

Chapitre X

Recherche de bosons de Higgs supplémentaires de haute masse se désintégrant en paire de leptons τ

Sommaire

1	Introduction	2
2	Corrections	3
2.1	Efficacité des chemins de déclenchement	3
2.2	Impulsions des particules générées	4
3	Sélection des événements	5
3.1	Sélection pour l'enregistrement des données	6
3.2	Obtention du dilepton et sélection finale	6
3.3	Sélection des jets	10
4	Modélisation du bruit de fond	11
4.1	Méthode des données encapsulées ou <i>embedding</i>	13
4.2	Estimation du bruit de fond QCD dans le canal $e\mu$	15
4.3	Méthode des facteurs de faux ou <i>fake factors</i>	16
5	Catégorisation des événements et variables discriminantes	21
5.1	Catégories « BSM »	21
5.2	Catégories « SM »	24
5.3	Combinaison des catégories SM et BSM	27
6	Incertitudes systématiques	28
6.1	Incertitudes de normalisation	28
6.2	Incertitudes de forme	29
7	Extraction du signal	32
7.1	Modèle de vraisemblance	32
7.2	Méthode CL_s	35
7.3	Limites indépendantes du modèle	36
7.4	Interprétation dans les scénarios du MSSM	36
8	Résultats et interprétations	38
8.1	Limites indépendantes du modèle	38
8.2	Limites du scénario M_h^{125}	39
8.3	Limites du scénario $M_{H_1}^{125}$ avec violation de CP	40
9	Conclusion	40

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 [1-5]. 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 [6]. 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 [7, 8].

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) [9-12] ainsi qu'avec les données récoltées en 2016 avec une énergie de $\sqrt{s} = 13$ TeV [13]. La désintégration en paire de b est également exploitée [14, 15], ainsi que celle en paire de muons [16]. L'expérience ATLAS mène des recherches similaires [17-19].

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 ($\tau_h \tau_h$), semi-leptoniques ($\mu \tau_h$, $e \tau_h$) 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 leurs 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 [20].

Une note d'analyse [21] interne à la collaboration CMS est déjà disponible et un article est en préparation [22]. 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 [23]. Je me suis concentré sur les canaux semi-leptoniques et plus particulièrement le canal $\mu \tau_h$ [24]. Les événements étaient analysés à l'aide d'un code basé sur HEPPY [25], 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 combinaison statistique utilisé par la collaboration CMS et basé sur ROOSTATS [26]. Cette erreur a été comprise et corrigée comme présenté dans la section 7.1.3. Le correctif [27] 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 [21] et de la publication [22] 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-rapacité du lepton par une méthode « balise et sonde » (*tag and probe*) sur des événements $Z \rightarrow \ell\ell$ [28] 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

$$\epsilon = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}} \quad (\text{X.1})$$

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 [28].

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 [29]. Les événements exploités sont ceux contenant un boson W hautement virtuel, tel que $m_{W^*} \geq 200 \text{ GeV}$, se désintégrant selon $W^* \rightarrow \tau \nu_\tau$. 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}_T^{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_T^{miss} [21, 29]. Les modélisations du signal $W^* \rightarrow \tau \nu_\tau$ 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 $W^* \rightarrow \tau\nu_\tau$. Les quantités d'événements $W^* \rightarrow \tau\nu_\tau$ 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

$$\epsilon = \frac{N_{\text{pass}}}{N_{\text{total}}} \quad (\text{X.2})$$

avec N_{pass} le nombre d'événements $W^* \rightarrow \tau\nu_\tau$ 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

$$SF = \frac{\epsilon(\text{données réelles})}{\epsilon(\text{données simulées})}. \quad (\text{X.3})$$

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 de 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

$$\epsilon = \epsilon(1\tau_h) + \epsilon(1\ell) - \epsilon(1\tau_h) \times \epsilon(1\ell) \quad (\text{X.4})$$

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 »,

$$\begin{aligned} \epsilon = & \epsilon(2\tau_h) + \epsilon(\tau_h1) + \epsilon(\tau_h2) \\ & - \epsilon(2\tau_h + \tau_h1) - \epsilon(2\tau_h + \tau_h2) - \epsilon(\tau_h1 + \tau_h2) \\ & + \epsilon(2\tau_h + \tau_h1 + \tau_h2), \end{aligned} \quad (\text{X.5})$$

où $2\tau_h$ signifie « double τ_h », τ_h1 « τ_h seul » appliqué au τ_h de plus haut p_T et τ_h2 « 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 $Z \rightarrow \mu\mu$ 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

$$\omega = \exp\left(0,088 - 8,7 \times 10^{-4} \times p_T + 9,2 \times 10^{-7} \times p_T^2\right). \quad (\text{X.6})$$

Le poids total à appliquer aux événements $t\bar{t}$ contenant deux quarks t est alors

$$\omega(\text{total}) = \sqrt{\omega(1) \times \omega(2)}. \quad (\text{X.7})$$

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 X.1.

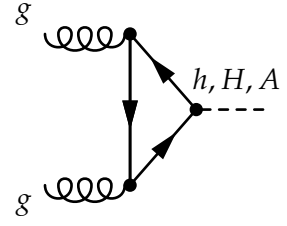


Figure X.1 – Diagramme de Feynman de production de boson de Higgs dans le cadre du MSSM par fusion de gluons (ggH).

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 [30, 31].

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 [32], les gerbes partoniques, l'hadronisation et l'événement sous-jacent sont simulés par PYTHIA [33] et la modélisation du détecteur est traitée par GEANT4 [34-36]. 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

$$\begin{aligned} \sigma_{\text{MSSM}} = & \left(\frac{y_{t,\text{MSSM}}}{y_{t,2\text{HDM}}} \right)^2 \sigma_{t,2\text{HDM}}(Q_t) + \left(\frac{y_{b,\text{MSSM}}}{y_{b,2\text{HDM}}} \right)^2 \sigma_{b,2\text{HDM}}(Q_b) \\ & + \left(\frac{y_{t,\text{MSSM}} y_{b,\text{MSSM}}}{y_{t,2\text{HDM}} y_{b,2\text{HDM}}} \right) [\sigma_{t+b,2\text{HDM}}(Q_{tb}) - \sigma_{t,2\text{HDM}}(Q_{tb}) - \sigma_{b,2\text{HDM}}(Q_{tb})] \end{aligned} \quad (\text{X.8})$$

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 [30, 31]. 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 $H \rightarrow \tau\tau$ 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 $\mu \tau_h = L_1 L_2$. 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 $L_1 L_2$, sont formés à partir de ces deux listes de candidats. Pour être compatibles avec le cas de figure $H \rightarrow \tau\tau \rightarrow L_1 L_2$, 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 $L_1 L_2$ 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) $< p_T^\mu \leq 23$ (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

$$m_T^\mu = m_T(\mu, E_T^{\text{miss}}) = \sqrt{2 p_T^\mu E_T^{\text{miss}} (1 - \cos \Delta\phi)} \quad (\text{X.9})$$

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 $W + \text{jets}$ 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 : $p_T^{\tau_h} < 120$ (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 : $p_T^{\tau_h} > 120$ (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 $p_T^{\tau_h} > 25$ (35) GeV en 2016 (2017, 2018);
- vérifier $|\eta^{\tau_h}| < 2,1$.

Masse transverse de l'électron La masse transverse de l'électron, définie par

$$m_T^e = m_T(e, E_T^{\text{miss}}) = \sqrt{2 p_T^e E_T^{\text{miss}} (1 - \cos \Delta\phi)} \quad (\text{X.10})$$

avec $\Delta\phi = \phi^e - \phi^{E_T^{\text{miss}}}$ 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 $W + \text{jets}$ 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 $p_T^\mu > 10$ GeV, $|\eta^\mu| < 2,4$, passant le point de fonctionnement *medium* du *muonID* et d'isolation $I^\mu < 0,3 p_T^\mu$;
- de second 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 d'électrons de charges opposées avec $\Delta R > 0,15$, tous deux vérifiant $p_T^e > 15$ GeV, $|\eta^e| < 2,5$, passant le *cut-based veto ID*, de paramètres d'impact $d_z < 0,2$ cm et $d_{xy} < 0,045$ cm et d'isolation $I^e < 0,3 p_T^e$.

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 :

- $p_T^\mu > 15$ GeV;
- $|\eta^\mu| < 2,4$;
- paramètres d'impact $d_z < 0,2$ cm et $d_{xy} < 0,045$ cm;
- $I^\mu < 0,2 p_T^\mu$;
- 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 :

- $p_T^e > 15$ GeV;
- $|\eta^e| < 2,4$;
- 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* 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

$$SF(d_z) = \frac{\mathcal{L}_{B-F}}{\mathcal{L}_{B-H}} + 0,953 \frac{\mathcal{L}_{G-H}}{\mathcal{L}_{B-H}} = 0,979 \quad (\text{X.11})$$

avec \mathcal{L}_x la luminosité intégrée sur la période x , notion introduite dans le chapitre 2.

Coupe sur D_ζ La variable D_ζ est définie selon

$$D_\zeta = p_\zeta^{\text{miss}} - 0,85 p_\zeta^{(\tau\tau)} \quad (\text{X.12})$$

avec

$$p_\zeta^{\text{miss}} = \vec{E}_T^{\text{miss}} \cdot \hat{\zeta}, \quad p_\zeta^{(\tau\tau)} = \vec{p}_T^{(\tau\tau)} \cdot \hat{\zeta} \quad (\text{X.13})$$

où $\hat{\zeta}$ est la direction bisectionnelle entre l'électron et le muon dans le plan transverse [37] et

$$\vec{p}_T^{(\tau\tau)} = \vec{p}_T^e + \vec{p}_T^\mu \quad (\text{X.14})$$

comme illustré sur la figure X.2. Il est requis que $D_\zeta \geq -35$ 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 coupe permet de plus de rejeter ce bruit de fond dans la région de signal.

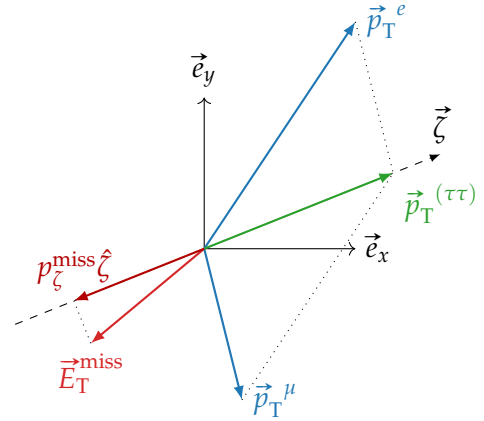


Figure X.2 – Illustration de la définition de $\hat{\zeta}$ [37]. Le plan de ce schéma est le plan transverse.

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 $p_T^\mu > 10 \text{ GeV}$, $|\eta^\mu| < 2,4$, passant le point de fonctionnement *medium* du *muonID* et d'isolation $I^\mu < 0,3 p_T^\mu$;
- de second 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.

