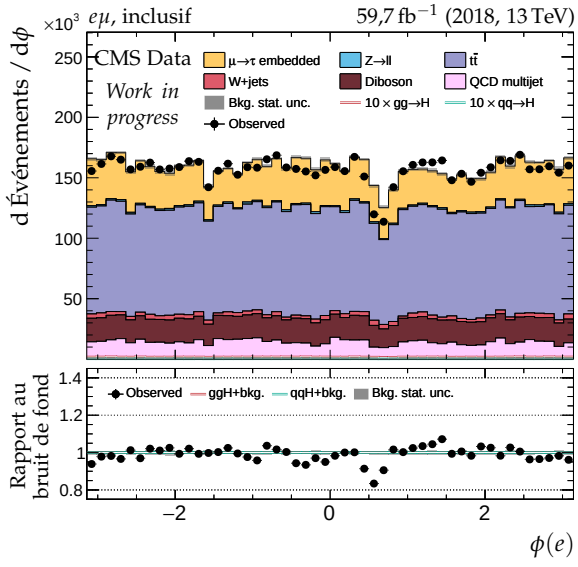
(b) *Impulsion transverse du lepton 2 (μ).*



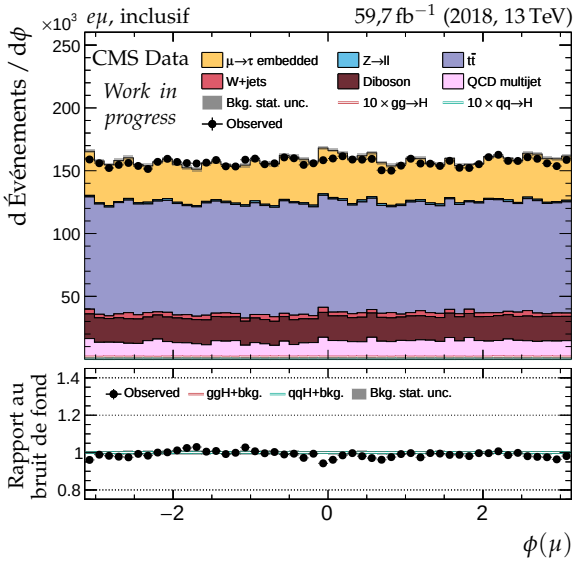
(c) *Pseudo-rapidité du lepton 1 (e).*



(d) *Pseudo-rapidité du lepton 2 (μ).*

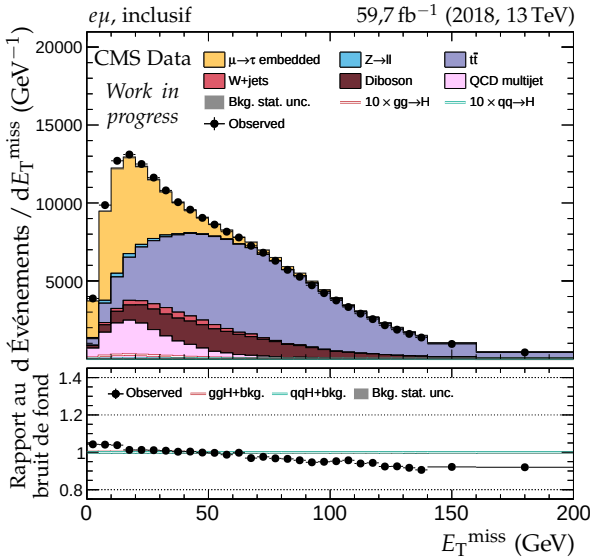


(e) *Angle azimutal du lepton 1 (e).*

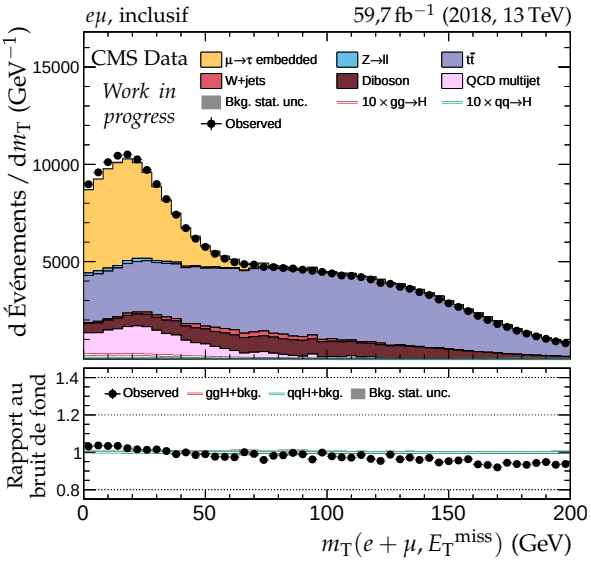


(f) *Angle azimutal du lepton 2 (μ).*

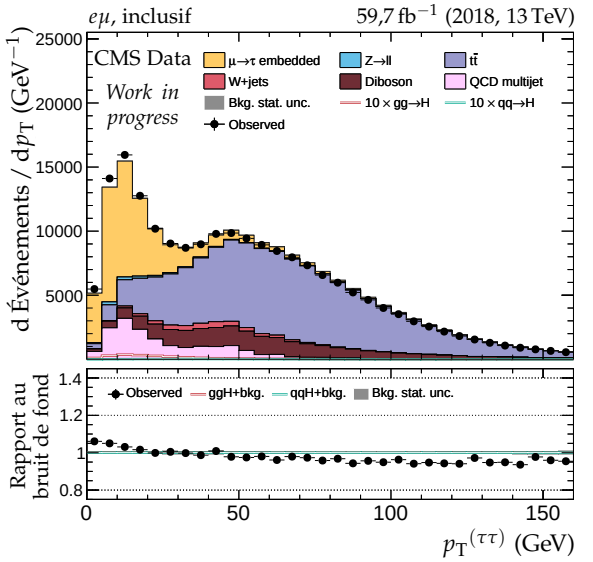
Figure G.69 – *Canal eμ, 2018 : cinématique des leptons (e, μ).*



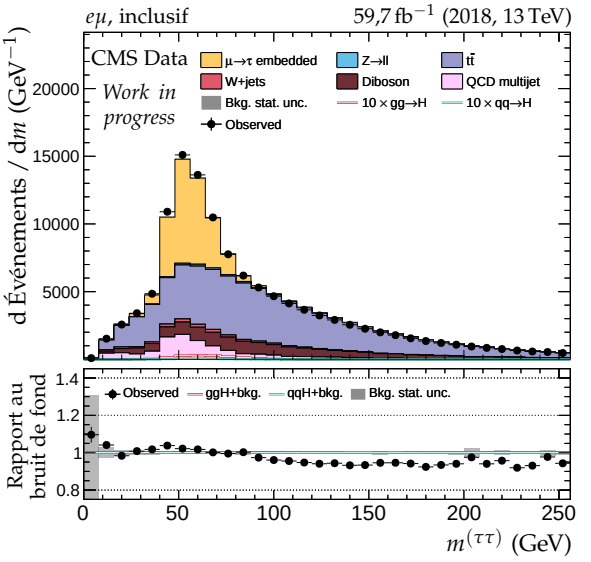
(a) *Énergie transverse manquante.*



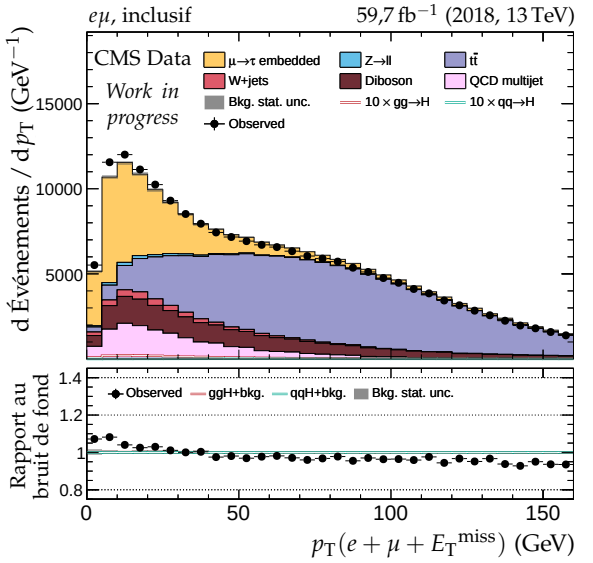
(b) *Masse transverse du dilepton.*



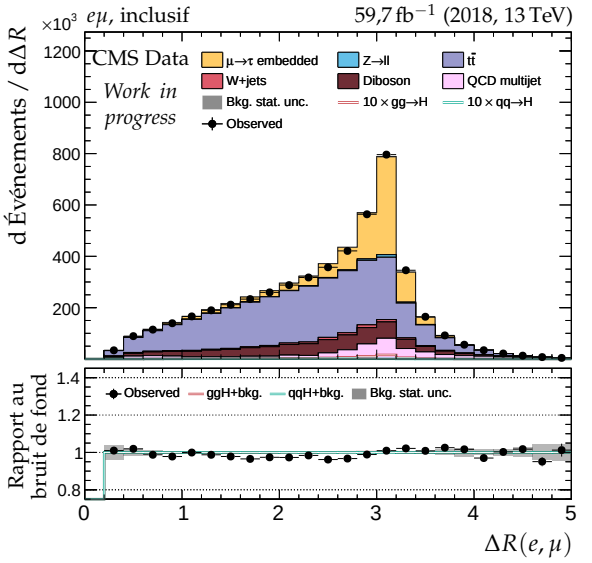
(c) *Impulsion transverse du dilepton.*



(d) *Masse visible du dilepton.*

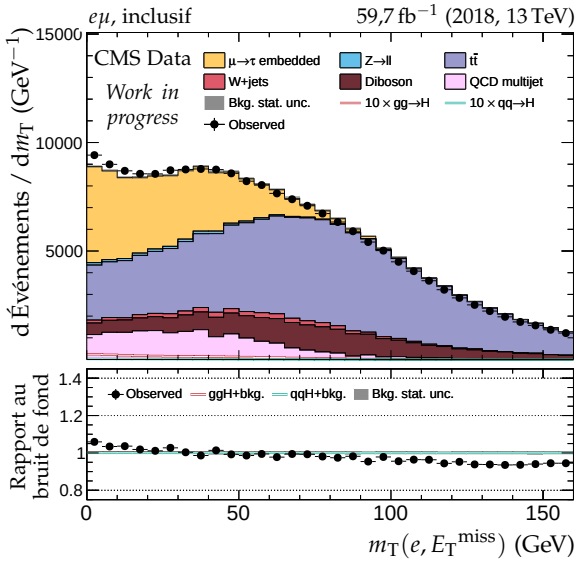


(e) *Impulsion transverse du système di-τ.*

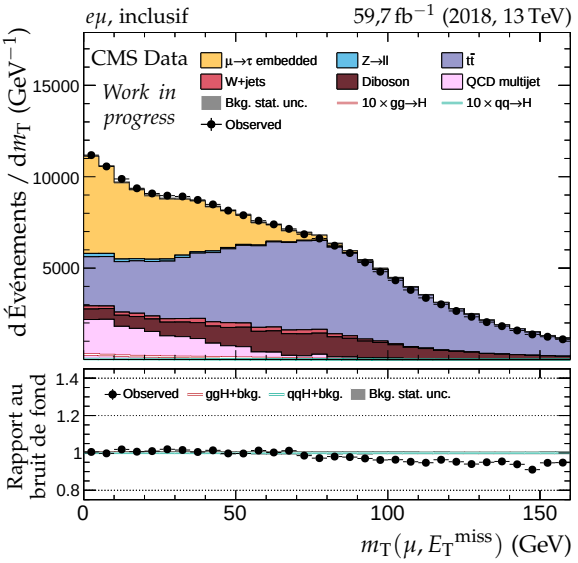


(f) *Distance ΔR entre les leptons (e, μ).*

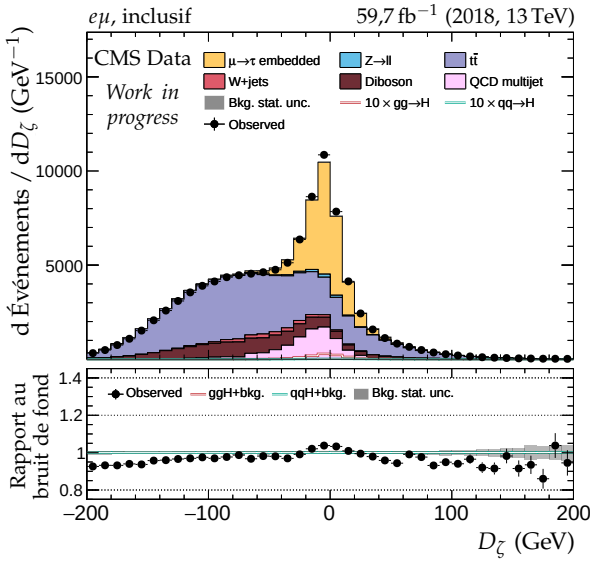
Figure G.70 – *Canal eμ, 2018 : dilepton et énergie transverse manquante.*



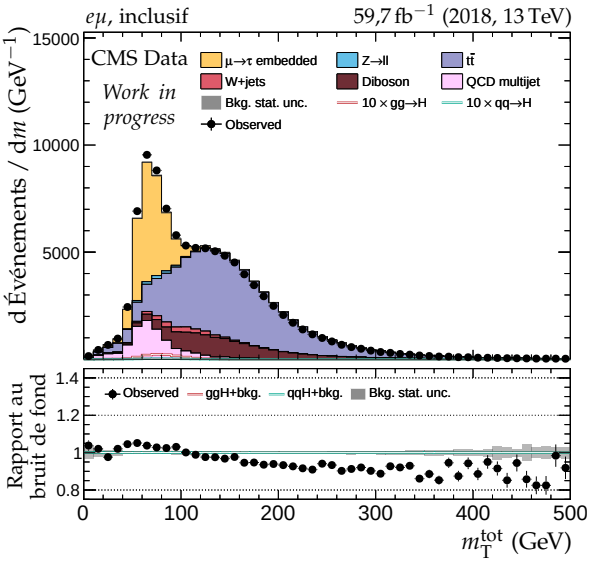
(a) *Masse transverse du lepton 1 (e).*



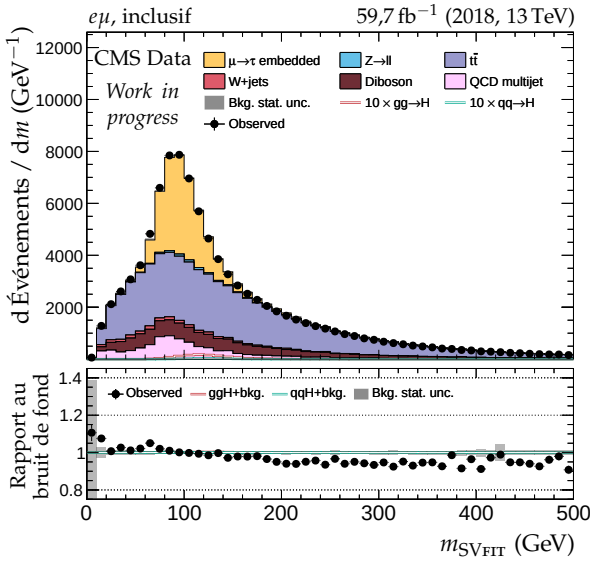
(b) *Masse transverse du lepton 2 (μ).*



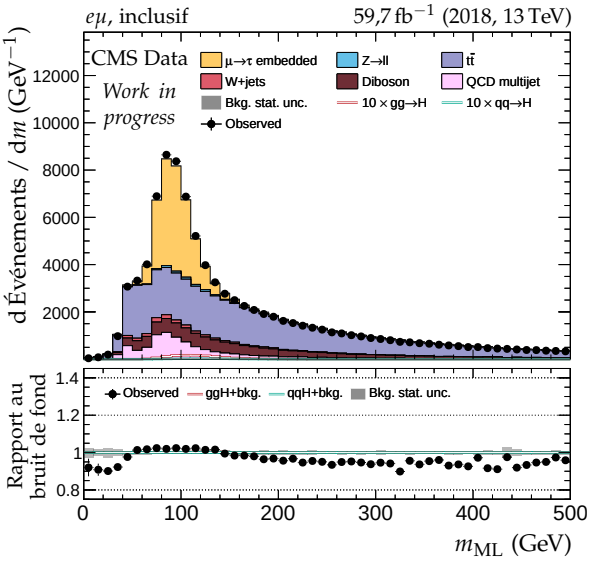
(c) *Valeur de D_ζ.*



(d) *Masse transverse totale.*

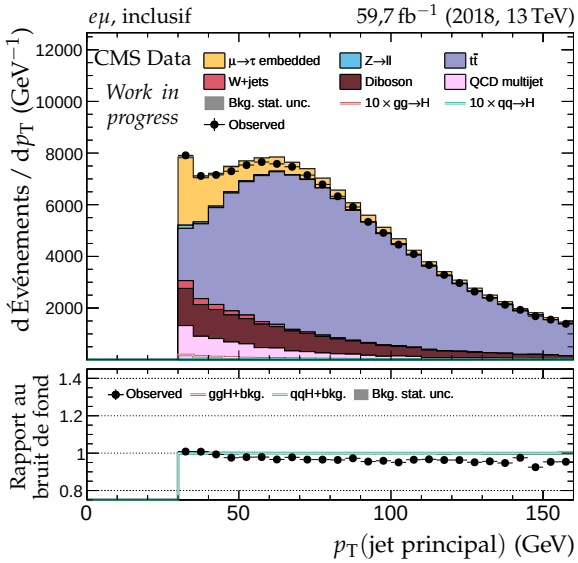


(e) *Masse du système di-τ d'après* SVFIT.

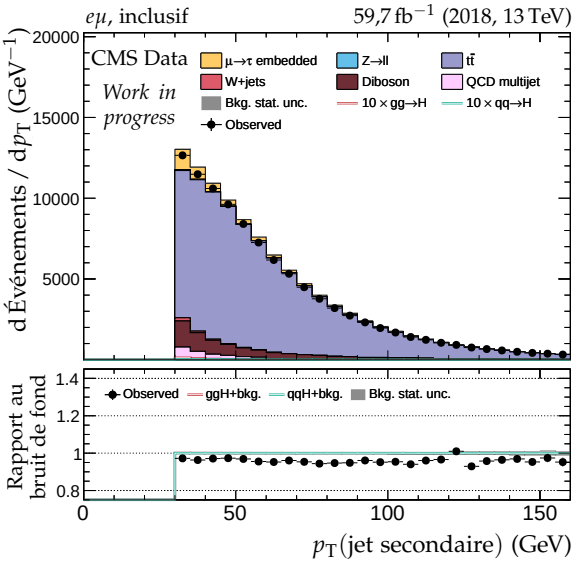


(f) *Masse du système di-τ d'après le ML.*

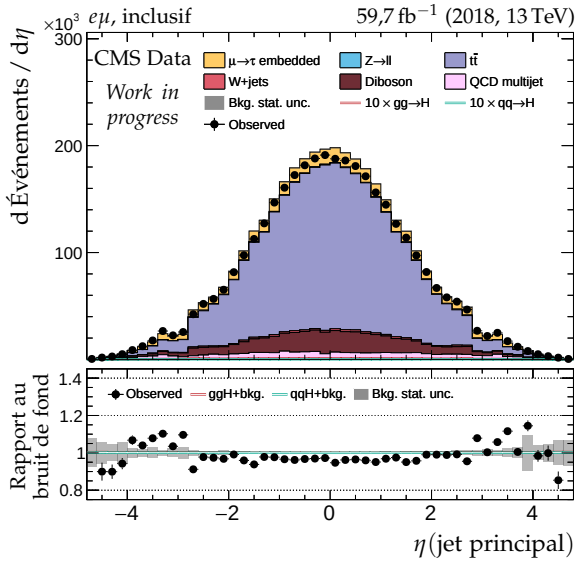
Figure G.71 – *Canal eμ, 2018 : masses transverses, D_ζ et masses.*



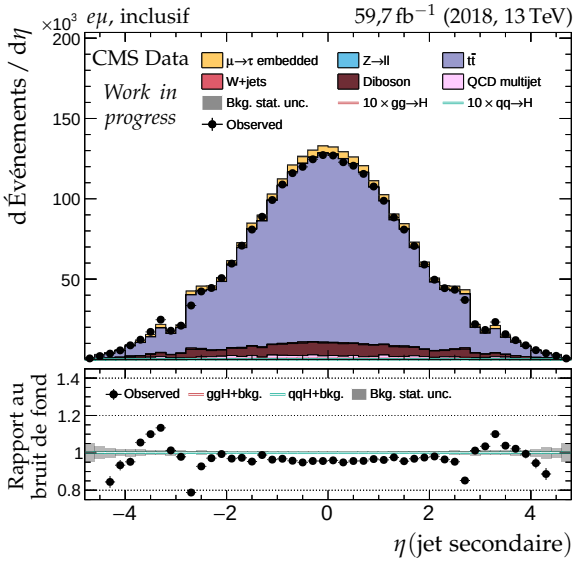
(a) *Impulsion transverse du jet principal.*



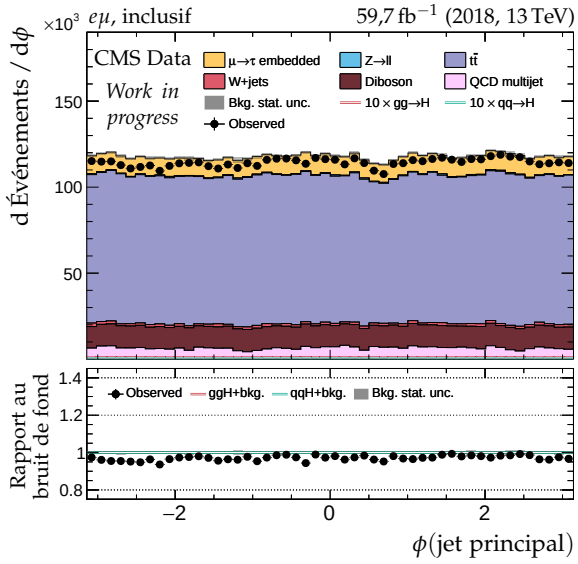
(b) *Impulsion transverse du jet secondaire.*



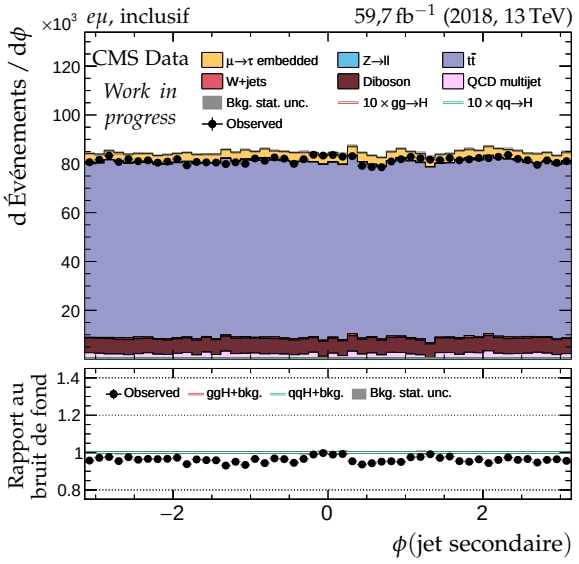
(c) *Pseudo-rapidité du jet principal.*



(d) *Pseudo-rapidité du jet secondaire.*

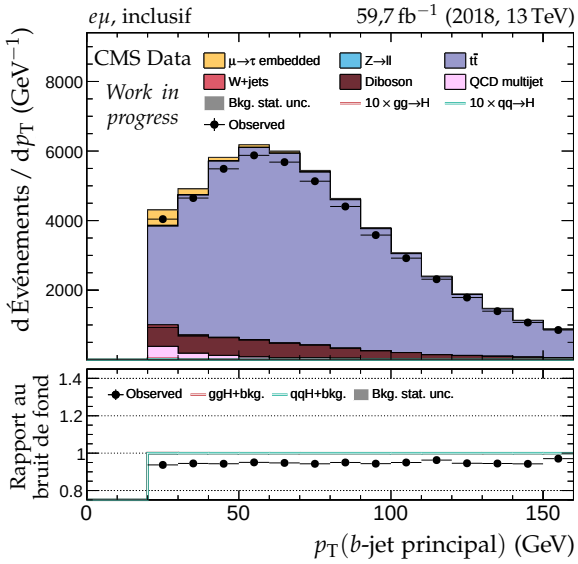


(e) *Angle azimutal du jet principal.*

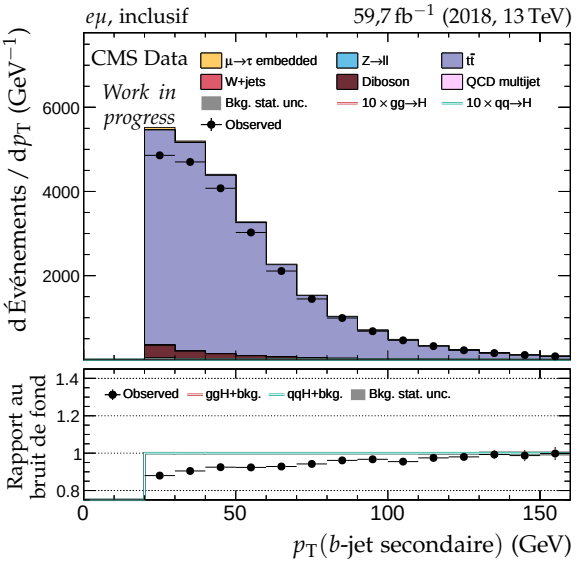


(f) *Angle azimutal du jet secondaire.*

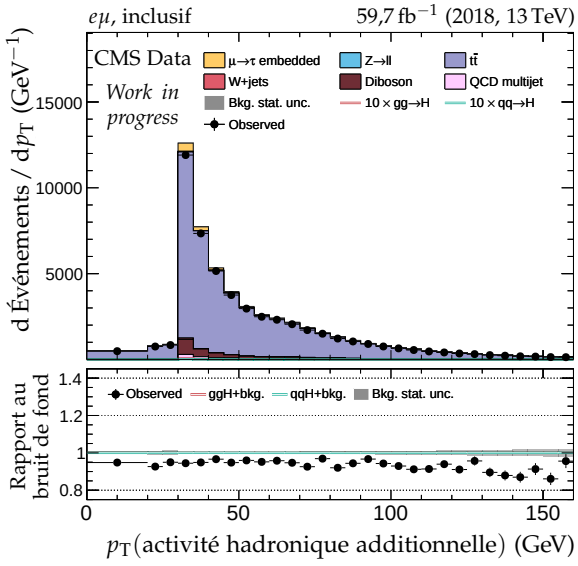
Figure G.72 – *Canal eμ, 2018 : cinématique des deux jets principaux.*



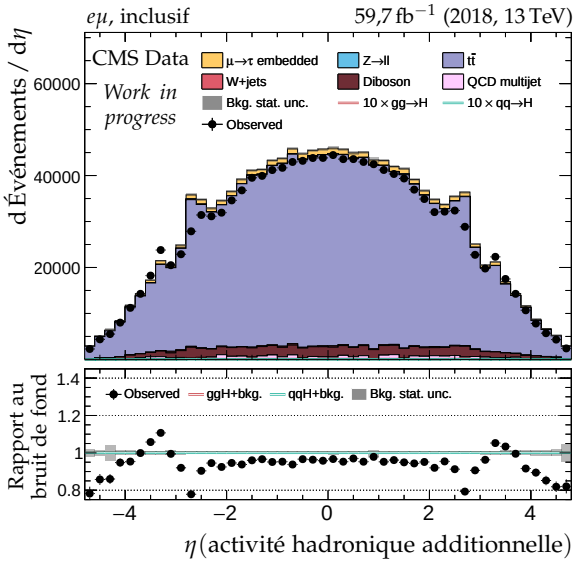
(a) *Impulsion transverse du b-jet principal.*



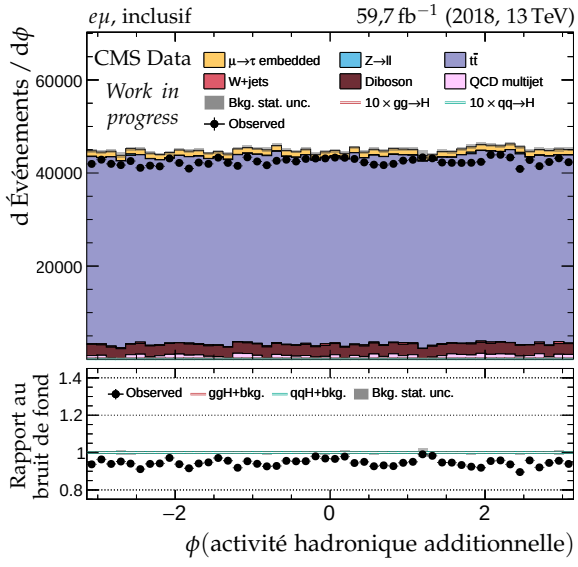
(b) *Impulsion transverse du b-jet secondaire.*



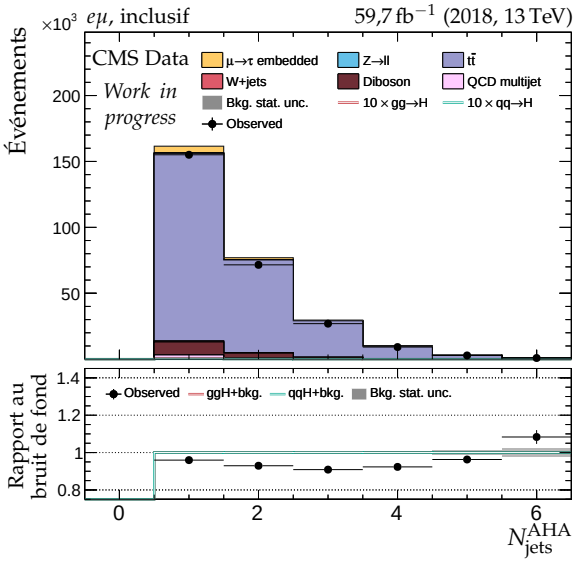
(c) *Impulsion transverse de l'AHA.*



(d) *Pseudo-rapidité de l'AHA.*

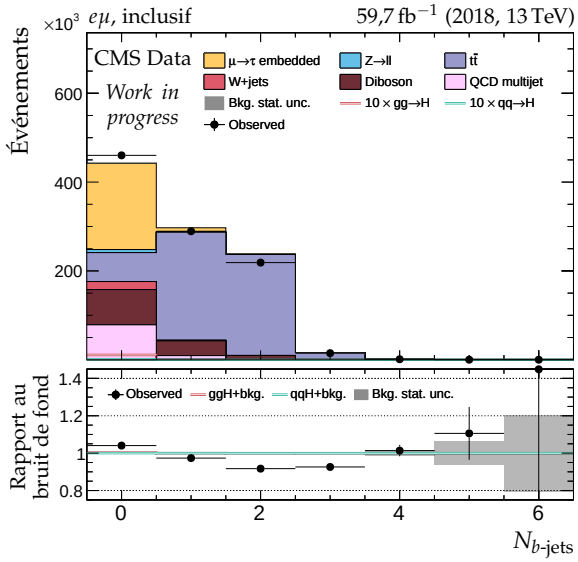


(e) *Angle azimutal de l'AHA.*

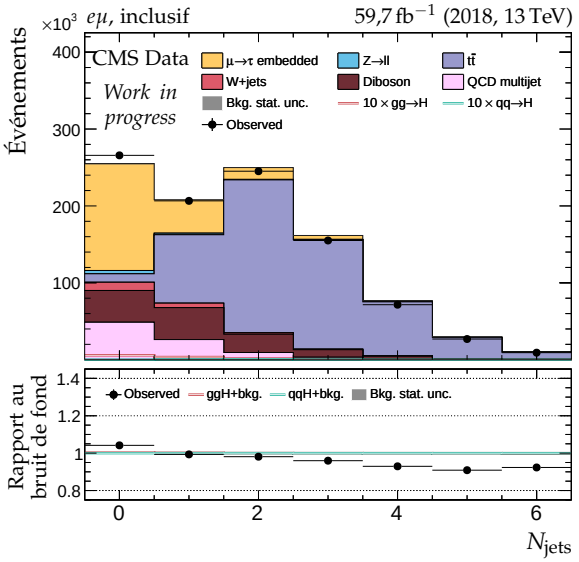


(f) *Nombre de jets dans l'AHA.*

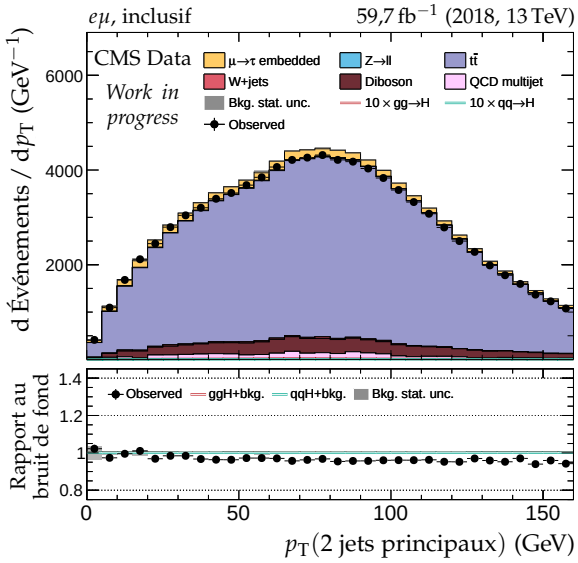
Figure G.73 – *Canal eμ, 2018 : b-jets et activité hadronique additionnelle.*



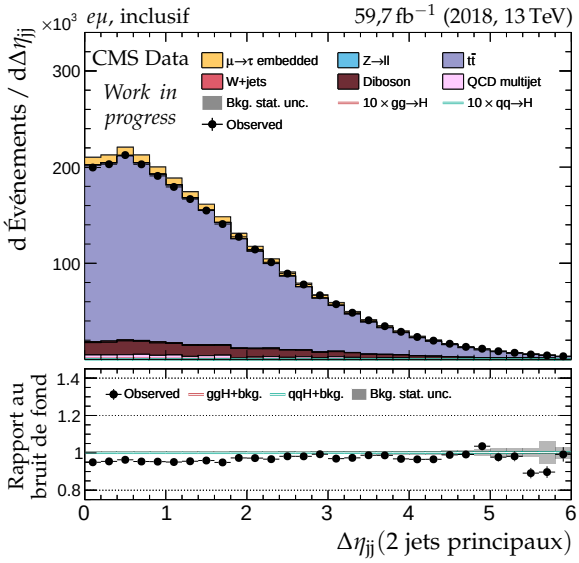
(a) *Nombre de b-jets.*



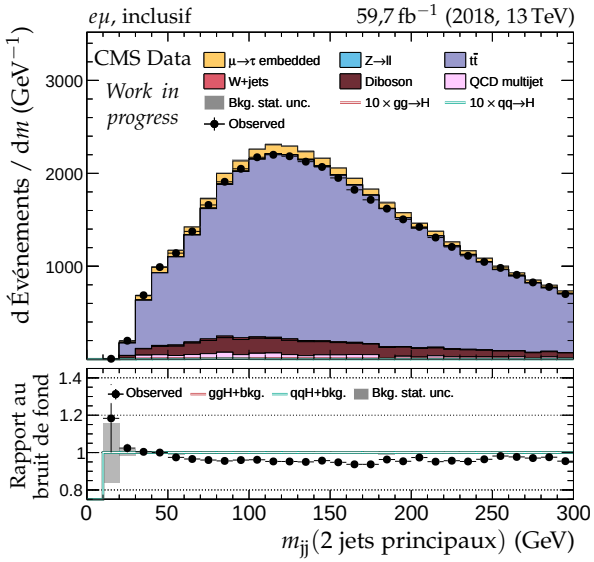
(b) *Nombre de jets.*



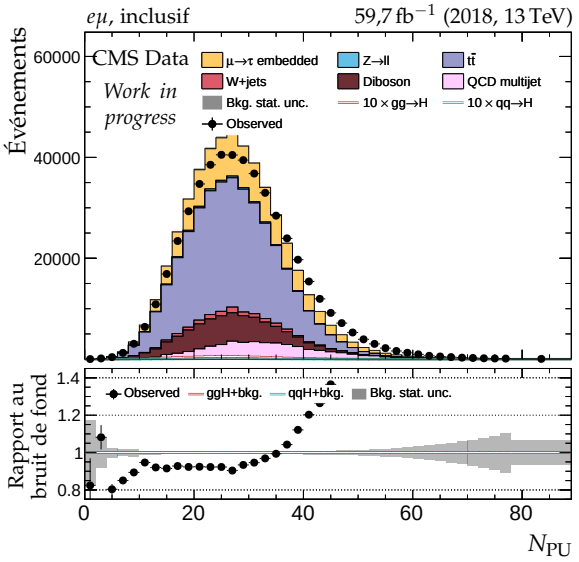
(c) *Impulsion transverse du système des deux jets.*



(d) *Distance en η entre les deux jets.*



(e) *Masse invariante du système des deux jets.*



(f) *Nombre de vertex d'empilement.*

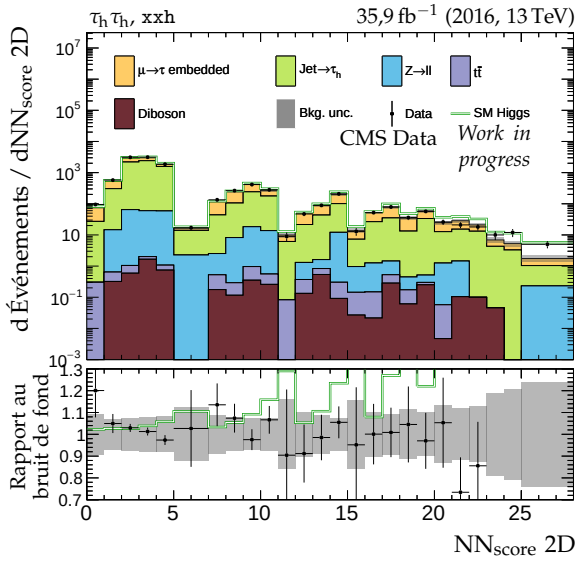
Figure G.74 – *Canal eμ, 2018 : nombre de jets, système des deux jets principaux et empilement.*

Annexe H
Distributions des variables discriminantes – $H \to \tau\tau$

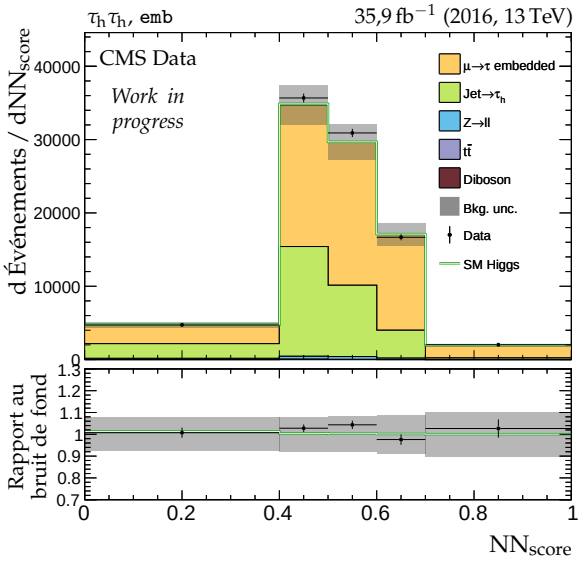
Cette annexe présente des distributions des variables discriminantes avant ajustement des paramètres de nuisance sur les événements utilisés dans l'analyse des événements $H \to \tau\tau$ présentée dans le chapitre 4. Cet ajustement dépend en effet de l'étude exacte menée (limites indépendantes du modèle, dépendantes d'un modèle, duquel, etc.). Les distributions avant ajustement sont en revanche communes aux différentes études menées au chapitre 4.

Pour chacune des trois années de prise de données (2016, 2017, 2018), chacun des quatre canaux ($\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$, $e\mu$) et chacune des catégories définies au chapitre 4, la distribution de la variable discriminante associée est donnée.

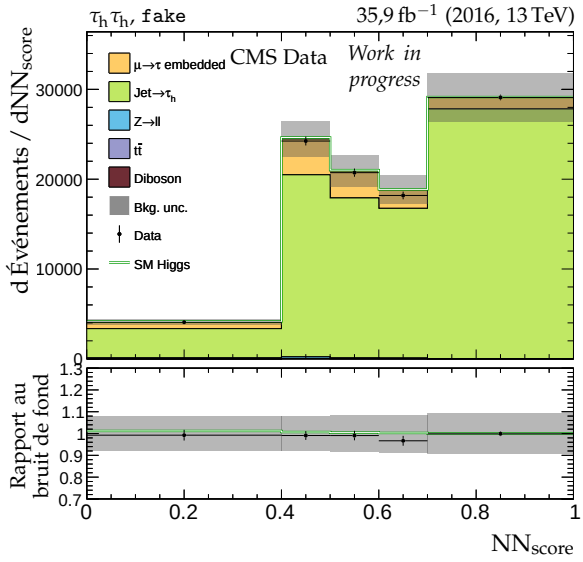
Dans chacun des graphiques, les données observées (points noirs) sont comparées à la modélisation des bruits de fond (histogrammes remplis en couleur et empilés). Les bandes grisées correspondent à l'incertitude totale (statistique et systématique) sur le bruit de fond avant ajustement des paramètres de nuisance par COMBINE. Le rapport au bruit de fond est donné dans la partie inférieure des graphiques.



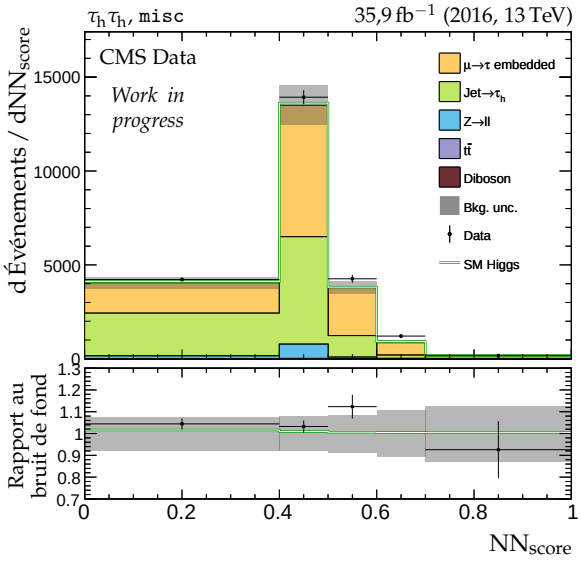
(a) *Catégorie* `xxh`.



(b) *Catégorie* `emb`.

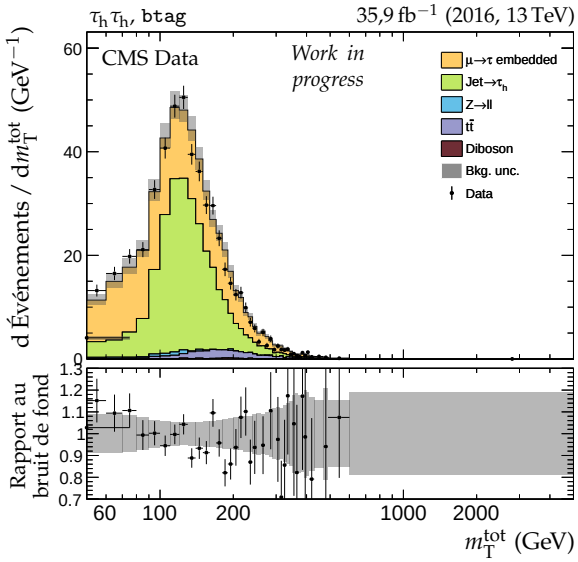


(c) *Catégorie* `fake`.

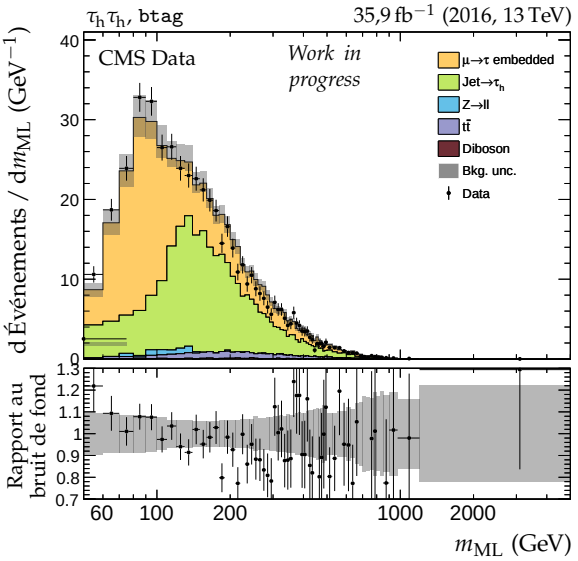


(d) *Catégorie* `misc`.

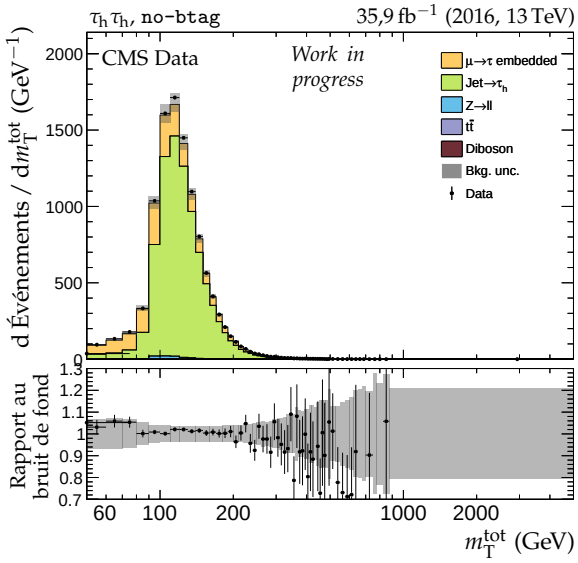
Figure H.1 – *Distributions de* $\mathrm{NN_{score}}$ *en 2016 dans le canal* $\tau_\mathrm{h}\tau_\mathrm{h}$.



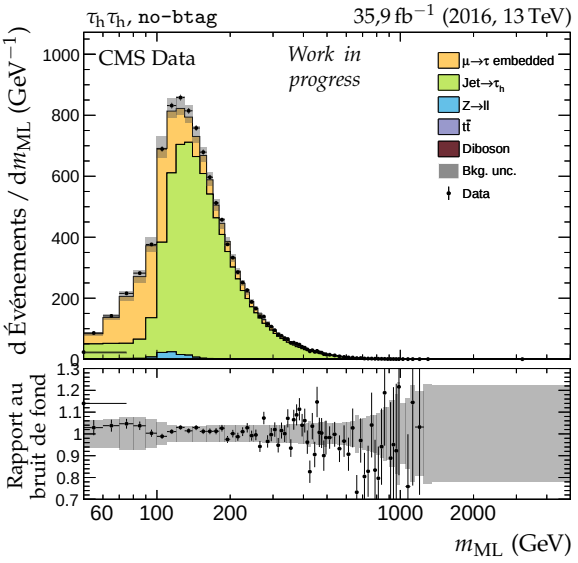
(a) *Catégorie* `btag`, m_T^{tot}.



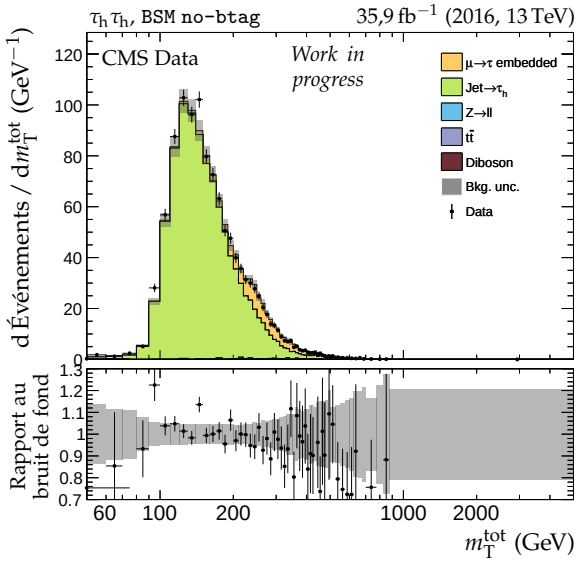
(b) *Catégorie* `btag`, m_{ML}.



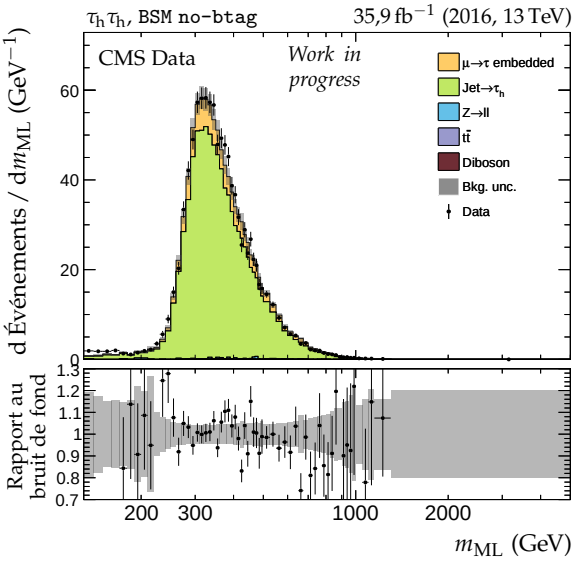
(c) *Catégorie* `no-btag`, m_T^{tot}.



(d) *Catégorie* `no-btag`, m_{ML}.

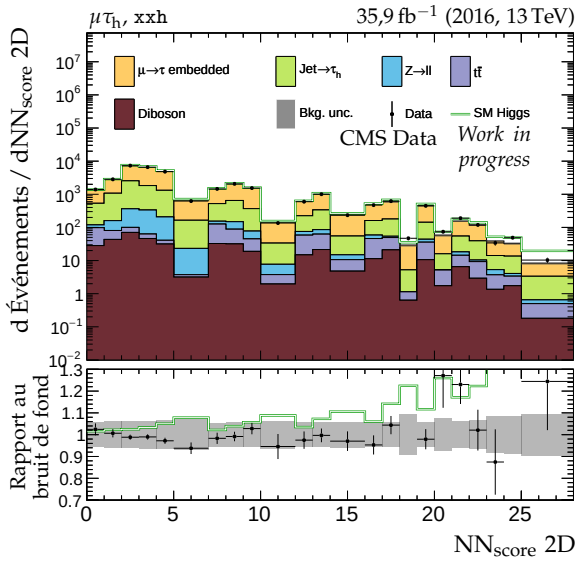


(e) *Catégorie BSM* `no-btag`, m_T^{tot}.

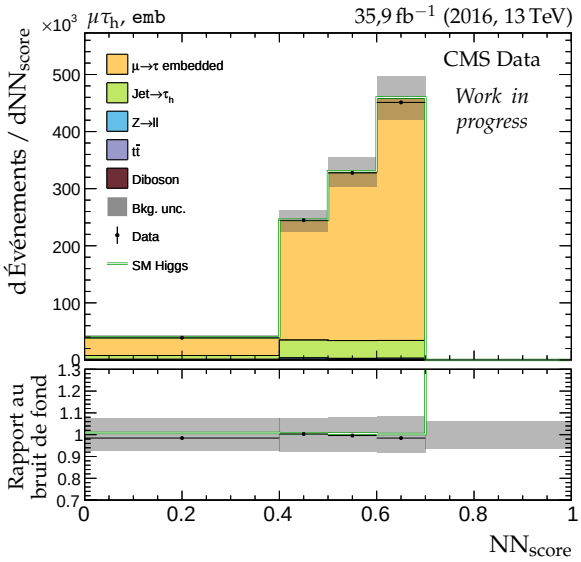


(f) *Catégorie BSM* `no-btag`, m_{ML}.

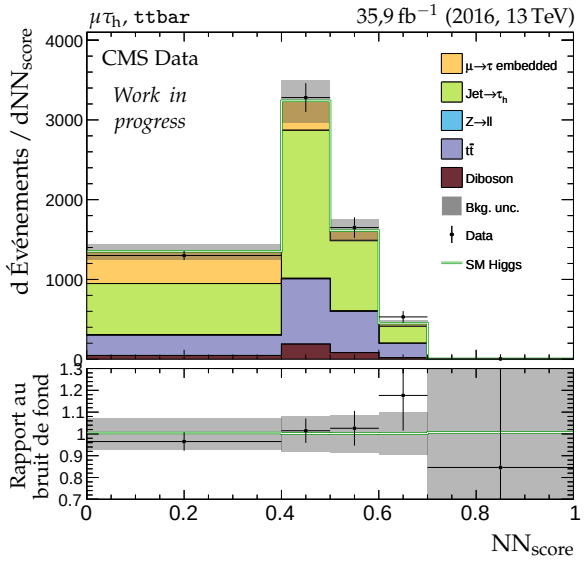
Figure H.2 – *Distributions de m_T^{tot} et m_{ML} en 2016 dans le canal $\tau_h\tau_h$.*



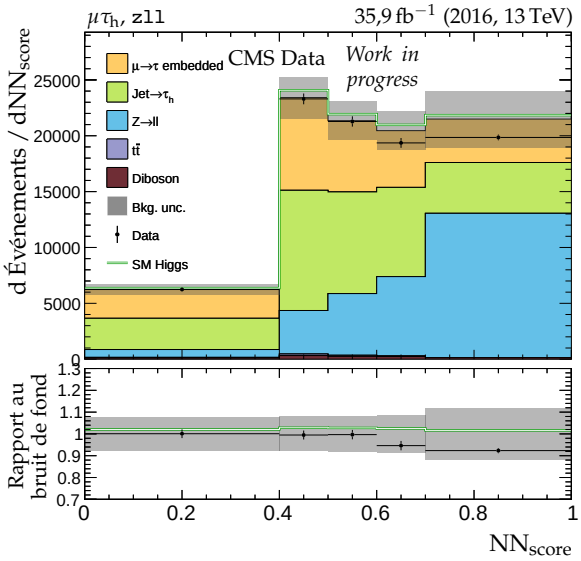
(a) *Catégorie* `xxh`.



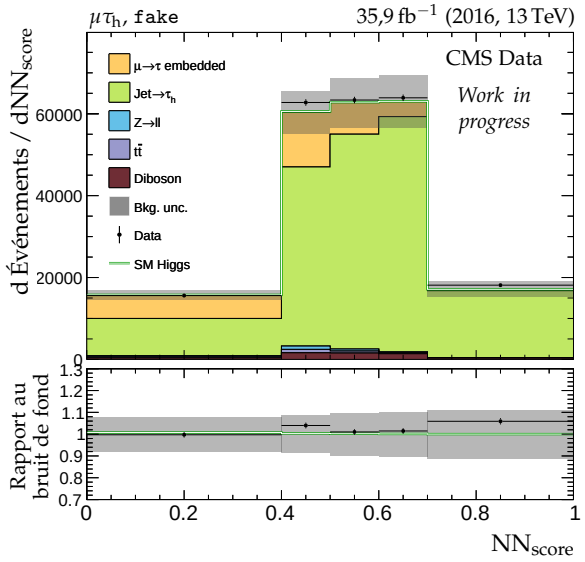
(b) *Catégorie* `emb`.



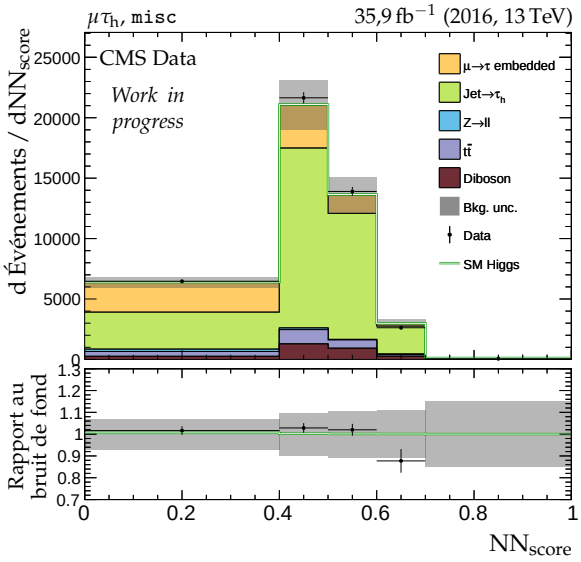
(c) *Catégorie* `ttbar`.



(d) *Catégorie* `zll`.

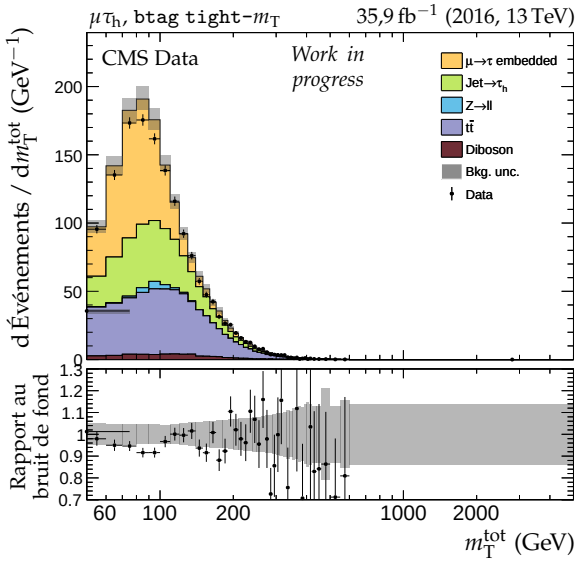


(e) *Catégorie* `fake`.

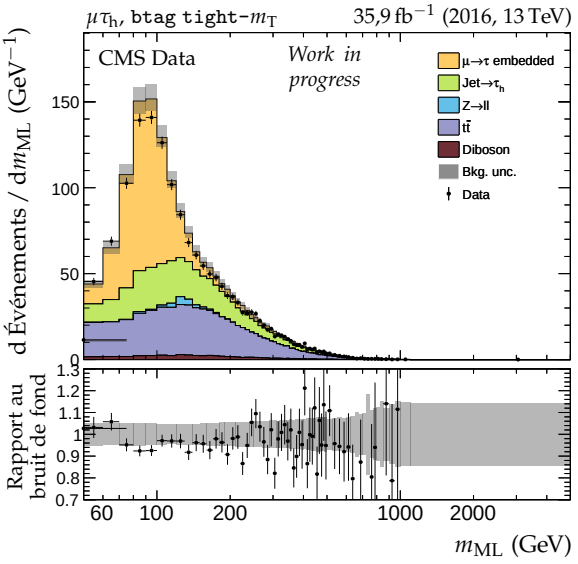


(f) *Catégorie* `misc`.

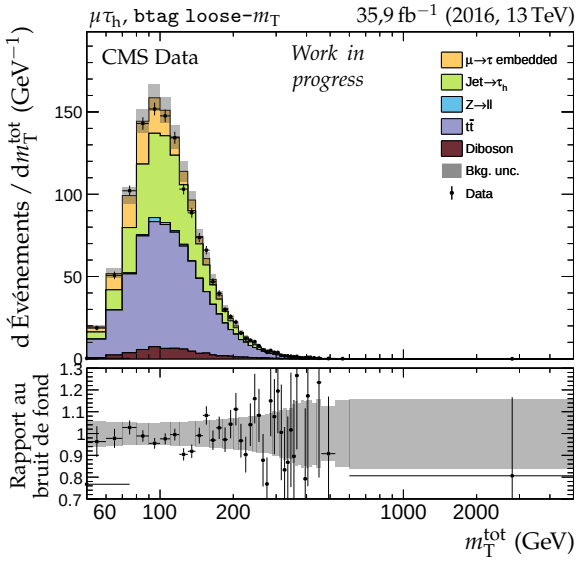
Figure H.3 – *Distributions de* $\mathrm{NN_{score}}$ *en 2016 dans le canal* $\mu\tau_\mathrm{h}$.



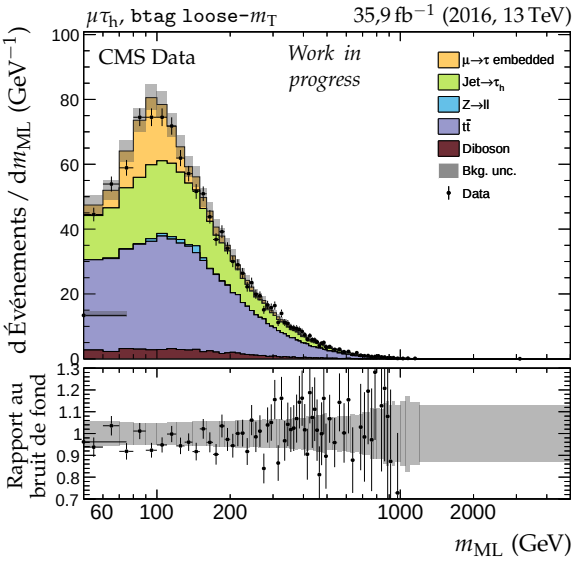
(a) *Catégorie* `btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `btag tight-`m_{T}, m_{ML}.

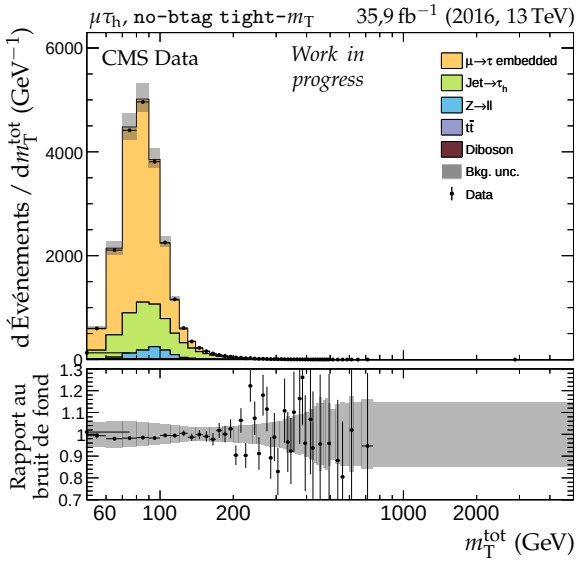


(c) *Catégorie* `btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

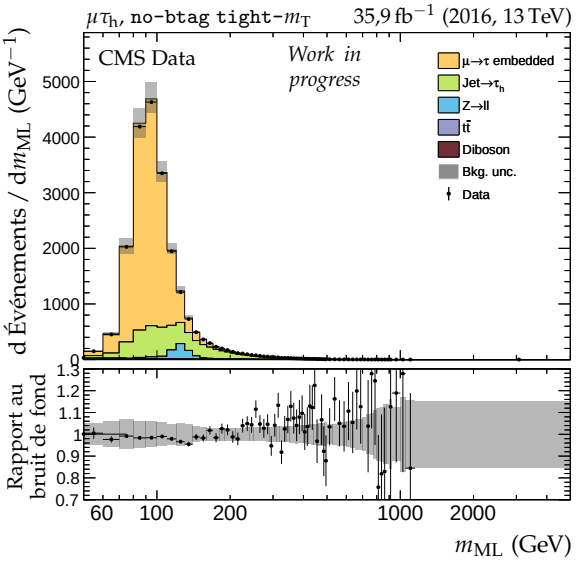


(d) *Catégorie* `btag loose-`m_{T}, m_{ML}.

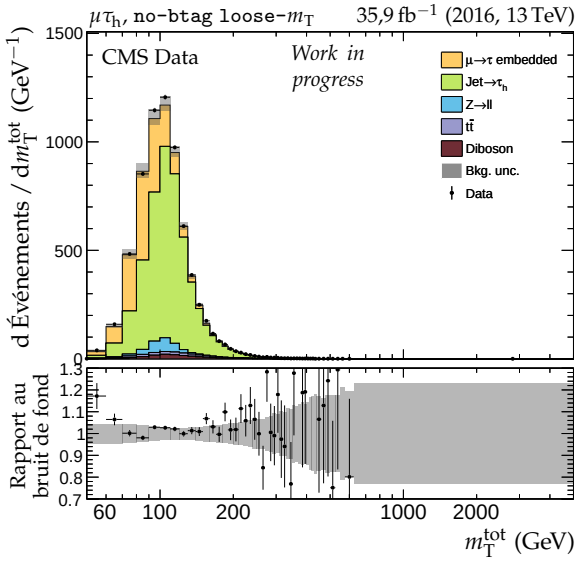
Figure H.4 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2016 dans le canal* $\mu\tau_{\mathrm{h}}$*, catégories* `btag`.



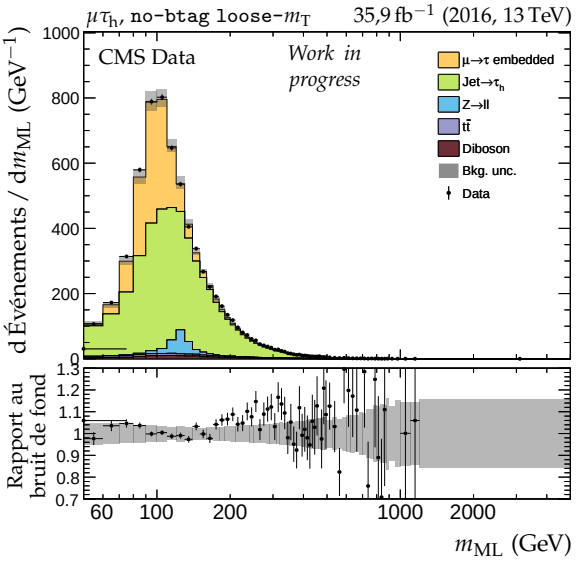
(a) *Catégorie* `no-btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `no-btag tight-`m_{T}, m_{ML}.

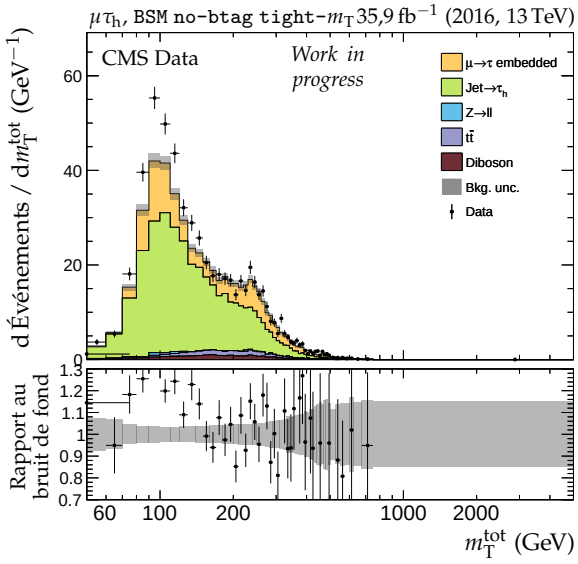


(c) *Catégorie* `no-btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

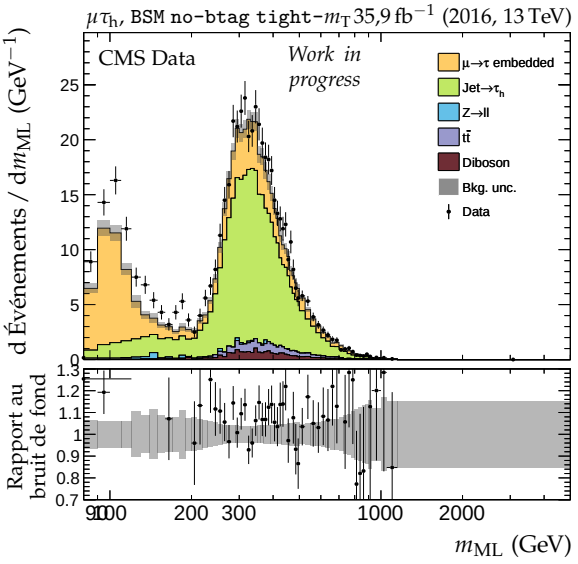


(d) *Catégorie* `no-btag loose-`m_{T}, m_{ML}.

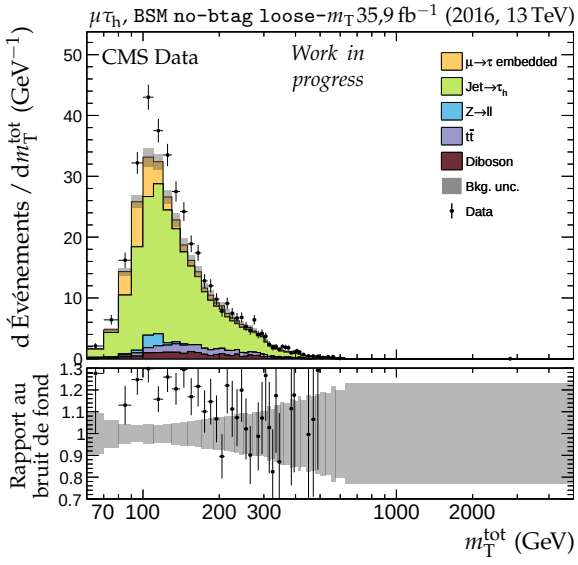
Figure H.5 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2016 dans le canal* $\mu\tau_{\mathrm{h}}$*, catégories* `no-btag`.



