

DÉPARTEMENT DE PHYSIQUE
FACULTÉ DES SCIENCES

Étude des leptons taus et recherche de nouvelle physique avec le détecteur CMS

Mémoire présenté en vue de l'obtention du diplôme de master en
sciences physiques

DIEGO BEGHIN

Promotrice : Pr. Barbara CLERBAUX

Co-promotrice : Dr. Cécile CAILLOL

Année académique 2015-2016

Résumé

Deux analyses de données ont été réalisées dans le cadre de ce mémoire. L'une concerne l'étude de l'efficacité de l'algorithme *Hadrons Plus Stips* (HPS) de CMS, dont le rôle est de reconstruire et d'identifier les *taus hadroniques* (τ_h), les leptons taus qui se désintègrent semi-leptoniquement en neutrino tau et hadrons. L'autre analyse consiste en la recherche d'une nouvelle résonance neutre massive dans le canal de désintégration $\tau^+\tau^-$, dans l'état final particulier $\tau_\mu\tau_h$ où un lepton tau se désintègre en muon et neutrinos et l'autre lepton tau est un tau hadronique. Les données utilisées dans ce mémoire sont celles qui ont été récoltées en 2015 par le détecteur CMS situé au collisionneur LHC. Les données sont issues de collisions proton-proton. Elles ont été collectées à une énergie du centre de masse $\sqrt{s} = 13$ TeV et correspondent à une luminosité intégrée de 2.2 fb^{-1} . La première partie de ce mémoire est motivée par la nécessité de connaître l'efficacité de l'algorithme HPS pour toutes les analyses de données de CMS qui étudient des états finaux contenant un ou plusieurs taus hadroniques, comme les recherches de nouvelle physique dans le canal $\tau^+\tau^-$ mais aussi la mesure du couplage du boson scalaire du Modèle Standard aux leptons taus. L'étude de la reconstruction et de l'identification des leptons taus est particulièrement importante car elle est effectuée pour la première fois avec les données de haute énergie ($\sqrt{s} = 13$ TeV), et avec les nouveaux outils de reconstruction et d'identification de CMS. L'efficacité de l'algorithme HPS a d'abord été déterminée dans les simulations, avant que le rapport r de l'efficacité dans les données sur l'efficacité dans les simulations ne soit calculé. Ce calcul a été réalisé par la méthode du rapport $Z \rightarrow \tau_\mu\tau_h / Z \rightarrow \mu\mu$, qui consiste en un ajustement simultané de deux régions des données, l'une riche en événements $Z \rightarrow \tau_\mu\tau_h$ et l'autre riche en événements $Z \rightarrow \mu\mu$. Des valeurs de r ont été trouvées pour les trois sélections standards de l'algorithme HPS, respectivement $r = 0.929 \pm 0.055$, $r = 0.924 \pm 0.058$ et $r = 0.911 \pm 0.053$ pour les sélections *loose*, *medium* et *tight*. La deuxième analyse de ce mémoire est justifiée par l'existence de nombreux modèles de nouvelle physique qui prédisent l'existence d'une nouvelle résonance neutre massive. Ces modèles cherchent à résoudre des problèmes non-résolus dans le Modèle Standard, comme la nature des constituants de la matière noire et l'unification de la gravitation aux trois autres forces fondamentales. Le signal utilisé dans cette deuxième analyse est un boson Z' du Modèle Standard Séquentiel (MSS). Le spectre de masse $\tau_\mu\tau_h$ a été étudié et aucun signe de nouvelle physique n'a été observé : le nombre d'événements dans les données est en accord avec les prédictions du Modèle Standard. Une méthode statistique a été utilisée afin de déterminer les limites attendue et observée à un niveau de confiance de 95% sur la masse du boson Z' du MSS, en tenant compte des résultats du calcul de l'efficacité de l'algorithme HPS réalisé précédemment. La limite attendue obtenue est de 1.37 ± 0.13 TeV, et la limite observée est de 1.18 TeV.

Mots clés : physique des particules, nouvelle physique, boson Z' , leptons taus, CERN, LHC, CMS.

Abstract

Two data analyses are presented in this masters thesis. The first one is a study of the CMS *Hadrons Plus Strips* (HPS) algorithm, which reconstructs and identifies tau leptons that decay into a tau neutrino and one or more hadrons. Those are the so-called *hadronic taus* (τ_h). The second analysis is a search for a new neutral massive resonance in the $\tau^+\tau^-$ decay channel, in the particular final state $\tau_\mu\tau_h$ where one of the tau leptons is a hadronic tau and the other decays into a muon and two neutrinos. The data used in this thesis were collected in the year 2015 by the CMS detector at the LHC collider. The data were collected in proton-proton collisions at a center-of-mass energy $\sqrt{s} = 13$ TeV, and correspond to an integrated luminosity of 2.2 fb^{-1} . The first part of this thesis is motivated by the essential role the HPS algorithm efficiency plays in all CMS data analyses that study final states containing one or more hadronic taus, such as new physics searches in the $\tau^+\tau^-$ channel and the measure of the Standard Model scalar boson coupling to tau leptons. The study of the tau lepton reconstruction and identification efficiency is especially important because this is the first time it is done with high energy ($\sqrt{s} = 13$ TeV) data and the new CMS identification and reconstruction tools. The HPS algorithm efficiency was first determined in the simulations, and the ratio r of the efficiency in the data over the efficiency in the simulations was calculated, using the $Z \rightarrow \tau_\mu\tau_h / Z \rightarrow \mu\mu$ ratio method. This method determines the value of r by a simultaneous adjustment of event yields in two data regions, one of them rich in $Z \rightarrow \tau_\mu\tau_h$ events and the other rich in $Z \rightarrow \mu\mu$ events. Three values of r were calculated, one for each working point of the HPS algorithm. The values $r = 0.929 \pm 0.055$, $r = 0.924 \pm 0.058$ and $r = 0.911 \pm 0.053$ were found for, respectively, the loose, medium and tight working points. The second analysis presented in this thesis is justified by the existence of numerous new physics models which predict the existence of a new neutral massive resonance. These models were built in order to solve some of the open problems left by the Standard Model, such as the nature of the dark matter components and the unification of gravity to the three other fundamental forces. The signal used as a reference in this second analysis is the Sequential Standard Model (SSM) Z' boson. The $\tau_\mu\tau_h$ invariant mass spectrum was studied and no sign of new physics was found : the number of data events is in agreement with Standard Model predictions. A statistical method was used in order to determine both the expected and observed 95% confidence limits for the SSM Z' boson mass, taking into account the results obtained previously for the HPS algorithm efficiency. The expected limit is 1.37 ± 0.13 TeV, and the observed limit is 1.18 TeV.

Keywords : particle physics, new physics, Z' boson, tau leptons, CERN, LHC, CMS.

Remerciements

Je remercie de tout coeur Barbara Clerbaux, pour son travail d'encadrement, pour m'avoir soutenu lors de la présentation de mes résultats au groupe de travail de CMS sur les leptons taus, pour avoir relu le manuscrit ici présent et suggéré des améliorations, et surtout pour m'avoir motivé à travailler sur ce mémoire et à poursuivre d'autres projets d'avenir. J'ai pu profiter de toute son expertise en analyses de données de physique des particules.

Je remercie aussi Cécile Caillol, pour avoir répondu à mes nombreuses questions sur les leptons taus, pour avoir encadré mon stage en juillet dernier au CERN, qui a servi d'inspiration pour ce mémoire, et aussi pour avoir relu ce manuscrit. Son expertise sur les leptons taus m'a été d'une aide précieuse tout au long de l'année.

Je tiens à remercier Jean-Marie Frère, pour son cours sur la physique au-delà du Modèle Standard, pour son aide concernant la rédaction de la partie théorique de ce manuscrit et pour avoir accepté d'être l'un des lecteurs de ce mémoire. Je remercie aussi Julian Heeck pour la discussion fructueuse que nous avons eue sur les modèles de nouvelle physique où les leptons taus jouent un rôle important.

Je remercie Gilles de Lentdecker pour avoir accepté d'être l'un de mes lecteurs. Son cours sur la détection des particules et l'acquisition et l'analyse de données, donné conjointement avec Pascal Vanlaer, m'a également été fort utile lorsqu'il m'a fallu décrire le détecteur CMS.

Je remercie Laurent Thomas, qui m'a partagé son code de calcul de la section efficace des bosons Z' du Modèle Standard Séquentiel, et d'avoir répondu à toutes mes questions concernant l'utilisation du code. Sans lui, la dernière partie de ce mémoire n'aurait pas été possible.

Je suis aussi très reconnaissant envers tous mes collègues de bureau, passés et présents. Aidan Randle-Conde, Giuseppe Fasanella et David Vannerom m'ont apporté une aide essentielle en partageant leurs trucs et astuces sur l'utilisation de ROOT et de latex.

Table des matières

Table des abréviations	8
Introduction	12
1 Contexte	14
1.1 Contexte théorique	14
1.1.1 Le modèle standard de la physique des particules	14
1.1.2 Modèles BSM et nouvelles résonances neutres massives	17
1.1.3 Modèle Standard Séquentiel et recherche expérimentale d'un boson Z'	21
1.2 Le collisionneur LHC	23
1.2.1 Le complexe d'accélérateurs du CERN	23
1.2.2 Paramètres du LHC	25
1.2.3 Design du LHC	27
1.2.4 Calendrier du LHC	28
1.3 Le détecteur CMS	29
1.3.1 Le détecteur de traces	30
1.3.2 Les calorimètres	31
1.3.3 Les chambres à muons	32
1.3.4 Le système de déclenchement	32
1.4 Données et simulations	34
1.4.1 Données	34
1.4.2 Simulations	36
1.5 Algorithmes de reconstruction et d'identification de particules de CMS	38
1.5.1 Muons	39
1.5.2 Électrons	39
1.5.3 Jets et énergie transverse manquante	40
1.5.4 Leptons taus	41
2 Étude de l'efficacité de l'algorithme de reconstruction et d'identification des leptons taus	48
2.1 Détermination de l'efficacité dans les simulations	48
2.1.1 Cinématique des événements tau-antitau	49
2.1.2 Efficacité en fonction de l'impulsion transverse générée du τ_h	51
2.2 Détermination de l'efficacité à partir des données	51
2.2.1 Sélections dans la région $\tau_\mu\tau_h$	54
2.2.2 Détermination du facteur d'échelle pour les processus W +jets	56
2.2.3 Estimation du processus multijets à partir des données	60
2.2.4 Histogrammes de contrôle dans le canal $\tau_\mu\tau_h$	62
2.2.5 Sélections dans la région $\mu\mu$	62

2.2.6	Ajustement du rapport entre efficacité dans les données et efficacité dans les simulations	63
3	Spectre à haute masse invariante du système $\tau_\mu\tau_h$	70
3.1	Sélection des événements $\tau_\mu\tau_h$	70
3.2	Choix de la variable de masse reconstruite	70
3.2.1	Présentation des variables	71
3.2.2	Comparaison entre les variables	72
3.3	Analyse des données	78
3.3.1	Méthode statistique utilisée	80
3.3.2	Acceptance et efficacité	82
3.3.3	Limites sur la section efficace et conséquences pour les modèles prédisant un boson Z'	82
	Conclusion et perspectives	85
	Bibliographie	87

Table des abréviations

η Pseudo-rapacité. $\eta = -\ln(\tan \frac{\theta}{2})$, où θ est l'angle polaire dans le référentiel de CMS. Pour des particules sans masse, les différences de pseudo-rapacité sont invariantes sous des boosts dans la direction du tube à vide.

ϕ Angle azimutal dans le référentiel de CMS, définit une direction dans le plan transverse (le plan perpendiculaire au tube à vide).

τ_μ Tau muonique, désigne un lepton tau qui se désintègre en muon et neutrinos.

τ_e Tau électronique, désigne un lepton tau qui se désintègre en électron et neutrinos.

τ_h Tau hadronique, désigne un lepton tau qui se désintègre en hadrons et neutrino.

Anti- k_T Algorithme de reconstruction de jets de hadrons utilisé dans la plupart des analyses de données au LHC.

ATLAS A Toroidal LHC ApparatuS. Détecteur multifonction installé au LHC, similaire à CMS.

BSM Beyond the Standard Model. Expression qui désigne tout modèle prédisant une nouvelle physique au-delà du Modèle Standard de la physique des particules.

CERN Organisation Européenne pour la Recherche Nucléaire. Organisation scientifique européenne ayant pour mission la recherche en physique nucléaire et en physique des particules. On y trouve notamment le collisionneur LHC.

CMS Compact Muon Solenoid. Détecteur multifonction installé au collisionneur LHC, où les données analysées dans ce mémoire ont été collectées.

DAQ Data Acquisition System. Système électronique d'acquisition de données, qui permet de transformer les interactions des particules avec le détecteur en données qui puissent être utilisées par un ordinateur.

DMF Decay Mode Finding. Composante de l'algorithme HPS qui reconstruit le mode de désintégration du tau hadronique (τ_h), en fonction du nombre de "strips" (correspondant à des π^0) et de hadrons chargés trouvés.

ECal Electromagnetic Calorimeter. Calorimètre électromagnétique, mesure en particulier l'énergie des photons et électrons incidents. Cette mesure est destructive, la particule

initiale n'existe plus à la sortie du calorimètre.

HCal Hadronic Calorimeter. Calorimètre à hadrons, mesure en particulier l'énergie des particules interagissant par la force forte, c'est-à-dire les hadrons. Cette mesure est destructive (cf. ECal).

HL-LHC High Luminosity LHC. Renouvellement prévu du collisionneur LHC afin de lui permettre de collecter des données à une luminosité instantanée plus élevée.

HLT High Level Trigger. Deuxième système de déclenchement de CMS. Sur base des données de tous les sous-détecteurs de CMS, il prend la décision de garder ou non les événements acceptés par le premier système de déclenchement. Les événements acceptés par le HLT sont définitivement sauvegardés et disponibles pour les analyses de données.

HPS Hadrons Plus Strips. Algorithme utilisé dans les analyses de CMS pour la reconstruction et l'identification des taus hadroniques (τ_h). L'étude de l'efficacité de cet algorithme est l'un des sujets principaux de ce mémoire.

LHC Large Hadron Collider. Collisionneur du CERN où se trouve le détecteur CMS. Il s'agit du plus grand collisionneur de particules jamais construit et permettra à terme d'atteindre des énergies du centre de masse de 14 TeV pour les collisions proton-proton. Les collisions de 2015 ont été réalisées à une énergie de 13 TeV.

LO Leading Order. Ordre dominant d'un calcul en théorie des perturbations.

MC Monte-Carlo. Désigne les simulations d'événements générées par une méthode Monte-Carlo, c'est-à-dire l'utilisation répétée de procédés aléatoires. Les simulations Monte-Carlo sont universellement utilisées dans les analyses de physique de particule.

MET Missing Transverse Energy. Énergie transverse manquante. Lors d'une collision au LHC, la quantité de mouvement totale dans le plan transverse (\vec{p}_T^{tot}) est nulle. Si la \vec{p}_T^{tot} mesurée est non-nulle, un vecteur d'énergie transverse manquante est construit afin que la somme vectorielle $\vec{p}_T^{tot} +$ vecteur de MET soit nulle. Le plan transverse est le plan perpendiculaire au tube à vide.

MS Modèle Standard de la physique des particules. Modèle décrivant l'ensemble des particules élémentaires découvertes jusqu'à présent et leurs interactions par les forces électromagnétique, faible et forte. Aucune description de la gravitation n'y est incluse. Jusqu'à présent, aucune expérience n'a permis de le remettre en question.

MSS Modèle Standard Séquentiel. Modèle postulant l'existence d'un boson Z' ayant les mêmes couplages que le boson Z du MS, mais plus lourd. Ce modèle n'est motivé par aucun argument théorique particulier, il est simplement utilisé afin de comparer les différentes expériences qui recherchent l'existence d'un boson Z' .

NLO, NNLO, etc. Next-to-Leading Order, Next-to-Next-to-Leading Order, etc. Ordre sous-dominant et ordre sous-sous-dominant, etc. d'un calcul en théorie des perturbations.

PDF Probability Density Function. Fonction de distribution de probabilité.

PF Particle Flow. Algorithme utilisé dans les analyses de données de CMS, qui réalise une reconstruction complète de chaque événement à partir des informations de tous les sous-détecteurs.

p_T Quantité de mouvement transverse. Il s'agit de la quantité de mouvement dans le plan perpendiculaire à la direction du tube à vide.

PU Pile Up. Empilement, ou nombre de collisions proton-proton simultanées lors d'un croisement de faisceaux dans CMS.

RF RadioFréquence. Les cavités RF servent à accélérer les faisceaux dans un accélérateur de particules.

ROC Les courbes ROC sont utilisées lorsque l'on cherche à déterminer la meilleure coupure sur une certaine variable. Elles permettent d'exprimer l'efficacité du signal en fonction de l'efficacité du bruit de fond, pour différentes coupures appliquées sur cette variable.

WP Working Point. Sélection standard. Les algorithmes d'identification de particules ont souvent plusieurs sélections standards au lieu d'une seule, car les meilleures valeurs pour l'efficacité du signal et du bruit de fond peuvent varier en fonction de l'analyse de données réalisée. Chaque sélection standard correspond alors à des valeurs différentes pour les efficacités du signal et du bruit de fond.

Introduction

Le développement des accélérateurs de particules à partir des années 1950 a permis d'étudier en profondeur les constituants élémentaires de la matière. Les découvertes expérimentales réalisées dans cette période ont motivé l'élaboration de modèles théoriques permettant de les expliquer. Un modèle particulier, développé dans le courant des années 1960 et 1970, a connu un succès expérimental exceptionnel : le Modèle Standard de la physique des particules. Aucune expérience réalisée jusqu'à présent ne permet de le remettre fondamentalement en question. Cependant, il existe de bonnes raisons de penser que le Modèle Standard n'est pas le dernier mot en physique des particules.

Tout d'abord, le Modèle Standard ne décrit pas la gravitation, l'une des quatre forces fondamentales de la nature. Les effets de la gravitation sont négligeables pour les interactions entre particules aux échelles d'énergie qui ont été étudiées expérimentalement jusqu'à présent, mais il existe une échelle d'énergie où la gravitation devient comparable aux trois autres forces. L'un des principaux sujets de recherche en physique théorique est un modèle décrivant les quatre forces fondamentales dans un même cadre théorique. Des théories plus fondamentales sont aussi motivées par le problème de la matière noire (quels sont les particules qui composent la matière noire ?) et des problèmes plus techniques comme le problème de la hiérarchie (pourquoi la gravitation est-elle tellement plus faible que les autres forces ?) et le nombre de paramètres libres du Modèle Standard que certains jugent excessif. Ces théories impliquent typiquement l'existence d'interactions supplémentaires entre les particules. L'existence de ces interactions peut être testée expérimentalement : en physique des particules chaque force est transmise par une ou plusieurs particules médiatrices appelées bosons, de nouvelles forces impliquent l'existence de nouveaux bosons. Une analyse statistique des données permet de déterminer si ces nouveaux bosons existent ou pas.

Le collisionneur LHC a commencé en 2015 à réaliser des collisions proton-proton à une énergie du centre de masse record de 13 TeV. L'énergie la plus élevée atteinte précédemment dans un collisionneur était de 8 TeV en 2012 au LHC. Si les particules massives prédites par les modèles de nouvelle physique existent, ce gain d'énergie rend leur production plus probable. Ce moment est donc idéal pour tester leur existence.

La partie finale de ce mémoire est consacrée à la recherche d'un nouveau boson, appelé Z' , dans les données du détecteur CMS, l'un des quatre détecteurs principaux installés dans le collisionneur LHC. L'appellation *boson Z'* désigne toute une classe de particules neutres, semblables au boson Z du Modèle Standard, et prédites par des modèles de nouvelle physique. Les modèles prédisent que le boson Z' doit avoir un temps de vie très court et qu'il se désintègre, entre autres, en paires lepton-antilepton (l^+l^-). Ces canaux de désintégration sont intéressants à étudier car ils se caractérisent par un état final simple et ne souffrent que d'une faible contamination de bruit de fond provenant de processus connus du Modèle Stan-

dard. Comparé aux canaux $Z' \rightarrow e^+e^-$ et $Z' \rightarrow \mu^+\mu^-$, le canal $Z' \rightarrow \tau^+\tau^-$ est plus complexe et souffre d'une contamination plus importante d'événements de bruit de fond. Cependant, il est important de ne pas limiter l'analyse aux canaux les plus simples, en effet le boson Z' pourrait a priori avoir des taux de désintégration différents dans chacun de ces canaux, certaines théories prédisent par exemple l'existence d'un boson Z' qui ne se couple pas du tout aux électrons ou aux muons.

Le canal $\tau^+\tau^-$ est celui qui a été choisi pour la recherche du boson Z' dans ce mémoire. Les leptons taus sont des particules relativement difficiles à identifier, car contrairement aux électrons, qui sont des particules stables, et les muons, qui ont un temps de vie suffisamment long pour traverser l'ensemble du détecteur, les leptons taus ont une durée de vie de $2.9 \cdot 10^{-13}$ s, et ils se désintègrent à l'intérieur du tube à vide du détecteur. Ce ne sont donc pas les leptons taus eux-mêmes qui sont reconstruits mais leurs produits de désintégration, et des algorithmes particuliers sont nécessaires pour remonter des produits de désintégration reconstruits à l'identification des leptons taus. L'étude du fonctionnement de ces algorithmes de reconstruction et d'identification est un sujet de recherche en soi. Les applications sont nombreuses et concernent toutes les analyses de données qui considèrent des événements avec un ou plusieurs leptons taus dans l'état final. Le deuxième chapitre de ce mémoire étudie l'un de ces algorithmes, appelé *Hadrons Plus Strips*. Cet algorithme est utilisé dans les analyses de données de CMS afin de reconstruire et d'identifier les leptons taus dont les produits de désintégration contiennent un ou plusieurs hadrons. Ces leptons taus sont appelés taus hadroniques. L'efficacité de leur reconstruction et de leur identification par l'algorithme HPS est étudiée en détail, dans les simulations et les données. Les résultats finaux de cette étude sont les différentes valeurs du rapport r de l'efficacité dans les données sur l'efficacité dans les simulations, en fonction du point de fonctionnement choisi pour l'algorithme HPS. Ce travail a été réalisé en étroite collaboration avec le groupe d'étude des leptons taus de CMS.