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) [38] 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 [39] 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 [40, 41]. 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 $h, H, A \rightarrow \tau\tau$. Plusieurs processus y contribuent. En effet, ce derniers peuvent donner des états finaux similaires à ceux attendus avec le signal recherché, comme illustré sur la figure X.3. Ils peuvent également produire des objets physiques pouvant être interprétés comme des produits de désintégration de leptons τ . Ces

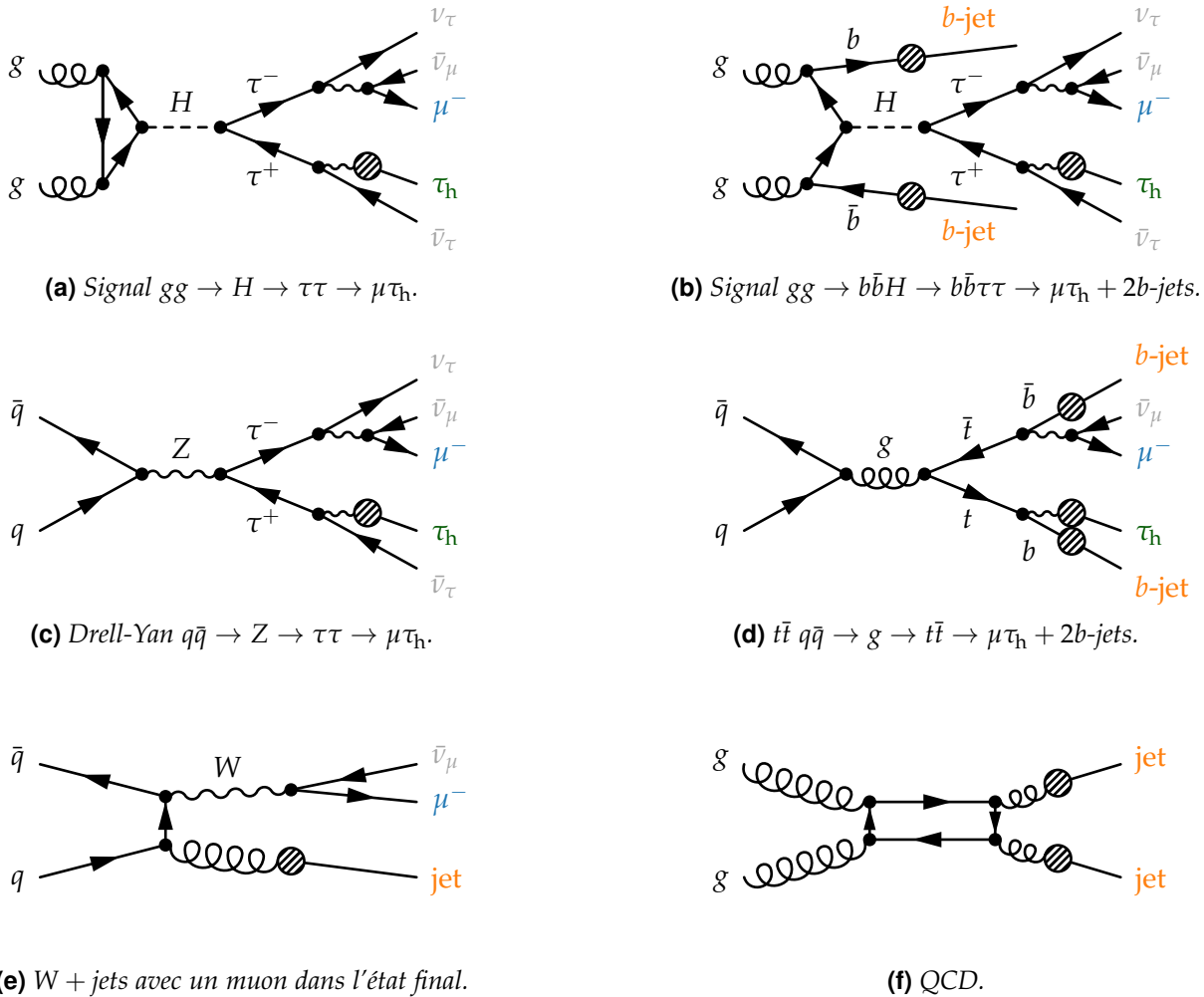


Figure X.3 – Diagrammes de Feynman complets des signaux ggH (X.3a) et $b\bar{b}H$ (X.3b) et bruits de fond Drell-Yan (X.3c), $t\bar{t}$ (X.3d), $W + \text{jets}$ (X.3e) et QCD (X.3f) de l'analyse illustrés dans le cas du canal $\mu\tau_h$.

processus, résumés dans le tableau X.1 avec les pourcentages de leurs contributions au bruit de fond total, sont :

- $Z \rightarrow \tau\tau, Z \rightarrow \ell\ell$** La désintégration du boson Z en paire de leptons τ ($Z \rightarrow \tau\tau$), ainsi qu'en paire de muons ou d'électrons ($Z \rightarrow \ell\ell$) 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 X.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.
- $W + \text{jets}$** 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 X.3e. Le W peut être produit par annihilation d'une paire de quarks, comme sur la figure X.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 X.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 $W, h \rightarrow W^+W^-$, constitue également un bruit de fond vis-à-vis de l'analyse $h \rightarrow \tau\tau$. 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 .

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 $Z \rightarrow \tau\tau$ 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 donnés 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 $\Delta R = 0,2$.

Bruit de fond	Canal			
	$\tau_h\tau_h$	$\mu\tau_h$	$e\tau_h$	$e\mu$
$Z \rightarrow \tau\tau$	33	46	27	20
$Z \rightarrow \ell\ell, \ell \in \{e, \mu\}$	~ 1	2	9	1
$t\bar{t}$	< 1	13	18	54
$W + \text{jets}$	< 1	36	42	3
QCD	66			11
Diboson	< 1	3	4	11

Tableau X.1 – Contributions en pourcent des bruits de fond aux canaux étudiés.

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 X.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*) $\gamma \rightarrow \ell^+ \ell^-$. 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 X.2, pour L_1 et L_2 selon les coupures données dans le tableau X.3.

gen_match	Type de particule	Propriétés de l'objet au niveau générateur
1	électron natif	<code> pdgID = 11</code> , $p_T > 8$ GeV, <code>IsPrompt == True</code>
2	muon natif	<code> pdgID = 13</code> , $p_T > 8$ GeV, <code>IsPrompt == True</code>
3	$\tau \rightarrow e$	<code> pdgID = 11</code> , $p_T > 8$ GeV, <code>IsDirectPromptTauDecayProduct == True</code>
4	$\tau \rightarrow \mu$	<code> pdgID = 13</code> , $p_T > 8$ GeV, <code>IsDirectPromptTauDecayProduct == True</code>
5	$\tau \rightarrow \tau_h$	Tau hadronique généré
6	Faux τ_h , τ_h de l'empilement	Tout objet ne rentrant pas dans les catégories 1 à 5

Tableau X.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 X.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 détails dans la référence [42], se déroule en quatre étapes, résumées sur la figure X.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 $Z \rightarrow \mu\mu$, $t\bar{t}$ 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 [42]. 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.

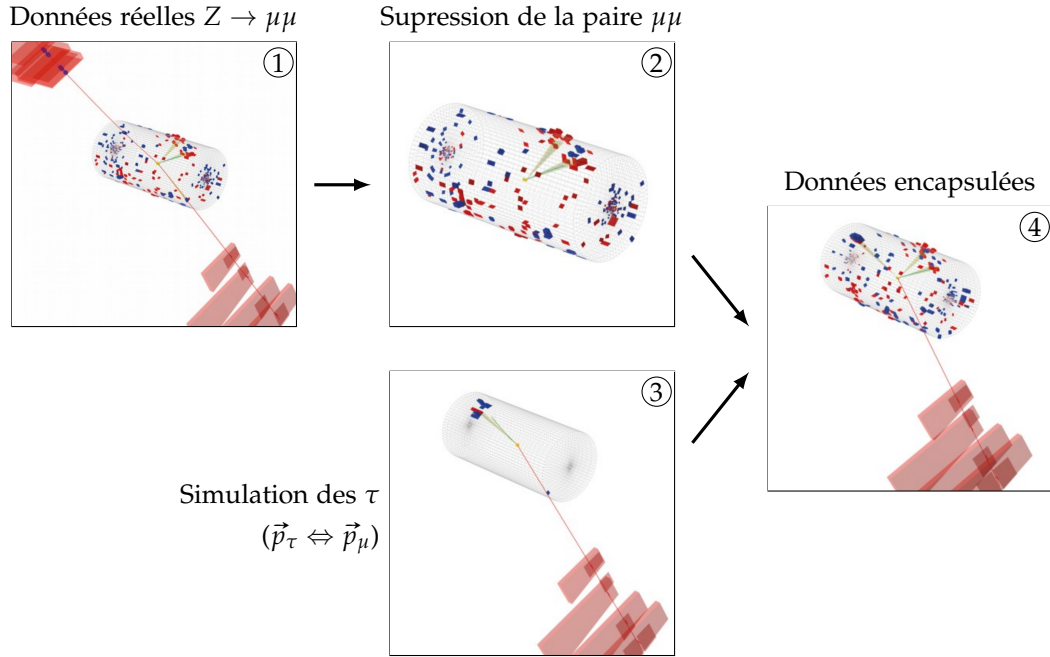


Figure X.4 – Schéma récapitulatif de la méthode des données encapsulées [42], illustrée dans le cas de l'état final $\mu\tau_h$.

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 X.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 $E_T^{\text{miss}} < 60 \text{ GeV}$.

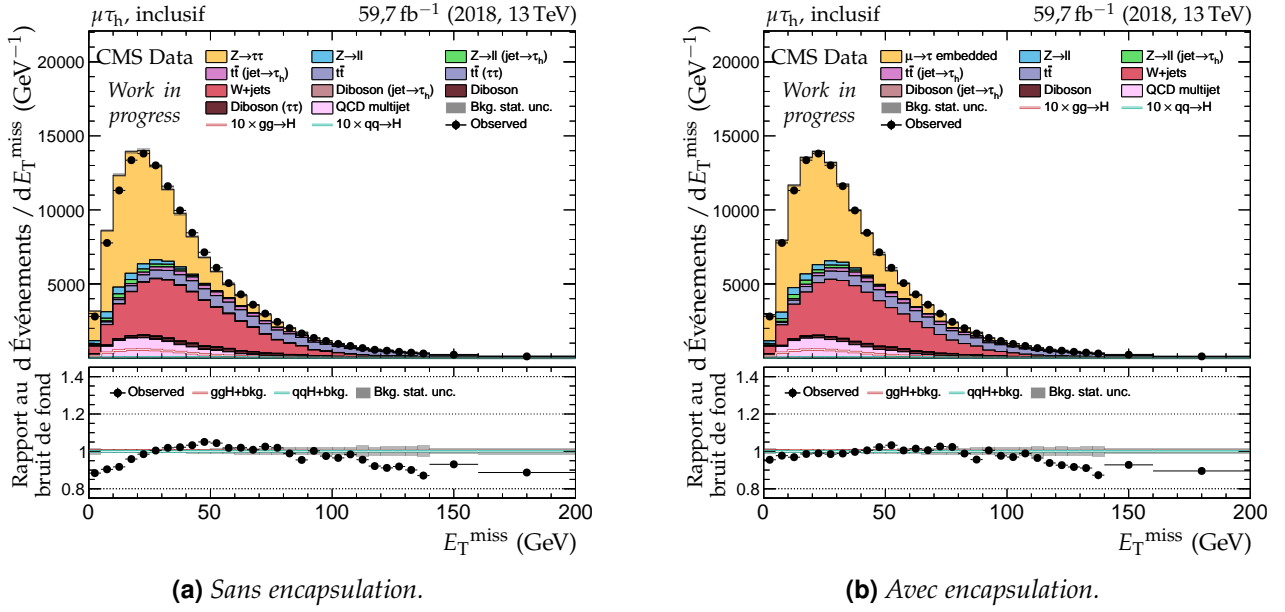


Figure X.5 – Distributions de E_T^{miss} pour le canal $\mu\tau_h$ en 2018. Les données observées (points noirs) sont comparées à la modélisation des bruits de fond (histogrammes remplis en couleur et empilés). Les bandes grisées correspondent à l'incertitude statistique totale sur le bruit de fond. 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 X.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, $0,2 \leq I_{\text{rel}}^\mu < 0,5$ au lieu de $I_{\text{rel}}^\mu < 0,2$;
- 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 X.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,

$$A = B \times \frac{C}{D} \Leftrightarrow h_v^A = h_v^B \times \frac{\int h_v^C}{\int h_v^D} \quad (\text{X.15})$$

où h_v^X correspond à la distribution de la variable v dans la région X et $\int h_v^X$ 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 [43] 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_T^e ;
- l'impulsion transverse du muon, p_T^μ ;

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 ($0,15 \leq I_{\text{rel}}^e < 0,5$ au lieu de $I_{\text{rel}}^e < 0,15$) 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 ($\tau_h\tau_h$) et semi-leptoniques ($\ell\tau_h$ 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 $\tau_h\tau_h$, 38 % dans le canal $\mu\tau_h$ et 68 % dans le canal $e\tau_h$ [44, 45]. Les processus physiques responsables des faux τ_h sont majoritairement QCD, W + jets et $t\bar{t}$. Dans le canal $\tau_h\tau_h$, près de 93 % des faux τ_h proviennent du bruit de fond QCD. Dans les canaux $\ell\tau_h$, 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 $Z \rightarrow \tau\tau$ 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 [44, 45]. 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

$$\text{FF} = \frac{n_{\text{iso}}}{n_{\text{anti-iso}}} = \frac{n(\text{Medium})}{n(\text{VVVLoose} \&\& \text{!Medium})} \quad (\text{X.16})$$

avec :

- $n(\text{Medium})$ 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;
- $n(\text{VVVLoose} \&\& \text{!Medium})$ 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^{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 $n_{j \rightarrow \tau_h}$, est alors obtenue par produit en croix avec la quantité d'événements dans la AR, notée n_{AR} , selon

$$n_{j \rightarrow \tau_h} = n_{\text{AR}} \cdot \text{FF}, \quad (\text{X.17})$$

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 (X.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

$$\text{FF} = \sum_i f_i \cdot \text{FF}_i, \quad f_i = \frac{n_{\text{AR}}^i}{\sum_j n_{\text{AR}}^j}, \quad i, j \in \{ \text{QCD}, W + \text{jets}, t\bar{t} \}. \quad (\text{X.18})$$

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 (X.24), m_T^{tot} .