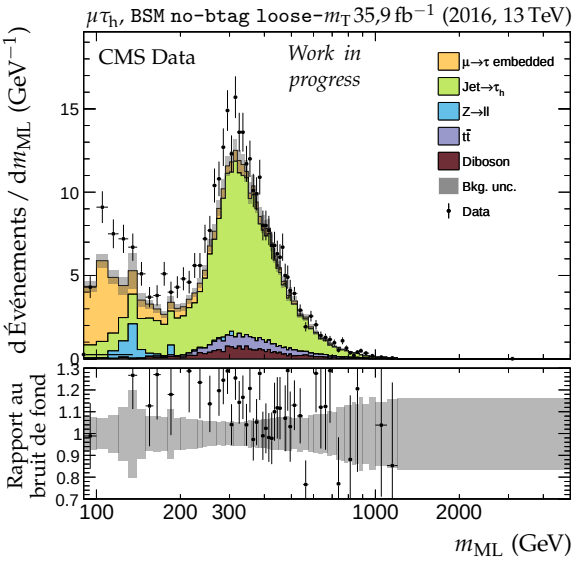
(a) *Catégorie* BSM `no-btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* BSM `no-btag tight-`m_{T}, m_{ML}.

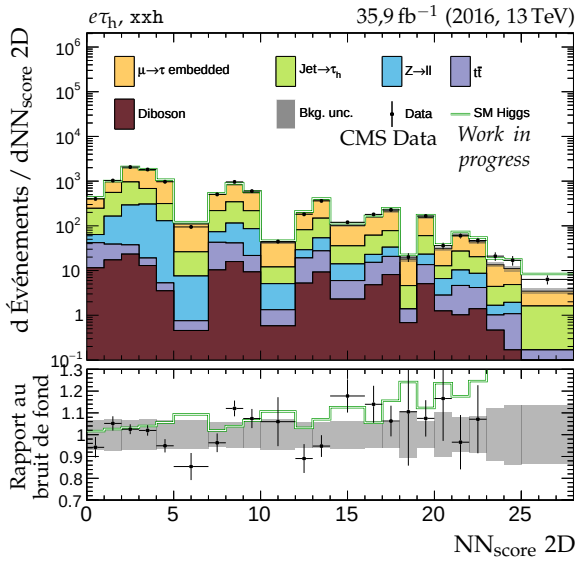


(c) *Catégorie* BSM `no-btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

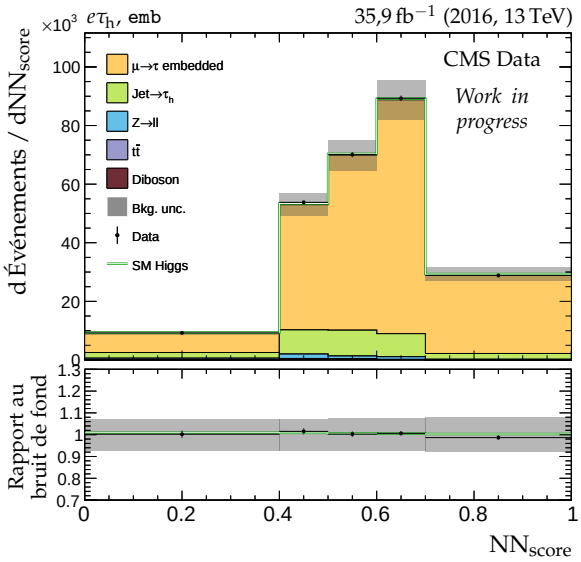


(d) *Catégorie* BSM `no-btag loose-`m_{T}, m_{ML}.

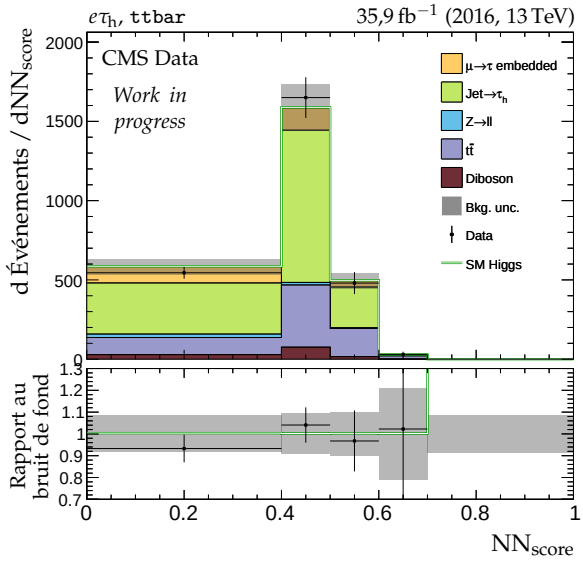
Figure H.6 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2016 dans le canal* $\mu\tau_{\mathrm{h}}$, *catégories* `no-btag` *avec* $m_{\mathrm{SVFIT}} \geqslant 250\,\mathrm{GeV}$.



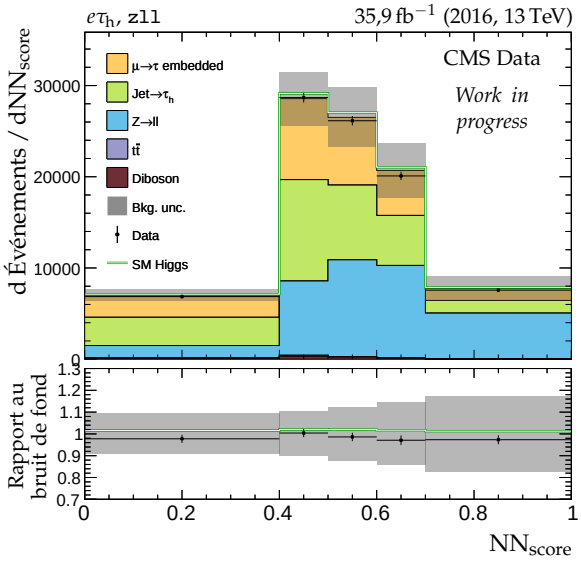
(a) *Catégorie* `xxh`.



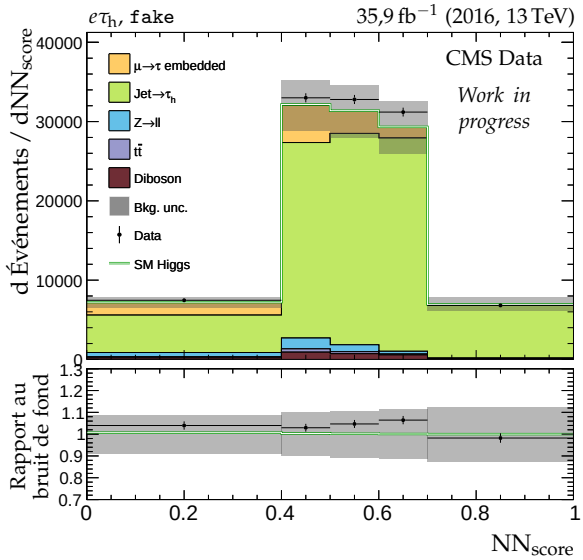
(b) *Catégorie* `emb`.



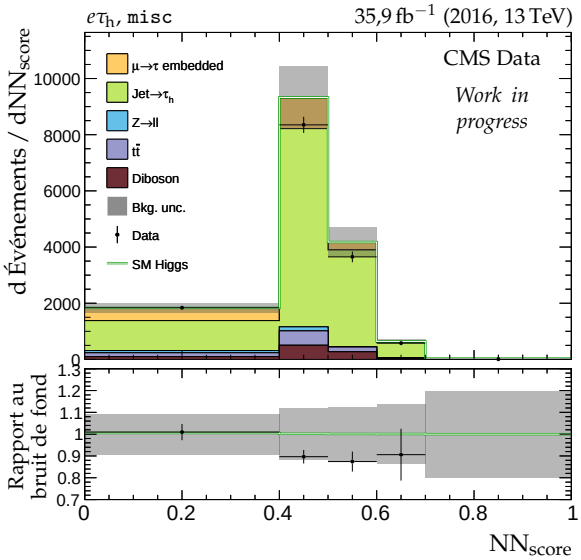
(c) *Catégorie* `ttbar`.



(d) *Catégorie* `zll`.

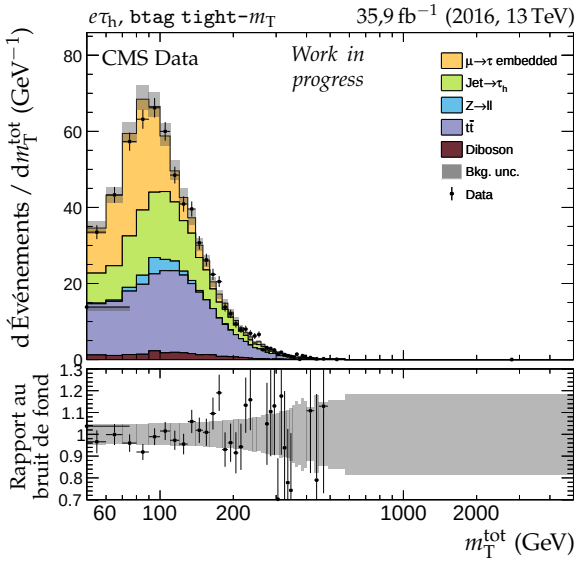


(e) *Catégorie* `fake`.

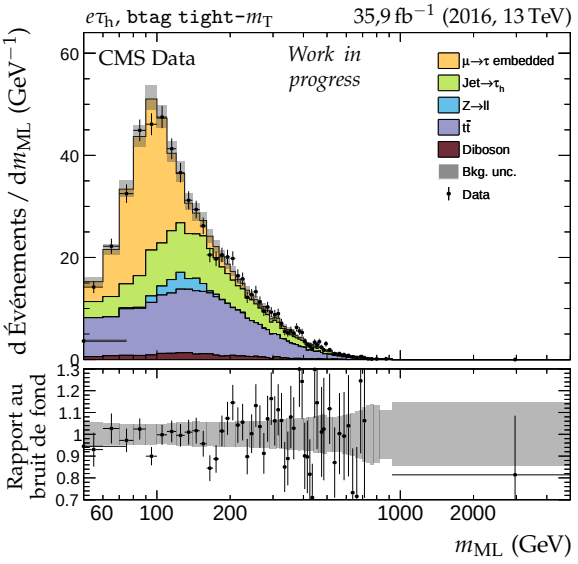


(f) *Catégorie* `misc`.

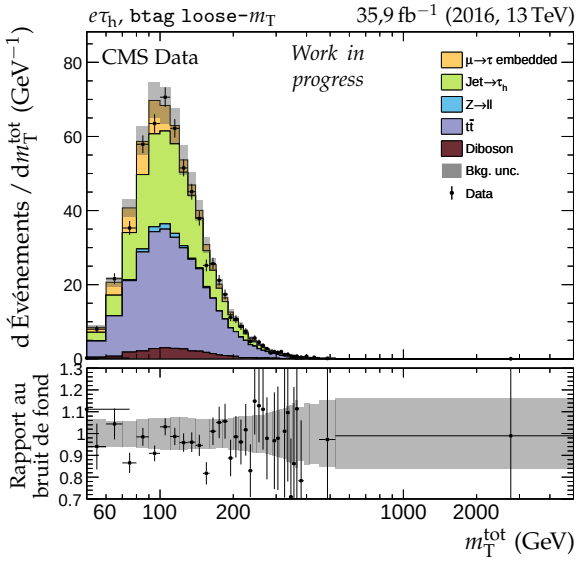
Figure H.7 – *Distributions de* $\mathrm{NN}_{\mathrm{score}}$ *en 2016 dans le canal* $e\tau_{\mathrm{h}}$.



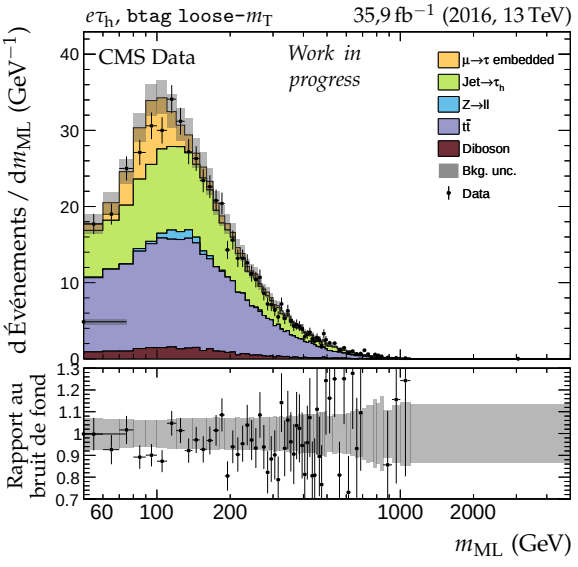
(a) *Catégorie* `btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `btag tight-`m_{T}, m_{ML}.

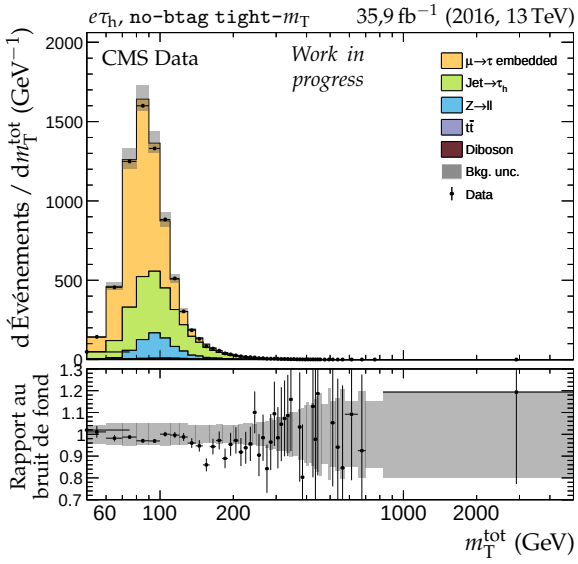


(c) *Catégorie* `btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

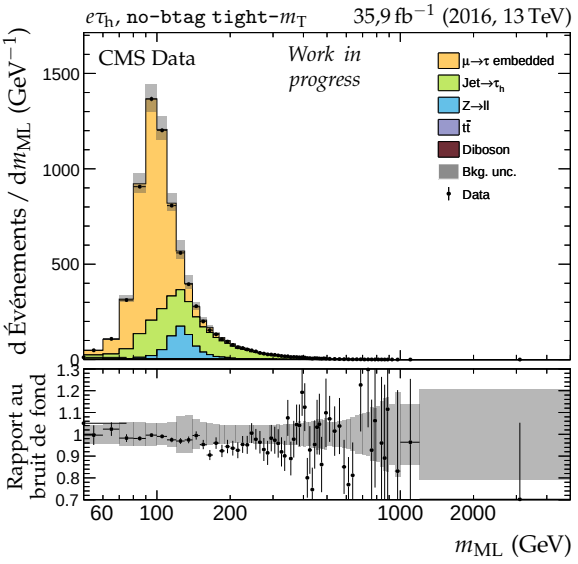


(d) *Catégorie* `btag loose-`m_{T}, m_{ML}.

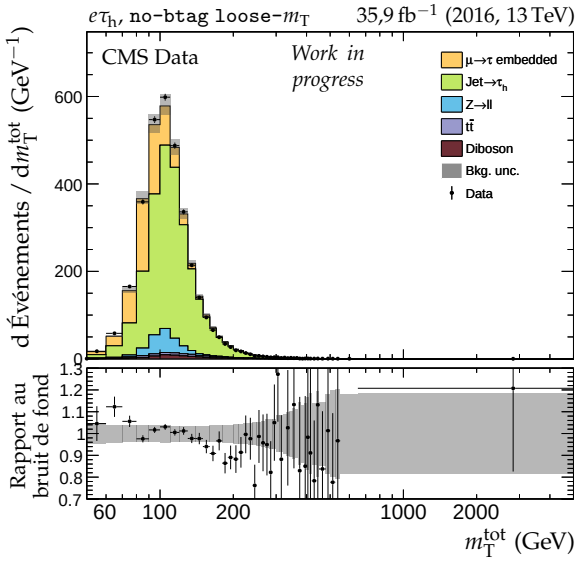
Figure H.8 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2016 dans le canal* $e\tau_{\mathrm{h}}$*, catégories* `btag`.



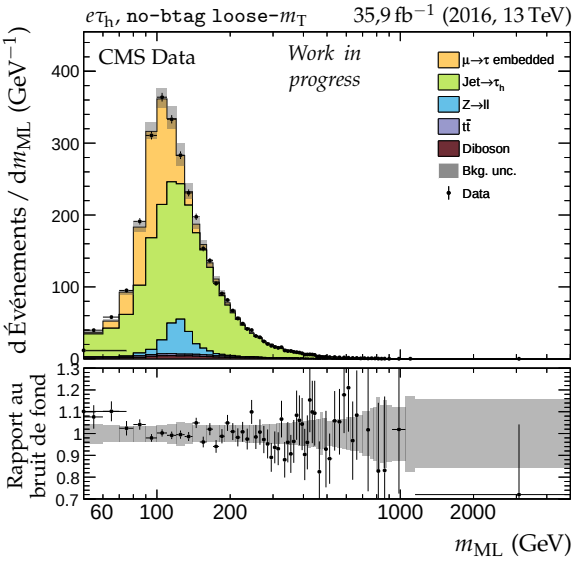
(a) *Catégorie* `no-btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `no-btag tight-`m_{T}, m_{ML}.

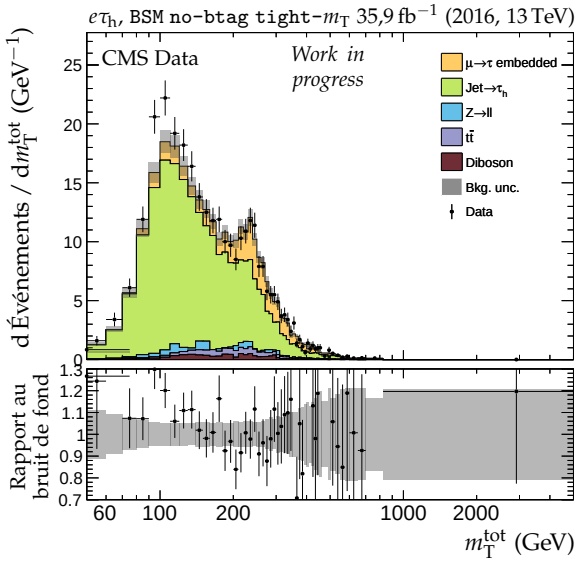


(c) *Catégorie* `no-btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

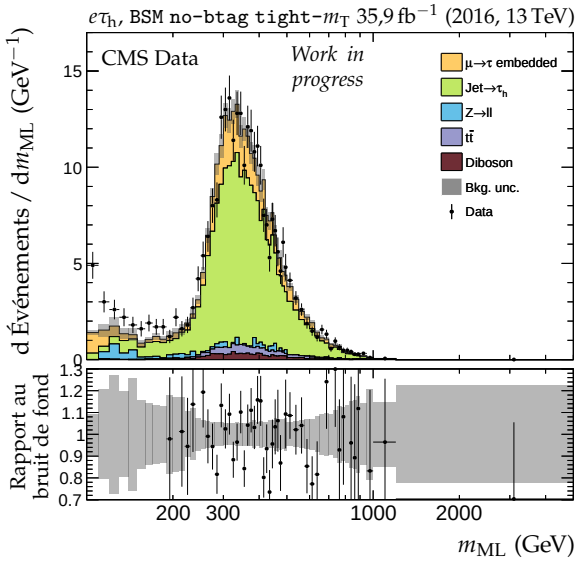


(d) *Catégorie* `no-btag loose-`m_{T}, m_{ML}.

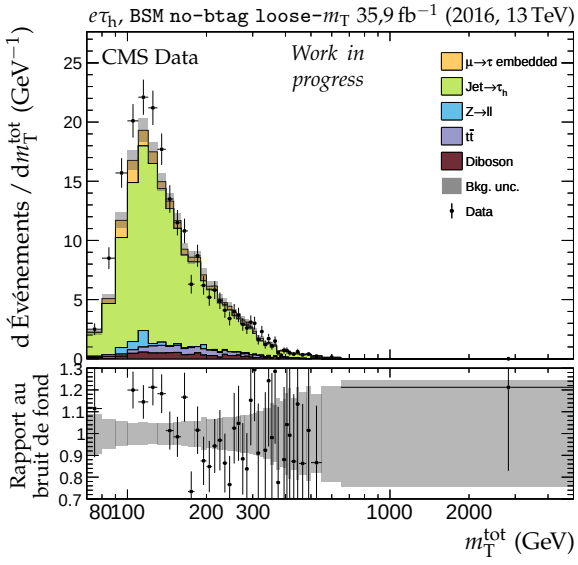
Figure H.9 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2016 dans le canal* $e\tau_{\mathrm{h}}$, *catégories* `no-btag`.



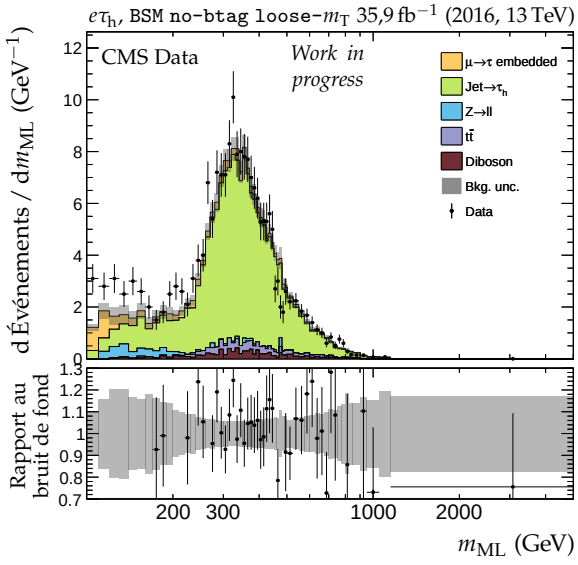
(a) *Catégorie* BSM `no-btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* BSM `no-btag tight-`m_{T}, m_{ML}.

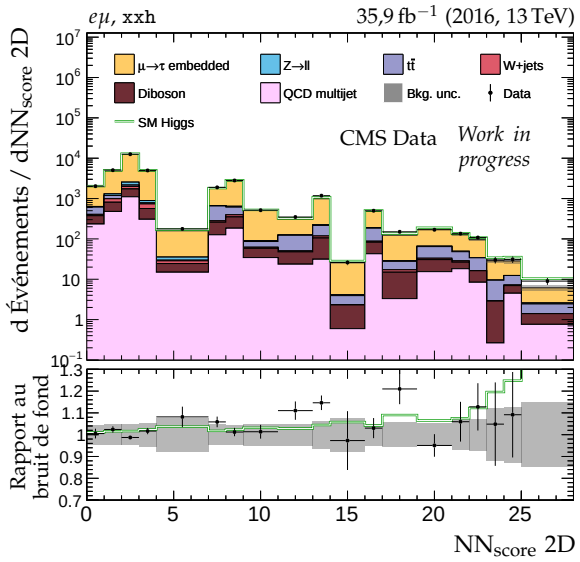


(c) *Catégorie* BSM `no-btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

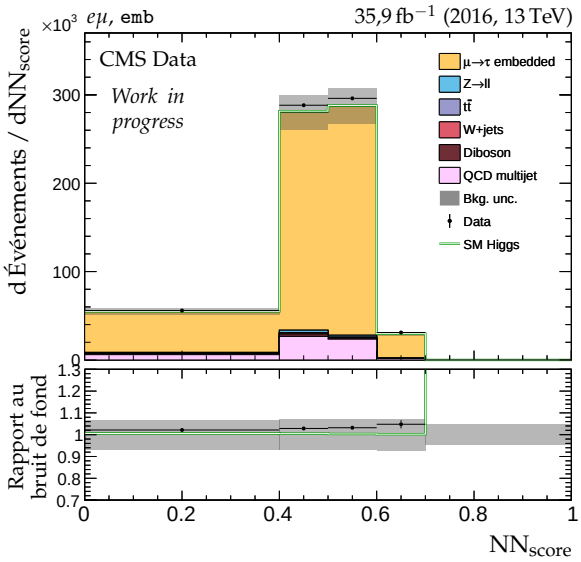


(d) *Catégorie* BSM `no-btag loose-`m_{T}, m_{ML}.

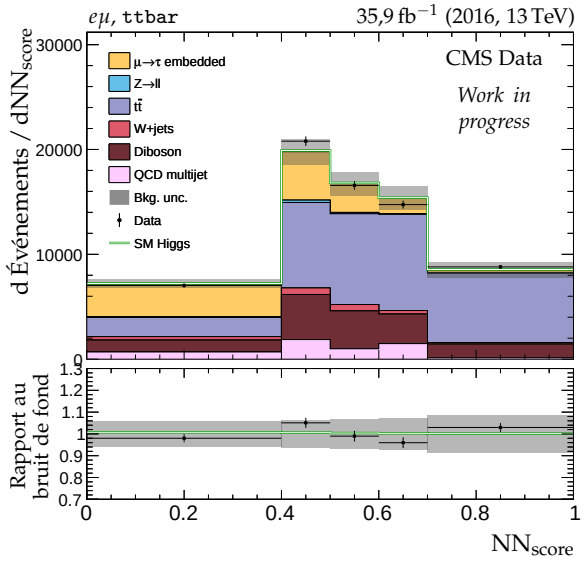
Figure H.10 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2016 dans le canal* $e\tau_{\mathrm{h}}$, *catégories* `no-btag` *avec* $m_{\mathrm{SV\,FIT}} \geqslant 250\,\mathrm{GeV}$.



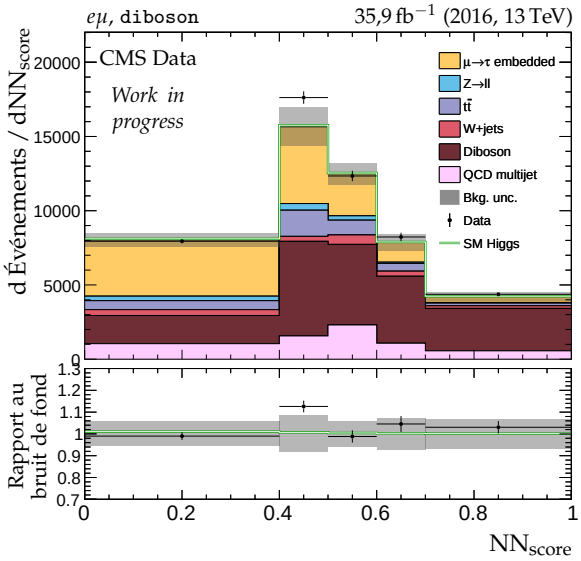
(a) *Catégorie* `xxh`.



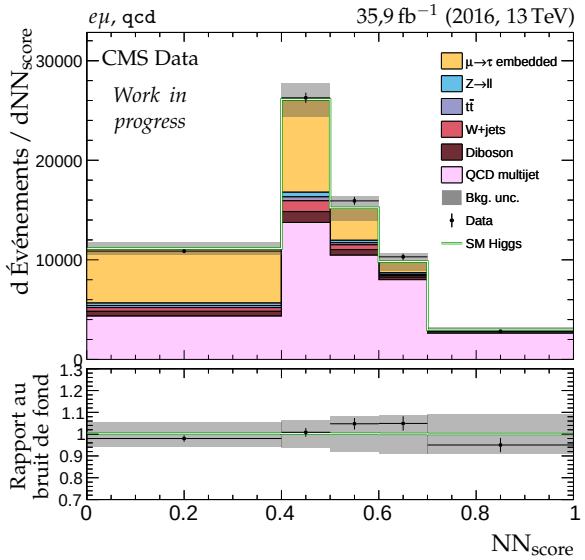
(b) *Catégorie* `emb`.



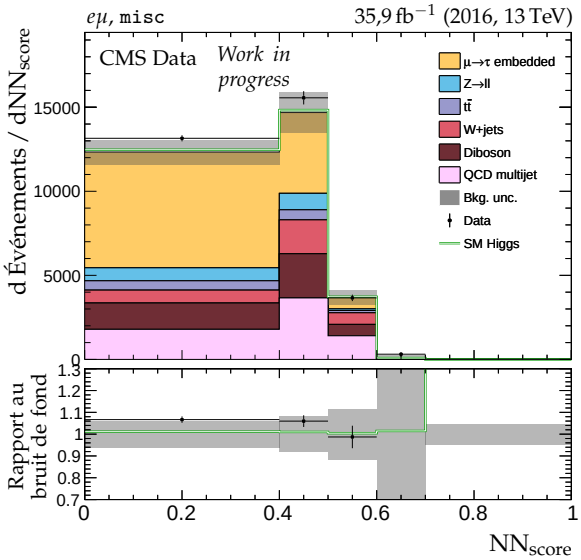
(c) *Catégorie* `ttbar`.



(d) *Catégorie* `diboson`.

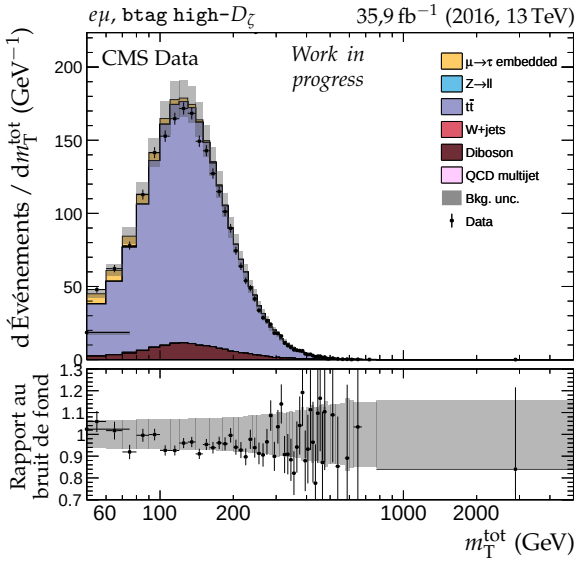


(e) *Catégorie* `qcd`.

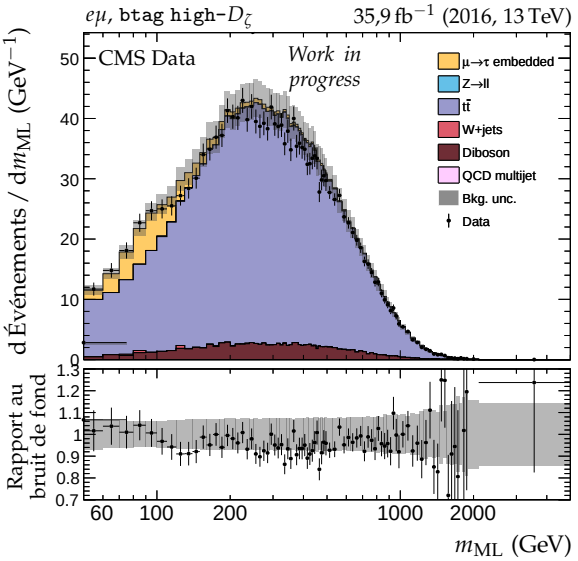


(f) *Catégorie* `misc`.

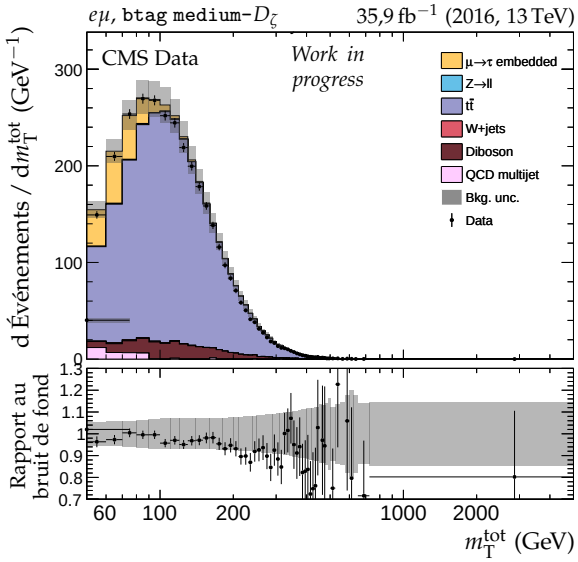
Figure H.11 – *Distributions de* NN$_{score}$ *en 2016 dans le canal eμ.*



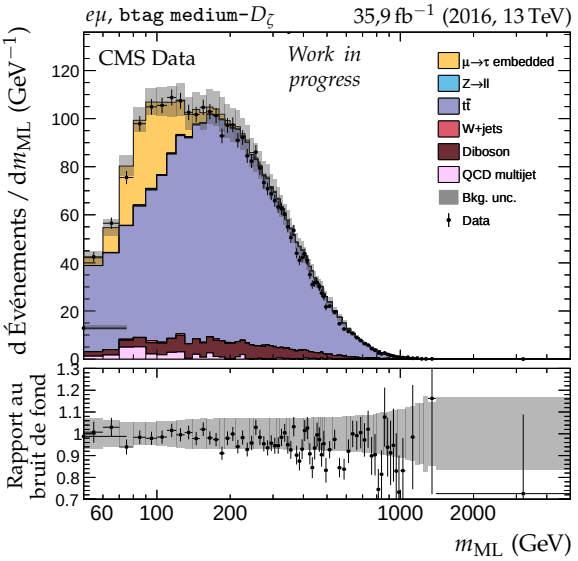
(a) *Catégorie* `btag high-`D_ζ, m_T^{tot}.



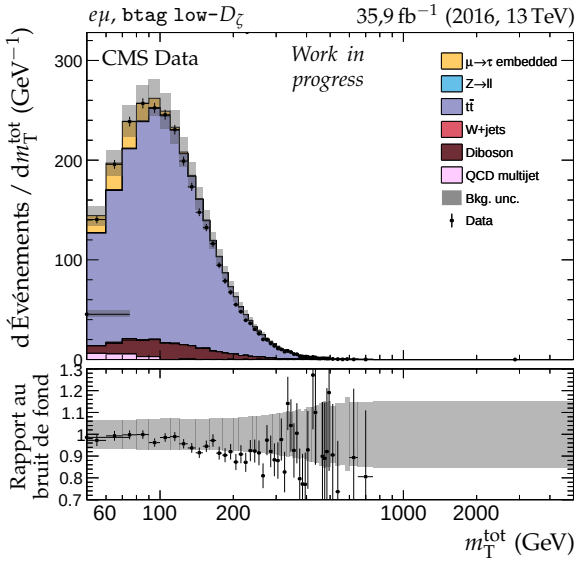
(b) *Catégorie* `btag high-`D_ζ, m_{ML}.



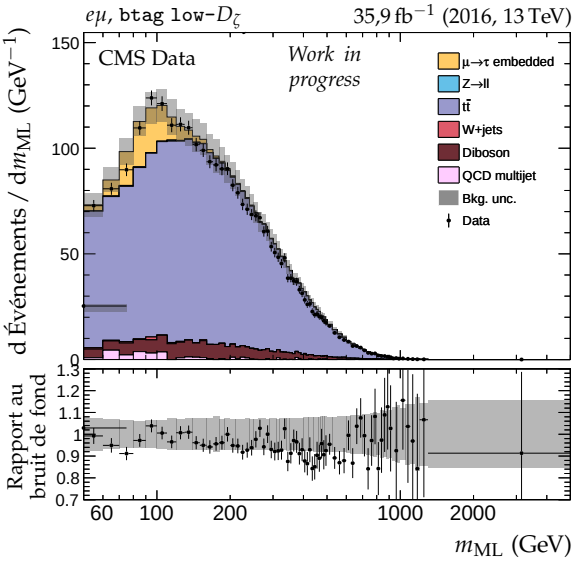
(c) *Catégorie* `btag medium-`D_ζ, m_T^{tot}.



(d) *Catégorie* `btag medium-`D_ζ, m_{ML}.

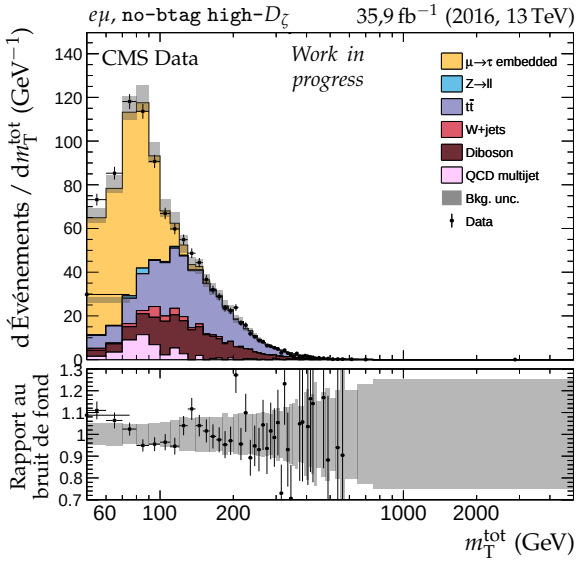


(e) *Catégorie* `btag low-`D_ζ, m_T^{tot}.

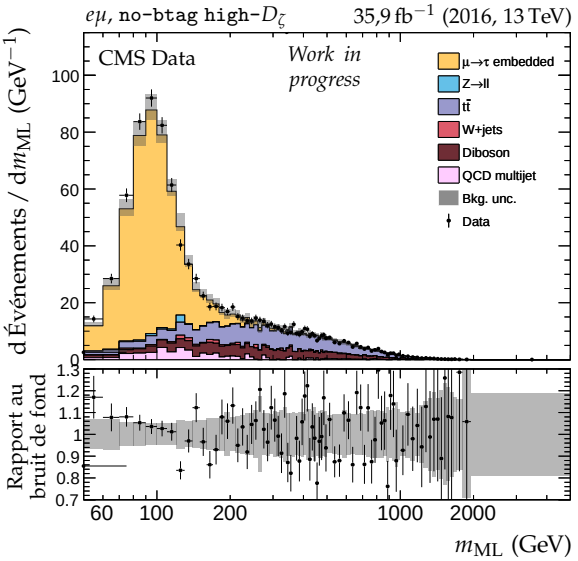


(f) *Catégorie* `btag low-`D_ζ, m_{ML}.

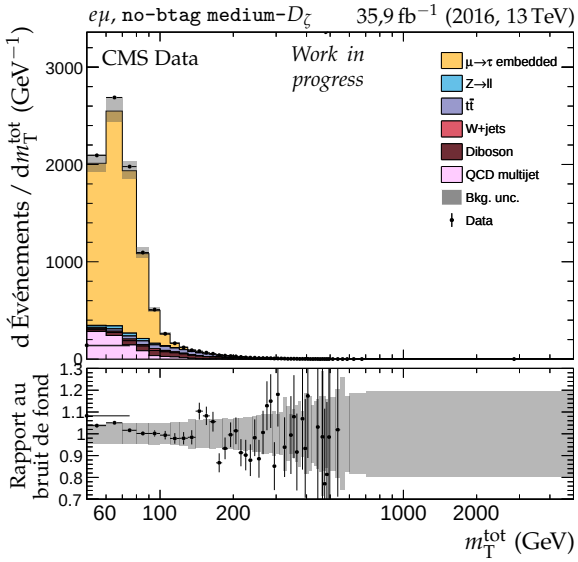
Figure H.12 – *Distributions de m_T^{tot} et m_{ML} en 2016 dans le canal eμ, catégories* `btag`.



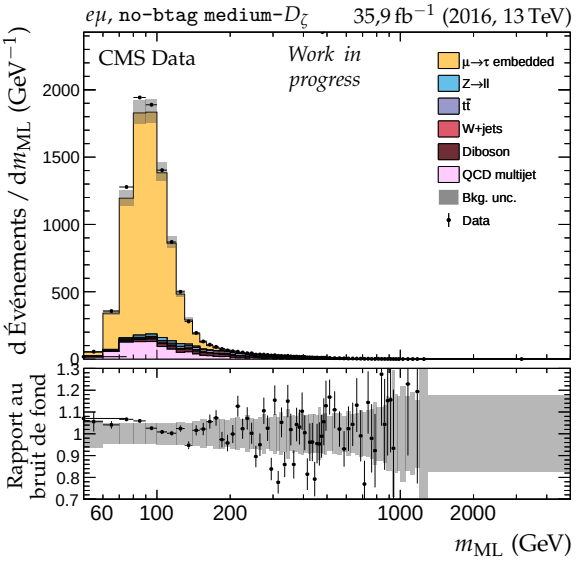
(a) *Catégorie* `no-btag high-`D_ζ, m_T^{tot}.



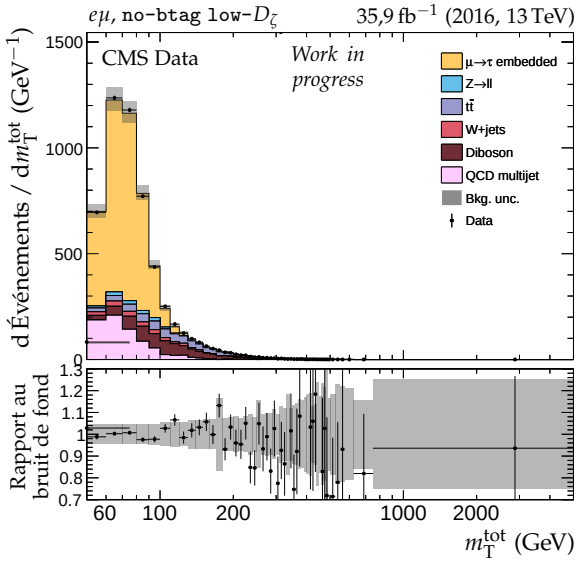
(b) *Catégorie* `no-btag high-`D_ζ, m_{ML}.



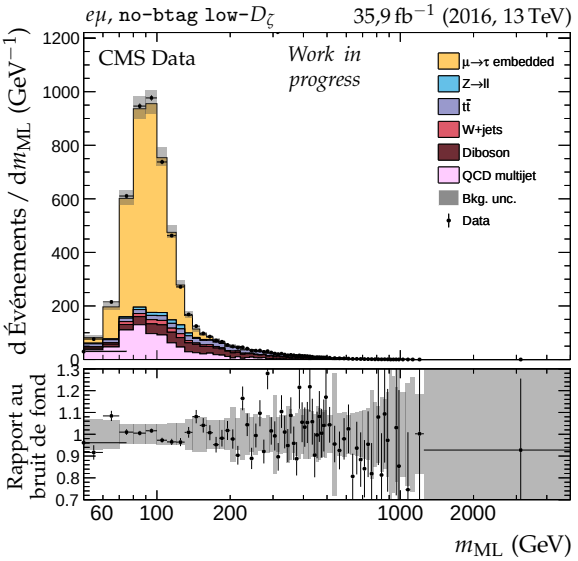
(c) *Catégorie* `no-btag medium-`D_ζ, m_T^{tot}.



(d) *Catégorie* `no-btag medium-`D_ζ, m_{ML}.

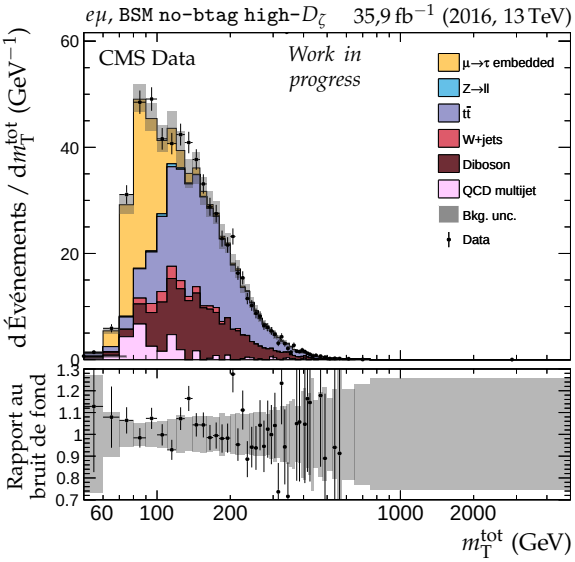


(e) *Catégorie* `no-btag low-`D_ζ, m_T^{tot}.

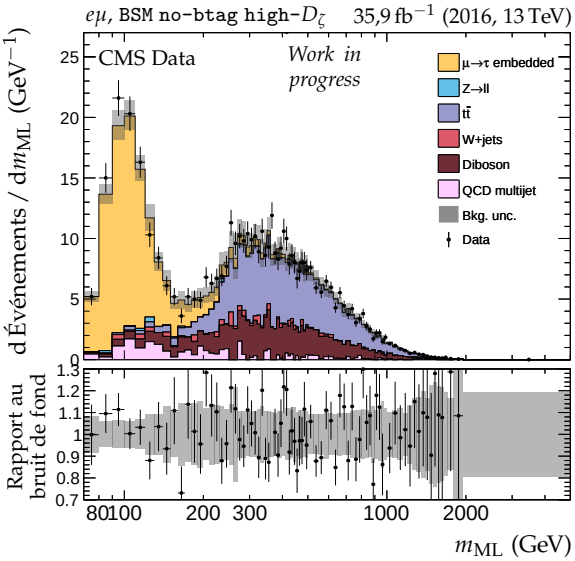


(f) *Catégorie* `no-btag low-`D_ζ, m_{ML}.

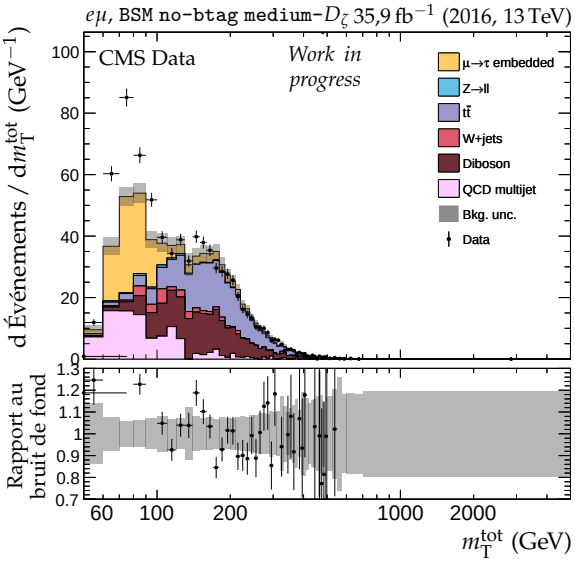
Figure H.13 – *Distributions de m_T^{tot} et m_{ML} en 2016 dans le canal eμ, catégories* `no-btag`.



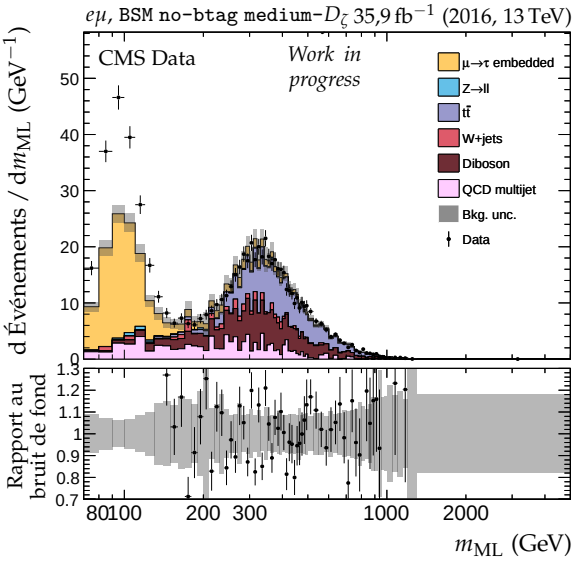
(a) *Catégorie BSM* `no-btag high-`D_ζ, m_T^{tot}.



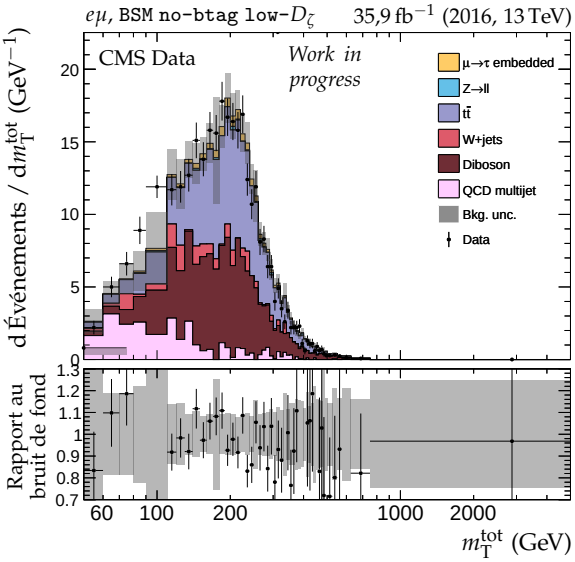
(b) *Catégorie BSM* `no-btag high-`D_ζ, m_{ML}.



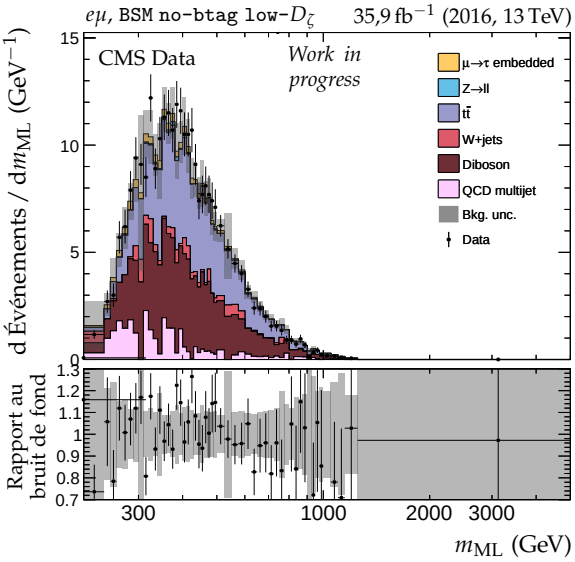
(c) *Catégorie BSM* `no-btag medium-`D_ζ, m_T^{tot}.



(d) *Catégorie BSM* `no-btag medium-`D_ζ, m_{ML}.

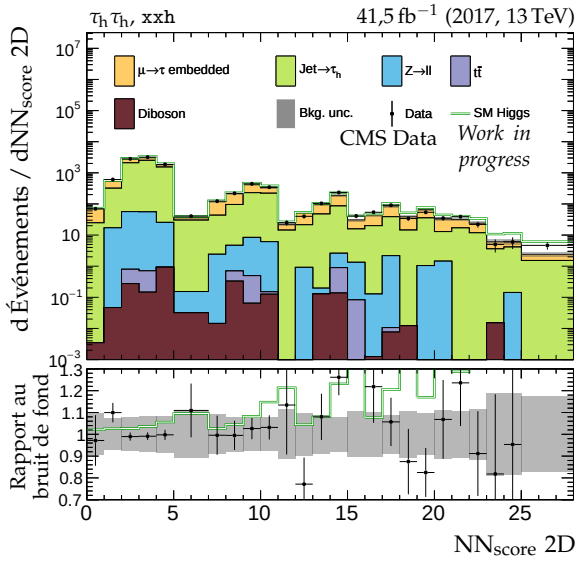


(e) *Catégorie BSM* `no-btag low-`D_ζ, m_T^{tot}.

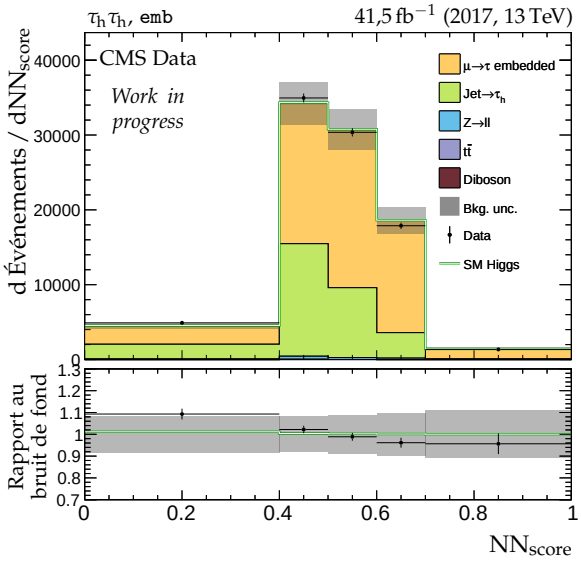


(f) *Catégorie BSM* `no-btag low-`D_ζ, m_{ML}.

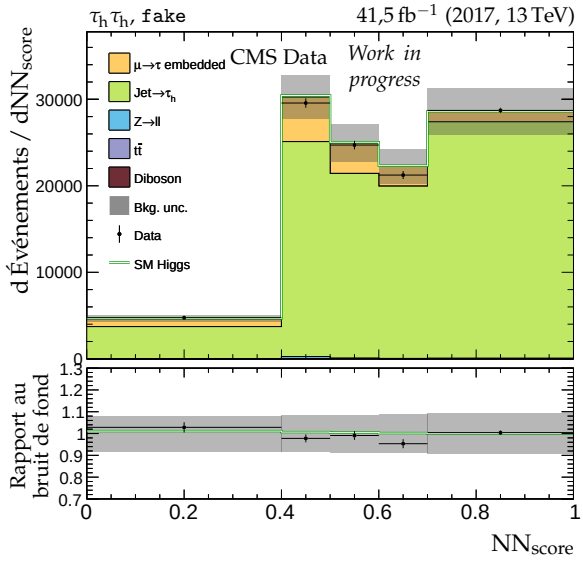
Figure H.14 – *Distributions de* m_T^{tot} *et* m_{ML} *en 2016 dans le canal eμ, catégories* `no-btag` *avec* $m_{SVFIT} \geqslant 250\,\text{GeV}$.



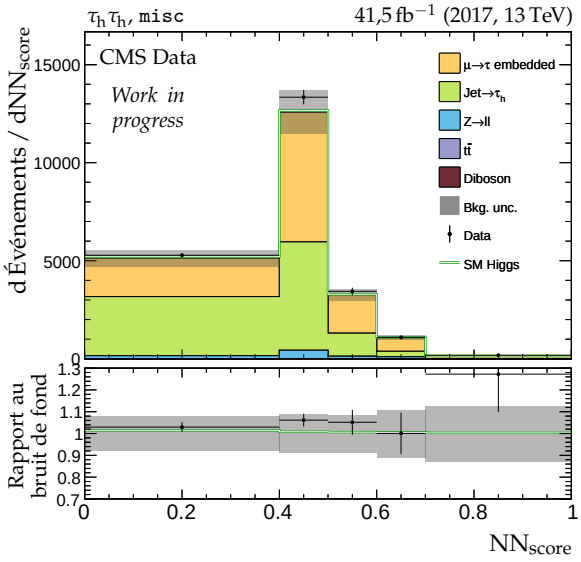
(a) *Catégorie* `xxh`.



(b) *Catégorie* `emb`.

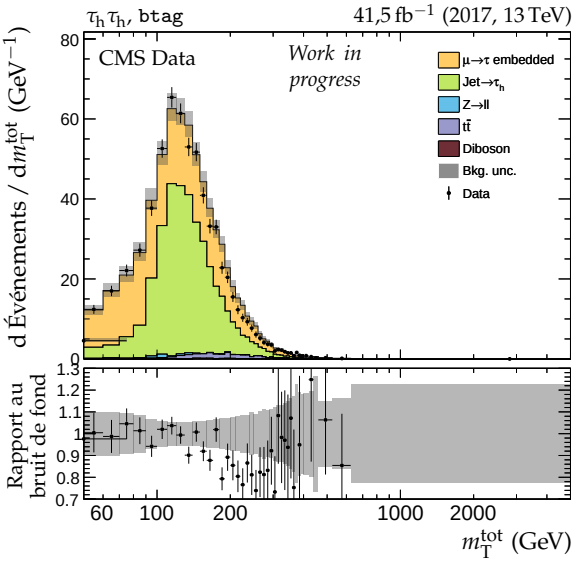


(c) *Catégorie* `fake`.

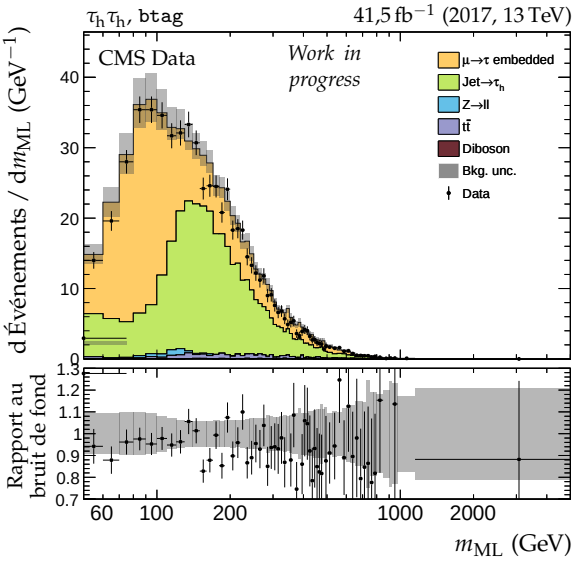


(d) *Catégorie* `misc`.

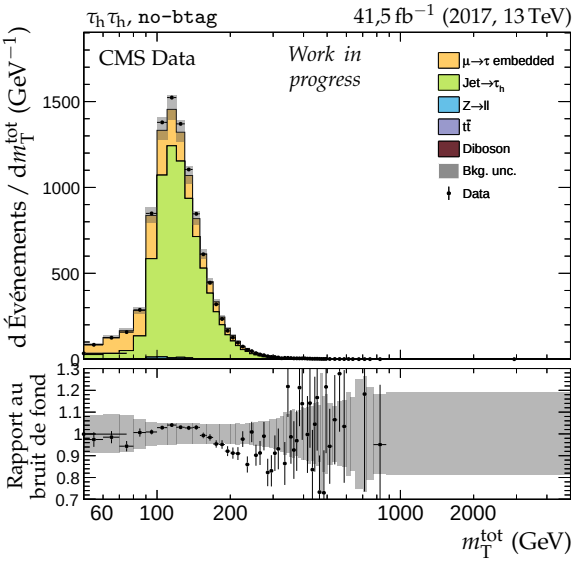
Figure H.15 – *Distributions de* $\mathrm{NN_{score}}$ *en 2017 dans le canal* $\tau_h\tau_h$.



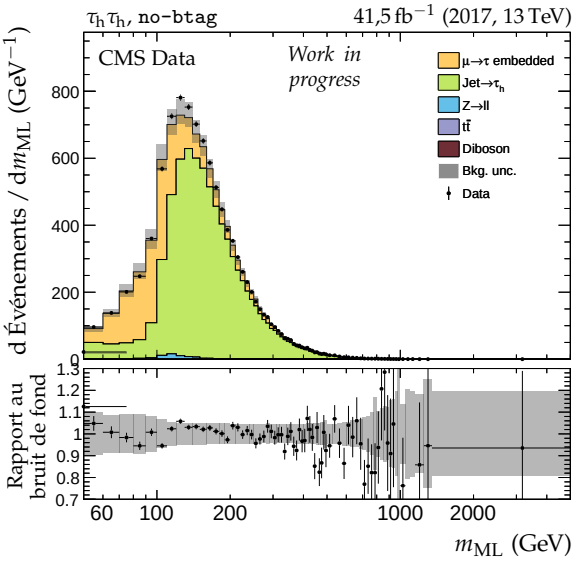
(a) *Catégorie* `btag`, m_T^{tot}.



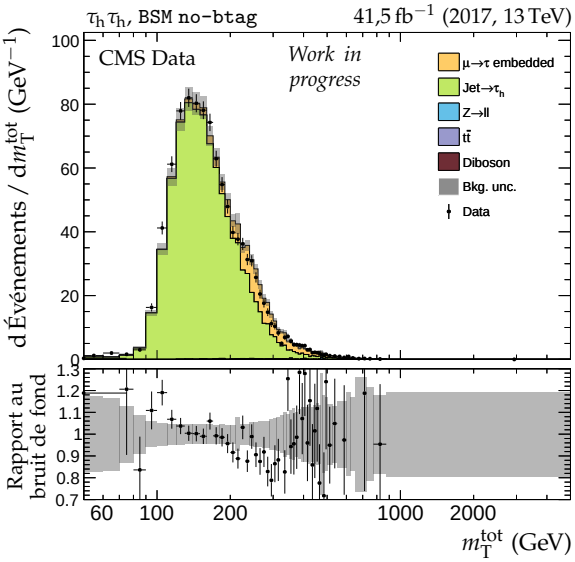
(b) *Catégorie* `btag`, m_{ML}.



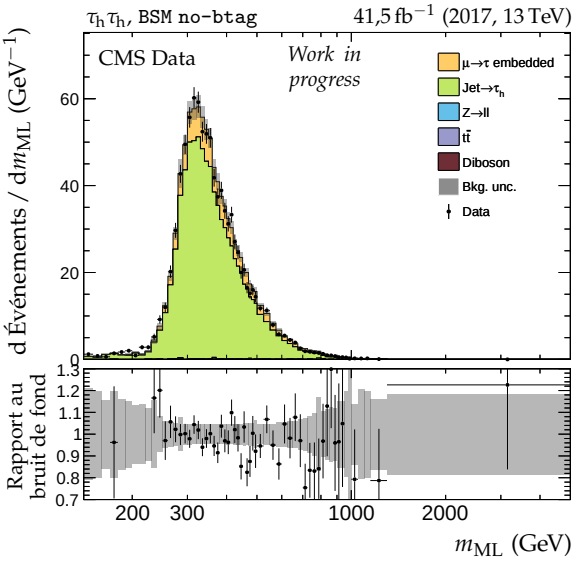
(c) *Catégorie* `no-btag`, m_T^{tot}.



(d) *Catégorie* `no-btag`, m_{ML}.

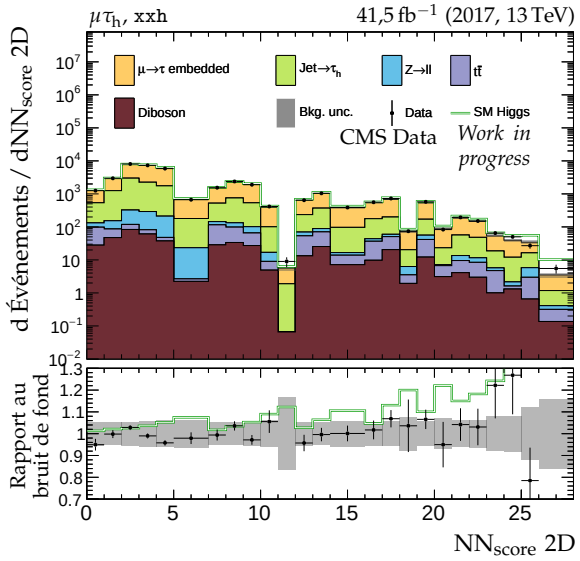


(e) *Catégorie BSM* `no-btag`, m_T^{tot}.

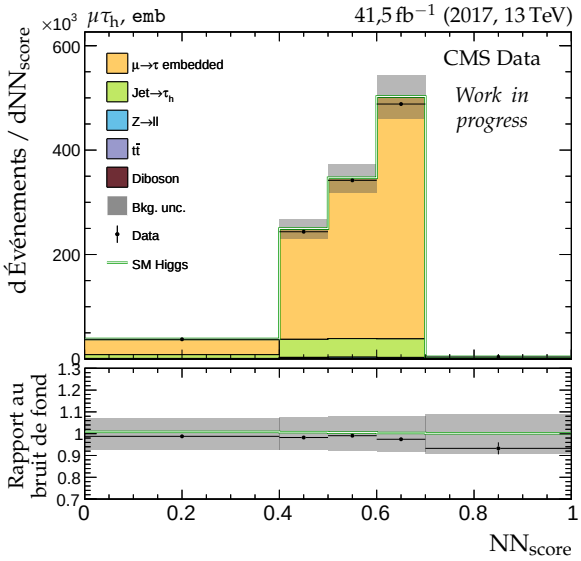


(f) *Catégorie BSM* `no-btag`, m_{ML}.

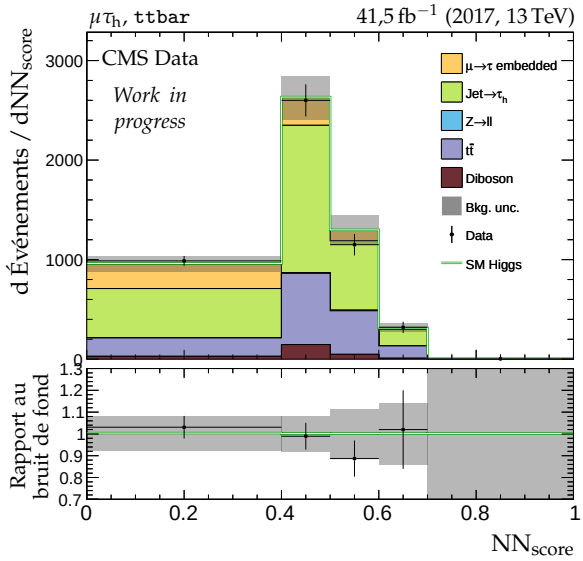
Figure H.16 – *Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal $\tau_h\tau_h$.*



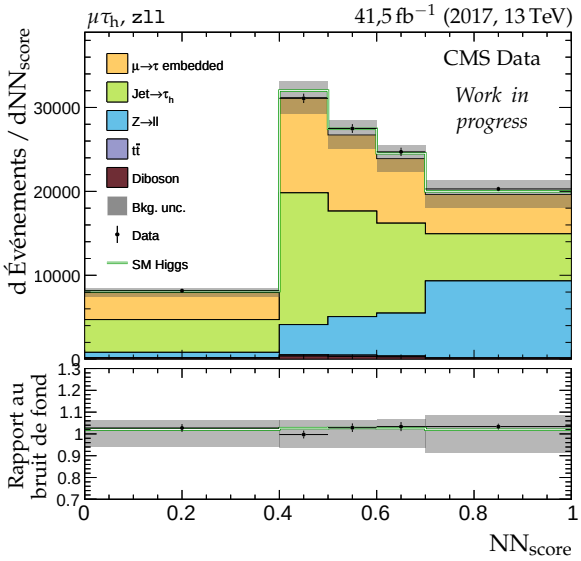
(a) *Catégorie* `xxh`.



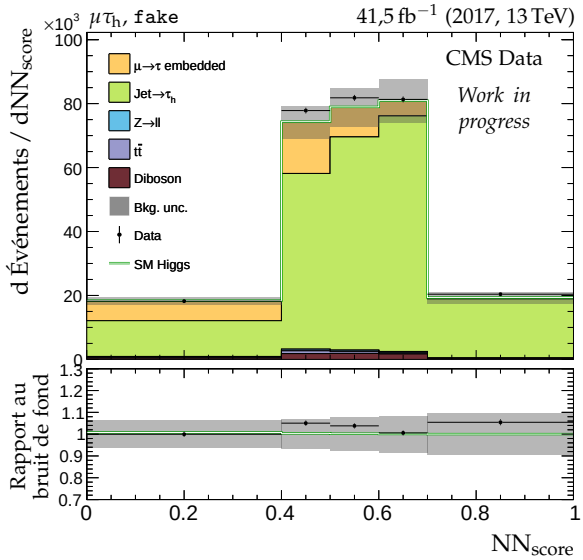
(b) *Catégorie* `emb`.



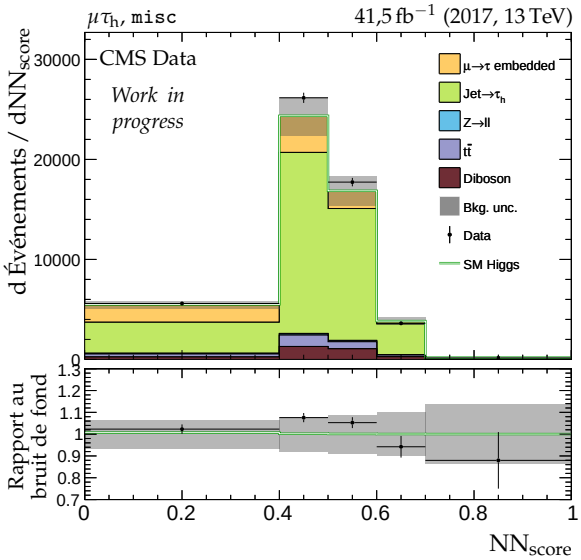
(c) *Catégorie* `ttbar`.



(d) *Catégorie* `zll`.