Les résultats de l'étude de l'efficacité du chapitre 2 sont ensuite utilisés dans le chapitre 3 pour la recherche du boson Z' . Une analyse statistique est réalisée afin de poser une limite sur la section efficace de production du boson Z' et d'exclure les modèles incompatibles avec les données.

Avant de passer au chapitre 1, précisons quelques conventions utilisées dans ce mémoire :

- La distinction entre particules et antiparticules n'est pas toujours faite. Sauf mention contraire, le nom de la particule est utilisé indifféremment pour les particules et antiparticules. Par exemple, le mot "électrons" désigne tout aussi bien les électrons (de charge négative) que les antiélectrons (de charge positive).
- Les unités dites naturelles sont utilisées, où $c = \hbar = 1$. Dans les unités du système internationale, la vitesse de la lumière dans le vide c vaut $2.998 \cdot 10^8$ m \cdot s $^{-1}$ et la constante de Planck réduite \hbar vaut $1.055 \cdot 10^{-34}$ kg m 2 s $^{-1}$. Les unités naturelles sont utilisées par commodité, elles permettent notamment d'exprimer les masses et quantités de mouvement en unités d'énergie. Il est toujours possible de convertir une valeur exprimée en unités naturelles en une valeur exprimée en unités SI, en ajoutant les facteurs c et/ou \hbar appropriés.
- Les énergies sont exprimées en multiples d'électron-Volts (eV), comme c'est la convention en physique des particules. La facteur de conversion entre l'électron-Volt et le Joule, unité SI, est $1 \text{ eV} = 1.602 \cdot 10^{-19}$ J.

Chapitre 1

Contexte

Ce premier chapitre explique le contexte théorique et expérimental des analyses réalisées dans les chapitres 2 et 3. La première section introduit le Modèle Standard de la physique des particules, et les modèles de nouvelle physique qui motivent les recherches réalisées dans la suite de ce mémoire. Ensuite, le collisionneur LHC et le détecteur CMS sont présentés. L'avant-dernière section concerne les données et les simulations utilisées dans les analyses des chapitres suivants, et la dernière section explique les algorithmes de reconstruction et d'identification des particules étudiés dans ce mémoire, avec une attention toute particulière à l'algorithme concernant les leptons taus.

1.1 Contexte théorique

1.1.1 Le modèle standard de la physique des particules

Les propriétés des particules élémentaires et leurs interactions sont décrites par le Modèle Standard (MS) de la physique des particules. Le MS est une théorie quantique des champs, c'est-à-dire c'est une théorie où les objets fondamentaux sont des champs, chaque particule correspond à l'excitation du champ qui lui est associé. Le MS est aussi une théorie relativiste, compatible avec la relativité restreinte d'Einstein.

Trois interactions fondamentales sont modélisées par le MS : l'électromagnétisme, la force faible et la force forte. Le MS n'inclut pas de description de la gravitation, dont les effets sont négligeables dans toutes les expériences de physique des particules réalisées jusqu'à présent, que ce soit à l'aide d'accélérateurs, de collisionneurs ou de rayons cosmiques. Chacune des trois forces du MS est transmise par un ou plusieurs bosons vectoriels appelés *bosons de jauge*. Elles n'agissent pas forcément sur toutes les particules du MS.

- L'**électromagnétisme** agit sur toutes les particules qui portent une charge électrique et est transmis par le photon, une particule neutre et sans masse.
- La **force faible** est transmise par trois bosons différents : le W^+ (chargé positivement), son antiparticule le W^- (chargé négativement), et le boson Z (neutre). Ces trois bosons sont massifs, ce qui limite la portée de la force faible. Celle-ci a en outre la particularité de ne pas conserver la parité (invariance par réflexion dans un miroir).
- La **force forte** n'agit que sur les particules qui portent une charge de couleur, et est transmise par huit gluons différents. Les gluons sont tous neutres et sans masse, par

contre ils portent chacun une charge de couleur différente. La constante de couplage des interactions fortes a la particularité de devenir plus faible à des échelles d'énergie plus élevées. À basse énergie ($\lesssim 1$ GeV), sa valeur est trop grande pour qu'un calcul perturbatif soit possible, mais à de plus hautes énergies, sa valeur diminue et l'approche perturbative est justifiée.

La matière est incluse dans la théorie par des fermions de spin 1/2 : les leptons et les quarks.

- Les **leptons** sont soit chargés (c'est le cas des électrons, muons et taus) soit neutres (c'est le cas des neutrinos). Tous les leptons sont sensibles à la force faible, et ceux qui sont chargés sont aussi sensibles à la force électromagnétique. Aucun lepton n'est sensible à la force forte.
- Il y a 6 types (ou *savours*) de **quarks** différents, et chacune de ces saveurs vient en trois couleurs différentes, pour un total de 18 quarks. Trois des saveurs ont une charge $+2/3$ et les trois autres une charge $-1/3$. Les quarks sont sensibles aux trois interactions du MS : la force forte (car les quarks portent une charge de couleur), la force faible et l'électromagnétisme.

Les fermions sont distribués en trois générations différentes. Chaque génération comporte un lepton chargé et le neutrino qui lui est associé, trois quarks de charge $+2/3$ (un de chaque couleur) et trois quarks de charge $-1/3$ (idem). Les particules de même type de deux génération différentes, par exemple le muon et le tau, ont la même charge, sont sensibles aux mêmes forces, mais ont une masse et un temps de vie différents. Les trois générations sont classées en ordre croissant de masse.

À basse énergie, la force forte entre dans son régime non-perturbatif, et confine les quarks à l'intérieur de particules composites, les *hadrons*. Les quarks ne sont jamais observés individuellement, seulement les hadrons qu'ils composent sont accessibles à l'étude expérimentale. Les gluons individuels ne peuvent pas non plus être observés et plus généralement, aucune charge de couleur n'est observée directement à cause de ce phénomène de confinement.

Toutes les particules élémentaires du MS sont présentées dans la table 1.1. Chacune de ces particules vient avec sa propre antiparticule de même masse, même temps de vie, mais avec une charge opposée. Le photon, le boson Z et le boson H sont leurs propres antiparticules.

Le MS a été construit progressivement au cours des années 1960 et 1970. Salam et Ward proposent en 1964 un modèle d'interaction de jauge $SU(2) \times U(1)$ pour les forces électromagnétique et faible [1], mais il ne contient pas encore de mécanisme satisfaisant pour briser la symétrie $SU(2)$ associée à l'interaction faible. Ce problème est résolu la même année par le mécanisme de Brout-Englert-Higgs (BEH) [2, 3] de brisure spontanée de symétrie. La symétrie est brisée par l'introduction d'un doublet scalaire complexe (portant quatre composantes) ayant une valeur non-nulle dans le vide. Trois des composantes de ce doublet sont absorbées dans une redéfinition des bosons de jauge faibles (qui acquièrent alors une masse) et la quatrième composante correspond au boson H. Dans les années suivantes une théorie complète des interactions électromagnétique et faible, appelée théorie électrofaible et incorporant le mécanisme BEH de brisure spontanée de symétrie, est construite par Weinberg [4], Salam et Glashow. L'autre composante du MS est la chromodynamique quantique

		Fermions			Bosons de jauge	Boson scalaire
		Génération I	Génération II	Génération III		
Quarks	u up 2/3 $\simeq 2$ MeV	c charm 2/3 1.3 GeV	t top 2/3 173 GeV	W^+ boson W 1 80 GeV	H boson H 0 125 GeV	
	d down -1/3 $\simeq 4$ MeV	s strange -1/3 $\simeq 100$ MeV	b bottom -1/3 4.7 GeV	Z boson Z 0 91 GeV		
Leptons	e électron -1 511 keV	μ muon -1 106 MeV	τ tau -1 1.777 GeV	γ photon 0 0		
	ν_e neutrino électronique 0 < 1 eV	ν_μ neutrino muonique 0 < 1 eV	ν_τ neutrino tau 0 < 1 eV	g gluon 0 0		

TABLE 1.1 – Particules du Modèle Standard. Au sein de chaque case, la première ligne donne le symbole de la particule, la deuxième son nom, la troisième sa charge électrique (en multiple de la charge élémentaire $e \simeq 1,6 \cdot 10^{-19}$ C), et la quatrième sa masse. Chacune des 6 saveurs des quarks existe sous trois couleurs différentes.

(QCD), une théorie de jauge SU(3) décrivant l'interaction forte entre quarks, introduite pour la première fois par Nambu et Han [5] en 1965. Le groupe de jauge total du MS est alors $SU(3)_C \times SU(2)_L \times U(1)_Y$, où $SU(3)_C$ représente la symétrie de jauge associée à la QCD, et $SU(2)_L \times U(1)_Y$ celle de la théorie électrofaible.

Le MS n'a jusqu'à présent pas été directement remis en question par des tests expérimentaux. La seule exception est due aux masses non-nulles des neutrinos, qui ne sont pas incluses dans le MS. Cependant, une extension minimale du MS est possible, avec l'introduction de termes de masses pour les neutrinos de la même façon que pour les autres fermions, à l'aide d'une interaction avec le champ de Brout-Englert-Higgs. Toutes les particules qui composent le MS ont été découvertes, la dernière en date étant le boson H, découvert en 2012 par les collaborations ATLAS [6] et CMS [7], dont les détecteurs se situent auprès du collisionneur LHC. Les masses et les couplages des particules ont été mesurés, et aucune incompatibilité avec le MS n'a été trouvée.

Cependant, certains problèmes encore non résolus suggèrent l'existence de physique au-delà du MS :

1. Le MS ne modélise pas la gravitation, l'une des quatre interactions fondamentales de la Nature. Cela ne pose pas de problème pour toutes les expériences réalisées jusqu'à

présent, où les effets de la gravitation sont négligeables, mais il existe une échelle d'énergie (la masse de Planck) où la gravitation et les trois autres interactions ont une force comparable. Un modèle plus fondamental est nécessaire afin de décrire les quatre interactions fondamentales dans le même cadre théorique.

2. Le problème de la hiérarchie. La masse de Planck ($\sim 10^{19}$ GeV) est plusieurs ordres de grandeur plus élevée que celle de la brisure spontanée de la symétrie électrofaible par le mécanisme de BEH (environ 100 GeV). Certains théoriciens considèrent cet état des lieux peu naturel, en particulier en ce qui concerne le problème de la renormalisation de la masse du boson H. Les boucles quantiques donnent des corrections quadratiques à cette masse. À cause de l'énorme différence d'énergie entre l'échelle de Planck et l'échelle électrofaible, la renormalisation requiert l'ajustement précis de certains paramètres, ce qui rend la théorie non-naturelle [8]. Notons cependant que d'autres théoriciens interprètent l'ajustement précis comme une conséquence fortuite du choix de la procédure de renormalisation, et considèrent qu'il n'a aucune réalité physique [9].
3. La matière noire. Des effets gravitationnels montrent que la matière visible ne représente qu'une partie (environ 20%) de la totalité de la matière dans l'univers. La matière non-visible, dite noire, ne peut pas être composée des particules connues du MS, cette possibilité est notamment exclue par l'étude des anisotropies du rayonnement cosmologique de fond (CMB)[10]. Les candidats principaux pour la matière noire sont des nouvelles particules appelées *WIMP* (Weakly Interacting Massive Particles), des particules stables lourdes qui n'interagissent avec les particules du MS que par la force faible (et la gravité).
4. L'asymétrie matière-antimatière de l'univers. Le modèle cosmologique standard prédit que la matière et l'antimatière sont produites en quantité égales peu après le Big Bang. Bien que des processus introduisant des asymétries entre matière et antimatière existent dans le MS, ils ne sont pas suffisants pour expliquer la totalité de l'asymétrie observée [11].
5. Il y a au moins 19 paramètres libres [12] dans le MS, plus si l'on inclut des termes de masse pour les neutrinos. Le nombre de générations, trois, n'est pas expliqué. Existe-t-il une théorie des interactions fondamentales avec moins de paramètres libres ?

De nombreux modèles de physique au-delà du Modèle Standard (BSM) ont été construits pour répondre à certaines de ces questions.

1.1.2 Modèles BSM et nouvelles résonances neutres massives

De nombreux modèles BSM prédisent l'existence de nouveaux bosons neutres massifs : notamment les théories de grande unification (*Grand Unified Theory* - GUT), les théories avec des dimensions spatiales supplémentaires, et les théories supersymétriques (SUSY). Il est aussi possible de construire des théories où un nouveau boson neutre se couple préférentiellement aux leptons taus.

Théories de grande unification

Les théories GUT prédisent l'unification des trois forces du MS (électromagnétisme, force faible, force forte) en une force plus fondamentale, à des échelles d'énergie de grande unification, typiquement de l'ordre de 10^{15} - 10^{16} GeV [13]. La nouvelle force est due à un nouveau

groupe de jauge G , qui inclut le groupe $SU(3) \times SU(2) \times U(1)$ du MS. Il n'y aurait plus trois couplages différents et indépendants, un pour chaque force du MS, mais un seul couplage associé à la nouvelle force. La recherche sur les théories GUT était à l'origine motivée par la convergence apparente entre les couplages des trois forces du MS¹ à des énergies élevées. Aujourd'hui, on sait que cette triple convergence n'est pas exacte si l'on ne considère que le MS, mais elle devient possible si l'on considère une extension SUSY (cf. ci-dessous) du MS.

La symétrie associée au groupe G doit être brisée à une certaine échelle d'énergie, mais un groupe plus large que celui du MS peut subsister, par exemple un groupe $U(1)$ supplémentaire, et être brisé à plus basse énergie. À cette symétrie $U(1)$ supplémentaire est associée un nouveau boson de jauge, appelé Z' par analogie au boson Z du modèle standard. La masse de ce boson Z' est typiquement un paramètre libre du modèle. Dans le cas général, les couplages de ce nouveau boson aux fermions dépendent de la génération des fermions et de leur chiralité (gauche ou droite). Il y a donc a priori 24 couplages différents, mais les couplages mélangeant plusieurs générations (appelés *couplages non-universels*) peuvent donner lieu à des FCNC (*Flavor Changing Neutral Currents* - courants neutres violant la saveur) via une boucle de Z' virtuel. Les FCNC sont fortement contraints par l'expérience, ce qui impose soit une limite inférieure très élevée à l'échelle de nouvelle physique soit un mécanisme fort de suppression des FCNC [14].

Les nouveaux bosons neutres massifs prédits par les théories GUT ont donc un spin 1, mais leur masse et leurs 24 couplages aux fermions varient selon les modèles.

Théories avec des dimensions spatiales supplémentaires

Certaines théories cherchent à résoudre le problème de la hiérarchie entre l'échelle d'énergie de la gravitation et celle de la brisure de l'interaction électrofaible en postulant l'existence de dimensions spatiales supplémentaires compactifiées (de dimension finie). Toutes les forces et les particules du MS sont confinées dans un espace-temps de 4 dimensions, la gravitation étant la seule force pouvant se propager dans les dimensions supplémentaires. Celles-ci "diluent" la gravitation par rapport aux autres forces, ce qui expliquerait pourquoi l'échelle de Planck se trouve à une énergie beaucoup plus élevée que celle de l'unification électrofaible. Dans le cas où il y a n dimensions supplémentaires compactes, leur volume V_n permet de faire le lien entre la masse de Planck habituelle (en 4 dimensions) $M_{Pl} = 2 \cdot 10^{18}$ GeV et la masse de Planck fondamentale en $n+4$ dimensions M [15] :

$$M_{Pl}^2 = M^{n+2} V_n$$

M peut être ramené à l'échelle électrofaible (~ 100 GeV) si V_n est suffisamment grand. D'autres échelles peuvent également être choisies pour M (quelques TeV par exemple) car V_n est essentiellement un paramètre libre.

Randall et Sundrum [15] ont proposé un autre modèle, contenant une seule dimension supplémentaire, plus petite, mais avec un facteur de déformation exponentiel (*warp factor*) dans la cinquième dimension.

1. Les couplages de l'électromagnétisme, de la force faible et de la force forte ne sont pas constants, ils dépendent de l'énergie à laquelle l'interaction a lieu.

Dans tous ces modèles, les dimensions supplémentaires sont compactes et peuvent donc contenir des ondes stationnaires, correspondant à des particules. Le graviton serait le mode fondamental de ces ondes stationnaires. Il existe aussi des variantes de ces modèles où les bosons de jauge peuvent aussi se propager dans les dimensions supplémentaires. Les bosons de jauge du MS seraient alors aussi des modes fondamentaux de ces ondes stationnaires. Des excitations plus massives existeraient aussi, correspondant aux modes avec 1, 2, 3, etc. noeuds. Ces excitations sont appelées excitations de Kaluza-Klein. L'existence de nouveaux bosons massifs, certains de spin 1 (excitations de bosons de jauge) et d'autres de spin 2 (excitations du graviton) est alors prédite, tout aussi bien dans des modèles avec des grandes dimensions supplémentaires [16] que dans des modèles de Randall-Sundrum [17].

Les nouveaux bosons neutres massifs prédits par les modèles avec dimensions supplémentaires sont donc les excitations du photon et du boson Z, de spin 1, et les excitations du graviton, de spin 2. La masse de ces excitations n'est pas fixée, elle dépend de certains paramètres du modèle, notamment la taille des dimensions supplémentaires, le nombre de dimension supplémentaire et le warp factor dans le modèle de Randall-Sundrum. Les couplages dépendent aussi du modèle étudié, mais les mêmes remarques sur les FCNC que nous avons faites dans la section précédentes s'appliquent ici.

Théories supersymétriques

Le MS est une théorie relativiste, c'est-à-dire invariante sous les transformations de l'algèbre de Poincaré. La supersymétrie (SUSY) est une extension de l'algèbre de Poincaré en une *superalgèbre*, qui contient non-seulement des opérateurs vectoriels (les translations, rotations et boosts) mais aussi des opérateurs de spin 1/2. Ces opérateurs de spin 1/2 mélangent fermions et bosons, de sorte que dans les théories SUSY tout boson est associé à un fermion et vice-versa, formant ce qu'on appelle des *supermultiplets*. Chaque fermion du MS a pour *superpartenaire* un boson scalaire, et chaque boson a pour superpartenaire un fermion de spin 1/2 [18]. Si la supersymétrie était exacte les superpartenaires auraient exactement les mêmes masses, charges électriques et temps de vie. Mais la supersymétrie ne peut pas être exacte, elle doit être brisée car sinon les partenaires bosoniques de tous les fermions du MS auraient déjà été découverts, vu leurs petites masses. Mais même brisée, la supersymétrie répond à de nombreuses questions [18] :

- Elle résout le problème de la renormalisation de la masse du boson H. Si la supersymétrie était exacte, les corrections quadratiques sont annulées, et il ne resterait plus que des corrections logarithmiques, qui ne présentent pas de problème d'ajustement précis. Une brisure dite *soft* de la supersymétrie permet de préserver cette propriété.
- La supersymétrie explique pourquoi l'énergie du vide (la constante cosmologique) est si petite comparée à l'échelle de Planck. Si elle était exacte, l'énergie du vide serait exactement nulle. La brisure de SUSY implique une correction à l'énergie du vide, qui peut alors être petite mais non-nulle.
- Les modèles SUSY préservant la R-parité prédisent l'existence d'une nouvelle particule stable et neutre (la plus légère particule supersymétrique) qui serait un candidat idéal pour la matière noire.

Les masses des superpartenaires dépendent du mécanisme de brisure de symétrie et de l'échelle d'énergie à laquelle celle-ci a lieu. Par contre, il y a des contraintes importantes sur les couplages de ces superpartenaires. Comme les nouvelles particules sont dans des supermultiplets avec des particules connues du MS, leurs couplages ne sont pas indépendants, ils peuvent être déduits des autres paramètres de la théorie. Par exemple, dans un supermultiplet avec un boson scalaire ϕ et un fermion ψ (tous les fermions du MS sont dans ce genre de supermultiplet dans les théories SUSY), le couplage de Yukawa $\lambda_{\phi\psi\psi}$ entre le scalaire et le fermion n'est pas indépendant au couplage quartique λ_{ϕ^4} du scalaire :

$$\lambda_{\phi^4} = |\lambda_{\phi\psi\psi}|^2 \quad (1.1)$$

Pour une théorie supersymétrique, le mécanisme le plus simple de brisure spontanée de l'interaction électrofaible ne fonctionne plus. Au lieu d'un seul doublet scalaire complexe, comme dans le mécanisme de Brout-Englert-Higgs, il en faut deux. Cela est dû au fait que le boson scalaire issu du champ de BEH possède un partenaire fermionique dans une théorie SUSY. L'ajout d'un fermion supplémentaire au MS peut poser problème à cause des anomalies introduites par les fermions. Les anomalies sont des symétries conservées classiquement qui sont brisées dans une théorie quantique, par des boucles de particules virtuelles. Cela pose problème pour des théories de jauge car les symétries de jauge doivent être exactes afin de donner un sens à la théorie. Tous les fermions introduisent des anomalies dans les symétries de jauge du MS, cependant les anomalies de certains fermions compensent celles introduites par d'autres de sorte à ce que, globalement, les anomalies du MS soient nulles. L'introduction d'un nouveau fermion rend ces anomalies non-nulles, et il faut inclure encore d'autres fermions pour les compenser. D'où la nécessité d'introduire un deuxième doublet scalaire, avec son partenaire fermionique, dans les théories SUSY [13].