La figure X.7 présente ces fractions en fonction de m_T^{tot} pour $m_T^\mu < 40 \text{ GeV}$ et $N_{b\text{-jets}} \in \{0, \geq 1\}$ pour le canal $\mu \tau_h$ en 2018. La méthode des facteurs de faux est résumée figure X.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 $I_{\text{rel}}^\ell > 0,05$ 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 $N_{\text{pre } b\text{-jets}} \geq 1$ et dans chacun de ces deux cas pour :

- $p_T^{\text{jet}} / p_T^{\tau_h} < 1,25$;
- $1,25 \leq p_T^{\text{jet}} / p_T^{\tau_h} < 1,5$;
- $1,5 \leq p_T^{\text{jet}} / p_T^{\tau_h}$.

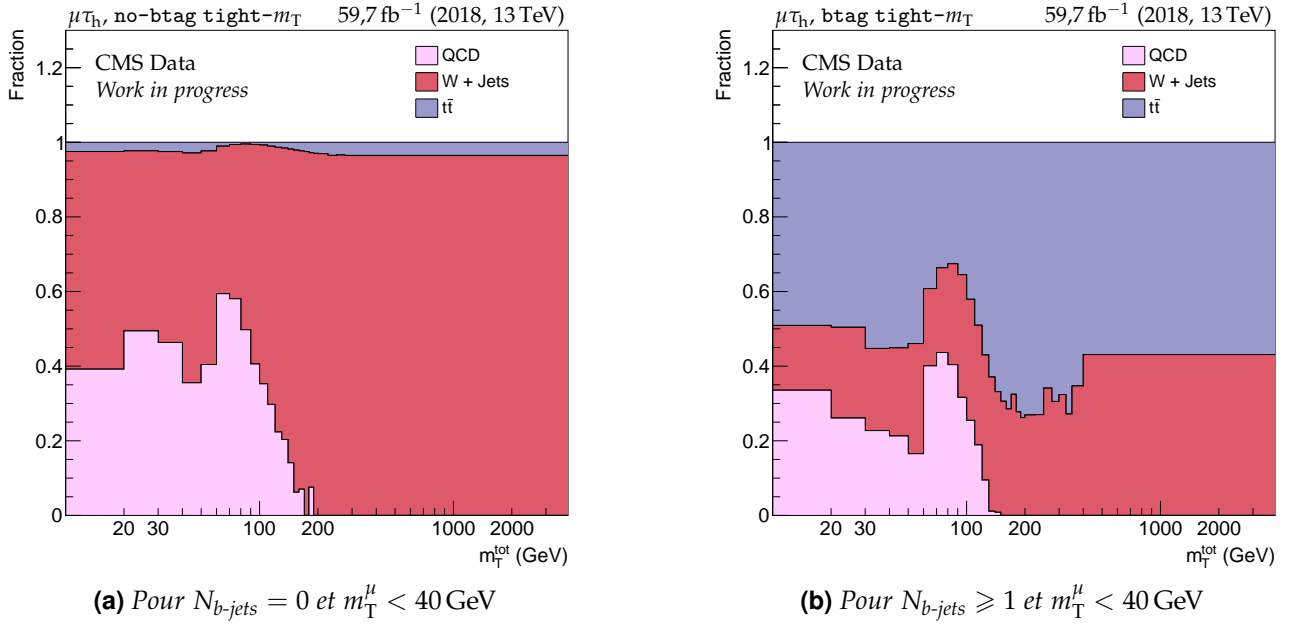


Figure X.7 – Fractions des sources de faux τ_h dans le canal $\mu\tau_h$ en 2018 [21].

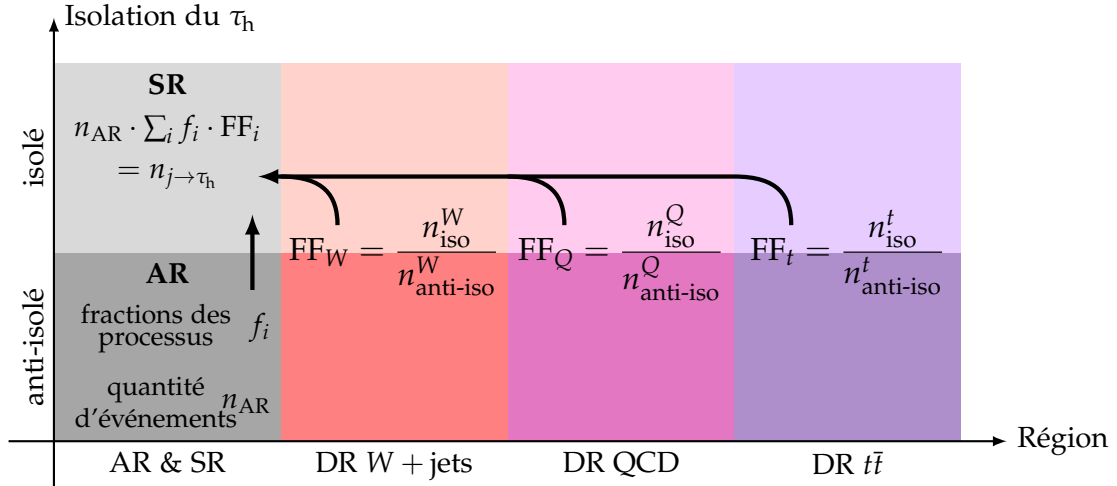


Figure X.8 – Illustration de la méthode des facteurs de faux. Les facteurs de faux sont obtenus à partir du nombre d'événements avec des τ_h isolés et anti-isolés dans les régions de détermination (DR) de chaque processus contribuant significativement au bruit de fond contenant des faux τ_h . La quantité 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 $p_{T^{\tau_h}}$ de FF_Q est modélisée par un polynôme de degré 3 ajusté aux mesures pour $p_{T^{\tau_h}} < 200$ GeV.

Dans le cas du canal $\tau_h\tau_h$, peu d'événements sont disponibles pour $p_{T^{\tau_h}} \geq 200$ GeV. Le facteur de faux FF_Q est ainsi fixé à la valeur mesurée pour $p_{T^{\tau_h}}^{\text{jet}}/p_{T^{\tau_h}} \geq 1,5$ et à la valeur à 200 GeV du polynôme pour $p_{T^{\tau_h}}^{\text{jet}}/p_{T^{\tau_h}} < 1,5$. Pour les canaux $\ell\tau_h$, la situation est similaire à partir de $p_{T^{\tau_h}} \geq 140$ GeV. La valeur utilisée à haut $p_{T^{\tau_h}}$ suit la logique suivante :

- si l'erreur relative sur FF_Q pour $p_{T^{\tau_h}} \geq 200$ GeV est inférieure à 0,5 :
 - si l'erreur relative sur FF_Q pour $p_{T^{\tau_h}} \in [140, 200]$ GeV est inférieure à 0,5, les valeurs mesurées sont utilisées sur les intervalles $[140, 200]$ GeV et $[200, \infty[$ GeV,
 - si l'erreur relative sur FF_Q pour $p_{T^{\tau_h}} \in [140, 200]$ GeV est supérieure à 0,5, la valeur mesurée est utilisée sur l'intervalle $[200, \infty[$ GeV ;
- si l'erreur relative sur FF_Q pour $p_{T^{\tau_h}} \geq 200$ GeV est supérieure à 0,5 et inférieure à 0,5 pour $p_{T^{\tau_h}} \in [140, 200]$ GeV, la valeur mesurée est utilisée sur l'intervalle $[140, \infty[$ 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.

Le facteur de faux FF_W est mesuré séparément pour $N_{\text{pre } b\text{-jets}} = 0$ et $N_{\text{pre } b\text{-jets}} \geq 1$ et dans chacun de ces deux cas pour :

- $p_T^{\text{jet}}/p_T^{\tau_h} < 1,25$;
- $1,25 \leq p_T^{\text{jet}}/p_T^{\tau_h} < 1,5$;
- $1,5 \leq p_T^{\text{jet}}/p_T^{\tau_h}$.

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 $p_T^{\tau_h} \geq 140$ 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 X.9 pour $N_{\text{pre } b\text{-jets}} \geq 1$ et $1,25 \leq p_T^{\text{jet}}/p_T^{\tau_h} < 1,5$. Il y apparaît l'effet du traitement de la région à haut $p_T^{\tau_h}$.

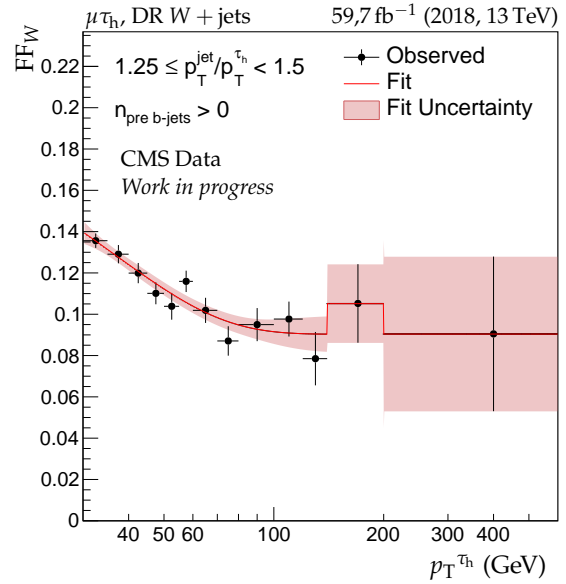


Figure X.9 – Ajustement de FF_W dans le canal $\mu\tau_h$ en 2018 [21].

$t\bar{t}$ 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 :

- $p_T^{\text{jet}}/p_T^{\tau_h} < 1,25$;
- $1,25 \leq p_T^{\text{jet}}/p_T^{\tau_h} < 1,5$;
- $1,5 \leq p_T^{\text{jet}}/p_T^{\tau_h}$;

sans séparation en $N_{\text{pre } b\text{-jets}}$, la majorité des événements $t\bar{t}$ vérifiant $N_{\text{pre } b\text{-jets}} \geq 1$. L'ajustement obtenu pour FF_t sur le canal $\mu\tau_h$ en 2018 est illustré figure X.10 pour $p_T^{\text{jet}}/p_T^{\tau_h} < 1,25$.

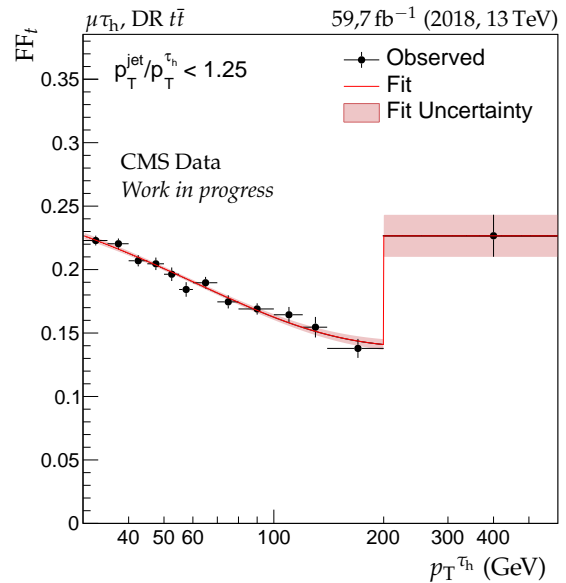


Figure X.10 – Ajustement de FF_t dans le canal $\mu\tau_h$ en 2018 [21].

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_T^ℓ ;
- 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 ,

$$C_Q = \frac{E_T^{\text{miss}}}{p_T^{\tau_h}} \cos\left(\Delta\phi(\vec{E}_T^{\text{miss}}, \vec{p}_T^{\tau_h})\right) ; \quad (\text{X.19})$$

- de la fraction d'énergie transverse manquante alignée avec le τ_h pour les événements W + jets, C_W ,

$$C_W = \frac{\|\vec{E}_T^{\text{miss}} + \vec{p}_T^\ell\|}{p_T^{\tau_h}} \cos\left(\Delta\phi(\vec{E}_T^{\text{miss}} + \vec{p}_T^\ell, \vec{p}_T^{\tau_h})\right), \quad (\text{X.20})$$

dont la définition est semblable à celle de C_Q mais où \vec{E}_T^{miss} est remplacé par $\vec{E}_T^{\text{miss}} + \vec{p}_T^\ell$ afin de prendre en compte la contribution à E_T^{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 $\mu\tau_h$ en 2018 sont illustrées sur la figure X.11.