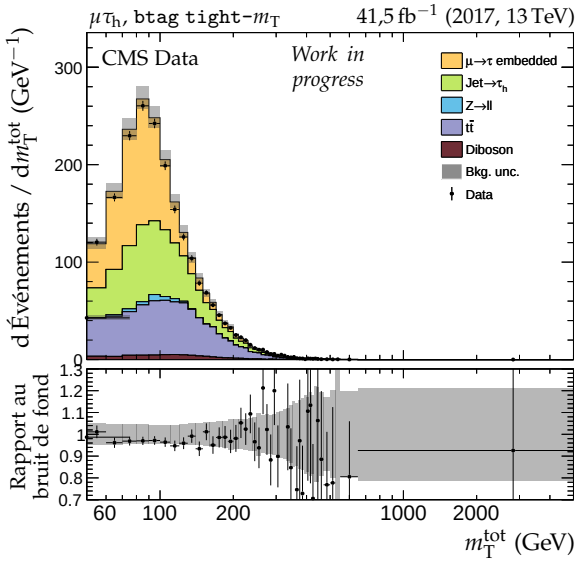


(e) *Catégorie* `fake`.

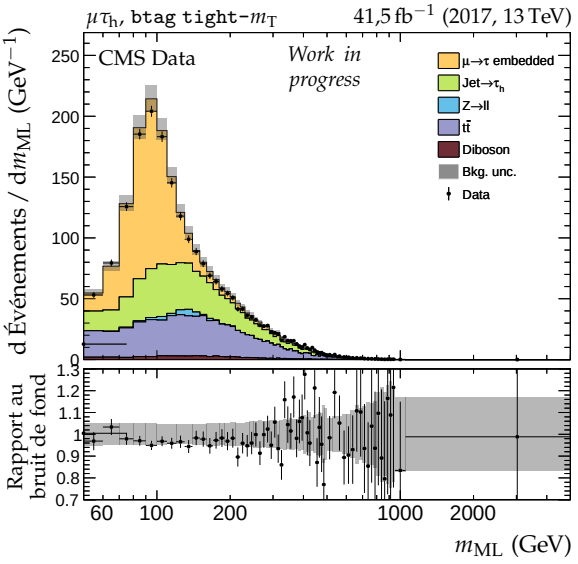


(f) *Catégorie* `misc`.

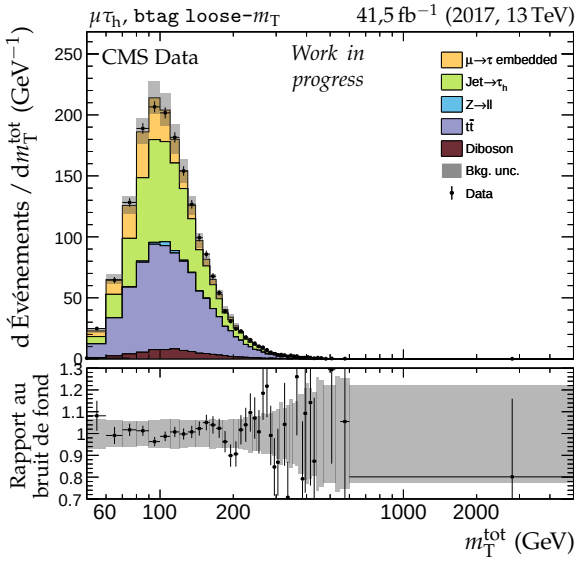
Figure H.17 – *Distributions de* $\mathrm{NN_{score}}$ *en 2017 dans le canal* $\mu\tau_h$.



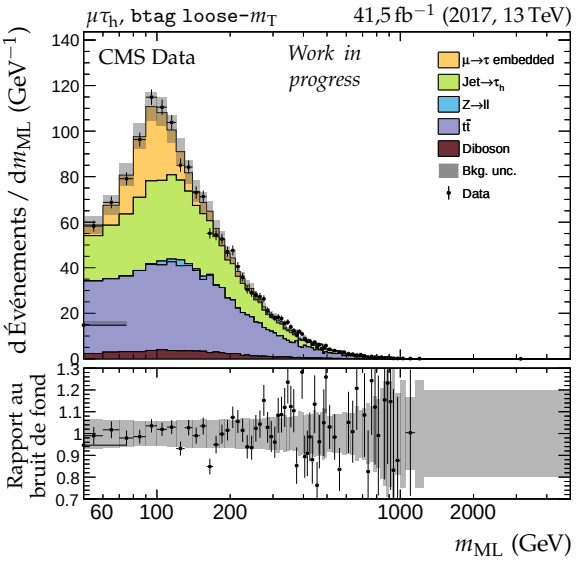
(a) *Catégorie* `btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `btag tight-`m_{T}, m_{ML}.

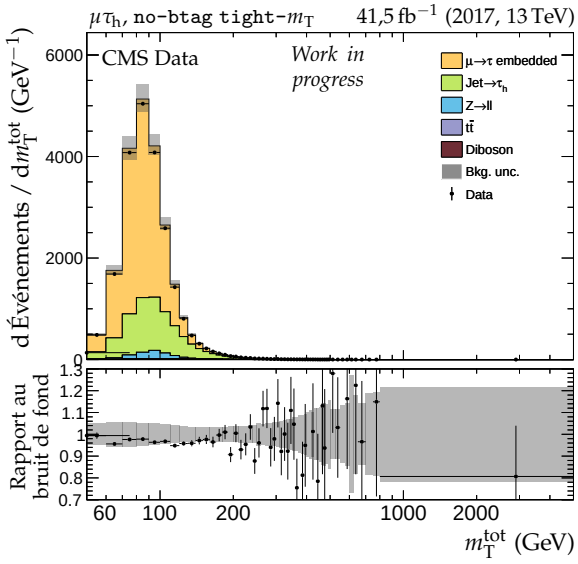


(c) *Catégorie* `btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

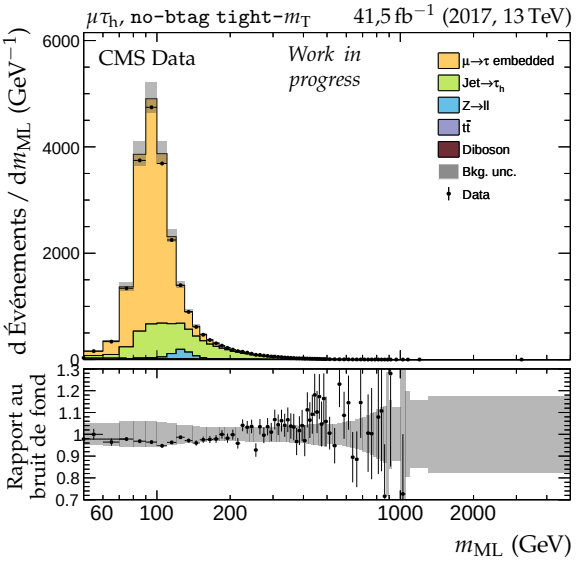


(d) *Catégorie* `btag loose-`m_{T}, m_{ML}.

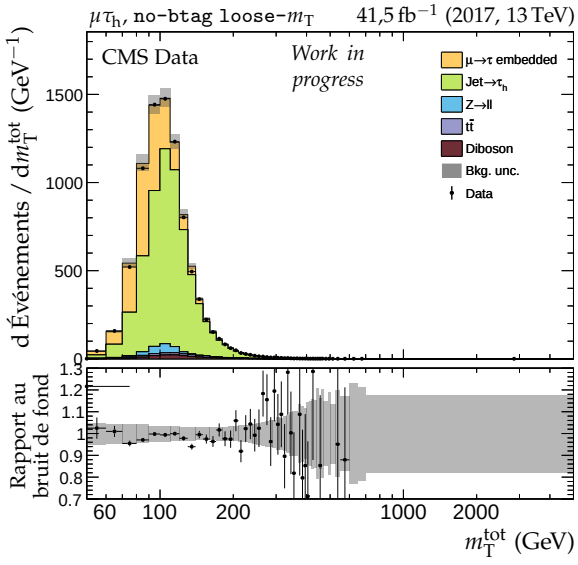
Figure H.18 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2017 dans le canal* $\mu\tau_{\mathrm{h}}$, *catégories* `btag`.



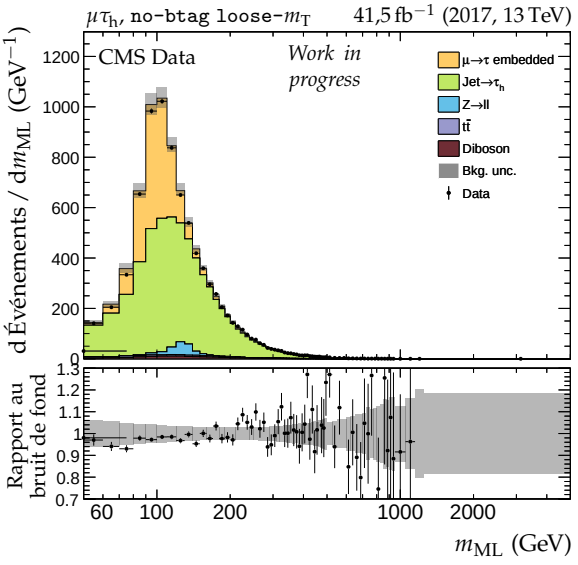
(a) *Catégorie* `no-btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `no-btag tight-`m_{T}, m_{ML}.

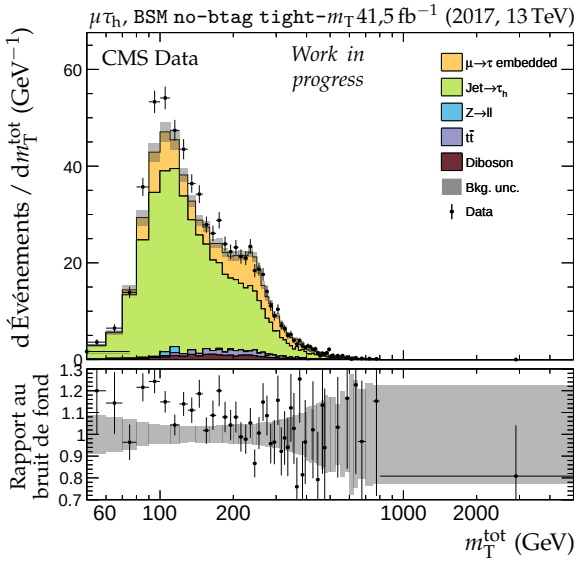


(c) *Catégorie* `no-btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

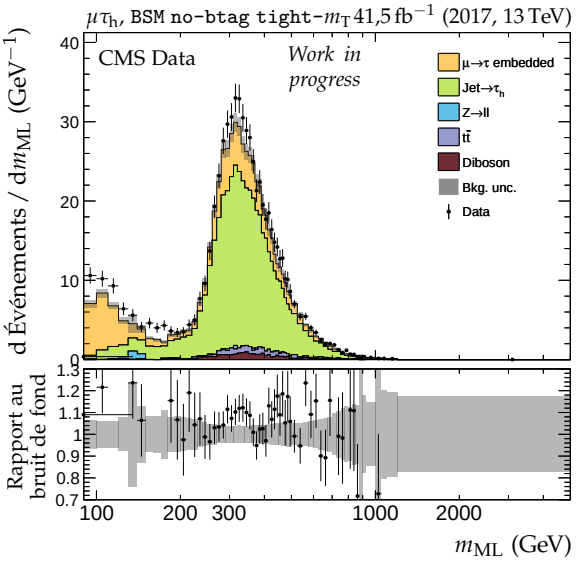


(d) *Catégorie* `no-btag loose-`m_{T}, m_{ML}.

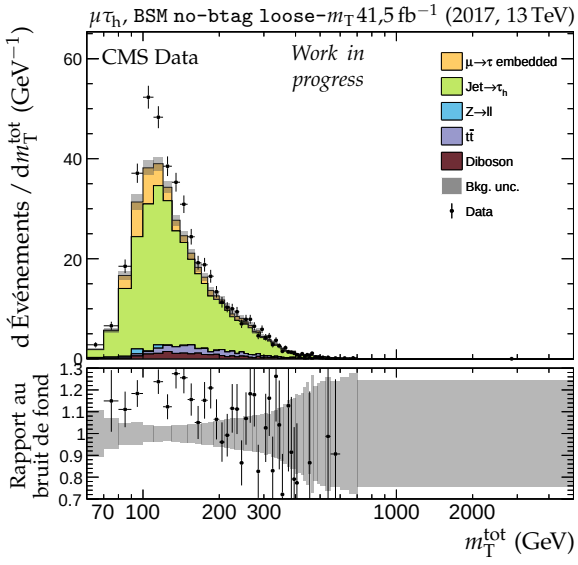
Figure H.19 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2017 dans le canal* $\mu\tau_{\mathrm{h}}$, *catégories* `no-btag`.



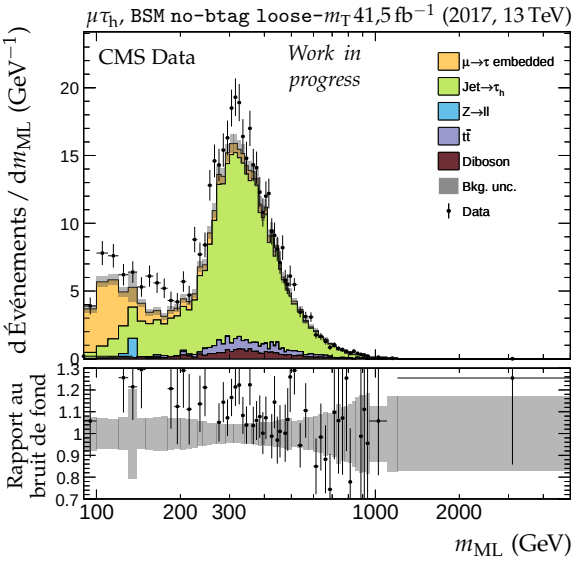
(a) *Catégorie* BSM `no-btag tight-`m_T, m_T^{tot}.



(b) *Catégorie* BSM `no-btag tight-`m_T, m_{ML}.

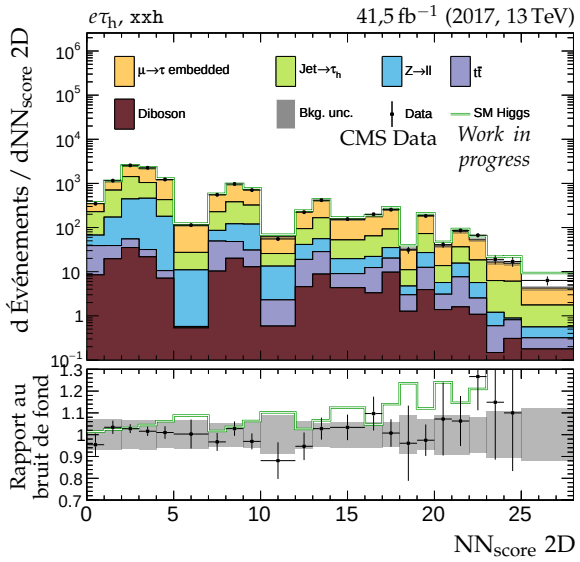


(c) *Catégorie* BSM `no-btag loose-`m_T, m_T^{tot}.

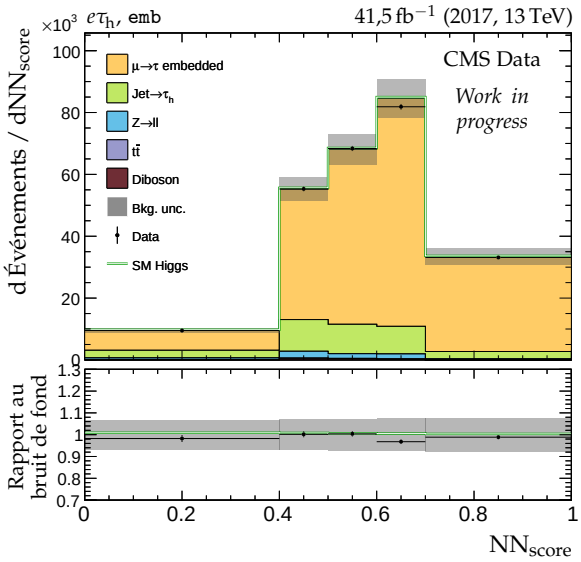


(d) *Catégorie* BSM `no-btag loose-`m_T, m_{ML}.

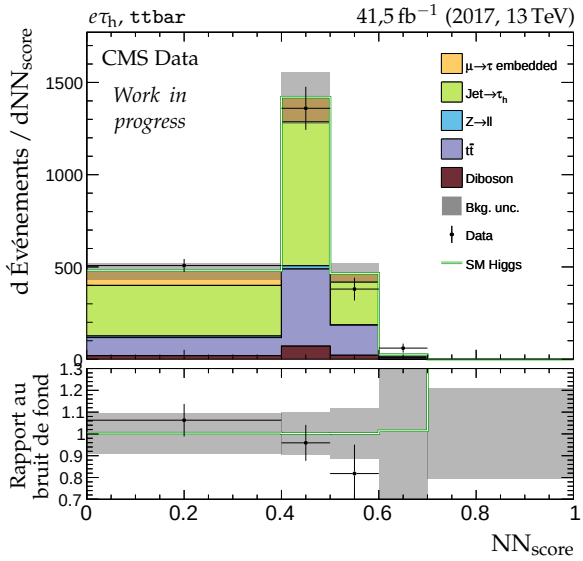
Figure H.20 – *Distributions de* m_T^{tot} *et* m_{ML} *en 2017 dans le canal* $\mu\tau_h$*, catégories* `no-btag` *avec* $m_{SVFIT} \geqslant 250\,\text{GeV}$.



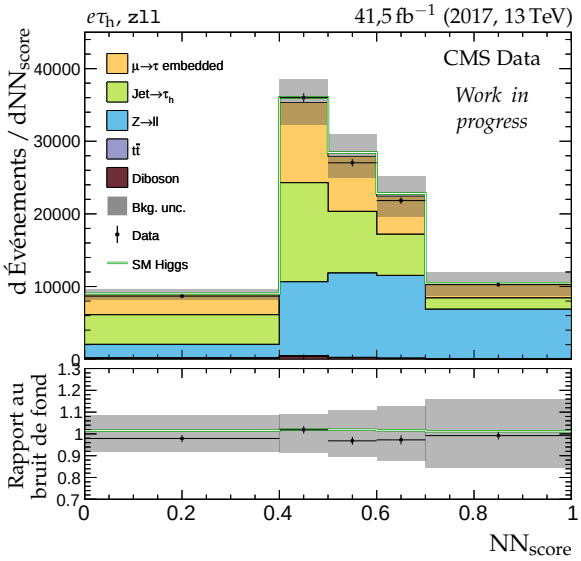
(a) *Catégorie* `xxh`.



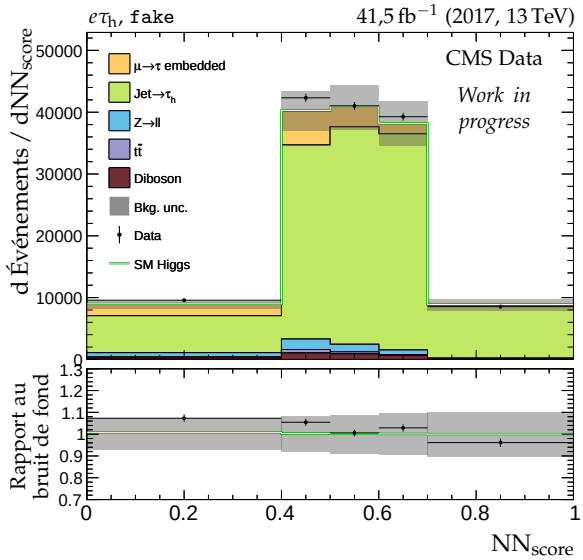
(b) *Catégorie* `emb`.



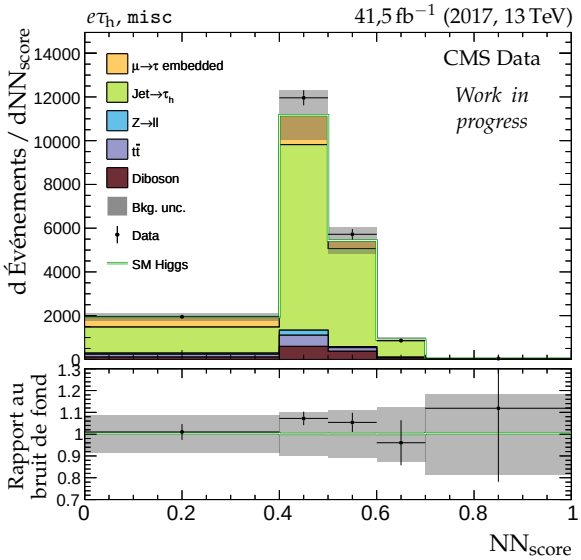
(c) *Catégorie* `ttbar`.



(d) *Catégorie* `zll`.

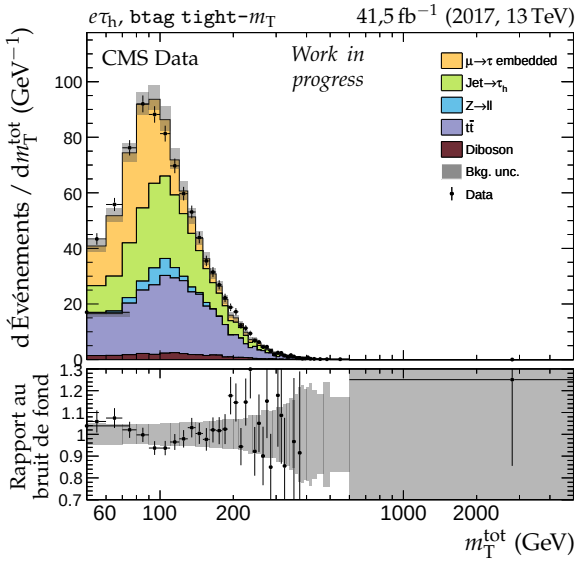


(e) *Catégorie* `fake`.

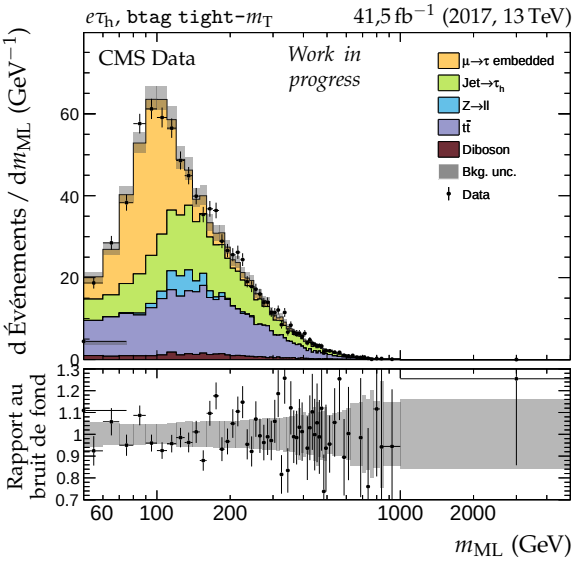


(f) *Catégorie* `misc`.

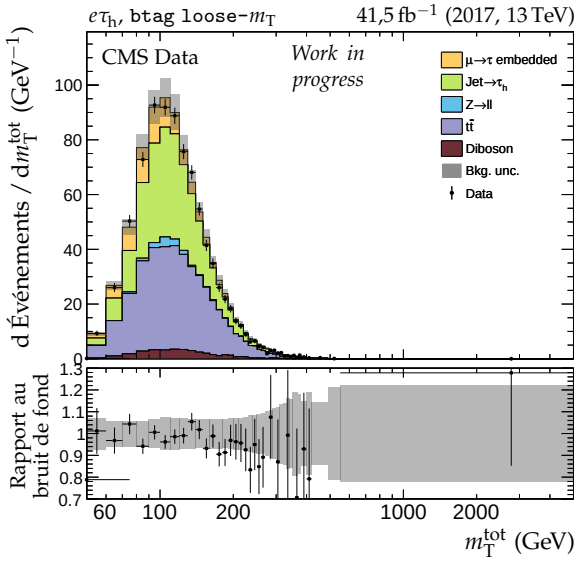
Figure H.21 – *Distributions de* $\mathrm{NN}_{\mathrm{score}}$ *en 2017 dans le canal* $e\tau_{\mathrm{h}}$.



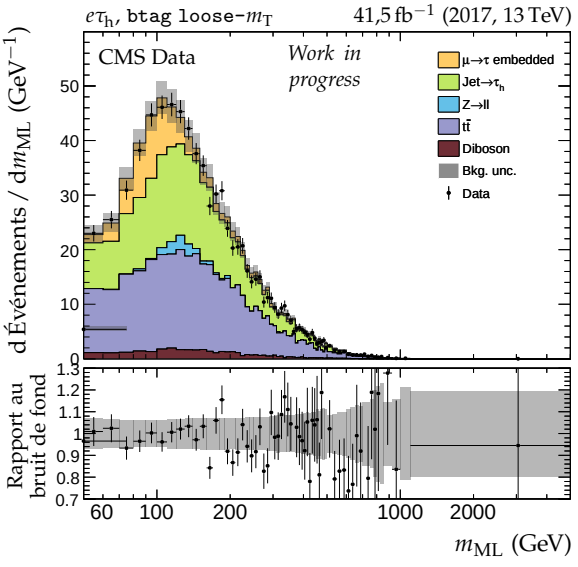
(a) *Catégorie* `btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `btag tight-`m_{T}, m_{ML}.

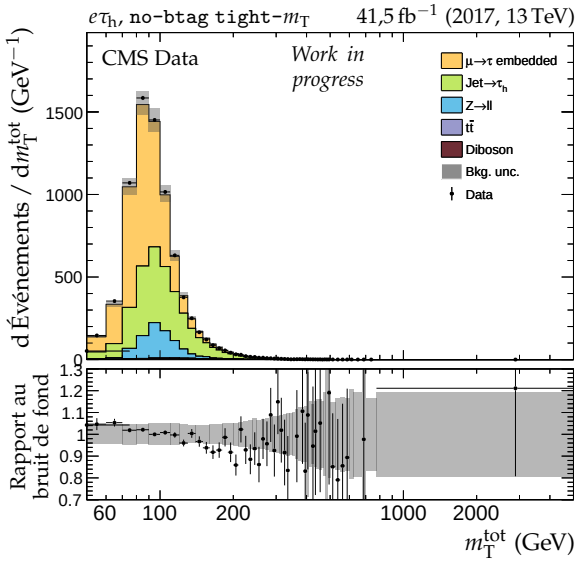


(c) *Catégorie* `btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

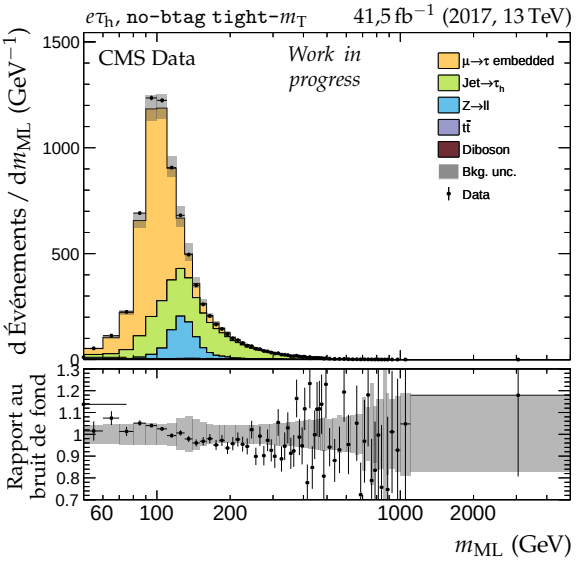


(d) *Catégorie* `btag loose-`m_{T}, m_{ML}.

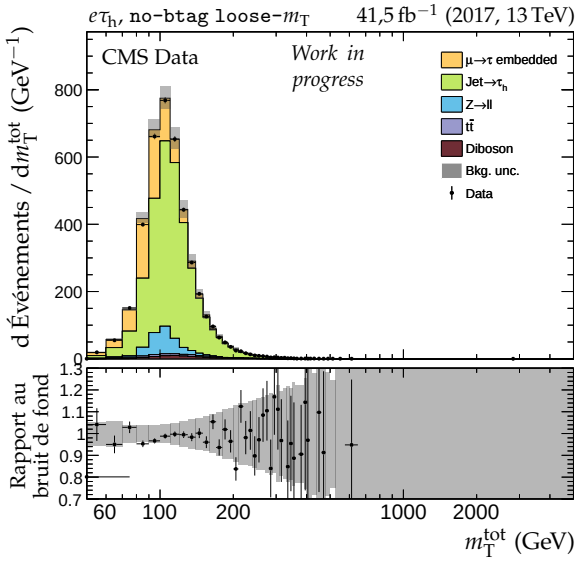
Figure H.22 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2017 dans le canal* $e\tau_{\mathrm{h}}$*, catégories* `btag`.



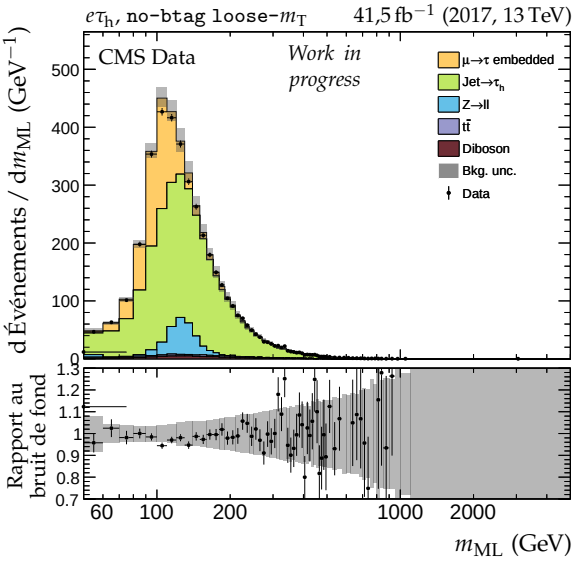
(a) *Catégorie* `no-btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `no-btag tight-`m_{T}, m_{ML}.

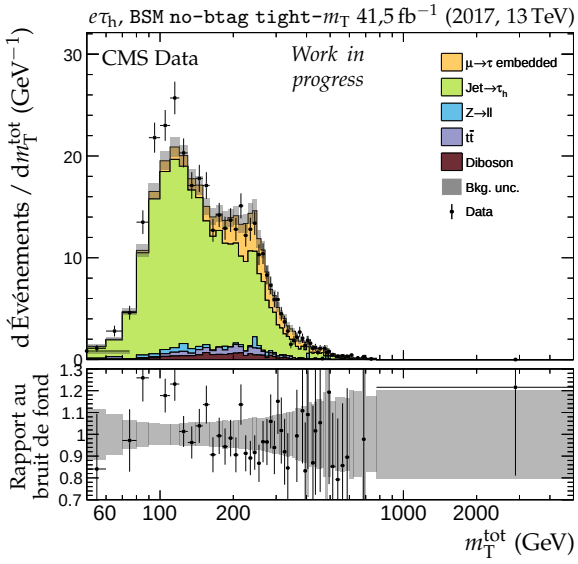


(c) *Catégorie* `no-btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

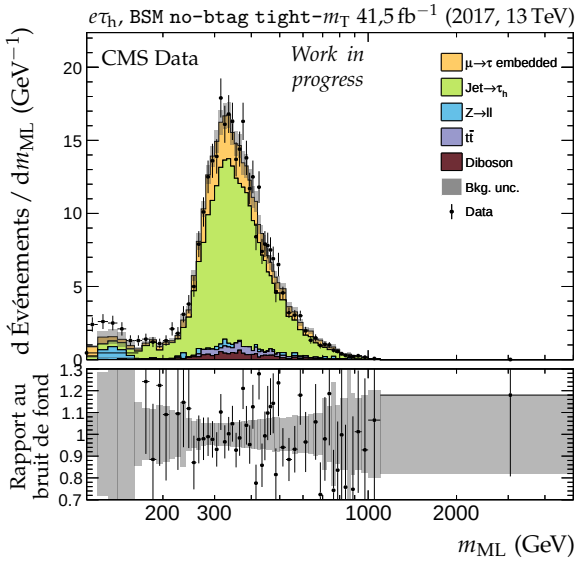


(d) *Catégorie* `no-btag loose-`m_{T}, m_{ML}.

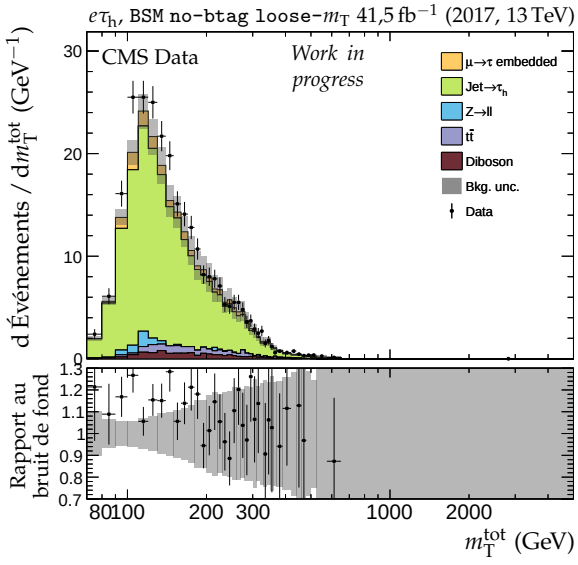
Figure H.23 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2017 dans le canal* $e\tau_{\mathrm{h}}$, *catégories* `no-btag`.



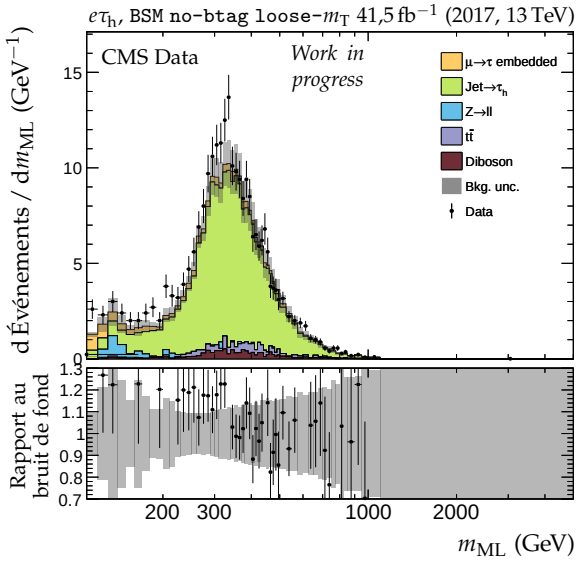
(a) *Catégorie* BSM `no-btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* BSM `no-btag tight-`m_{T}, m_{ML}.

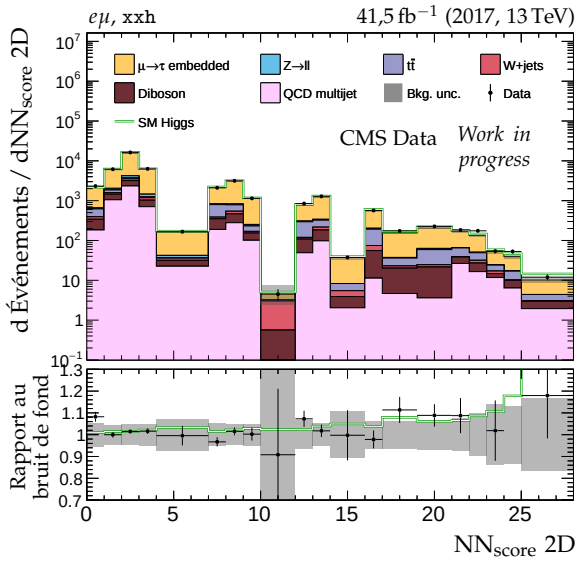


(c) *Catégorie* BSM `no-btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

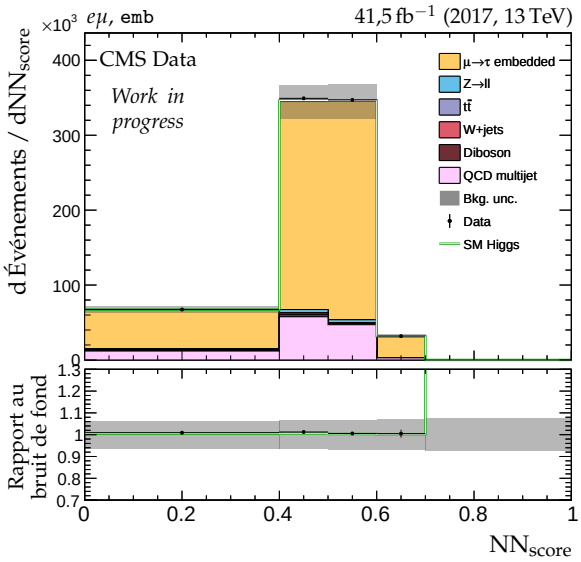


(d) *Catégorie* BSM `no-btag loose-`m_{T}, m_{ML}.

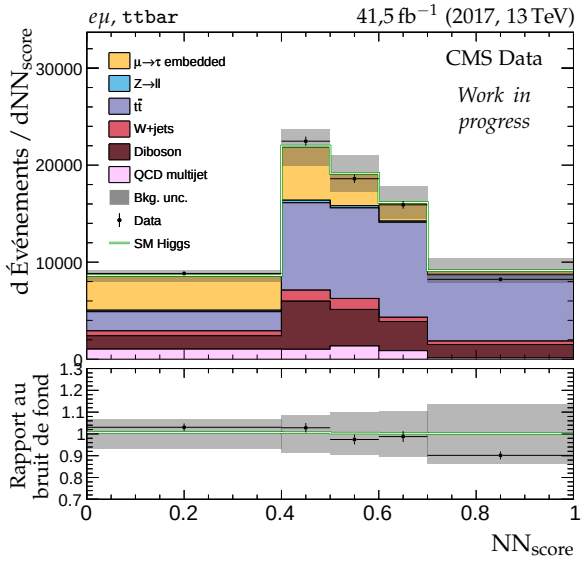
Figure H.24 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2017 dans le canal* $e\tau_{\mathrm{h}}$, *catégories* `no-btag` *avec* $m_{\mathrm{SV\,FIT}} \geqslant 250\,\mathrm{GeV}$.



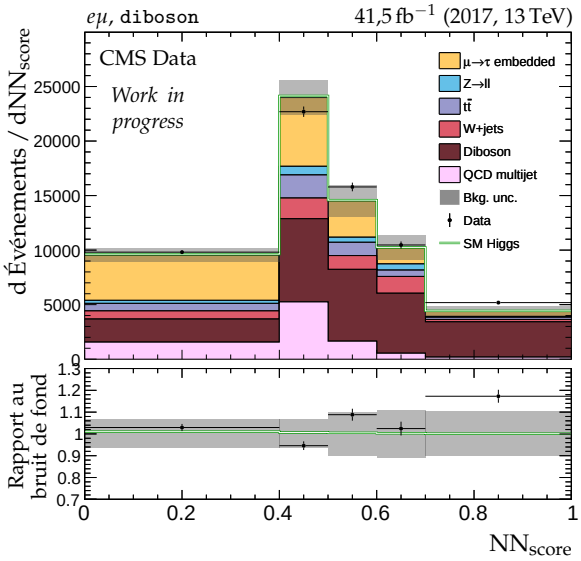
(a) *Catégorie* `xxh`.



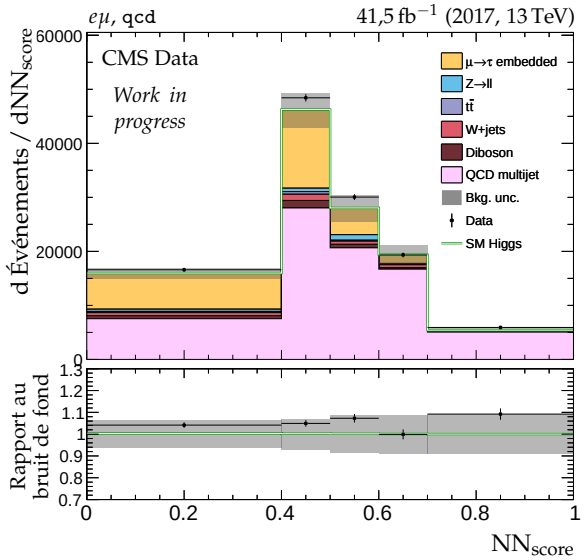
(b) *Catégorie* `emb`.



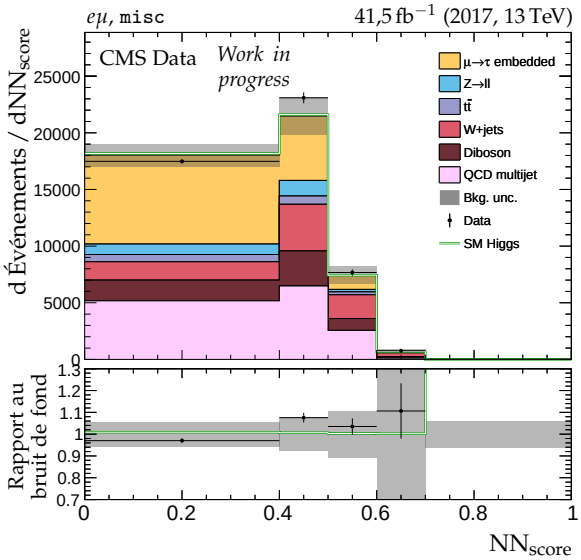
(c) *Catégorie* `ttbar`.



(d) *Catégorie* `diboson`.

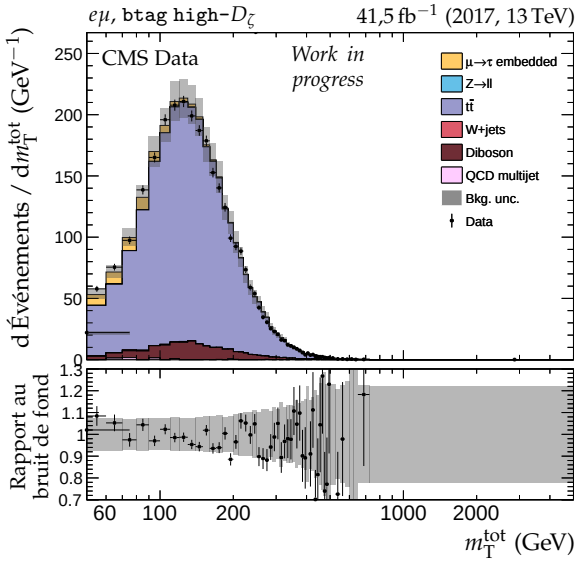


(e) *Catégorie* `qcd`.

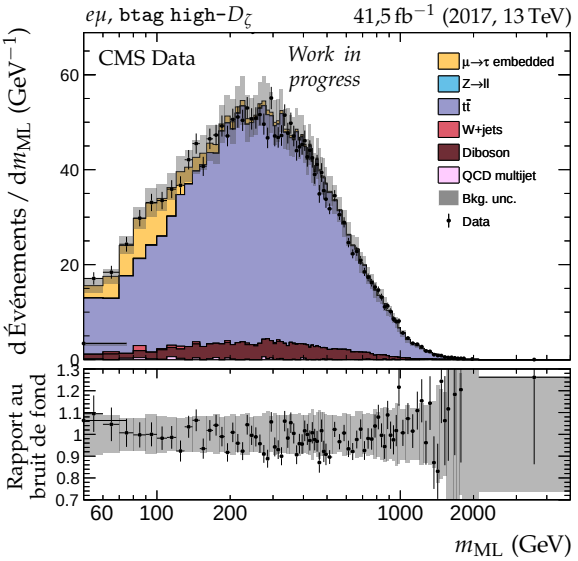


(f) *Catégorie* `misc`.

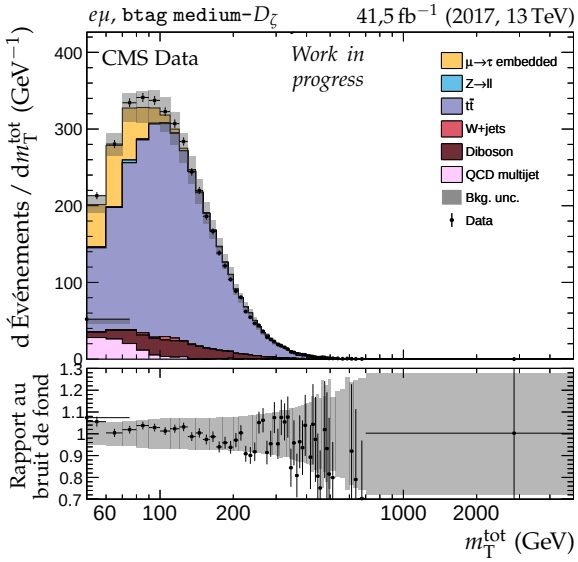
Figure H.25 – *Distributions de* NN$_{score}$ *en 2017 dans le canal eμ.*



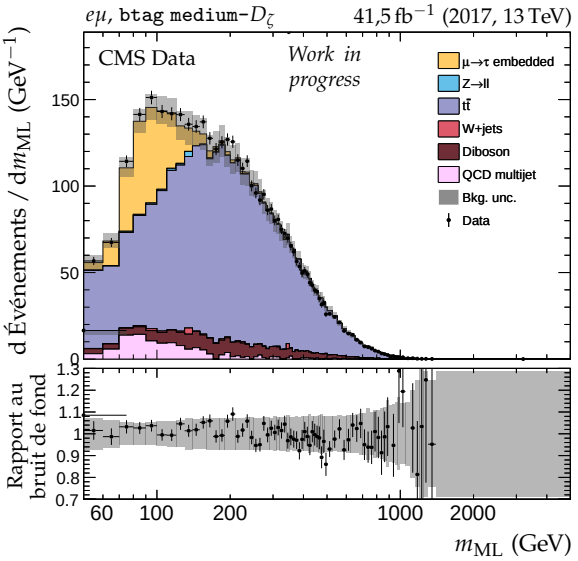
(a) *Catégorie* `btag high-`D_ζ, m_T^{tot}.



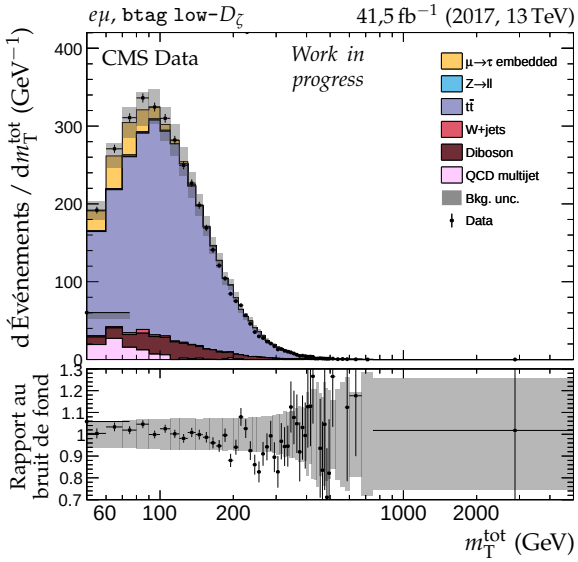
(b) *Catégorie* `btag high-`D_ζ, m_{ML}.



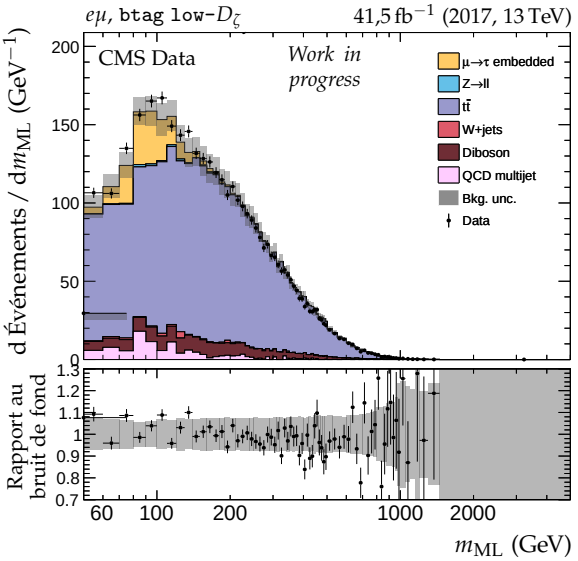
(c) *Catégorie* `btag medium-`D_ζ, m_T^{tot}.



(d) *Catégorie* `btag medium-`D_ζ, m_{ML}.

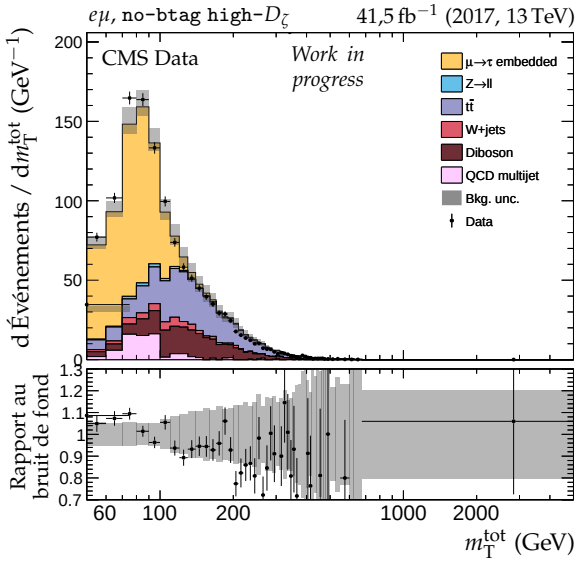


(e) *Catégorie* `btag low-`D_ζ, m_T^{tot}.

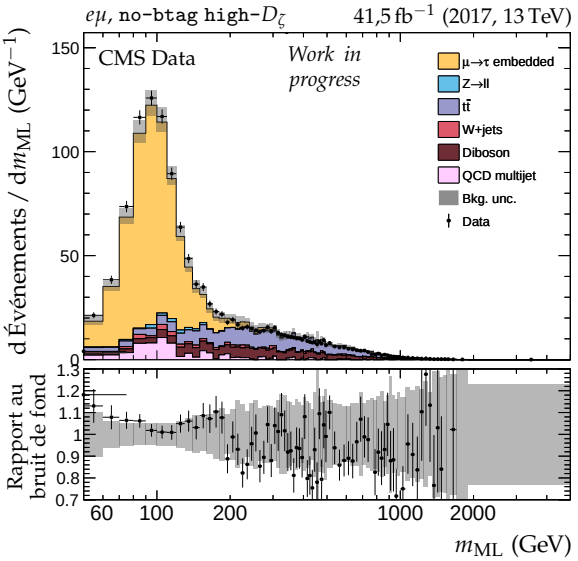


(f) *Catégorie* `btag low-`D_ζ, m_{ML}.

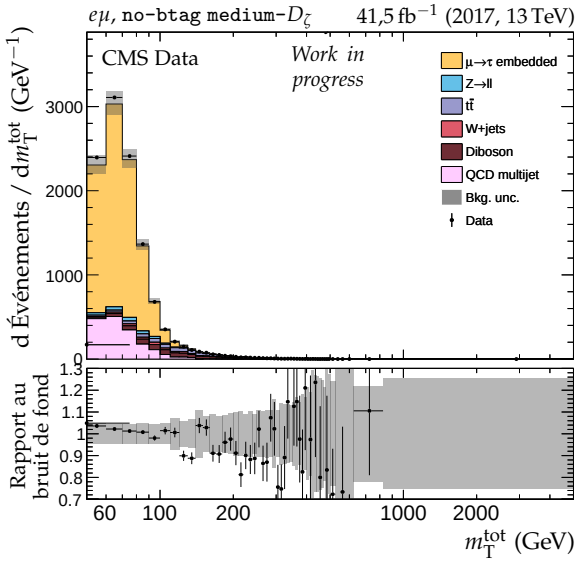
Figure H.26 – *Distributions de* m_T^{tot} *et* m_{ML} *en 2017 dans le canal eμ, catégories* `btag`.



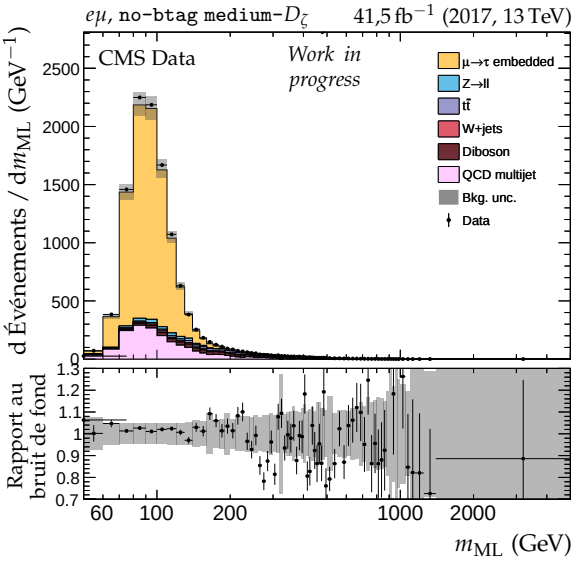
(a) *Catégorie* `no-btag high-`D_ζ, m_T^{tot}.



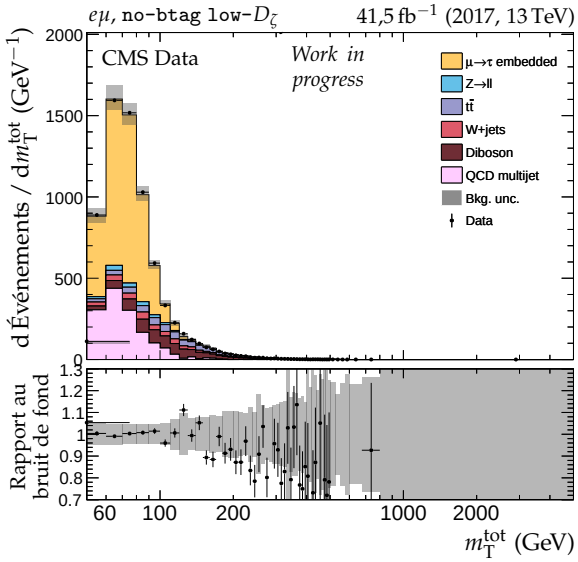
(b) *Catégorie* `no-btag high-`D_ζ, m_{ML}.



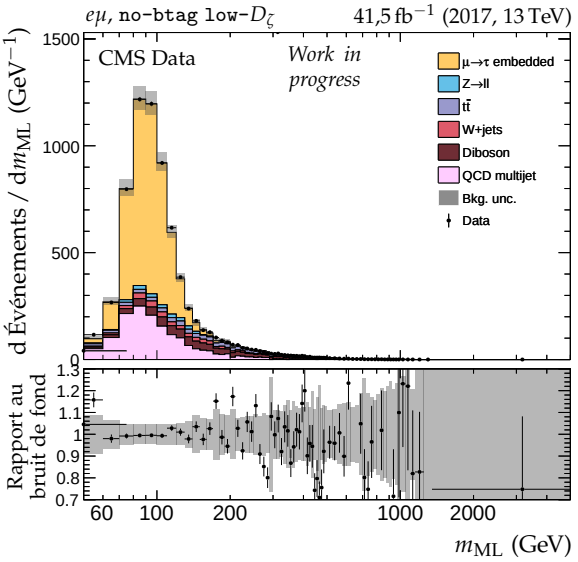
(c) *Catégorie* `no-btag medium-`D_ζ, m_T^{tot}.



(d) *Catégorie* `no-btag medium-`D_ζ, m_{ML}.

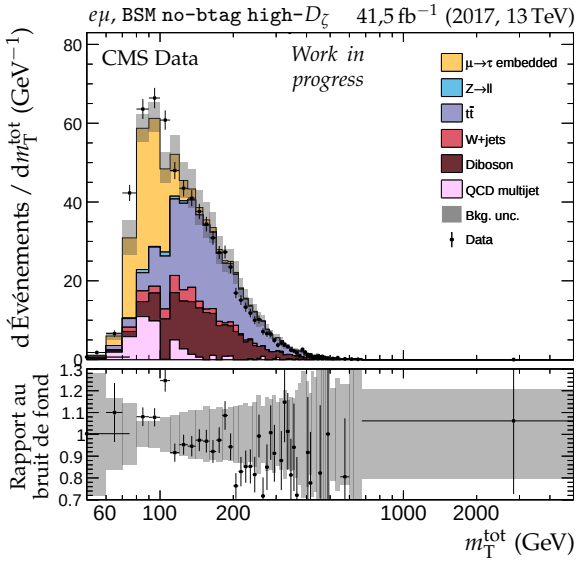


(e) *Catégorie* `no-btag low-`D_ζ, m_T^{tot}.

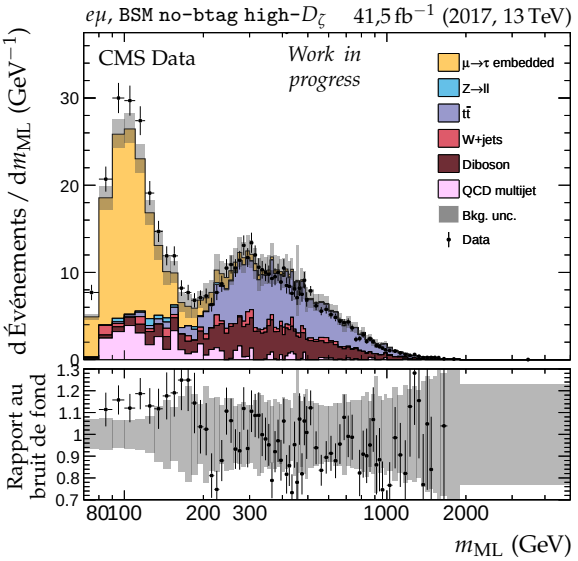


(f) *Catégorie* `no-btag low-`D_ζ, m_{ML}.

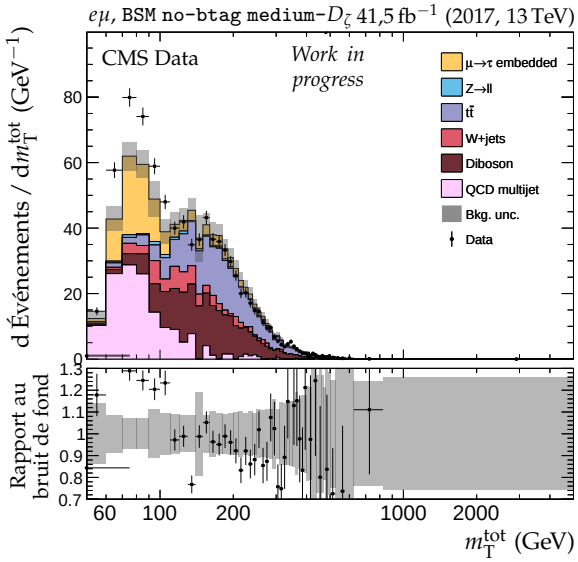
Figure H.27 – *Distributions de m_T^{tot} et m_{ML} en 2017 dans le canal eμ, catégories* `no-btag`.



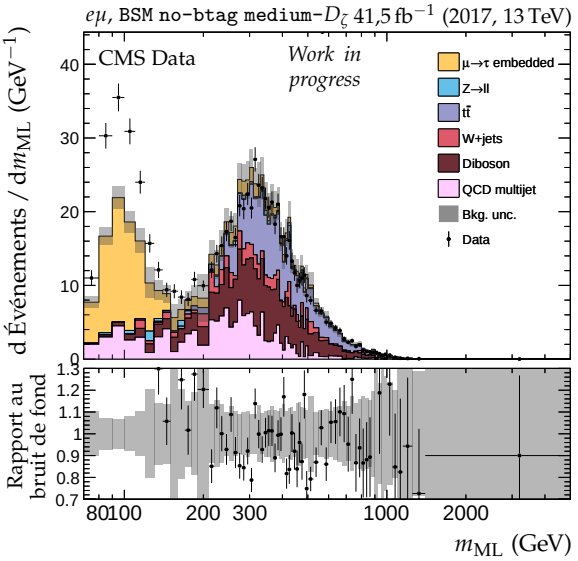
(a) *Catégorie BSM* `no-btag high-`D_ζ, m_T^{tot}.



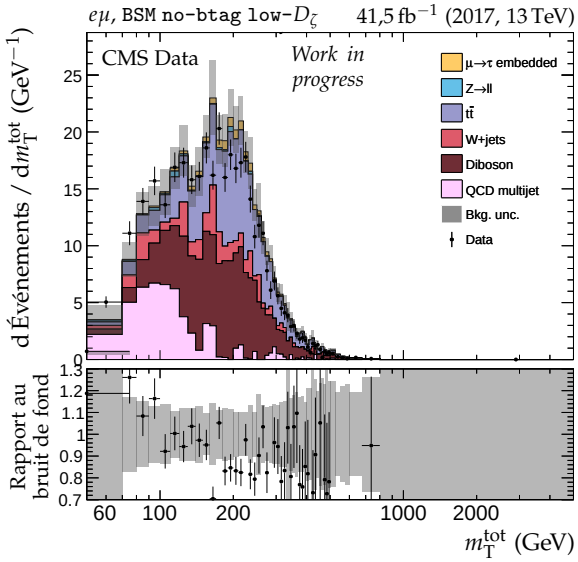
(b) *Catégorie BSM* `no-btag high-`D_ζ, m_{ML}.



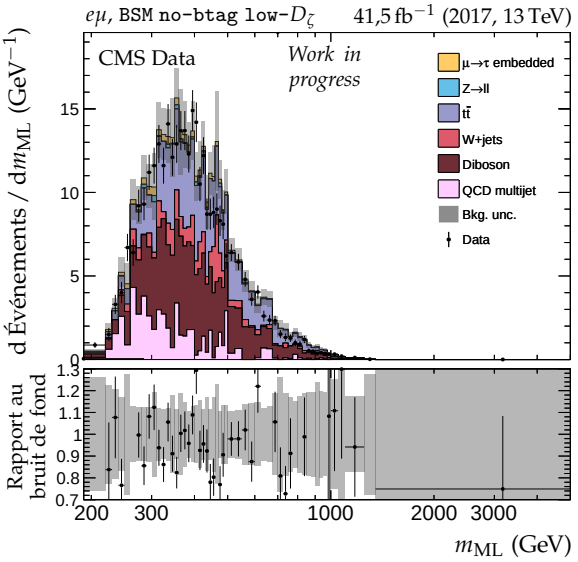
(c) *Catégorie BSM* `no-btag medium-`D_ζ, m_T^{tot}.



(d) *Catégorie BSM* `no-btag medium-`D_ζ, m_{ML}.

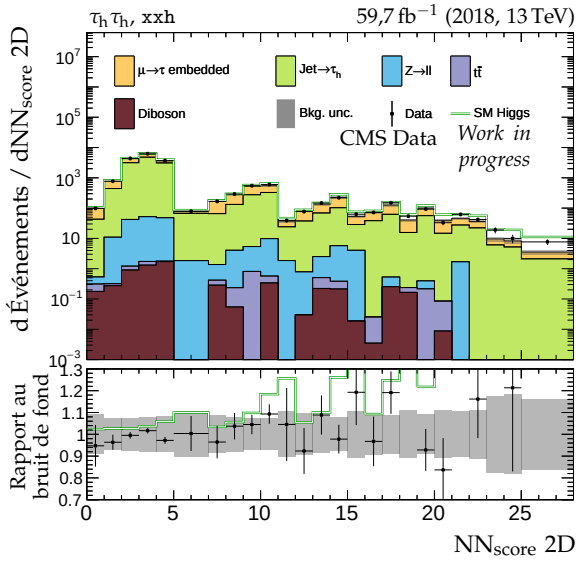


(e) *Catégorie BSM* `no-btag low-`D_ζ, m_T^{tot}.

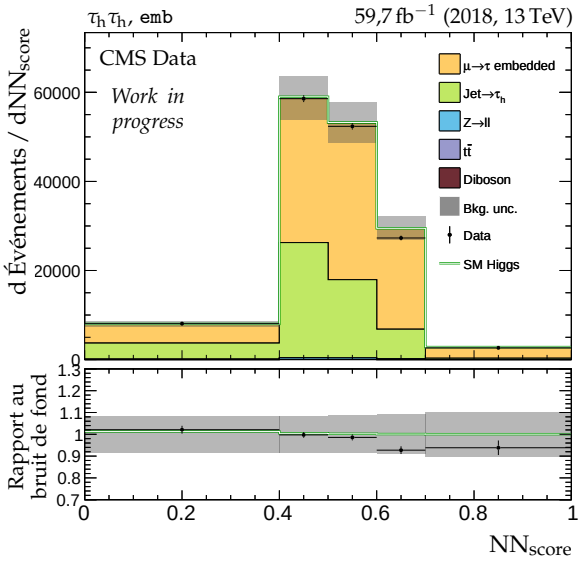


(f) *Catégorie BSM* `no-btag low-`D_ζ, m_{ML}.

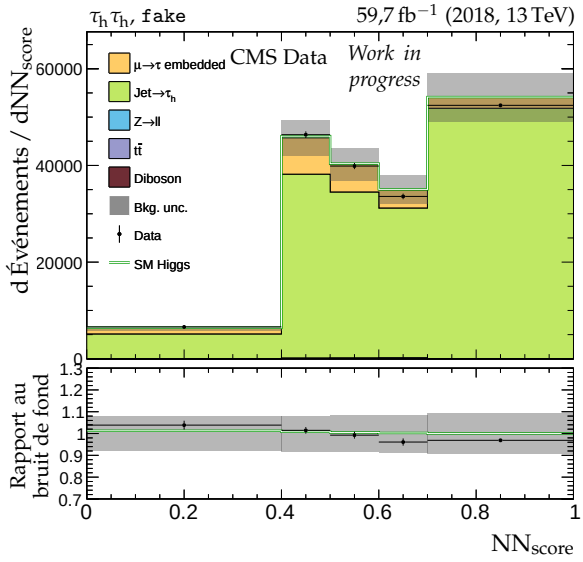
Figure H.28 – *Distributions de* m_T^{tot} *et* m_{ML} *en 2017 dans le canal eμ, catégories* `no-btag` *avec* $m_{SVFIT} \geqslant 250$ GeV.



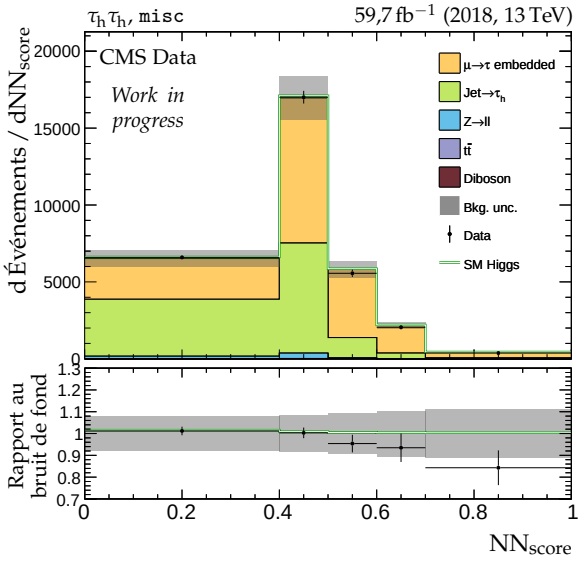
(a) *Catégorie* `xxh`.



(b) *Catégorie* `emb`.

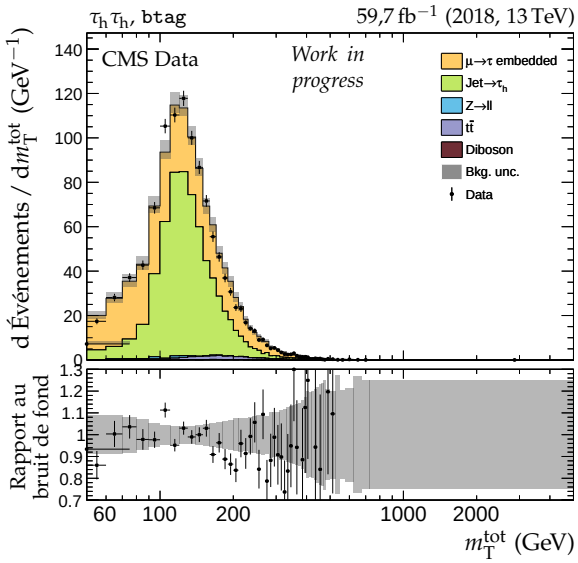


(c) *Catégorie* `fake`.