La présence de ce deuxième doublet implique l'existence de 4 nouveaux bosons scalaires, deux chargés et deux neutres. D'autres bosons neutres massifs sont prédits par la supersymétrie : les superpartenaires scalaires des neutrinos, les sneutrinos. Tous les nouveaux bosons prédits par SUSY ont un spin 0, leur masse est indéterminée, mais leurs couplages sont fortement contraints par la théorie.

Modèles avec un couplage particulier aux leptons taus

Il existe des modèles de nouvelle physique où les leptons taus jouent un rôle particulier. C'est le cas par exemple du modèle $U(1)_{B-3L_\tau}$ [19]. Il s'agit d'un modèle introduisant une nouvelle symétrie de jauge, associée à la conservation du nombre quantique $B - 3L_\tau$, où B est le nombre baryonique et L_τ le nombre leptonique tauique. Le nombre baryonique vaut $1/3$ pour les quarks, $-1/3$ pour les antiquarks et 0 pour les autres particules, le nombre leptonique tauique vaut 1 pour les leptons taus (chargés et neutres), -1 pour les antileptons taus (chargés et neutres) et 0 pour les autres particules.

L'existence d'une nouvelle symétrie de jauge entraîne l'existence d'un nouveau boson de jauge. Comme il s'agit d'une symétrie $U(1)$, le nouveau boson de jauge est neutre et est donc une sorte de boson Z' . Ce boson Z' ne se couple qu'aux quarks et aux leptons taus. L'annulation des anomalies dans ce modèle impose à la charge des quarks sous cette nouvelle symétrie d'être 9 fois inférieure à celle des leptons taus. Par convention, une charge de $1/3$ est choisie pour les quarks et de 3 pour le lepton tau. Si l'on suppose que le boson Z' est

beaucoup plus lourd que le lepton tau et que tous les quarks, le rapport de branchement $R_{\tau\tau}$ du Z' en paire $\tau^+\tau^-$ peut s'approcher par la formule suivante :

$$R_{\tau\tau} = \frac{2 \cdot (-3)^2}{3 \cdot (-3)^2 + 6 \cdot 3 \cdot 2 \cdot (1/3)^2} = \frac{54}{93} \quad (1.2)$$

Ce rapport de branchement est beaucoup plus grand que celui du boson Z du MS en $\tau^+\tau^-$, qui est de l'ordre de 3.3%.

1.1.3 Modèle Standard Séquentiel et recherche expérimentale d'un boson Z'

Nous allons nous concentrer ici sur l'hypothèse d'un nouveau boson de jauge neutre massif, un Z' . Comme il existe toute une série de modèles prédisant un boson Z' , un modèle de référence appelé *Modèle Standard Séquentiel* (MSS) est utilisé dans les recherches de Z' . Il s'agit d'un modèle où le boson Z' aurait les mêmes caractéristiques que le boson Z du MS (notamment les mêmes rapports de branchement), les seuls paramètres qui les distinguent sont leurs masses et leurs sections efficaces. Précisons que le MSS n'est motivé par aucune théorie particulière, il sert simplement de point de comparaison utile entre les expériences qui recherchent des bosons Z' . Des limites sur des modèles particuliers de boson Z' peuvent être déduites à partir des limites sur le MSS.

Un canal de désintégration possible du Z' est la paire lepton-antilepton. Ce boson apparaîtrait alors au LHC sous la forme d'une nouvelle résonance dans le spectre de masse invariante des deux leptons de l'état final du processus $pp \rightarrow l^+l^- + X$. Une résonance représente un excès d'événements localisé (un "pic") dans un certain intervalle de masse invariante. Une particule se désintégrant en paire lepton-antilepton est électriquement neutre, ne porte pas de charge de couleur et est sa propre antiparticule [20].

Il est intéressant de rechercher le Z' dans les trois générations de leptons, dans les canaux de désintégration e^+e^- , $\mu^+\mu^-$ et $\tau^+\tau^-$. En effet, ses couplages aux électrons, muons et taus peuvent être différents. Nous avons donné un exemple d'une théorie BSM où le boson Z' se couple exclusivement aux quarks et aux leptons taus, ce genre de particule ne pourra jamais être découvert dans les canaux e^+e^- et $\mu^+\mu^-$. C'est pourquoi nous nous intéressons plus particulièrement à la recherche d'un boson Z' dans le canal $\tau^+\tau^-$ malgré le fait que l'analyse soit plus compliquée que dans les canaux e^+e^- et $\mu^+\mu^-$.

Les leptons taus ont un temps de vie court de $2.9 \cdot 10^{-13}$ s [21]. Pour un lepton ayant une énergie de 20 GeV, cela correspond à un parcours de seulement 980 μm depuis son lieu de production. Ce ne sont donc pas les leptons taus eux-mêmes qui sont identifiés dans les détecteurs mais leurs produits de désintégrations. Nous nous concentrerons sur le cas particulier où l'un des leptons taus se désintègre en hadron et neutrino tau, et l'autre se désintègre en muon et neutrinos.

Le défi principal de l'analyse est de reconstruire et d'identifier efficacement les leptons taus produits dans le détecteur CMS. Les leptons taus sont des objets particulièrement difficiles à reconstruire et identifier, en particulier dans le cas d'une désintégration en neutrino et hadrons. Cela demande un algorithme particulier, expliqué en détail dans la section 1.5

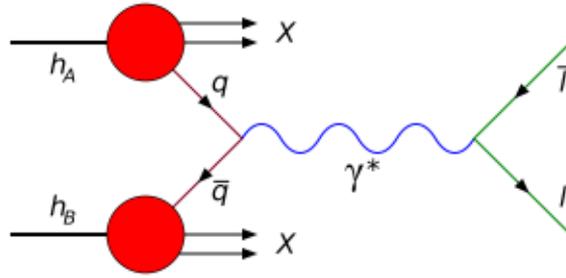


FIGURE 1.1 – Exemple de diagramme de Feynman d'un processus de Drell-Yan lors d'une collision hadron-hadron [22].

et étudié dans l'ensemble du chapitre 2. Une autre difficulté est de maîtriser les différents événements de bruits de fond. Ce sont des événements de processus du MS qui passent les sélections utilisées pour la recherche du boson Z' . Certains bruits de fond sont réductibles en affinant les sélections, d'autres ne le sont pas car ils correspondent au même état final que celui étudié pour la recherche du boson Z' . Pour découvrir un boson Z' , il faut observer un excès d'événements par rapport à ce qui est prédit par le MS étant donné les sélections. Voici une liste des principaux bruits de fond :

- Processus Drell-Yan (DY) du MS. Il s'agit de l'annihilation d'une paire quark-antiquark, qui produit un boson Z ou un photon virtuel, qui donne une paire lepton-antilepton (cf. le diagramme de Feynman dans la figure 1.1). Ceci est le seul bruit de fond irréductible de la liste.
- Processus multijets. Il s'agit de la production d'un certain nombre de quarks et de gluons énergétiques lors de la collision proton-proton. Chacun de ces quarks et gluons énergétiques sont à l'origine d'un "jet" de hadrons dans l'état final (plus d'explications sur les jets dans la section 1.5.3). Pour qu'un événement de ce processus contribue au bruit de fond, il faut que l'un des jets soit mal identifié comme un muon et un autre comme un tau hadronique. La section efficace des processus multijets est très élevée, ce qui fait que ceci reste un bruit important malgré l'application de sélections strictes.
- Production d'un boson W accompagné de jets (W +jets). La figure 1.2 donne un exemple de ce processus. Le boson W peut se désintégrer en deux leptons, l'un chargé et l'autre neutre. Un événement de ce processus contribue au bruit de fond si le lepton chargé est un muon et que l'un des jets est mal identifié en tant que tau hadronique.
- Processus diboson. Ce processus consiste en la production combinée de deux bosons vecteur (Z et/ou W). L'un des deux bosons peut se désintégrer en leptons alors que l'autre se désintègre en jets, ce qui peut provoquer une mauvaise identification en tant qu'événement $\tau_\mu\tau_h$.
- Production d'une paire top-antitop ($t\bar{t}$). Les quarks top se désintègrent presque exclusivement en boson W et quark b . Le boson W peut se désintégrer en muon et neutrino

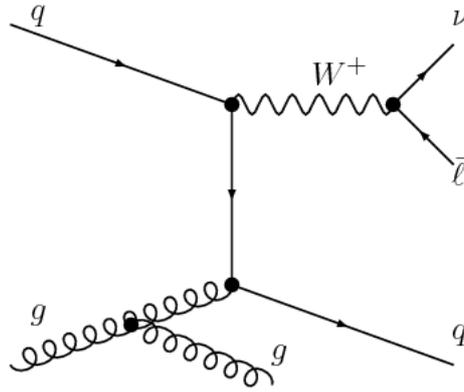


FIGURE 1.2 – Exemple de diagramme de Feynman d'un processus $W+2$ jets [23].

alors que le quark b est à l'origine d'un jet qui peut être mal identifié comme un tau hadronique.

Une analyse statistique est réalisée afin de déterminer s'il y a un excès de données par rapport au nombre d'événements de bruit de fond attendus. Les modèles prédisant un nombre de données trop élevé, incompatible avec ce qui a été observé, sont exclus. Ceci est expliqué en détail dans le chapitre 3 du mémoire.

1.2 Le collisionneur LHC

Le LHC est un collisionneur de hadrons situé au CERN, l'Organisation Européenne Pour La Recherche Nucléaire. Deux types de particules peuvent être accélérées par le LHC : les protons et les ions de Plomb. Les données utilisées dans ce mémoire concernent les collisions proton-proton, et c'est donc dans ce contexte que le LHC est décrit ci-dessous.

Le LHC est une machine circulaire de 27 km de circonférence qui sert à accélérer des faisceaux de protons et à les faire entrer en collision. Des cavités radiofréquence (RF) de 400 MHz [24] accélèrent les protons, des aimants supraconducteurs dipolaires pouvant atteindre des champs magnétiques de plus de 8 T [25] guident les particules dans le tracé de l'accélérateur, des aimants supraconducteurs quadripolaires les focalisent dans le plan transverse à la direction de l'accélération. Deux faisceaux de protons sont accélérés dans deux chambres à vide différentes, dans des sens opposés, et entrent en collision en 4 points de la machine, où se trouvent les expériences.

Mais avant que les protons n'entrent dans le LHC, ils sont accélérés par une série de machines différentes, appelées pré-accélerateurs ou injecteurs.

1.2.1 Le complexe d'accélérateurs du CERN

Le CERN entretient un grand complexe d'accélérateurs de particules, dont le LHC est le dernier composant (cf. la figure 1.3). La source des protons accélérés dans le LHC est un réservoir où de l'hydrogène gazeux est chauffé jusqu'à ce qu'il devienne du plasma. Une longue

CERN's Accelerator Complex

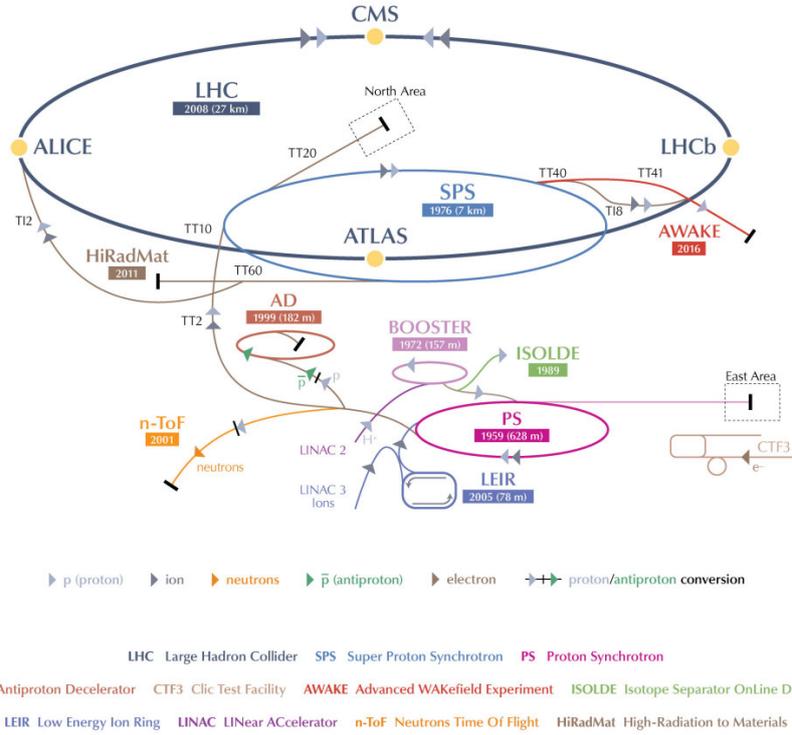


FIGURE 1.3 – Complexe d'accélérateurs du CERN [26].

chaîne d'injecteurs accélère les protons avant qu'ils n'entrent dans le LHC : le Linac2 fournit une première accélération de 50 MeV, ensuite le PSB (PS Booster) porte leur énergie à 1.4 GeV. Les protons sont ensuite injectés dans le synchrotron à protons (PS), où ils atteignent une énergie de 25 GeV avant d'être envoyés vers le supersynchrotron à protons (SPS) qui les accélère à 450 GeV. C'est à cette énergie qu'ils sont injectés dans le LHC.

Cette longue série d'accélération à travers 4 machines différentes est nécessaire car les machines circulaires possèdent toutes une énergie minimale d'injection, due à leur champ magnétique minimal d'injection. En effet la relation entre la quantité de mouvement p d'une particule, sa charge q , le rayon de courbure ρ de sa trajectoire et le champ magnétique B dans laquelle elle est plongée est donnée par l'équation

$$p = qB\rho \quad (1.3)$$

Si les aimants de courbure ont un champ rémanent B_{min} et que l'accélérateur circulaire a un rayon R , les particules doivent être injectées à une quantité de mouvement minimale donnée par

$$p_{min} = qB_{min}R \quad (1.4)$$

En plus de l'accélération des faisceaux à une énergie minimale acceptable pour le LHC, les injecteurs forment des *paquets* de protons. Les faisceaux ne sont pas continus, ils sont formés d'une série discrète de paquets de particules. Cela est dû aux cavités RF utilisées

pour l'accélération : elles fournissent un champ électrique sinusoïdal, qui n'accélère que les particules qui se trouvent dans la cavité pendant le bon intervalle de phase. Les protons se trouvant dans la cavité pendant la mauvaise phase auront une trajectoire instable et seront perdus par des collisions avec la paroi du tube à vide ou avec un collimateur. Les paquets sont formés de particules ayant des phases qui garantissent une trajectoire stable sous l'accélération par les cavités RF.

Pour recueillir un maximum de données, on souhaite que le LHC ait le plus de paquets possible. C'est pourquoi les injecteurs opèrent plusieurs divisions de paquets. Les divisions de paquets sont des opérations qui consistent à diviser en deux ou en trois chacun des paquets présents dans la machine. On introduit une harmonique de la fréquence utilisée dans la cavité RF, l'intensité de l'harmonique est augmentée progressivement alors que l'intensité de la fréquence fondamentale est diminuée. Cela se fait de sorte à ce que chacun des paquets résultant ait environ le même nombre de particules [27].

1.2.2 Paramètres du LHC

Le LHC a été conçu afin de maximiser deux paramètres : l'énergie du centre de masse proton-proton (\sqrt{s}) et sa *luminosité instantanée*.

L'énergie nominale du LHC est $\sqrt{s} = 14$ TeV, même si en pratique les données de 2015 ont été prises à une énergie de 13 TeV. L'énergie de 14 TeV correspond à deux fois l'énergie de 7 TeV de chacun des protons des deux faisceaux. En guise de comparaison, l'énergie au repos d'un proton est d'environ 1 GeV seulement, c'est-à-dire un 7000ème de l'énergie d'un proton au LHC. La formule relativiste qui relie l'énergie en mouvement E et énergie au repos mc^2 est :

$$E = \gamma mc^2 \tag{1.5}$$

où c est la vitesse de la lumière et γ est le facteur relativiste

$$\gamma = \frac{1}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \tag{1.6}$$

où v est la vitesse du proton. De 1.6, l'équation suivante pour v peut être déduite :

$$v = c \cdot \frac{\sqrt{\gamma^2 - 1}}{\gamma} = c \cdot \left(1 - \frac{1}{2\gamma^2} + O\left(\frac{1}{\gamma^4}\right)\right) \tag{1.7}$$

Comme nous venons de le voir, $\gamma \sim 7000$ au LHC, ce qui implique que la vitesse des protons est extrêmement proche de celle de la lumière, à plus de 99.9999% de la valeur de c . Par la suite, l'approximation $v \simeq c$ sera utilisée.

La luminosité instantanée est un paramètre qui indique le taux d'interactions qui peuvent se produire dans un détecteur, sa définition est liée à celle d'une quantité appelée *section efficace*. Chaque processus en physique des particules est caractérisé par une section efficace : plus un processus est probable, plus sa section efficace est grande. Si l'on multiplie la luminosité instantanée l de la machine par la section efficace σ du processus, on obtient la fréquence $\frac{dN}{dt}$ à laquelle ce processus aura lieu (cf. l'équation 1.8). Plus la luminosité est élevée, et plus

l'observation de processus rares est probable, et plus l'on pourra étudier précisément des processus plus fréquents.

$$\frac{dN}{dt} = \sigma \cdot l \quad (1.8)$$

L'équation 1.9 donne une valeur approchée² de la luminosité instantanée dans le cas où les particules dans un paquet sont distribués selon une loi normale dans les deux directions transverse x et y [28] :

$$l = \frac{N^2 N_b f}{4\pi \sigma_x \sigma_y} \quad (1.9)$$

où N est le nombre de particules dans un paquet, N_b est le nombre de paquets, f est la fréquence de révolution des particules dans le collisionneur circulaire, et σ_x et σ_y sont les écarts-types des distributions des particules au point d'interaction dans les directions transverses x et y . Voici ce que valent ces différentes variables au LHC :

- $f = \frac{v}{2\pi R} \simeq \frac{c}{2\pi R} \simeq 1.1 \cdot 10^4$ Hz est un paramètre fixé par la géométrie de la machine.
- $N \simeq 1.1 \cdot 10^{11}$. Le nombre de particule par paquet est limité par la répulsion électrostatique entre les protons.
- $N_b = 2808$, c'est le nombre maximal de paquets que le LHC peut contenir. La plupart de ces paquets se suivent à 25 ns d'intervalle, mais il faut laisser un certain nombre de "trous" dans le train de paquets. Un *abort gap* est nécessaire afin de pouvoir éjecter le faisceau en toute sécurité. Il correspond au temps nécessaire pour que les aimants d'éjection du faisceau atteignent leur voltage maximal. D'autres intervalles plus petits sont également nécessaires, ils correspondent à des besoins similaires des aimants d'injection du SPS et du LHC [29]. Le taux moyen de collision est de 31.6 MHz.
- $\sigma_x \simeq \sigma_y \simeq 14 - 16 \mu\text{m}$ [30]. Ces paramètres correspondent à la largeur du faisceau au point d'interaction. Ils sont minimisés avec un *squeeze* : la largeur du faisceau est augmentée de part et d'autre de la région d'interaction afin qu'elle puisse être diminuée au point d'interaction. Cette opération est limitée par la taille de la chambre à vide de part et d'autre du point d'interaction, et tout comme le nombre de particules par paquet, la taille transverse du faisceau est limitée par la répulsion électrostatique entre protons.

Ces paramètres donnent une luminosité instantanée de l'ordre de $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. La luminosité instantanée varie au cours du temps : lorsque les faisceaux ont été accélérés à leur énergie maximale et qu'ils commencent à entrer en collision, elle est maximale. Elle diminue ensuite au fur et à mesure que les particules du faisceau se perdent dans des collisions avec des parois de la machine et les collimateurs.

La luminosité instantanée peut être augmentée soit en augmentant le taux de croisement de paquets, soit en augmentant le nombre de collisions proton-proton par paquet. Le LHC ne peut pas contenir plus de 2808 paquets. Une fois que l'on a rempli le LHC, la seule façon d'augmenter la luminosité est d'augmenter le nombre de collisions par paquet. Ce nombre est appelé empilement ou *pile up* (PU), il vaut environ 20 pour une luminosité de l'ordre

2. On ne considère pas la répulsion électrique entre les deux faisceaux, ni la répulsion électrique entre les protons au sein du même faisceau.

$10^{34} \text{ cm}^{-2}\text{s}^{-1}$.

La luminosité instantanée est un paramètre important, mais le paramètre qui détermine le nombre total d'événements lors d'une prise de donnée est la *luminosité intégrée* L , qui est l'intégrale sur un certain intervalle de temps de la luminosité instantanée. Lorsque la luminosité intégrée est multipliée par la section efficace d'un certain processus, elle donne le nombre total de fois N où ce processus s'est produit lors du laps de temps considéré (cf. l'équation 1.10, version intégrée de 1.8).

$$N = \sigma \cdot L \tag{1.10}$$

L'unité de base utilisée pour les sections efficaces est le *barn* b (avec $1 b = 10^{-28} \text{ m}^2$), la luminosité intégrée est alors exprimée en multiples de b^{-1} .

Dans la suite de ce mémoire, "luminosité" sans plus de précision désigne la luminosité intégrée, la luminosité instantanée sera toujours appelée par son nom complet.

1.2.3 Design du LHC

Le LHC se trouve dans la catégorie des collisionneurs circulaires, mais précisons qu'aucun collisionneur n'est parfaitement circulaire. Le LHC contient des sections droites relativement longues séparées par des sections courbes. Les sections courbes sont composées de mailles qui se répètent périodiquement. Au LHC il s'agit de mailles "FODO", où le F désigne une section contenant un champ magnétique dipolaire pour la courbure et un champ quadripolaire focalisant dans la direction horizontale du plan transverse (et défocalisant dans la direction verticale), le O désigne une courte section droite, et le D un champ dipolaire accompagné d'un champ quadripolaire défocalisant dans la direction horizontale du plan transverse (et focalisant dans la direction verticale) [28]. L'alternance entre sections focalisantes et défocalisantes donne une focalisation globale forte, ce qui diminue la largeur du faisceau. Les quantités σ_x et σ_y de l'équation 1.9 sont donc plus petites, ce qui augmente la luminosité instantanée.