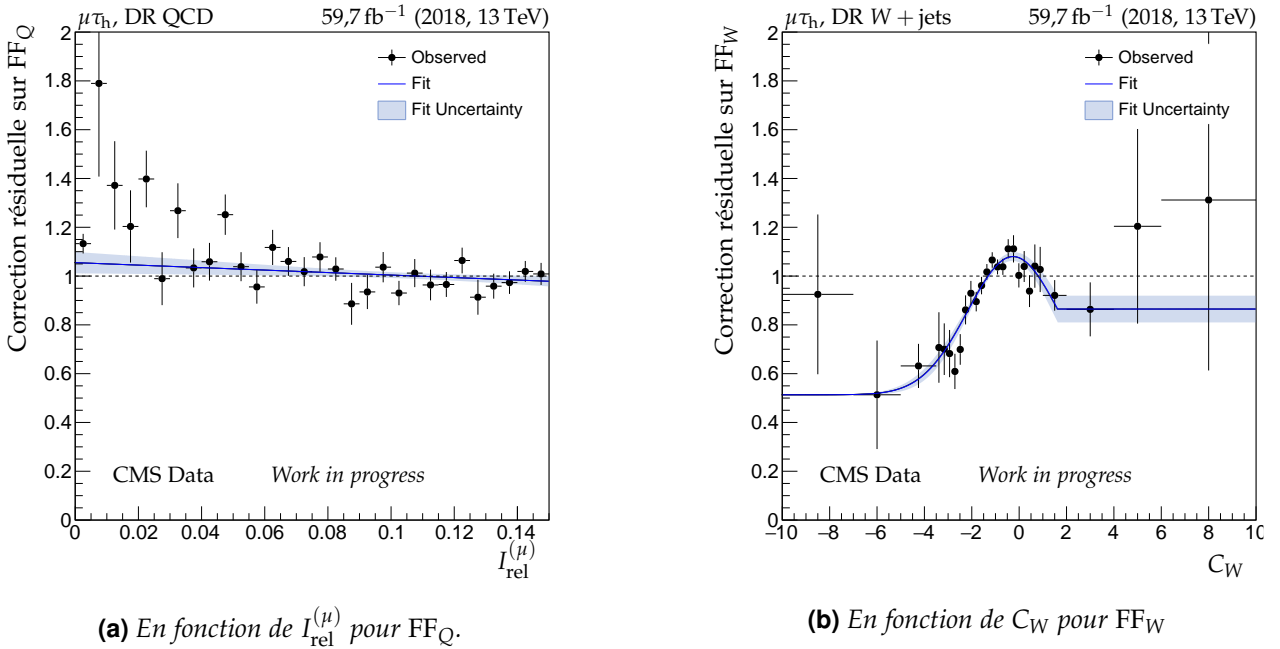


Figure X.11 – Corrections résiduelles des facteurs de faux dans le canal $\mu\tau_h$ en 2018 [21].

L'amélioration de la description des données ainsi obtenue grâce aux facteurs de faux est visible sur la figure X.12, où les distributions de la masse transverse de l'électron dans le canal $e\tau_h$ 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.

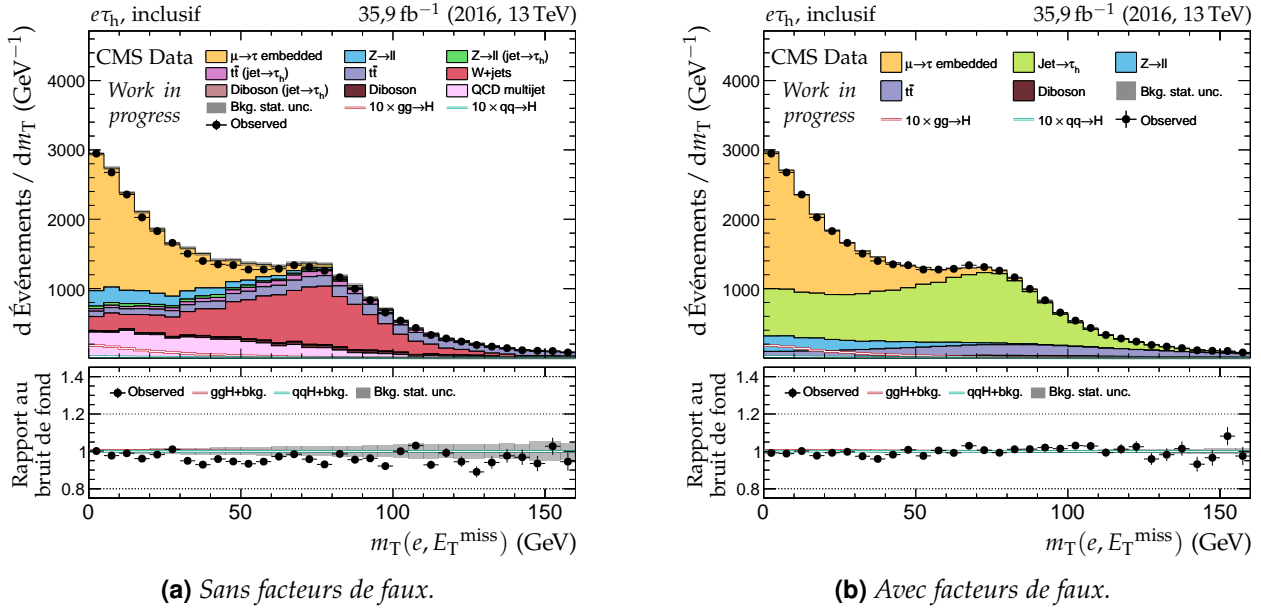


Figure X.12 – Distributions de la masse transverse de l'électron pour le canal $e\tau_h$ en 2016.

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 [13]. 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 $h \rightarrow \tau\tau$ dans le cadre du SM [46–48]. 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 [13], 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 : $N_{b\text{-jets}} \geq 1$.

Dans le cas des canaux $\mu\tau_h$, $e\tau_h$ et $e\mu$, chacune de ces deux catégories est à nouveau subdivisée.

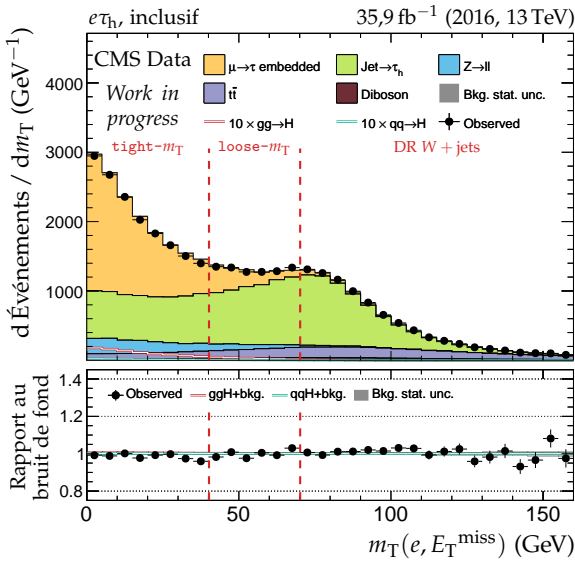
Canaux $\mu\tau_h$ et $e\tau_h$ Dans ces deux canaux, la masse transverse de L_1 (le muon ou l'électron, notés ℓ), définie par

$$m_T^\ell = m_T(\ell, E_T^{\text{miss}}) = \sqrt{2 p_T^\ell E_T^{\text{miss}} (1 - \cos \Delta\phi)} \quad (\text{X.21})$$

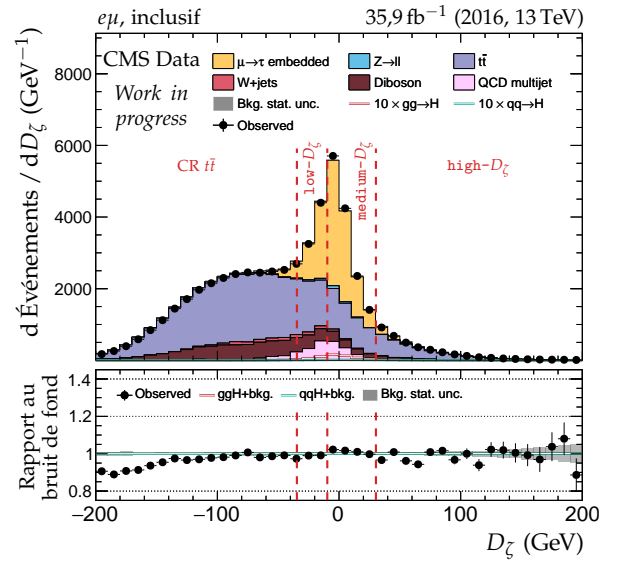
avec $\Delta\phi = \phi^\ell - \phi^{E_T^{\text{miss}}}$, est utilisée afin de définir deux catégories :

- **tight- m_T** : $m_T^\ell < 40 \text{ GeV}$;
- **loose- m_T** : $40 \text{ GeV} \leq m_T^\ell < 70 \text{ GeV}$;

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 $W + \text{jets}$. 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 X.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 X.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 (X.12) :

- **low- D_ζ** : $-35 \text{ GeV} \leq D_\zeta < -10 \text{ GeV}$;
- **medium- D_ζ** : $-10 \text{ GeV} \leq D_\zeta < 30 \text{ GeV}$;
- **high- D_ζ** : $30 \text{ GeV} \leq D_\zeta$;

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 X.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 X.14 pour les quatre canaux considérés.

5.1.2 Variable discriminante

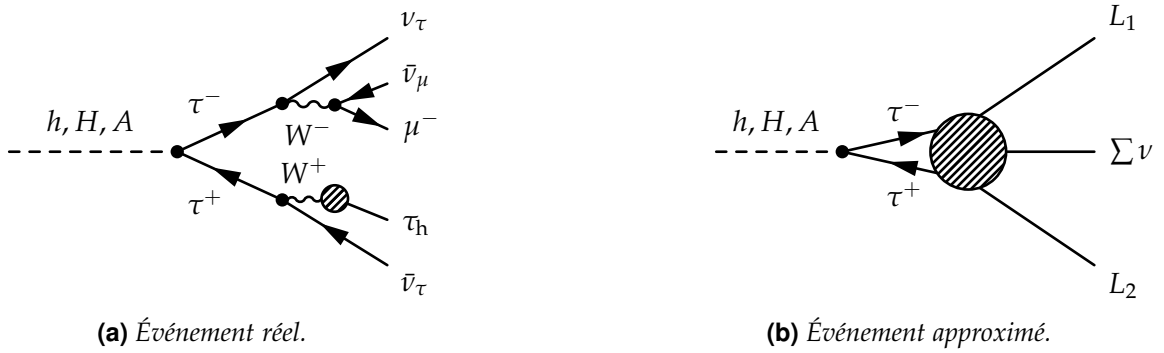
La masse invariante d'une particule peut être obtenue par un calcul de relativité restreinte à partir des propriétés cinématiques de chacun de ses produits de désintégration. Cette observable doit

	no-btag			btag		
$H \rightarrow \tau\tau \rightarrow \tau_h \tau_h$						
$H \rightarrow \tau\tau \rightarrow \mu \tau_h$	tight- m_T	loose- m_T		tight- m_T	loose- m_T	
$H \rightarrow \tau\tau \rightarrow e \tau_h$	tight- m_T	loose- m_T		tight- m_T	loose- m_T	
$H \rightarrow \tau\tau \rightarrow e \mu$	low- D_ζ	medium- D_ζ	high- D_ζ	low- D_ζ	medium- D_ζ	high- D_ζ

Figure X.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 $h \rightarrow ZZ \rightarrow 4\ell$ [49].

Cependant, dans l'analyse menée ici, l'état final comporte deux à quatre neutrinos issus des désintégrations des leptons τ . La figure X.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.

Figure X.15 – Diagrammes de Feynman d'un événement $H \rightarrow \tau\tau \rightarrow \mu \tau_h$, 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,

$$\vec{E}_T^{\text{miss}} = \sum_{\nu_i} \vec{p}_T^{\nu_i}, \quad (\text{X.22})$$

la somme se faisant sur tous les neutrinos de l'événement dont ceux de l'état final du processus $H \rightarrow \tau\tau$.