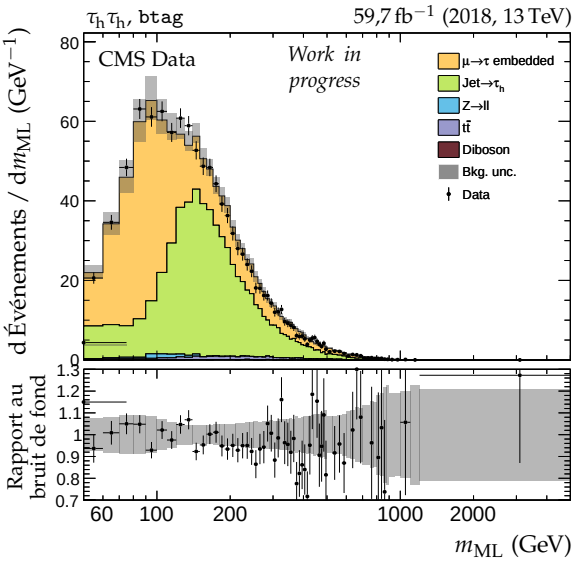


(d) *Catégorie* `misc`.

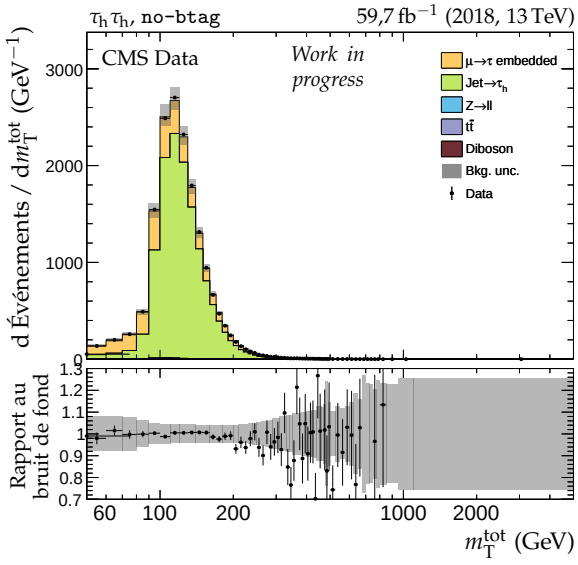
Figure H.29 – *Distributions de* $\mathrm{NN_{score}}$ *en 2018 dans le canal* $\tau_h\tau_h$.



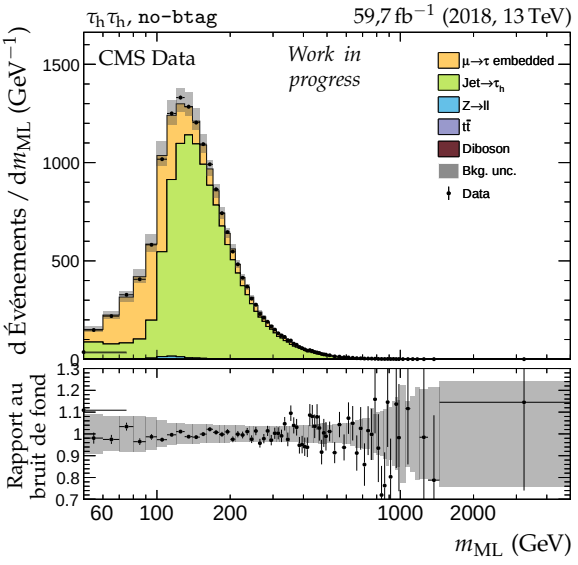
(a) *Catégorie* `btag`, m_T^{tot}.



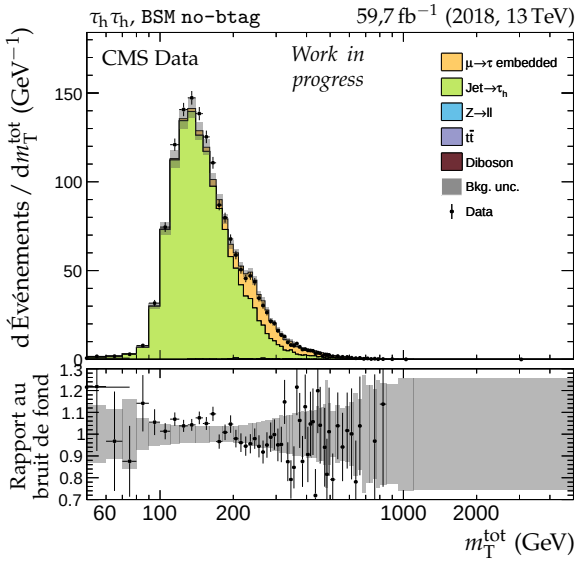
(b) *Catégorie* `btag`, m_{ML}.



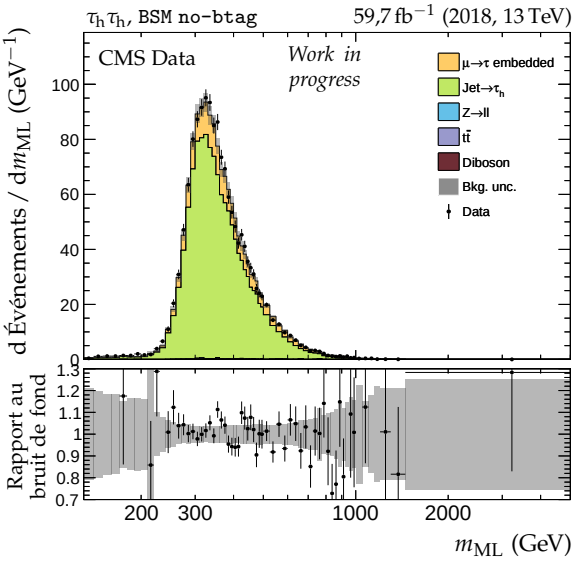
(c) *Catégorie* `no-btag`, m_T^{tot}.



(d) *Catégorie* `no-btag`, m_{ML}.

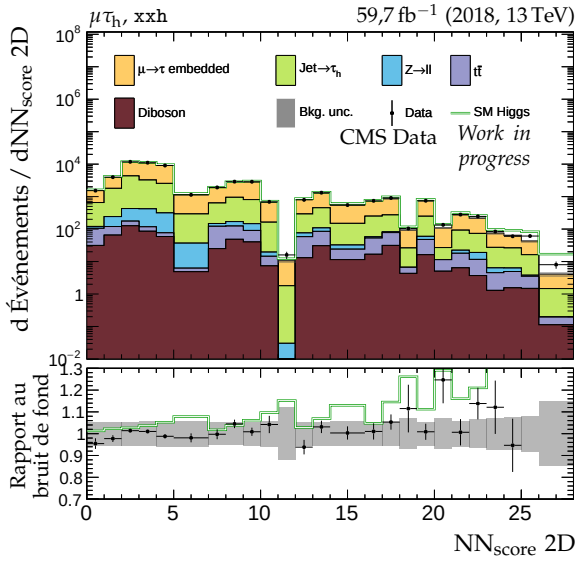


(e) *Catégorie* `BSM no-btag`, m_T^{tot}.

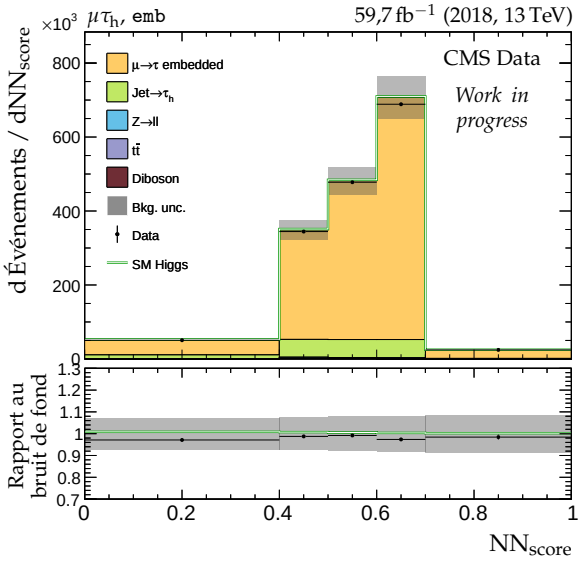


(f) *Catégorie* `BSM no-btag`, m_{ML}.

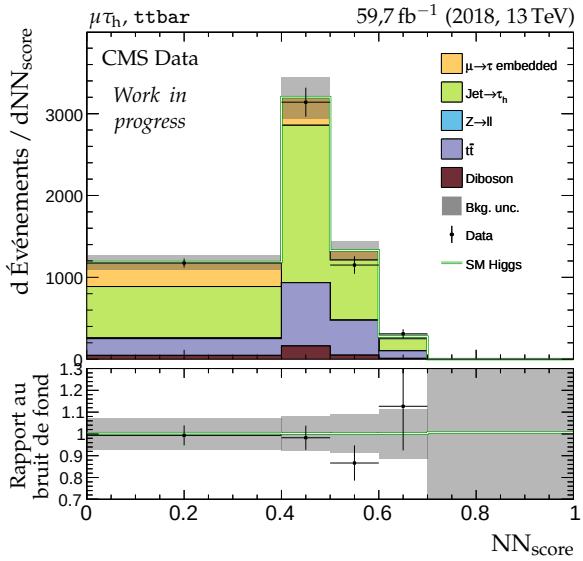
Figure H.30 – *Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal $\tau_h\tau_h$.*



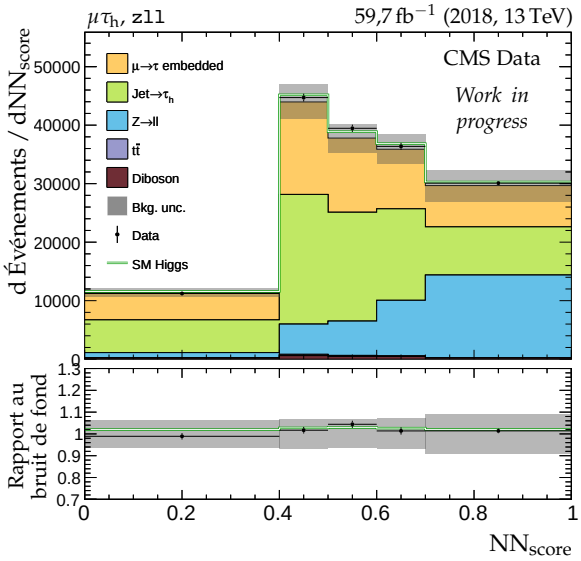
(a) *Catégorie* `xxh`.



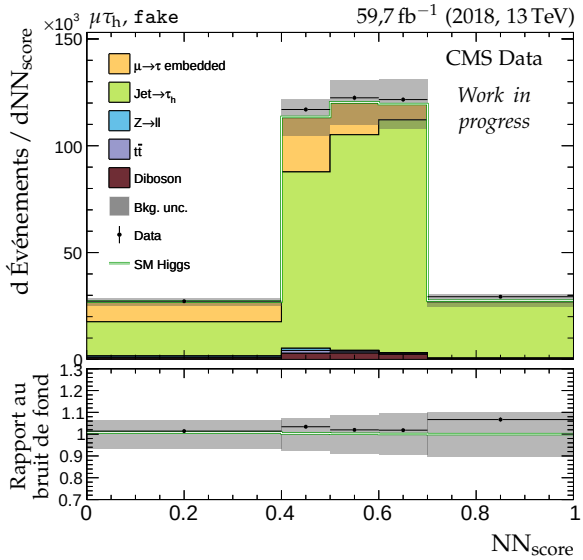
(b) *Catégorie* `emb`.



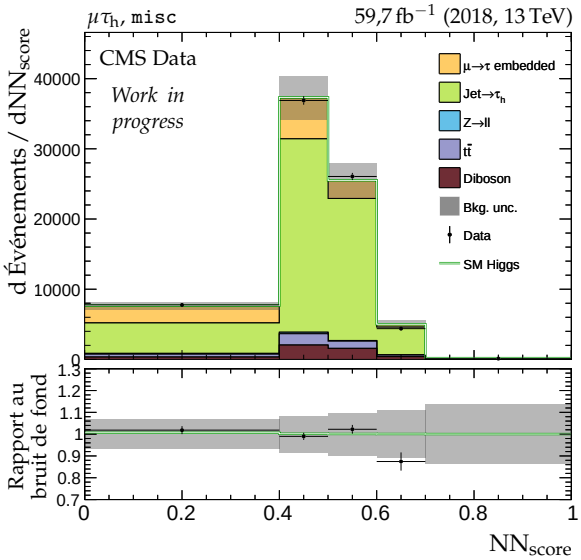
(c) *Catégorie* `ttbar`.



(d) *Catégorie* `zll`.

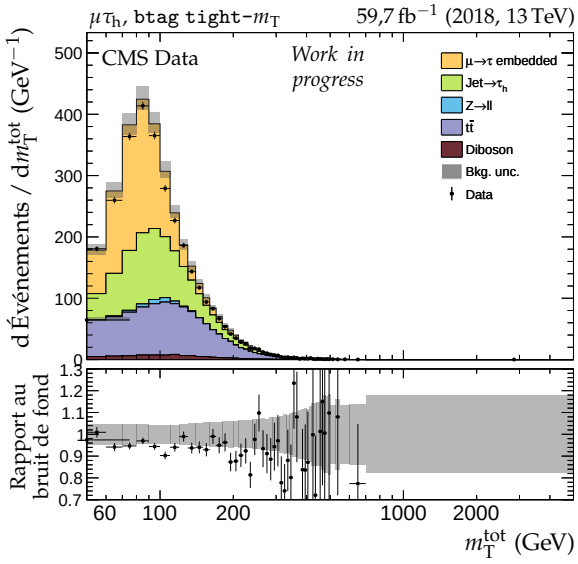


(e) *Catégorie* `fake`.

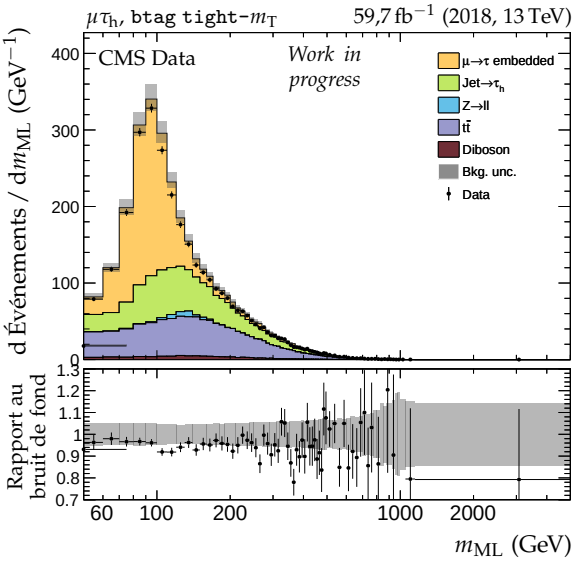


(f) *Catégorie* `misc`.

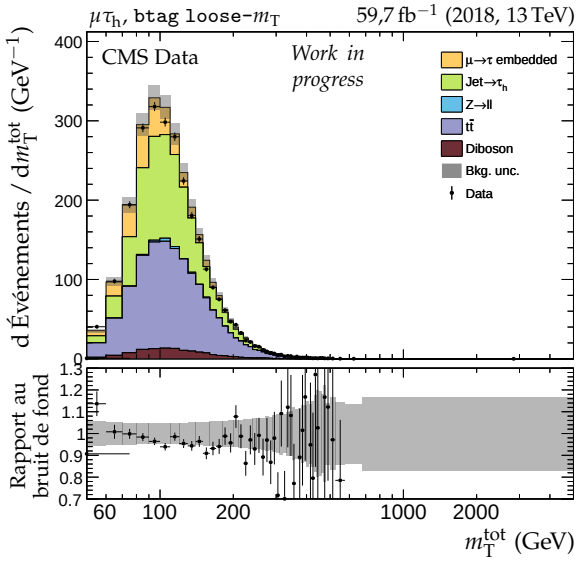
Figure H.31 – *Distributions de* $\mathrm{NN_{score}}$ *en 2018 dans le canal* $\mu\tau_h$.



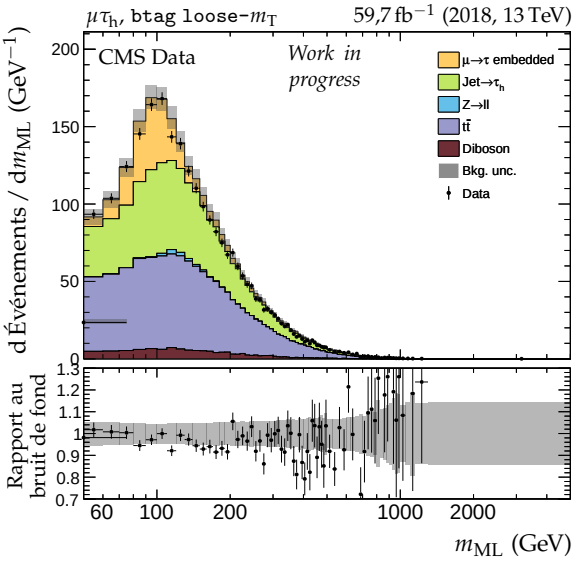
(a) *Catégorie* `btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `btag tight-`m_{T}, m_{ML}.

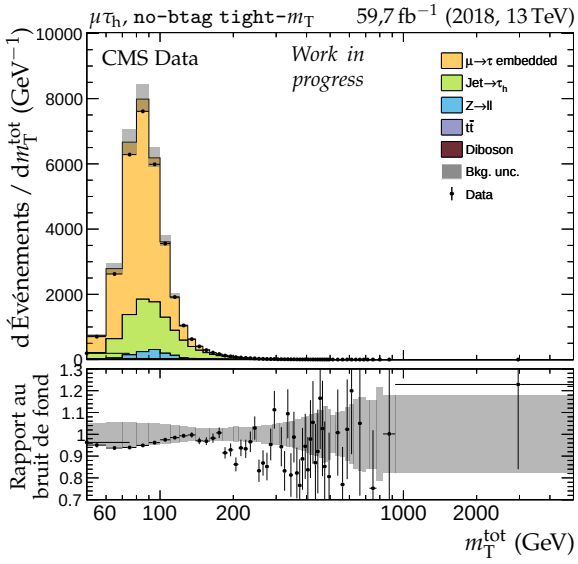


(c) *Catégorie* `btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

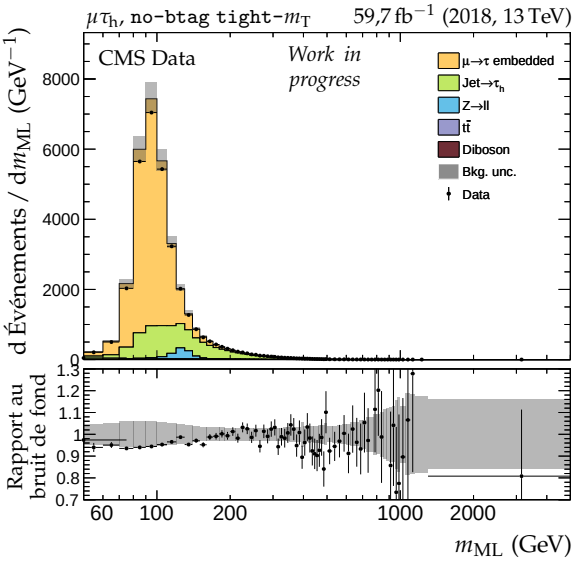


(d) *Catégorie* `btag loose-`m_{T}, m_{ML}.

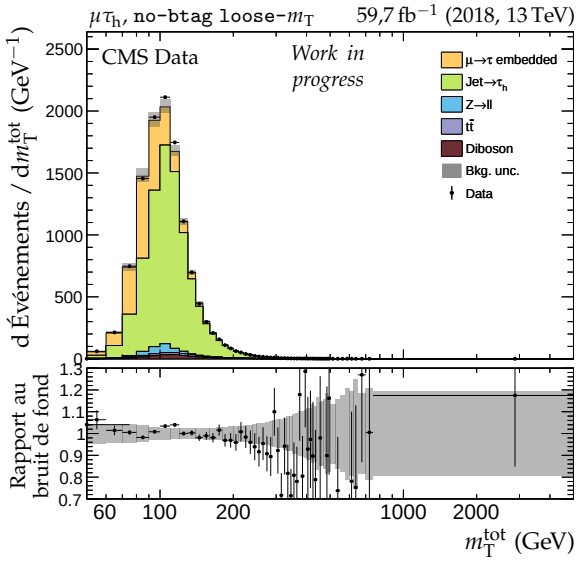
Figure H.32 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2018 dans le canal* $\mu\tau_{\mathrm{h}}$, *catégories* `btag`.



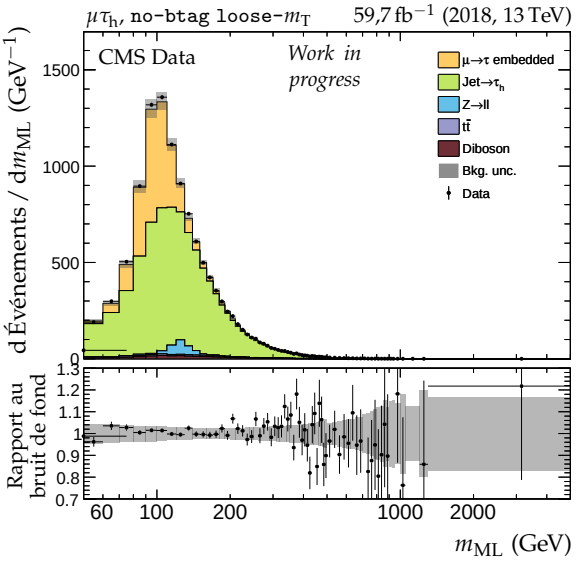
(a) *Catégorie* `no-btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `no-btag tight-`m_{T}, m_{ML}.

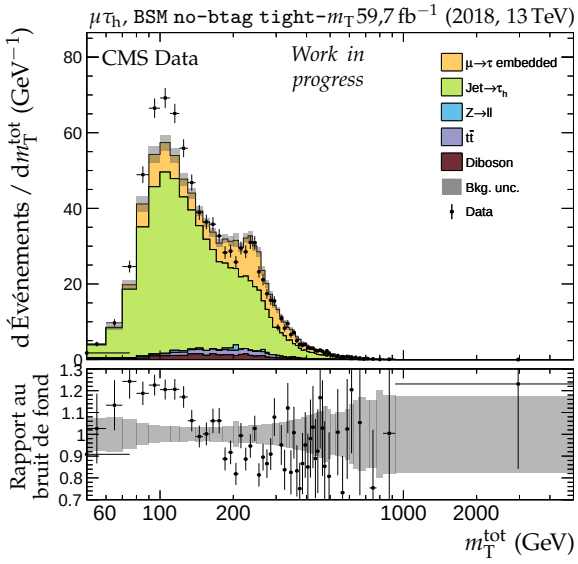


(c) *Catégorie* `no-btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

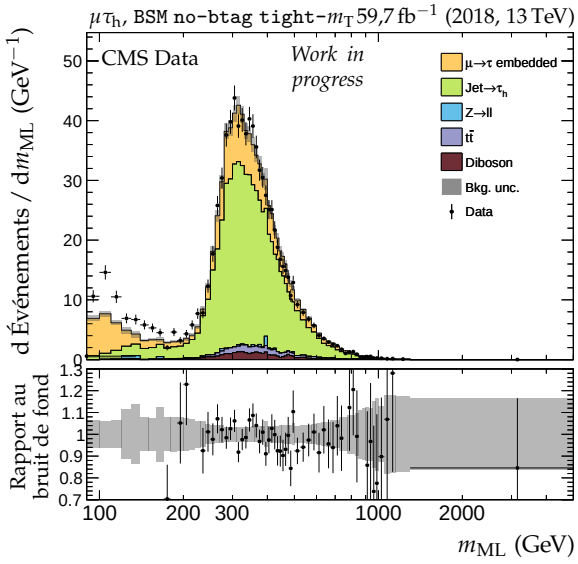


(d) *Catégorie* `no-btag loose-`m_{T}, m_{ML}.

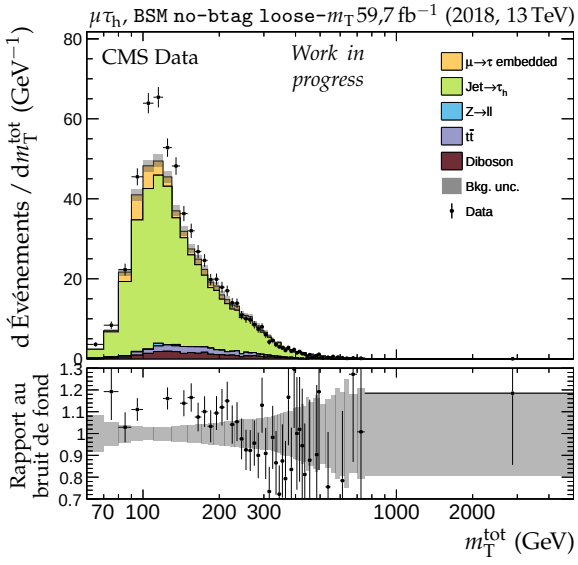
Figure H.33 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2018 dans le canal* $\mu\tau_{\mathrm{h}}$, *catégories* `no-btag`.



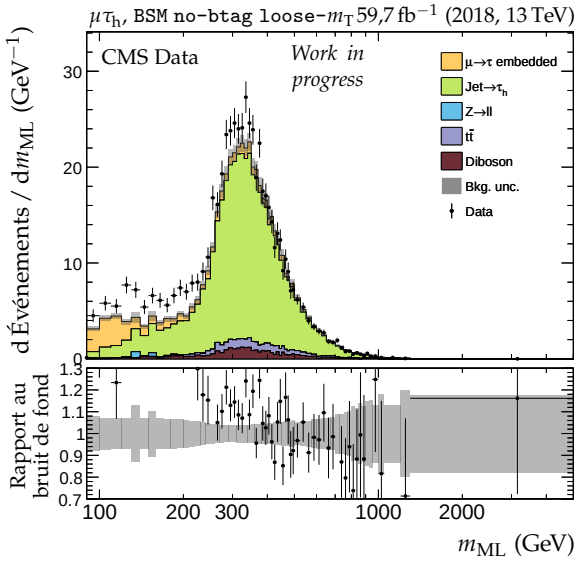
(a) *Catégorie* BSM `no-btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* BSM `no-btag tight-`m_{T}, m_{ML}.

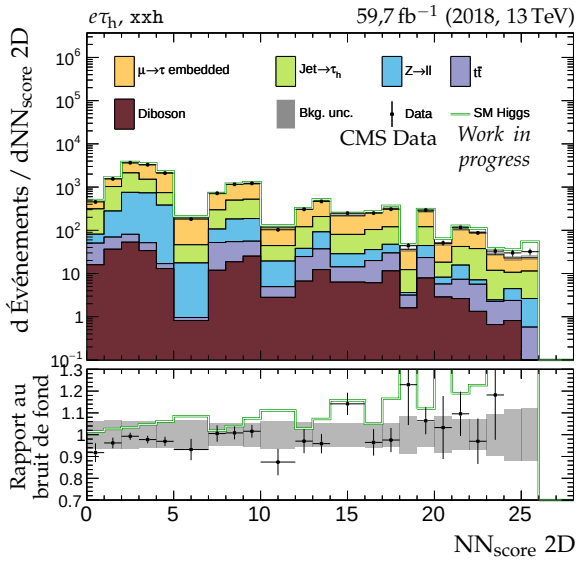


(c) *Catégorie* BSM `no-btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

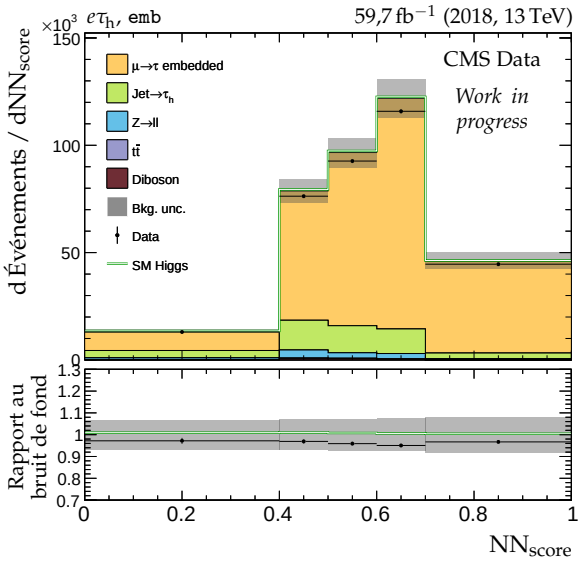


(d) *Catégorie* BSM `no-btag loose-`m_{T}, m_{ML}.

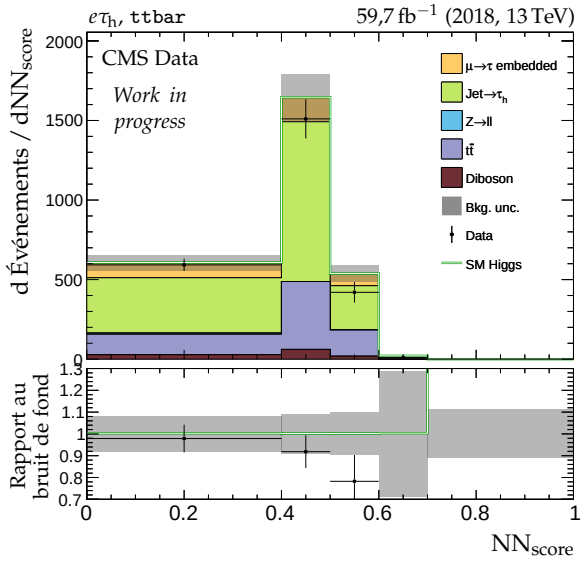
Figure H.34 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2018 dans le canal* $\mu\tau_{\mathrm{h}}$, *catégories* `no-btag` *avec* $m_{\mathrm{SVFIT}} \geqslant 250\,\mathrm{GeV}$.



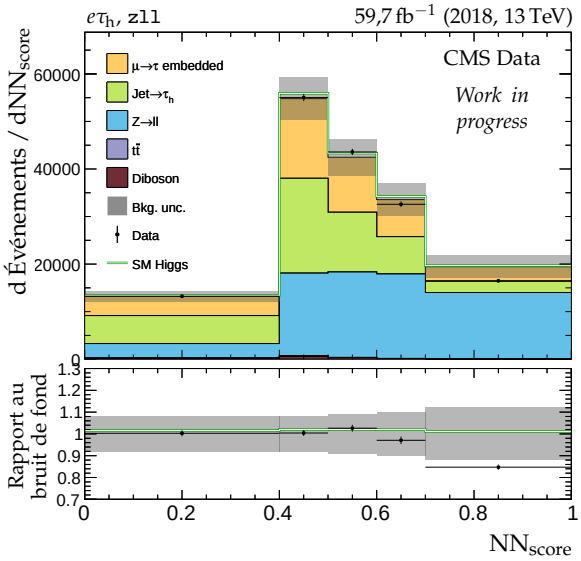
(a) *Catégorie* `xxh`.



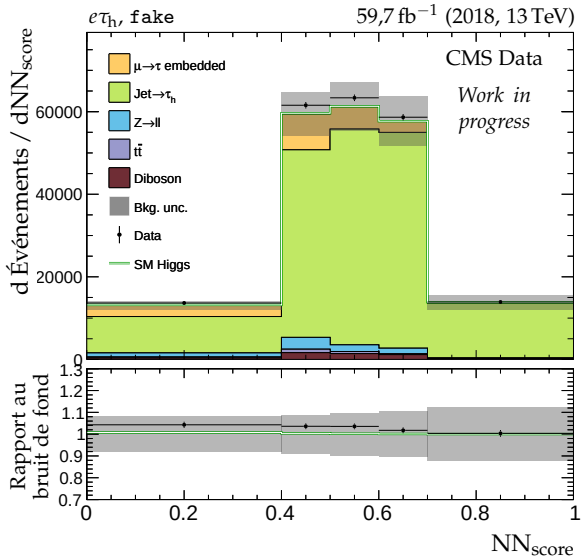
(b) *Catégorie* `emb`.



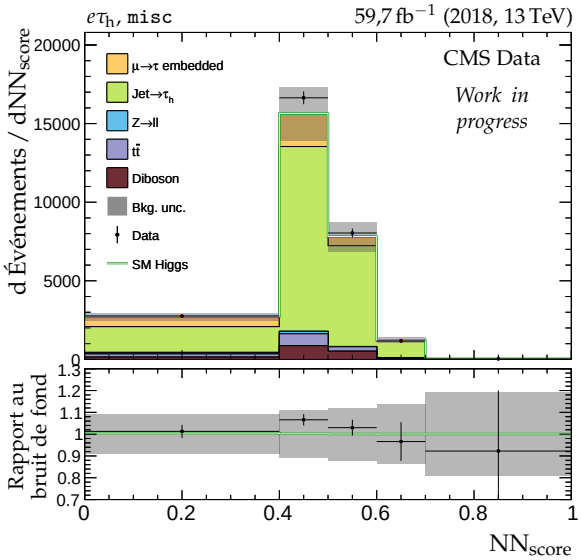
(c) *Catégorie* `ttbar`.



(d) *Catégorie* `zll`.

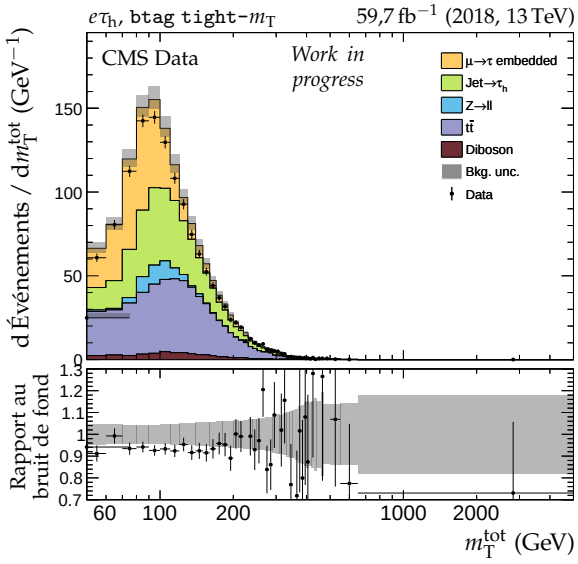


(e) *Catégorie* `fake`.

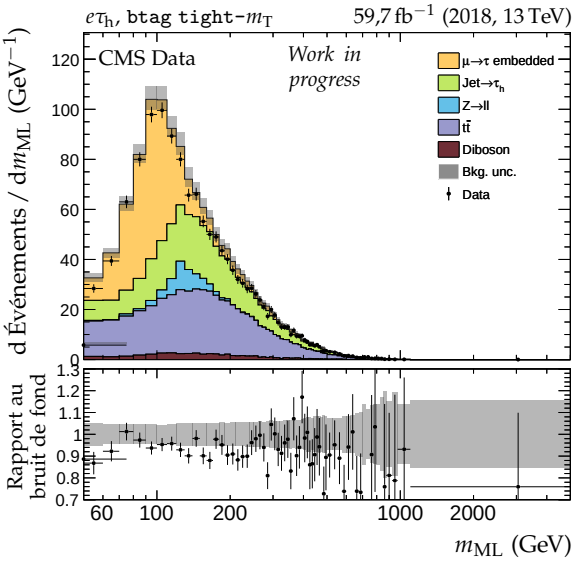


(f) *Catégorie* `misc`.

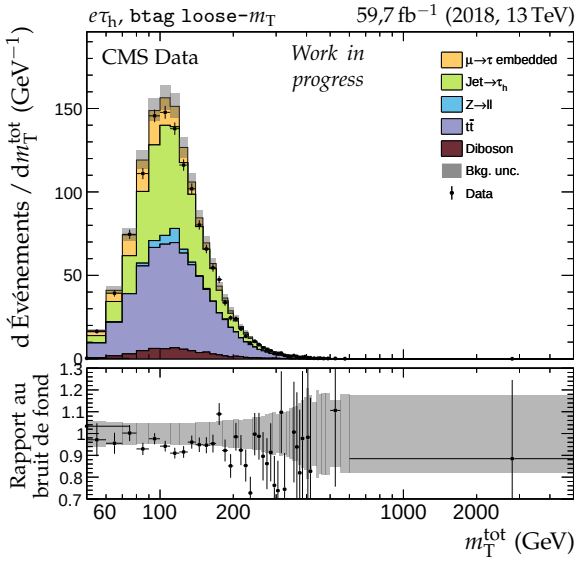
Figure H.35 – *Distributions de* $\mathrm{NN}_{\mathrm{score}}$ *en 2018 dans le canal* $e\tau_{\mathrm{h}}$.



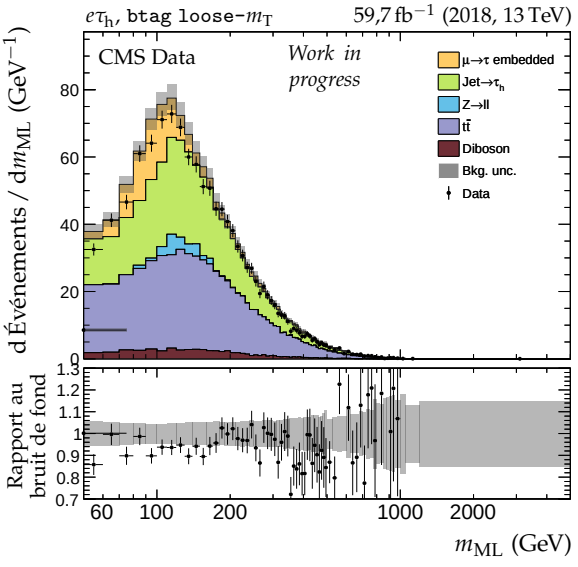
(a) *Catégorie* `btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `btag tight-`m_{T}, m_{ML}.

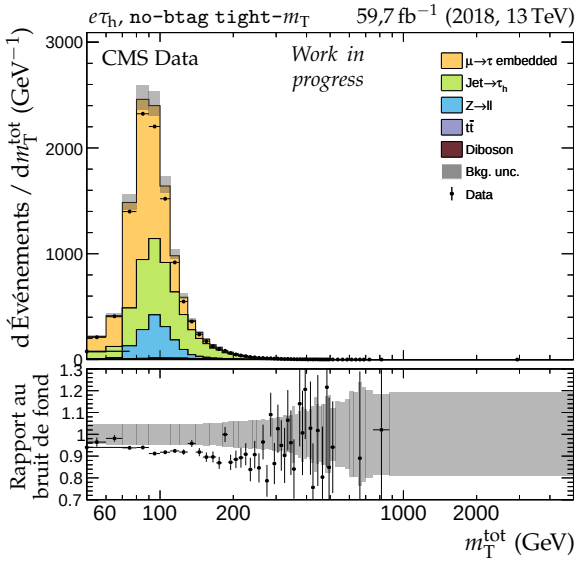


(c) *Catégorie* `btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

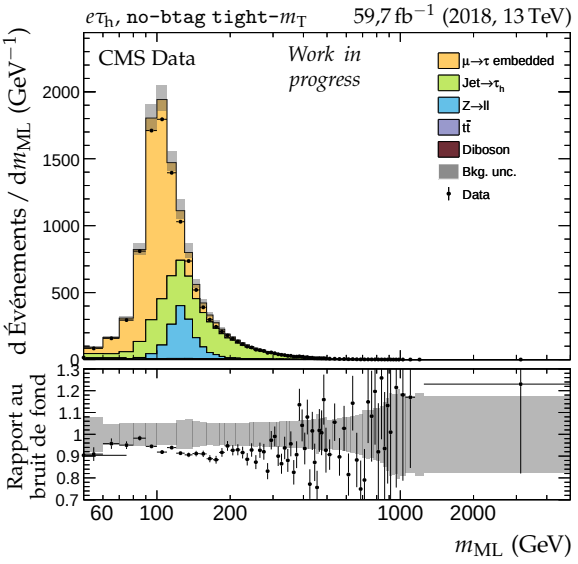


(d) *Catégorie* `btag loose-`m_{T}, m_{ML}.

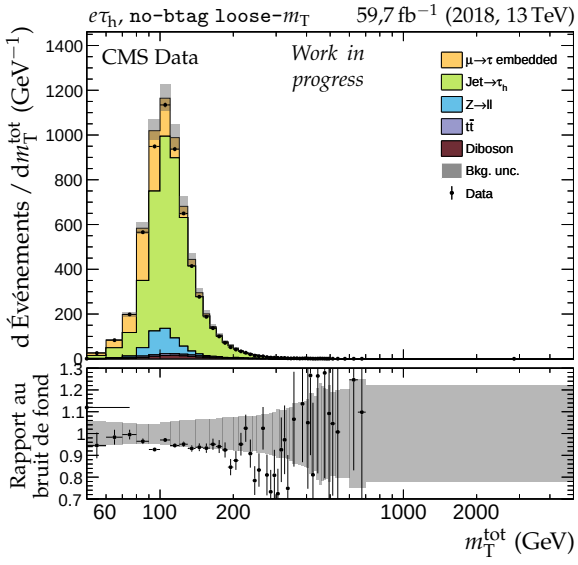
Figure H.36 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2018 dans le canal* $e\tau_{\mathrm{h}}$*, catégories* `btag`.



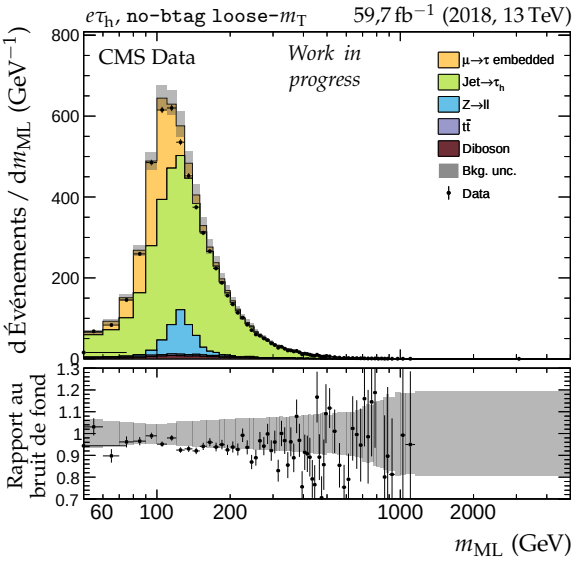
(a) *Catégorie* `no-btag tight-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.



(b) *Catégorie* `no-btag tight-`m_{T}, m_{ML}.

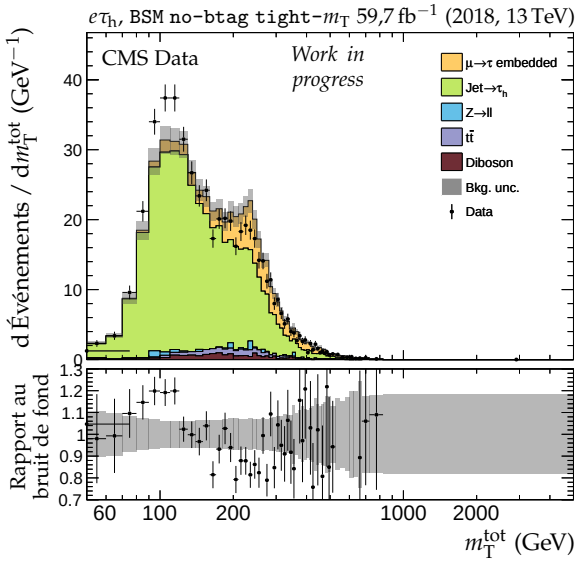


(c) *Catégorie* `no-btag loose-`m_{T}, $m_{\mathrm{T}}^{\mathrm{tot}}$.

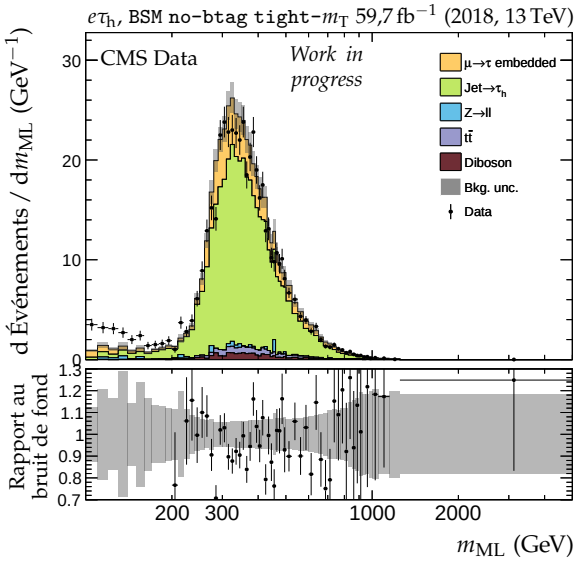


(d) *Catégorie* `no-btag loose-`m_{T}, m_{ML}.

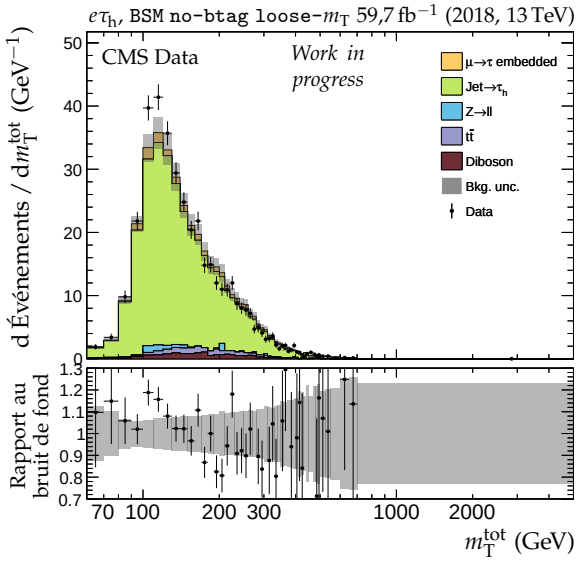
Figure H.37 – *Distributions de* $m_{\mathrm{T}}^{\mathrm{tot}}$ *et* m_{ML} *en 2018 dans le canal* $e\tau_{\mathrm{h}}$, *catégories* `no-btag`.



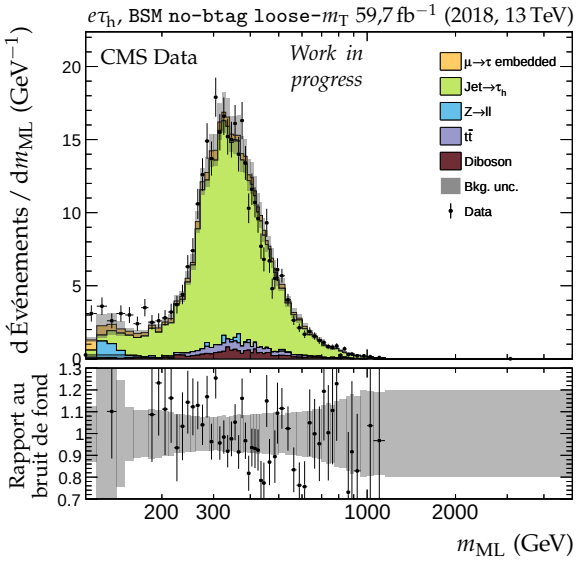
(a) *Catégorie* BSM `no-btag tight-`m_T, m_T^{tot}.



(b) *Catégorie* BSM `no-btag tight-`m_T, m_{ML}.

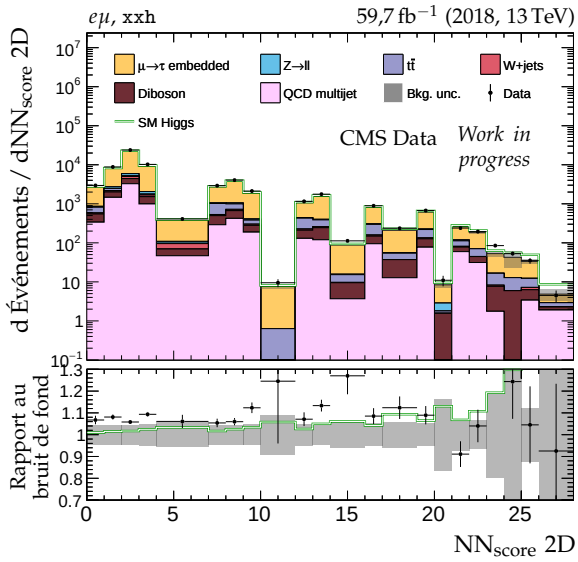


(c) *Catégorie* BSM `no-btag loose-`m_T, m_T^{tot}.

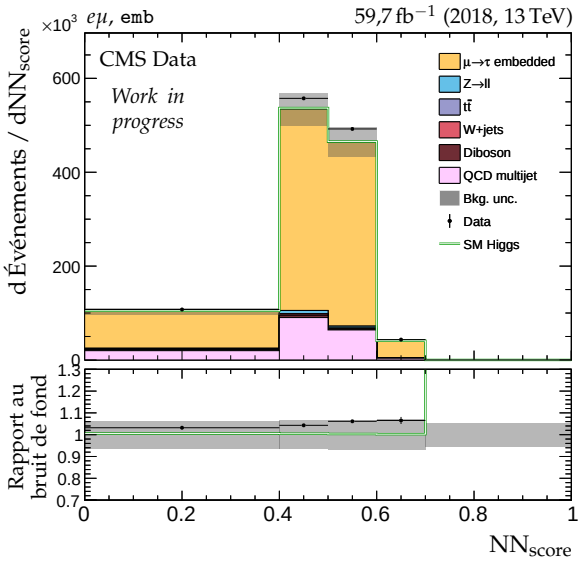


(d) *Catégorie* BSM `no-btag loose-`m_T, m_{ML}.

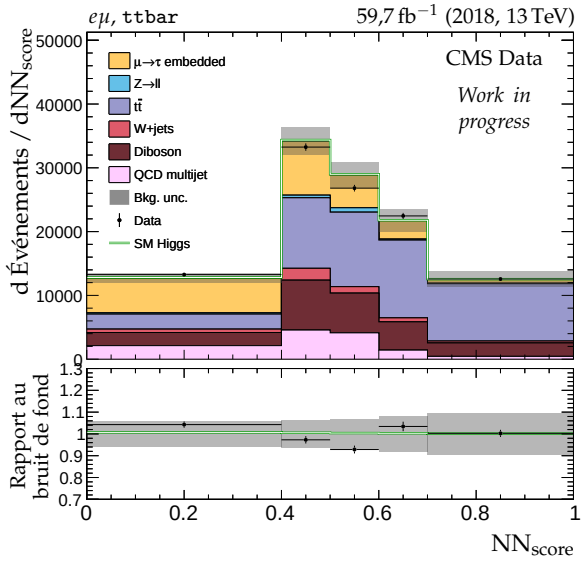
Figure H.38 – *Distributions de* m_T^{tot} *et* m_{ML} *en 2018 dans le canal* $e\tau_h$*, catégories* `no-btag` *avec* $m_{SV FIT} \geqslant 250\,\text{GeV}$.



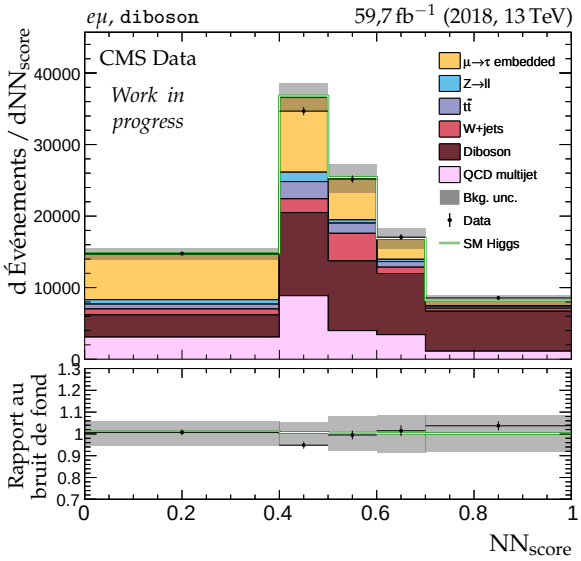
(a) *Catégorie* `xxh`.



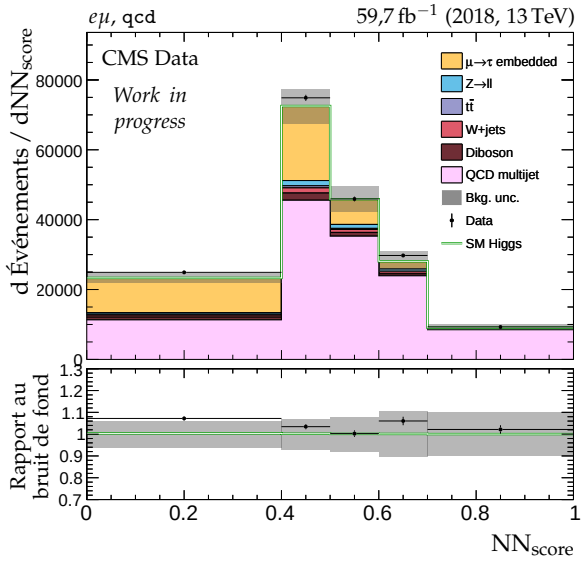
(b) *Catégorie* `emb`.



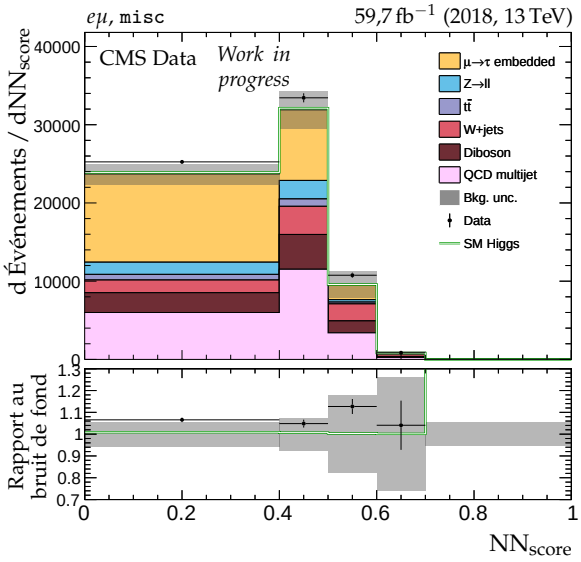
(c) *Catégorie* `ttbar`.



(d) *Catégorie* `diboson`.

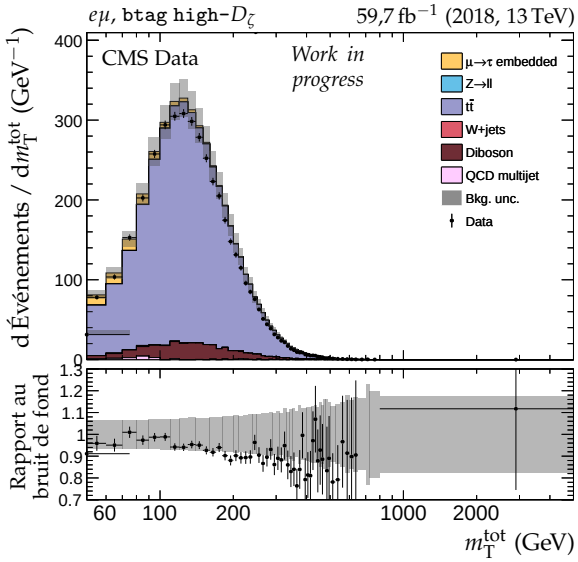


(e) *Catégorie* `qcd`.

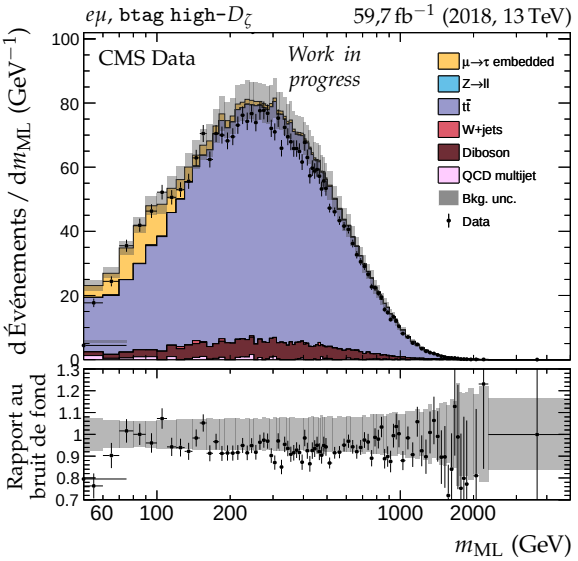


(f) *Catégorie* `misc`.

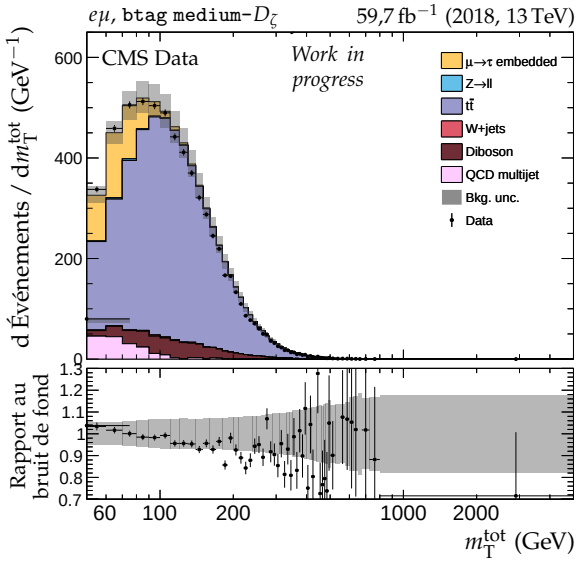
Figure H.39 – *Distributions de* NN$_{score}$ *en 2018 dans le canal eμ.*



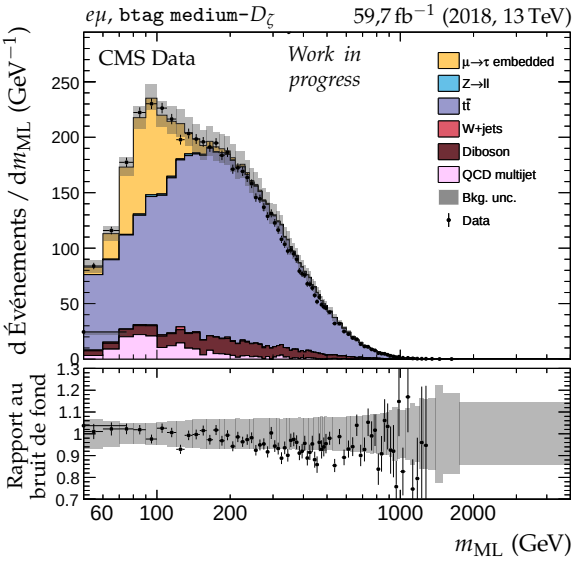
(a) *Catégorie* `btag high-`D_ζ, m_T^{tot}.



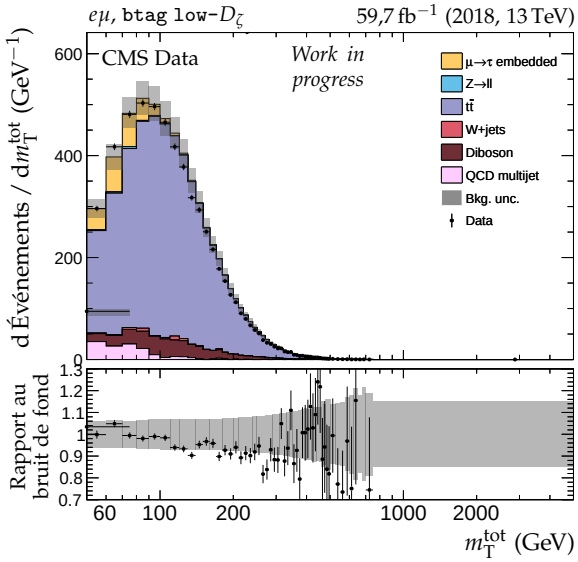
(b) *Catégorie* `btag high-`D_ζ, m_{ML}.



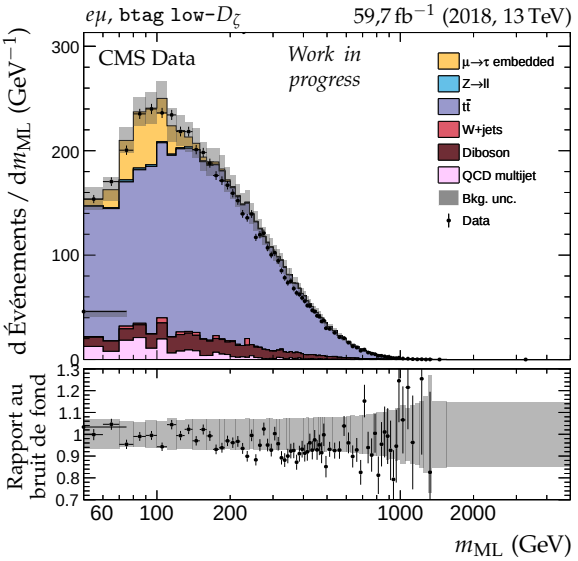
(c) *Catégorie* `btag medium-`D_ζ, m_T^{tot}.



(d) *Catégorie* `btag medium-`D_ζ, m_{ML}.

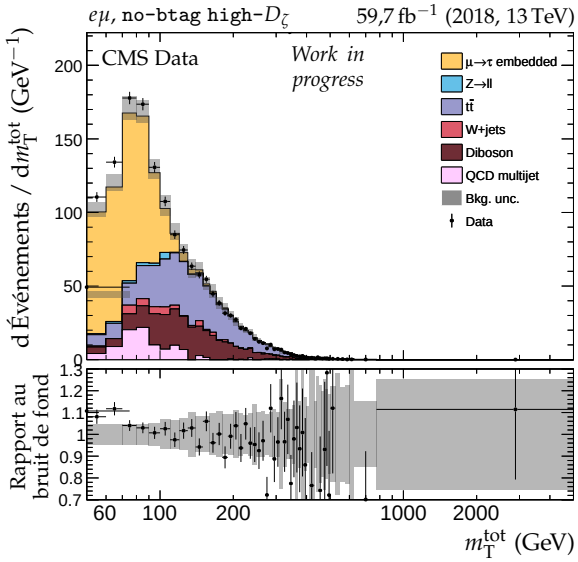


(e) *Catégorie* `btag low-`D_ζ, m_T^{tot}.

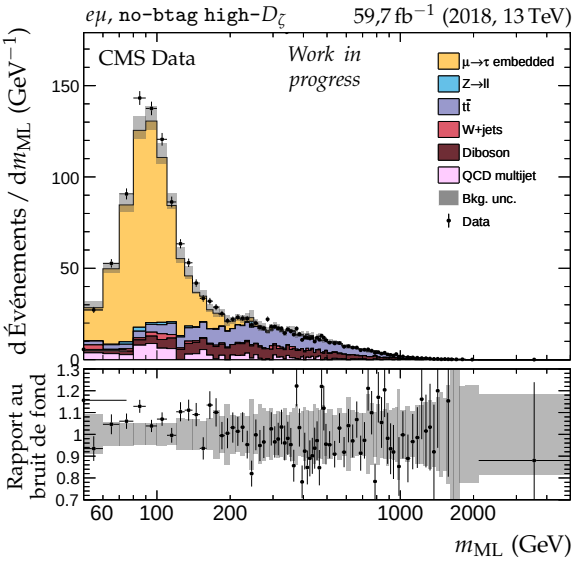


(f) *Catégorie* `btag low-`D_ζ, m_{ML}.

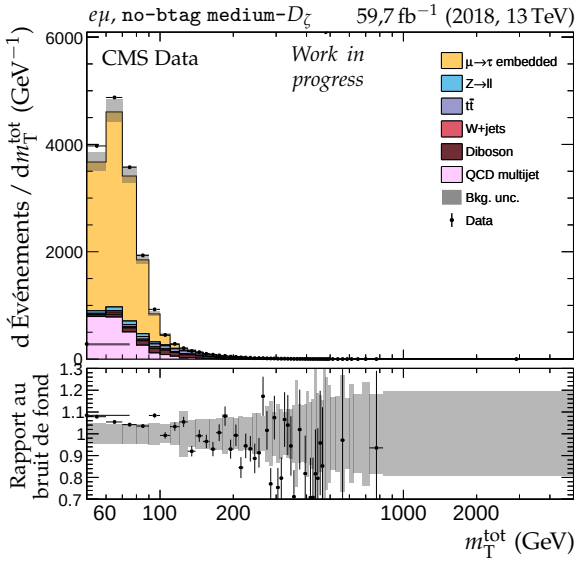
Figure H.40 – *Distributions de* m_T^{tot} *et* m_{ML} *en 2018 dans le canal eμ, catégories* `btag`.



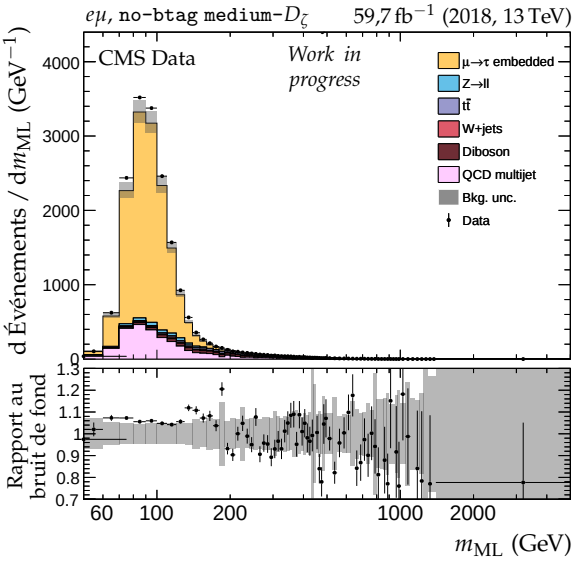
(a) *Catégorie* `no-btag high-`D_ζ, m_T^{tot}.



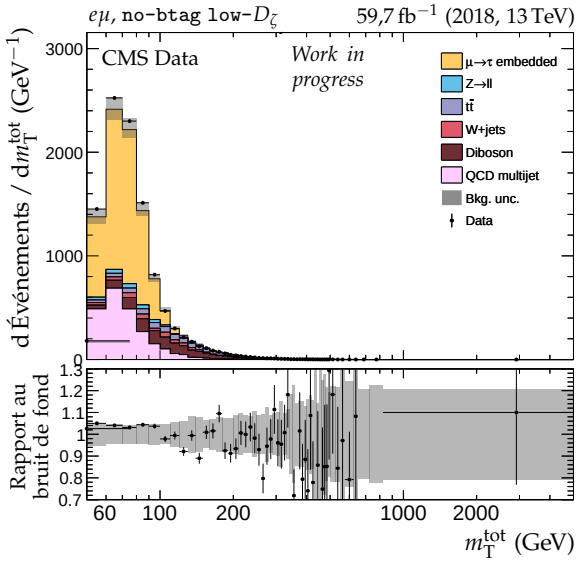
(b) *Catégorie* `no-btag high-`D_ζ, m_{ML}.



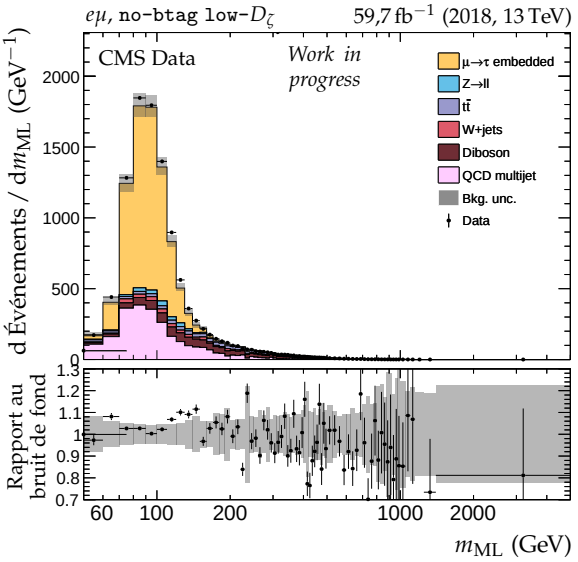
(c) *Catégorie* `no-btag medium-`D_ζ, m_T^{tot}.



(d) *Catégorie* `no-btag medium-`D_ζ, m_{ML}.

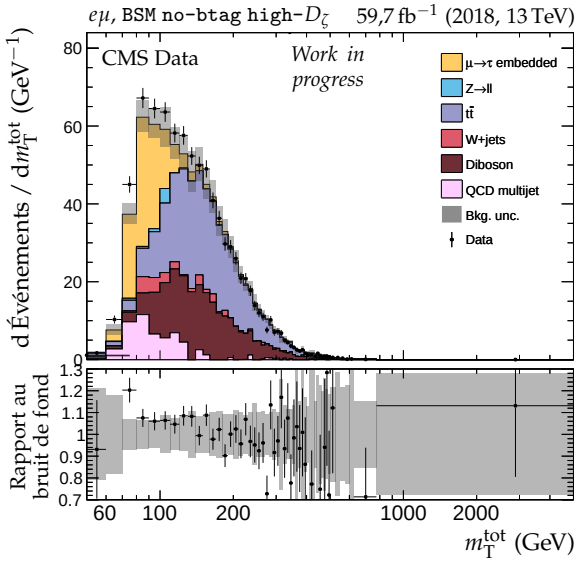


(e) *Catégorie* `no-btag low-`D_ζ, m_T^{tot}.