Il y a huit sections droites au LHC [31], numérotées de 1 à 8 à partir de l'entrée principale du CERN (Meyrin), dans le sens des aiguilles d'une montre :

- Point 1 : Expérience ATLAS (A Toroidal LHC ApparatuS). C'est l'une des deux grandes expériences généralistes du LHC. ATLAS s'intéresse à des études plus précises des particules du MS mais aussi à la recherche de nouvelle physique.
- Point 2 : Équipement d'injection du faisceau tournant dans le sens des aiguilles d'une montre et expérience ALICE (A Large Ion Collider Experiment). ALICE est spécialisée dans les collisions d'ions lourds.
- Point 3 : Collimateurs, pour contrôler les oscillations des particules autour de leur point de référence.
- Point 4 : Cavités RF, pour l'accélération des faisceaux.
- Point 5 : Expérience CMS (Compact Muon Solenoid), l'autre grande expérience généraliste du LHC. CMS partage les mêmes objectifs que ATLAS, mais avec un design différent pour les détecteurs. Les données de CMS sont celles qui sont utilisées pour ce mémoire.

- Point 6 : Équipement d'éjection du faisceau. En cas d'instabilité des faisceaux, ils sont éjectés de façon contrôlée dans cette section de la machine.
- Point 7 : Collimateurs, cf. Point 3.
- Point 8 : Équipement d'injection de l'autre faisceau et expérience LHCb. LHCb s'intéresse plus particulièrement à la physique du quark b et des mésons associés.

1.2.4 Calendrier du LHC

En 2011 et 2012, pendant son Run 1, le LHC a pris des données à des énergies du centre de masse $\sqrt{s} = 7$ et 8 TeV. Ces données correspondent à une luminosité intégrée d'environ 5 fb^{-1} pour les collisions à 7 TeV, et d'environ 20 fb^{-1} pour les collisions à 8 TeV. Un arrêt technique ou *Long Shutdown 1* (LS1) a suivi la fin du Run 1, il a servi à faire certaines réparations et à l'installation de nouvelles technologies dans les détecteurs. Le LS1 a également permis de préparer la montée en énergie à 13-14 TeV pour le Run 2.

Le LHC est à présent dans son Run 2, 4.2 fb^{-1} de données ont déjà été fournies à $\sqrt{s} = 13 \text{ TeV}$ en 2015 [32]. L'objectif pour l'entièreté du Run 2 (2015-2018) est de fournir 150 fb^{-1} [33] de données à des énergies du centre de masse de 13 et 14 TeV. Un LS2 suivra le Run 2, come le LS1 il permettra de réparer et d'améliorer les détecteurs. Le Run 3 comptera déjà avec une luminosité plus élevée que le Run 2, il devrait collecter 350 fb^{-1} de données entre 2020 et 2022.

À plus long terme, un LS3 suivra le Run 3 afin de préparer la mise à jour du LHC en *High Luminosity LHC* (HL-LHC), une machine capable d'atteindre des luminosités beaucoup plus élevées. Le Run 4 serait le premier Run du HL-LHC, avec un objectif ambitieux de récolter 3000 fb^{-1} de données en 4 ans (cf. la figure 1.4 pour plus de détails).

LHC Schedule: The Big Picture

2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Run 1 7-8 TeV, 0.7×10^{34} ($\mu=20$), 25 fb^{-1}		LS1		Run 2 13-14 TeV, 1.6×10^{34} ($\mu=43$), 150 fb^{-1}				LS2 Phase-I Install	
<i>LS = Long Shutdown</i>									
2020	2021	2022	2023	2024	2025	2026	2027	2028	2029
Run 3 14 TeV, 2.3×10^{34} ($\mu=50-80$), 350 fb^{-1}			LS3 – Phase-II Install		Run 4 14 TeV, 5.7×10^{34} ($\mu=140-200$), 3000 fb^{-1}				LS4

- *The HL-LHC running starts in 2025 and continues beyond LS4 until 2035*

FIGURE 1.4 – Calendrier à long terme du LHC [34]. Les cadres verts contiennent les paramètres des différents Runs du LHC : d'abord l'énergie du centre de masse, ensuite le *pile-up* μ (nombre de collisions proton-proton par croisement de paquets) et enfin la luminosité intégrée fournie par le LHC au cours de l'entièreté du Run. Les cadres beiges indiquent les arrêts techniques de longue durée (Long Shutdown - LS).

Dans ce mémoire, nous analyserons les données collectées en 2015 par le détecteur CMS, qui est décrit plus en détail dans la section suivante.

1.3 Le détecteur CMS

CMS est un détecteur multifonction spécialisé dans les collisions proton-proton. Il possède une symétrie cylindrique et est composé d'un tonneau central et de deux bouchons, un de chaque côté du tonneau. Le détecteur pèse près de 14 000 tonnes et est équipé d'un aimant solénoïde supraconducteur au niobium-titane (NbTi) pouvant fournir un champ de 3.8 Teslas [35]. Le solénoïde se trouve à l'intérieur d'un compartiment froid maintenu à 4.5 K afin de préserver la supraconductivité du composé NbTi. Le champ magnétique est renvoyé à travers la culasse de l'aimant dans les parties extérieures du détecteur. CMS contient plusieurs sous-détecteurs, qui sont distribués en couches concentriques autour de la région d'interaction (cf. les figures 1.5 et 1.6) : le détecteur de traces, le calorimètre électromagnétique, le calorimètre hadronique et le système de chambres à muons.

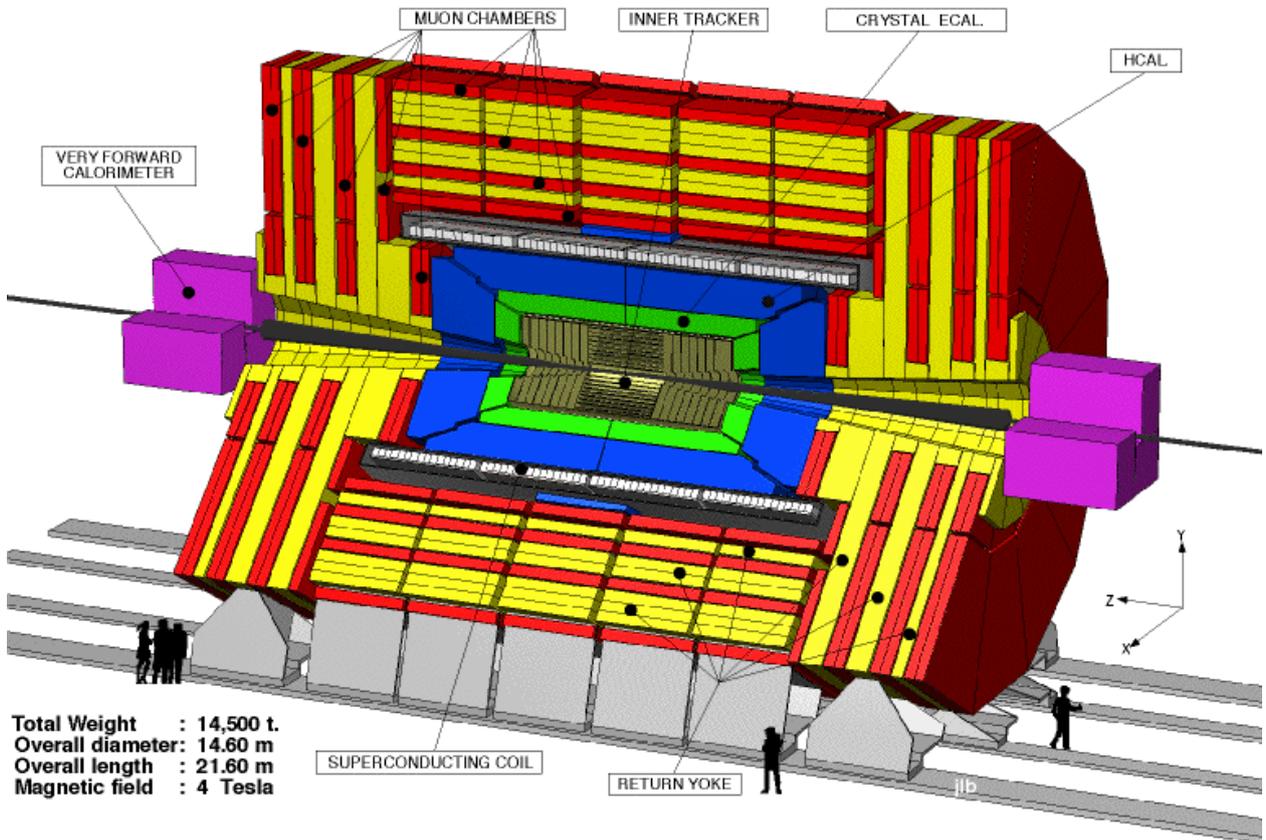


FIGURE 1.5 – Vue en 3D du détecteur CMS [38].

Le système de coordonnées utilisé dans CMS consiste d'un axe x horizontal et orienté vers le centre du LHC, d'un axe y vertical et orienté vers le haut, et d'un axe z dans la direction du tube à vide dont le sens est choisi de sorte à ce que le repère soit dextrogyre. L'angle azimutal ϕ est défini comme l'angle par rapport à l'axe x dans le plan xy et l'angle polaire θ est défini comme l'angle par rapport à l'axe z . Plutôt que l'angle θ , la pseudorapi-

dité $\eta = -\ln(\tan \frac{\theta}{2})$ est utilisée car dans la limite où la particule étudiée n'a pas de masse, les différences de pseudo-rapidités $\Delta\eta$ sont invariantes sous des boosts selon la direction z [36].

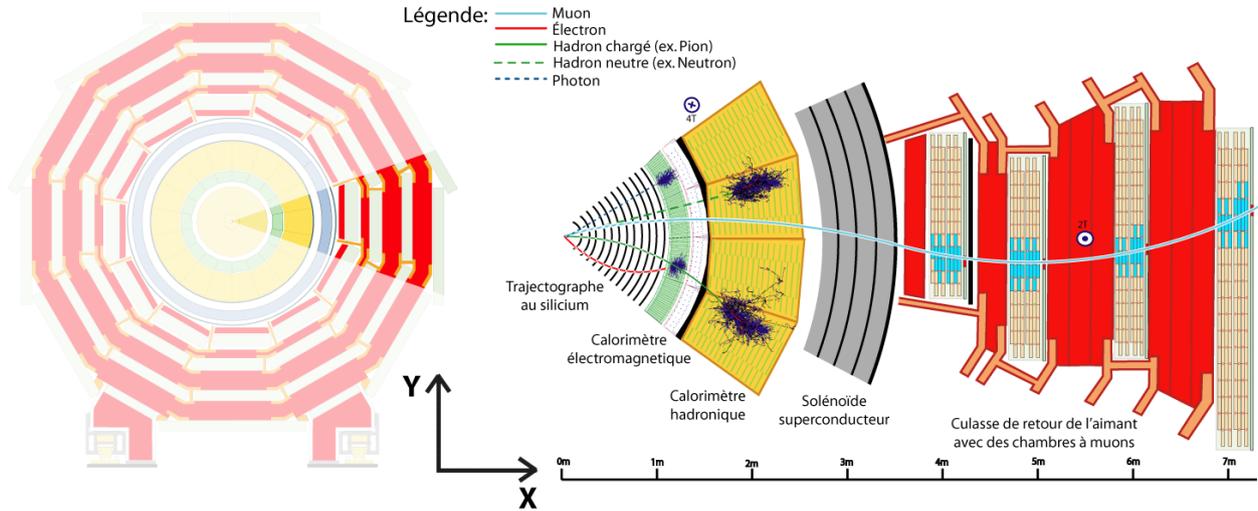


FIGURE 1.6 – Coupe transversale du tonneau de CMS [39].

1.3.1 Le détecteur de traces

Le détecteur de traces (tracker) permet de mesurer les trajectoires des particules chargées provenant des collisions de la région d'interaction, et à reconstruire d'éventuels vertex secondaires. Les précisions atteintes sont de l'ordre de quelques dizaines de microns. La technologie choisie pour la construction du tracker est celle des semi-conducteurs au silicium, qui permettent d'obtenir une granularité élevée, une bonne résistance à la radiation et une vitesse d'intégration rapide. Une coupe transverse de ce détecteur est présentée dans la figure 1.7.

Le tracker est divisé en deux parties : un détecteur à pixels dans la partie intérieure, et un détecteur à pistes (strips) à l'extérieur. Le détecteur à pixels a une meilleure résolution car il permet d'avoir les trois coordonnées de la particule qui le traverse en même temps, mais il faut un système d'acquisition de données pour chacun des pixels, ce qui peut provoquer un réchauffement considérable du détecteur dû aux pertes par effet Joule. C'est pourquoi il ne peut être utilisé qu'au centre du détecteur, où la surface à couvrir est moins grande. Pour le détecteur à pistes, il ne faut qu'un système d'acquisition de données pour chaque piste, par contre il faut changer l'orientation des pistes d'environ 10 degrés entre deux couches successives, afin d'obtenir une trajectoire tridimensionnelle de la particule tout en évitant le problème de "particules fantômes" que l'on aurait si l'angle était de 90 degrés [37].

Un champ magnétique uniforme de 3.8 T est présent dans tout le volume du tracker grâce à l'aimant de CMS. Le champ incurve les trajectoires de toutes les particules chargées d'une façon qui dépend de leur quantité de mouvement (cf. l'équation 1.3), celle-ci peut alors être mesurée. L'angle azimutal ϕ et la pseudo-rapidité η de la particule sont aussi mesurés à

l'entrée du détecteur.

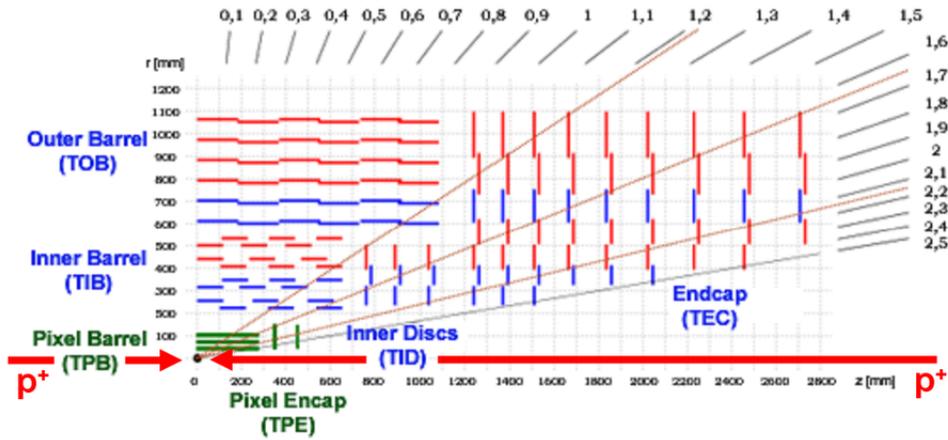


FIGURE 1.7 – Coupe longitudinale du détecteur de traces de CMS [40].

1.3.2 Les calorimètres

L'expérience CMS possède deux calorimètres : le calorimètre électromagnétique et le calorimètre hadronique.

Le calorimètre électromagnétique (ECal) mesure l'énergie des électrons et photons, appelés particules électromagnétiques. Le calorimètre est composé de plusieurs dizaines de milliers de cristaux de tungstate de plomb (PbWO_4) transparents et homogènes, qui provoquent des réactions en cascade pour les particules électromagnétiques qui le traversent. Dans ce milieu dense, les électrons émettent des photons par Bremsstrahlung et les photons se convertissent en paires électron-positron, cette réaction en chaîne est appelée gerbe électromagnétique. En-dessous d'une certaine énergie critique, les photons ne peuvent plus se convertir en paires d'électrons et la gerbe s'arrête. Les particules de la gerbe excitent le cristal en le traversant, celui-ci se désexcite en scintillant des photons du spectre visible. Ces photons sont collectés par des photodiodes à avalanche dans la partie tonneau de l'ECal, et par des phototriodes à vide dans les bouchons [37]. L'énergie de la particule incidente est proportionnelle au nombre de particules finales de la gerbe. Un calorimètre homogène comme celui de CMS permet d'atteindre de très bonnes résolutions en énergie.

Le calorimètre hadronique (HCal) mesure l'énergie des particules qui interagissent par la force forte. C'est un calorimètre à échantillonnage, où les matériaux absorbants sont du laiton et en moindre mesure de l'acier, et le matériel actif des scintillateurs plastiques. Le matériel absorbant dense sert à provoquer des réactions nucléaires et donc des gerbes hadroniques, et le matériel actif sert à mesurer le nombre et l'énergie des particules produites dans le matériel absorbant. Les calorimètres hadroniques souffrent de nombreux problèmes : non-linéarité, fraction importante d'énergie sous forme électromagnétique, grande largeur et profondeur des gerbes hadroniques. La résolution en énergie du HCal de CMS est beaucoup moins bonne que pour l'ECal.

La résolution d'un calorimètre s'écrit comme la somme en quadrature de trois termes (respectivement le terme stochastique et d'échantillonnage, le terme de bruit et le terme constant) :

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (1.11)$$

Le terme stochastique (a) correspond à la fluctuation du nombre de particules produites dans la gerbe. Il est modifié par un terme d'échantillonnage dans le cas d'un calorimètre non-homogène car une partie des particules produites restent dans le matériel absorbant, n'atteignent jamais le matériel actif, et ne sont donc jamais observées. Le terme de bruit (b) vient des imprécisions de mesure dans le système électronique d'acquisition de données. Le terme constant (c) est dû aux inhomogénéités du détecteur qui n'ont pas été prises en compte dans la calibration.

Pour l'ECal de CMS, $a \simeq 2.8 \sqrt{\text{GeV}} \cdot \%$, $b \simeq 0.12 \text{ GeV} \cdot \%$, $c \simeq 0.30\%$. Pour le HCal, $a \simeq 93.8 \sqrt{\text{GeV}} \cdot \%$, $c \simeq 4.4\%$ et b est négligeable [37].

1.3.3 Les chambres à muons

À l'extérieur du solénoïde se trouvent les chambres à muons, qui servent à identifier les muons et à mesurer plus précisément leurs quantités de mouvement. Ces détecteurs sont nécessaires car les muons sont très pénétrants. Les électrons interagissent par Bremsstrahlung, les photons se convertissent en paires e^+e^- et les hadrons subissent des interactions nucléaires, mais les muons ne subissent aucune de ces interactions³. Les muons sont typiquement les seules particules chargées à sortir des deux calorimètres, ils sont identifiés grâce aux chambres à muons extérieures. Celles-ci sont placées entre plusieurs sections de la culasse de l'aimant, qui fournit le champ magnétique nécessaire pour courber la trajectoire des muons et mesurer leur quantité de mouvement.

La nécessité de couvrir un grand volume à l'extérieur du solénoïde par les chambres à muons impose l'utilisation de détecteurs gazeux, qui ont un moindre coût. Le temps d'intégration de données est plus long, mais cela ne pose pas de problème car chaque chambre ne couvre qu'un angle solide petit par rapport aux sous-détecteurs qui se trouvent plus près du tube à vide, et le taux d'événement attendu pour chaque chambre à muon est donc plus petit. Trois technologies différentes ont été utilisées pour ces détecteurs :

- Des chambres à dérive dans la région du tonneau.
- Des chambres à pistes cathodiques dans les bouchons.
- Des chambres à plaques résistives dans le tonneau et les deux bouchons.

1.3.4 Le système de déclenchement

Comme nous l'avons vu dans la section 1.2.2, le taux moyen de croisement de paquets au LHC est de 31.6 MHz. Stocker la totalité des données produites lors des collisions au LHC

³. Pour être plus précis, les muons interagissent par Bremsstrahlung, mais en bien moindre mesure que les électrons.

est impossible, il faut un système de déclenchement qui ne sauvegarde que les événements les plus intéressants pour une analyse de physique. Il y a deux systèmes de déclenchement dans le détecteur CMS : un premier, moins complexe mais très rapide, diminue le taux d'événements à ~ 100 kHz ; un deuxième, plus sophistiqué, diminue le taux d'événements à ~ 100 Hz. Ces événements-là sont sauvegardés.

Les événements collectés par le système d'acquisition de données (DAQ) de CMS passent d'abord par un premier système de déclenchement appelé niveau 1 ou *Level 1* (Lvl-1). Les algorithmes du Lvl-1 sont implémentés dans des circuits intégrés spécialisés, et prennent la décision de sauvegarder ou pas l'événement en $2-3 \mu\text{s}$. Ces algorithmes analysent les données des calorimètres et des chambres à muons afin de chercher des électrons, muons, photons ou jets isolés, et de calculer l'énergie transverse manquante (MET)⁴ de l'événement. Les données du tracker ne sont pas utilisées car cela prendrait trop de temps.

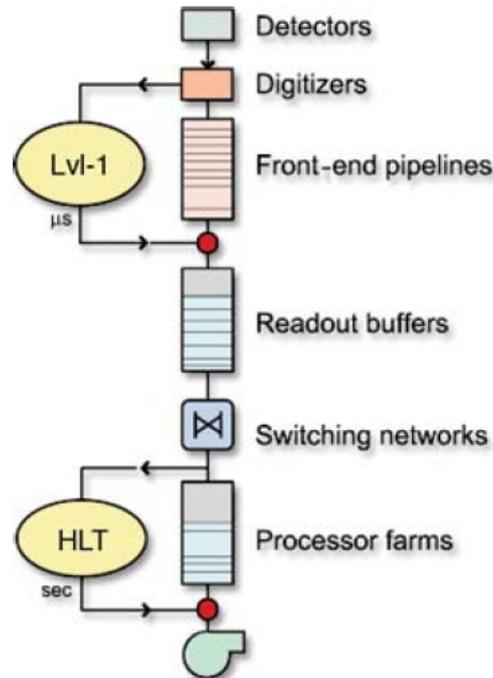


FIGURE 1.8 – Système de déclenchement (*trigger*) de CMS [41]. Les données collectées par CMS sont d'abord numérisées par les *Digitizers*, ensuite elles sont sauvegardées temporairement dans les *Front-end pipelines* tout en étant analysées par le Lvl-1. Si elles sont acceptées, elle passent au tampon (*Readout buffer*) et ensuite au centre de calcul (*Processor Farms*) où la sélection au niveau HLT est implémentée.