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 X.15a ;
- L_2 la partie visible de la désintégration de l'autre lepton τ , par exemple le τ_h résultant de $W^+ \rightarrow u\bar{d}$ sur la figure X.15a ;
- $\Sigma \nu \simeq E_T^{\text{miss}}$ 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^{miss} ;

comme illustré figure X.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

$$\begin{aligned}
 m^2 &= E^2 - p^2 = \left(\sum_{i \in \{L_1, L_2, \Sigma \nu\}} E_i \right)^2 - \left(\sum_{i \in \{L_1, L_2, \Sigma \nu\}} \vec{p}_T^i \right)^2 \\
 &= (E_{L_1} + E_{L_2} + E_T^{\text{miss}})^2 - (\vec{p}_T^{L_1} + \vec{p}_T^{L_2} + \vec{E}_T^{\text{miss}})^2 \\
 &= E_{L_1}^2 + E_{L_2}^2 + E_T^{\text{miss}^2} + 2(E_{L_1}E_{L_2} + E_{L_1}E_T^{\text{miss}} + E_{L_2}E_T^{\text{miss}}) \\
 &\quad - (\vec{p}_T^{L_1^2} + \vec{p}_T^{L_2^2} + \vec{E}_T^{\text{miss}^2}) - 2(\vec{p}_T^{L_1} \cdot \vec{p}_T^{L_2} + \vec{p}_T^{L_1} \cdot \vec{E}_T^{\text{miss}} + \vec{p}_T^{L_2} \cdot \vec{E}_T^{\text{miss}}) \\
 &= 2 \left[p_T^{L_1} p_T^{L_2} (1 - \cos \Delta\phi^{(L_1, L_2)}) \right. \\
 &\quad \left. + p_T^{L_1} E_T^{\text{miss}} (1 - \cos \Delta\phi^{(L_1, E_T^{\text{miss}})}) + p_T^{L_2} E_T^{\text{miss}} (1 - \cos \Delta\phi^{(L_2, E_T^{\text{miss}})}) \right] \\
 &= m_T^2(L_1, L_2) + m_T^2(L_1, E_T^{\text{miss}}) + m_T^2(L_2, E_T^{\text{miss}}) = m_T^{\text{tot}^2}.
 \end{aligned} \tag{X.23}$$

La variable discriminante utilisée dans les catégories BSM est ainsi m_T^{tot} , la masse transverse totale, définie telle que

$$m_T^{\text{tot}} = \sqrt{m_T^2(L_1, L_2) + m_T^2(L_1, E_T^{\text{miss}}) + m_T^2(L_2, E_T^{\text{miss}})} \tag{X.24}$$

avec

$$m_T(A, B) = \sqrt{2 p_T^A p_T^B (1 - \cos \Delta\phi^{(A, B)})}. \tag{X.25}$$

À titre d'illustration, les distributions obtenues pour la catégorie *btag tight- m_T* du canal $\mu\tau_h$ et *btag medium- D_ζ* du canal $e\mu$ en 2017 sont représentées en figure X.16.

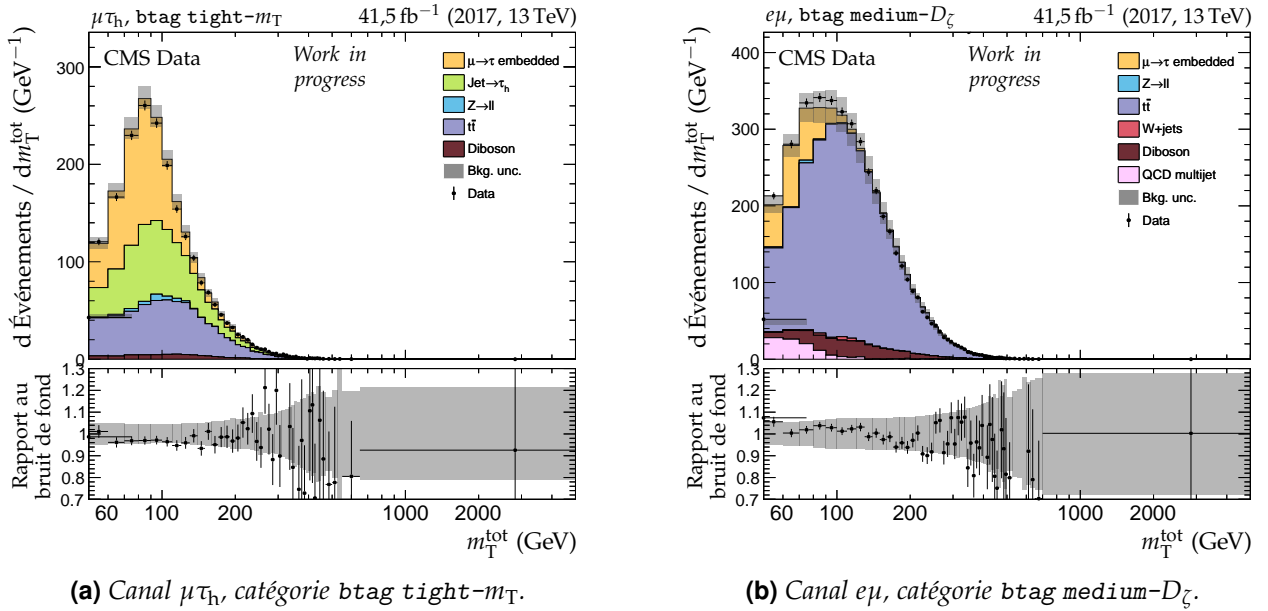


Figure X.16 – Distributions de m_T^{tot} en 2017 pour deux catégories et canaux. Les données observées (points noirs) sont comparées à la modélisation des bruits de fond (histogrammes remplis en couleur et empilés). Les bandes grises correspondent à l'incertitude totale (statistique et systématique) 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 [47, 48], 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 [48].

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 $e\mu$ ($m_T(\vec{p}_T^e + \vec{p}_T^\mu, \vec{E}_T^{\text{miss}})$);
- 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 $\Delta\eta_{jj}$;
- l'impulsion transverse totale des deux principaux jets de l'événement;
- la masse du dilepton estimée par SVFIT [50], 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, $p_T^{(\tau\tau)}$.

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*,

$$\text{Softmax}(\vec{z})_j = \frac{\exp(z_j)}{\sum_{k=1}^n \exp(z_k)}, \quad j \in \{1, \dots, K\}, \quad (\text{X.26})$$

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 différenciables d'autres bruits de fond, une catégorie *misc* est également définie. Les différentes catégories ainsi possibles sont listées dans le tableau X.4. Le canal $\tau_h \tau_h$ ne devant pas contenir d'électron ni de muons, les processus $Z \rightarrow \ell\ell$ ($\ell \in \{e, \mu\}$), $t\bar{t}$ 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 $\tau_h \tau_h$. Celle-ci couvre donc les processus $Z \rightarrow \ell\ell$, $t\bar{t}$, W + jets et Diboson dans le canal $\tau_h \tau_h$; Diboson dans les canaux $\mu\tau_h$ et $e\tau_h$; $Z \rightarrow \ell\ell$ et W + jets dans le canal $e\mu$. Les processus EWK $Z \rightarrow LL$ et EWK $Z \rightarrow \nu\nu$, introduits dans l'annexe E, sont également associés à la catégorie *misc* dans les canaux $\tau_h \tau_h$, $\mu\tau_h$ et $e\tau_h$.

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,

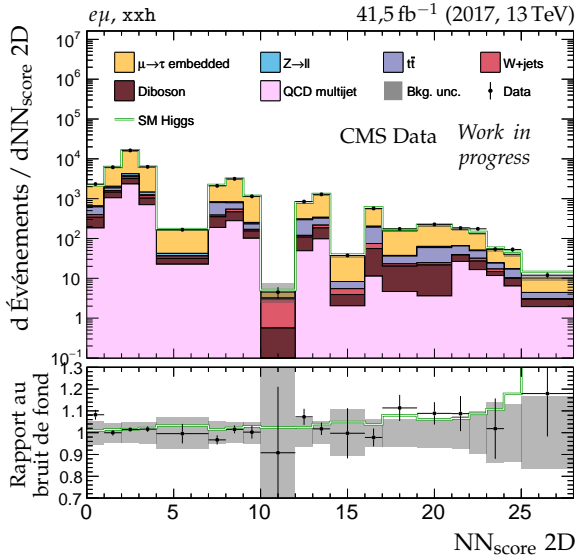
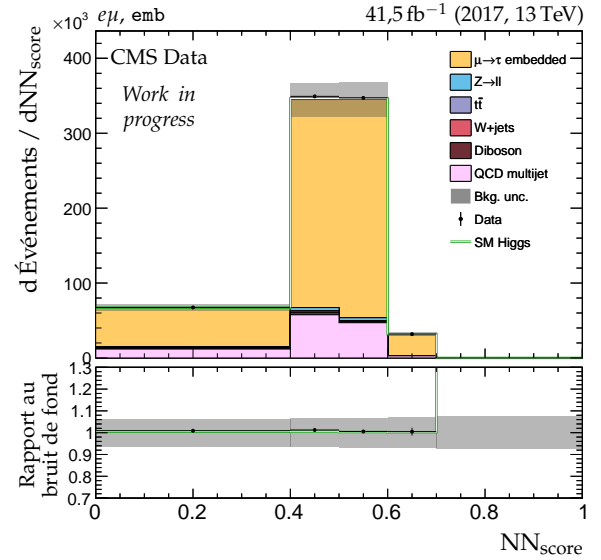
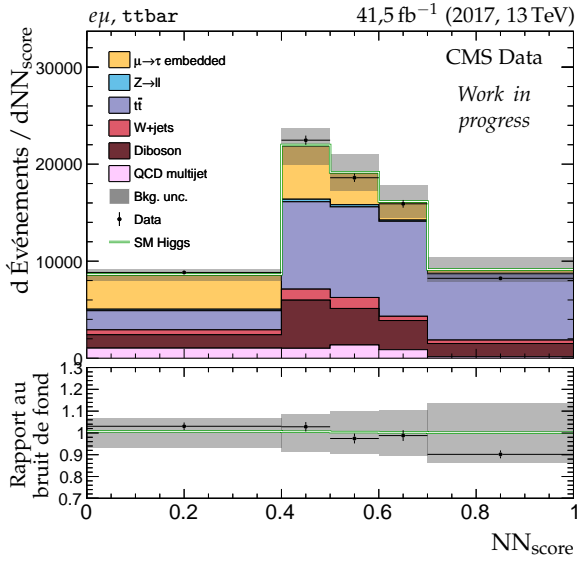
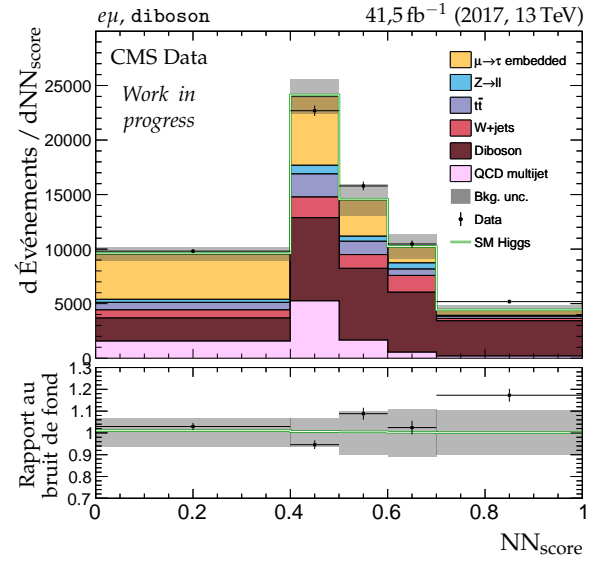
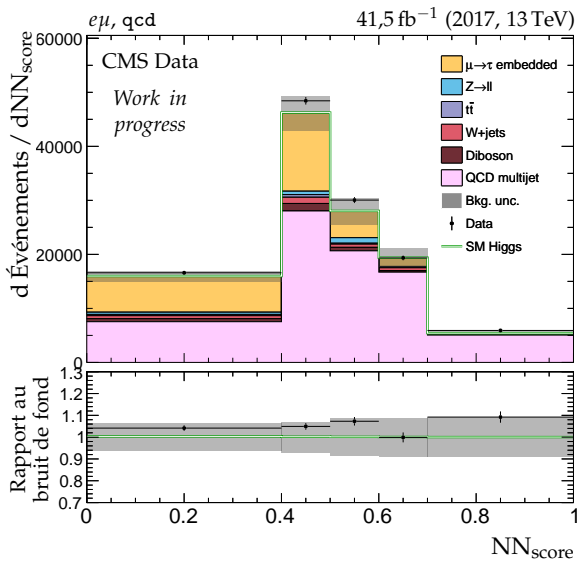
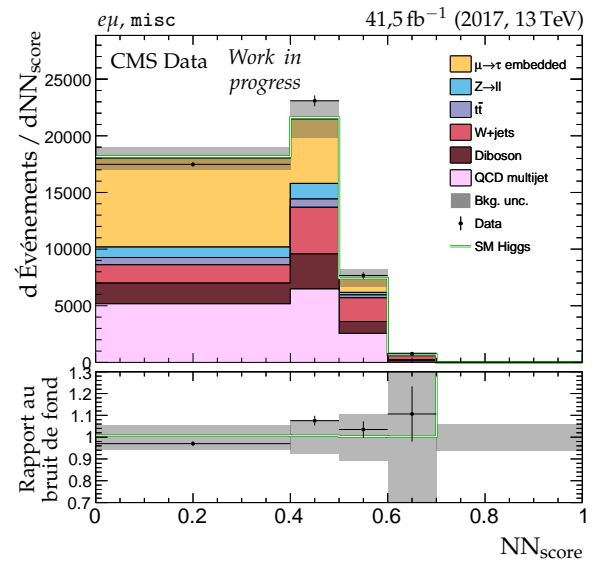

 (a) Catégorie xxh .