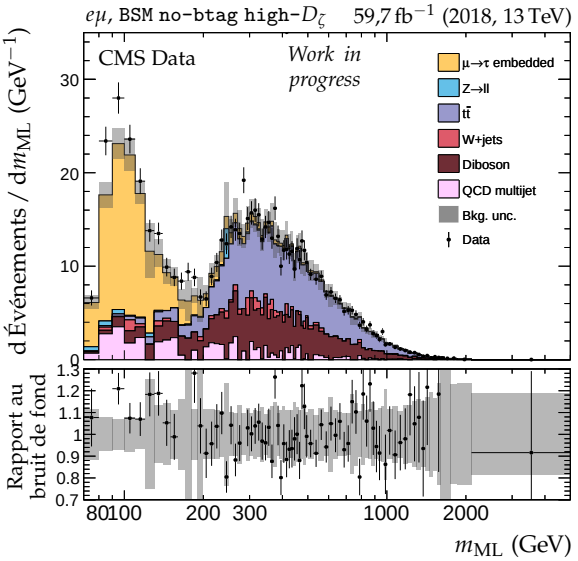


(f) *Catégorie* `no-btag low-`D_ζ, m_{ML}.

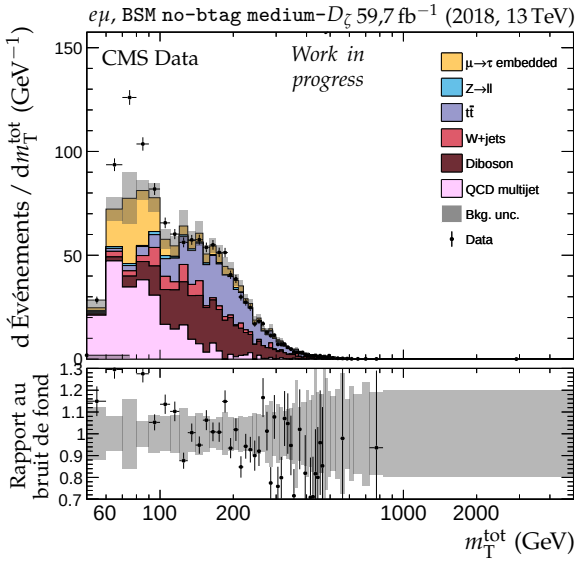
Figure H.41 – *Distributions de m_T^{tot} et m_{ML} en 2018 dans le canal eμ, catégories* `no-btag`.



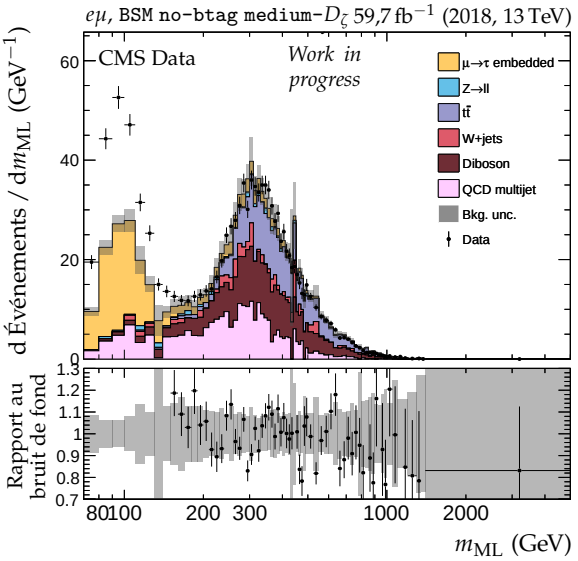
(a) *Catégorie BSM* `no-btag high-`D_ζ, m_T^{tot}.



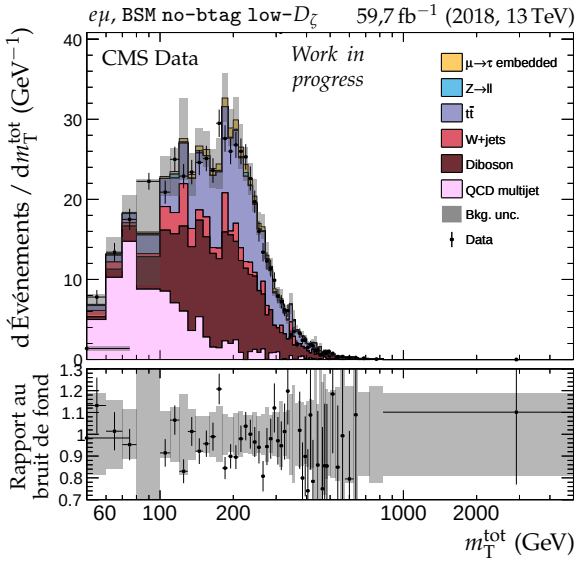
(b) *Catégorie BSM* `no-btag high-`D_ζ, m_{ML}.



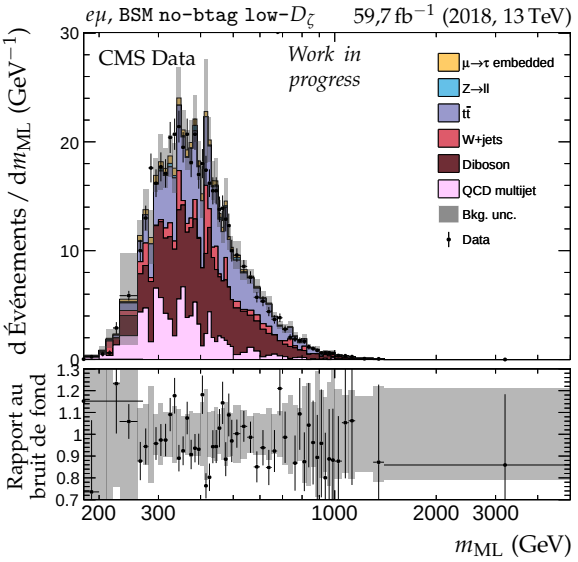
(c) *Catégorie BSM* `no-btag medium-`D_ζ, m_T^{tot}.



(d) *Catégorie BSM* `no-btag medium-`D_ζ, m_{ML}.



(e) *Catégorie BSM* `no-btag low-`D_ζ, m_T^{tot}.



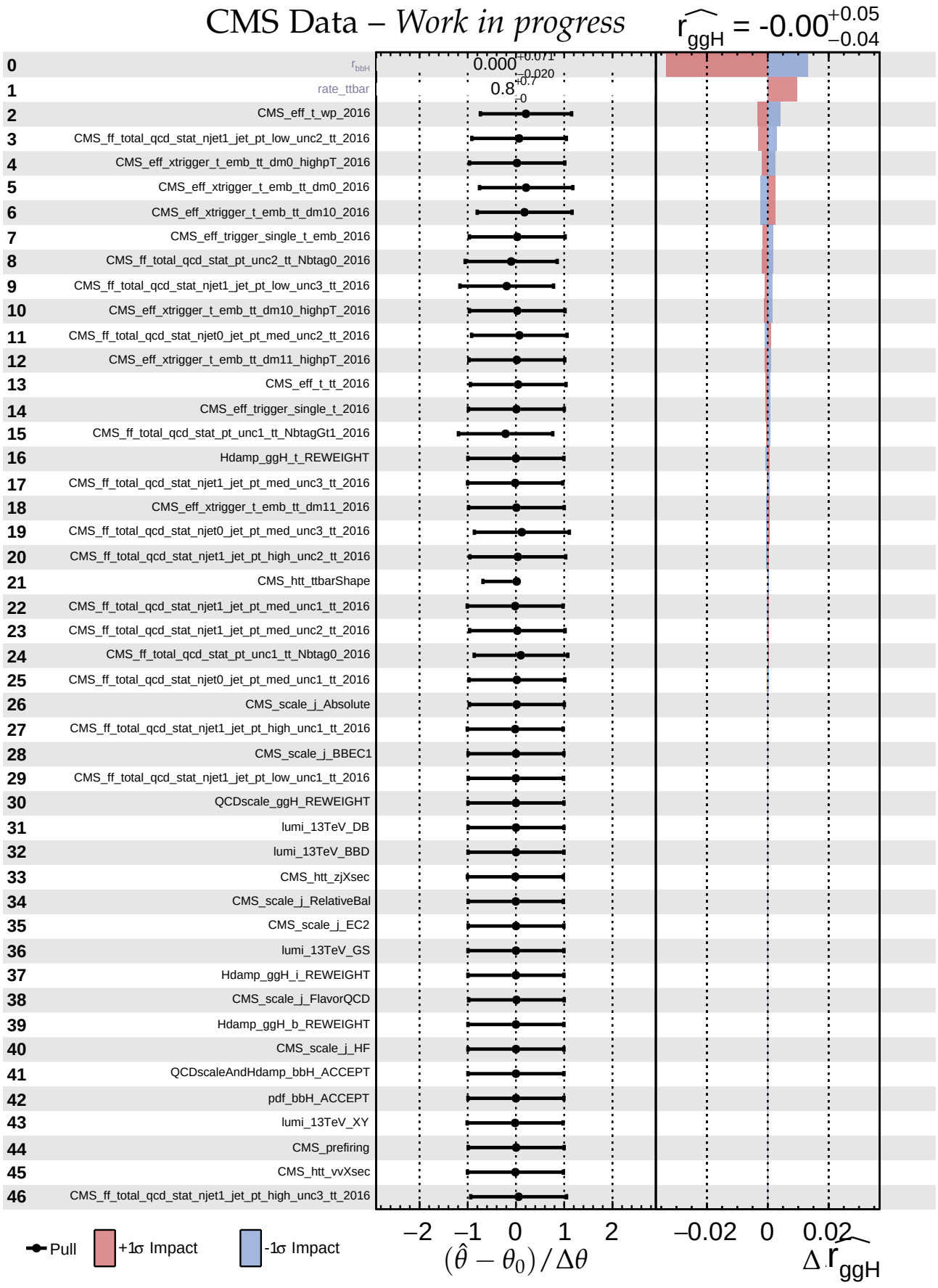
(f) *Catégorie BSM* `no-btag low-`D_ζ, m_{ML}.

Figure H.42 – *Distributions de* m_T^{tot} *et* m_{ML} *en 2018 dans le canal eμ, catégories* `no-btag` *avec* $m_{SVFIT} \geqslant 250\,\text{GeV}$.

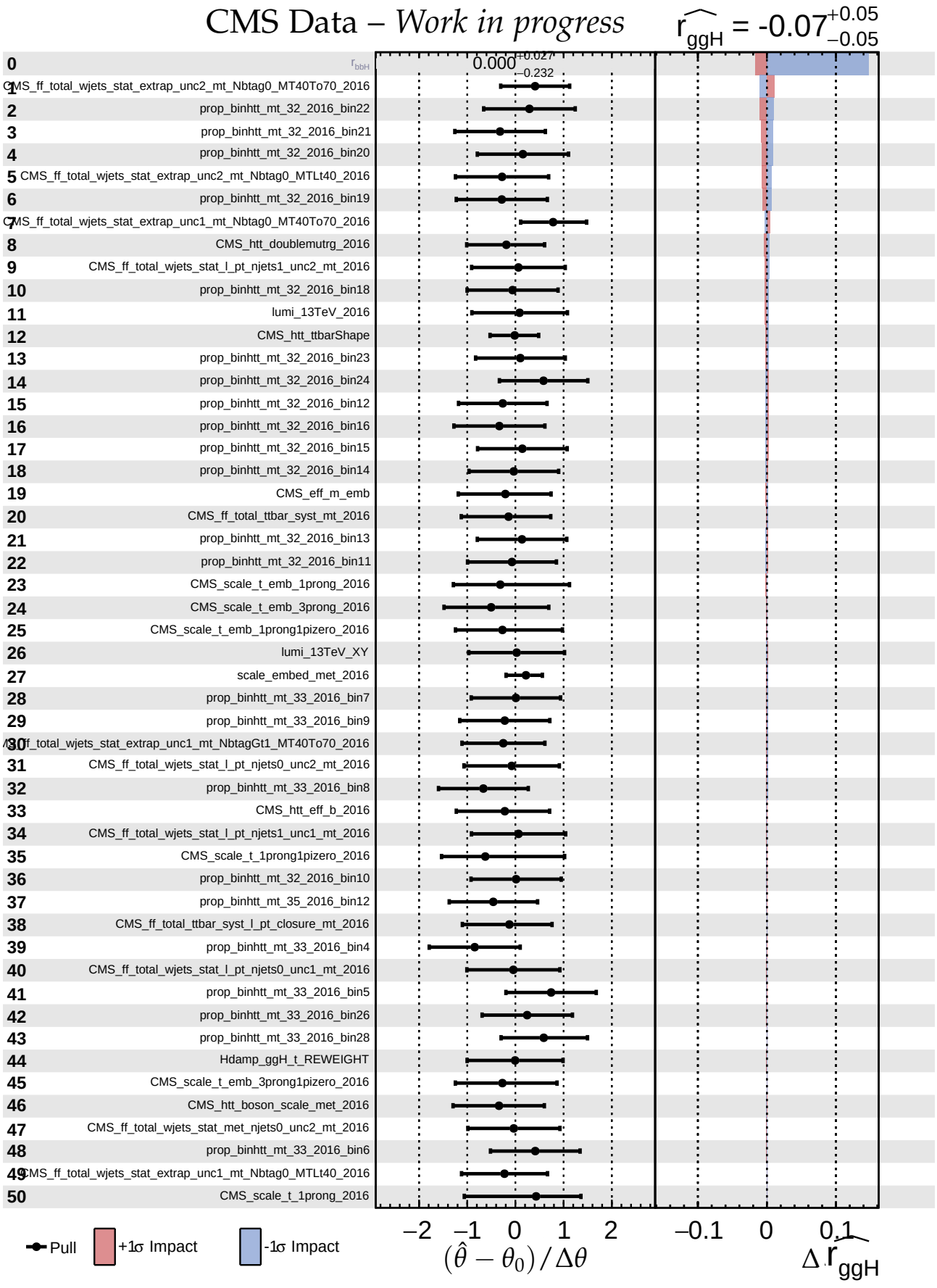
Annexe I
Ajustement des paramètres de nuisance – $H \to \tau\tau$

Les paramètres de nuisance introduits au chapitre 4 sont ajustés par COMBINE, l'outil de combinaison statistique utilisé par la collaboration CMS basé sur ROOSTATS [214].

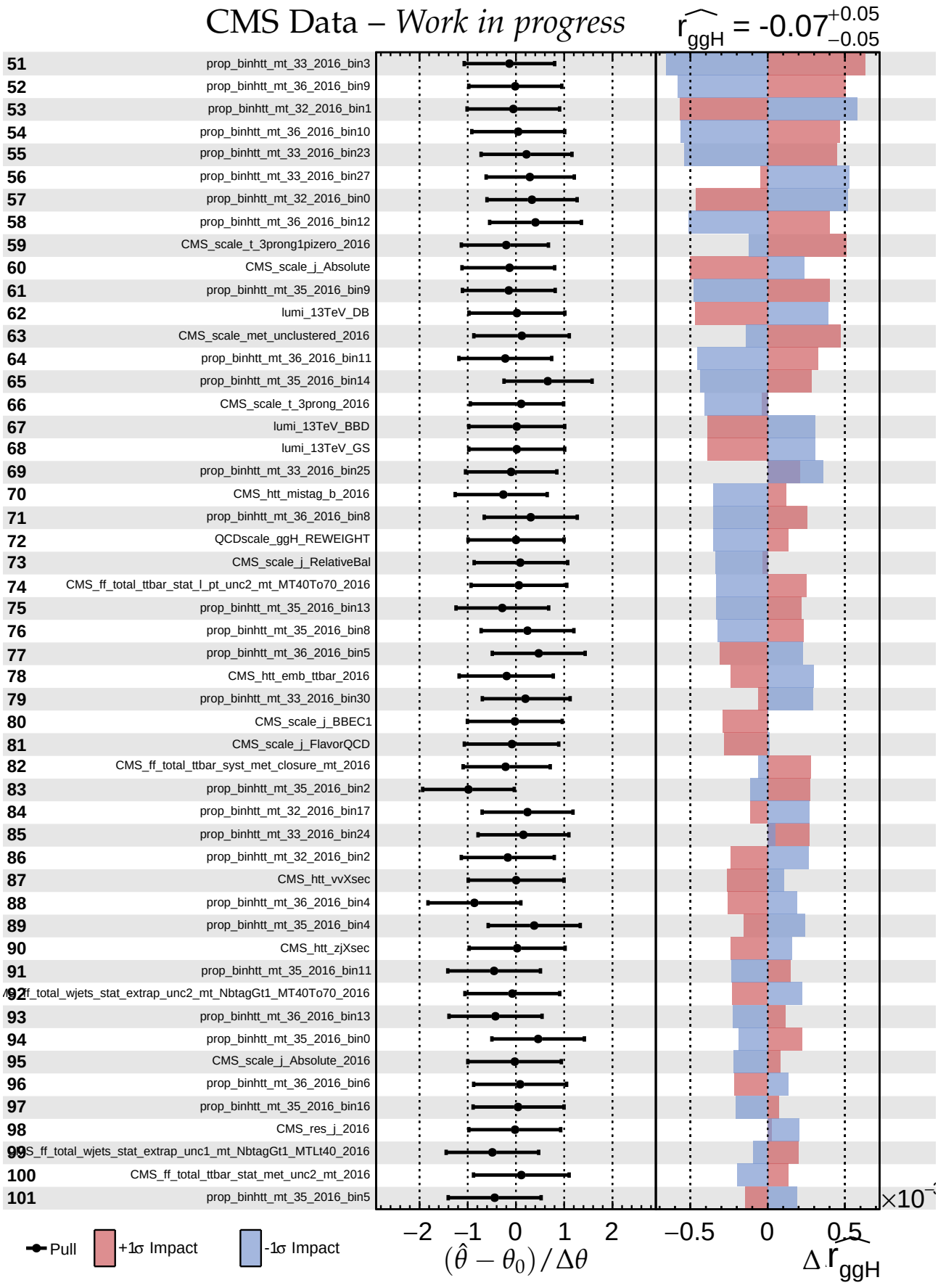
Dans cette annexe, les valeurs prises par ces paramètres après ajustement pour l'obtention des limites d'exclusion indépendantes du modèle $\hat{\theta}$ (*pulls*) ainsi que l'incertitude sur cette valeur sont données en termes de nombre de déviations standard $\Delta\theta$ par rapport à la valeur nominale θ_0. Leur effet sur le paramètre d'intérêt, en l'occurrence le modificateur d'intensité du signal $\mu = r_{ggH}$, est donné par les bandes rouges et bleues. Elles correspondent respectivement aux effets corrélé et anti-corrélé du paramètre de nuisance sur μ ($\pm 1\sigma$ impact).



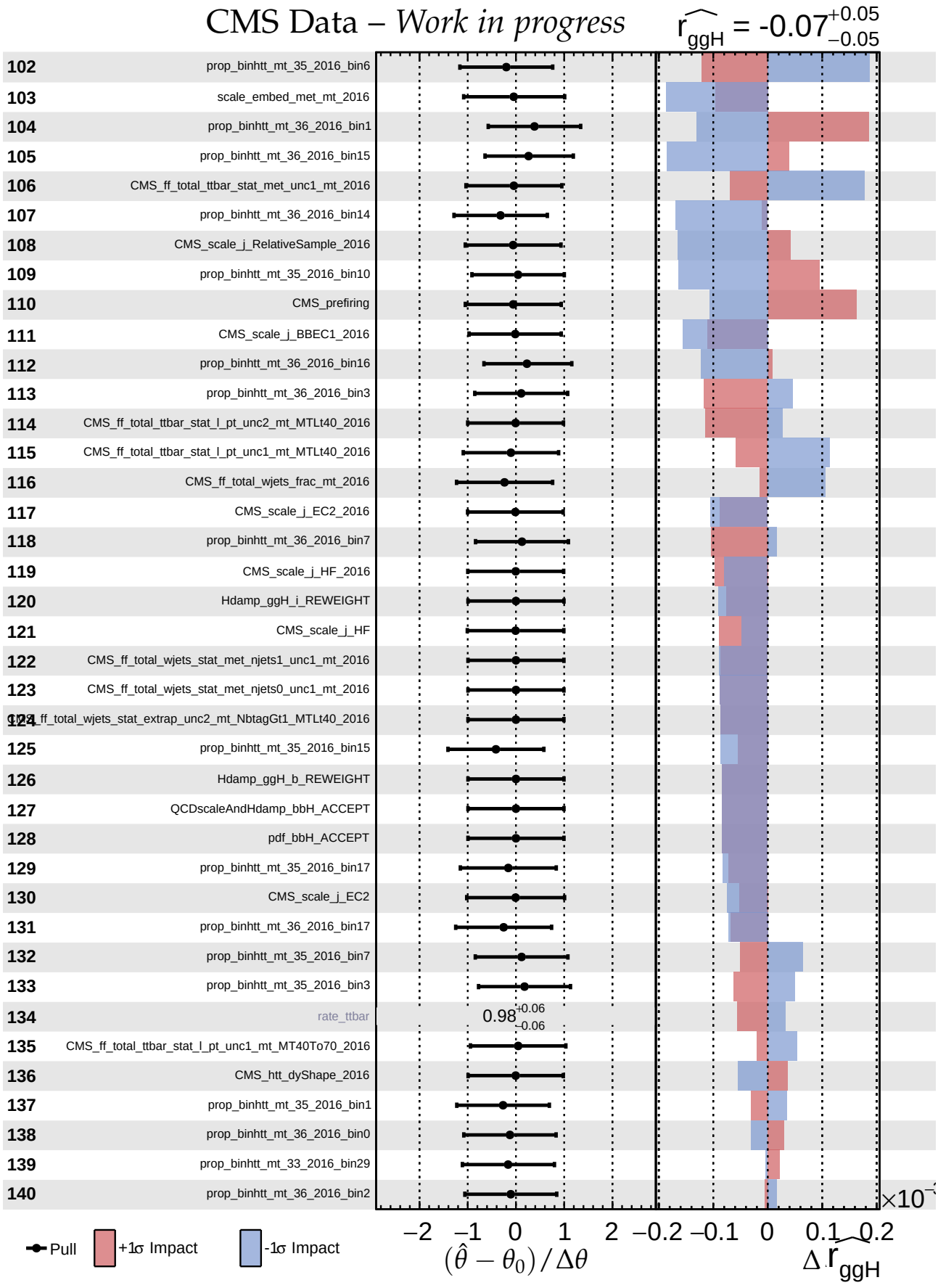
Année 2016, canal $\tau_h\tau_h$, page 1/1



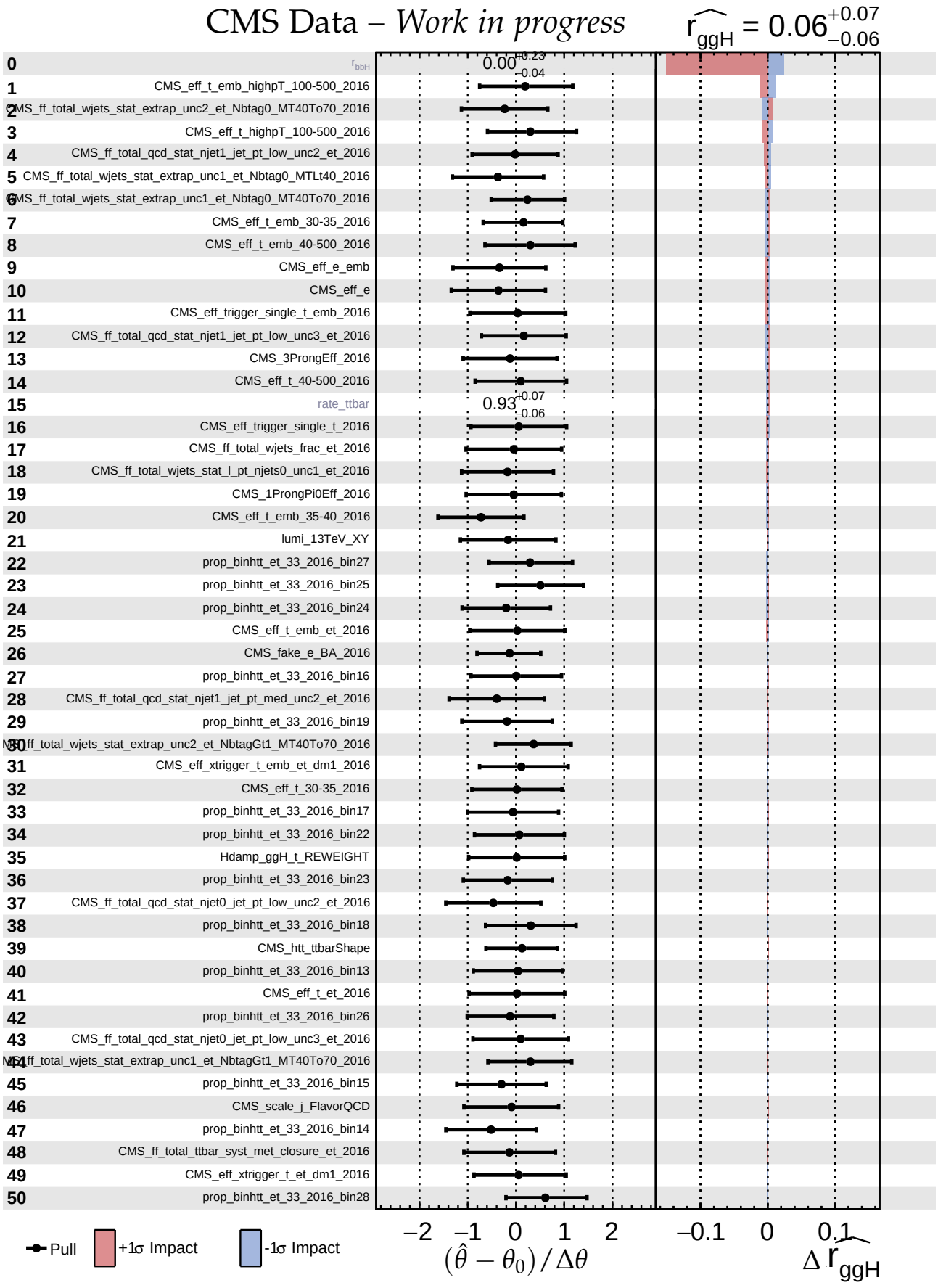
Année 2016, canal $\mu\tau_h$, page 1/3



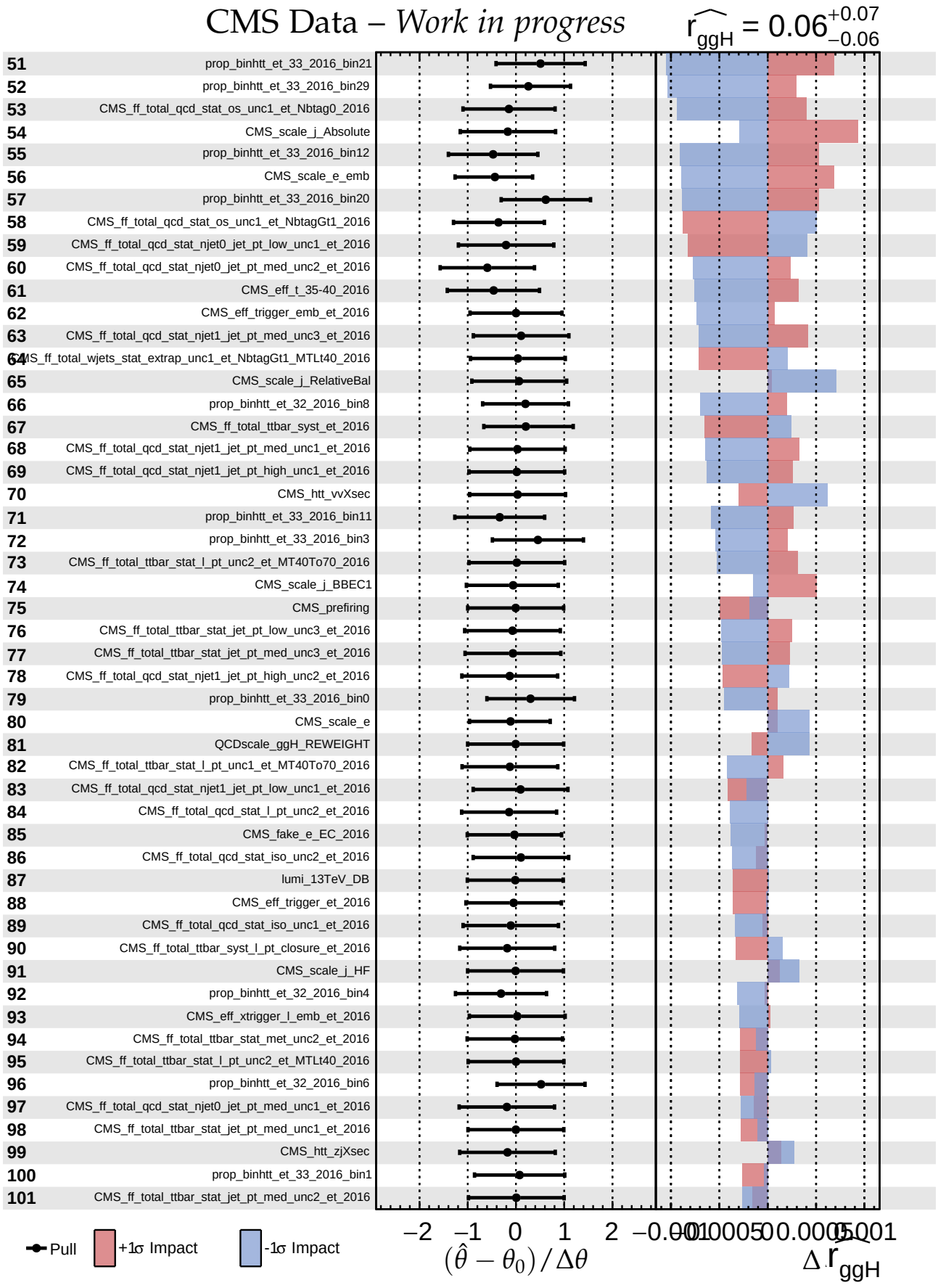
Année 2016, canal $\mu\tau_h$, page 2/3

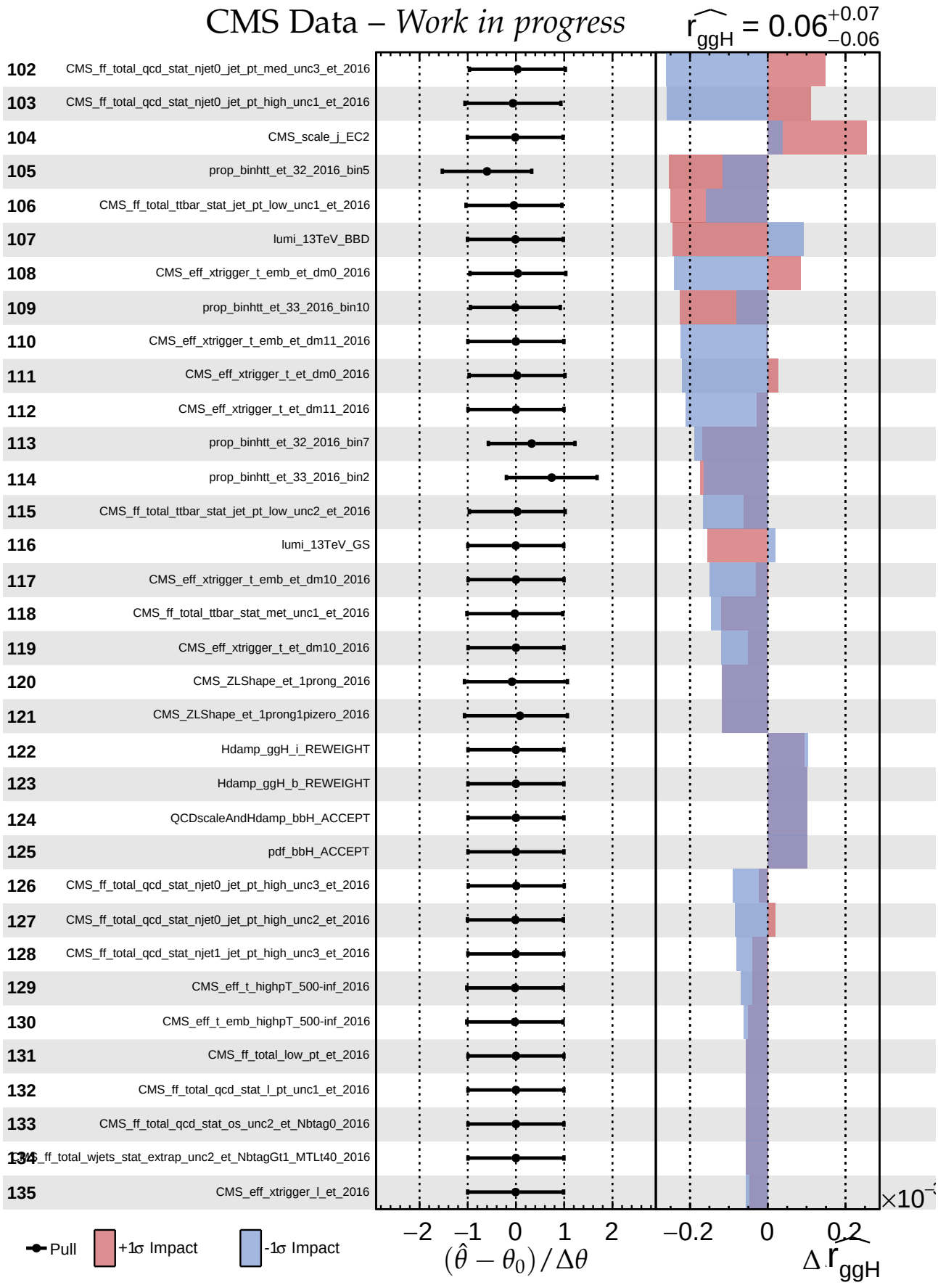


Année 2016, canal $\mu\tau_h$, page 3/3

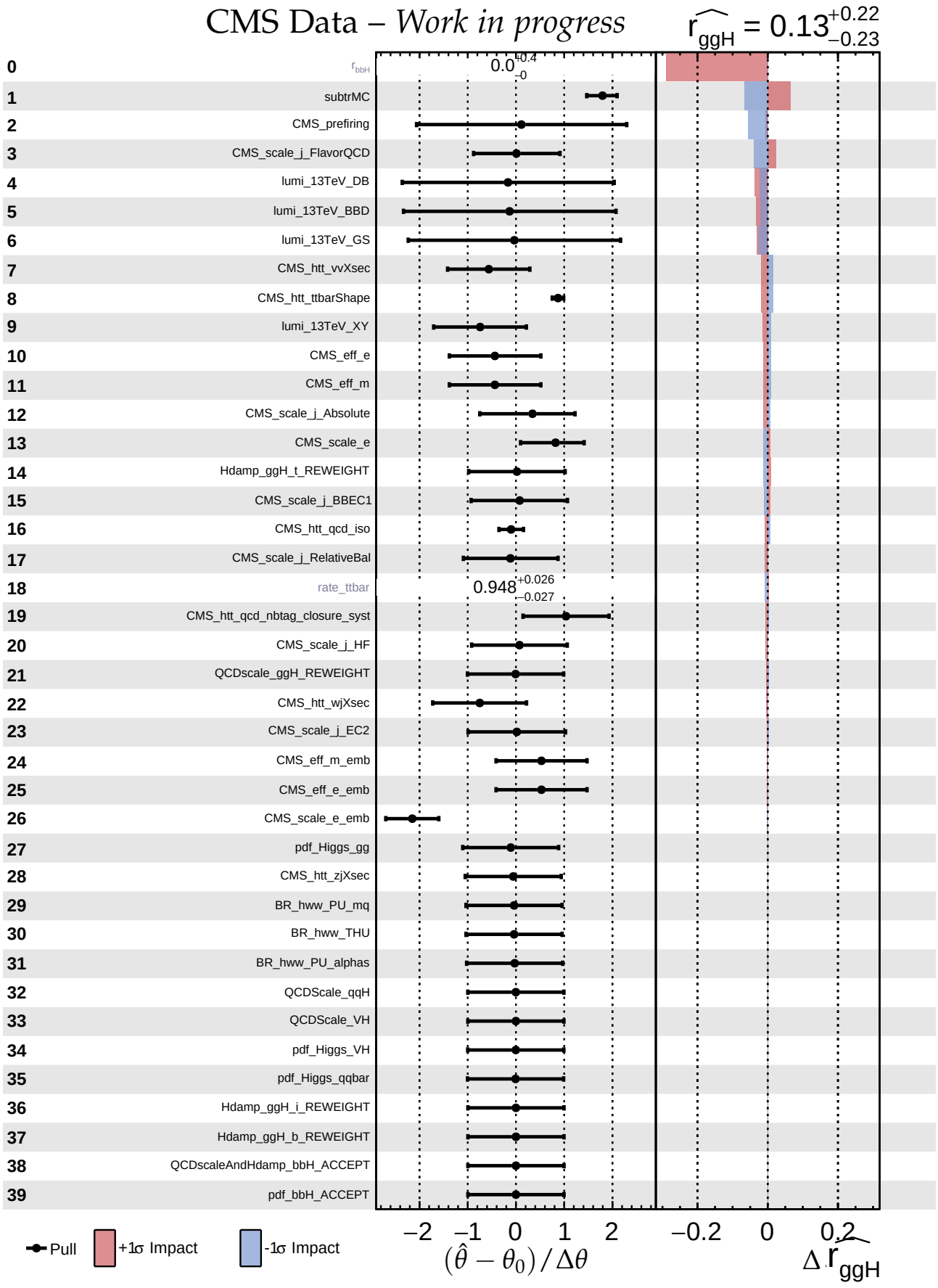


Année 2016, canal $e\tau_h$, page 1/3

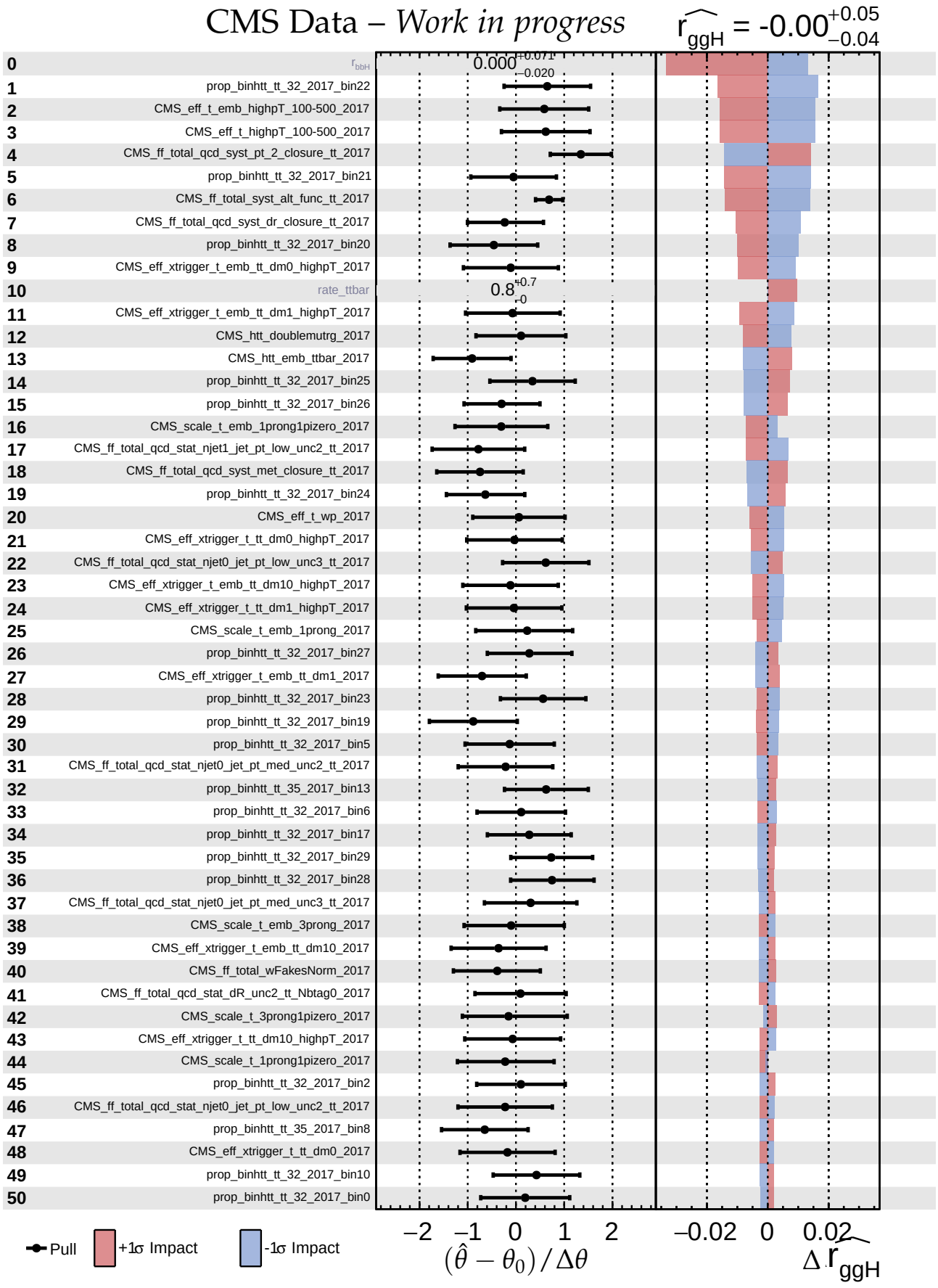

CMS Data – Work in progress
$\widehat{r_{ggH}} = 0.06^{+0.07}_{-0.06}$
51 prop_binhtt_et_33_2016_bin21
52 prop_binhtt_et_33_2016_bin29
53 CMS_ff_total_qcd_stat_os_unc1_et_Nbtag0_2016
54 CMS_scale_j_Absolute
55 prop_binhtt_et_33_2016_bin12
56 CMS_scale_e_emb
57 prop_binhtt_et_33_2016_bin20
58 CMS_ff_total_qcd_stat_os_unc1_et_NbtagGt1_2016
59 CMS_ff_total_qcd_stat_njet0_jet_pt_low_unc1_et_2016
60 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc2_et_2016
61 CMS_eff_t_35-40_2016
62 CMS_eff_trigger_emb_et_2016
63 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc3_et_2016
64 CMS_ff_total_wjets_stat_extrap_unc1_et_NbtagGt1_MTLt40_2016
65 CMS_scale_j_RelativeBal
66 prop_binhtt_et_32_2016_bin8
67 CMS_ff_total_ttbar_syst_et_2016
68 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc1_et_2016
69 CMS_ff_total_qcd_stat_njet1_jet_pt_high_unc1_et_2016
70 CMS_htt_vvXsec
71 prop_binhtt_et_33_2016_bin11
72 prop_binhtt_et_33_2016_bin3
73 CMS_ff_total_ttbar_stat_l_pt_unc2_et_MT40To70_2016
74 CMS_scale_j_BBEC1
75 CMS_prefiring
76 CMS_ff_total_ttbar_stat_jet_pt_low_unc3_et_2016
77 CMS_ff_total_ttbar_stat_jet_pt_med_unc3_et_2016
78 CMS_ff_total_qcd_stat_njet1_jet_pt_high_unc2_et_2016
79 prop_binhtt_et_33_2016_bin0
80 CMS_scale_e
81 QCDscale_ggH_REWEIGHT
82 CMS_ff_total_ttbar_stat_l_pt_unc1_et_MT40To70_2016
83 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc1_et_2016
84 CMS_ff_total_qcd_stat_l_pt_unc2_et_2016
85 CMS_fake_e_EC_2016
86 CMS_ff_total_qcd_stat_iso_unc2_et_2016
87 lumi_13TeV_DB
88 CMS_eff_trigger_et_2016
89 CMS_ff_total_qcd_stat_iso_unc1_et_2016
90 CMS_ff_total_ttbar_syst_l_pt_closure_et_2016
91 CMS_scale_j_HF
92 prop_binhtt_et_32_2016_bin4
93 CMS_eff_xtrigger_l_emb_et_2016
94 CMS_ff_total_ttbar_stat_met_unc2_et_2016
95 CMS_ff_total_ttbar_stat_l_pt_unc2_et_MTLt40_2016
96 prop_binhtt_et_32_2016_bin6
97 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc1_et_2016
98 CMS_ff_total_ttbar_stat_jet_pt_med_unc1_et_2016
99 CMS_htt_zjXsec
100 prop_binhtt_et_33_2016_bin1
101 CMS_ff_total_ttbar_stat_jet_pt_med_unc2_et_2016
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.001 0.0005 0.0005 0.001
Δr_{ggH}



Année 2016, canal $e\tau_h$, page 3/3



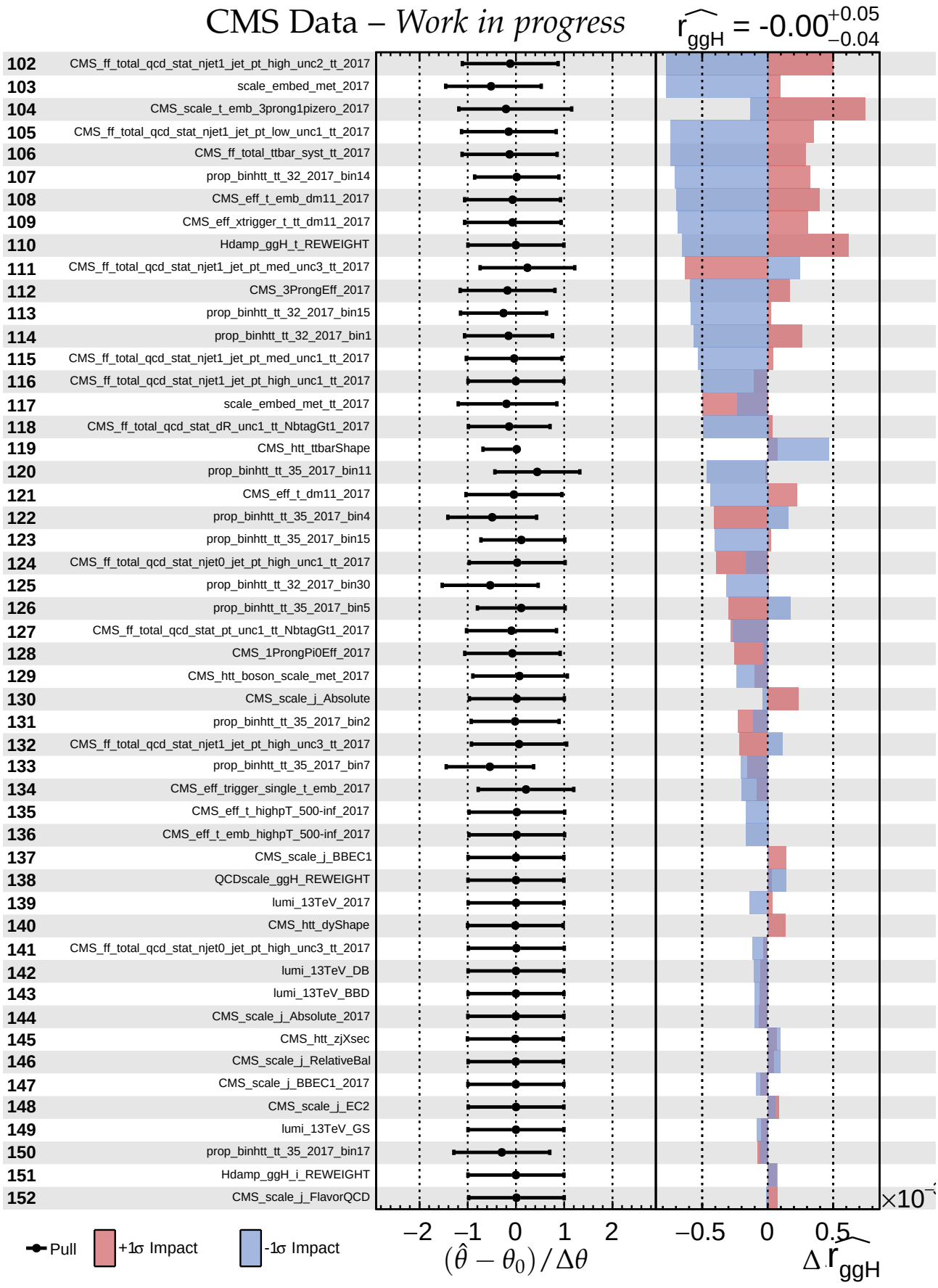
Année 2016, canal $e\mu$, page 1/1



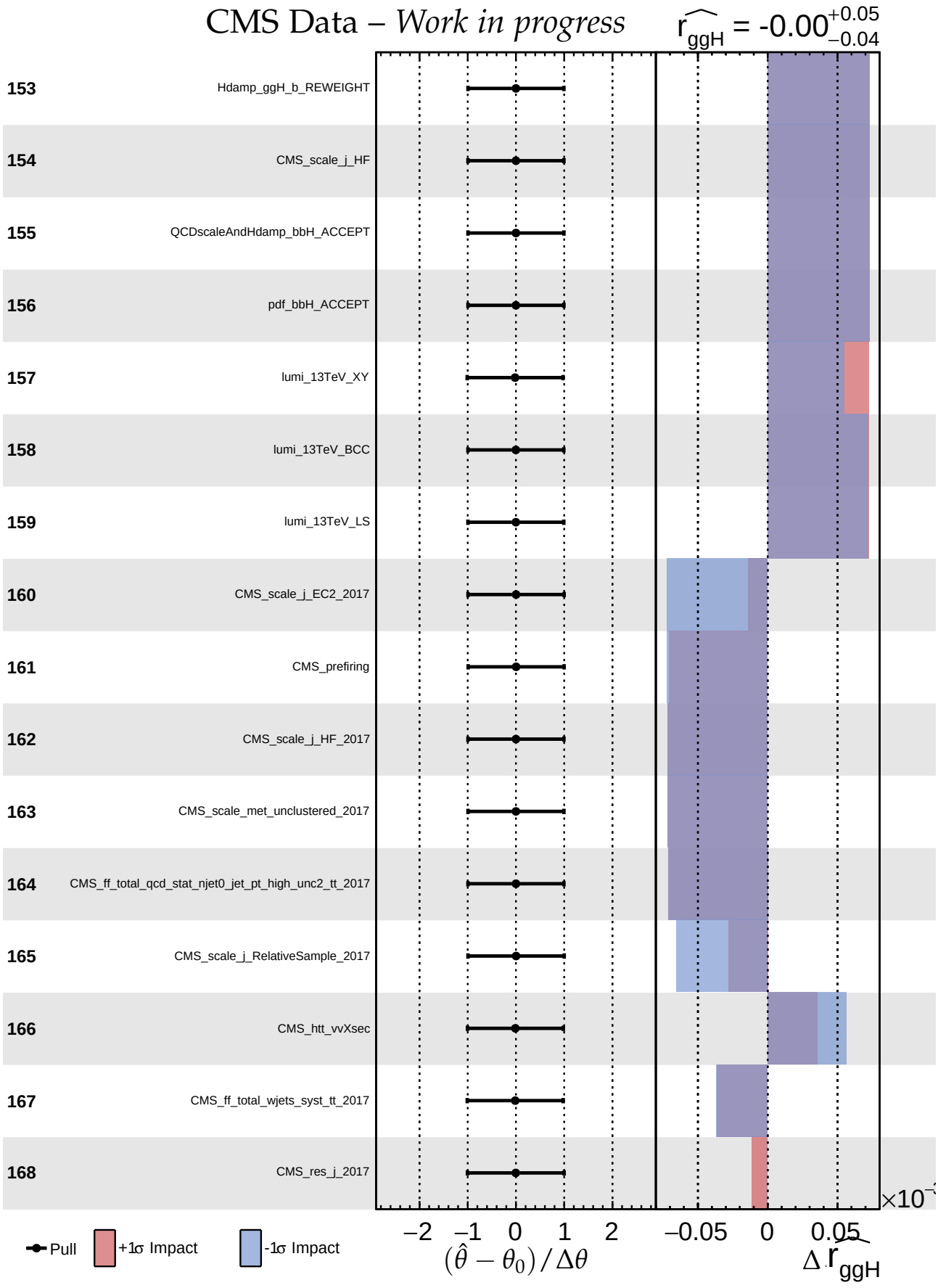
Année 2017, canal $\tau_h\tau_h$, page 1/4



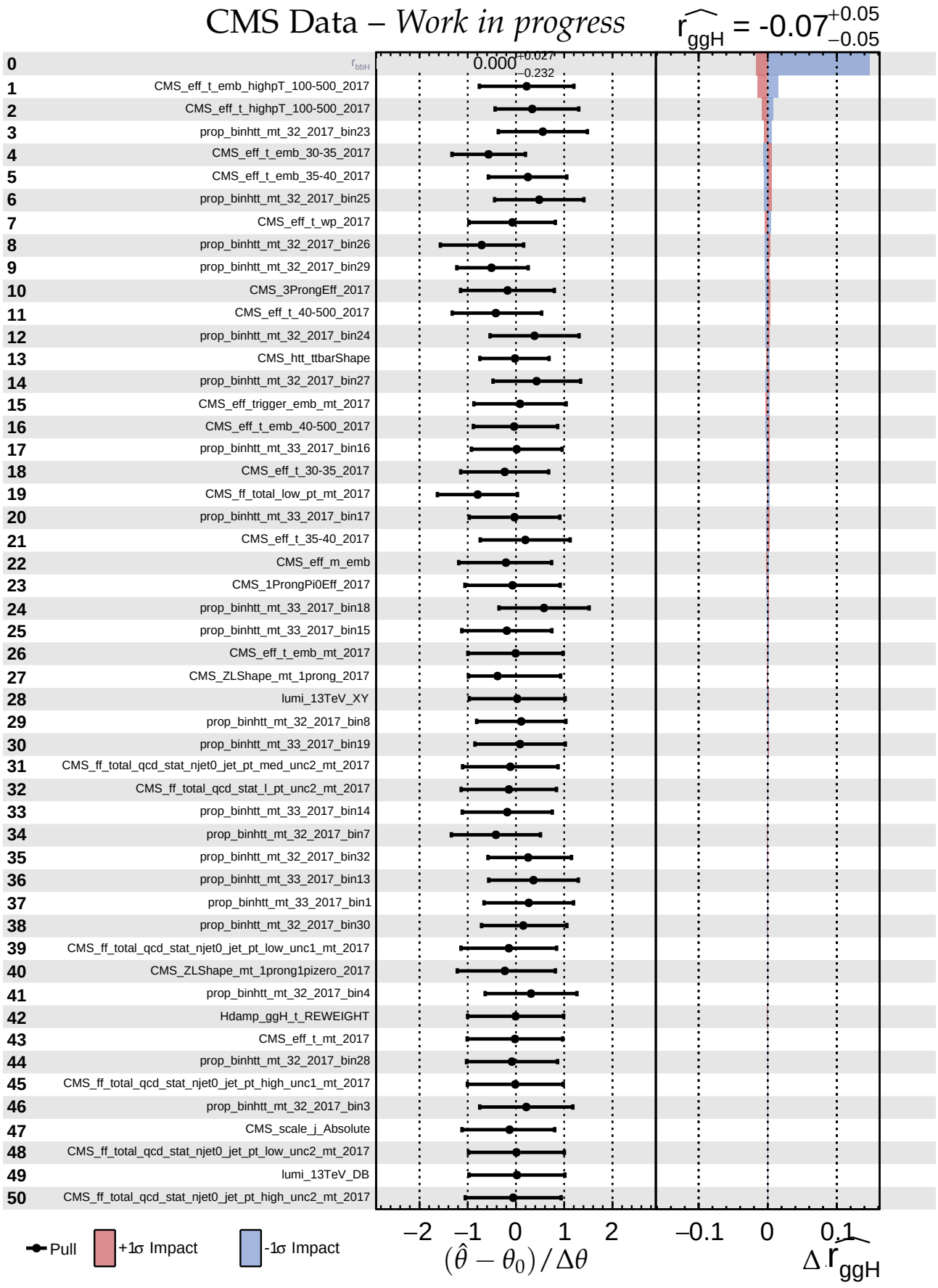
Année 2017, canal $\tau_h\tau_h$, page 2/4



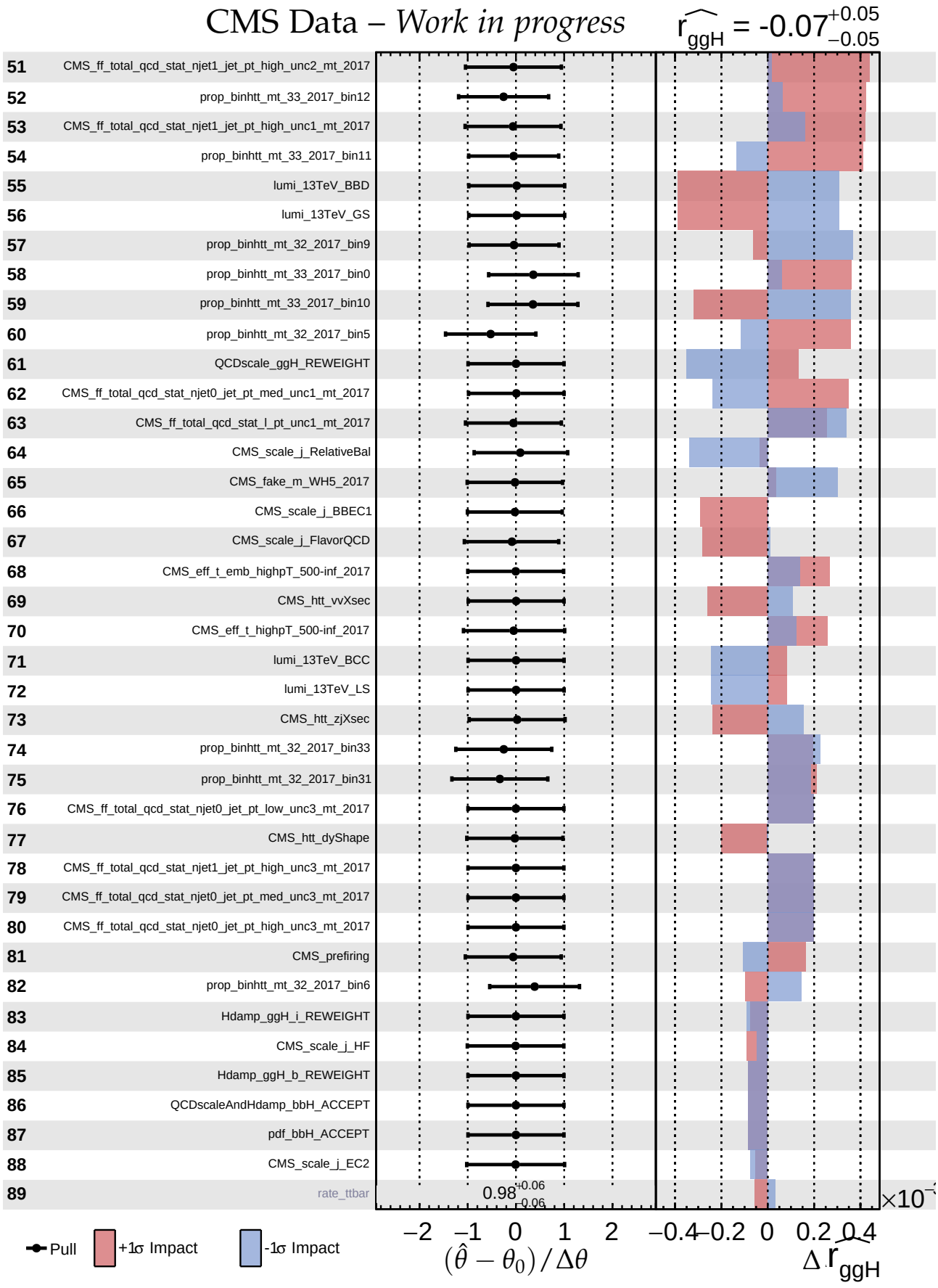
Année 2017, canal $\tau_h\tau_h$, page 3/4



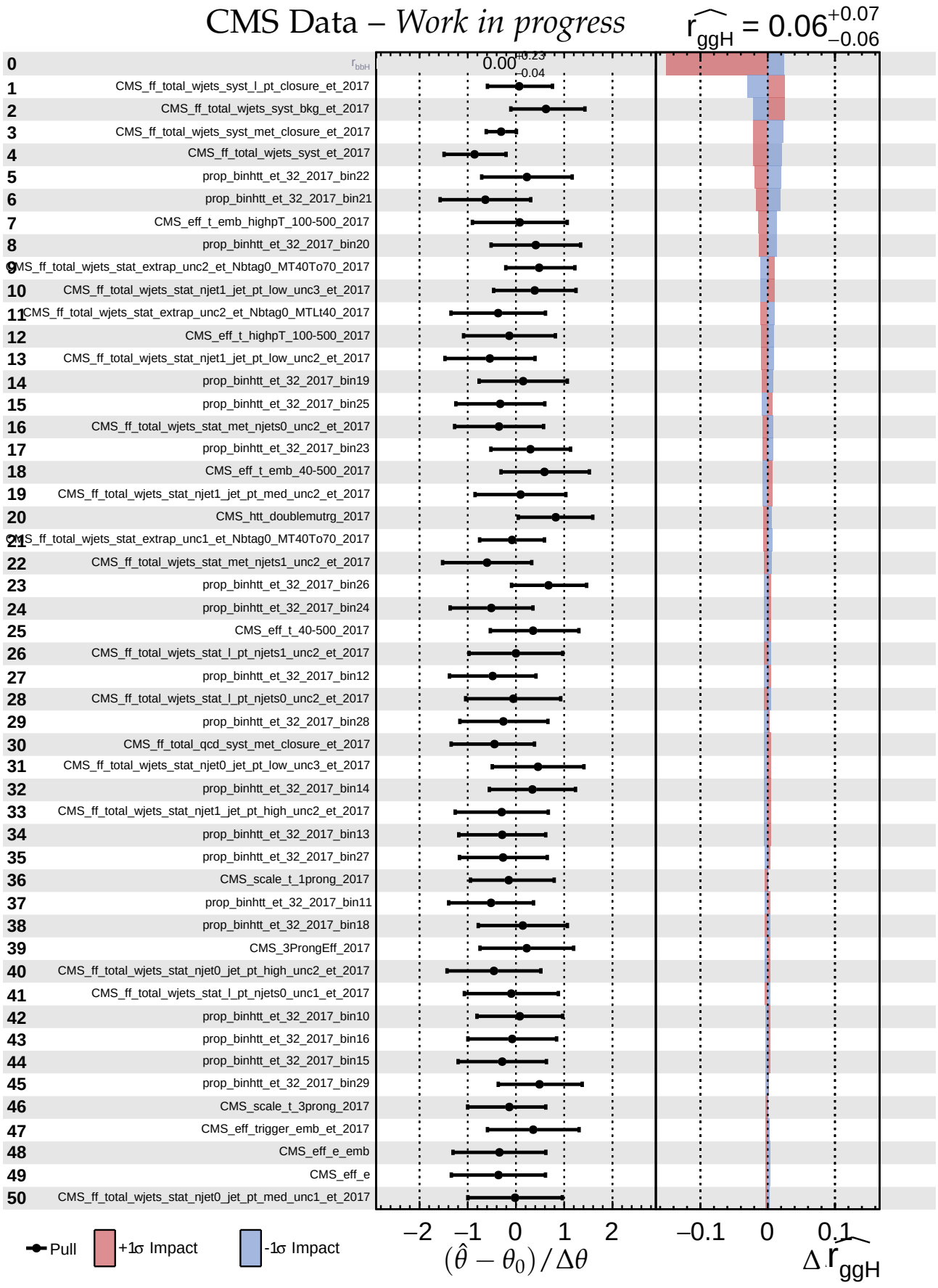
Année 2017, canal $\tau_h\tau_h$, page 4/4