Il y a un croisement de paquets toutes les 25 ns au LHC, il est donc nécessaire de stocker temporairement les données d'une centaine de croisements de paquets afin de donner le temps nécessaire au système de déclenchement de prendre sa décision. Ce stockage temporaire se fait dans un composant électronique appelé *pipeline*. À la sortie du pipeline, les données rejetées par le Lvl-1 sont effacées, celles qui sont acceptées entrent dans un tampon. Le flux de données sortant du Lvl-1 est irrégulier car seuls les événements "intéressants"

4. Pour une définition de l'énergie transverse manquante, cf. la section 1.5.3.

sont sélectionnés, le tampon permet de régulariser ce flux avant que les données ne soient analysées par le deuxième système de déclenchement, le déclenchement de haut niveau *High Level Trigger* (HLT).

La sélection au niveau HLT se fait grâce à un centre de calcul très performant. Les algorithmes du HLT utilisent des logiciels de reconstruction et des sélections d'événements similaires à ceux utilisés dans les analyses de données. Chaque événement est reconstruit dans son ensemble, en considérant les données de tous les sous-détecteurs, et est sauvegardé s'il remplit toutes les conditions de l'un ou l'autre *menu de déclenchement*. Les menus de déclenchement sélectionnent des événements qui ont des particules isolées très énergétiques, ou une énergie transverse manquante importante, etc. Il existe plusieurs menus de déclenchement différents, qui sont utiles à des analyses différentes. Comme pour le Lvl-1, un pipeline sauvegarde temporairement les données des événements le temps que le HLT prenne sa décision (cf. la figure 1.8). Les événements acceptés par le HLT sont définitivement sauvegardés.

1.4 Données et simulations

Une analyse en physique des particules se base à la fois sur des données et des simulations. Les données possèdent de nombreuses caractéristiques importantes, entre autres leur luminosité, et l'énergie du centre de masse à laquelle elles ont été prises. Les simulations sont quant à elles essentielles afin de corriger les données pour certains effets propres au détecteur, comme son acceptation géométrique et l'efficacité des reconstructions des particules, et de comparer les mesures aux prédictions théoriques. Cette section contient des détails sur les données et simulations utilisées dans ce mémoire.

1.4.1 Données

L'étude de l'efficacité de l'algorithme d'identification des leptons taus et la recherche d'un boson Z' dans le spectre à haute masse invariante de deux leptons taus se basent sur les données de collisions proton-proton collectées en 2015 par CMS. Il s'agit du début du Run 2. Les collisions ont eu lieu à une énergie du centre de masse $\sqrt{s} = 13$ TeV.

En 2015, le LHC a fourni une luminosité totale d'environ 4.2 fb^{-1} aux expériences CMS et ATLAS [32]. La prise de données a commencé en juin avec des paquets de protons se suivant à des intervalles de 50 ns. Ensuite, les faisceaux ont commencé à circuler avec des intervalles de 25 ns entre les paquets, et la largeur des faisceaux σ_x et σ_y a été progressivement réduite par des focalisations de plus en plus fortes. La luminosité instantanée fournie par le LHC a augmenté au cours des mois, partant de $2 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ en juin et atteignant $50 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ en novembre. La figure 1.9 donne l'évolution au cours de l'année 2015 de la luminosité instantanée et de la luminosité intégrée fournies par le LHC. Au total, environ 100 pb^{-1} de données ont été fournies avec une séparation de 50 ns entre les paquets [42], et le reste (4.1 fb^{-1}) avec une séparation de 25 ns.

Il existe une différence entre la luminosité intégrée totale fournie par le LHC et celle qui est effectivement collectée par les expériences, car les différents sous-détecteurs ne sont pas toujours en état de fonctionnement. L'expérience CMS a collecté parmi les 4.2 fb^{-1} de

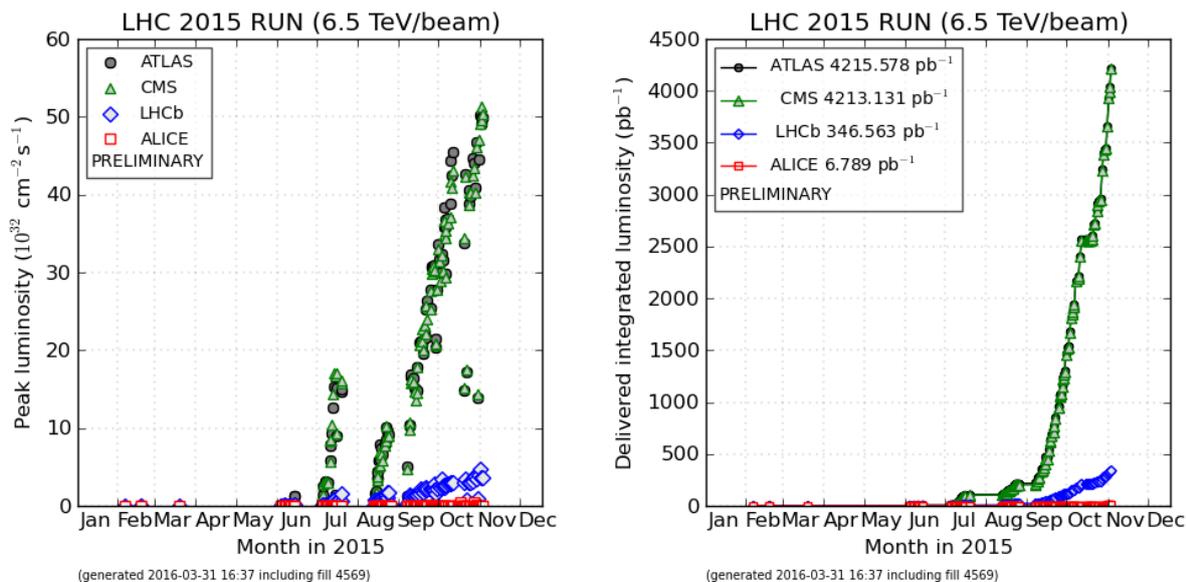


FIGURE 1.9 – À gauche, la luminosité instantanée maximale fournie par le LHC en 2015 à ses quatre expériences principales. À droite, la luminosité intégrée [32]. La plupart des données ont été collectées entre septembre et novembre.

données fournies par le LHC 3.8 fb^{-1} avec la totalité ou la quasi-totalité des sous-détecteurs en bon état de fonctionnement.

L'expérience CMS n'a pas toujours pu collecter les données avec le champ magnétique allumé. Il y avait une fuite dans la chambre froide de l'aimant solénoïde. Cette fuite provoquait parfois un réchauffement de l'aimant au-delà de la température à laquelle il est supraconducteur, et le solénoïde devait alors être éteint. Les données collectées par CMS avec le champ magnétique allumé correspondent à 2.7 fb^{-1} , environ $3/4$ du total [43].

Le champ magnétique sert à incurver les traces des particules chargées produites lors des collisions, ce qui permet de mesurer leurs quantités de mouvement (cf. l'équation 1.3). L'absence de champ magnétique pose un certain nombre de problèmes :

- La reconstruction⁵ de l'énergie des hadrons chargés est moins précise. Le tracker permet d'estimer la quantité de mouvement d'une particule (et donc son énergie, avec une hypothèse sur la masse de la particule) en mesurant la courbure de sa trajectoire. Or toutes les trajectoires sont droites en l'absence de champ magnétique. La reconstruction de l'énergie doit donc se baser entièrement sur les données des calorimètres, qui sont peu précises pour les hadrons.
- La reconstruction de l'énergie des muons devient très difficile. Les muons sont des particules pénétrantes, qui interagissent relativement peu avec la matière : ils traversent notamment les deux calorimètres de CMS en n'y déposant que quelques centièmes de leur énergie totale. Les chambres à muons et le tracker sont les sous-détecteurs qui permettent de mesurer la quantité de mouvement des muons, mais cela n'est possible

5. Une explication détaillée des algorithmes de reconstruction est donnée dans la section 1.5.

que si leur trajectoire est courbée par le champ magnétique.

L'analyse réalisée dans ce mémoire est basée sur l'état final $\tau^+\tau^-$, avec un lepton tau se désintégrant en muon et neutrinos et l'autre en hadrons et neutrino. Elle est impossible en l'absence de champ magnétique, car elle repose sur une bonne reconstruction des muons et dans une moindre mesure une bonne reconstruction des hadrons. Pour ce mémoire, seules des données avec l'ensemble des sous-détecteurs en bon état de fonctionnement, avec le champ magnétique allumé et avec un espacement de 25 ns entre les paquets de protons ont été considérées. Cela correspond à une luminosité intégrée de 2.2 fb^{-1} .

1.4.2 Simulations

Lors de la mesure d'un paramètre physique, il est nécessaire de travailler avec des simulations d'événements, afin de corriger les données pour une série d'effets (acceptance géométrique du détecteur, efficacité des critères de sélections, etc.). Les mesures corrigées peuvent alors être comparées à des expériences différentes ou à des modèles théoriques. Lors de la recherche d'une nouvelle particule prédite par un modèle BSM, les données sont comparées à deux simulations différentes : l'une correspondant aux prédictions du MS, l'autre à celle du modèle BSM qui contient la nouvelle particule. Des méthodes statistiques permettent de déterminer l'écart entre les données et les prédictions du MS, et déterminer si cet écart est compatible avec le modèle BSM.

Les simulations sont générées par des méthodes *Monte-Carlo* (MC). Le nom Monte-Carlo désigne une classe d'algorithmes qui servent, entre autres, à estimer la forme d'une fonction ou d'une distribution de probabilité par l'utilisation répétée de procédés aléatoires. La précision de ces méthodes s'améliore en raison inverse de la racine carrée du nombre d'événements simulés, quel que soit le nombre de dimensions considérées. Cela est très utile en physique des particules, car il est nécessaire de simuler simultanément une grande quantité de paramètres différents (identité, énergie et quantité de mouvement de chaque particule, MET de l'événement, etc.), ce qui requiert donc un espace paramétrique avec un grand nombre de dimensions.

La modélisation des interactions proton-proton est un problème compliqué. La plupart des interactions sont des interactions de basse énergie, dites *soft*, dans le régime non-perturbatif de l'interaction forte. Mais les événements qui nous intéressent dans ce mémoire sont les collisions énergétiques entre les constituants du proton⁶, les processus *hard*, qui sont beaucoup plus rares. Afin de ne pas simuler inutilement des événements inintéressants, les simulations sont construites autour des processus *hard*, et c'est l'utilisateur de la simulation qui choisit le processus *hard* particulier qui l'intéresse. Les processus *hard* peuvent être calculés directement à partir de la théorie par des méthodes perturbatives [44]. La désintégration des particules instables à court temps de vie produites par le processus *hard* (comme le boson Z, le quark top, etc.) est simulée ensuite.

6. Les constituants du proton sont des quarks et des gluons. Il y a non seulement les trois quarks de valence (2 up et 1 down) qui déterminent les nombres quantiques du proton, mais aussi toute une série de quarks dits de la mer et de gluons issus de l'interaction forte entre les trois quarks de valence.

Parmi les particules de l'état final du processus *hard* se trouvent des quarks et des gluons. Les quarks et gluons s'assemblent dans des particules composites sans charge de couleur, les hadrons. Ce processus de hadronisation est peu compris dans le cadre de la chromodynamique quantique (QCD), mais des modèles phénoménologiques permettent de le simuler.

Lorsque l'ensemble de l'événement a été généré, l'interaction des particules de l'état final avec le détecteur peut être simulée. Dans CMS, cela se fait à l'aide de l'algorithme GEANT4. La géométrie et l'ensemble des différents matériaux présents dans tous les sous-détecteurs y sont simulés. Les particules générées sont propagées à travers ce détecteur virtuel, en tenant compte de leurs interactions avec la matière [45]. La réponse des différents sous-détecteurs est modélisée, et à partir de ces données les particules de l'état final sont alors reconstruites avec les algorithmes de reconstruction décrits dans la section 1.5.

Les événements dont on ne simule pas l'interaction avec le détecteur sont appelés *événements générés*. Après simulation de l'interaction avec le détecteur et application des algorithmes de reconstruction, ils sont appelés *événements reconstruits*.

Dans ce mémoire, les simulations suivantes ont été utilisées :

- Une simulation du processus $DY \rightarrow l^+l^-$ générée par le logiciel madgraph [48] à l'ordre dominant (*Leading Order* ou LO) et avec une limite inférieure de 50 GeV sur la masse invariante générée de l'état final.
- Une simulation du processus W+jets générée par le logiciel madgraph à l'ordre dominant.
- Une simulation du processus $t\bar{t}$ générée par le logiciel pythia [49] à l'ordre dominant.
- Deux simulations du processus diboson générée par le logiciel aMC@NLO [50] à l'ordre suivant de l'ordre dominant (NLO). L'une des simulations concerne le processus $VV \rightarrow 2l2\nu$ (où V représente un boson W ou un boson Z et l un lepton chargé) et l'autre le processus $WW \rightarrow l\nu$.
- Quatre simulations pour le processus $Z' \rightarrow \tau^+\tau^-$ du Modèle Standard Séquentiel (MSS), pour quatre hypothèses de masse différentes (1000, 1500, 2000 et 2500 GeV), générées par le logiciel madgraph à l'ordre dominant. Les termes d'interférence avec le photon et le boson Z n'ont pas été pris en compte.

Plus d'informations sur ces simulations sont disponibles dans la table 1.2.

Repondération des événements en fonction de l'empilement

Les événements simulés sont générés avec une certaine hypothèse sur le nombre moyen de collisions simultanées dans le détecteur, l'empilement (pile up ou PU, cf. la section 1.2.2). Les simulations sont construites avant la prise de données, et elles n'ont donc pas la même distribution en PU que les données. Pour corriger cela, il faut appliquer une pondération différente aux événements simulés en fonction du nombre d'événements de PU. Cette procédure est construite de sorte à ne pas changer le poids moyen accordés aux événements, seuls les poids relatifs entre événements sont changés. Tous les simulations utilisées dans ce mémoire ont été repondérées sur base des données de 2015.

Processus simulé	Nombre d'événements simulés	Section efficace de la simulation [pb]
$DY \rightarrow l^+l^-$ ($M > 50$ GeV)	$9.04 \cdot 10^6$	6025
W+jets	$72.4 \cdot 10^6$	61527
$t\bar{t}$	$96.8 \cdot 10^6$	832
$VV \rightarrow 2l2\nu$	$8.22 \cdot 10^6$	11.95
$WW \rightarrow l\nu$	$5.04 \cdot 10^6$	49.995
$Z' \rightarrow \tau^+\tau^-$ (1000 GeV)	$0.34 \cdot 10^6$	0.502
$Z' \rightarrow \tau^+\tau^-$ (1500 GeV)	$0.36 \cdot 10^6$	0.08525
$Z' \rightarrow \tau^+\tau^-$ (2000 GeV)	$0.36 \cdot 10^6$	0.01990
$Z' \rightarrow \tau^+\tau^-$ (2500 GeV)	$0.37 \cdot 10^6$	$6.13 \cdot 10^{-3}$

TABLE 1.2 – Nombres d'événements simulés pour chaque processus et section efficace de la simulation. Les sections efficaces pour les cinq premiers processus sont calculées aux ordres suivants de l'ordre dominant (NLO et NNLO), voir la référence [46] pour des détails. La section efficace des processus $Z' \rightarrow \tau^+\tau^-$ correspond à la section efficace calculée à l'ordre dominant multipliée par un facteur 1.3 représentant les effets dus aux ordres supérieures. Cette correction est standard dans CMS [47].

1.5 Algorithmes de reconstruction et d'identification de particules de CMS

Les particules produites lors des collisions proton-proton dans CMS doivent être *reconstruites et identifiées* avant qu'elles ne puissent être utilisées dans une analyse. L'étape de reconstruction consiste à déterminer la trajectoire, l'énergie et la quantité de mouvement de la particule dans le détecteur. L'identification correspond à une série de sélections appliquées sur la particule reconstruite afin de déterminer son identité (muon, électron, hadron, photon, etc.).

Les algorithmes de reconstruction et d'identification de chaque particule se caractérisent par une efficacité et par des taux de mauvaise identification. Si l'on cherche à identifier une particule A, l'efficacité correspond à la proportion de particules A passant les critères d'identification de la particule A, parmi l'ensemble des particules A produites. Le taux de mauvaise identification $B \rightarrow A$ correspond à la proportion de particules B passant les critères d'identification de la particule A, parmi l'ensemble des particules B produites. Les algorithmes d'identification sont conçus de sorte à maximiser leur efficacité et minimiser leurs taux de mauvaise identification.

Cette section décrit les algorithmes développés par CMS pour reconstruire et identifier les différentes particules créées lors des collisions proton-proton : les photons, les leptons chargés (électrons, muons et taus), les hadrons. CMS possède un algorithme sophistiqué appelé *Particle Flow* (PF) qui réalise une reconstruction complète de l'événement basée sur l'information de tous les sous-détecteurs. Les particules sont classées dans des catégories mutuellement exclusives, afin d'éviter les erreurs de double comptage [51]. Pour le lepton tau, qui a un court temps de vie moyen et qui se désintègre à l'intérieur du tube à vide du détecteur, un algorithme particulier est nécessaire pour sa reconstruction et son identification.

1.5.1 Muons

Les muons sont reconstruits de deux façons différentes : *Global Muon* et *Tracker Muon* [52].

- La méthode *Global Muon* part d'une trace dans le système des chambres à muons et cherche une trajectoire compatible dans le tracker. Une trace globale du muon est construite à partir des traces des deux sous-détecteurs par la méthode du filtre de Kalman. Il s'agit d'une méthode itérative linéaire qui sert à ajuster des paramètres de trace [53]. L'estimation de ces paramètres est non-biaisée et optimale si les erreurs de mesure et les incertitudes dans la propagation des particules sont gaussiennes.
- L'approche *Tracker Muon* part d'une trace dans le tracker, l'extrapole et vérifie sa compatibilité avec l'une des traces du système de chambres à muons.

Dans ce mémoire, nous ne considérerons que des muons globaux, dont l'identification est meilleure si les muons ont une impulsion plus grande que 5 GeV [52]. Les efficacités de la reconstruction des muons dans le tracker et les chambres à muons sont très grandes, plus de 99% des muons sont correctement reconstruits s'ils sont dans l'acceptance du détecteur et s'ils ont une quantité de mouvement suffisamment élevée (quelques GeV). Les taux de mauvaise identification (ou *fake rates*) sont très faibles (<1% [54]) car peu de particules autres que des muons atteignent le système de chambres à muons.

Pour les analyses de données réalisées dans ce mémoire, les muons doivent passer la sélection standard de l'algorithme PF, ainsi que d'autres sélections plus strictes, ce qui fait que l'efficacité de reconstruction et d'identification des muons est légèrement inférieure à 99%.

1.5.2 Électrons

L'approximation linéaire du filtre de Kalman n'est pas adaptée pour la reconstruction de trace des électrons. Contrairement aux muons, dont l'interaction avec les matériaux constituant le détecteur se fait principalement par un processus de diffusion coulombienne multiple qui peut raisonnablement s'approximer par une gaussienne, pour les électrons le mode dominant de perte d'énergie dans la matière est l'émission d'un rayonnement de freinage (Bremsstrahlung) [37], dont la fonction de distribution de probabilité (PDF) n'est pas une gaussienne mais une fonction de Bethe-Heitler [55].

Un photon de Bremsstrahlung peut emporter une part importante de l'énergie de l'électron, et la direction de propagation de l'électron peut changer de façon considérable. Il faut tenir compte de ces changements brusques de direction afin de bien reconstruire la trace d'un électron. L'algorithme utilisé pour reconstruire les traces des électrons dans CMS s'appelle *Gaussian-sum filter* (GSF), il approxime la PDF de Bethe-Heitler par un mélange gaussien. Le photon de Bremsstrahlung peut être émis dans une large gamme d'énergie, et chacune des composantes du mélange gaussien représente une certaine échelle d'énergie pour le photon. La trajectoire se construit progressivement, point par point, chaque point correspondant à l'intersection de la trace de la particule avec un appareil de mesure. Des étapes de propagation (extrapolation du vecteur d'état et de ses incertitudes basée sur l'information des points précédents) sont alternées avec des étapes de mise à jour avec les données d'un nouveau

point. Le vecteur d'état obtenu après mise à jour est une convolution entre l'effet statistique de la perte d'énergie par Bremsstrahlung (modélisée par un mélange gaussien) et l'extrapolation du vecteur d'état (qui est lui aussi un mélange gaussien). À chaque étape, le nombre de paramètres déterminant le vecteur d'état de l'électron ne cesse de croître. Afin d'éviter une explosion combinatoire, les composantes du mélange gaussien sont regroupées en *clusters*, chacun des *clusters* étant décrit par une simple gaussienne.

La trace de l'électron ainsi reconstruite est comparée à l'activité observée dans l'ECal, en tenant compte de l'énergie emportée par les photons de Bremsstrahlung. Un algorithme cherche des compatibilités entre traces et dépôts d'énergie dans l'ECal. Les informations provenant de ces sous-détecteurs sont regroupées pour former un seul objet, l'électron. Son énergie est la moyenne pondérée par les incertitudes de mesure de l'énergie mesurée à partir du tracker et celle mesurée dans l'ECal.