 (b) Catégorie emb .

 (c) Catégorie $t\bar{t}bar$.

 (d) Catégorie $diboson$.

 (e) Catégorie qcd .

 (f) Catégorie $misc$.

 Figure X.17 – Distributions de NN_{score} en 2017 dans le canal $e\mu$.

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 X.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 NN_{score} . À titre d’illustration, les distributions obtenues pour le canal $e\mu$ en 2017 sont représentées en figure X.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 [47] 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 X.18.

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 NN_{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.

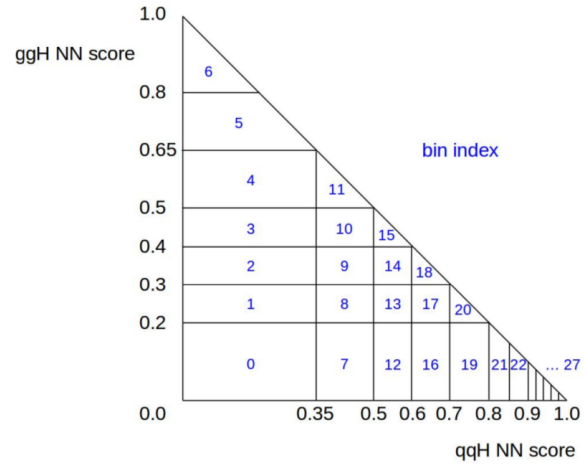


Figure X.18 – Réduction à une dimension de la catégorie xxh [47].

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 $H \rightarrow \tau\tau$ dans le cadre du MSSM [11-13]. 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 [51].

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 [50]. La région BSM, quant à elle, concerne les événements contenant des jets issus de quark b ($N_{b\text{-jets}} \geq 1$) ou tels que $m_{\text{SVFIT}} \geq 250 \text{ GeV}$. Les deux régions ainsi obtenues ne se recouvrent pas et peuvent se résumer selon le schéma de la figure X.19.

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 [52], 2,3 % en 2017 [53] et 2,5 % en 2018 [54]. Elle est donnée par le POG Lumi. Plus de détails sont disponibles dans la référence [55]. L'incertitude pour l'année 2016 a récemment été abaissée à 1,2 % [56], 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 $\ell \rightarrow \tau_h$ ($\ell \rightarrow \tau_h$ fake rate) L'incertitude, fournie par le POG tau, dépend de la pseudo-rapacité 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 [57] 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.

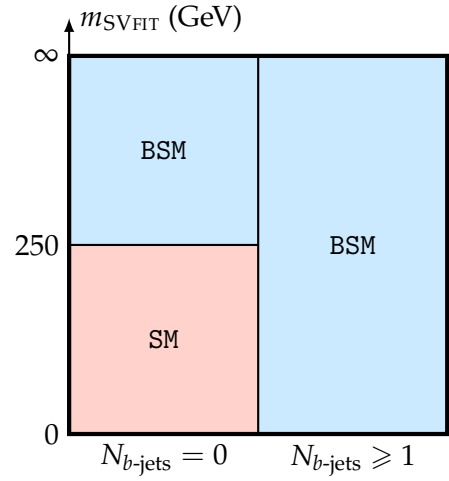


Figure X.19 – Définition des deux régions utilisant des catégories différentes.

Bruits de fond simulés Les incertitudes de normalisation sur les bruits de fond simulés sont :

- 2 % sur les processus $Z \rightarrow \ell\ell$, due à l’incertitude sur la section efficace Drell-Yan [58], corrélée entre les années ;
- 4 % sur les processus $W + \text{jets}$, due à l’incertitude sur leurs sections efficaces [58], corrélée entre les années ;
- 5 % sur les processus Diboson et *Single top*, due à l’incertitude sur leurs sections efficaces [58], 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 [59] sont appliquées ;

Incertaines 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 incluses. Elles sont fournies par le groupe LHC Higgs [60].

Dans le cas du processus $gg \rightarrow b\bar{b}\Phi \rightarrow \tau\tau$ 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 $\frac{1}{2} \leq \frac{\mu_R}{\mu_F} \leq 2$. 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* [61] 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 décorrelée 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 $Z \rightarrow \ell\ell$ dans tous les canaux. Les paramètres de nuisance des années 2017 et 2018 sont corrélés, car les mêmes réglages de simulation sont utilisés (CP5 [62]). Celui de l'année 2016 est décorrelé.

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* [63]. 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 propagée à E_T^{miss} , ainsi qu'aux variable 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 [64] 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* [61].

Énergie des leptons identifiés comme τ_h ($\ell \rightarrow \tau_h$ *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 $W + \text{jets}$ et $t\bar{t}$. L'incertitude est de 20 % pour les événements $W + \text{jets}$ 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.

Incertainces de segmentation (*bin-by-bin uncertainties*) Les incertainces 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.

Incertainces théoriques Dans le cas du processus $gg \rightarrow \Phi \rightarrow \tau\tau$, 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 $\mathfrak{s}\mathfrak{b}$.

Le test statistique de ces deux hypothèses est fait par une approche fréquentiste modifiée connue sous le nom de méthode CL_s [65-67], implémentée dans COMBINE.

Le modèle de vraisemblance utilisé est détaillé dans la section 7.1. La méthode CL_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 poissonniennes $\mathfrak{P}(n_i|v_i(\mu, \vec{\theta}))$ d'observer n_i événements dans chaque segment i de l'histogramme de la variable discriminante utilisée selon

$$\mathfrak{L}(\{n_i\}|\mu, \vec{\theta}) = \prod_i \mathfrak{P}(n_i|v_i(\mu, \vec{\theta})) \cdot \prod_j \mathfrak{C}(\theta_j, \tilde{\theta}_j), \quad \mathfrak{P}(n_i|v_i(\mu, \vec{\theta})) = \frac{v_i^{n_i}}{n_i!} e^{-v_i} \quad (\text{X.27})$$

où

- v_i est le nombre d'événements attendus dans ce segment dans l'hypothèse $\mathfrak{s}\mathfrak{b}$, c'est-à-dire

$$v_i(\mu, \vec{\theta}) = \mu s_i(\vec{\theta}) + b_i(\vec{\theta}) \quad (\text{X.28})$$

avec s le nombre d'événements de signal et b le nombre d'événements de bruit de fond. Lorsque $\mu = 0$, v_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 poissonniennes \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,

$$\mathfrak{C}(\theta, \tilde{\theta})|_{\text{facteurs}} = \frac{1}{\sqrt{2\pi \ln \kappa}} \frac{1}{\theta} \exp\left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln \kappa)^2}\right) \quad (\text{X.29})$$

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,

$$\mathfrak{C}(\theta, \tilde{\theta})|_{\text{stat}} = \frac{1}{\kappa \Gamma(\tilde{\theta} + 1)} \left(\frac{\theta}{\kappa}\right)^{\tilde{\theta}} \exp\left(-\frac{\theta}{\kappa}\right) \quad (\text{X.30})$$

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 $\pm 1\sigma$ 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

$$h(\vec{\lambda}) = h_0 + \sum_j \left(a(\lambda_j) h_j^+ + b(\lambda_j) h_0 + c(\lambda_j) h_j^- \right) \quad (\text{X.31})$$

avec

$$a = \begin{cases} \lambda(\lambda + 1)/2, & |\lambda| \leq 1, \\ 0, & \lambda < -1, \\ \lambda, & \lambda > +1, \end{cases} \quad b = \begin{cases} -\lambda^2, & |\lambda| \leq 1, \\ -|\lambda|, & |\lambda| > 1, \end{cases} \quad c = \begin{cases} \lambda(\lambda - 1)/2, & |\lambda| \leq 1, \\ |\lambda|, & \lambda < -1, \\ 0, & \lambda > +1. \end{cases} \quad (\text{X.32})$$

L'interpolation (X.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 [68, 69]. 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

$$e_{\text{tot}}^2 = \sum_{j \in \{i\}} e_j^2. \quad (\text{X.33})$$

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 X.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 [27] dans le code de COMBINE.

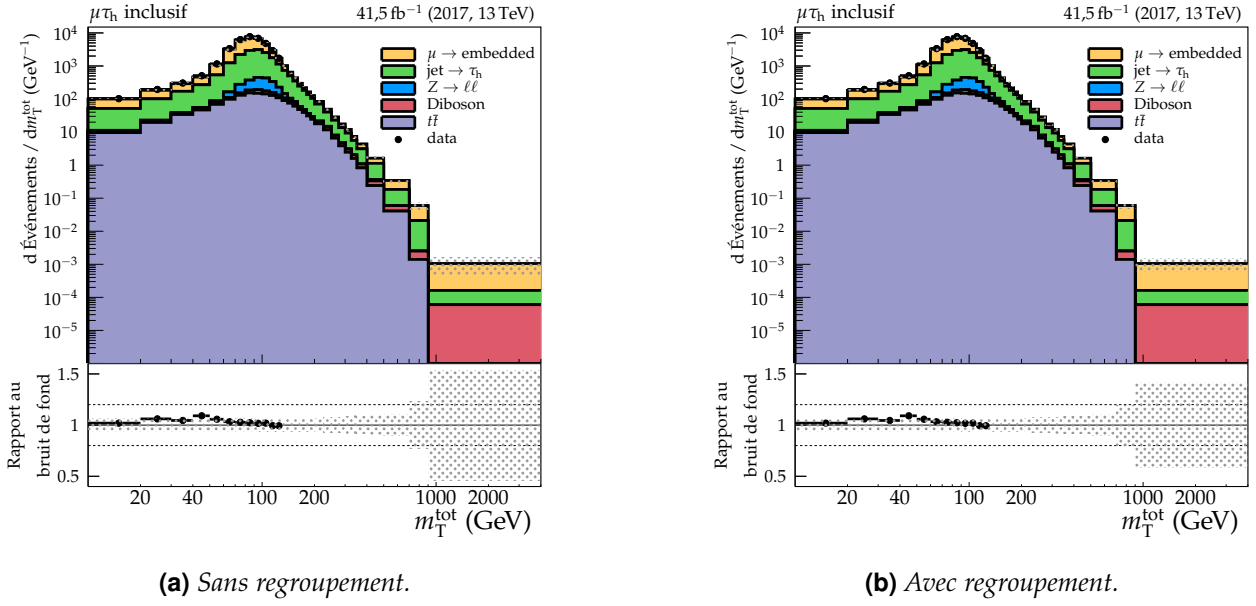


Figure X.20 – Distributions de m_T^{tot} avec et sans regroupement des incertitudes pour le canal $\mu\tau_h$ en 2017. Le tracé des données s'arrête à 130 GeV, avant la zone où le signal est attendu.

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 $0 < x_i \leq e_i$;

groupe D incertitude à conserver et $0 < x_i \leq e_i$.