Année 2017, canal $\mu\tau_h$, page 1/2

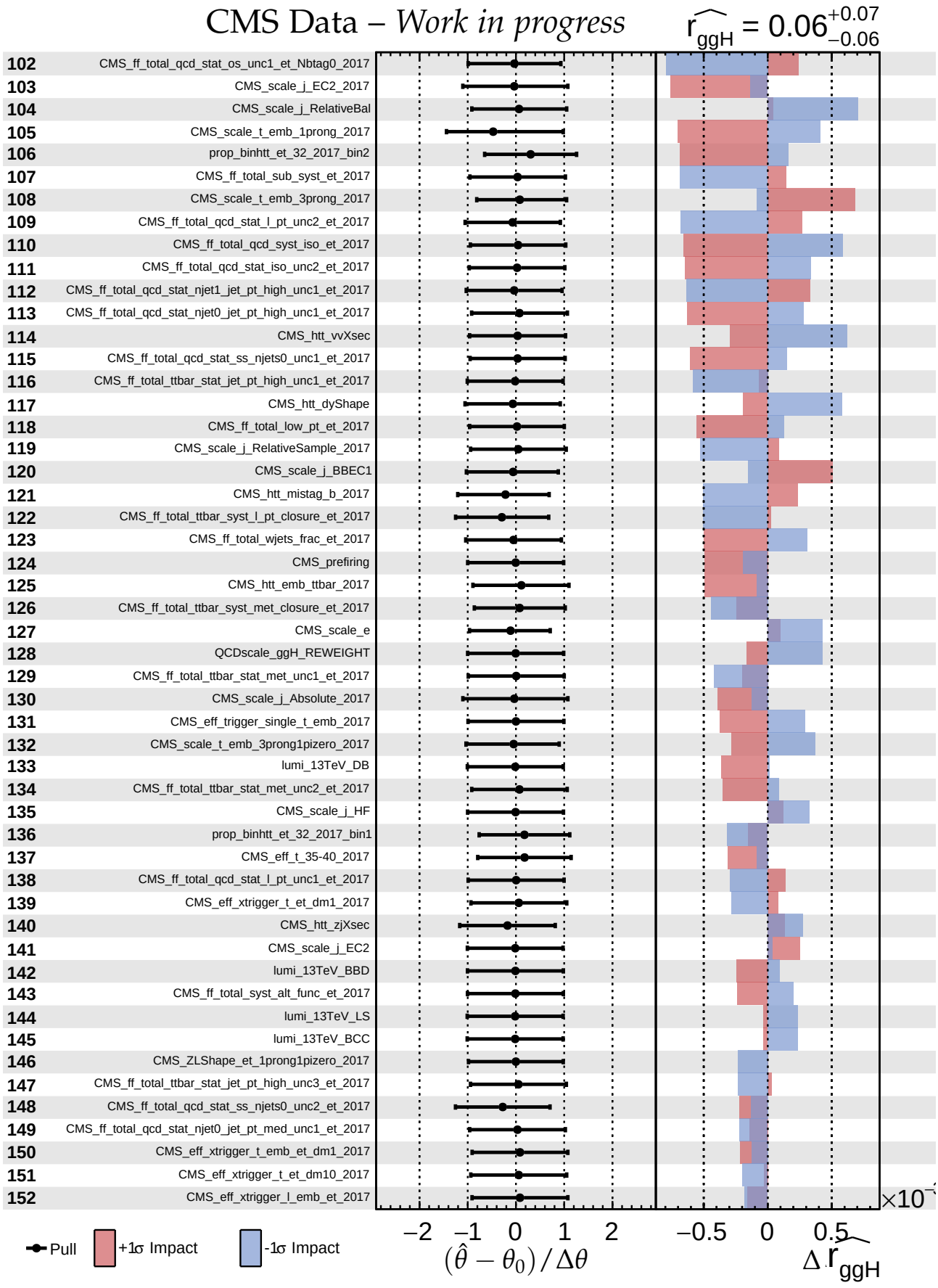


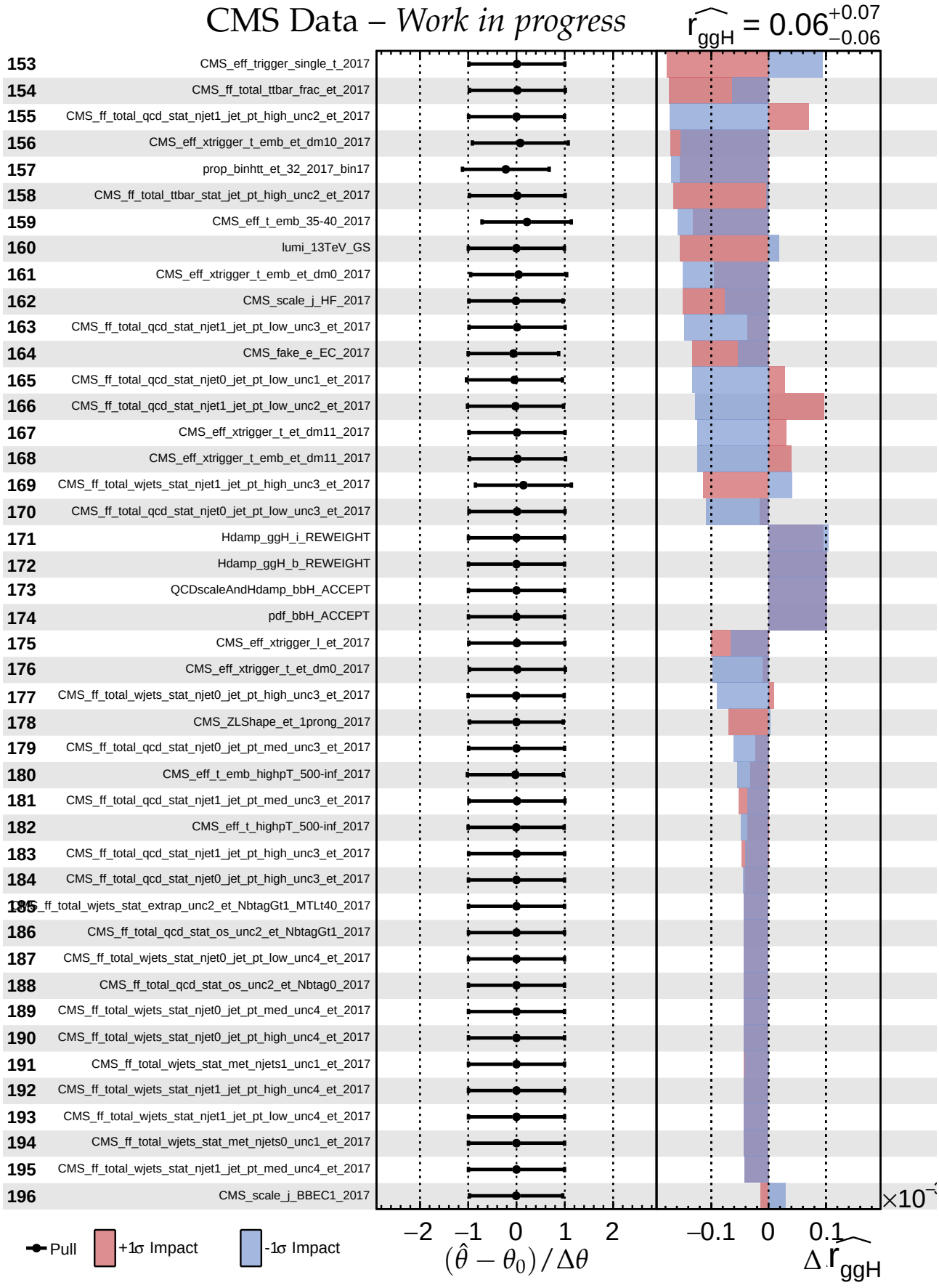
Année 2017, canal $\mu\tau_h$, page 2/2

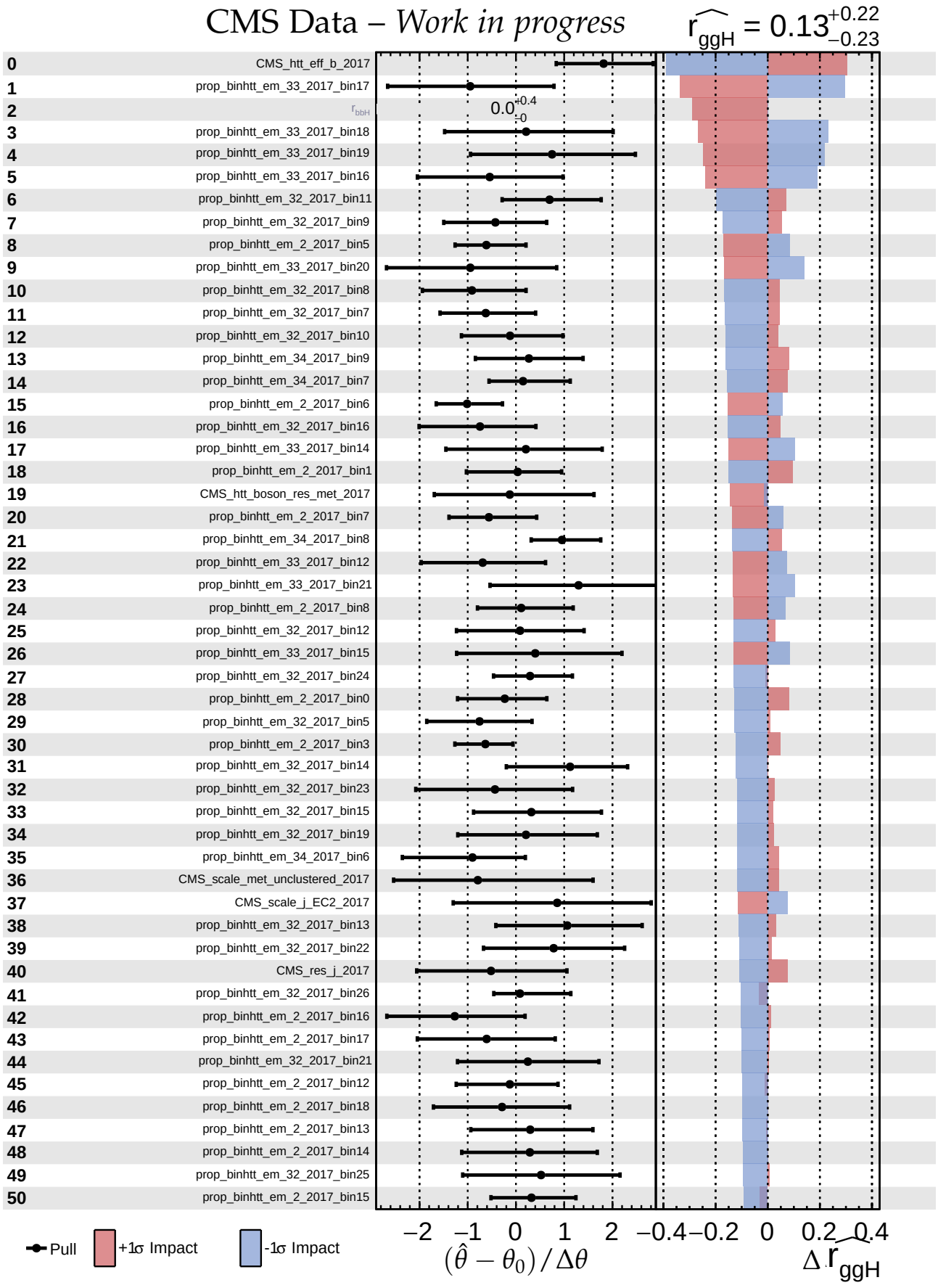


Année 2017, canal $e\tau_h$, page 1/4

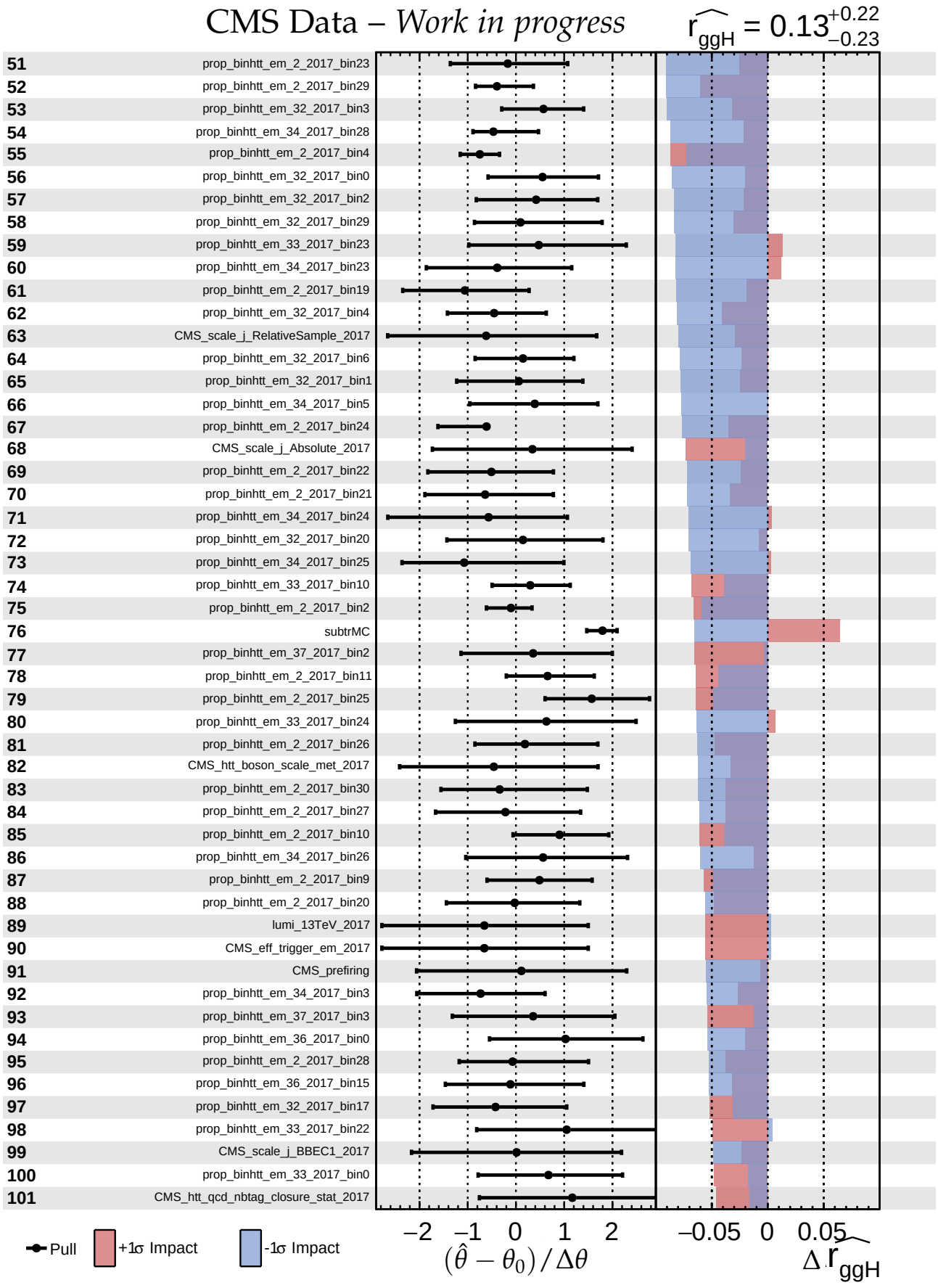

CMS Data – Work in progress
$\hat{r}_{ggH} = 0.06^{+0.07}_{-0.06}$
51 CMS_ff_total_wjets_stat_l_pt_njets1_unc1_et_2017
52 CMS_ff_total_wjets_stat_njet0_jet_pt_high_unc1_et_2017
53 CMS_ff_total_wjets_stat_njet1_jet_pt_med_unc1_et_2017
54 rate_ttbar
$0.93^{+0.07}_{-0.06}$
55 CMS_ff_total_wjets_stat_extrap_unc1_et_Nbtag0_MTLt40_2017
56 CMS_ff_total_wjets_stat_njet0_jet_pt_low_unc1_et_2017
57 CMS_eff_t_emb_30-35_2017
58 CMS_ff_total_wjets_stat_njet0_jet_pt_low_unc2_et_2017
59 CMS_ff_total_qcd_stat_njet0_jet_pt_low_unc2_et_2017
60 CMS_ff_total_wjets_stat_extrap_unc1_et_NbtagGt1_MTLt40_2017
61 CMS_ff_total_wjets_stat_njet1_jet_pt_high_unc1_et_2017
62 CMS_1ProngPi0Eff_2017
63 CMS_ff_total_wjets_stat_extrap_unc1_et_NbtagGt1_MT40To70_2017
64 CMS_ff_total_qcd_stat_njet0_jet_pt_high_unc2_et_2017
65 CMS_scale_t_3prong1pizero_2017
66 CMS_ff_total_qcd_syst_et_2017
67 CMS_ff_total_wjets_stat_njet1_jet_pt_low_unc1_et_2017
68 CMS_htt_eff_b_2017
69 CMS_ff_total_ttbar_syst_et_2017
70 lumi_13TeV_XY
71 CMS_ff_total_wjets_stat_njet0_jet_pt_med_unc3_et_2017
72 CMS_ff_total_qcd_stat_ss_njets1_unc2_et_2017
73 CMS_ff_total_wjets_stat_njet1_jet_pt_med_unc3_et_2017
74 CMS_ff_total_wjets_stat_njet0_jet_pt_med_unc2_et_2017
75 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc2_et_2017
76 CMS_ff_total_qcd_stat_os_unc1_et_NbtagGt1_2017
77 CMS_htt_boson_res_met_2017
78 CMS_scale_t_1prong1pizero_2017
79 Hdamp_ggH_t_REWEIGHT
80 CMS_eff_t_30-35_2017
81 CMS_eff_t_emb_et_2017
82 CMS_htt_ttbarShape
83 CMS_res_j_2017
84 CMS_scale_t_emb_1prong1pizero_2017
85 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc1_et_2017
86 CMS_ff_total_qcd_syst_bkg_et_2017
87 CMS_scale_j_FlavorQCD
88 CMS_ff_total_qcd_stat_ss_njets1_unc1_et_2017
89 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc2_et_2017
90 CMS_scale_met_unclustered_2017
91 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc1_et_2017
92 prop_binhtt_et_32_2017_bin0
93 CMS_ff_total_wjets_stat_extrap_unc2_et_NbtagGt1_MT40To70_2017
94 CMS_scale_j_Absolute
95 CMS_htt_boson_scale_met_2017
96 CMS_scale_e_emb
97 lumi_13TeV_2017
98 CMS_eff_trigger_et_2017
99 CMS_fake_e_BA_2017
100 CMS_ff_total_qcd_stat_iso_unc1_et_2017
101 CMS_eff_t_et_2017
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.002 0 0.002
$\Delta \hat{r}_{ggH}$

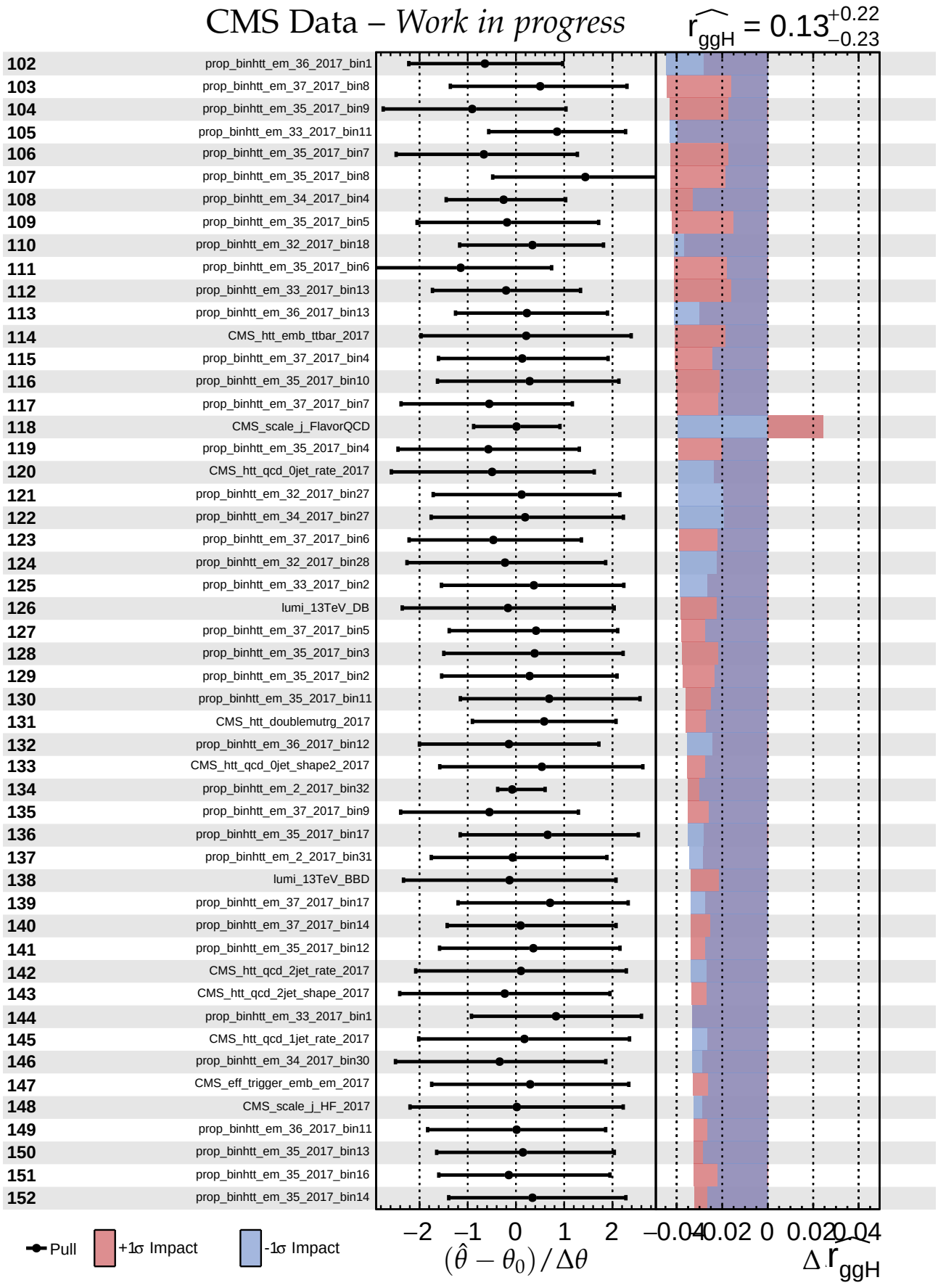

CMS Data – Work in progress
$\hat{r}_{ggH} = 0.06^{+0.07}_{-0.06}$
102 CMS_ff_total_qcd_stat_os_unc1_et_Nbtag0_2017
103 CMS_scale_j_EC2_2017
104 CMS_scale_j_RelativeBal
105 CMS_scale_t_emb_1prong_2017
106 prop_binhtt_et_32_2017_bin2
107 CMS_ff_total_sub_syst_et_2017
108 CMS_scale_t_emb_3prong_2017
109 CMS_ff_total_qcd_stat_l_pt_unc2_et_2017
110 CMS_ff_total_qcd_syst_iso_et_2017
111 CMS_ff_total_qcd_stat_iso_unc2_et_2017
112 CMS_ff_total_qcd_stat_njet1_jet_pt_high_unc1_et_2017
113 CMS_ff_total_qcd_stat_njet0_jet_pt_high_unc1_et_2017
114 CMS_htt_vvXsec
115 CMS_ff_total_qcd_stat_ss_njets0_unc1_et_2017
116 CMS_ff_total_ttbar_stat_jet_pt_high_unc1_et_2017
117 CMS_htt_dyShape
118 CMS_ff_total_low_pt_et_2017
119 CMS_scale_j_RelativeSample_2017
120 CMS_scale_j_BBEC1
121 CMS_htt_mistag_b_2017
122 CMS_ff_total_ttbar_syst_l_pt_closure_et_2017
123 CMS_ff_total_wjets_frac_et_2017
124 CMS_prefiring
125 CMS_htt_emb_ttbar_2017
126 CMS_ff_total_ttbar_syst_met_closure_et_2017
127 CMS_scale_e
128 QCDscale_ggH_REWEIGHT
129 CMS_ff_total_ttbar_stat_met_unc1_et_2017
130 CMS_scale_j_Absolute_2017
131 CMS_eff_trigger_single_t_emb_2017
132 CMS_scale_t_emb_3prong1pizero_2017
133 lumi_13TeV_DB
134 CMS_ff_total_ttbar_stat_met_unc2_et_2017
135 CMS_scale_j_HF
136 prop_binhtt_et_32_2017_bin1
137 CMS_eff_t_35-40_2017
138 CMS_ff_total_qcd_stat_l_pt_unc1_et_2017
139 CMS_eff_xtrigger_t_et_dm1_2017
140 CMS_htt_zjXsec
141 CMS_scale_j_EC2
142 lumi_13TeV_BBD
143 CMS_ff_total_syst_alt_func_et_2017
144 lumi_13TeV_LS
145 lumi_13TeV_BCC
146 CMS_ZLShape_et_1prong1pizero_2017
147 CMS_ff_total_ttbar_stat_jet_pt_high_unc3_et_2017
148 CMS_ff_total_qcd_stat_ss_njets0_unc2_et_2017
149 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc1_et_2017
150 CMS_eff_xtrigger_t_emb_et_dm1_2017
151 CMS_eff_xtrigger_t_et_dm10_2017
152 CMS_eff_xtrigger_l_emb_et_2017
×10⁻¹
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0)/\Delta\theta$
−0.5 0 0.5
$\Delta \hat{r}_{ggH}$

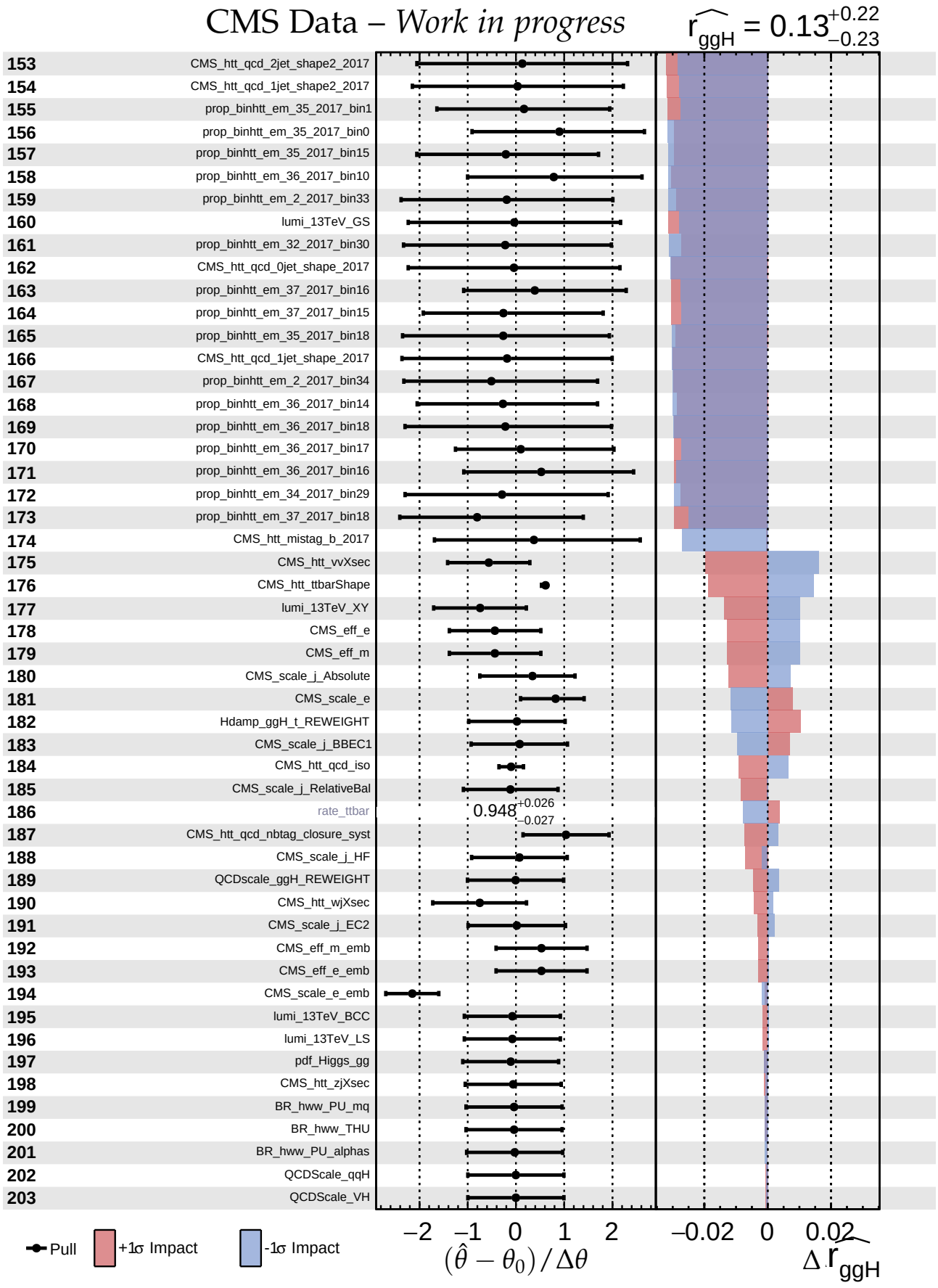

CMS Data – Work in progress
$\hat{r}_{ggH} = 0.06^{+0.07}_{-0.06}$
153 CMS_eff_trigger_single_t_2017
154 CMS_ff_total_ttbar_frac_et_2017
155 CMS_ff_total_qcd_stat_njet1_jet_pt_high_unc2_et_2017
156 CMS_eff_xtrigger_t_emb_et_dm10_2017
157 prop_binhtt_et_32_2017_bin17
158 CMS_ff_total_ttbar_stat_jet_pt_high_unc2_et_2017
159 CMS_eff_t_emb_35-40_2017
160 lumi_13TeV_GS
161 CMS_eff_xtrigger_t_emb_et_dm0_2017
162 CMS_scale_j_HF_2017
163 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc3_et_2017
164 CMS_fake_e_EC_2017
165 CMS_ff_total_qcd_stat_njet0_jet_pt_low_unc1_et_2017
166 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc2_et_2017
167 CMS_eff_xtrigger_t_et_dm11_2017
168 CMS_eff_xtrigger_t_emb_et_dm11_2017
169 CMS_ff_total_wjets_stat_njet1_jet_pt_high_unc3_et_2017
170 CMS_ff_total_qcd_stat_njet0_jet_pt_low_unc3_et_2017
171 Hdamp_ggH_i_REWEIGHT
172 Hdamp_ggH_b_REWEIGHT
173 QCDscaleAndHdamp_bbH_ACCEPT
174 pdf_bbH_ACCEPT
175 CMS_eff_xtrigger_l_et_2017
176 CMS_eff_xtrigger_t_et_dm0_2017
177 CMS_ff_total_wjets_stat_njet0_jet_pt_high_unc3_et_2017
178 CMS_ZLShape_et_1prong_2017
179 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc3_et_2017
180 CMS_eff_t_emb_highpT_500-inf_2017
181 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc3_et_2017
182 CMS_eff_t_highpT_500-inf_2017
183 CMS_ff_total_qcd_stat_njet1_jet_pt_high_unc3_et_2017
184 CMS_ff_total_qcd_stat_njet0_jet_pt_high_unc3_et_2017
185 CMS_ff_total_wjets_stat_extrap_unc2_et_NbtagGt1_MTLt40_2017
186 CMS_ff_total_qcd_stat_os_unc2_et_NbtagGt1_2017
187 CMS_ff_total_wjets_stat_njet0_jet_pt_low_unc4_et_2017
188 CMS_ff_total_qcd_stat_os_unc2_et_Nbtag0_2017
189 CMS_ff_total_wjets_stat_njet0_jet_pt_med_unc4_et_2017
190 CMS_ff_total_wjets_stat_njet0_jet_pt_high_unc4_et_2017
191 CMS_ff_total_wjets_stat_met_njets1_unc1_et_2017
192 CMS_ff_total_wjets_stat_njet1_jet_pt_high_unc4_et_2017
193 CMS_ff_total_wjets_stat_njet1_jet_pt_low_unc4_et_2017
194 CMS_ff_total_wjets_stat_met_njets0_unc1_et_2017
195 CMS_ff_total_wjets_stat_njet1_jet_pt_med_unc4_et_2017
196 CMS_scale_j_BBEC1_2017
×10
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.1 0 0.1
$\Delta \hat{r}_{ggH}$

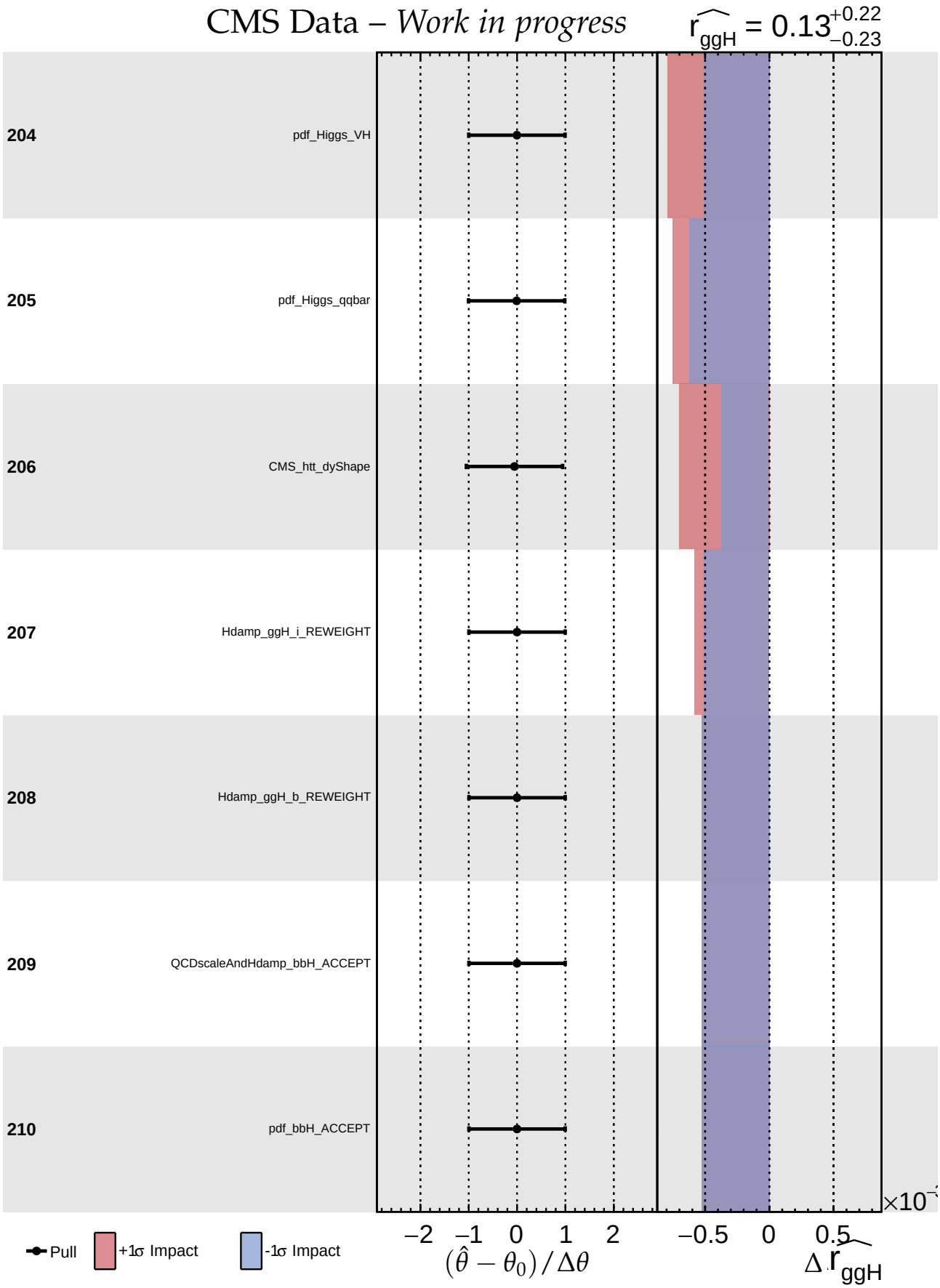


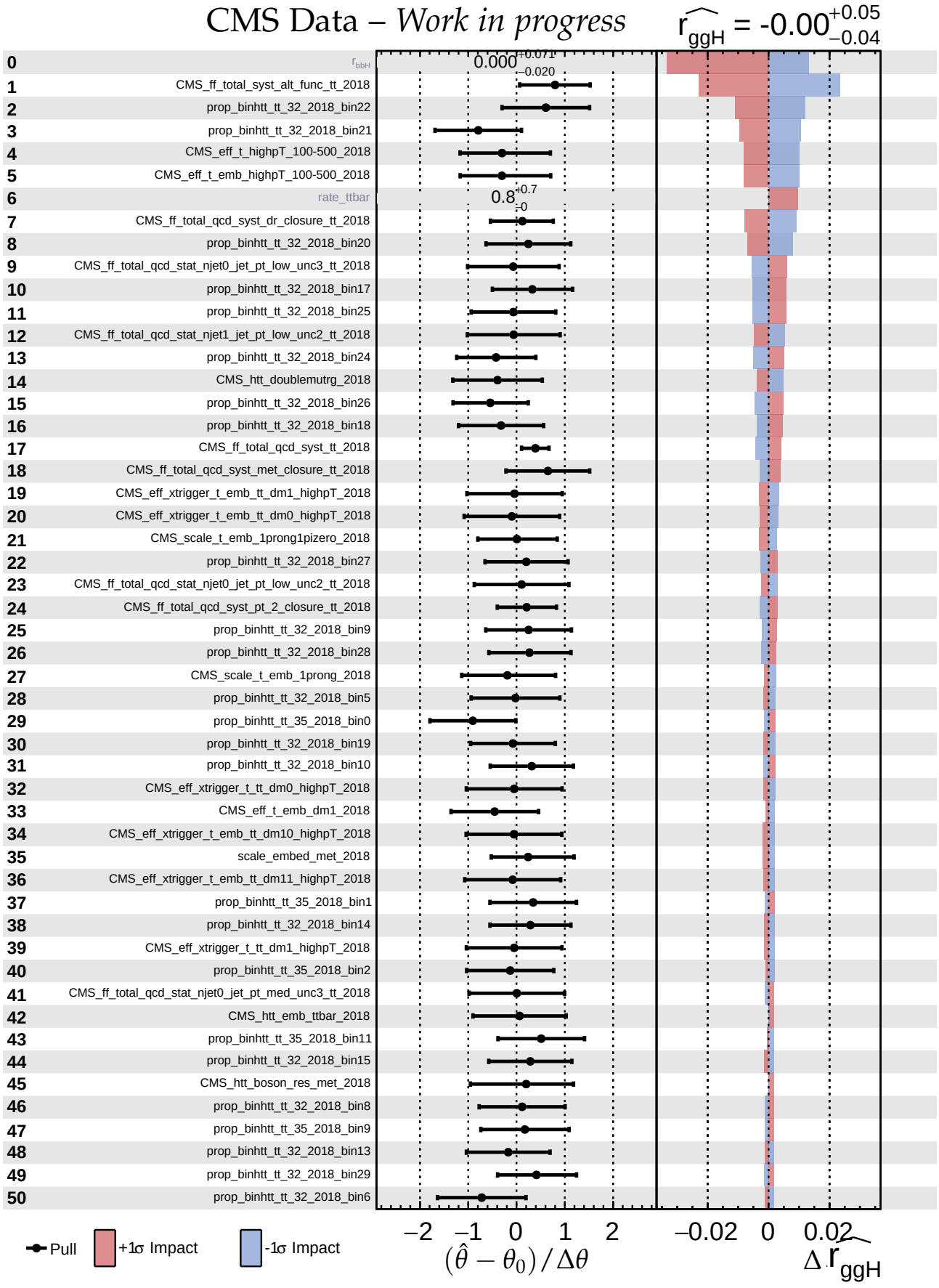
Année 2017, canal $e\mu$, page 1/5


CMS Data – Work in progress
$\hat{r}_{ggH} = 0.13^{+0.22}_{-0.23}$
51 prop_binhtt_em_2_2017_bin23
52 prop_binhtt_em_2_2017_bin29
53 prop_binhtt_em_32_2017_bin3
54 prop_binhtt_em_34_2017_bin28
55 prop_binhtt_em_2_2017_bin4
56 prop_binhtt_em_32_2017_bin0
57 prop_binhtt_em_32_2017_bin2
58 prop_binhtt_em_32_2017_bin29
59 prop_binhtt_em_33_2017_bin23
60 prop_binhtt_em_34_2017_bin23
61 prop_binhtt_em_2_2017_bin19
62 prop_binhtt_em_32_2017_bin4
63 CMS_scale_j_RelativeSample_2017
64 prop_binhtt_em_32_2017_bin6
65 prop_binhtt_em_32_2017_bin1
66 prop_binhtt_em_34_2017_bin5
67 prop_binhtt_em_2_2017_bin24
68 CMS_scale_j_Absolute_2017
69 prop_binhtt_em_2_2017_bin22
70 prop_binhtt_em_2_2017_bin21
71 prop_binhtt_em_34_2017_bin24
72 prop_binhtt_em_32_2017_bin20
73 prop_binhtt_em_34_2017_bin25
74 prop_binhtt_em_33_2017_bin10
75 prop_binhtt_em_2_2017_bin2
76 subtrMC
77 prop_binhtt_em_37_2017_bin2
78 prop_binhtt_em_2_2017_bin11
79 prop_binhtt_em_2_2017_bin25
80 prop_binhtt_em_33_2017_bin24
81 prop_binhtt_em_2_2017_bin26
82 CMS_htt_boson_scale_met_2017
83 prop_binhtt_em_2_2017_bin30
84 prop_binhtt_em_2_2017_bin27
85 prop_binhtt_em_2_2017_bin10
86 prop_binhtt_em_34_2017_bin26
87 prop_binhtt_em_2_2017_bin9
88 prop_binhtt_em_2_2017_bin20
89 lumi_13TeV_2017
90 CMS_eff_trigger_em_2017
91 CMS_prefiring
92 prop_binhtt_em_34_2017_bin3
93 prop_binhtt_em_37_2017_bin3
94 prop_binhtt_em_36_2017_bin0
95 prop_binhtt_em_2_2017_bin28
96 prop_binhtt_em_36_2017_bin15
97 prop_binhtt_em_32_2017_bin17
98 prop_binhtt_em_33_2017_bin22
99 CMS_scale_j_BBEC1_2017
100 prop_binhtt_em_33_2017_bin0
101 CMS_htt_qcd_nbtag_closure_stat_2017
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.05 0 0.05
$\Delta \hat{r}_{ggH}$

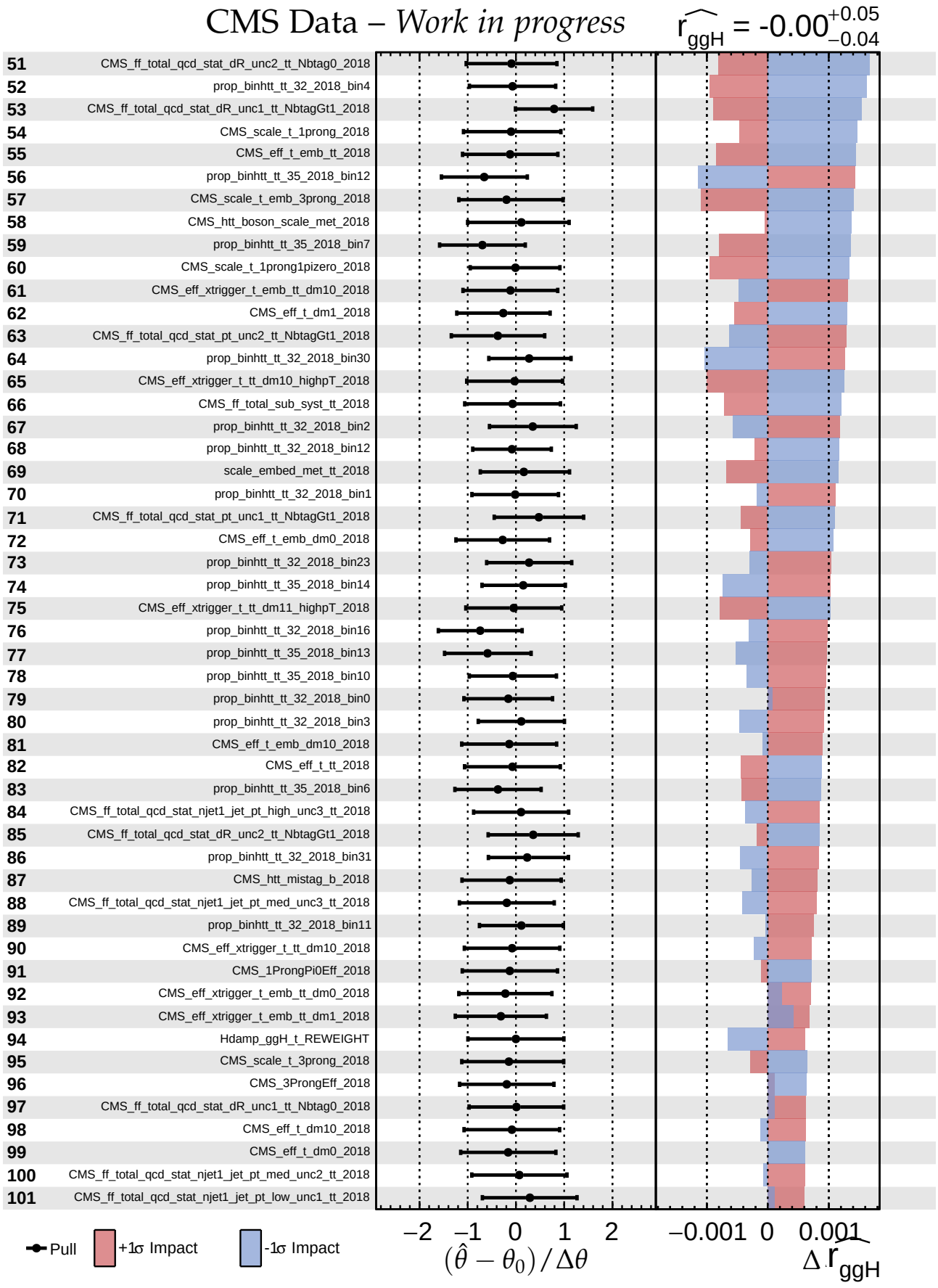

CMS Data – Work in progress
$\hat{r}_{ggH} = 0.13^{+0.22}_{-0.23}$
102 prop_binhtt_em_36_2017_bin1
103 prop_binhtt_em_37_2017_bin8
104 prop_binhtt_em_35_2017_bin9
105 prop_binhtt_em_33_2017_bin11
106 prop_binhtt_em_35_2017_bin7
107 prop_binhtt_em_35_2017_bin8
108 prop_binhtt_em_34_2017_bin4
109 prop_binhtt_em_35_2017_bin5
110 prop_binhtt_em_32_2017_bin18
111 prop_binhtt_em_35_2017_bin6
112 prop_binhtt_em_33_2017_bin13
113 prop_binhtt_em_36_2017_bin13
114 CMS_htt_emb_ttbar_2017
115 prop_binhtt_em_37_2017_bin4
116 prop_binhtt_em_35_2017_bin10
117 prop_binhtt_em_37_2017_bin7
118 CMS_scale_j_FlavorQCD
119 prop_binhtt_em_35_2017_bin4
120 CMS_htt_qcd_0jet_rate_2017
121 prop_binhtt_em_32_2017_bin27
122 prop_binhtt_em_34_2017_bin27
123 prop_binhtt_em_37_2017_bin6
124 prop_binhtt_em_32_2017_bin28
125 prop_binhtt_em_33_2017_bin2
126 lumi_13TeV_DB
127 prop_binhtt_em_37_2017_bin5
128 prop_binhtt_em_35_2017_bin3
129 prop_binhtt_em_35_2017_bin2
130 prop_binhtt_em_35_2017_bin11
131 CMS_htt_doublemutrg_2017
132 prop_binhtt_em_36_2017_bin12
133 CMS_htt_qcd_0jet_shape2_2017
134 prop_binhtt_em_2_2017_bin32
135 prop_binhtt_em_37_2017_bin9
136 prop_binhtt_em_35_2017_bin17
137 prop_binhtt_em_2_2017_bin31
138 lumi_13TeV_BBD
139 prop_binhtt_em_37_2017_bin17
140 prop_binhtt_em_37_2017_bin14
141 prop_binhtt_em_35_2017_bin12
142 CMS_htt_qcd_2jet_rate_2017
143 CMS_htt_qcd_2jet_shape_2017
144 prop_binhtt_em_33_2017_bin1
145 CMS_htt_qcd_1jet_rate_2017
146 prop_binhtt_em_34_2017_bin30
147 CMS_eff_trigger_emb_em_2017
148 CMS_scale_j_HF_2017
149 prop_binhtt_em_36_2017_bin11
150 prop_binhtt_em_35_2017_bin13
151 prop_binhtt_em_35_2017_bin16
152 prop_binhtt_em_35_2017_bin14
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.04 −0.02 0 0.02 0.04
$\Delta \hat{r}_{ggH}$
Pull
+1σ Impact
-1σ Impact

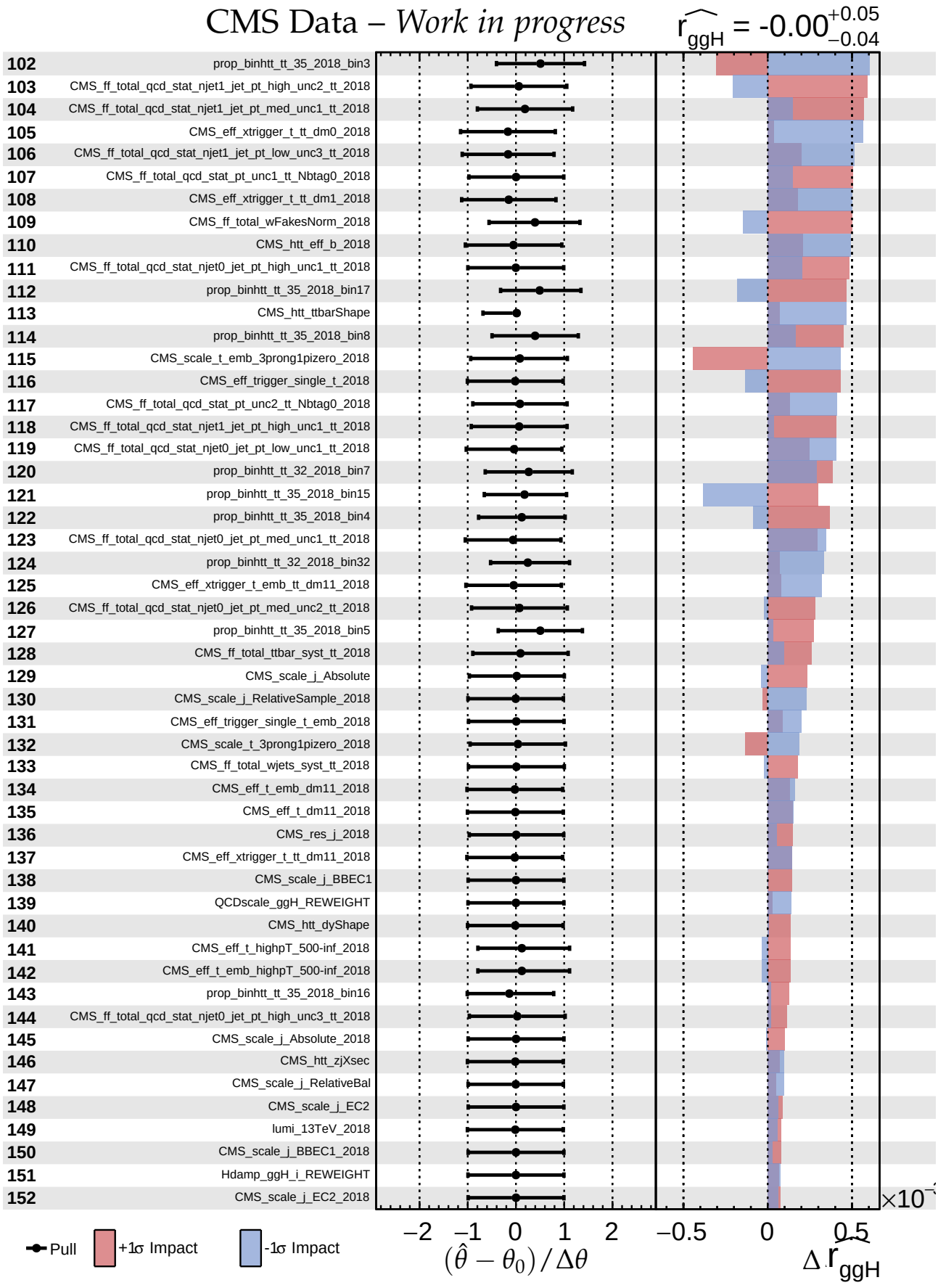

CMS Data – Work in progress
$\hat{r}_{ggH} = 0.13^{+0.22}_{-0.23}$
153 CMS_htt_qcd_2jet_shape2_2017
154 CMS_htt_qcd_1jet_shape2_2017
155 prop_binhtt_em_35_2017_bin1
156 prop_binhtt_em_35_2017_bin0
157 prop_binhtt_em_35_2017_bin15
158 prop_binhtt_em_36_2017_bin10
159 prop_binhtt_em_2_2017_bin33
160 lumi_13TeV_GS
161 prop_binhtt_em_32_2017_bin30
162 CMS_htt_qcd_0jet_shape_2017
163 prop_binhtt_em_37_2017_bin16
164 prop_binhtt_em_37_2017_bin15
165 prop_binhtt_em_35_2017_bin18
166 CMS_htt_qcd_1jet_shape_2017
167 prop_binhtt_em_2_2017_bin34
168 prop_binhtt_em_36_2017_bin14
169 prop_binhtt_em_36_2017_bin18
170 prop_binhtt_em_36_2017_bin17
171 prop_binhtt_em_36_2017_bin16
172 prop_binhtt_em_34_2017_bin29
173 prop_binhtt_em_37_2017_bin18
174 CMS_htt_mistag_b_2017
175 CMS_htt_vvXsec
176 CMS_htt_ttbarShape
177 lumi_13TeV_XY
178 CMS_eff_e
179 CMS_eff_m
180 CMS_scale_j_Absolute
181 CMS_scale_e
182 Hdamp_ggH_t_REWEIGHT
183 CMS_scale_j_BBEC1
184 CMS_htt_qcd_iso
185 CMS_scale_j_RelativeBal
186 rate_ttbar
$0.948^{+0.026}_{-0.027}$
187 CMS_htt_qcd_nbtag_closure_syst
188 CMS_scale_j_HF
189 QCDscale_ggH_REWEIGHT
190 CMS_htt_wjXsec
191 CMS_scale_j_EC2
192 CMS_eff_m_emb
193 CMS_eff_e_emb
194 CMS_scale_e_emb
195 lumi_13TeV_BCC
196 lumi_13TeV_LS
197 pdf_Higgs_gg
198 CMS_htt_zjXsec
199 BR_hww_PU_mq
200 BR_hww_THU
201 BR_hww_PU_alphas
202 QCDScale_qqH
203 QCDScale_VH
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0)/\Delta\theta$
−0.02 0 0.02
$\Delta\hat{r}_{ggH}$


CMS Data – Work in progress
$\hat{r}_{ggH} = 0.13^{+0.22}_{-0.23}$
204
pdf_Higgs_VH
205
pdf_Higgs_qqbar
206
CMS_htt_dyShape
207
Hdamp_ggH_i_REWEIGHT
208
Hdamp_ggH_b_REWEIGHT
209
QCDscaleAndHdamp_bbH_ACCEPT
210
pdf_bbH_ACCEPT
×10
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0)/\Delta\theta$
−0.5 0 0.5
$\Delta \hat{r}_{ggH}$



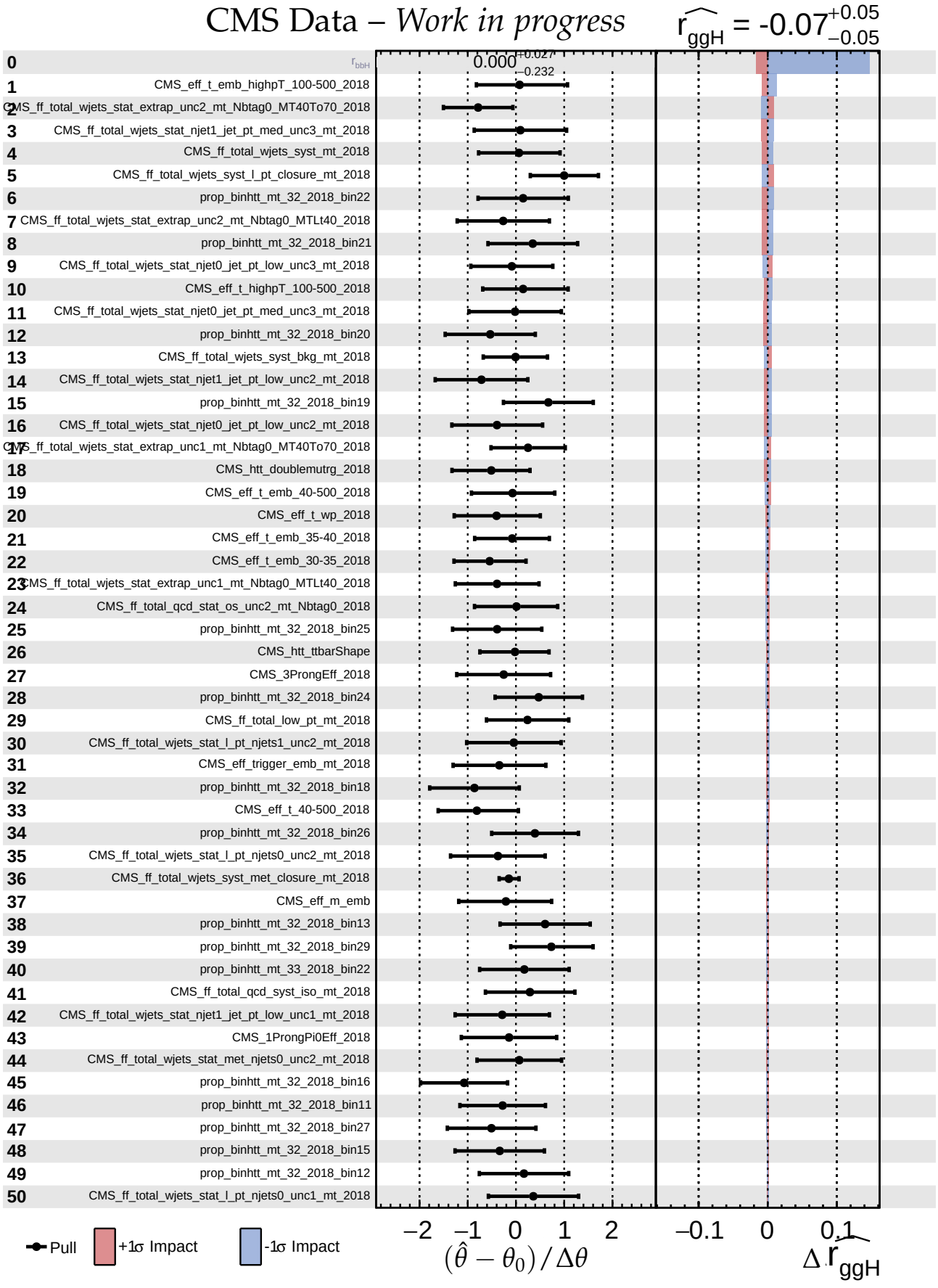
Année 2018, canal $\tau_h\tau_h$, page 1/4


CMS Data – Work in progress
$\hat{r}_{ggH} = -0.00^{+0.05}_{-0.04}$
51 CMS_ff_total_qcd_stat_dR_unc2_tt_Nbtag0_2018
52 prop_binhtt_tt_32_2018_bin4
53 CMS_ff_total_qcd_stat_dR_unc1_tt_NbtagGt1_2018
54 CMS_scale_t_1prong_2018
55 CMS_eff_t_emb_tt_2018
56 prop_binhtt_tt_35_2018_bin12
57 CMS_scale_t_emb_3prong_2018
58 CMS_htt_boson_scale_met_2018
59 prop_binhtt_tt_35_2018_bin7
60 CMS_scale_t_1prong1pizero_2018
61 CMS_eff_xtrigger_t_emb_tt_dm10_2018
62 CMS_eff_t_dm1_2018
63 CMS_ff_total_qcd_stat_pt_unc2_tt_NbtagGt1_2018
64 prop_binhtt_tt_32_2018_bin30
65 CMS_eff_xtrigger_t_tt_dm10_highpT_2018
66 CMS_ff_total_sub_syst_tt_2018
67 prop_binhtt_tt_32_2018_bin2
68 prop_binhtt_tt_32_2018_bin12
69 scale_embed_met_tt_2018
70 prop_binhtt_tt_32_2018_bin1
71 CMS_ff_total_qcd_stat_pt_unc1_tt_NbtagGt1_2018
72 CMS_eff_t_emb_dm0_2018
73 prop_binhtt_tt_32_2018_bin23
74 prop_binhtt_tt_35_2018_bin14
75 CMS_eff_xtrigger_t_tt_dm11_highpT_2018
76 prop_binhtt_tt_32_2018_bin16
77 prop_binhtt_tt_35_2018_bin13
78 prop_binhtt_tt_35_2018_bin10
79 prop_binhtt_tt_32_2018_bin0
80 prop_binhtt_tt_32_2018_bin3
81 CMS_eff_t_emb_dm10_2018
82 CMS_eff_t_tt_2018
83 prop_binhtt_tt_35_2018_bin6
84 CMS_ff_total_qcd_stat_njet1_jet_pt_high_unc3_tt_2018
85 CMS_ff_total_qcd_stat_dR_unc2_tt_NbtagGt1_2018
86 prop_binhtt_tt_32_2018_bin31
87 CMS_htt_mistag_b_2018
88 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc3_tt_2018
89 prop_binhtt_tt_32_2018_bin11
90 CMS_eff_xtrigger_t_tt_dm10_2018
91 CMS_1ProngPi0Eff_2018
92 CMS_eff_xtrigger_t_emb_tt_dm0_2018
93 CMS_eff_xtrigger_t_emb_tt_dm1_2018
94 Hdamp_ggH_t_REWEIGHT
95 CMS_scale_t_3prong_2018
96 CMS_3ProngEff_2018
97 CMS_ff_total_qcd_stat_dR_unc1_tt_Nbtag0_2018
98 CMS_eff_t_dm10_2018
99 CMS_eff_t_dm0_2018
100 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc2_tt_2018
101 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc1_tt_2018
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.001 0 0.001
$\Delta \hat{r}_{ggH}$

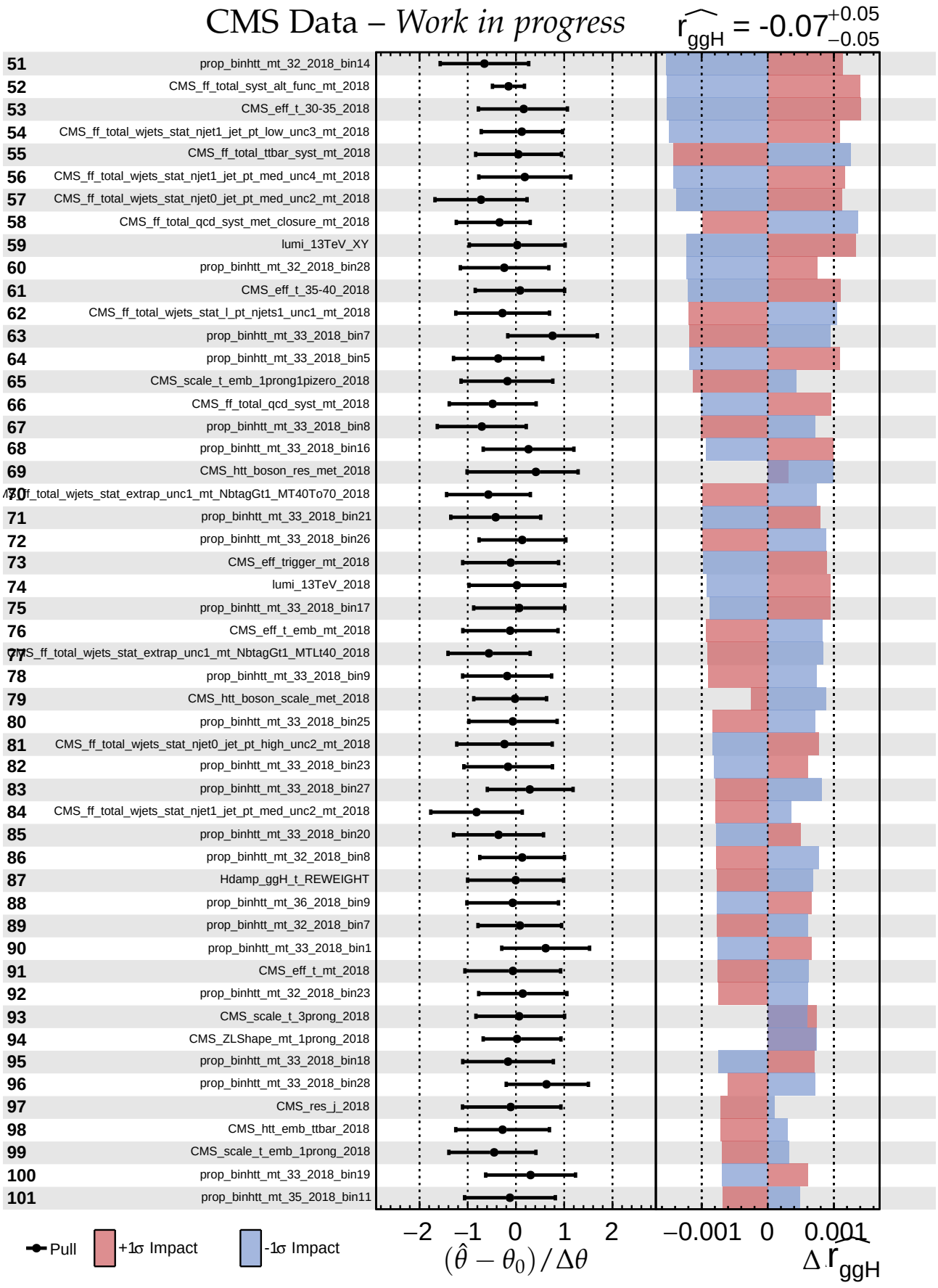

CMS Data – Work in progress
$\widehat{r_{ggH}} = -0.00^{+0.05}_{-0.04}$
102 prop_binhtt_tt_35_2018_bin3
103 CMS_ff_total_qcd_stat_njet1_jet_pt_high_unc2_tt_2018
104 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc1_tt_2018
105 CMS_eff_xtrigger_t_tt_dm0_2018
106 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc3_tt_2018
107 CMS_ff_total_qcd_stat_pt_unc1_tt_Nbtag0_2018
108 CMS_eff_xtrigger_t_tt_dm1_2018
109 CMS_ff_total_wFakesNorm_2018
110 CMS_htt_eff_b_2018
111 CMS_ff_total_qcd_stat_njet0_jet_pt_high_unc1_tt_2018
112 prop_binhtt_tt_35_2018_bin17
113 CMS_htt_ttbarShape
114 prop_binhtt_tt_35_2018_bin8
115 CMS_scale_t_emb_3prong1pizero_2018
116 CMS_eff_trigger_single_t_2018
117 CMS_ff_total_qcd_stat_pt_unc2_tt_Nbtag0_2018
118 CMS_ff_total_qcd_stat_njet1_jet_pt_high_unc1_tt_2018
119 CMS_ff_total_qcd_stat_njet0_jet_pt_low_unc1_tt_2018
120 prop_binhtt_tt_32_2018_bin7
121 prop_binhtt_tt_35_2018_bin15
122 prop_binhtt_tt_35_2018_bin4
123 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc1_tt_2018
124 prop_binhtt_tt_32_2018_bin32
125 CMS_eff_xtrigger_t_emb_tt_dm11_2018
126 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc2_tt_2018
127 prop_binhtt_tt_35_2018_bin5
128 CMS_ff_total_ttbar_syst_tt_2018
129 CMS_scale_j_Absolute
130 CMS_scale_j_RelativeSample_2018
131 CMS_eff_trigger_single_t_emb_2018
132 CMS_scale_t_3prong1pizero_2018
133 CMS_ff_total_wjets_syst_tt_2018
134 CMS_eff_t_emb_dm11_2018
135 CMS_eff_t_dm11_2018
136 CMS_res_j_2018
137 CMS_eff_xtrigger_t_tt_dm11_2018
138 CMS_scale_j_BBEC1
139 QCDscale_ggH_REWEIGHT
140 CMS_htt_dyShape
141 CMS_eff_t_highpT_500-inf_2018
142 CMS_eff_t_emb_highpT_500-inf_2018
143 prop_binhtt_tt_35_2018_bin16
144 CMS_ff_total_qcd_stat_njet0_jet_pt_high_unc3_tt_2018
145 CMS_scale_j_Absolute_2018
146 CMS_htt_zjXsec
147 CMS_scale_j_RelativeBal
148 CMS_scale_j_EC2
149 lumi_13TeV_2018
150 CMS_scale_j_BBEC1_2018
151 Hdamp_ggH_i_REWEIGHT
152 CMS_scale_j_EC2_2018
×10⁻³
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.5 0 0.5
$\Delta\widehat{r_{ggH}}$



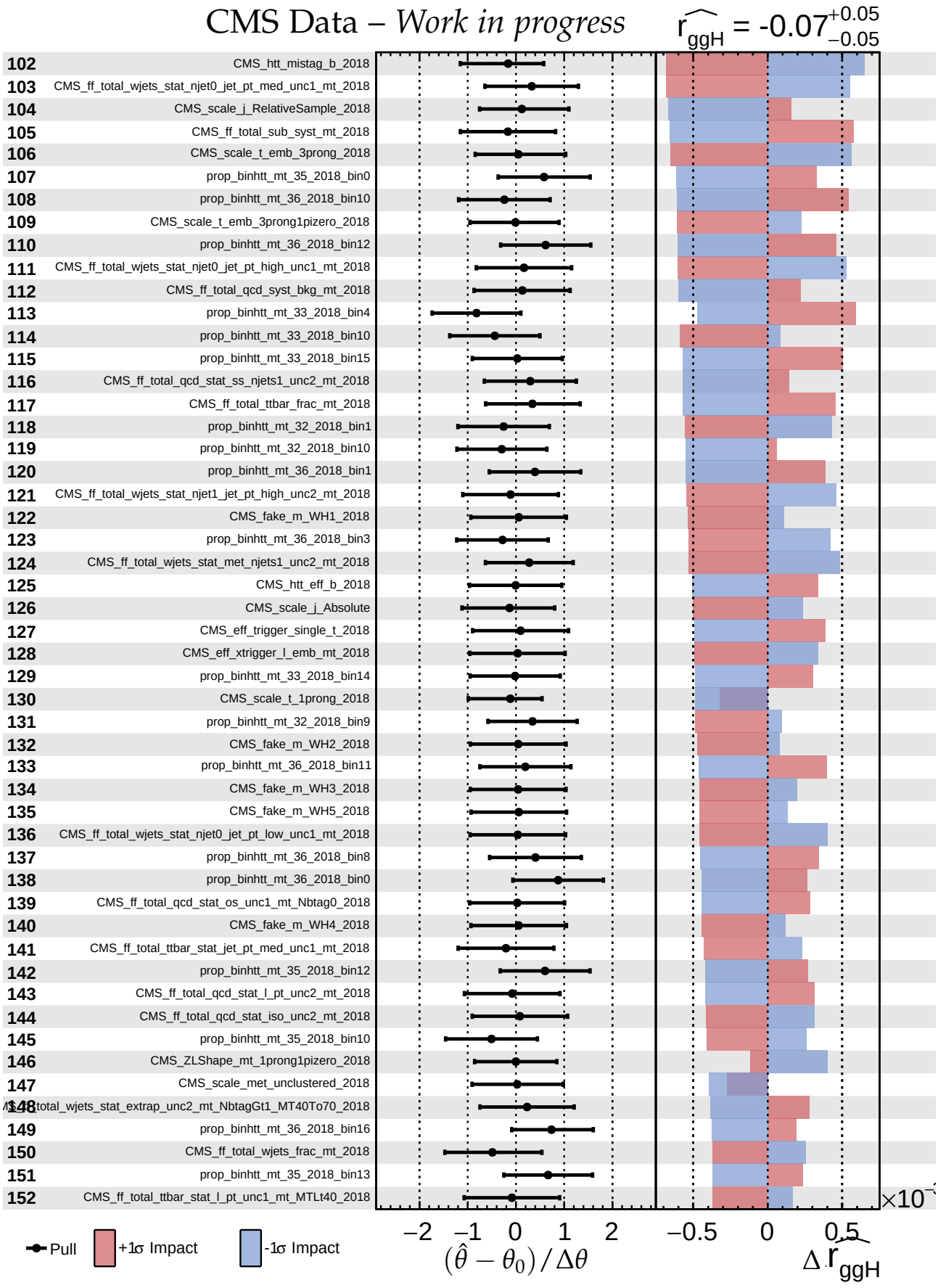
Année 2018, canal $\tau_h\tau_h$, page 4/4