Pour les analyses de ce mémoire, l'identification des électrons est utilisée afin de rejeter tout événement contenant des électrons. Les électrons rejetés sont ceux qui passent la sélection standard de CMS (connue sous le nom de *WP 90*, ou point de fonctionnement à une efficacité de 90%).

1.5.3 Jets et énergie transverse manquante

Une fois que les électrons et muons ont été identifiés, les traces restantes présentes dans le tracker sont interprétées comme provenant de hadrons chargés. Des dépôts dans le HCal sans trace correspondante dans le tracker sont attribués à des hadrons neutres. Une étape de vérification existe afin de s'assurer que l'énergie totale restante collectée dans les calorimètres est supérieure à l'énergie des particules chargées restantes mesurée dans le tracker, après élimination des dépôts associés aux électrons et/ou muons [51].

Les hadrons sont ensuite regroupés en jets de particules par l'algorithme *anti- k_T* [56]. L'intérêt de regrouper les hadrons par "jets" est d'essayer de retrouver les quarks et gluons (quarks et gluons sont désignés ensemble sous le nom de *partons*) énergétiques produits lors de l'interaction proton-proton. Ces partons énergétiques peuvent émettre d'autres partons, ce qui donne lieu à une gerbe partonique. Ensuite tous les partons de l'état final forment des hadrons. Un seul parton énergétique produit lors de l'interaction proton-proton peut donc donner plusieurs hadrons dans l'état final. Ces hadrons sont plus ou moins colinéaires et forment un jet.

L'algorithme *anti- k_T* commence par définir des distances d_{ij} entre les entités i et j (particules ou collections de particules) et des distances d_{iB} entre l'entité i et le faisceau B . L'algorithme cherche la distance la plus petite, si elle est de type d_{ij} , i et j sont regroupés dans une nouvelle entité, si elle est de type d_{iB} , i est déclaré être un jet et est enlevé de la liste des entités. Les distances sont recalculées et ce processus continue jusqu'à ce qu'il ne reste plus aucune entité. Dans l'algorithme *anti- k_T* , les distances sont définies ainsi :

$$d_{ij} = \min \left(\frac{1}{k_{Ti}^2}, \frac{1}{k_{Tj}^2} \right) \cdot \frac{\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2}{R^2} \quad (1.12)$$

$$d_{iB} = \frac{1}{k_{T_i}^2} \quad (1.13)$$

où k_{T_i} est la quantité de mouvement transverse de l'entité i , $\Delta\eta_{ij}$ est la différence entre les deux pseudo-rapidités, $\Delta\phi_{ij}$ la différence entre les deux angles azimutaux, et R est le paramètre de distance, dont on peut choisir la valeur. On prend typiquement $R = 0.4$ ou 0.5 .

Anti- k_T est un algorithme de reconstruction de jets dit *soft-resilient*, c'est-à-dire, la forme du jet n'est pas influencée par la présence de radiation de faible énergie.

Le vecteur d'énergie transverse totale est obtenu en calculant la somme vectorielle des impulsions transverses visibles⁷, le vecteur opposé à cette somme est alors le vecteur d'énergie transverse manquante (MET). Celui-ci est caractérisé par deux paramètres : son module, et son angle azimutal. L'existence d'un vecteur de MET non-nul indique la présence de particules "invisibles" qui n'interagissent que faiblement (comme les neutrinos ou une nouvelle particule hypothétique) et/ou d'une erreur dans la mesure des quantités de mouvement visibles. Il s'agit d'une variable importante lorsque l'on étudie des états finaux avec un ou plusieurs neutrinos, elle sera utile dans notre analyse surtout pour rejeter un bruit de fond particulier, la production d'un boson W qui se désintègre ensuite en lepton et neutrino.

1.5.4 Leptons taus

Le lepton tau n'est pas une particule stable à l'échelle du détecteur CMS : sa durée de vie est de $2.9 \cdot 10^{-13}$ s [21], un lepton tau de 20 GeV se désintègre après un parcours d'environ 980 μm . L'identification d'un lepton tau demande donc la reconstruction de ses produits de désintégration, et il faut déterminer ensuite si ceux-ci proviennent effectivement d'un lepton tau. La désintégration d'un lepton tau produit toujours un neutrino tau par la conservation de la saveur leptonique. La masse du tau est d'environ 1.777 GeV [21], ce qui lui permet non seulement d'avoir des modes de désintégrations entièrement leptoniques (électron et neutrinos ou muon et neutrinos) mais aussi des modes semi-leptoniques en hadrons et neutrino.

Les neutrinos sont des particules qui n'interagissent que par la force faible. Leur probabilité d'interaction avec la matière est si faible que les neutrinos sont indétectables par CMS ou n'importe quelle autre expérience du LHC. Par commodité, les modes de désintégration du lepton tau sont nommés en fonction de leur(s) produit(s) de désintégration *visible(s)*, c'est-à-dire en ignorant les neutrinos. Il y a ainsi un mode électronique ($\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$), un mode muonique ($\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$) et plusieurs modes hadroniques ($\tau^- \rightarrow h^- \nu_\tau$, $\tau^- \rightarrow h^- h^0 \nu_\tau$ etc.). Dans la suite de ce mémoire, nous écrirons $\tau \rightarrow e$ pour le mode électronique, $\tau \rightarrow \mu$ pour le mode muonique et $\tau \rightarrow h$ pour le mode hadronique, la présence des neutrinos étant implicite.

Les différents rapports de branchement se trouvent dans la table 1.3, nous pouvons y constater que les modes hadroniques sont nettement dominants avec un rapport de branchement presque 4 fois supérieur au mode électronique ou au mode muonique.

7. Les particules visibles sont définies comme toutes les particules neutres et chargées reconstruites par CMS à l'aide des données de tous les sous-détecteurs.

Mode de désintégration du τ^-	Résonance (masse en MeV)	R en %
$e^- \bar{\nu}_e \nu_\tau$		17,8
$\mu^- \bar{\nu}_\mu \nu_\tau$		17,4
$h^- \nu_\tau$		11,5
$h^- \pi^0 \nu_\tau$	$\rho(770)$	26,0
$h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	10,8
$h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9,8
Autres modes hadroniques		6,6
Tous modes hadroniques confondus		64,8

TABLE 1.3 – Modes de désintégration du lepton tau. R est le rapport de branchement. Les modes hadroniques sont des désintégrations en pions ou kaons. Ici $h^{+/-}$ représente des pions et des kaons chargés. Les éventuelles résonances dans la colonne du milieu correspondent à des hadrons instables produits lors de la désintégration du tau. Ces résonances ont une durée de vie très courte et se désintègrent en deux ou trois hadrons plus légers.

La reconstruction et l'identification des leptons taus qui se désintègrent en électron (appelés τ_e) ou en muon (τ_μ) utilise les algorithmes standards de CMS pour les électrons et les muons. L'identification des taus qui se désintègrent en hadrons (appelés taus hadroniques ou τ_h par la suite) se fait avec l'algorithme *Hadrons Plus Strips* (HPS) de CMS. Toute la deuxième partie de ce mémoire concerne l'étude de l'efficacité d'identification des taus hadroniques par l'algorithme HPS.

L'algorithme HPS utilise comme point de départ des jets reconstruits par la méthode *anti- k_T* avec un paramètre de distance $R = 0.5$ [57]. Les particules composant le jet sont issues de l'algorithme Particle Flow (PF). L'algorithme HPS procède en deux étapes : il reconstruit d'abord les pions neutres (π^0) et hadrons chargés composant le jet, et combine ces particules afin de déterminer le mode de désintégration du tau et de reconstruire sa quadri-impulsion et son isolation (l'isolation est une quantité qui caractérise combien d'autres particules énergétiques sont présentes dans un certain cône autour de la particule étudiée). Ensuite des discriminateurs contre les électrons, muons et jets QCD sont appliqués afin de réduire les taux de mauvaise identification. Ces deux étapes, appelées respectivement *Decay Mode Finding* et *Identification*, sont décrites plus en détail ci-dessous.

Decay Mode Finding

Pour trouver le mode de désintégration du tau hadronique, l'algorithme HPS doit identifier les π^0 et les hadrons chargés. La reconstruction de ces deux types de particules se fait de façon très différente.

Les π^0 ont un temps de vie très court, de l'ordre de $8.5 \cdot 10^{-17}$ s, et se désintègrent quasi-exclusivement (rapport de branchement de 98,8%) en deux photons [21]. Il est donc important de bien identifier les photons. L'algorithme tient compte des conversions $\gamma \rightarrow e^+e^-$ qui peuvent se produire par interaction avec le tracker. Les électrons et positrons ainsi produits peuvent à leur tour émettre un photon par Bremsstrahlung. Ces gerbes électromagnétiques issues d'un seul photon ont une largeur importante en ϕ car les électrons et positrons sont courbés par le champ magnétique de 3.8 T de CMS. C'est pourquoi les photons sont recons-

truits en "strips" par HPS : une bande étroite en η ($\Delta\eta = 0.05$) et large en ϕ ($\Delta\phi = 0.20$). Un strip est reconstruit de la manière suivante [58] :

1. Le strip est centré sur la particule électromagnétique la plus énergétique du jet. Le quadrivecteur du strip est alors par définition le quadrivecteur de cette particule.
2. Si d'autres particules électromagnétiques sont trouvées à l'intérieur du strip, la plus énergétique d'entre elles est associée au strip. Le strip est alors recentré et son quadrivecteur est recalculé sur base de l'ensemble des particules du strip.
3. Répéter le point précédent jusqu'à ce qu'il ne soit plus possible d'ajouter des particules au strip.
4. Seuls les strips ayant une quantité de mouvement transverse $p_T > 2.5$ GeV sont gardés.

Les strips et les traces de hadrons chargés sont combinés et une hypothèse sur le mode de désintégration du tau hadronique est formulée. Sur base de cette hypothèse, la quadri-impulsion du candidat- τ_h est calculée, en sommant les quadri-impulsions des strips et hadrons composant le candidat- τ_h . Dans le cas où deux hypothèses différentes de reconstruction sont valables, celle qui donne le plus grand p_T du tau est retenue. Une série de conditions sont imposées :

- La collimation. Les particules formant le candidat- τ_h doivent se trouver à l'intérieur d'un cône (de taille variable selon le p_T du tau) centré autour du quadrivecteur impulsion du candidat- τ_h .
- La masse du candidat- τ_h doit être compatible avec celle d'un τ . Seules les candidats- τ_h ayant une masse incluse dans un certain intervalle sont acceptés. Les bornes de l'intervalle de masse varient en fonction du p_T du candidat- τ_h et de l'hypothèse sur le mode de désintégration.
- Si le mode de désintégration inclut une résonance intermédiaire, la masse invariante des hadrons et strips doit être compatible avec la masse de cette résonance (cf. la table 1.3).
- Les traces des particules chargées composant le candidat- τ_h doivent être compatibles avec le vertex de production présumé du τ_h . Ceci diminue considérablement les effets du pile up.

Ensuite, les candidats- τ_h restants doivent passer des discriminateurs anti-électron, anti-muon et anti-jet dans l'étape d'identification, décrite ci-après.

Identification

L'objectif de l'étape d'identification est de diminuer autant que possible les taux de mauvaise identification $\text{jet} \rightarrow \tau_h$, $e \rightarrow \tau_h$ et $\mu \rightarrow \tau_h$ tout en gardant une efficacité d'identification des τ_h relativement élevée. Cela se fait à l'aide de discriminateurs anti-électrons, anti-muons et anti-jets.

Discriminateur anti-électrons

Pour éliminer les électrons, une méthode multivariée (MVA) basée sur l'apprentissage par arbre de décision a été développée. L'apprentissage par arbre de décision est un type d'apprentissage automatique. Dans le cas du discriminateur anti-électron, son but est de créer un modèle qui sépare d'un côté les candidats- τ_h qui ont une forte probabilité de contenir un électron (considérés comme faisant partie du bruit de fond) et de l'autre tous les autres candidats- τ_h (considérés comme faisant partie du signal). Une série de variables sont fournies afin de discriminer les candidats contenant des électrons de ceux qui n'en ont pas. Cette discrimination se fait par des questions binaires successives ("noeuds"), selon un arbre logique : on part du noeud-"racine" et l'on suit la branche de gauche ou de droite en fonction de la réponse à la première question. Ensuite, on passe à la deuxième question, qui dépend de la réponse à la première question. On suit l'une ou l'autre branche en fonction de la réponse à cette deuxième question, et l'on continue à répondre aux questions jusqu'à ce qu'il n'y en ait plus et qu'on atteigne un verdict final (une "feuille" de l'arbre), à savoir déterminer si le candidat- τ_h contient ou pas un électron. Chacune des questions est basée sur l'une des variables fournies à l'algorithme : pour chacune des variables du candidat- τ_h , il y a une question qui évalue si elle est supérieure ou inférieure à une certaine valeur seuil (cf. la figure 1.10 pour un exemple d'arbre de décision). Ces valeurs seuils sont ensuite optimisées afin de maximiser l'efficacité d'identification du τ_h tout en minimisant le taux de mauvaise identification $e \rightarrow \tau_h$. En pratique, on optimise les valeurs seuils dans plusieurs arbres à la fois, et à la fin de ce processus on choisit le meilleur arbre.

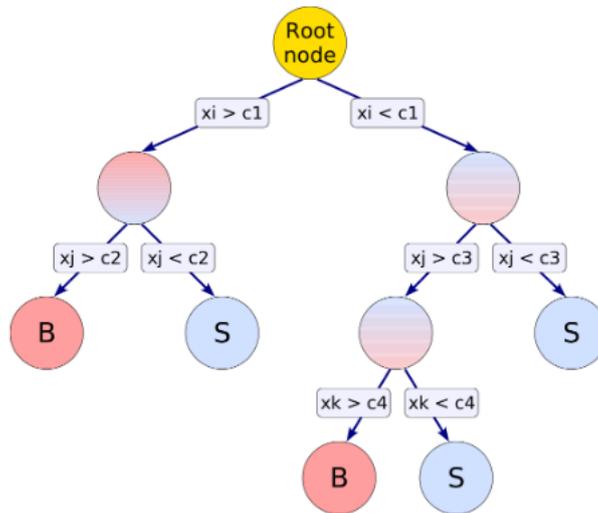


FIGURE 1.10 – Exemple d'un arbre de décision [59]. L'algorithme commence à la racine (*root node*) et évalue si la première variable (x_i) est supérieure ou inférieure à la valeur seuil c_1 . Selon la réponse à cette première question, une valeur seuil différente (c_2 ou c_3) sera utilisée pour la question portant sur la deuxième variable x_j . Après un nombre variable de questions, on atteint l'une des feuilles, qui est soit marquée d'un B (pour *background* ou bruit de fond) ou d'un S (pour signal).

Le discriminateur anti-électron se base sur des variables qui caractérisent l'étendue et la

forme des dépôts d'énergie dans l'ECal (ce qui permet de distinguer les gerbes hadroniques des gerbes électromagnétiques), ainsi que des variables liées au nombre total de particules et au Bremsstrahlung de la trace principale [58]. Plusieurs sélections standards différentes (en anglais *Working Point* ou WP) sont disponibles. Le choix du WP est un compromis entre l'efficacité d'un côté et le taux de mauvaise identification de l'autre. Une efficacité plus élevée s'obtient au prix d'un taux de mauvaise identification également plus élevé. Pour l'analyse réalisée dans ce mémoire, le WP *Very Loose* (efficacité très élevée) a été choisi, car il est déjà suffisant pour réduire le taux de mauvaise identification $e \rightarrow \tau_h$ à des proportions négligeables.

Discriminateur anti-muons

Pour le discriminateur anti-muons, une méthode basée sur des coupures sur chaque variable est utilisée. Cette méthode rejette des candidats- τ_h si des signaux sont détectés dans les chambres à muons dans une direction proche de celle du candidat- τ_h [58]. Comme pour le discriminateur anti-électrons, plusieurs sélections standards différentes sont disponibles. La sélection *tight* a été choisi pour ce mémoire, il est particulièrement important d'avoir un faible taux de mauvaise identification $\mu \rightarrow \tau_h$ car le bruit de fond $Z \rightarrow \mu\mu$ est important dans l'analyse réalisée dans le chapitre 2.

Discriminateur anti-jets ou isolation du tau hadronique

Pour discriminer contre les jets, un algorithme de sélection est également employé. L'algorithme accepte ou rejette des candidats- τ_h en fonction de son isolation. L'isolation est une variable qui représente l'énergie qui a été déposée dans une région voisine à la particule dans le détecteur. Une isolation élevée signifie que la particule étudiée est entourée par un grand nombre d'autres particules énergétiques. Ce genre de configuration est beaucoup plus probable si la particule étudiée fait partie d'un jet, une coupure sur l'isolation est donc un bon moyen de discriminer contre les jets.

L'isolation est une variable sensible au pile up (PU), le nombre de collisions proton-proton par croisement de paquets. En effet, les événements de PU constituent une activité parasite qui contamine le calcul de l'isolation : des particules provenant d'une autre collision proton-proton peuvent se trouver par hasard près d'un tau hadronique isolé. Une correction sensible au PU est donc nécessaire. Le nombre d'événements de PU est estimé à partir du nombre de vertex primaires reconstruit dans l'événement.

Pour calculer l'isolation, il faut d'abord sommer les p_T de toutes les particules chargées reconstruites par l'algorithme PF qui se trouvent dans un cône d'isolation de $\Delta R < 0.5$ autour du candidat- τ_h , les particules utilisées pour construire le candidat- τ_h sont exclues de cette somme, ainsi que les particules ayant un $p_T < 0.5$ GeV. L'effet du PU est réduit en excluant de la somme les particules ayant des trajectoires non-compatibles avec le vertex de production du τ_h . On somme ensuite le p_T de tous les photons ayant une énergie transverse $E_T > 0.5$ et se trouvant dans le même cône d'isolation $\Delta R < 0.5$ autour du candidat- τ_h (cf. l'équation 1.14). L'effet du PU sur l'isolation des photons est corrigé par la fonction $\Delta\beta$ (cf. l'équation 1.15), qui est construite à partir des particules chargées incompatibles avec le

vertex de désintégration du τ_h [58].

$$\text{Isolation} = \sum p_T^{\text{charged}}(\text{vertex compatible}) + \max\{\sum p_T^\gamma - \Delta\beta, 0\} \quad (1.14)$$

$$\Delta\beta = \alpha \cdot \sum p_T^{\text{charged}}(\text{vertex incompatible}) \quad (1.15)$$

Le paramètre α est choisi de sorte à ce que l'efficacité d'identification soit indépendante du nombre d'événements de PU. Les quantités p_T^{charged} et p_T^γ sont respectivement les impulsions transverses des particules chargées et des photons.

Trois sélections standards (*Working Points* ou WP), avec des efficacités décroissantes, sont disponibles pour le discriminateur anti-jet : *loose*, *medium*, et *tight*. La deuxième partie de ce mémoire consiste à évaluer les efficacités de ces trois WP, d'abord en utilisant des simulations et ensuite en utilisant les données collectées en 2015 par CMS.

Si le candidat- τ_h passe les discriminateurs anti-électrons, anti-muons et anti-jets, il est identifié comme un τ_h , et peut alors être utilisé dans des analyses de physique.

Chapitre 2

Étude de l'efficacité de l'algorithme de reconstruction et d'identification des leptons taus

Plusieurs facteurs déterminent si un lepton tau produit lors d'une collision proton-proton dans CMS est reconstruit et identifié en tant que tel. Un premier facteur est l'*acceptance* géométrique du détecteur : celui-ci ne peut pas recouvrir tout l'angle solide autour de la région d'interaction à cause de la présence du tube à vide. Certaines particules ne vont donc jamais interagir avec le détecteur, car elles se propagent dans la direction du tube à vide. Un deuxième facteur est l'énergie de la particule : si son énergie est trop petite, l'interaction de la particule avec le détecteur peut devenir du même ordre de grandeur que les erreurs de mesure, ce qui la rend indétectable.

Mais même les particules qui sont dans l'acceptance du détecteur et qui ont une énergie suffisamment élevée peuvent ne pas être identifiées. On parle alors de l'*efficacité* de l'identification, c'est-à-dire la proportion de particules correctement identifiées parmi toutes les particules qui sont dans l'acceptance du détecteur et qui ont une énergie suffisamment élevée.

Dans ce chapitre, l'efficacité de la reconstruction et de l'identification des taus hadroniques dans CMS est estimée, d'abord dans les simulations et ensuite dans les données. Plus précisément, nous nous intéressons à l'efficacité combinée de la reconstruction (Decay Mode Finding) et des trois sélections standards (loose, medium et tight) du discriminateur anti-jets de l'algorithme CMS, décrit dans la section 1.5.4.

2.1 Détermination de l'efficacité dans les simulations

L'efficacité de l'identification des taus hadroniques est déterminée dans des événements simulés Drell-Yan (DY) $q\bar{q} \rightarrow Z/\gamma^* \rightarrow \tau^+\tau^-$. Comme il s'agit d'une simulation, tous les événements sont connus au niveau généré. Pour déterminer l'efficacité, il suffit donc de comparer le nombre de taus hadroniques identifiés lors de la reconstruction au nombre de taus hadroniques générés.

La simulation utilisée génère des événements DY ayant pour état final une paire lepton-

antilepton :

$$q\bar{q} \rightarrow Z/\gamma^* \rightarrow l^+l^-$$

Seuls des événements ayant une masse invariante au-dessus de 50 GeV sont générés. Environ $9 \cdot 10^5$ événements ont été générés, mais il n'y a qu'une partie de ces événements qui nous intéressent : ceux qui contiennent des leptons taus, et en particulier ceux qui contiennent des taus hadroniques. Étant donné les rapports de branchement du Z et du lepton tau, $\sim 3 \cdot 10^5$ événements contenant un lepton tau ont été générés, et 88% d'entre eux contiennent au moins un τ_h . La table 2.1 permet de vérifier que, dans la simulation, les nombres d'événements sont les mêmes pour chaque mode de désintégration en paire lepton-antilepton du boson Z, et que les rapports de branchement des modes de désintégration du lepton tau sont bien compatibles avec ceux qui avaient été donnés dans la table 1.3.