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

$$\text{expand} = \text{std}::\text{sqrt}(1. / (1. - (\text{removed} / \text{tot_bbb_added}))) \Leftrightarrow E = \sqrt{\frac{1}{1 - \frac{R}{T}}} \quad (\text{X.34})$$

avec

$$R = \text{removed} = \sum_{i \in \{A,C\}} e_i^2, \quad T = \text{tot_bbb_added} = \sum_{i \in \{A,B\}} e_i^2. \quad (\text{X.35})$$

Ainsi, l'incertitude totale après regroupement s'exprime en fonction des incertitudes de chaque processus i avant regroupement selon

$$\begin{aligned} e_{\text{tot,après}}^2 &= \sum_{i \in \{A\}} (E \times 0 \times e_i)^2 + \sum_{i \in \{B\}} (E \times e_i)^2 + \sum_{i \in \{C\}} (0 \times e_i)^2 + \sum_{i \in \{D,Z\}} (e_i)^2 \\ &= E^2 \sum_{i \in \{B\}} e_i^2 + \sum_{i \in \{D,Z\}} e_i^2. \end{aligned} \quad (\text{X.36})$$

Or,

$$E^2 = \frac{1}{1 - \frac{R}{T}} = \frac{T}{T - R} = \frac{\sum_{i \in \{A,B\}} e_i^2}{\sum_{i \in \{A,B\}} e_i^2 - \sum_{i \in \{A,C\}} e_i^2} = \frac{\sum_{i \in \{A,B\}} e_i^2}{\sum_{i \in \{B\}} e_i^2 - \sum_{i \in \{C\}} e_i^2} \quad (\text{X.37})$$

soit

$$e_{\text{tot,après}}^2 = \frac{\sum_{i \in \{A,B\}} e_i^2}{\sum_{i \in \{B\}} e_i^2 - \sum_{i \in \{C\}} e_i^2} \times \sum_{i \in \{B\}} e_i^2 + \sum_{i \in \{D,Z\}} e_i^2 \quad (\text{X.38})$$

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 [27] 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 (X.38) se réécrit

$$e_{\text{tot,après}}^2 = \frac{\sum_{i \in \{A,B\}} e_i^2}{\sum_{i \in \{B\}} e_i^2} \times \sum_{i \in \{B\}} e_i^2 + \sum_{i \in \{D,Z\}} e_i^2 = \sum_{i \in \{A,B\}} e_i^2 + \sum_{i \in \{D,Z\}} e_i^2 = e_{\text{tot,avant}}^2, \quad (\text{X.39})$$

l'incertitude totale est donc bien conservée.

Dans le cas d'étude de la figure X.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_s

7.2.1 Approche fréquentiste classique

Afin de déterminer quantitativement quelle hypothèse, entre b et 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*),

$$q_\mu = -2 \ln \left(\frac{\mathcal{L}(\text{données} | \mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{données} | \hat{\mu}, \hat{\theta}_{\hat{\mu}})} \right), \quad 0 \leq \hat{\mu} \leq \mu \quad (\text{X.40})$$

où « données » se réfère aux quantités d'événements n_i dans chaque segment i des distributions des variables discriminantes dans chaque catégorie et $\hat{\theta}_x$ est l'ensemble des paramètres de nuisance maximisant \mathcal{L} pour $\mu = x$. L'ensemble $(\hat{\mu}, \hat{\theta}_{\hat{\mu}})$ donne le maximum global de \mathcal{L} . La contrainte $0 \leq \hat{\mu}$ impose une fréquence du signal positive, c'est-à-dire que μ a une interprétation physique. De plus, $\hat{\mu} \leq \mu$ 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 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 sb que celle effectivement réalisée, est définie par

$$CL_{sb} = \int_{q_\mu^{\text{obs}}}^{+\infty} f(q_\mu | \mu, \hat{\theta}_\mu) dq_\mu \quad (\text{X.41})$$

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 $\alpha = 1 - CL_{sb}$. 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 [65-67]. En plus de CL_{sb} définie précédemment, la probabilité de réaliser une observation moins compatible avec l'hypothèse b que celle effectivement réalisée, définie comme

$$CL_b = \int_{q_\mu^{\text{obs}}}^{+\infty} f(q_\mu | 0, \hat{\theta}_0) dq_\mu \quad (\text{X.42})$$

est déterminée. La quantité CL_5 est le rapport

$$CL_5 = \frac{CL_{sb}}{CL_b}. \quad (\text{X.43})$$

Dans l'approche fréquentiste modifiée, l'exclusion à 95 % de confiance est obtenue lorsque $CL_5 \leq 0,05$. Plus de détails sur la méthode CL_5 sont disponibles dans la référence [70].

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 $CL_5 = 0,05$. 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 :

- $gg \rightarrow \Phi \rightarrow \tau\tau$;
- $gg \rightarrow b\bar{b}\Phi \rightarrow \tau\tau$.

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 $gg \rightarrow \Phi \rightarrow \tau\tau$, 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 $Z \rightarrow \tau\tau$, $Z \rightarrow \ell\ell$, $W + \text{jets}$, $t\bar{t}$, Diboson, *Single top* et QCD, le boson de Higgs du SM h dans ses désintégrations $h \rightarrow \tau\tau$ et $h \rightarrow WW$ est également considéré comme un bruit de fond. Le signal du MSSM est ainsi constitué des processus $gg \rightarrow \Phi \rightarrow \tau\tau$ et $gg \rightarrow b\bar{b}\Phi \rightarrow \tau\tau$ 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 [20] 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 [20]. 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 [20] :

- 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 [60].

Le signal du MSSM attendu (hypothèse **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 $\Phi \in \{h_{\text{MSSM}}, H, A\} \rightarrow \tau\tau$ 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

$$\sin^2(\beta - \alpha) \times \frac{\mathcal{BR}(h_{\text{MSSM}} \rightarrow \tau\tau)}{\mathcal{BR}(h_{\text{SM}} \rightarrow \tau\tau)}. \quad (\text{X.44})$$

Dans la limite découplée, $\sin^2(\beta - \alpha) \simeq 1$, seuls les rapport de branchement ont donc un effet significatif.

L'hypothèse **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 $v_i(\mu, \vec{\theta})$ donnant le nombre d'événements attendus,

$$v_i(\mu, \vec{\theta}) = \mu s_i(\vec{\theta}) + b_i(\vec{\theta}), \quad (\text{X.45})$$

est réécrit sous la forme

$$v_i(\mu, \vec{\theta}) = \mu s_i^{\text{MSSM}}(\vec{\theta}) + (1 - \mu) s_i^{\text{SM}}(\vec{\theta}) + b_i(\vec{\theta}) \quad (\text{X.46})$$

avec s_i^{MSSM} le signal $\Phi \in \{h_{\text{MSSM}}, H, A\} \rightarrow \tau\tau$ 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 $h_{\text{SM}} \rightarrow \tau\tau$ 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 modèle 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,

$$q_\mu = -2 \ln \left(\frac{\mathcal{L}(\text{données} | \mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{données} | 0, \hat{\theta}_0)} \right), \quad 0 \leq \mu. \quad (\text{X.47})$$

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 X.5. L'utilisation des catégories SM permet d'avoir une sensibilité spécifique à h avec les modes ggh , VBF et VH.

Catégorie		$ggh, \text{VBF, VH}$	$b\bar{b}h$	H, A
no-btag, $m_{\text{SVFIT}} < 250 \text{ GeV}$	(SM)	✓	✗	✗
no-btag, $m_{\text{SVFIT}} \geq 250 \text{ GeV}$	(BSM)	✗	✗	✓
btag	(BSM)	✓	✓	✓

Tableau X.5 – Présences attendues des contributions au signal dans les catégories.

7.4.2 Scénario $M_{H_1}^{125}$ (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 $M_{H_1}^{125}$ (CPV) correspond à ce cas de figure. Les paramètres fixes sont [20] :

- 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 $\mu = 1,65$ TeV ;
- les paramètres de masse des *gauginos* $M_1 = 1$ TeV, $M_2 = 1$ TeV et $M_3 = 2,5$ 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

$$|A_t| = \mu \cot \beta + 2,8 \text{ TeV}, \quad \phi_{A_t} = \frac{2\pi}{15}, \quad A_b = A_\tau = |A_t|. \quad (\text{X.48})$$

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 [71] à l'aide des méthodes du paramètres d'impact [72], du plan de désintégration [73] ou du vecteur polarimétrique [74], basées sur les propriétés cinématiques des particules de l'état final.

Dans le cadre de l'analyse menée dans cette thèse, le scénario $M_{H_1}^{125}$ (CPV) est testé de manière similaire à M_h^{125} . Cependant, le signal du MSSM considéré est $\Phi \in \{H_1, H_2, H_3\} \rightarrow \tau\tau$. 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 $\tan \beta$ [60] 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 $H/A \rightarrow \tau\tau$, 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 X.21 pour les processus $gg \rightarrow \Phi \rightarrow \tau\tau$ (figure X.21a) et $gg \rightarrow b\bar{b}\Phi \rightarrow \tau\tau$ (figure X.21b). Les précédents résultats obtenus par la collaboration CMS avec les données récoltées en 2016 [13] ainsi que les derniers résultats de la collaboration ATLAS sur l'intégralité du Run II [19] 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

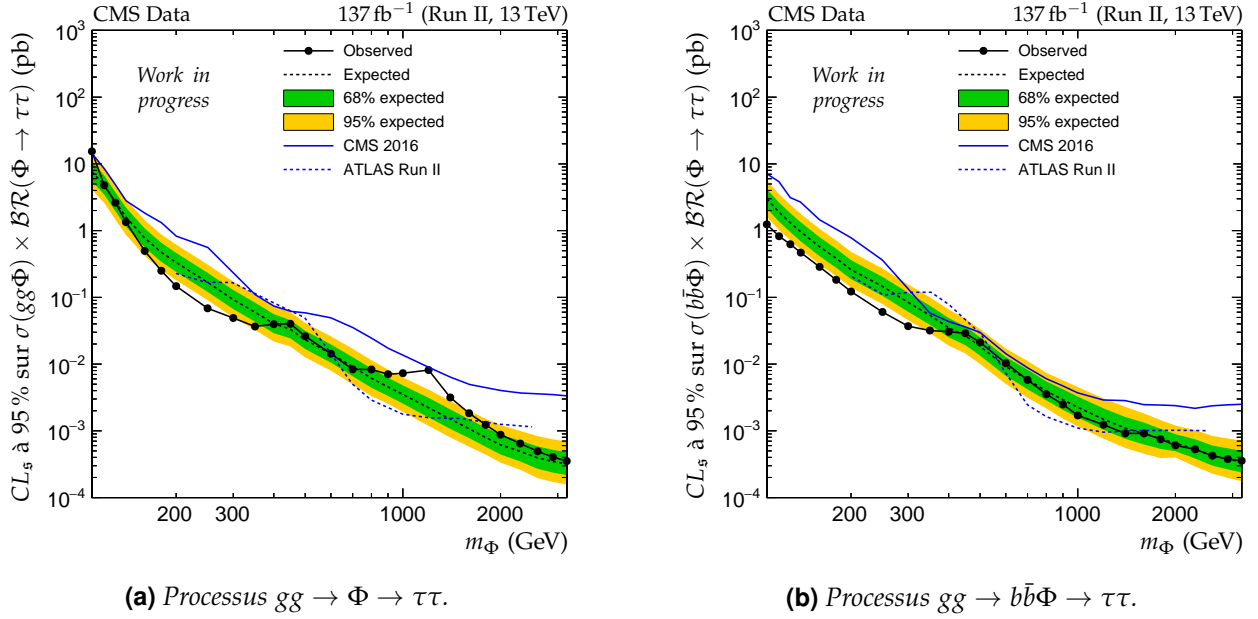


Figure X.21 – Limites d'exclusion indépendantes du modèle physique obtenues avec l'intégralité des données du Run II récoltées par CMS. Les précédents résultats [13] obtenus avec les données de l'année 2016 sont également donnés (CMS 2016) ainsi que ceux de la collaboration ATLAS sur l'intégralité du Run II [19] (ATLAS Run II).