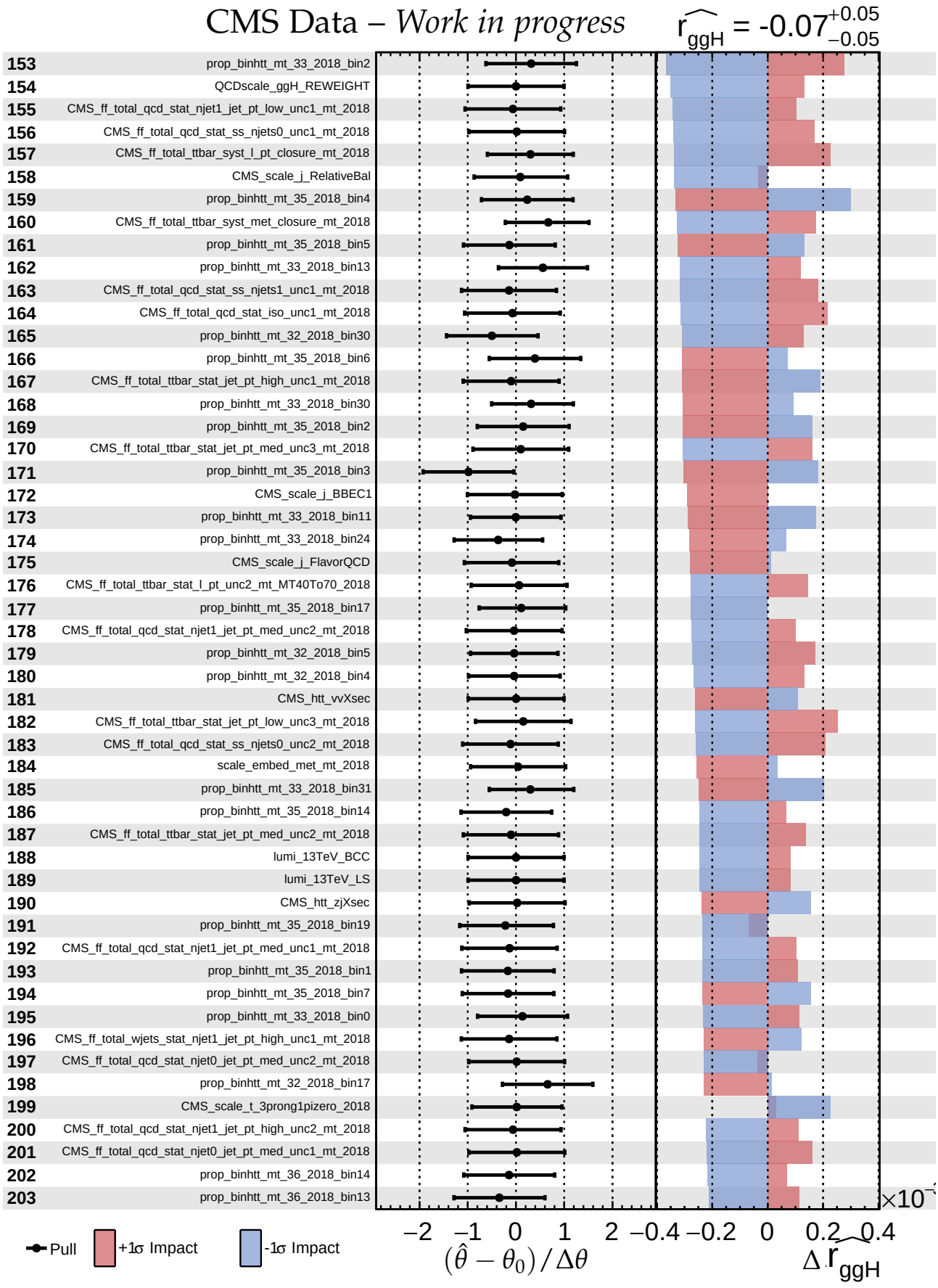


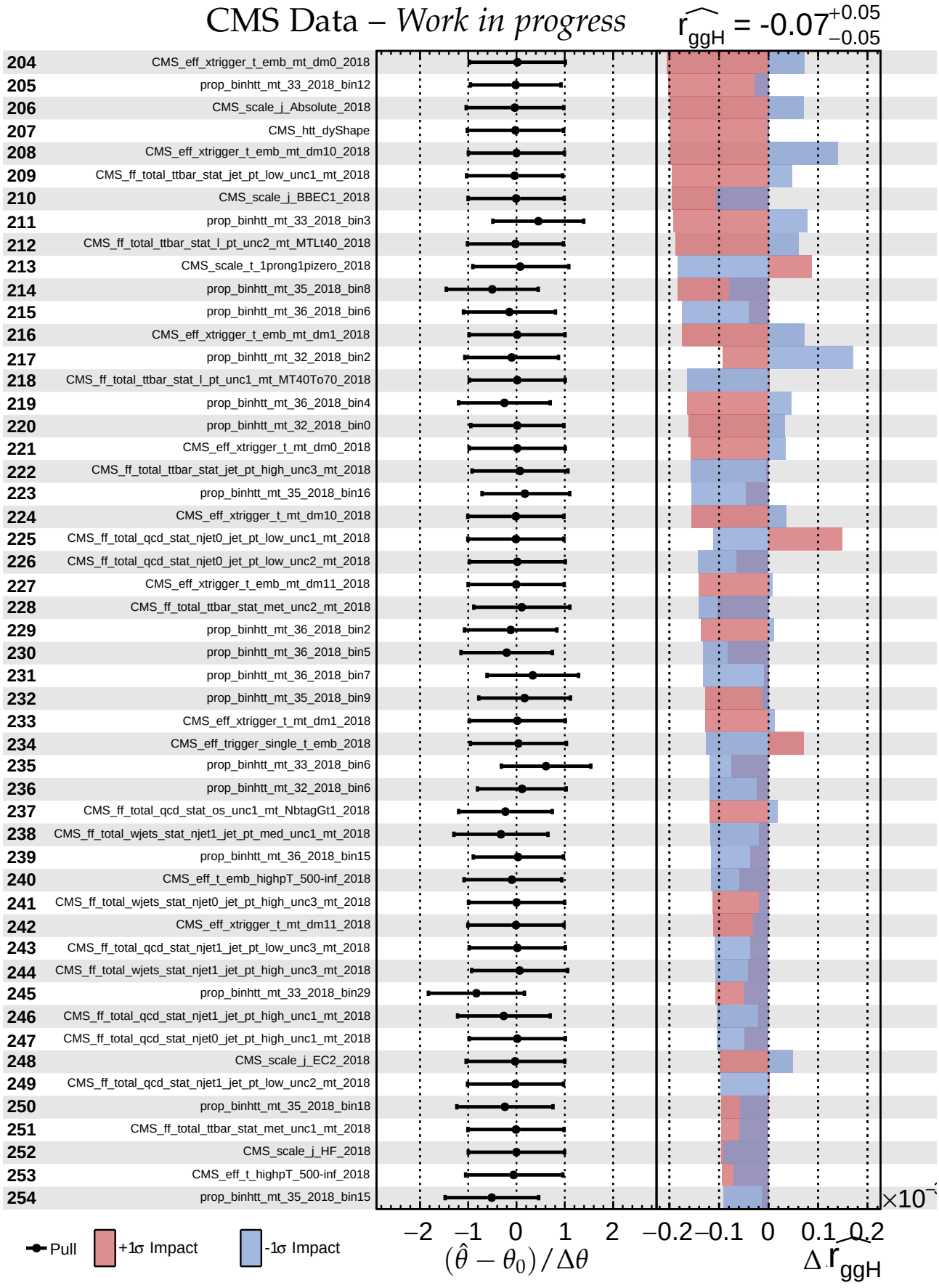
Année 2018, canal $\mu\tau_h$, page 1/6

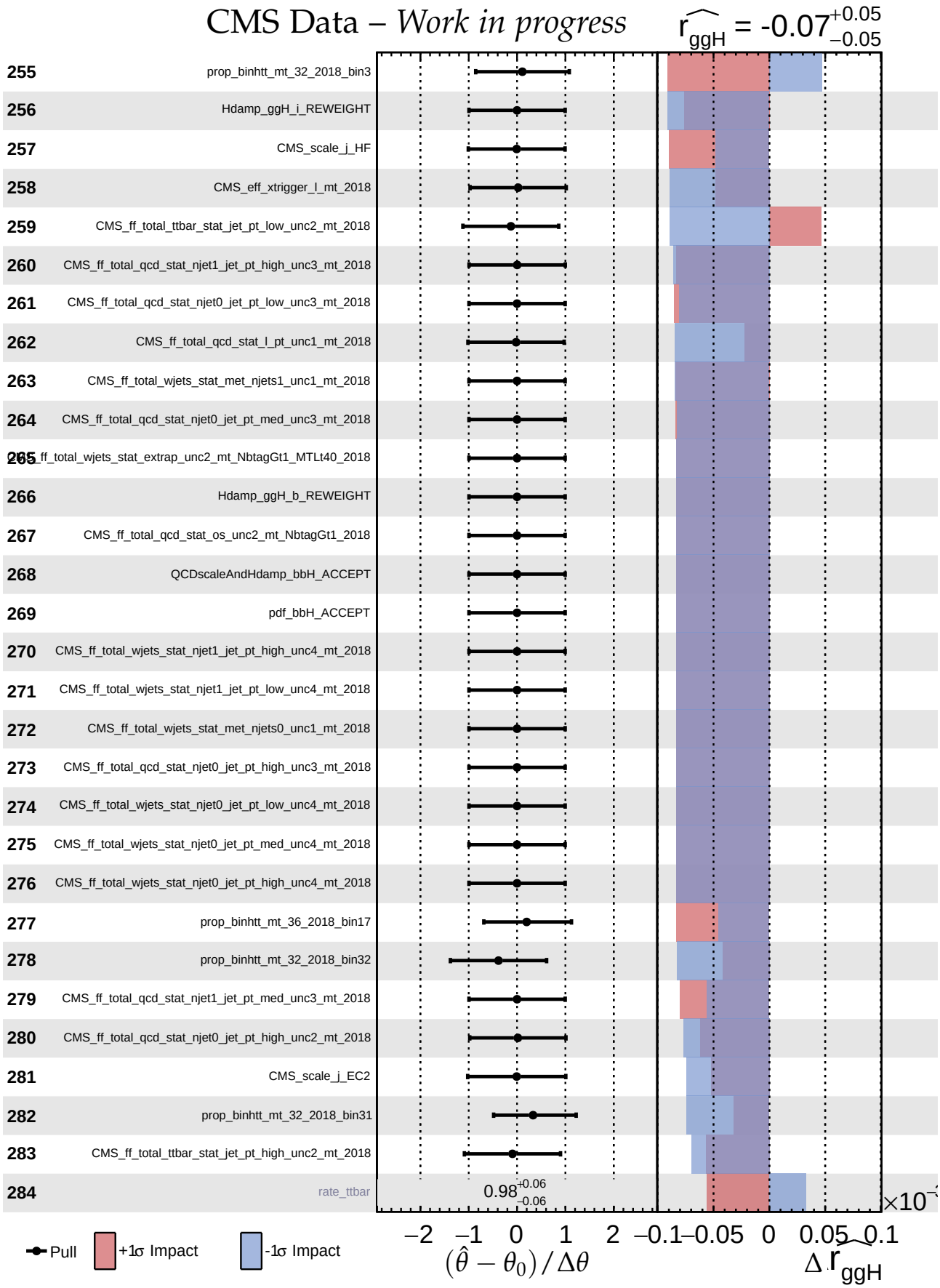


Année 2018, canal $\mu\tau_h$, page 2/6

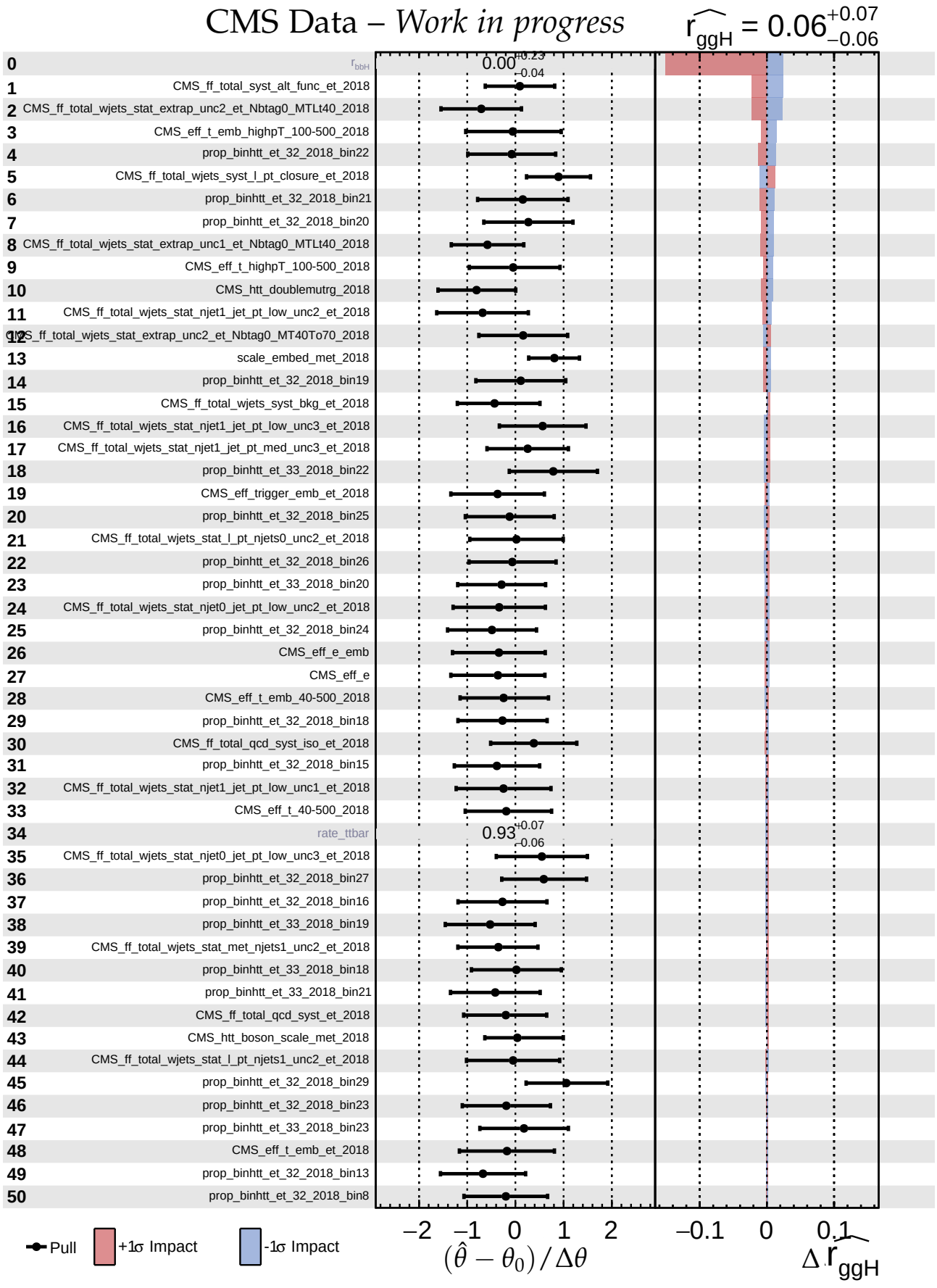

CMS Data – Work in progress
$\widehat{r_{ggH}} = -0.07^{+0.05}_{-0.05}$
102 CMS_htt_mistag_b_2018
103 CMS_ff_total_wjets_stat_njet0_jet_pt_med_unc1_mt_2018
104 CMS_scale_j_RelativeSample_2018
105 CMS_ff_total_sub_syst_mt_2018
106 CMS_scale_t_emb_3prong_2018
107 prop_binhtt_mt_35_2018_bin0
108 prop_binhtt_mt_36_2018_bin10
109 CMS_scale_t_emb_3prong1pizero_2018
110 prop_binhtt_mt_36_2018_bin12
111 CMS_ff_total_wjets_stat_njet0_jet_pt_high_unc1_mt_2018
112 CMS_ff_total_qcd_syst_bkg_mt_2018
113 prop_binhtt_mt_33_2018_bin4
114 prop_binhtt_mt_33_2018_bin10
115 prop_binhtt_mt_33_2018_bin15
116 CMS_ff_total_qcd_stat_ss_njets1_unc2_mt_2018
117 CMS_ff_total_ttbar_frac_mt_2018
118 prop_binhtt_mt_32_2018_bin1
119 prop_binhtt_mt_32_2018_bin10
120 prop_binhtt_mt_36_2018_bin1
121 CMS_ff_total_wjets_stat_njet1_jet_pt_high_unc2_mt_2018
122 CMS_fake_m_WH1_2018
123 prop_binhtt_mt_36_2018_bin3
124 CMS_ff_total_wjets_stat_met_njets1_unc2_mt_2018
125 CMS_htt_eff_b_2018
126 CMS_scale_j_Absolute
127 CMS_eff_trigger_single_t_2018
128 CMS_eff_xtrigger_l_emb_mt_2018
129 prop_binhtt_mt_33_2018_bin14
130 CMS_scale_t_1prong_2018
131 prop_binhtt_mt_32_2018_bin9
132 CMS_fake_m_WH2_2018
133 prop_binhtt_mt_36_2018_bin11
134 CMS_fake_m_WH3_2018
135 CMS_fake_m_WH5_2018
136 CMS_ff_total_wjets_stat_njet0_jet_pt_low_unc1_mt_2018
137 prop_binhtt_mt_36_2018_bin8
138 prop_binhtt_mt_36_2018_bin0
139 CMS_ff_total_qcd_stat_os_unc1_mt_Nbtag0_2018
140 CMS_fake_m_WH4_2018
141 CMS_ff_total_ttbar_stat_jet_pt_med_unc1_mt_2018
142 prop_binhtt_mt_35_2018_bin12
143 CMS_ff_total_qcd_stat_l_pt_unc2_mt_2018
144 CMS_ff_total_qcd_stat_iso_unc2_mt_2018
145 prop_binhtt_mt_35_2018_bin10
146 CMS_ZLShape_mt_1prong1pizero_2018
147 CMS_scale_met_unclustered_2018
148 CMS_ff_total_wjets_stat_extrap_unc2_mt_NbtagGt1_MT40To70_2018
149 prop_binhtt_mt_36_2018_bin16
150 CMS_ff_total_wjets_frac_mt_2018
151 prop_binhtt_mt_35_2018_bin13
152 CMS_ff_total_ttbar_stat_l_pt_unc1_mt_MTLt40_2018
×10
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.5 0 0.5
$\Delta \widehat{r_{ggH}}$


CMS Data – Work in progress
$\hat{r}_{ggH} = -0.07^{+0.05}_{-0.05}$
153 prop_binhtt_mt_33_2018_bin2
154 QCDscale_ggH_REWEIGHT
155 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc1_mt_2018
156 CMS_ff_total_qcd_stat_ss_njets0_unc1_mt_2018
157 CMS_ff_total_ttbar_syst_l_pt_closure_mt_2018
158 CMS_scale_j_RelativeBal
159 prop_binhtt_mt_35_2018_bin4
160 CMS_ff_total_ttbar_syst_met_closure_mt_2018
161 prop_binhtt_mt_35_2018_bin5
162 prop_binhtt_mt_33_2018_bin13
163 CMS_ff_total_qcd_stat_ss_njets1_unc1_mt_2018
164 CMS_ff_total_qcd_stat_iso_unc1_mt_2018
165 prop_binhtt_mt_32_2018_bin30
166 prop_binhtt_mt_35_2018_bin6
167 CMS_ff_total_ttbar_stat_jet_pt_high_unc1_mt_2018
168 prop_binhtt_mt_33_2018_bin30
169 prop_binhtt_mt_35_2018_bin2
170 CMS_ff_total_ttbar_stat_jet_pt_med_unc3_mt_2018
171 prop_binhtt_mt_35_2018_bin3
172 CMS_scale_j_BBEC1
173 prop_binhtt_mt_33_2018_bin11
174 prop_binhtt_mt_33_2018_bin24
175 CMS_scale_j_FlavorQCD
176 CMS_ff_total_ttbar_stat_l_pt_unc2_mt_MT40To70_2018
177 prop_binhtt_mt_35_2018_bin17
178 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc2_mt_2018
179 prop_binhtt_mt_32_2018_bin5
180 prop_binhtt_mt_32_2018_bin4
181 CMS_htt_vvXsec
182 CMS_ff_total_ttbar_stat_jet_pt_low_unc3_mt_2018
183 CMS_ff_total_qcd_stat_ss_njets0_unc2_mt_2018
184 scale_embed_met_mt_2018
185 prop_binhtt_mt_33_2018_bin31
186 prop_binhtt_mt_35_2018_bin14
187 CMS_ff_total_ttbar_stat_jet_pt_med_unc2_mt_2018
188 lumi_13TeV_BCC
189 lumi_13TeV_LS
190 CMS_htt_zjXsec
191 prop_binhtt_mt_35_2018_bin19
192 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc1_mt_2018
193 prop_binhtt_mt_35_2018_bin1
194 prop_binhtt_mt_35_2018_bin7
195 prop_binhtt_mt_33_2018_bin0
196 CMS_ff_total_wjets_stat_njet1_jet_pt_high_unc1_mt_2018
197 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc2_mt_2018
198 prop_binhtt_mt_32_2018_bin17
199 CMS_scale_t_3prong1pizero_2018
200 CMS_ff_total_qcd_stat_njet1_jet_pt_high_unc2_mt_2018
201 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc1_mt_2018
202 prop_binhtt_mt_36_2018_bin14
203 prop_binhtt_mt_36_2018_bin13
×10
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.4 −0.2 0 0.2 0.4
$\Delta \hat{r}_{ggH}$

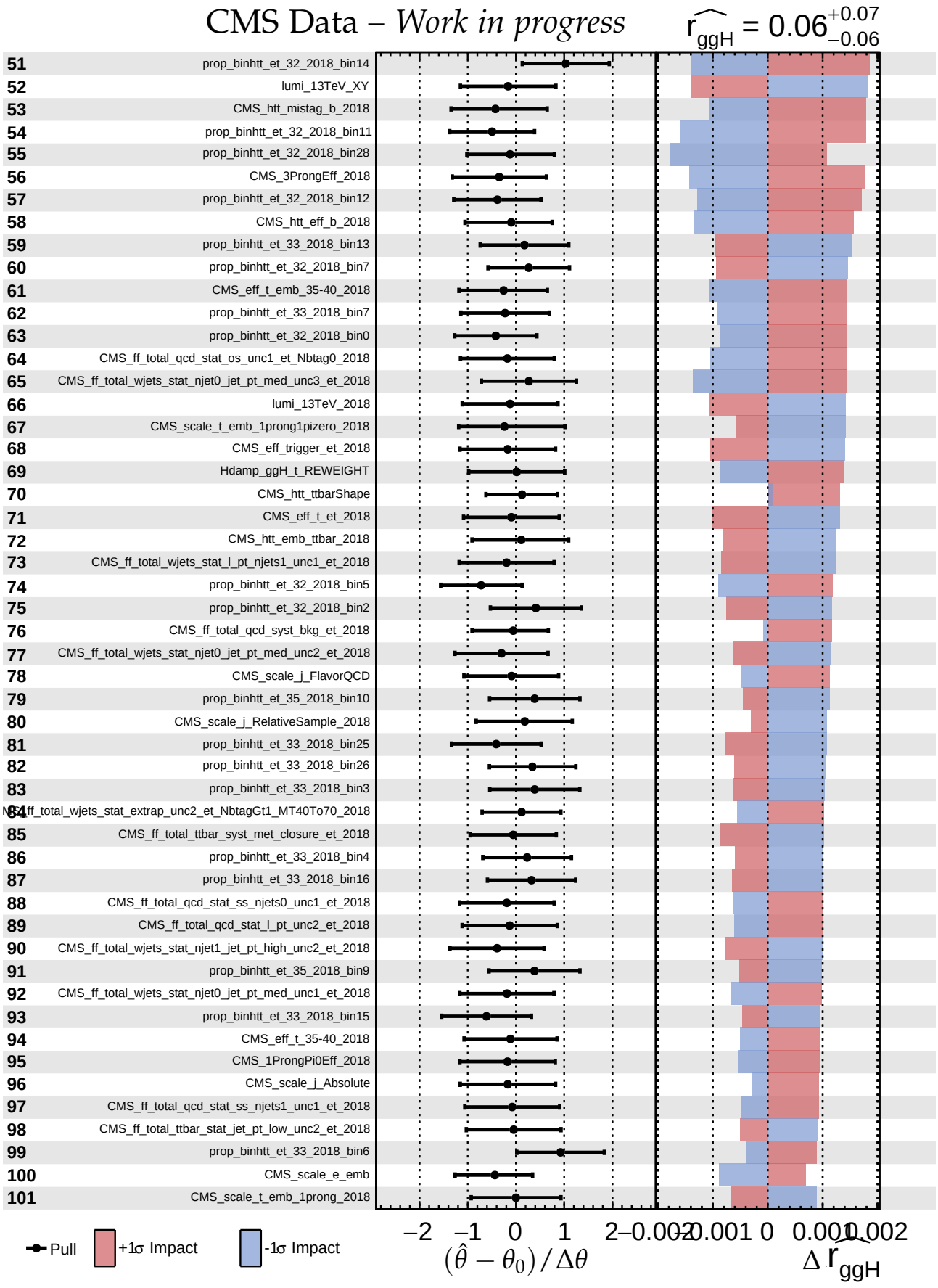

CMS Data – Work in progress
$\widehat{r_{ggH}} = -0.07^{+0.05}_{-0.05}$
204 CMS_eff_xtrigger_t_emb_mt_dm0_2018
205 prop_binhtt_mt_33_2018_bin12
206 CMS_scale_j_Absolute_2018
207 CMS_htt_dyShape
208 CMS_eff_xtrigger_t_emb_mt_dm10_2018
209 CMS_ff_total_ttbar_stat_jet_pt_low_unc1_mt_2018
210 CMS_scale_j_BBEC1_2018
211 prop_binhtt_mt_33_2018_bin3
212 CMS_ff_total_ttbar_stat_l_pt_unc2_mt_MTLt40_2018
213 CMS_scale_t_1prong1pizero_2018
214 prop_binhtt_mt_35_2018_bin8
215 prop_binhtt_mt_36_2018_bin6
216 CMS_eff_xtrigger_t_emb_mt_dm1_2018
217 prop_binhtt_mt_32_2018_bin2
218 CMS_ff_total_ttbar_stat_l_pt_unc1_mt_MT40To70_2018
219 prop_binhtt_mt_36_2018_bin4
220 prop_binhtt_mt_32_2018_bin0
221 CMS_eff_xtrigger_t_mt_dm0_2018
222 CMS_ff_total_ttbar_stat_jet_pt_high_unc3_mt_2018
223 prop_binhtt_mt_35_2018_bin16
224 CMS_eff_xtrigger_t_mt_dm10_2018
225 CMS_ff_total_qcd_stat_njet0_jet_pt_low_unc1_mt_2018
226 CMS_ff_total_qcd_stat_njet0_jet_pt_low_unc2_mt_2018
227 CMS_eff_xtrigger_t_emb_mt_dm11_2018
228 CMS_ff_total_ttbar_stat_met_unc2_mt_2018
229 prop_binhtt_mt_36_2018_bin2
230 prop_binhtt_mt_36_2018_bin5
231 prop_binhtt_mt_36_2018_bin7
232 prop_binhtt_mt_35_2018_bin9
233 CMS_eff_xtrigger_t_mt_dm1_2018
234 CMS_eff_trigger_single_t_emb_2018
235 prop_binhtt_mt_33_2018_bin6
236 prop_binhtt_mt_32_2018_bin6
237 CMS_ff_total_qcd_stat_os_unc1_mt_NbtagGt1_2018
238 CMS_ff_total_wjets_stat_njet1_jet_pt_med_unc1_mt_2018
239 prop_binhtt_mt_36_2018_bin15
240 CMS_eff_t_emb_highpT_500-inf_2018
241 CMS_ff_total_wjets_stat_njet0_jet_pt_high_unc3_mt_2018
242 CMS_eff_xtrigger_t_mt_dm11_2018
243 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc3_mt_2018
244 CMS_ff_total_wjets_stat_njet1_jet_pt_high_unc3_mt_2018
245 prop_binhtt_mt_33_2018_bin29
246 CMS_ff_total_qcd_stat_njet1_jet_pt_high_unc1_mt_2018
247 CMS_ff_total_qcd_stat_njet0_jet_pt_high_unc1_mt_2018
248 CMS_scale_j_EC2_2018
249 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc2_mt_2018
250 prop_binhtt_mt_35_2018_bin18
251 CMS_ff_total_ttbar_stat_met_unc1_mt_2018
252 CMS_scale_j_HF_2018
253 CMS_eff_t_highpT_500-inf_2018
254 prop_binhtt_mt_35_2018_bin15
×10⁻²
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.2 −0.1 0 0.1 0.2
$\Delta\, r_{ggH}$

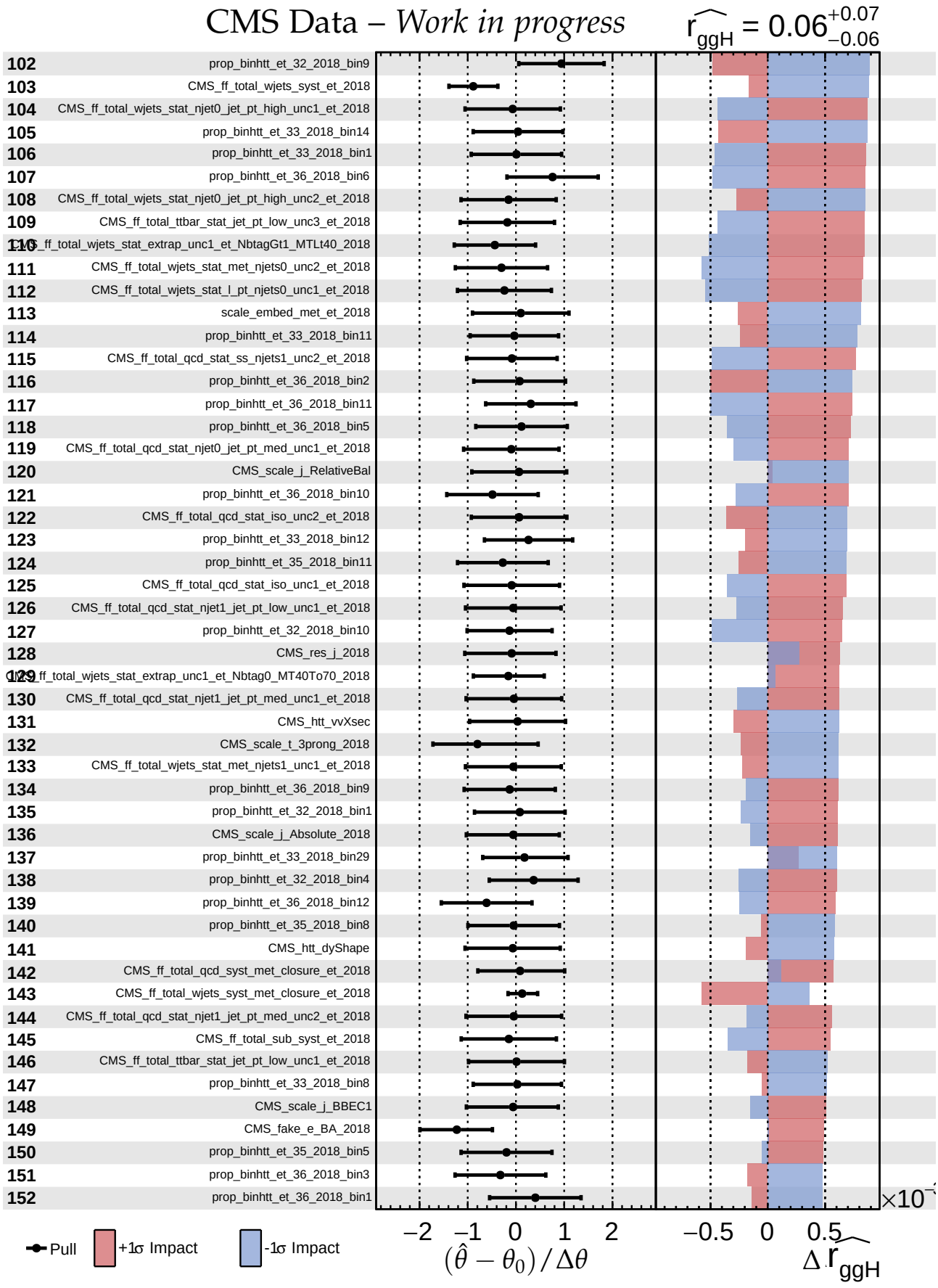


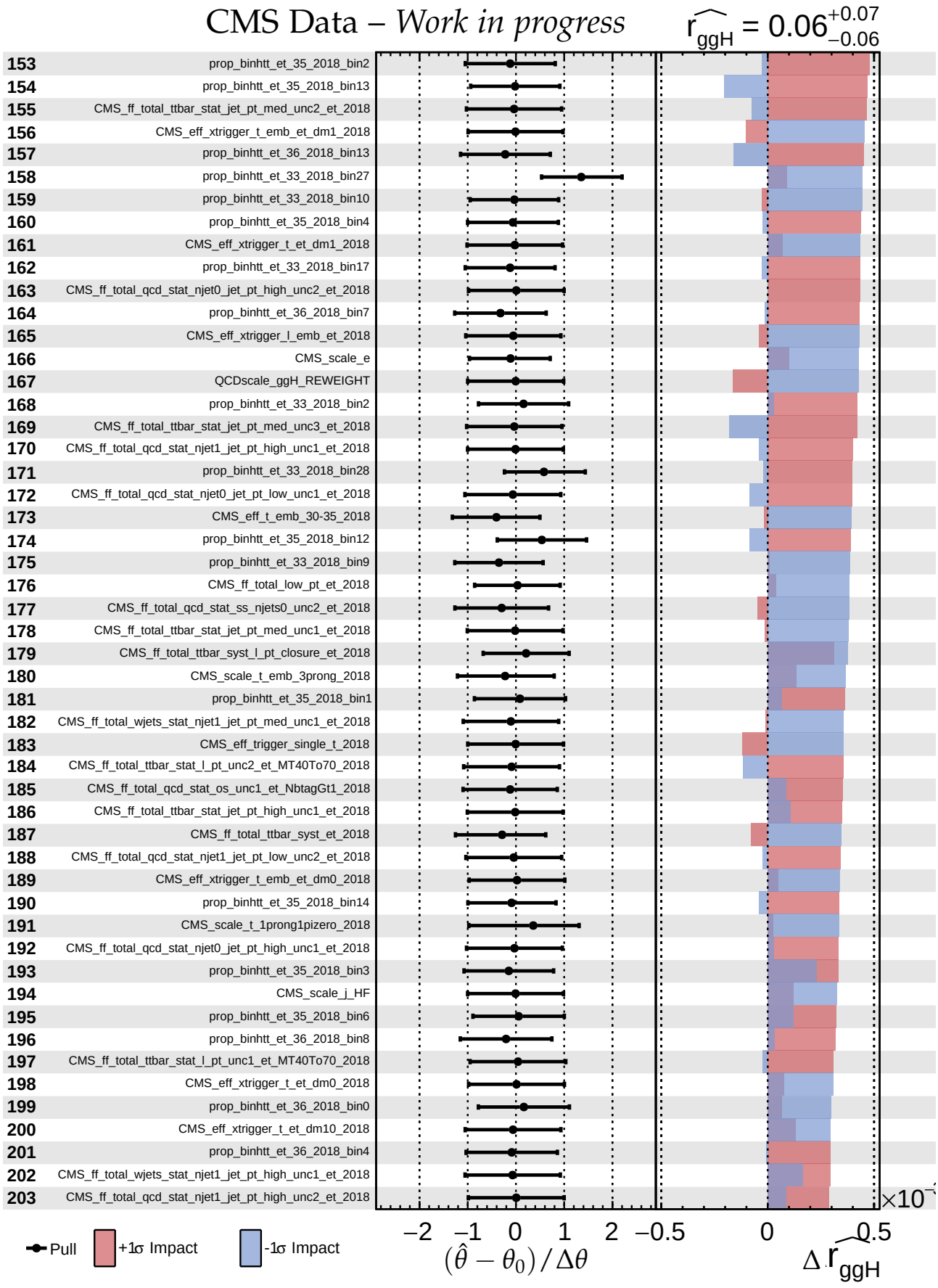
Année 2018, canal $\mu\tau_h$, page 6/6



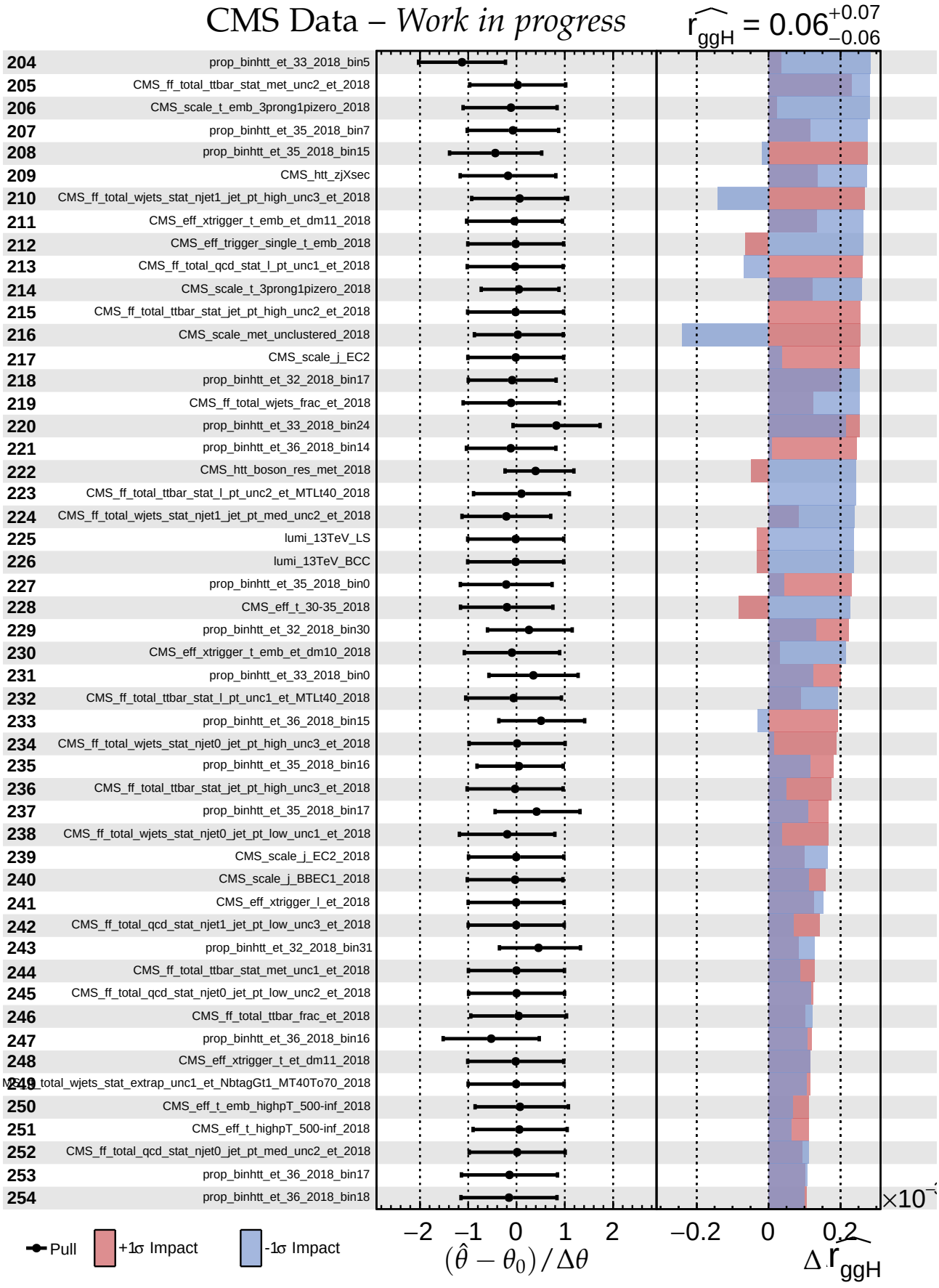
Année 2018, canal $e\tau_h$, page 1/6


CMS Data – Work in progress
$\hat{r}_{ggH} = 0.06^{+0.07}_{-0.06}$
51 prop_binhtt_et_32_2018_bin14
52 lumi_13TeV_XY
53 CMS_htt_mistag_b_2018
54 prop_binhtt_et_32_2018_bin11
55 prop_binhtt_et_32_2018_bin28
56 CMS_3ProngEff_2018
57 prop_binhtt_et_32_2018_bin12
58 CMS_htt_eff_b_2018
59 prop_binhtt_et_33_2018_bin13
60 prop_binhtt_et_32_2018_bin7
61 CMS_eff_t_emb_35-40_2018
62 prop_binhtt_et_33_2018_bin7
63 prop_binhtt_et_32_2018_bin0
64 CMS_ff_total_qcd_stat_os_unc1_et_Nbtag0_2018
65 CMS_ff_total_wjets_stat_njet0_jet_pt_med_unc3_et_2018
66 lumi_13TeV_2018
67 CMS_scale_t_emb_1prong1pizero_2018
68 CMS_eff_trigger_et_2018
69 Hdamp_ggH_t_REWEIGHT
70 CMS_htt_ttbarShape
71 CMS_eff_t_et_2018
72 CMS_htt_emb_ttbar_2018
73 CMS_ff_total_wjets_stat_l_pt_njets1_unc1_et_2018
74 prop_binhtt_et_32_2018_bin5
75 prop_binhtt_et_32_2018_bin2
76 CMS_ff_total_qcd_syst_bkg_et_2018
77 CMS_ff_total_wjets_stat_njet0_jet_pt_med_unc2_et_2018
78 CMS_scale_j_FlavorQCD
79 prop_binhtt_et_35_2018_bin10
80 CMS_scale_j_RelativeSample_2018
81 prop_binhtt_et_33_2018_bin25
82 prop_binhtt_et_33_2018_bin26
83 prop_binhtt_et_33_2018_bin3
84 CMS_ff_total_wjets_stat_extrap_unc2_et_NbtagGt1_MT40To70_2018
85 CMS_ff_total_ttbar_syst_met_closure_et_2018
86 prop_binhtt_et_33_2018_bin4
87 prop_binhtt_et_33_2018_bin16
88 CMS_ff_total_qcd_stat_ss_njets0_unc1_et_2018
89 CMS_ff_total_qcd_stat_l_pt_unc2_et_2018
90 CMS_ff_total_wjets_stat_njet1_jet_pt_high_unc2_et_2018
91 prop_binhtt_et_35_2018_bin9
92 CMS_ff_total_wjets_stat_njet0_jet_pt_med_unc1_et_2018
93 prop_binhtt_et_33_2018_bin15
94 CMS_eff_t_35-40_2018
95 CMS_1ProngPi0Eff_2018
96 CMS_scale_j_Absolute
97 CMS_ff_total_qcd_stat_ss_njets1_unc1_et_2018
98 CMS_ff_total_ttbar_stat_jet_pt_low_unc2_et_2018
99 prop_binhtt_et_33_2018_bin6
100 CMS_scale_e_emb
101 CMS_scale_t_emb_1prong_2018
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.002 −0.001 0 0.001 0.002
$\Delta \hat{r}_{ggH}$


CMS Data – Work in progress
$\hat{r}_{ggH} = 0.06^{+0.07}_{-0.06}$
102 prop_binhtt_et_32_2018_bin9
103 CMS_ff_total_wjets_syst_et_2018
104 CMS_ff_total_wjets_stat_njet0_jet_pt_high_unc1_et_2018
105 prop_binhtt_et_33_2018_bin14
106 prop_binhtt_et_33_2018_bin1
107 prop_binhtt_et_36_2018_bin6
108 CMS_ff_total_wjets_stat_njet0_jet_pt_high_unc2_et_2018
109 CMS_ff_total_ttbar_stat_jet_pt_low_unc3_et_2018
110 CMS_ff_total_wjets_stat_extrap_unc1_et_NbtagGt1_MTLt40_2018
111 CMS_ff_total_wjets_stat_met_njets0_unc2_et_2018
112 CMS_ff_total_wjets_stat_l_pt_njets0_unc1_et_2018
113 scale_embed_met_et_2018
114 prop_binhtt_et_33_2018_bin11
115 CMS_ff_total_qcd_stat_ss_njets1_unc2_et_2018
116 prop_binhtt_et_36_2018_bin2
117 prop_binhtt_et_36_2018_bin11
118 prop_binhtt_et_36_2018_bin5
119 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc1_et_2018
120 CMS_scale_j_RelativeBal
121 prop_binhtt_et_36_2018_bin10
122 CMS_ff_total_qcd_stat_iso_unc2_et_2018
123 prop_binhtt_et_33_2018_bin12
124 prop_binhtt_et_35_2018_bin11
125 CMS_ff_total_qcd_stat_iso_unc1_et_2018
126 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc1_et_2018
127 prop_binhtt_et_32_2018_bin10
128 CMS_res_j_2018
129 CMS_ff_total_wjets_stat_extrap_unc1_et_Nbtag0_MT40To70_2018
130 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc1_et_2018
131 CMS_htt_vvXsec
132 CMS_scale_t_3prong_2018
133 CMS_ff_total_wjets_stat_met_njets1_unc1_et_2018
134 prop_binhtt_et_36_2018_bin9
135 prop_binhtt_et_32_2018_bin1
136 CMS_scale_j_Absolute_2018
137 prop_binhtt_et_33_2018_bin29
138 prop_binhtt_et_32_2018_bin4
139 prop_binhtt_et_36_2018_bin12
140 prop_binhtt_et_35_2018_bin8
141 CMS_htt_dyShape
142 CMS_ff_total_qcd_syst_met_closure_et_2018
143 CMS_ff_total_wjets_syst_met_closure_et_2018
144 CMS_ff_total_qcd_stat_njet1_jet_pt_med_unc2_et_2018
145 CMS_ff_total_sub_syst_et_2018
146 CMS_ff_total_ttbar_stat_jet_pt_low_unc1_et_2018
147 prop_binhtt_et_33_2018_bin8
148 CMS_scale_j_BBEC1
149 CMS_fake_e_BA_2018
150 prop_binhtt_et_35_2018_bin5
151 prop_binhtt_et_36_2018_bin3
152 prop_binhtt_et_36_2018_bin1
×10
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.5 0 0.5
$\Delta \hat{r}_{ggH}$

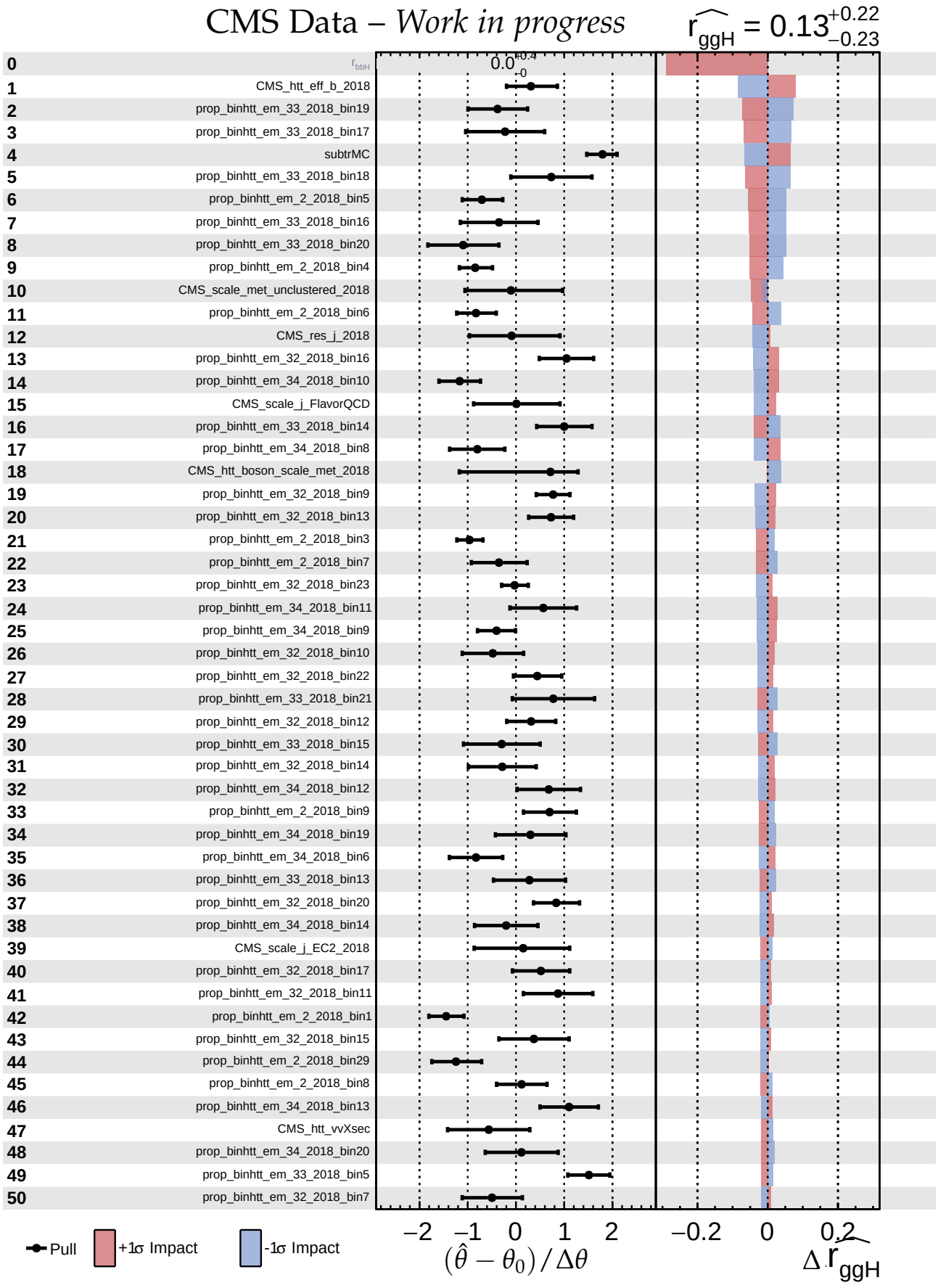


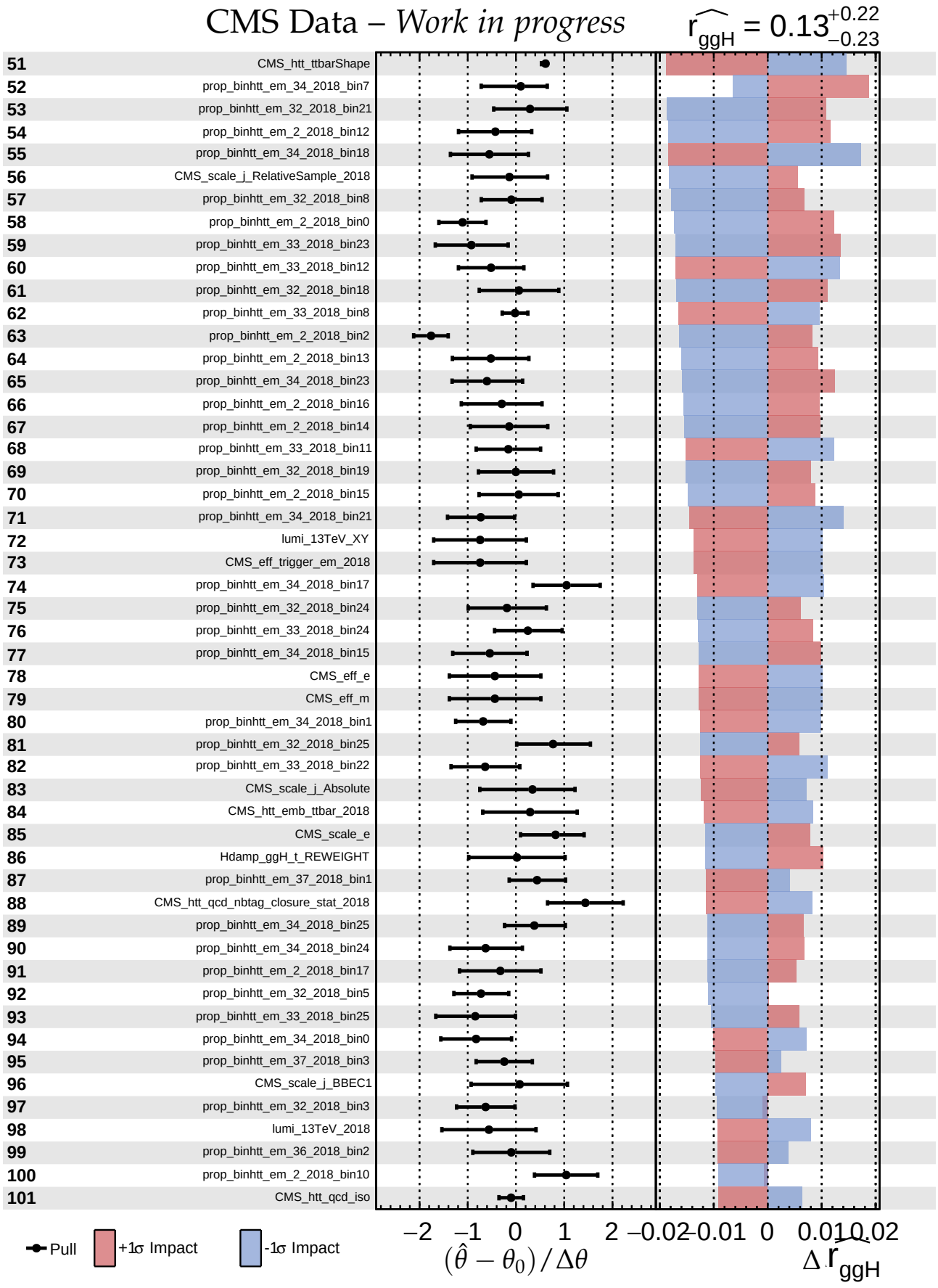
Année 2018, canal $e\tau_h$, page 4/6


CMS Data – Work in progress
$\hat{r}_{ggH} = 0.06^{+0.07}_{-0.06}$
204 prop_binhtt_et_33_2018_bin5
205 CMS_ff_total_ttbar_stat_met_unc2_et_2018
206 CMS_scale_t_emb_3prong1pizero_2018
207 prop_binhtt_et_35_2018_bin7
208 prop_binhtt_et_35_2018_bin15
209 CMS_htt_zjXsec
210 CMS_ff_total_wjets_stat_njet1_jet_pt_high_unc3_et_2018
211 CMS_eff_xtrigger_t_emb_et_dm11_2018
212 CMS_eff_trigger_single_t_emb_2018
213 CMS_ff_total_qcd_stat_l_pt_unc1_et_2018
214 CMS_scale_t_3prong1pizero_2018
215 CMS_ff_total_ttbar_stat_jet_pt_high_unc2_et_2018
216 CMS_scale_met_unclustered_2018
217 CMS_scale_j_EC2
218 prop_binhtt_et_32_2018_bin17
219 CMS_ff_total_wjets_frac_et_2018
220 prop_binhtt_et_33_2018_bin24
221 prop_binhtt_et_36_2018_bin14
222 CMS_htt_boson_res_met_2018
223 CMS_ff_total_ttbar_stat_l_pt_unc2_et_MTLt40_2018
224 CMS_ff_total_wjets_stat_njet1_jet_pt_med_unc2_et_2018
225 lumi_13TeV_LS
226 lumi_13TeV_BCC
227 prop_binhtt_et_35_2018_bin0
228 CMS_eff_t_30-35_2018
229 prop_binhtt_et_32_2018_bin30
230 CMS_eff_xtrigger_t_emb_et_dm10_2018
231 prop_binhtt_et_33_2018_bin0
232 CMS_ff_total_ttbar_stat_l_pt_unc1_et_MTLt40_2018
233 prop_binhtt_et_36_2018_bin15
234 CMS_ff_total_wjets_stat_njet0_jet_pt_high_unc3_et_2018
235 prop_binhtt_et_35_2018_bin16
236 CMS_ff_total_ttbar_stat_jet_pt_high_unc3_et_2018
237 prop_binhtt_et_35_2018_bin17
238 CMS_ff_total_wjets_stat_njet0_jet_pt_low_unc1_et_2018
239 CMS_scale_j_EC2_2018
240 CMS_scale_j_BBEC1_2018
241 CMS_eff_xtrigger_l_et_2018
242 CMS_ff_total_qcd_stat_njet1_jet_pt_low_unc3_et_2018
243 prop_binhtt_et_32_2018_bin31
244 CMS_ff_total_ttbar_stat_met_unc1_et_2018
245 CMS_ff_total_qcd_stat_njet0_jet_pt_low_unc2_et_2018
246 CMS_ff_total_ttbar_frac_et_2018
247 prop_binhtt_et_36_2018_bin16
248 CMS_eff_xtrigger_t_et_dm11_2018
249 CMS_ff_total_wjets_stat_extrap_unc1_et_NbtagGt1_MT40To70_2018
250 CMS_eff_t_emb_highpT_500-inf_2018
251 CMS_eff_t_highpT_500-inf_2018
252 CMS_ff_total_qcd_stat_njet0_jet_pt_med_unc2_et_2018
253 prop_binhtt_et_36_2018_bin17
254 prop_binhtt_et_36_2018_bin18
×10⁻¹
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0)/\Delta\theta$
−0.2 0 0.2
$\Delta\hat{r}_{ggH}$

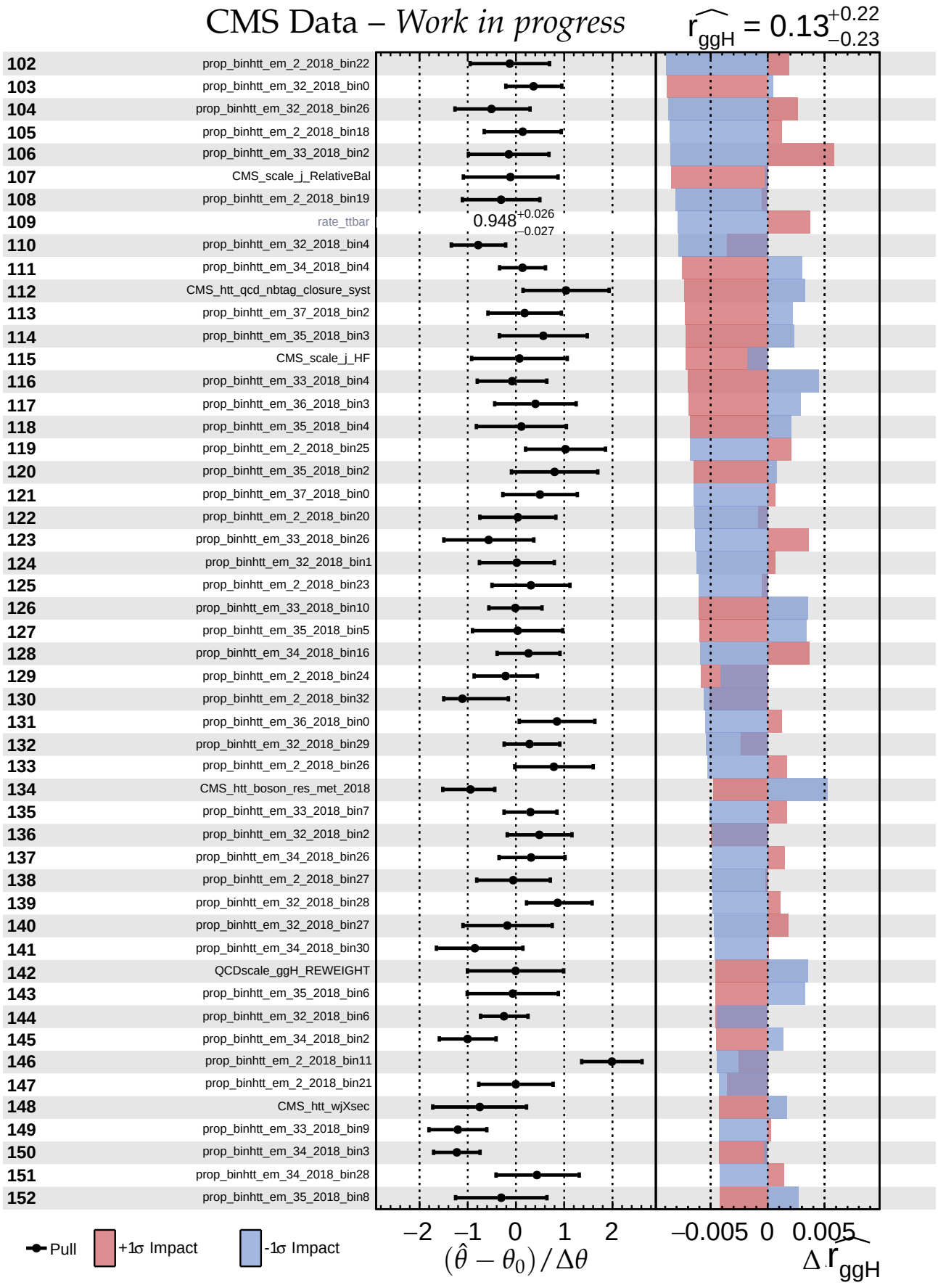


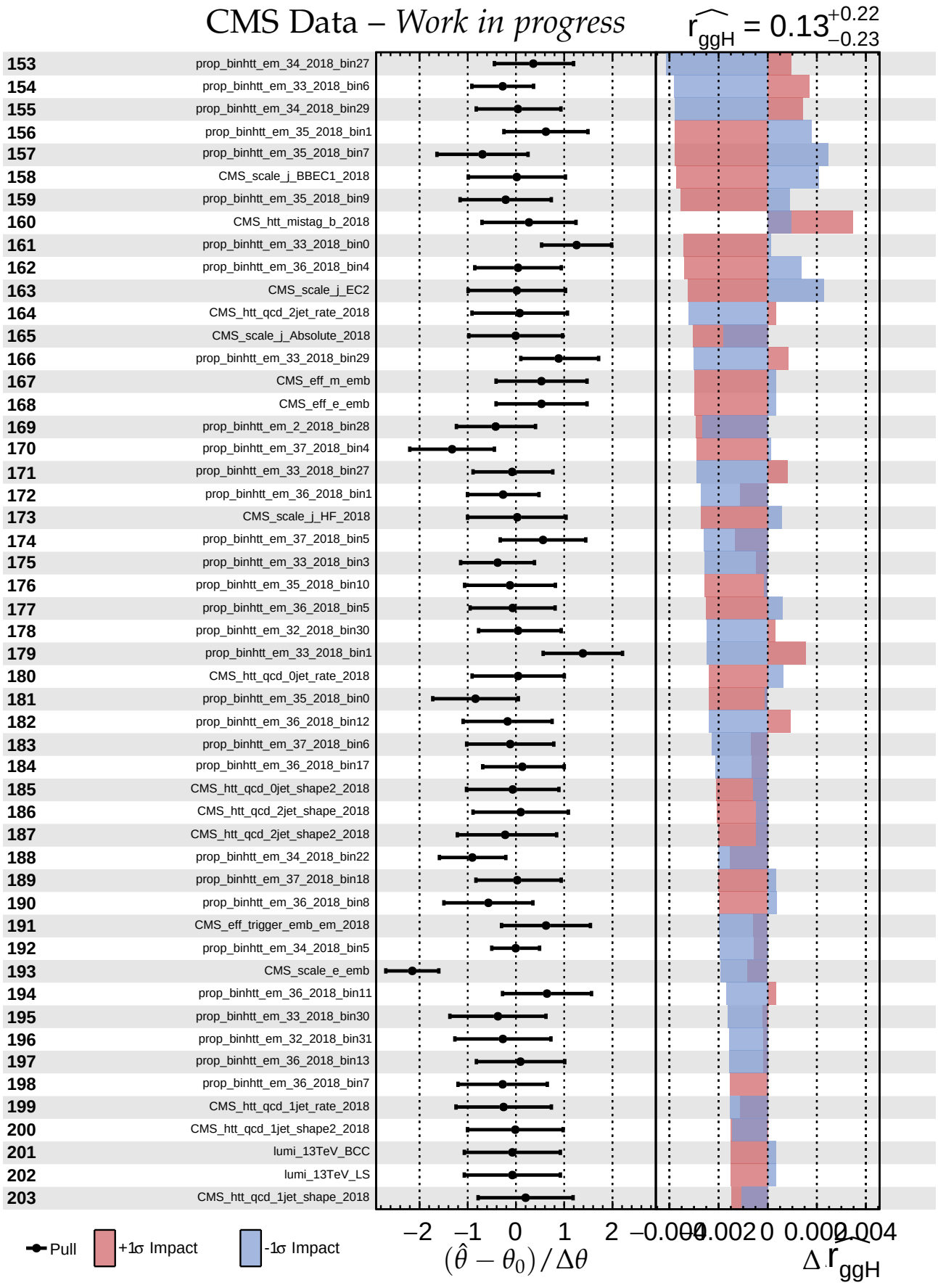
Année 2018, canal $e\tau_h$, page 6/6

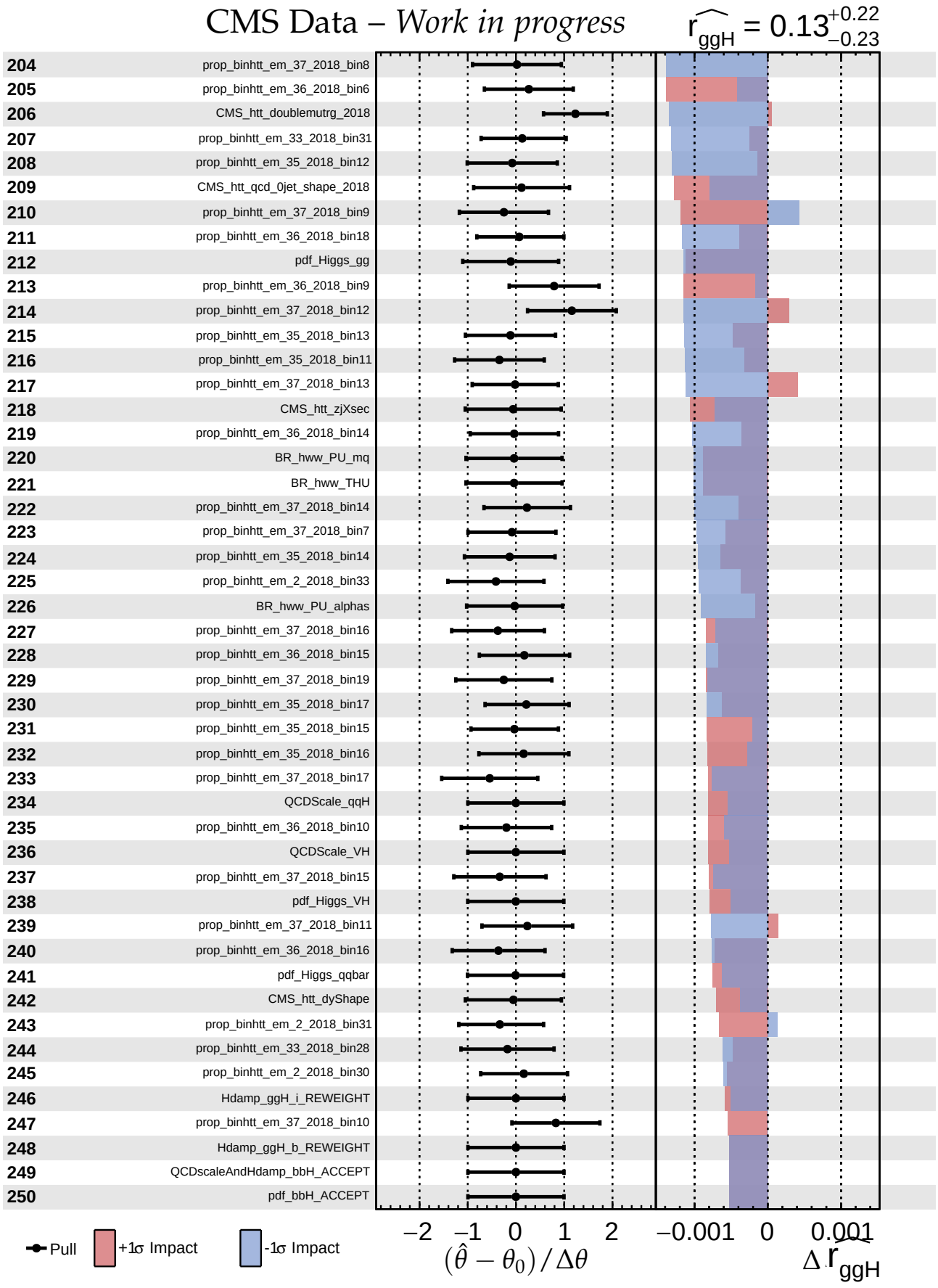

CMS Data – Work in progress
$\hat{r}_{ggH} = 0.13^{+0.22}_{-0.23}$
0 r_{bbH} $0.0^{+0.4}_{-0}$
1 CMS_htt_eff_b_2018
2 prop_binhtt_em_33_2018_bin19
3 prop_binhtt_em_33_2018_bin17
4 subtrMC
5 prop_binhtt_em_33_2018_bin18
6 prop_binhtt_em_2_2018_bin5
7 prop_binhtt_em_33_2018_bin16
8 prop_binhtt_em_33_2018_bin20
9 prop_binhtt_em_2_2018_bin4
10 CMS_scale_met_unclustered_2018
11 prop_binhtt_em_2_2018_bin6
12 CMS_res_j_2018
13 prop_binhtt_em_32_2018_bin16
14 prop_binhtt_em_34_2018_bin10
15 CMS_scale_j_FlavorQCD
16 prop_binhtt_em_33_2018_bin14
17 prop_binhtt_em_34_2018_bin8
18 CMS_htt_boson_scale_met_2018
19 prop_binhtt_em_32_2018_bin9
20 prop_binhtt_em_32_2018_bin13
21 prop_binhtt_em_2_2018_bin3
22 prop_binhtt_em_2_2018_bin7
23 prop_binhtt_em_32_2018_bin23
24 prop_binhtt_em_34_2018_bin11
25 prop_binhtt_em_34_2018_bin9
26 prop_binhtt_em_32_2018_bin10
27 prop_binhtt_em_32_2018_bin22
28 prop_binhtt_em_33_2018_bin21
29 prop_binhtt_em_32_2018_bin12
30 prop_binhtt_em_33_2018_bin15
31 prop_binhtt_em_32_2018_bin14
32 prop_binhtt_em_34_2018_bin12
33 prop_binhtt_em_2_2018_bin9
34 prop_binhtt_em_34_2018_bin19
35 prop_binhtt_em_34_2018_bin6
36 prop_binhtt_em_33_2018_bin13
37 prop_binhtt_em_32_2018_bin20
38 prop_binhtt_em_34_2018_bin14
39 CMS_scale_j_EC2_2018
40 prop_binhtt_em_32_2018_bin17
41 prop_binhtt_em_32_2018_bin11
42 prop_binhtt_em_2_2018_bin1
43 prop_binhtt_em_32_2018_bin15
44 prop_binhtt_em_2_2018_bin29
45 prop_binhtt_em_2_2018_bin8
46 prop_binhtt_em_34_2018_bin13
47 CMS_htt_vvXsec
48 prop_binhtt_em_34_2018_bin20
49 prop_binhtt_em_33_2018_bin5
50 prop_binhtt_em_32_2018_bin7
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.2 0 0.2
$\Delta \hat{r}_{ggH}$