	Mode de désintégration	Nombre d'événements	Rapport de branchement en %
Z	e^+e^-	$(2.8352 \pm 0.0017) \cdot 10^6$	3.3
	$\mu^+\mu^-$	$(2.8314 \pm 0.0017) \cdot 10^6$	3.3
	$\tau^+\tau^-$	$(2.8310 \pm 0.0017) \cdot 10^6$	3.3
τ	$e\nu_e\nu_\tau$	$(1.0263 \pm 0.0010) \cdot 10^6$	17.8
	$\mu\nu_\mu\nu_\tau$	$(0.9829 \pm 0.0010) \cdot 10^6$	17.4
	$\nu_\tau + \text{hadrons}$	$(3.6527 \pm 0.0019) \cdot 10^6$	64.8

TABLE 2.1 – Nombres d'événements simulés pour chaque mode de désintégration du lepton tau et chaque mode l^+l^- du boson Z. Les rapports de branchement de chaque mode sont donnés en guise de comparaison.

2.1.1 Cinématique des événements tau-antitau

Avant de passer au calcul de l'efficacité, une analyse cinématique des événements $\tau_\mu\tau_h$ est réalisée. L'objectif de cette analyse est de mieux comprendre les variables caractérisant l'état final $\tau_\mu\tau_h$: les quantités de mouvement transverses (p_T) du τ_h et du muon, leurs directions en η , la masse invariante du système $\mu\tau_h$, le vecteur d'énergie transverse manquante (MET) et la séparation ΔR dans la surface η - ϕ entre les quantités de mouvement du muon et du τ_h .

Une part importante de la quantité de mouvement des leptons taus est invisible car il y a toujours un ou deux neutrinos parmi leurs produits de désintégration. Les variables cinématiques de chaque événement sont donc différentes si l'on ne considère que les particules *visibles* (sans tenir compte des neutrinos) ou si l'on considère toutes les particules générées (neutrinos compris). La distribution des variables cinématiques visibles est aussi différente selon le mode de désintégration des leptons taus parce que l'énergie transverse manquante est plus grande en moyenne lorsque l'un des leptons taus se désintègre de façon entièrement leptonique, car il y a un neutrino de plus dans l'état final. La figure 2.1 donne la distribution de quelques variables au niveau généré, avec une attention toute particulière à l'état final $\tau_\mu\tau_h$, où un lepton tau se désintègre en muon et neutrinos et l'autre en hadrons et neutrino.

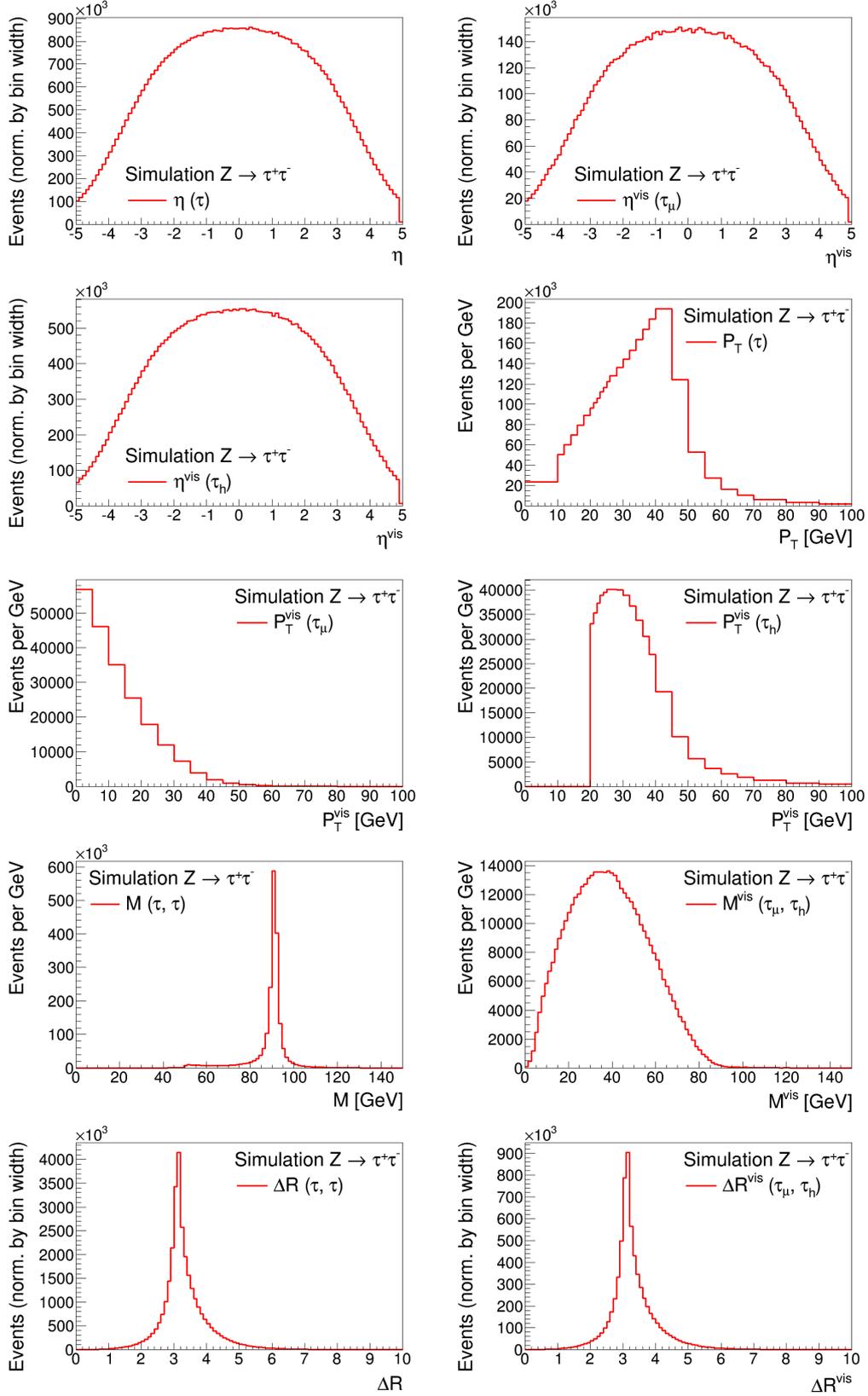


FIGURE 2.1 – Quelques distributions de variable au niveau g n r . De haut en bas et de gauche   droite : η total du tau, η visible du τ_μ , η visible du τ_h , p_T total du tau, p_T visible du τ_μ , p_T visible du τ_h , masse invariante totale du syst me $\tau^+\tau^-$, masse visible du syst me $\tau_\mu\tau_h$, ΔR du syst me $\tau^+\tau^-$, ΔR du syst me visible $\tau_\mu\tau_h$.

2.1.2 Efficacité en fonction de l'impulsion transverse générée du τ_h

Les candidats- τ_h sont reconstruits par l'étape de *Decay Mode Finding* (DMF) de l'algorithme HPS. Un candidat- τ_h est identifié comme un τ_h s'il passe le discriminateur anti-jets¹.

Dans cette section, l'efficacité combinée du DMF et du discriminateur anti-jets dans les événements simulés est calculée. Pour cela, il faut comparer le nombre de taus hadroniques identifiés au nombre de taus hadroniques générés, en tenant compte de l'acceptance du détecteur et de l'énergie minimale pour laquelle l'algorithme d'identification HPS a été conçu. L'algorithme HPS est optimisé pour des τ_h ayant un p_T visible supérieur à 20 GeV et une pseudo-rapacité η dont le module est inférieur à 2.3. Déterminer l'efficacité du discriminateur anti-jets revient donc à compter, parmi les taus hadroniques générés avec $|\eta^{vis}| < 2.3$ et $p_T^{vis} > 20$ GeV, combien d'entre eux ont effectivement été identifiés comme des taus hadroniques par l'algorithme HPS au niveau reconstruit. Il faut aussi tenir compte du fait que les valeurs de η^{vis} et p_T^{vis} sont différentes au niveau généré et au niveau reconstruit. Finalement, l'efficacité combinée ϵ du DMF et du discriminateur anti-jets est définie par l'équation suivante :

$$\epsilon = \frac{\text{sél. du dén.} \wedge p_{T\text{reco}}^{vis} > 20 \text{ GeV} \wedge |\eta_{\text{reco}}^{vis}| < 2.3 \wedge \text{DMF} \wedge \text{discr. anti jets}}{p_{T\text{gen}}^{vis} > 20 \text{ GeV} \wedge |\eta_{\text{gen}}^{vis}| < 2.3 \wedge \Delta R(\text{gen, reco}) < 0.3} \quad (2.1)$$

où le superscript *vis* désigne les quantités visibles, le subscript *reco* désigne des quantités reconstruites, le subscript *gen* désigne des quantités générées et $\Delta R(\text{gen, reco})$ désigne la séparation dans la surface η - ϕ entre les quantités de mouvement du τ_h généré et du τ_h reconstruit. L'efficacité est donnée par le nombre de τ_h passant les coupures du numérateur divisé par le nombre de τ_h passant les coupures du dénominateur.

Il est important de regarder plus particulièrement l'efficacité en fonction de l'impulsion transverse visible du τ_h , afin de s'assurer qu'elle ne varie pas excessivement avec l'énergie du τ_h . C'est ce qui est fait dans la figure 2.2, qui montre l'efficacité combinée de DMF et du discriminateur anti-jets en fonction de l'impulsion transverse visible générée du τ_h . Cette efficacité est globalement constante pour des leptons taus ayant des quantités de mouvement transverses supérieures à 30 GeV. Elle est d'environ 56% pour le WP loose, 50% pour le WP medium et 42% pour le WP tight sur ce plateau. Mais l'efficacité est plus petite pour des $p_{T\text{gen}}^{vis}$ proches de 20 GeV. Cela est dû à la coupure sur le $p_{T\text{reco}}^{vis}$, qui est aussi de 20 GeV. Certains τ_h générés avec un $p_{T\text{gen}}^{vis}$ d'un peu plus de 20 GeV sont reconstruits avec un $p_{T\text{reco}}^{vis}$ de moins de 20 GeV, et ne passent donc pas les sélections du numérateur de l'équation 2.1.

2.2 Détermination de l'efficacité à partir des données

Plusieurs méthodes sont disponibles pour déterminer l'efficacité de reconstruction et d'identification des taus hadroniques, celle qui a été retenue pour ce mémoire est la méthode du rapport $Z \rightarrow \tau_\mu \tau_h / Z \rightarrow \mu\mu$.

La méthode du rapport $Z \rightarrow \tau_\mu \tau_h / Z \rightarrow \mu\mu$ consiste à comparer le nombre d'événements Drell-Yan dans les canaux $Z \rightarrow \tau_\mu \tau_h$ et $Z \rightarrow \mu\mu$ afin de déterminer le rapport r entre

1. Les discriminateurs anti-électron et anti-muons décrits dans la section 1.5.4 sont optionnels.

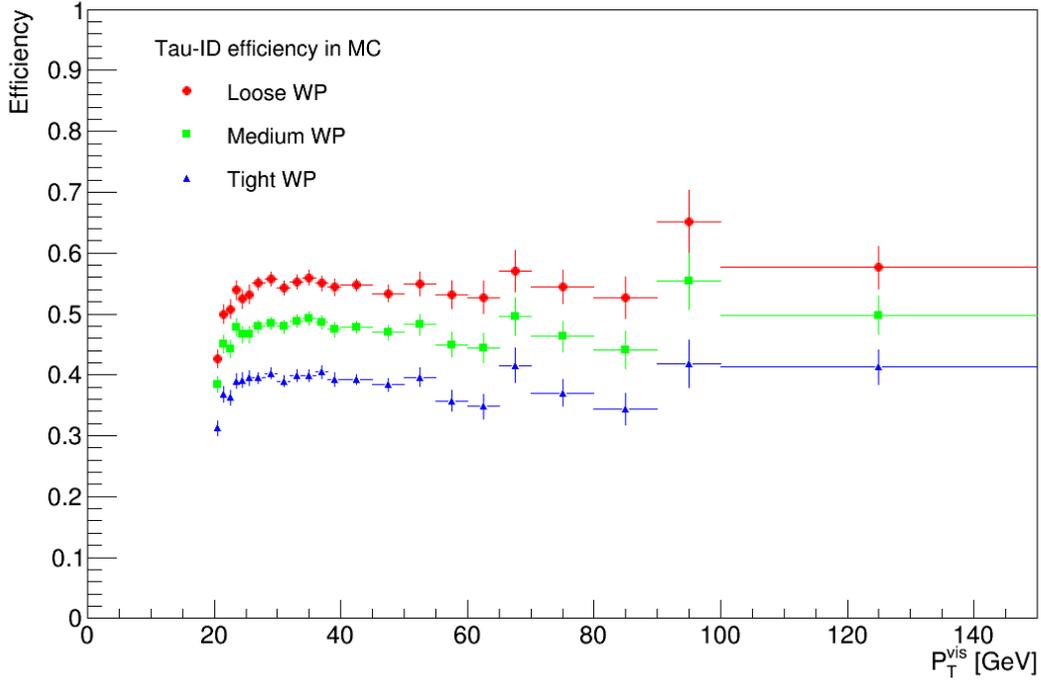


FIGURE 2.2 – Efficacité combiné du DMF et du discriminateur anti-jets pour l’identification des taus hadroniques dans des événements DY simulés. L’efficacité est exprimée en fonction du p_T visible généré du τ_h . L’efficacité atteint un plateau pour des valeurs de p_T supérieures à 30 GeV.

l’efficacité d’identification du τ_h dans les données et cette même efficacité dans les simulations MC :

$$r = \frac{\epsilon_{data}}{\epsilon_{MC}} \quad (2.2)$$

où ϵ_{data} est l’efficacité dans les données et ϵ_{MC} l’efficacité dans les simulations.

Pour des événements générés $N_{Z \rightarrow \tau_\mu \tau_h}^{gen}$ et $N_{Z \rightarrow \mu\mu}^{gen}$, l’équation ci-dessous est vraie :

$$\frac{N_{Z \rightarrow \tau_\mu \tau_h}^{gen}}{N_{Z \rightarrow \mu\mu}^{gen}} \cdot \frac{R(Z \rightarrow \mu\mu)}{R(Z \rightarrow \tau_\mu \tau_h)} = 1 \quad (2.3)$$

où R est le rapport de branchement du canal considéré. Remarquons que :

$$\frac{R(Z \rightarrow \tau_\mu \tau_h)}{R(Z \rightarrow \mu\mu)} = \frac{2 \cdot R(Z \rightarrow \tau\tau) \cdot R(\tau \rightarrow \tau_\mu) \cdot R(\tau \rightarrow \tau_h)}{R(Z \rightarrow \mu\mu)} = 2 \cdot R(\tau \rightarrow \tau_\mu) \cdot R(\tau \rightarrow \tau_h) \simeq 0.225 \quad (2.4)$$

où nous avons utilisé l’universalité des couplages du boson Z aux leptons : $R(Z \rightarrow \tau\tau) = R(Z \rightarrow \mu\mu)$, et les valeurs de la table 1.3 pour les rapports de branchement du lepton tau. Le facteur 2 est un facteur combinatoire qui apparaît car tout aussi bien le tau que l’antitau produits lors de la désintégration du Z peuvent se désintégrer en muon (l’autre se désintégrant alors en hadrons).

Lorsque l'on calcule le membre de gauche de l'équation 2.3 avec des événements extraits des données, tout écart par rapport à 1 doit nécessairement venir de variables qui sont *différentes* pour les deux canaux. Toute variable identique est éliminée car c'est le rapport entre les deux canaux qui est calculé. Il y a des différences dans l'acceptance et l'efficacité des sélections entre les deux canaux, mais la différence principale est due à l'algorithme de reconstruction et d'identification des τ_h , qui a une efficacité proche de 50 %, contre une efficacité de plus de 95 % pour l'identification des muons (cf. la section 2.2.1). L'efficacité de reconstruction et d'identification des τ_h peut donc être déduite à partir du rapport du nombre d'événements dans les deux canaux. Une première approximation attribuerait la totalité de l'écart par rapport à 1 à cette efficacité :

$$\epsilon_{\tau_h} = \frac{N_{Z \rightarrow \tau_\mu \tau_h}^{data}}{N_{Z \rightarrow \mu\mu}^{data}} \cdot \frac{R(Z \rightarrow \mu\mu)}{R(Z \rightarrow \tau_\mu \tau_h)} \quad (2.5)$$

où ϵ_{τ_h} est l'efficacité de reconstruction et d'identification du τ_h et N^{data} représente le nombre d'événements dans les données. Une approche plus rigoureuse consiste à considérer les autres différences entre les deux canaux. Les nombres d'événements dans les deux canaux peuvent s'écrire par les équations suivantes :

$$N_{Z \rightarrow \tau_\mu \tau_h}^{data} = N_{Z \rightarrow X}^{tot} \cdot R(Z \rightarrow \tau_\mu \tau_h) \cdot \alpha_{\tau_\mu \tau_h} \cdot \epsilon_{\tau_\mu \tau_h \text{ other}} \cdot \epsilon_{\tau_h} \quad (2.6)$$

$$N_{Z \rightarrow \mu\mu}^{data} = N_{Z \rightarrow X}^{tot} \cdot R(Z \rightarrow \mu\mu) \cdot \alpha_{\mu\mu} \cdot \epsilon_{\mu\mu} \quad (2.7)$$

où $N_{Z \rightarrow X}^{tot}$ est le nombre total de désintégrations de boson Z, $\alpha_{\mu\mu}$ et $\alpha_{\tau_\mu \tau_h}$ sont respectivement les acceptances des canaux $\mu\mu$ et $\tau_\mu \tau_h$, $\epsilon_{\tau_\mu \tau_h \text{ other}}$ représente l'efficacité des autres sélections dans le canal $\tau_\mu \tau_h$ qui ne sont pas appliquées dans le canal $\mu\mu$, et $\epsilon_{\mu\mu}$ est l'efficacité des sélections dans le canal $\mu\mu$ qui ne sont pas appliquées dans le canal $\tau_\mu \tau_h$. En divisant la première équation par la deuxième et en réarrangeant les termes, on obtient :

$$\epsilon_{\tau_h} = \frac{N_{Z \rightarrow \tau_\mu \tau_h}^{data}}{N_{Z \rightarrow \mu\mu}^{data}} \cdot \frac{R(Z \rightarrow \mu\mu)}{R(Z \rightarrow \tau_\mu \tau_h)} \cdot \frac{\epsilon_{\mu\mu}}{\epsilon_{\tau_\mu \tau_h \text{ other}}} \cdot \frac{\alpha_{\mu\mu}}{\alpha_{\tau_\mu \tau_h}} \quad (2.8)$$

En pratique, ce rapport n'est pas calculé, mais une approche similaire est utilisée : l'efficacité est déterminée par un ajustement simultané des deux canaux. L'ajustement se fait avec une méthode de maximum de vraisemblance. La vraisemblance représente ici l'accord entre les simulations et les données. Un certain nombre de *paramètres de nuisance* sont introduits. Ce sont des paramètres qui, au cours de l'ajustement, peuvent varier autour d'une valeur centrale, avec un certain écart-type. Toutes choses étant égales par ailleurs, plus leur variation est grande, moins la vraisemblance associée sera grande. Le paramètre r est le *paramètre d'intérêt*, et n'a pas de contrainte sur ses variations. L'ajustement consiste à varier les paramètres de nuisance et le paramètre d'intérêt et de trouver la configuration qui maximise la vraisemblance. Le canal $Z \rightarrow \mu\mu$ sert à contraindre les paramètres de nuisance qui sont communs aux deux canaux. L'ajustement dans le canal $Z \rightarrow \tau_\mu \tau_h$ permet d'obtenir le paramètre d'intérêt, étant donné les paramètres de nuisances contraints par le canal $Z \rightarrow \mu\mu$.

Cet ajustement simultané des deux canaux a été réalisé avec l'outil statistique *Combine* de CMS. Les paramètres de nuisance utilisés et d'autres détails sur l'ajustement sont donnés dans la section 2.2.6.

2.2.1 Sélections dans la région $\tau_\mu\tau_h$

Les principaux bruits de fonds dans cette analyse sont les processus DY autres que $Z \rightarrow \tau_\mu\tau_h$ (en particulier le processus $Z \rightarrow \mu\mu$), ainsi que les processus multijets, W +jets et $t\bar{t}$ (cf. la section 1.1.3 pour plus de détails sur ces processus). Une série de sélections sont utilisées afin de minimiser le nombre d'événements de bruit de fond par rapport aux événements de signal $Z \rightarrow \tau_\mu\tau_h$.

Système de déclenchement

Le système de déclenchement utilisé demande la présence d'un muon isolé avec $p_T > 18$ GeV et $|\eta| < 2.1$.

Coupages sur le muon

Le processus qui nous intéresse, $Z \rightarrow \tau_\mu\tau_h$, contient un muon dans l'état final. Les coupures appliquées sur le muon sont strictes afin de profiter de l'excellente efficacité de la reconstruction et de l'identification des muons dans CMS (toujours $> 95\%$, les chiffres exacts sont donnés ci-dessous), et des faibles taux de mauvaise identification ($< 1\%$).