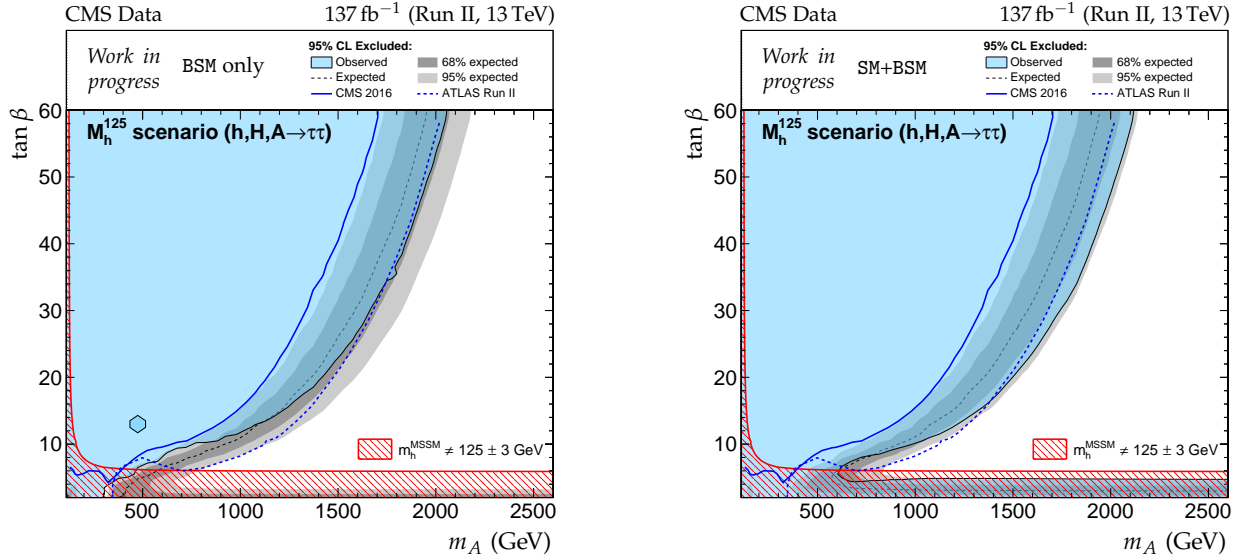
effectivement observées. Lorsqu'elles sont égales, comme c'est le cas pour $m_\Phi \gtrsim 400$ GeV en figure X.21b, les observations sont en accord avec les prédictions du modèle. Lorsque les limites observées sont plus basses que les limites attendues, cela signifie que les données observées permettent 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 $m_T^{\text{tot}} \gtrsim 300$ GeV [23] et donnait des limites d'exclusion observées plus basses que celles attendues, ce qui a été corrigé depuis. Sur les figures X.21a et X.21b, les limites observées sont plus basses d'environ 2σ que celles attendues à basse masse ($m_\Phi \lesssim 300$ 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 X.21a, la limite d'exclusion observée à $m_\Phi = 1200$ GeV est de 8×10^{-3} pb contre 2×10^{-3} pb attendus. Cet écart est d'environ 4σ . Cependant, les derniers résultats de la collaboration ATLAS [19] 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 X.22. Sur la figure X.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 [13]. Dans le cas de la figure X.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 ($m_{h_{\text{MSSM}}} \neq 125 \pm 3$ 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 s_b correspondant au MSSM est rejetée en faveur de l'hypothèse b correspondant au SM d'après les données observées.

La comparaison des figures X.22a et X.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 X.22 – Limites d'exclusion du scénario M_h^{125} obtenues avec l'intégralité des données du Run II récoltées par CMS. Les précédents résultats [13] 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 [19] (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 limites attendues à 2σ peu sensibles à la combinaison des catégories dans cette région de l'espace de phases. Sur la figure X.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 X.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.

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 ($\tau_h \tau_h$, $\mu \tau_h$, $e \tau_h$, $e \mu$).

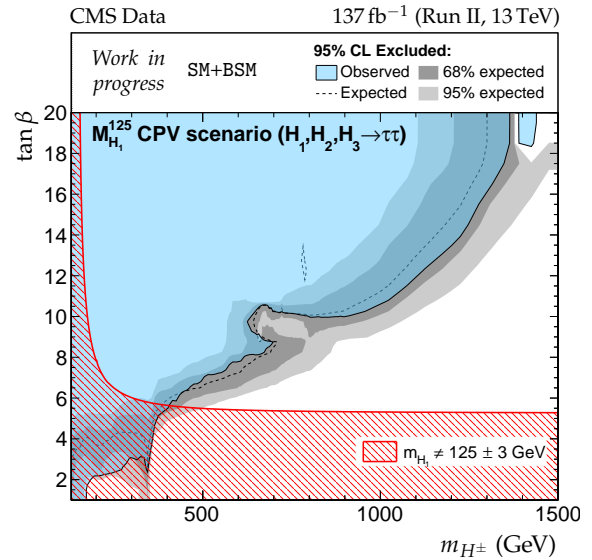


Figure X.23 – Limites d'exclusion du scénario $M_{H_1}^{125}$ (CPV) dans le plan $(m_{H^\pm}, \tan \beta)$ obtenues avec l'intégralité des données du Run II récoltées par CMS.

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 $Z \rightarrow \tau\tau$. 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 ($gg \rightarrow \Phi \rightarrow \tau\tau$) et en association avec des quarks b ($gg \rightarrow b\bar{b}\Phi \rightarrow \tau\tau$). 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 $\ell\tau_h$ ($\mu\tau_h, e\tau_h$) 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, $M_{H_1}^{125}$ (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 ce manuscrit 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) [75] 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.

Références

- [1] 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 B* **716.1** (2012), p. 1-29. DOI : [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).

- [2] 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](https://doi.org/10.1016/j.physletb.2012.08.021).
- [3] 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](https://doi.org/10.1007/jhep06(2013)081).
- [4] 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](https://doi.org/10.1103/physrevlett.114.191803).
- [5] 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](https://doi.org/10.1007/jhep08(2016)045).
- [6] 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](https://doi.org/10.1140/epjc/s2006-02569-7). arXiv : [hep-ex/0602042](https://arxiv.org/abs/hep-ex/0602042) [hep-ex].
- [7] 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](https://doi.org/10.1103/PhysRevLett.103.201801). arXiv : [0906.1014](https://arxiv.org/abs/0906.1014) [hep-ex].
- [8] 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](https://doi.org/10.1016/j.physletb.2011.12.050). arXiv : [1106.4555](https://arxiv.org/abs/1106.4555) [hep-ex].
- [9] 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](https://doi.org/10.1016/j.physletb.2012.05.028). arXiv : [1202.4083](https://arxiv.org/abs/1202.4083) [hep-ex].
- [10] 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](https://doi.org/10.1007/jhep10(2014)160).
- [11] The CMS Collaboration. « Higgs to tau tau (MSSM) ». CMS-PAS-HIG-13-021 (2013). URL : <http://cds.cern.ch/record/1623367>.
- [12] 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>.
- [13] The CMS Collaboration. « Search for additional neutral MSSM Higgs bosons in the di-tau final state in pp collisions at $\sqrt{s} = 13$ TeV ». *Journal of High Energy Physics* **09.007** (sept. 2018). DOI : [10.1007/JHEP09\(2018\)007](https://doi.org/10.1007/JHEP09(2018)007).
- [14] 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](https://doi.org/10.1016/j.physletb.2013.04.017). arXiv : [1302.2892](https://arxiv.org/abs/1302.2892) [hep-ex].
- [15] 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](https://doi.org/10.1007/JHEP11(2015)071). arXiv : [1506.08329](https://arxiv.org/abs/1506.08329) [hep-ex].
- [16] 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](https://doi.org/10.1016/j.physletb.2015.11.042). arXiv : [1508.01437](https://arxiv.org/abs/1508.01437) [hep-ex].
- [17] 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](https://doi.org/10.1007/JHEP02(2013)095). arXiv : [1211.6956](https://arxiv.org/abs/1211.6956) [hep-ex].

- [18] 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 $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector ». *Journal of High Energy Physics* **1** (jan. 2018). DOI : [10.1007/jhep01\(2018\)055](https://doi.org/10.1007/jhep01(2018)055).
- [19] The ATLAS Collaboration. « Search for Heavy Higgs Bosons Decaying into Two Tau Leptons with the ATLAS Detector Using pp Collisions at $\sqrt{s} = 13\text{ TeV}$ ». *Physical Review Letters* **125** (5 juil. 2020), p. 051801. DOI : [10.1103/PhysRevLett.125.051801](https://doi.org/10.1103/PhysRevLett.125.051801).
- [20] 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](https://doi.org/10.1140/epjc/s10052-019-7114-8).
- [21] J. ANDREJKOVIC & coll. « BSM $H \rightarrow \tau\tau$ analysis on full Run 2 CMS data at $\sqrt{s} = 13\text{ TeV}$ ». *CMS analysis Note* (2021). URL : https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2020/218.
- [22] The CMS Collaboration. « MSSM $H/A \rightarrow \tau\tau$ 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>.
- [23] 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>.
- [24] 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>.
- [25] C. BERNET. *Heppy : a python framework for high-energy physics data analysis*. URL : <https://github.com/cbernet/heppy>.
- [26] 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.
- [27] L. TORTEROTOT, C. BERNET & E. AŞILAR. *MergeBinErrors fix on total bin error*. Fév. 2020. URL : <https://github.com/cms-analysis/CombineHarvester/pull/243>.
- [28] 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.
- [29] 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.
- [30] 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](https://doi.org/10.1007/JHEP01(2016)056).
- [31] 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](https://doi.org/10.1007/JHEP01(2016)090).
- [32] 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](https://doi.org/10.1007/jhep06(2010)043). arXiv : [1002.2581 \[hep-ph\]](https://arxiv.org/abs/1002.2581).
- [33] 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](https://doi.org/10.1016/j.cpc.2015.01.024). arXiv : [1410.3012 \[hep-ph\]](https://arxiv.org/abs/1410.3012).
- [34] 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](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [35] 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](https://doi.org/10.1109/tns.2006.869826).

- [36] 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](https://doi.org/10.1016/j.nima.2016.06.125).
- [37] D. JANG. « Search for MSSM Higgs decaying to tau pairs in $p\bar{p}$ collision at $\sqrt{s} = 1,96$ 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>.
- [38] The CMS Collaboration. *Pileup Removal Algorithms*. Rapp. tech. CMS-PAS-JME-14-001. Geneva : CERN, 2014. URL : <https://cds.cern.ch/record/1751454>.
- [39] 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](https://doi.org/10.1088/1126-6708/2008/04/063). arXiv : [0802.1189](https://arxiv.org/abs/0802.1189) [hep-ph].
- [40] 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](https://doi.org/10.1088/1748-0221/13/05/p05011).
- [41] The CMS Collaboration. *DeepJet : deep learning based on physics objects for jet reconstruction*. URL : <https://twiki.cern.ch/twiki/bin/viewauth/CMS/DeepFlavour>.
- [42] 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](https://doi.org/10.1088/1748-0221/14/06/p06032).
- [43] 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>.
- [44] 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.
- [45] 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.
- [46] 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>.
- [47] 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.
- [48] 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.
- [49] The CMS Collaboration. *Properties of the Higgs-like boson in the decay $H \rightarrow ZZ \rightarrow 4\ell$ 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>.
- [50] L. BIANCHINI & coll. « Reconstruction of the Higgs mass in $H \rightarrow \tau\tau$ Events by Dynamical Likelihood techniques ». *Journal of Physics : Conference Series* **513**.2 (juin 2014). DOI : [10.1088/1742-6596/513/2/022035](https://doi.org/10.1088/1742-6596/513/2/022035).
- [51] 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>.
- [52] 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>.

- [53] 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>.
- [54] 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>.
- [55] The CMS Collaboration. *Luminosity Physics Object Group (Lumi POG)*. URL : <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TWikiLUM>.
- [56] 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>.
- [57] The CMS Collaboration. *Recommendation for Using b -tag Objects in Physics Analyses*. URL : <https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation>.
- [58] The CMS Collaboration. *Standard Model Cross Sections for CMS at 13 TeV*. URL : <https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat13TeV>.
- [59] 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>.
- [60] The CMS Collaboration. *MSSM Neutral Higgs*. URL : <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGMSSMNeutral>.
- [61] The CMS Collaboration. *Tau ID recommendations for Run-2*. URL : <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauIDRecommendationForRun2>.
- [62] The CMS Collaboration. « Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements ». *European Physical Journal C* **80** (mar. 2019). URL : <https://cds.cern.ch/record/2669320>.
- [63] The CMS Collaboration. *MET Uncertainties*. URL : https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETRun2Corrections#MET_Uncertainties.
- [64] The CMS Collaboration. *Jet Energy Resolution*. URL : <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>.
- [65] T. JUNK. « Confidence level computation for combining searches with small statistics ». *Nuclear Instruments and Methods in Physics Research A* **434.2-3** (sept. 1999), p. 435-443. DOI : [10.1016/S0168-9002\(99\)00498-2](https://doi.org/10.1016/S0168-9002(99)00498-2). arXiv : [hep-ex/9902006](https://arxiv.org/abs/hep-ex/9902006) [hep-ex].
- [66] 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>.
- [67] A. L. READ. « Presentation of search results : The $CL(s)$ technique ». *Journal of Physics G* **28.10** (sept. 2002), p. 2693-2704. DOI : [10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- [68] 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](https://doi.org/10.1016/0010-4655(93)90005-W).
- [69] J. S. CONWAY. « Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra » (2011), p. 115-120. DOI : [10.5170/CERN-2011-006.115](https://doi.org/10.5170/CERN-2011-006.115).
- [70] 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>.
- [71] 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>.

- [72] 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](https://doi.org/10.1016/j.physletb.2013.11.006).
- [73] 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](https://doi.org/10.1016/j.physletb.2003.10.074).
- [74] 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](https://doi.org/10.21468/SciPostPhysProc.1.018).
- [75] The FCC Collaboration. *Future Circular Collider conceptual design report*. URL : <https://fcc-cdr.web.cern.ch/>.