Année 2018, canal $e\mu$, page 2/5


CMS Data – Work in progress
$\hat{r}_{ggH} = 0.13^{+0.22}_{-0.23}$
102 prop_binhtt_em_2_2018_bin22
103 prop_binhtt_em_32_2018_bin0
104 prop_binhtt_em_32_2018_bin26
105 prop_binhtt_em_2_2018_bin18
106 prop_binhtt_em_33_2018_bin2
107 CMS_scale_j_RelativeBal
108 prop_binhtt_em_2_2018_bin19
109 rate_ttbar
$0.948^{+0.026}_{-0.027}$
110 prop_binhtt_em_32_2018_bin4
111 prop_binhtt_em_34_2018_bin4
112 CMS_htt_qcd_nbtag_closure_syst
113 prop_binhtt_em_37_2018_bin2
114 prop_binhtt_em_35_2018_bin3
115 CMS_scale_j_HF
116 prop_binhtt_em_33_2018_bin4
117 prop_binhtt_em_36_2018_bin3
118 prop_binhtt_em_35_2018_bin4
119 prop_binhtt_em_2_2018_bin25
120 prop_binhtt_em_35_2018_bin2
121 prop_binhtt_em_37_2018_bin0
122 prop_binhtt_em_2_2018_bin20
123 prop_binhtt_em_33_2018_bin26
124 prop_binhtt_em_32_2018_bin1
125 prop_binhtt_em_2_2018_bin23
126 prop_binhtt_em_33_2018_bin10
127 prop_binhtt_em_35_2018_bin5
128 prop_binhtt_em_34_2018_bin16
129 prop_binhtt_em_2_2018_bin24
130 prop_binhtt_em_2_2018_bin32
131 prop_binhtt_em_36_2018_bin0
132 prop_binhtt_em_32_2018_bin29
133 prop_binhtt_em_2_2018_bin26
134 CMS_htt_boson_res_met_2018
135 prop_binhtt_em_33_2018_bin7
136 prop_binhtt_em_32_2018_bin2
137 prop_binhtt_em_34_2018_bin26
138 prop_binhtt_em_2_2018_bin27
139 prop_binhtt_em_32_2018_bin28
140 prop_binhtt_em_32_2018_bin27
141 prop_binhtt_em_34_2018_bin30
142 QCDscale_ggH_REWEIGHT
143 prop_binhtt_em_35_2018_bin6
144 prop_binhtt_em_32_2018_bin6
145 prop_binhtt_em_34_2018_bin2
146 prop_binhtt_em_2_2018_bin11
147 prop_binhtt_em_2_2018_bin21
148 CMS_htt_wjXsec
149 prop_binhtt_em_33_2018_bin9
150 prop_binhtt_em_34_2018_bin3
151 prop_binhtt_em_34_2018_bin28
152 prop_binhtt_em_35_2018_bin8
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.005 0 0.005
$\Delta \hat{r}_{ggH}$


CMS Data – Work in progress
$\hat{r}_{ggH} = 0.13^{+0.22}_{-0.23}$
153 prop_binhtt_em_34_2018_bin27
154 prop_binhtt_em_33_2018_bin6
155 prop_binhtt_em_34_2018_bin29
156 prop_binhtt_em_35_2018_bin1
157 prop_binhtt_em_35_2018_bin7
158 CMS_scale_j_BBEC1_2018
159 prop_binhtt_em_35_2018_bin9
160 CMS_htt_mistag_b_2018
161 prop_binhtt_em_33_2018_bin0
162 prop_binhtt_em_36_2018_bin4
163 CMS_scale_j_EC2
164 CMS_htt_qcd_2jet_rate_2018
165 CMS_scale_j_Absolute_2018
166 prop_binhtt_em_33_2018_bin29
167 CMS_eff_m_emb
168 CMS_eff_e_emb
169 prop_binhtt_em_2_2018_bin28
170 prop_binhtt_em_37_2018_bin4
171 prop_binhtt_em_33_2018_bin27
172 prop_binhtt_em_36_2018_bin1
173 CMS_scale_j_HF_2018
174 prop_binhtt_em_37_2018_bin5
175 prop_binhtt_em_33_2018_bin3
176 prop_binhtt_em_35_2018_bin10
177 prop_binhtt_em_36_2018_bin5
178 prop_binhtt_em_32_2018_bin30
179 prop_binhtt_em_33_2018_bin1
180 CMS_htt_qcd_0jet_rate_2018
181 prop_binhtt_em_35_2018_bin0
182 prop_binhtt_em_36_2018_bin12
183 prop_binhtt_em_37_2018_bin6
184 prop_binhtt_em_36_2018_bin17
185 CMS_htt_qcd_0jet_shape2_2018
186 CMS_htt_qcd_2jet_shape_2018
187 CMS_htt_qcd_2jet_shape2_2018
188 prop_binhtt_em_34_2018_bin22
189 prop_binhtt_em_37_2018_bin18
190 prop_binhtt_em_36_2018_bin8
191 CMS_eff_trigger_emb_em_2018
192 prop_binhtt_em_34_2018_bin5
193 CMS_scale_e_emb
194 prop_binhtt_em_36_2018_bin11
195 prop_binhtt_em_33_2018_bin30
196 prop_binhtt_em_32_2018_bin31
197 prop_binhtt_em_36_2018_bin13
198 prop_binhtt_em_36_2018_bin7
199 CMS_htt_qcd_1jet_rate_2018
200 CMS_htt_qcd_1jet_shape2_2018
201 lumi_13TeV_BCC
202 lumi_13TeV_LS
203 CMS_htt_qcd_1jet_shape_2018
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.004 −0.002 0 0.002 0.004
$\Delta \hat{r}_{ggH}$


CMS Data – Work in progress
$\widehat{r_{ggH}} = 0.13^{+0.22}_{-0.23}$
204 prop_binhtt_em_37_2018_bin8
205 prop_binhtt_em_36_2018_bin6
206 CMS_htt_doublemutrg_2018
207 prop_binhtt_em_33_2018_bin31
208 prop_binhtt_em_35_2018_bin12
209 CMS_htt_qcd_0jet_shape_2018
210 prop_binhtt_em_37_2018_bin9
211 prop_binhtt_em_36_2018_bin18
212 pdf_Higgs_gg
213 prop_binhtt_em_36_2018_bin9
214 prop_binhtt_em_37_2018_bin12
215 prop_binhtt_em_35_2018_bin13
216 prop_binhtt_em_35_2018_bin11
217 prop_binhtt_em_37_2018_bin13
218 CMS_htt_zjXsec
219 prop_binhtt_em_36_2018_bin14
220 BR_hww_PU_mq
221 BR_hww_THU
222 prop_binhtt_em_37_2018_bin14
223 prop_binhtt_em_37_2018_bin7
224 prop_binhtt_em_35_2018_bin14
225 prop_binhtt_em_2_2018_bin33
226 BR_hww_PU_alphas
227 prop_binhtt_em_37_2018_bin16
228 prop_binhtt_em_36_2018_bin15
229 prop_binhtt_em_37_2018_bin19
230 prop_binhtt_em_35_2018_bin17
231 prop_binhtt_em_35_2018_bin15
232 prop_binhtt_em_35_2018_bin16
233 prop_binhtt_em_37_2018_bin17
234 QCDScale_qqH
235 prop_binhtt_em_36_2018_bin10
236 QCDScale_VH
237 prop_binhtt_em_37_2018_bin15
238 pdf_Higgs_VH
239 prop_binhtt_em_37_2018_bin11
240 prop_binhtt_em_36_2018_bin16
241 pdf_Higgs_qqbar
242 CMS_htt_dyShape
243 prop_binhtt_em_2_2018_bin31
244 prop_binhtt_em_33_2018_bin28
245 prop_binhtt_em_2_2018_bin30
246 Hdamp_ggH_i_REWEIGHT
247 prop_binhtt_em_37_2018_bin10
248 Hdamp_ggH_b_REWEIGHT
249 QCDscaleAndHdamp_bbH_ACCEPT
250 pdf_bbH_ACCEPT
Pull
+1σ Impact
-1σ Impact
−2 −1 0 1 2
$(\hat{\theta} - \theta_0) / \Delta\theta$
−0.001 0 0.001
$\Delta\, r_{ggH}$

Bibliographie

[1] S. L. Glashow. « The renormalizability of vector meson interactions ». *Nuclear Physics* **10** (1959), p. 107-117. doi : 10.1016/0029-5582(59)90196-8.

[2] A. Salam & J. Ward. « Weak and electromagnetic interactions ». *Nuovo Cim* **11** (fév. 1959), p. 568-577. doi : 10.1007/BF02726525.

[3] S. Weinberg. « A model of leptons ». *Physical Review Letters* **19** (21 nov. 1967), p. 1264-1266. doi : 10.1103/PhysRevLett.19.1264.

[4] C. S. Wu & coll. « Experimental Test of Parity Conservation in Beta Decay ». *Physical Review* **105** (4 fév. 1957), p. 1413-1415. doi : 10.1103/PhysRev.105.1413.

[5] J. H. Christenson & coll. « Evidence for the 2π Decay of the K_2^0 Meson ». *Physical Review Letters* **13** (4 juil. 1964), p. 138-140. doi : 10.1103/PhysRevLett.13.138.

[6] M. Kobayashi & T. Maskawa. « CP-Violation in the Renormalizable Theory of Weak Interaction ». *Progress of Theoretical Physics* **49**.2 (fév. 1973), p. 652-657. doi : 10.1143/PTP.49.652.

[7] N. Cabibbo. « Unitary Symmetry and Leptonic Decays ». *Physical Review Letters* **10** (12 juin 1963), p. 531-533. doi : 10.1103/PhysRevLett.10.531.

[8] N. Cabibbo. « Unitary Symmetry and Nonleptonic Decays ». *Physical Review Letters* **12** (2 jan. 1964), p. 62-63. doi : 10.1103/PhysRevLett.12.62.

[9] Particle Data Group. « Review of Particle Physics ». *Progress of Theoretical and Experimental Physics* **8** (août 2020). doi : 10.1093/ptep/ptaa104.

[10] F. Englert & R. Brout. « Broken symmetry and the mass of gauge vector mesons ». *Physical Review Letters* **13**.9 (9 août 1964), p. 321-323. doi : 10.1103/PhysRevLett.13.321.

[11] P. W. Higgs. « Broken symmetries, massless particles and gauge fields ». *Physics Letters* **12**.2 (sept. 1964). doi : 10.1016/0031-9163(64)91136-9.

[12] P. W. Higgs. « Broken symmetries and the masses of gauge bosons ». *Physics Letters* **13**.16 (oct. 1964). doi : 10.1103/physrevlett.13.508.

[13] G. S. Guralnik, C. R. Hagen & T. W. B. Kibble. « Global Conservation Laws and Massless Particles ». *Physical Review Letters* **13**.20 (20 nov. 1964), p. 585-587. doi : 10.1103/PhysRevLett.13.585.

[14] T. W. B. Kibble. « Symmetry Breaking in Non-Abelian Gauge Theories ». *Physical Review* **155** (5 mar. 1967), p. 1554-1561. doi : 10.1103/PhysRev.155.1554.

[15] P. W. Higgs. « Spontaneous symmetry breakdown without massless bosons ». *Physical Review* **145** (4 mai 1966), p. 1156-1163. doi : 10.1103/PhysRev.145.1156.

[16] S. L. Glashow. « Partial symmetries of weak interactions ». *Nuclear Physics* **22**.4 (1961), p. 579-588. doi : 10.1016/0029-5582(61)90469-2.

[17] G. Arnison & coll. « Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540\,\text{GeV}$ ». *Physics Letters* **B122**.1 (1983), p. 103-116. doi : 10.1016/0370-2693(83)91177-2.

[18] M. Banner & coll. « Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider ». *Physics Letters* **B122**.5 (1983), p. 476-485. doi : 10.1016/0370-2693(83)91605-2.

[19] G. Arnison & coll. « Further evidence for charged intermediate vector bosons at the SPS collider ». *Physics Letters* **B129**.3 (1983), p. 273-282. DOI : 10.1016/0370-2693(83)90860-2.

[20] G. Arnison & coll. « Experimental observation of lepton pairs of invariant mass around 95 GeV·c^{-2} at the CERN SPS collider ». *Physics Letters* **B126**.5 (1983), p. 398-410. DOI : 10.1016/0370-2693(83)90188-0.

[21] P. Bagnaia & coll. « Evidence for $Z^0 \to e^+e^-$ at the CERN pp collider ». *Physics Letters* **B129**.1 (1983), p. 130-140. DOI : 10.1016/0370-2693(83)90744-X.

[22] The ATLAS Collaboration. « Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC ». *Physics Letters* **B716**.1 (2012), p. 1-29. DOI : 10.1016/j.physletb.2012.08.020.

[23] The CMS Collaboration. « Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC ». *Physics Letters* **B716**.1 (2012), p. 30-61. DOI : 10.1016/j.physletb.2012.08.021.

[24] The CMS Collaboration. « Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV ». *Journal of High Energy Physics* **06** (juin 2013). DOI : 10.1007/jhep06(2013)081.

[25] G. Aad & coll. « Combined measurement of the Higgs boson mass in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments ». *Physical Review Letters* **114**.19 (mai 2015). DOI : 10.1103/physrevlett.114.191803.

[26] G. Aad & coll. « 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 ». *Journal of High Energy Physics* **08** (août 2016). DOI : 10.1007/jhep08(2016)045.

[27] The CMS Collaboration. « A measurement of the Higgs boson mass in the diphoton decay channel ». *Physics Letters* **B805** (2020), p. 135425. DOI : 10.1016/j.physletb.2020.135425.

[28] The CMS Collaboration. « Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state ». *Physical Review* **D99** (11 juin 2019), p. 112003. DOI : 10.1103/PhysRevD.99.112003.

[29] The CMS Collaboration. « Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states ». *Journal of High Energy Physics* **96** (jan. 2014). DOI : 10.1007/JHEP01(2014)096.

[30] The CMS Collaboration. *Properties of the Higgs-like boson in the decay $H \to ZZ \to 4\ell$ in pp collisions at $\sqrt{s} = 7$ and 8 TeV*. Rapp. tech. CMS-PAS-HIG-13-002. Geneva : CERN, 2013. URL : https://cds.cern.ch/record/1523767.

[31] The CMS Collaboration. « Measurement of the properties of a Higgs boson in the four-lepton final state ». *Physical Review* **D89** (9 mai 2014), p. 092007. DOI : 10.1103/PhysRevD.89.092007.

[32] The CMS Collaboration. « Observation of the diphoton decay of the Higgs boson and measurement of its properties ». *The European Physical Journal* **C74** (oct. 2014). DOI : 10.1140/epjc/s10052-014-3076-z.

[33] The CMS Collaboration. « Search for a Higgs boson decaying into a b-quark pair and produced in association with b quarks in proton-proton collisions at 7 TeV ». *Physics Letters* **B722** (2013), p. 207-232. DOI : 10.1016/j.physletb.2013.04.017. arXiv : 1302.2892 [hep-ex].

[34] The CMS Collaboration. « Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks ». *Physical Review* **D89** (jan. 2014), p. 012003. DOI : 10.1103/PhysRevD.89.012003.

[35] The CMS Collaboration. « Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons ». *Journal of High Energy Physics* **05** (20 jan. 2014). DOI : 10.1007/JHEP05(2014)104. arXiv : 1401.5041v2 [hep-ex].

[36] The CMS Collaboration. « Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector ». *Physics Letters* **B779** (avr. 2018), p. 283-316. DOI : 10.1016/j.physletb.2018.02.004.

[37] The CMS Collaboration. *Measurement of Higgs boson decay to a pair of muons in proton-proton collisions at* $\sqrt{s} = 13\,\text{TeV}$. Rapp. tech. CMS-PAS-HIG-19-006. Geneva : CERN, 2020. URL : https://cds.cern.ch/record/2725423.

[38] The CMS Collaboration. « Observation of $t\bar{t}h$ Production ». *Physical Review Letters* **120** (23 juin 2018), p. 231801. DOI : 10.1103/PhysRevLett.120.231801.

[39] T. M. Hahn & coll. « Neutrons and Gamma-Rays from the Proton Bombardment of Beryllium ». *Physical Review* **85**.5 (mar. 1952), p. 934. DOI : 10.1103/PhysRev.85.934.

[40] H. L. Anderson & coll. « Total Cross Sections of Negative Pions in Hydrogen ». *Physical Review* **85**.5 (mar. 1952), p. 934-935. DOI : 10.1103/PhysRev.85.934.2.

[41] H. L. Anderson & coll. « Total Cross Sections of Positive Pions in Hydrogen ». *Physical Review* **85**.5 (mar. 1952), p. 936. DOI : 10.1103/PhysRev.85.936.

[42] V. E. Barnes & coll. « Observation of a Hyperon with Strangeness Minus Three ». *Physical Review Letters* **12**.8 (fév. 1964), p. 204-206. DOI : 10.1103/PhysRevLett.12.204.

[43] S. Dürr & coll. « Ab Initio Determination of Light Hadron Masses ». *Science* **322**.5905 (nov. 2008), p. 1224-1227. DOI : 10.1126/science.1163233.

[44] G. P. Salam. *Elements of QCD for hadron colliders*. 2010. arXiv : 1011.5131 [hep-ph]. URL : https://arxiv.org/pdf/1011.5131.pdf.

[45] D. J. Gross & F. Wilczek. « Ultraviolet Behavior of Non-Abelian Gauge Theories ». *Physical Review Letters* **30** (26 juin 1973), p. 1343-1346. DOI : 10.1103/PhysRevLett.30.1343.

[46] H. D. Politzer. « Reliable Perturbative Results for Strong Interactions ? » *Physical Review Letters* **30** (26 juin 1973), p. 1346-1349. DOI : 10.1103/PhysRevLett.30.1346.

[47] J.-C. Winter, F. Krauss & G. Soff. « A modified cluster-hadronisation model ». *European Physical Journal* **C36**.3 (août 2004), p. 381-395. DOI : 10.1140/epjc/s2004-01960-8.

[48] B. Andersson & coll. « Parton fragmentation and string dynamics » (avr. 1983). URL : http://cds.cern.ch/record/143980.

[49] The DØ Collaboration. « Measurement of the $B^0{}_s$ lifetime in the exclusive decay channel $B^0{}_s \to J/\Psi\phi$ ». *Physical Review Letters* **94** (fév. 2005). DOI : 10.1103/physrevlett.94.042001.

[50] G. Bellini, I. Bigi & P. Dornan. « Lifetimes of charm and beauty hadrons ». *Physics Reports* **289**.1 (1997), p. 1-155. DOI : 10.1016/S0370-1573(97)00005-7.

[51] L. Torterotot. *CMSTransverseTikZ : event displays in the CMS transverse plane with TikZ*. URL : https://gitlab.com/lucastorterotot/cmstransversetikz.

[52] The CMS Collaboration. *Performance of quark/gluon discrimination in* 8 TeV *pp data*. Rapp. tech. CMS-PAS-JME-13-002. Geneva : CERN, 2013. URL : http://cds.cern.ch/record/1599732.

[53] The CDF Collaboration. « Observation of top quark production in $p\bar{p}$ collisions with the collider detector at Fermilab ». *Physical Review Letters* **74**.14 (avr. 1995), p. 2626-2631. DOI : 10.1103/physrevlett.74.2626.

[54] The DØ Collaboration. « Observation of the top quark ». *Physical Review Letters* **74**.14 (avr. 1995), p. 2632-2637. DOI : 10.1103/physrevlett.74.2632.

[55] S. Mele. « The Measurement of the Number of Light Neutrino Species at LEP ». *Advanced Series on Directions in High Energy Physics* **23** (2015), p. 89-106. URL : http://cds.cern.ch/record/2103251.

[56] The Super-Kamiokande Collaboration. « Evidence for oscillation of atmospheric neutrinos ». *Physical Review Letters* **81** (8 août 1998), p. 1562-1567. DOI : 10.1103/PhysRevLett.81.1562.

[57] The SNO Collaboration. « Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory ». *Physical Review Letters* **89** (1 juin 2002). DOI : 10.1103/PhysRevLett.89.011301.

[58] Z. Maki, M. Nakagawa & S. Sakata. « Remarks on the Unified Model of Elementary Particles ». *Progress of Theoretical Physics* **28**.5 (nov. 1962), p. 870-880. DOI : 10.1143/PTP.28.870.

[59] J. Schechter & J. W. F. Valle. « Neutrino masses in $SU(2) \times U(1)$ theories ». *Physical Review* **D22** (9 nov. 1980), p. 2227-2235. DOI : 10.1103/PhysRevD.22.2227.

[60] A. Davidson & K. C. Wali. « Family mass hierarchy from universal seesaw mechanism ». *Physical Review Letters* **60** (18 mai 1988), p. 1813-1816. DOI : 10.1103/PhysRevLett.60.1813.

[61] R. N. Mohapatra & G. Senjanović. « Neutrino masses and mixings in gauge models with spontaneous parity violation ». *Physical Review* **D23** (jan. 1981), p. 165-180. DOI : 10.1103/PhysRevD.23.165.

[62] R. N. Mohapatra & G. Senjanović. « Neutrino Mass and Spontaneous Parity Nonconservation ». *Physical Review Letters* **44** (14 avr. 1980), p. 912-915. DOI : 10.1103/PhysRevLett.44.912.

[63] M. Gell-Mann, P. Ramond & R. Slansky. « Complex Spinors and Unified Theories » (1979). arXiv : 1306.4669 [hep-th]. URL : http://cds.cern.ch/record/133618.

[64] E. Majorana. « Teoria simmetrica dell'elettrone e del positrone ». *Il Nuovo Cimento* **14**.171 (1937). DOI : 10.1007/BF02961314.

[65] E. Majorana & L. Maiani. « A symmetric theory of electrons and positrons ». *Ettore Majorana Scientific Papers* (2006). DOI : 10.1007/978-3-540-48095-2_10.

[66] W. Rodejohann. « Neutrino-less double beta decay and particle physics ». *International Journal of Modern Physics* **E20**.9 (2011). DOI : 10.1142/S0218301311020186.

[67] J. Schechter & J. W. F. Valle. « Neutrinoless double-β decay in $SU(2) \times U(1)$ theories ». *Physical Review* **D25** (11 juin 1982), p. 2951-2954. DOI : 10.1103/PhysRevD.25.2951.

[68] Y. Nagashima. *Beyond the Standard Model of Elementary Particle Physics*. Weinheim : Wiley-VCH, juin 2014. URL : http://cds.cern.ch/record/1620277.

[69] D. Clowe & coll. « A Direct Empirical Proof of the Existence of Dark Matter ». *Astrophysical Journal* **648**.2 (août 2006). DOI : 10.1086/508162.

[70] P. A. R. Ade & coll. « Planck 2013 results. I. Overview of products and scientific results ». *Astronomy & Astrophysics* **571** (oct. 2014). DOI : 10.1051/0004-6361/201321529.

[71] Y. A. Gol'fand & E. P. Likhtman. « Extension of the algebra of Poincare group generators and violation of P invariance ». *Journal of Experimental and Theoretical Physics Letters* **13**.8 (1971), p. 323-325. URL : http://cds.cern.ch/record/433516.

[72] J. Wess & B. Zumino. « Supergauge transformations in four dimensions ». *Nuclear Physics* **B70**.1 (1974), p. 39-50. DOI : 10.1016/0550-3213(74)90355-1.

[73] S. P. Martin. « A Supersymmetry primer ». *Advanced Series on Directions in High Energy Physics* (juil. 1998), p. 1-98. DOI : 10.1142/9789812839657_0001.

[74] J. F. Gunion & coll. *The Higgs hunter's guide*. T. **80**. Upton, NY : Brookhaven Nat. Lab., 1989. URL : https://cds.cern.ch/record/425736.

[75] J. F. Gunion & coll. *Errata for "The Higgs Hunter's Guide"*. Rapp. tech. hep-ph/9302272. Fév. 1993. URL : https://cds.cern.ch/record/559892.

[76] P. Fayet. « Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino ». *Nuclear Physics* **B90** (1975), p. 104-124. DOI : 10.1016/0550-3213(75)90636-7.

[77] P. Fayet. « Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions ». *Physics Letters* **B69**.4 (1977), p. 489-494. DOI : 10.1016/0370-2693(77)90852-8.

[78] M. Carena & coll. « MSSM Higgs boson searches at the LHC : benchmark scenarios after the discovery of a Higgs-like particle ». *European Physical Journal* **C73**.9 (sept. 2013). doi : 10.1140/epjc/s10052-013-2552-1.

[79] E. Bagnaschi & coll. « MSSM Higgs boson searches at the LHC : benchmark scenarios for Run 2 and beyond ». *The European Physical Journal* **C79**.7 (juil. 2019). doi : 10.1140/epjc/s10052-019-7114-8.

[80] The CMS Collaboration. « Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions ». *Journal of High Energy Physics* **10** (oct. 2014). doi : 10.1007/jhep10(2014)160.

[81] The CMS Collaboration. « Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ ». *Journal of High Energy Physics* **09**.007 (sept. 2018). doi : 10.1007/JHEP09(2018)007.

[82] G. Ridolfi, G. Ross & F. Zwirner. « Supersymmetry ». *Large Hadron Collider Workshop Proceedings*. **II**. CERN. Geneva : CERN, oct. 1990, p. 606-683.

[83] LHC Higgs Cross Section Working Group. « Inclusive Observables ». *Handbook of LHC Higgs Cross Sections*. **1**. CERN Yellow Reports : Monographs. Geneva : CERN, 2011. url : https://cds.cern.ch/record/1318996.

[84] LHC Higgs Cross Section Working Group. « Differential Distributions ». *Handbook of LHC Higgs Cross Sections*. **2**. CERN Yellow Reports : Monographs. Geneva : CERN, 2012. url : https://cds.cern.ch/record/1416519.

[85] LHC Higgs Cross Section Working Group. « Higgs Properties ». *Handbook of LHC Higgs Cross Sections*. **3**. CERN Yellow Reports : Monographs. Geneva : CERN, 2013. url : https://cds.cern.ch/record/1559921.

[86] LHC Higgs Cross Section Working Group. « Deciphering the Nature of the Higgs Sector ». *Handbook of LHC Higgs Cross Sections*. **4**. CERN Yellow Reports : Monographs. Geneva : CERN, oct. 2016. url : http://cds.cern.ch/record/2227475.

[87] P. Bolzoni & coll. « Vector boson fusion at next-to-next-to-leading order in QCD : Standard model Higgs boson and beyond ». *Physical Review* **D85** (3 fév. 2012). doi : 10.1103/PhysRevD.85.035002.

[88] S. Dawson, A. Djouadi & M. Spira. « QCD Corrections to Supersymmetric Higgs Boson Production : The Role of Squark Loops ». *Physical Review Letters* **77**.1 (juil. 1996), p. 16-19. doi : 10.1103/physrevlett.77.16.

[89] S. H. Neddermeyer & C. D. Anderson. « Note on the Nature of Cosmic-Ray Particles ». *Physical Review* **51** (10 mai 1937), p. 884-886. doi : 10.1103/PhysRev.51.884.

[90] O. S. Brüning & coll. « The LHC Main Ring ». *LHC Design Report*. **1**. CERN Yellow Reports : Monographs. Geneva : CERN, 2004. url : https://cds.cern.ch/record/782076.

[91] O. S. Brüning & coll. « The LHC Infrastructure and General Services ». *LHC Design Report*. **2**. CERN Yellow Reports : Monographs. Geneva : CERN, 2004. url : https://cds.cern.ch/record/815187.

[92] M. Benedikt & coll. « The LHC Injector Chain ». *LHC Design Report*. **3**. CERN Yellow Reports : Monographs. Geneva : CERN, 2004. url : https://cds.cern.ch/record/823808.

[93] *Site internet du CERN*. url : https://home.cern/.

[94] F. J. Hasert & coll. « Search for elastic muon neutrino electron scattering ». *Physics Letters* **B46**.1 (1973), p. 121-124. doi : 10.1016/0370-2693(73)90494-2. url : http://cds.cern.ch/record/243640.

[95] F. Hasert & coll. « Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment ». *Physics Letters* **B46**.1 (1973), p. 138-140. doi : 10.1016/0370-2693(73)90499-1.

[96] F. J. Hasert & coll. « Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment ». *Nuclear Physics* **B73**.1 (1974), p. 1-22. URL : http://cds.cern.ch/record/203096.

[97] CERN. *The World Wide Web Project*. 1989. URL : http://info.cern.ch/hypertext/WWW/TheProject.html.

[98] CERN. *The first touchscreen used at CERN*. URL : https://www.youtube.com/watch?v=tQe5dlzScwU.

[99] P. J. Dornan. « The ALEPH Experiment ». *Philosophical Transactions : Physical Sciences and Engineering* **336**.1642 (1991), p. 201-211. URL : http://www.jstor.org/stable/53783.

[100] P. S. L. Booth. « The DELPHI Experiment ». *Philosophical Transactions : Physical Sciences and Engineering* **336**.1642 (1991), p. 213-222. URL : http://www.jstor.org/stable/53784.

[101] D. P. Stickland. « The L3 Experiment ». *Philosophical Transactions : Physical Sciences and Engineering* **336**.1642 (1991), p. 223-236. URL : http://www.jstor.org/stable/53785.

[102] A. Michelini. « OPAL Detector Performance ». *Philosophical Transactions : Physical Sciences and Engineering* **336**.1642 (1991), p. 237-246. URL : http://www.jstor.org/stable/53786.

[103] CERN. *MapCERN*. URL : https://maps.web.cern.ch/.

[104] *Long term LHC schedule*. URL : https://lhc-commissioning.web.cern.ch/.

[105] J. Gao, L. Harland-Lang & J. Rojo. « The structure of the proton in the LHC precision era ». *Physics Reports* **742** (mai 2018). DOI : 10.1016/j.physrep.2018.03.002.

[106] R. D. Ball & coll. « Parton distributions for the LHC Run II ». *Journal of High Energy Physics* **4** (avr. 2015). DOI : 10.1007/jhep04(2015)040.

[107] J. Rojo & coll. « The PDF4LHC report on PDFs and LHC data : results from Run I and preparation for Run II ». *Journal of Physics* **G42**.10 (sept. 2015). DOI : 10.1088/0954-3899/42/10/103103.

[108] The CMS Collaboration. *CMS Luminosity Measurements for the 2016 Data Taking Period*. Rapp. tech. CMS-PAS-LUM-17-001. Geneva : CERN, 2017. URL : http://cds.cern.ch/record/2257069.

[109] The CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period at* $\sqrt{s} =$ 13 TeV. Rapp. tech. CMS-PAS-LUM-17-004. Geneva : CERN, 2018. URL : http://cds.cern.ch/record/2621960.

[110] The CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at* $\sqrt{s} =$ 13 TeV. Rapp. tech. CMS-PAS-LUM-18-002. Geneva : CERN, 2019. URL : http://cds.cern.ch/record/2676164.

[111] The CMS Collaboration. *CMS Luminosity – Public Results*. URL : https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[112] The ALICE Collaboration. « The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment ». *Journal of Instrumentation* **3**.S08002 (2008). URL : http://cds.cern.ch/record/1129812.

[113] The ATLAS Collaboration. « The ATLAS Experiment at the CERN Large Hadron Collider ». *Journal of Instrumentation* **3**.S08003 (2008). URL : http://cds.cern.ch/record/1129811.

[114] The CMS Collaboration. « The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment ». *Journal of Instrumentation* **3**.S08004 (2008). URL : http://cds.cern.ch/record/1129810.

[115] The LHCb Collaboration. « The LHCb Detector at the LHC ». *Journal of Instrumentation* **3**.S08005 (2008). URL : http://cds.cern.ch/record/1129809.

[116] The LHCf Collaboration. « The LHCf detector at the CERN Large Hadron Collider ». *Journal of Instrumentation* **3**.S08006 (2008). URL : http://cds.cern.ch/record/1129808.

[117] The TOTEM Collaboration. « The TOTEM Experiment at the CERN Large Hadron Collider ». *Journal of Instrumentation* **3**.S08007 (2008). URL : http://cds.cern.ch/record/1129807.

[118] The MoEDAL Collaboration. *Technical Design Report of the MoEDAL Experiment*. Rapp. tech. CERN-LHCC-2009-006. MoEDAL-TDR-001. Juin 2009. URL : https://cds.cern.ch/record/1181486.

[119] The CMS Collaboration. *CMS : letter of intent by the CMS Collaboration for a general purpose detector at LHC*. Rapp. tech. CERN-LHCC-92-003. LHCC-I-1. Geneva : CERN, 1992. URL : https://cds.cern.ch/record/290808.

[120] T. Sakuma. *3D SketchUp images of the CMS detector*. CMS Document Database. Sept. 2018. URL : https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=13631.

[121] The CMS Collaboration. *The CMS tracker system project : Technical Design Report*. Technical Design Report CMS. Geneva : CERN, 1997. URL : https://cds.cern.ch/record/368412.

[122] The CMS Collaboration. *The CMS electromagnetic calorimeter project : Technical Design Report*. Technical Design Report CMS. Geneva : CERN, 1997. URL : https://cds.cern.ch/record/349375.

[123] The CMS Collaboration. *The CMS hadron calorimeter project : Technical Design Report*. Technical Design Report CMS. Geneva : CERN, 1997. URL : https://cds.cern.ch/record/357153.

[124] The CMS Collaboration. *The CMS magnet project : Technical Design Report*. Technical Design Report CMS. Geneva : CERN, 1997. URL : http://cds.cern.ch/record/331056.

[125] The CMS Collaboration. *The CMS muon project : Technical Design Report*. Technical Design Report CMS. Geneva : CERN, 1997. URL : https://cds.cern.ch/record/343814.

[126] P. Arce & coll. « The network of photodetectors and diode lasers of the CMS Link alignment system ». *Nuclear Instruments and Methods in Physics Research* **A896** (2018), p. 1-23. URL : http://cds.cern.ch/record/2637152.

[127] The CMS Collaboration. « Detector Performance and Software ». *CMS Physics : Technical Design Report*. **1**. Technical Design Report CMS. Geneva : CERN, 2006. URL : http://cds.cern.ch/record/922757.

[128] The CMS Collaboration. « The CMS detector magnet ». *IEEE Transactions on Applied Superconductivity* **10**.1 (2000). URL : http://cds.cern.ch/record/438917.

[129] The CMS Collaboration. « Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays ». *Journal of Instrumentation* **5** (mar. 2010). DOI : 10.1088/1748-0221/5/03/t03021.

[130] A. Dominguez & coll. *CMS Technical Design Report for the Pixel Detector Upgrade*. Rapp. tech. Sept. 2012. URL : https://cds.cern.ch/record/1481838.

[131] The CMS Collaboration. « Description and performance of track and primary-vertex reconstruction with the CMS tracker ». *Journal of Instrumentation* **9** (mai 2014). URL : http://cds.cern.ch/record/1704291.

[132] The CMS Collaboration. « Precision measurement of the structure of the CMS inner tracking system using nuclear interactions ». *Journal of Instrumentation* **13** (juil. 2018). URL : https://cds.cern.ch/record/2629890.

[133] The CMS Collaboration. *The Phase-1 Upgrade of the CMS Pixel Detector*. Rapp. tech. CMS-CR-2017-135. 06. Geneva : CERN, mai 2017. URL : https://cds.cern.ch/record/2265423.

[134] The CMS Collaboration. « Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ ». *Journal of Instrumentation* **8** (juin 2013). URL : https://cds.cern.ch/record/1554142.

[135] The CMS Collaboration. « CMS ECAL Response to Laser Light » (mar. 2019). URL : https://cds.cern.ch/record/2668200.

[136] The CMS Collaboration. « ECAL 2016 refined calibration and Run2 summary plots » (avr. 2020). URL : https://cds.cern.ch/record/2717925.

[137] J. Mans & coll. *CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter*. Rapp. tech. Sept. 2012. URL : https://cds.cern.ch/record/1481837.

[138] The CMS Collaboration. « HCAL Energy Reconstruction Performance » (nov. 2016). URL : https://cds.cern.ch/record/2235509.

[139] The CMS Collaboration. « HCAL Out Of Time Pileup Subtraction and Energy Reconstruction » (mai 2018). URL : https://cds.cern.ch/record/2320408.

[140] The CMS Collaboration. « HCAL Calibration Status in Summer 2017 » (mai 2017). URL : https://cds.cern.ch/record/2281146.

[141] The CMS Collaboration. « The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC ». *Journal of Instrumentation* **8** (juin 2013). URL : https://cds.cern.ch/record/1558674.

[142] The CMS Collaboration. « Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV ». *Journal of Instrumentation* **13** (avr. 2018). URL : https://cds.cern.ch/record/2313130.

[143] The CMS Collaboration. *CMS TriDAS project : Technical Design Report*. T. **1**. Technical Design Report CMS. Geneva : CERN. URL : http://cds.cern.ch/record/706847.

[144] The CMS Collaboration. *CMS TriDAS project : Technical Design Report*. T. **2**. Technical Design Report CMS. Geneva : CERN, 2002. URL : http://cds.cern.ch/record/578006.

[145] The CMS Collaboration. « The CMS trigger system ». *Journal of Instrumentation* **12**.1 (jan. 2017). URL : https://cds.cern.ch/record/2212926.

[146] A. Tapper & D. Acosta. *CMS Technical Design Report for the Level-1 Trigger Upgrade*. Rapp. tech. Juin 2013. URL : https://cds.cern.ch/record/1556311.

[147] The CMS Collaboration. « Particle-flow reconstruction and global event description with the CMS detector ». *Journal of Instrumentation* **12**.10 (juin 2017). DOI : 10.1088/1748-0221/12/10/P10003. arXiv : 1706.04965v2 [physics.ins-det].

[148] The CMS Collaboration. « The CMS Particle Flow Algorithm ». *EPJ Web of Conferences* **191** (2018). URL : https://cds.cern.ch/record/2678077.

[149] T. Speer & coll. *Track Reconstruction in the CMS Tracker*. Rapp. tech. CMS-CR-2005-014. Geneva : CERN, juil. 2005. URL : http://cds.cern.ch/record/884424.

[150] R. Frühwirth. « Application of Kalman filtering to track and vertex fitting ». *Nuclear Instruments and Methods in Physics Research* **A262**.2 (1987), p. 444-450. DOI : 10.1016/0168-9002(87)90887-4.

[151] The CMS Collaboration. « Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV ». *Journal of Instrumentation* **10**.06 (juin 2015). DOI : 10.1088/1748-0221/10/06/p06005.

[152] D. Alspach & H. Sorenson. « Nonlinear Bayesian estimation using Gaussian sum approximations ». *IEEE Transactions on Automatic Control* **17**.4 (1972), p. 439-448. DOI : 10.1109/TAC.1972.1100034.

[153] H. A. Bethe. « Molière's Theory of Multiple Scattering ». *Physical Review* **89** (6 mar. 1953). DOI : 10.1103/PhysRev.89.1256.

[154] B. P. Nigam, M. K. Sundaresan & T.-Y. Wu. « Theory of Multiple Scattering : Second Born Approximation and Corrections to Molière's Work ». *Physical Review* **115** (3 août 1959). DOI : 10.1103/PhysRev.115.491.

[155] The CMS Collaboration. *Studies of Tracker Material*. Rapp. tech. CMS-PAS-TRK-10-003. 2010. URL : https://cds.cern.ch/record/1279138.

[156] The CMS Collaboration. « Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV ». *Journal of Instrumentation* **12**.02 (fév. 2017). DOI : 10.1088/1748-0221/12/02/p02014.

[157] J. L. Bentley. « Multidimensional Binary Search Trees Used for Associative Searching ». *Communications of the ACM* **18**.9 (sept. 1975), p. 509-517. doi : 10.1145/361002.361007.

[158] The CMS Collaboration. *Baseline muon selections for Run-II*. url : https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2.

[159] The CMS Collaboration. *Multivariate Electron Identification for Run2*. url : https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2.

[160] T. Chen & C. Guestrin. « XGBoost : A Scalable Tree Boosting System ». *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (août 2016). doi : 10.1145/2939672.2939785.

[161] J. Alwall & coll. « MadGraph 5 : Going Beyond ». *Journal of High Energy Physics* **06** (2011). doi : 10.1007/jhep06(2011)128. arXiv : 1106.0522 [hep-ph].

[162] M. Cacciari, G. P. Salam & G. Soyez. « The Anti-k_T jet clustering algorithm ». *Journal of High Energy Physics* **04** (avr. 2008). doi : 10.1088/1126-6708/2008/04/063. arXiv : 0802.1189 [hep-ph].

[163] M. Cacciari, G. P. Salam & G. Soyez. « fastjet user manual ». *European Physical Journal* **C72** (nov. 2012). doi : 10.1140/epjc/s10052-012-1896-2. arXiv : 1111.6097 [hep-ph].

[164] The CMS Collaboration. « Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018 » (avr. 2020). url : https://cds.cern.ch/record/2715872.

[165] The CMS Collaboration. « Identification of b-quark jets with the CMS experiment ». *Journal of Instrumentation* **8**.04 (avr. 2013). doi : 10.1088/1748-0221/8/04/p04013.

[166] D. Guest & coll. « Jet flavor classification in high-energy physics with deep neural networks ». *Physical Review* **D94**.11 (déc. 2016). doi : 10.1103/physrevd.94.112002.

[167] The CMS Collaboration. « Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV ». *Journal of Instrumentation* **13**.05 (mai 2018). doi : 10.1088/1748-0221/13/05/p05011.

[168] F. Chollet & coll. Keras. 2015. url : https://keras.io.

[169] M. Abadi & coll. TensorFlow *: Large-scale machine learning on heterogeneous distributed systems*. Software available from tensorflow.org. 2015. url : https://www.tensorflow.org/.

[170] The CMS Collaboration. *Performance of missing transverse momentum in pp collisions at* $\sqrt{s} =$ 13 TeV *using the CMS detector*. Rapp. tech. Geneva : CERN, 2018. url : https://cds.cern.ch/record/2628600.

[171] D. Bertolini & coll. « Pileup per particle identification ». *Journal of High Energy Physics* **10** (oct. 2014). doi : 10.1007/jhep10(2014)059.

[172] The CMS Collaboration. *MET Analysis*. url : https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis.

[173] The CMS Collaboration. « Reconstruction and identification of tau lepton decays to hadrons and tau neutrino at CMS ». *Journal of Instrumentation* **11**.1 (2016). doi : 10.1088/1748-0221/11/01/P01019. arXiv : 1510.07488 [physics.ins-det].

[174] The CMS Collaboration. « Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV ». *Journal of Instrumentation* **13**.10 (2018). doi : 10.1088/1748-0221/13/10/P10005. arXiv : 1809.02816 [hep-ex].

[175] I. Goodfellow, Y. Bengio & A. Courville. *Deep Learning*. MIT Press, 2016. url : http://www.deeplearningbook.org.

[176] The CMS Collaboration. « Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons » (oct. 2019). url : https://cds.cern.ch/record/2694158.

[177] K. Androsov. « DeepTau ID ». *Meeting calcolo CMS Italia*. Juin 2019. url : https://indico.cern.ch/event/819693/contributions/3438506/attachments/1858497/3053529/2019-06-07_DeepTau_ID.pdf.

[178] T. Sjöstrand, S. Mrenna & P. Skands. « pythia 6.4 physics and manual ». *Journal of High Energy Physics* **05** (mai 2006). doi : 10.1088/1126-6708/2006/05/026.

[179] T. Sjöstrand & coll. « An Introduction to pythia 8.2 ». *Computer Physics Communications* **191** (2015), p. 159-177. doi : 10.1016/j.cpc.2015.01.024. arXiv : 1410.3012 [hep-ph].

[180] S. Alioli & coll. « A general framework for implementing NLO calculations in shower Monte Carlo programs : the powheg BOX ». *Journal of High Energy Physics* **06** (2010). doi : 10.1007/jhep06(2010)043. arXiv : 1002.2581 [hep-ph].

[181] S. Frixione, P. Nason & B. R. Webber. « Matching NLO QCD and parton showers in heavy flavour production ». *Journal of High Energy Physics* **8** (août 2003). doi : 10.1088/1126-6708/2003/08/007.

[182] J. Bellm & coll. « herwig 7.0/herwig++ 3.0 release note ». *European Physical Journal* **C76**.196 (avr. 2016). doi : 10.1140/epjc/s10052-016-4018-8.

[183] A. Schälicke & coll. « An event generator for particle production in high-energy collisions ». *Progress in Particle and Nuclear Physics* **53**.1 (2004), p. 329-338. doi : 10.1016/j.ppnp.2004.02.031. url : http://www.sciencedirect.com/science/article/pii/S014664100400047X.

[184] S. Agostinelli & coll. « Geant4 – A simulation toolkit ». *Nuclear Instruments and Methods in Physics Research* **A506**.3 (2003), p. 250-303. doi : 10.1016/S0168-9002(03)01368-8.

[185] J. Allison & coll. « Geant4 developments and applications ». *IEEE Transactions on Nuclear Science* **53**.1 (fév. 2006), p. 270-278. doi : 10.1109/tns.2006.869826.

[186] J. Allison & coll. « Recent developments in Geant4 ». *Nuclear Instruments and Methods in Physics Research* **A835** (2016), p. 186-225. doi : 10.1016/j.nima.2016.06.125.

[187] The CMS Collaboration. *E/gamma Physics Object Group*. url : https://twiki.cern.ch/twiki/bin/view/CMS/EgammaPOG.

[188] The CMS Collaboration. *Recommendation for Using b-tag Objects in Physics Analyses*. url : https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation.

[189] The CMS Collaboration. *Tau ID recommendations for Run-2*. url : https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauIDRecommendationForRun2.

[190] A. Raspereza. *Recoil Corrections for the 2017 Dataset*. Oct. 2018. url : https://indico.cern.ch/event/762837/contributions/3172618/attachments/1731302/2798220/Recoils_20181010.pdf.

[191] The CMS Collaboration. « First results from the CMS SiPM-based hadronic endcap calorimeter ». Geneva, août 2018. url : https://cds.cern.ch/record/2636475.

[192] The CMS Collaboration. *Pileup Removal Algorithms*. Rapp. tech. CMS-PAS-JME-14-001. Geneva : CERN, 2014. url : https://cds.cern.ch/record/1751454.

[193] The CMS Collaboration. « Event generator tunes obtained from underlying event and multiparton scattering measurements ». *European Physical Journal* **C76**.3 (2016). doi : 10.1140/epjc/s10052-016-3988-x. arXiv : 1512.00815 [hep-ex].

[194] H. Lattaud. « Photon et jets avec l'expérience CMS du LHC : de la calibration à la mesure ». Thèse de doct. Université Claude Bernard Lyon 1, sept. 2019. url : https://tel.archives-ouvertes.fr/tel-02422226.

[195] The CMS Collaboration. « Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ ». *Journal of Instrumentation* **10** (fév. 2015). url : https://cds.cern.ch/record/1988093.

[196] The CMS Collaboration. « Measurement of differential cross sections for inclusive isolated-photon and photon+jets production in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ ». *European Physical Journal* **C79**.20 (juil. 2018). url : http://cds.cern.ch/record/2628267.

[197] The CMS Collaboration. « Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements ». *European Physical Journal* **C80** (mar. 2019). URL : https://cds.cern.ch/record/2669320.

[198] DELPHI, OPAL, ALEPH, LEP Working Group for Higgs Boson Searches, L3. « Search for neutral MSSM Higgs bosons at LEP ». *European Physical Journal* **C47** (2006), p. 547-587. DOI : 10.1140/epjc/s2006-02569-7. arXiv : hep-ex/0602042 [hep-ex].

[199] The CDF Collaboration. « Search for Higgs bosons predicted in two-Higgs-doublet models via decays to tau lepton pairs in 1,96 TeV $p\bar{p}$ collisions ». *Physical Review Letters* **103** (2009). DOI : 10.1103/PhysRevLett.103.201801. arXiv : 0906.1014 [hep-ex].

[200] The DØ Collaboration. « Search for Higgs bosons decaying to $\tau\tau$ pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1{,}96$ TeV ». *Physics Letters* **B707** (2012), p. 323-329. DOI : 10.1016/j.physletb.2011.12.050. arXiv : 1106.4555 [hep-ex].

[201] The CMS Collaboration. « Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV ». *Physics Letters* **B713** (2012), p. 68-90. DOI : 10.1016/j.physletb.2012.05.028. arXiv : 1202.4083 [hep-ex].

[202] The CMS Collaboration. « Higgs to tau tau (MSSM) ». CMS-PAS-HIG-13-021 (2013). URL : http://cds.cern.ch/record/1623367.

[203] The CMS Collaboration. « Search for additional neutral Higgs bosons decaying to a pair of tau leptons in pp collisions at $\sqrt{s} = 7$ and 8 TeV ». CMS-PAS-HIG-14-029 (2015). URL : https://cds.cern.ch/record/2041463.

[204] The CMS Collaboration. « Search for neutral MSSM Higgs bosons decaying into a pair of bottom quarks ». *Journal of High Energy Physics* **11** (2015). DOI : 10.1007/JHEP11(2015)071. arXiv : 1506.08329 [hep-ex].

[205] The CMS Collaboration. « Search for neutral MSSM Higgs bosons decaying to $\mu^+\mu^-$ in pp collisions at $\sqrt{s} = 7$ and 8 TeV ». *Physics Letters* **B752** (2016), p. 221-246. DOI : 10.1016/j.physletb.2015.11.042. arXiv : 1508.01437 [hep-ex].

[206] The ATLAS Collaboration. « Search for the neutral Higgs bosons of the Minimal Supersymmetric Standard Model in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector ». *Journal of High Energy Physics* **02** (2013). DOI : 10.1007/JHEP02(2013)095. arXiv : 1211.6956 [hep-ex].

[207] The ATLAS Collaboration. « Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector ». *Journal of High Energy Physics* **1** (jan. 2018). DOI : 10.1007/jhep01(2018)055.

[208] The ATLAS Collaboration. « Search for Heavy Higgs Bosons Decaying into Two Tau Leptons with the ATLAS Detector Using pp Collisions at $\sqrt{s} = 13$ TeV ». *Physical Review Letters* **125** (5 juil. 2020), p. 051801. DOI : 10.1103/PhysRevLett.125.051801.

[209] J. ANDREJKOVIC & coll. « BSM $H \to \tau\tau$ analysis on full Run 2 CMS data at $\sqrt{s} = 13$ TeV ». *CMS analysis Note* (2021). URL : https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2020/218.

[210] The CMS Collaboration. « MSSM $H/A \to \tau\tau$ search with full Run-2 data ». 2021. URL : https://cms.cern.ch/iCMS/analysisadmin/cadilines?line=HIG-21-001&tp=an&id=2409&ancode=HIG-21-001.

[211] G. TOUQUET. « Search for an additional neutral MSSM Higgs boson decaying to tau leptons with the CMS experiment ». Thèse de doct. Université Claude Bernard Lyon 1, oct. 2019. URL : https://hal.archives-ouvertes.fr/tel-02526393.

[212] L. TORTEROTOT. « Search for additional neutral Higgs bosons decaying to τ leptons pairs in the CMS experiment at the LHC ». *JRJC 2019. Book of Proceedings*. Août 2020, p. 53-56. URL : https://hal.archives-ouvertes.fr/hal-02971995.

[213] C. BERNET. *Heppy : a python framework for high-energy physics data analysis*. URL : https://github.com/cbernet/heppy.

[214] L. Moneta & coll. « The RooStats Project ». *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. 2010. URL : http://pos.sissa.it/archive/conferences/093/057/ACAT2010_057.pdf.

[215] L. Torterotot, C. Bernet & E. Aşılar. *MergeBinErrors fix on total bin error*. Fév. 2020. URL : https://github.com/cms-analysis/CombineHarvester/pull/243.

[216] D. Winterbottom, A. Dow & M. H. Hassanshahi. « Electron and muon scale factor measurements for Higgs to tau tau analyses ». *CMS analysis Note* (2020). URL : https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2020/014.

[217] T. Lenz, M. Meyer & A. Raspereza. « Tau ID scale factors for the DeepTau ID for full Run II data using high-p_T taus ». *CMS analysis Note* (nov. 2019). URL : https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2019/263.

[218] E. Bagnaschi & A. Vicini. « The Higgs transverse momentum distribution in gluon fusion as a multiscale problem ». *Journal of High Energy Physics* **01** (2016). DOI : 10.1007/JHEP01(2016)056.

[219] E. Bagnaschi & coll. « Resummation ambiguities in the Higgs transverse-momentum spectrum in the Standard Model and beyond ». *Journal of High Energy Physics* **01** (2016). DOI : 10.1007/JHEP01(2016)090.

[220] D. Jang. « Search for MSSM Higgs decaying to tau pairs in $p\bar{p}$ collision at $\sqrt{s} = 1{,}96\,\text{TeV}$ at CDF ». Thèse de doct. Rutgers, The State University of New Jersey, mai 2006. URL : https://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2006-11.pdf.

[221] The CMS Collaboration. *DeepJet : deep learning based on physics objects for jet reconstruction*. URL : https://twiki.cern.ch/twiki/bin/viewauth/CMS/DeepFlavour.

[222] The CMS Collaboration. « An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data ». *Journal of Instrumentation* **14**.06 (juin 2019). DOI : 10.1088/1748-0221/14/06/p06032.

[223] The CMS Collaboration. *Measurement of Higgs boson production and decay to the $\tau\tau$ final state*. Rapp. tech. CMS-PAS-HIG-18-032. Geneva : CERN, 2019. URL : https://cds.cern.ch/record/2668685.

[224] J. Andrejkovic & coll. « Data-driven background estimation of fake-tau backgrounds in di-tau final states with 2016 and 2017 data ». *CMS analysis Note* (oct. 2018). URL : https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2018/257.

[225] J. Andrejkovic & J. Bechtel. « Data-driven background estimation of fake-tau backgrounds in di-tau final states with the full Run-II dataset ». *CMS analysis Note* (juin 2020). URL : https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2019/170.

[226] The CMS Collaboration. *Measurement of Higgs boson production in the decay channel with a pair of τ leptons*. Rapp. tech. CMS-PAS-HIG-19-010. Geneva : CERN, 2020. URL : http://cds.cern.ch/record/2725590.

[227] J. Andrejkovic & coll. « Measurement of Higgs(125) boson properties in decays to a pair of tau leptons with full Run II data using Machine-Learning techniques ». *CMS analysis Note* (sept. 2020). URL : https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2019/177.

[228] J. Andrejkovic & coll. « Multi-class neural network architecture and training for measurements of Higgs(125) boson decays to two tau leptons on full Run II data ». *CMS analysis Note* (mai 2020). URL : https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2019/178.

[229] L. Bianchini & coll. « Reconstruction of the Higgs mass in $H \to \tau\tau$ Events by Dynamical Likelihood techniques ». *Journal of Physics : Conference Series* **513**.2 (juin 2014). DOI : 10.1088/1742-6596/513/2/022035.

[230] A. Gottmann. « Global Interpretation of $\tau\tau$ Events in the Context of the Standard Model and Beyond ». Thèse de doct. Fakultät für Physik des Karlsruher Instituts für Technologie (KIT), juin 2020. URL : https://publish.etp.kit.edu/record/22014.

[231] The CMS Collaboration. *Luminosity Physics Object Group (Lumi POG)*. URL : https://twiki.cern.ch/twiki/bin/viewauth/CMS/TWikiLUM.

[232] The CMS Collaboration. *Precision luminosity measurement in proton-proton collisions at* $\sqrt{s} =$ 13 TeV *in 2015 and 2016 at CMS*. Rapp. tech. CMS-PAS-LUM-17-003. Geneva : CERN, 2021. URL : http://cds.cern.ch/record/2759951.

[233] The CMS Collaboration. *Standard Model Cross Sections for CMS at 13 TeV*. URL : https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat13TeV.

[234] The CMS Collaboration. *MSSM Neutral Higgs*. URL : https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGMSSMNeutral.

[235] The CMS Collaboration. *MET Uncertainties*. URL : https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETRun2Corrections#MET_Uncertainties.

[236] The CMS Collaboration. *Jet Energy Resolution*. URL : https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution.

[237] T. Junk. « Confidence level computation for combining searches with small statistics ». *Nuclear Instruments and Methods in Physics Research* **A434**.2–3 (sept. 1999), p. 435-443. DOI : 10.1016/S0168-9002(99)00498-2. arXiv : hep-ex/9902006 [hep-ex].

[238] A. L. Read. « Modified frequentist analysis of search results (the CL_s method) ». *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000 : Proceedings*. CERN-OPEN-2000-205. Mai 2000. URL : http://cds.cern.ch/record/451614.

[239] A. L. Read. « Presentation of search results : The CL(s) technique ». *Journal of Physics* **G28**.10 (sept. 2002), p. 2693-2704. DOI : 10.1088/0954-3899/28/10/313.

[240] R. Barlow & C. Beeston. « Fitting using finite Monte Carlo samples ». *Computer Physics Communications* **77**.2 (1993), p. 219-228. DOI : 10.1016/0010-4655(93)90005-W.

[241] J. S. Conway. « Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra » (2011), p. 115-120. DOI : 10.5170/CERN-2011-006.115.

[242] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. Rapp. tech. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva : CERN, août 2011. URL : https://cds.cern.ch/record/1379837.

[243] G. Bourgatte. « Étude des propriétés *CP* du boson de Higgs dans le canal $\tau\tau$ dans l'expérience CMS auprès du LHC ». Thèse de doct. Université de Strasbourg, déc. 2020. URL : http://cds.cern.ch/record/2750735.

[244] S. Berge, W. Bernreuther & H. Spiesberger. « Higgs *CP* properties using the τ decay modes at the ILC ». *Physics Letters* **B727**.4 (2013), p. 488-495. DOI : 10.1016/j.physletb.2013.11.006.

[245] K. Desch & coll. « Probing the *CP* nature of the Higgs boson at linear colliders with τ spin correlations ; the case of mixed scalar–pseudoscalar couplings ». *Physics Letters* **B579** (jan. 2004), p. 157-164. DOI : 10.1016/j.physletb.2003.10.074.

[246] V. Cherepanov, E. Richter-Was & Z. Was. « Monte Carlo, fitting and Machine Learning for Tau leptons ». *SciPost Physics Proceedings* (2019), p. 18. DOI : 10.21468/SciPostPhysProc.1.018.

[247] The FCC Collaboration. *Future Circular Collider conceptual design report*. URL : https://fcc-cdr.web.cern.ch/.

[248] DeepMind. *AlphaGo*. URL : https://www.deepmind.com/research/case-studies/alphago-the-story-so-far.

[249] C. Bernet. *The Data Frog – Image Recognition : Dogs vs Cats !* URL : https://thedatafrog.com/en/articles/dogs-vs-cats/.

[250] M. Mir. *House Prices Prediction Using Deep Learning*. URL : https://towardsdatascience.com/house-prices-prediction-using-deep-learning-dea265cc3154.

[251] M. Scham. « Standard Model $H \to \tau\tau$ Analysis with a Neural Network Trained on a Mix of Simulation and Data Samples ». Mém. de mast. Fakultät für Physik des Karlsruher Instituts für Technologie (KIT), juin 2020. URL : https://publish.etp.kit.edu/record/21993.

[252] T. Kopf. « Recoil Calibration as a Neural Network Task ». Mém. de mast. Fakultät für Physik des Karlsruher Instituts für Technologie (KIT), fév. 2019. URL : https://publish.etp.kit.edu/record/21500.

[253] P. Baldi, P. Sadowski & D. Whiteson. « Enhanced Higgs Boson to $\tau^+\tau^-$ Search with Deep Learning ». *Physical Review Letters* **114**.11 (mar. 2015). DOI : 10.1103/physrevlett.114.111801.

[254] A. Elagin & coll. « A new mass reconstruction technique for resonances decaying to $\tau\tau$ ». *Nuclear Instruments and Methods in Physics Research* **A654**.1 (2011), p. 481-489. DOI : 10.1016/j.nima.2011.07.009.

[255] A. J. Barr & coll. « Speedy Higgs boson discovery in decays to tau lepton pairs : $h \to \tau\tau$ ». *Journal of High Energy Physics* **2011**.10 (oct. 2011). DOI : 10.1007/JHEP10(2011)080.

[256] B. Gripaios & coll. « Reconstruction of Higgs bosons in the di-tau channel via 3-prong decay ». *Journal of High Energy Physics* **2013**.3 (mar. 2013). DOI : 10.1007/JHEP03(2013)106.

[257] P. Bärtschi & coll. « Reconstruction of τ lepton pair invariant mass using an artificial neural network ». *Nuclear Instruments and Methods in Physics Research* **A929** (2019), p. 29-33. DOI : 10.1016/j.nima.2019.03.029.

[258] J. de Favereau & coll. « Delphes 3 : a modular framework for fast simulation of a generic collider experiment ». *Journal of High Energy Physics* **2** (fév. 2014). DOI : 10.1007/jhep02(2014)057.

[259] A. Mertens. « New features in Delphes 3 ». *Journal of Physics : Conference Series* **608**.1 (2015). Sous la dir. de L. Fiala, M. Lokajicek & N. Tumova. DOI : 10.1088/1742-6596/608/1/012045.

[260] S. Abdullin & coll. « The Fast Simulation of the CMS Detector at LHC ». *Journal of Physics : Conference Series* **331**.3 (déc. 2011). DOI : 10.1088/1742-6596/331/3/032049.

[261] A. Giammanco. « The Fast Simulation of the CMS Experiment ». *Journal of Physics : Conference Series* **513**.2 (juin 2014). DOI : 10.1088/1742-6596/513/2/022012.

[262] M. Komm. « Fast emulation of track reconstruction in the CMS simulation ». *Journal of Physics : Conference Series* **898** (oct. 2017). DOI : 10.1088/1742-6596/898/4/042034.

[263] S. Sekmen. *Recent Developments in CMS Fast Simulation*. 2017. arXiv : 1701.03850.

[264] E. Aşılar, L. Torterotot & C. Bernet. « Reconstruction of di-tau mass using deep neural networks ». *CMS analysis Note* (2021). URL : https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2021/054.

[265] E. Aşılar. *How to produce nanoAOD events of $h \to \tau\tau$ where Higgs has a* 130 GeV *mass*. URL : https://github.com/easilar/cmssw/blob/from-CMSSW_10_2_22/README.

[266] L. Torterotot, E. Aşılar & C. Bernet. *Reconstruction of di-tau mass using Machine Learning*. URL : https://github.com/lucastorterotot/DL_for_HTT_mass.

[267] L. Torterotot. *DiTau_ML_mass – Estimations of di-tau mass using Machine Learning*. URL : https://github.com/lucastorterotot/DiTau_ML_mass.

[268] A. Cauchy. « Méthode générale pour la résolution des systèmes d'équations simultannées ». *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*. **25**. 1847, p. 536-538. URL : https://gallica.bnf.fr/ark:/12148/bpt6k2982c/f3.item.

[269] L. Bottou. « Online Algorithms and Stochastic Approximations ». *Online Learning and Neural Networks*. Sous la dir. de D. Saad. Cambridge, UK : Cambridge University Press, 1998. URL : http://leon.bottou.org/papers/bottou-98x.

[270] *Kaggle Competitions*. URL : https://www.kaggle.com/competitions.

[271] W. Sarle. « Neural Networks and Statistical Models ». 1994. URL : https://people.orie.cornell.edu/davidr/or474/nn_sas.pdf.

[272] X. Glorot & Y. Bengio. « Understanding the difficulty of training deep feedforward neural networks ». *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Sous la dir. d'Y. W. Teh & M. Titterington. **9**. Proceedings of Machine Learning Research. PMLR, mai 2010, p. 249-256. URL : http://proceedings.mlr.press/v9/glorot10a.html.

[273] J. Duchi, E. Hazan & Y. Singer. « Adaptive Subgradient Methods for Online Learning and Stochastic Optimization ». *Journal of Machine Learning Research* **12**.61 (2011), p. 2121-2159. URL : http://jmlr.org/papers/v12/duchi11a.html.

[274] G. Hinton. *Neural Networks for Machine Learning*. Coursera Video Lectures, Academic Torrents. 2012. URL : https://archive.org/search.php?query=creator%3A%22Geoffrey+Hinton%22&and[]=year%3A%222012%22.

[275] D. P. Kingma & J. Ba. « Adam : A Method for Stochastic Optimization » (2017). arXiv : 1412.6980 [cs.LG].

[276] *Le système international d'unités*. 9e éd. Bureau international des poids et mesures. 2019. URL : https://www.bipm.org/utils/common/pdf/si-brochure/SI-Brochure-9.pdf.

[277] R. P. Feynman. « Space-Time Approach to Quantum Electrodynamics ». *Physical Review* **76**.6 (sept. 1949), p. 769-789. DOI : 10.1103/PhysRev.76.769.