- $p_T > 20$ GeV. Cette coupure se fait à une valeur de p_T plus grande que le seuil de déclenchement. Cela est dû au fait que l'efficacité du déclenchement diminue lorsque le p_T est proche du seuil. En effet, certains muons produits avec plus de 18 GeV lors d'une collision proton-proton sont reconstruits avec moins de 18 GeV et ne passent pas les critères de déclenchement. L'efficacité du déclenchement atteint un plateau pour des valeurs de p_T suffisamment élevées. La coupure de 20 GeV choisie se trouve au début de ce plateau.
- $|\eta| < 2.1$. Cette coupure est la même que pour le système de déclenchement.
- La reconstruction et l'identification standard des muons de CMS sont utilisées. L'efficacité de cette sélection est de plus de 99% pour des muons avec $p_T > 20$ GeV [52].
- L'isolation relative combinée I doit être inférieure à 0.1. Cette variable est calculée d'une façon similaire à celle de l'isolation du tau hadronique, dont la formule est donnée par l'équation 1.14. La coupure $I < 0.1$ a une efficacité d'environ 97% [52].

Coupages sur le tau hadronique

Les coupures précédentes servent à s'assurer que l'événement contient un muon. Il faut aussi s'assurer que l'événement contient un tau hadronique.

- $p_T > 20$ GeV. Cette coupure est utilisée car l'algorithme a été conçu pour $p_T > 20$ GeV. Ses résultats ne sont pas valides pour des p_T plus petits.
- $|\eta| < 2.3$. L'efficacité de la reconstruction des traces diminue près de la frontière de l'acceptance géométrique du détecteur, ce qui fait que le calcul de l'isolation du τ_h (cf. ci-dessous) devient moins précis à grand $|\eta|$.

- Discriminateur anti-muons, sélection *tight*. Cette sélection permet de réduire le nombre d'événements $Z \rightarrow \mu\mu$, où un muon serait mal identifié en tant que tau hadronique. Cette coupure est nécessaire car le nombre d'événements $Z \rightarrow \mu\mu$ attendus est élevé. En effet, le système de déclenchement et les autres sélections appliquées ici ne permettent pas d'éliminer ces événements car ils exigent la présence d'un muon énergétique, et le rapport de branchement du canal $Z \rightarrow \mu\mu$ est 4.4 fois plus élevé que celui du canal $Z \rightarrow \tau_\mu\tau_h$ (cf. l'équation 2.4).
- Discriminateur MVA anti-électrons, sélection *loose*. Cette sélection réduit le taux de mauvaise identification d'électrons en taus hadroniques. Un discriminateur plus strict contre les électrons n'est pas nécessaire car le nombre d'événements attendus contenant des électrons n'est pas grand. En effet, aucune sélection utilisée ici n'exige la présence d'un électron, contrairement au cas des muons. Et très peu d'événements de bruit de fond contiennent à la fois un électron et un muon dans l'état final, le principal processus éliminé par ce discriminateur est le processus $Z \rightarrow \tau_\mu\tau_e$.
- Discriminateur anti-jets. Il s'agit d'une coupure sur l'isolation du τ_h , présentée dans la section 1.5.4. Trois sélections ou WP sont disponibles pour ce discriminateur : *loose*, *medium* et *tight*. Son taux de mauvaise identification (jet $\rightarrow \tau_h$) est typiquement de l'ordre de 1% et varie entre $\sim 10^{-4}$ et 4% en fonction du pile up, et du p_T et du η et du jet [58]. L'efficacité de ce discriminateur est étudiée dans ce chapitre.

Il reste encore une quantité d'événements de bruit de fond du même ordre de grandeur que celui d'événements de signal, malgré les sélections sur le muon et le tau hadronique. Des coupures supplémentaires sont nécessaires.

Autres coupures

Les coupures suivantes sont utilisées afin de réduire encore le bruit de fond :

- Signes opposés pour les charges du muon et du tau hadronique. Dans le processus qui nous intéresse, $Z \rightarrow \tau_\mu\tau_h$, le muon et le tau hadronique proviennent de la désintégration d'un boson Z , une particule neutre. Dans ce cas, ils doivent forcément avoir des charges opposées.
- $\Delta R(\mu, \tau_h) > 0.5$. Cette coupure permet d'éliminer des événements où un jet identifié comme un tau hadronique aurait une particule identifiée comme un muon en son sein. Cela n'élimine pas beaucoup d'événements du signal car le τ_h et le muon ont tendance à être produits dos-à-dos (cf. la figure 2.1).
- Veto sur les électrons. Si l'événement contient au moins un électron identifié par le WP90 (90 % d'efficacité) et ayant un $p_T > 15$ GeV et $|\eta| < 2.4$, il est éliminé.
- Veto sur un second muon. Si l'événement contient au moins deux muons identifiés par la sélection standard de CMS et avec $p_T > 5$ GeV et $|\eta| < 2.4$, il est rejeté. La présence d'un deuxième muon énergétique est une caractéristique du processus $Z \rightarrow \mu\mu$ qui ne se retrouve pas dans le signal. Ce veto permet donc aussi d'éliminer une partie du bruit de fond avec peu de perte d'efficacité pour le signal.
- Veto sur les jets de quark b. Ce veto sert à rejeter les événements $t\bar{t}$. En effet, le quark top est instable et se désintègre avant de pouvoir s'hadroniser. Son mode de

désintégration principal est vers un quark b accompagné d'un boson W^2 . Les jets b sont identifiés par une sélection ayant une efficacité d'environ 60%.

- Masse transverse $M_T(\mu, \text{MET}) < 40$ GeV. Cette quantité correspond à la masse invariante du système énergie transverse manquante (MET) + composantes transverses de la quadri-impulsion du muon. Elle se calcule par la formule suivante :

$$M_T = \sqrt{2 \cdot p_T(\mu) \cdot E_T^{miss} \cdot (1 - \cos \Delta\phi)} \quad (2.9)$$

où $p_T(\mu)$ est la quantité de mouvement transverse du muon, E_T^{miss} la norme du vecteur de MET, et $\Delta\phi = \phi(\mu) - \phi(\text{MET})$ la différence entre la direction en ϕ de la quantité de mouvement du muon et celle du vecteur d'énergie transverse manquante. Cette coupure sur M_T est utilisée pour discriminer contre les événements W +jets. En effet, la plupart des événements W +jets qui passent les sélections présentées ici sont des événements où le boson W se désintègre en muon et neutrino muonique. Dans ce cas, la masse invariante du système muon - neutrino muonique correspond à la masse du W . Mais cette masse n'est pas directement calculable car les neutrinos sont invisibles pour un détecteur comme CMS. C'est pourquoi la masse transverse est calculée. Pour des bosons W produits dans une direction proche du plan transverse, cette masse transverse est élevée, proche de la masse du W , environ 80 GeV. Pour des bosons W avec des composantes longitudinales de l'impulsion plus grandes, cette masse transverse est plus petite. Par contre, la masse transverse n'est pas forcément élevée pour le processus $Z \rightarrow \tau_\mu \tau_h$ car trois neutrinos sont produits, dans des directions différentes, et le système muon + MET n'est pas issu pas d'une seule particule massive. La coupure sur M_T permet donc d'éliminer plus d'événements W +jets que de signal.

Dans la mesure du possible, nous cherchons à décrire au mieux les bruits de fond à partir des données plutôt que par des simulations. Les simulations des bruits de fond, après sélection, concernent un espace de phase particulier qui peut ne pas bien décrire les données, en particulier dans le cas des bruits de fond avec mauvaise identification jet $\rightarrow \tau_h$. Ci-dessous, nous expliquons comment les données sont utilisées pour modifier la description des processus W +jets et multijets. Pour le processus $t\bar{t}$, les simulations sont gardées sans modification car il ne contribue que faiblement au nombre d'événements qui passent les sélections.

2.2.2 Détermination du facteur d'échelle pour les processus W +jets

Toute simulation d'un processus contient un certain nombre d'événements générés N_{gen} . Tout histogramme obtenu dans une simulation doit être multiplié par un poids w afin de pouvoir décrire les données :

$$w = \frac{\sigma \cdot L}{N_{gen}} \quad (2.10)$$

où σ est la section efficace du processus et L la luminosité des données. La section efficace du processus W +jets est calculée avec une certaine erreur et pourrait ne pas être compatible avec le nombre d'événements observés dans les données. Le poids w est multiplié

2. Le rapport de branchement de $t \rightarrow Wb$ est de 91 ± 4 % parmi tous les modes de désintégration $t \rightarrow Wq$, où $q = b, c$ ou d [21].

par un facteur d'échelle f mesuré à partir des données afin d'ajuster la normalisation de la simulation au nombre d'événements observés dans les données :

$$w \rightarrow w' = f \cdot w \quad (2.11)$$

Pour calculer ce facteur d'échelle avec le moins d'incertitude possible, il faut sélectionner une région des données qui soit riche en événements W+jets et pauvre en événements de tous les autres processus. Une telle région peut s'obtenir en gardant toutes les sélections décrites dans la section précédente, sauf la coupure en M_T . Les événements W+jets sont largement dominants à grand M_T , nous choisissons donc la coupure $M_T > 80$ GeV pour cette partie du travail. L'autre sélection qui change par rapport à celles présentées dans la section 2.2.1 est le discriminateur anti-jets, qui n'est pas appliqué ici afin de garder un grand nombre d'événements W+jets. En résumé, les sélections appliquées pour calculer le facteur d'échelle de W+jets sont les suivantes :

- Coupure sur M_T changée vers $M_T > 80$ GeV.
- Discriminateurs anti-jets pas appliqué.
- Autres sélections : identiques à celles de la région $\tau_\mu\tau_h$.

Le facteur d'échelle s'obtient en comparant le nombre d'événements dans les données au nombre d'événements W+jets prévus dans les simulations. Le nombre d'événements dans les données doit être corrigé en soustrayant le nombre d'événements prévus pour les processus autres que W+jets. L'équation utilisée pour déterminer le facteur d'échelle f est la suivante :

$$f = \frac{N_{\text{data}} - N_{\text{DY}} - N_{t\bar{t}} - N_{\text{multijets}}}{N_{\text{Wjets}}} \quad (2.12)$$

où N_{data} est le nombre de données dans l'échantillon, et N_{DY} , $N_{t\bar{t}}$, $N_{\text{multijets}}$ et N_{Wjets} sont respectivement le nombre total d'événements Drell-Yan (tout aussi bien $\tau_\mu\tau_h$ que $\mu\mu$), $t\bar{t}$, multijets et W+jets prévus dans les simulations. La simulation du processus multijets est peu précise³, mais cela a peu d'influence sur le résultat ici car il y a peu d'événements multijets prévus.

Des fluctuations statistiques sur le nombre de données et le nombre d'événements dans chacun de ces processus contribuent à l'erreur sur le facteur d'échelle, mais ces erreurs statistiques sont négligeables face à l'incertitude sur la valeur de l'énergie transverse manquante, qui entre en jeu dans le calcul de M_T (cf. l'équation 2.9). L'erreur sur l'énergie transverse manquante est de l'ordre de 20% [60]. Pour déterminer l'incertitude sur le facteur d'échelle, nous calculons M_T de trois façons différentes : avec la valeur centrale de MET, telle qu'elle a été reconstruite par CMS (cf. la section 1.5.3), avec une valeur 20% plus grande et avec une valeur 20% plus petite. La masse transverse M_T prend trois valeurs différentes, et des événements différents seront sélectionnés selon la valeur de M_T . La valeur centrale du facteur d'échelle est celle calculée à partir de l'énergie transverse manquante telle qu'elle a été reconstruite par CMS, et l'erreur est estimée comme la plus grande différence entre la valeur centrale et l'une des deux autres valeurs.

3. Ce processus sera d'ailleurs entièrement estimé à partir des données, cf. la section suivante.

facteurs d'échelles		
MET centrale	MET -20%	MET +20%
1.03665	1.06511	1.02947

TABLE 2.2 – Facteurs d'échelles calculés avec des valeurs différentes de MET. D'abord avec la valeur centrale reconstruite par CMS et ensuite avec des variations de $\pm 20\%$. Le facteur d'échelle obtenu est $f = 1.037 \pm 0.028$.

La table 2.2 donne les trois valeurs de facteur d'échelle calculées. Le facteur d'échelle trouvé est :

$$f = 1.037 \pm 0.028 \quad (2.13)$$

Mais l'erreur de 0.028 ne sera pas utilisée dans la suite de ce mémoire, une erreur plus conservatrice de 10% est choisie, correspondant à l'erreur standard utilisée par CMS [61]. En effet notre estimation de l'erreur sur l'énergie transverse manquante est peu précise, il faudrait tenir compte du fait que cette erreur varie événement par événement. La précision sur l'énergie transverse manquante est notamment très sensible au nombre d'événements simultanés dans le détecteur (événements de pile up) [60]. Il reste aussi d'autres sources d'erreurs systématiques que nous n'avons pas considérées : les sélections sont différentes par rapport à la région $\tau_\mu\tau_h$, la coupure sur M_T a été changée et aucune coupure sur l'isolation du τ_h n'est appliquée. Cela peut introduire des différences dans la normalisation.

La figure 2.3 montre les distributions de quelques variables dans la région des données avec $M_T > 80$ GeV et aucune coupure sur l'isolation du τ_h , après application du facteur d'échelle donné dans l'équation 2.13.

Pour certains histogrammes, notamment celui du p_T du muon et celui de l'énergie transverse manquante, la forme des données n'est pas bien décrite. C'est le cas notamment pour des petites valeurs de MET. Cela s'explique par le fait que l'énergie transverse manquante est une variable difficile à décrire dans les simulations, surtout pour des petites valeurs. En effet, le vecteur d'énergie transverse manquante est obtenu en prenant l'opposé de la somme vectorielle des quantités de mouvement transverses de toutes les particules visibles. Toute erreur de mesure de ces quantités de mouvement se propage dans le calcul de MET. Ce calcul est encore compliqué par les événements de pile up. Le calcul de MET doit être corrigé en tenant compte du pile up, mais cette correction n'est pas sans erreur. La référence [60] a mesuré dans les données l'écart-type des composantes x et y de MET dans des événements où l'énergie manquante est censée être nulle (les sélections ont été optimisées pour des événements où toutes les particules de l'état final sont visibles). L'écart-type obtenu est de 15 GeV. Cela suggère que l'imprécision sur MET est de l'ordre de 50% - 100% pour des valeurs de 0 à 30 GeV.

Quant à l'histogramme du p_T du muon, le rapport données sur simulation augmente continûment avec le p_T . Cela est dû à certains facteurs de la reconstruction et de l'identification des muons dont l'efficacité dépend du p_T : le système de déclenchement, le calcul de l'isolation et l'identification d'un muon par l'algorithme Particle Flow. Il faudrait appliquer des corrections liées à ces facteurs afin de bien décrire la forme de la distribution du p_T du muon.

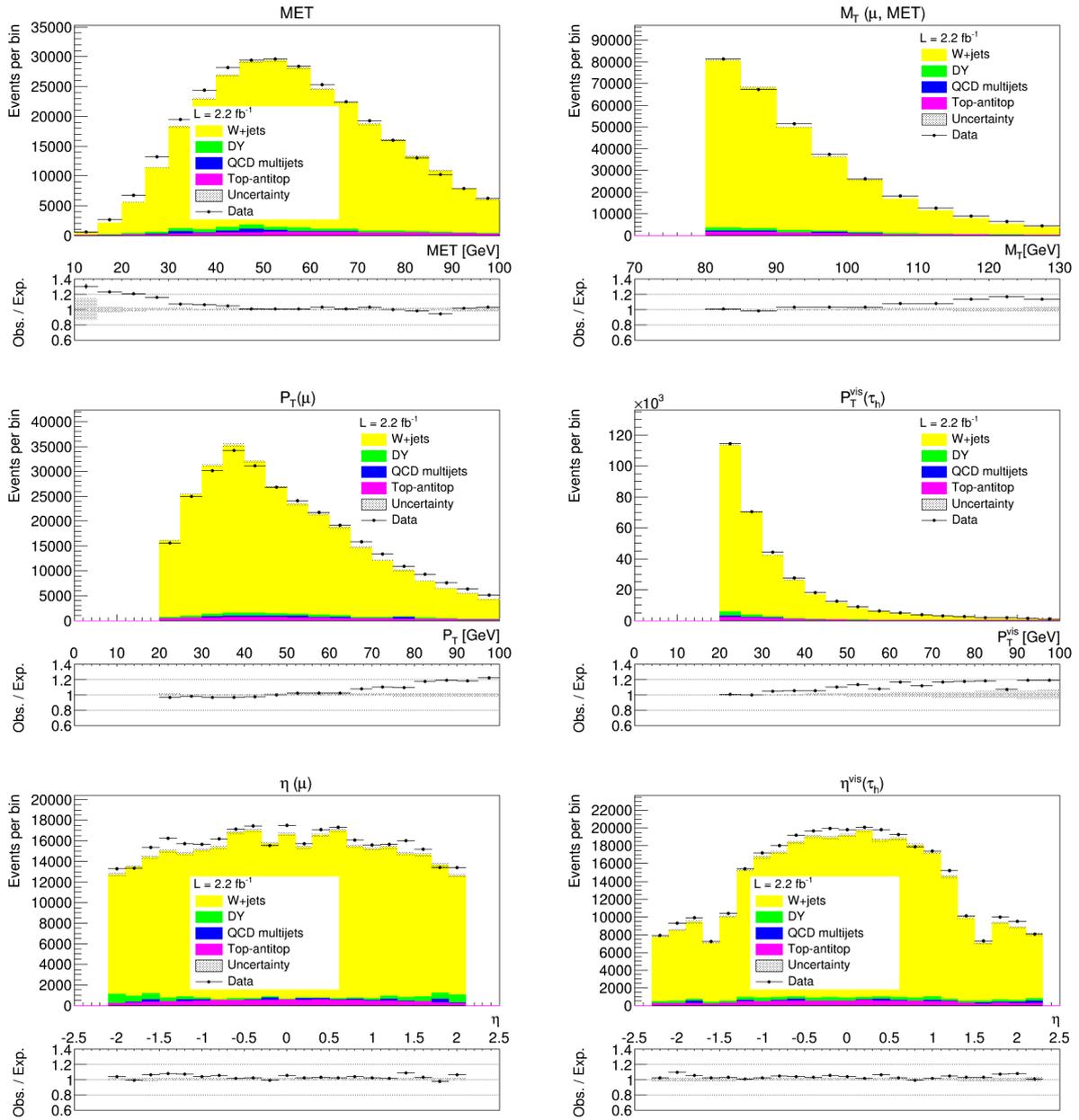


FIGURE 2.3 – Région des données avec $M_T > 80 \text{ GeV}$, sans coupure sur l'isolation du candidat- τ_h , et après correction de la normalisation du processus W+jets par le facteur d'échelle. De gauche à droite et de haut en bas, les histogrammes suivants sont présentés : celui de l'énergie transverse manquante (MET), celui de la masse transverse du système muon - énergie manquante (M_T), celui de la quantité de mouvement transverse (p_T) du muon, celui du p_T visible du tau hadronique, celui de la pseudo-rapidité du muon (η) et celui du η visible du tau hadronique. En-dessous de chaque histogramme, un graphique donne le rapport données sur simulations - le nombre d'événements observés sur le nombre d'événements attendus, *Observed (Obs.) over Expected (Exp.)*. Les incertitudes sur les simulations sont représentées par une zone hachurée grise. Le nombre d'événements W+jets représente environ 95% du total.

Il n'est cependant pas nécessaire de décrire précisément les formes des distributions, car elles ne jouent pas de rôle dans l'ajustement que nous réaliserons dans la section 2.2.6. En effet, nous cherchons à calculer un facteur r global pour le rapport de l'efficacité dans les données sur l'efficacité dans les simulations. Ce rapport ne dépend pas de la forme des différents processus, il ne dépend que de leur normalisation.

2.2.3 Estimation du processus multijets à partir des données

La simulation du processus multijets est peu fiable car elle implique des calculs compliqués de chromodynamique quantique (QCD). Les erreurs sur la normalisation de ce processus peuvent atteindre 50%, et il est préférable d'estimer directement ce processus à partir des données. Pour cela, il faut choisir une région des données enrichie en événements multijets et pauvre en autres processus. La région choisie est celle de *signes identiques* (SI) pour les charges du candidat- τ_h et du muon, contrairement au critère de *signes opposés* (SO) utilisé lorsque l'on cherche à sélectionner des événements $Z \rightarrow \tau_\mu \tau_h$. Toutes les autres sélections sont identiques à celles présentées dans la section 2.2.1.

Cette région des données est pauvre en événements issus de processus où les deux particules de l'état final ont des charges opposées, car le taux de mauvaise identification de la charge d'une particule est petit à des énergies en-dessous de 100 GeV (de l'ordre de 0.01-0.1% pour les muons [62]). Il y a donc peu d'événements Drell-Yan et top-antitop dans cette région des données. Il y a par contre beaucoup d'événements multijets et W+jets, car les charges du muon et du candidat- τ_h ne sont presque pas corrélées dans ces deux processus. Ces deux processus représentent ensemble plus de 95% des événements dans cette région des données, comme le montre la figure 2.4, qui contient la distribution de quelques variables dans la région de données SI.

Dans la région de signes identiques, pour chaque variable, la distribution des événements multijets ρ_{multi}^{SI} s'obtient en soustrayant à la distribution des données ρ_{data} les distributions des autres processus (Drell-Yan, $t\bar{t}$ et W+jets corrigé par le facteur d'échelle). La distribution multijets ainsi obtenue est ensuite corrigée par un facteur d'échelle SO/SI correspondant au rapport entre le nombre d'événements multijets de signes opposés (SO) et celui d'événements multijets de signes identiques (SI). Une distribution du processus multijets dans la région SO se calcule donc ainsi :

$$\rho_{multi}^{SO} = \text{SO/SI} \cdot \rho_{multi}^{SI} = \text{SO/SI} \cdot (\rho_{data} - \rho_{DY} - \rho_{t\bar{t}} - \rho_{Wjets}^{corr}) \quad (2.14)$$

où ρ_{DY} , $\rho_{t\bar{t}}$ et ρ_{Wjets}^{corr} sont respectivement les distributions dans la région SI de tous les événements Drell-Yan ($\tau_\mu \tau_h$ et $\mu\mu$), des événements top-antitop et des événements W+jets corrigés par leur facteur d'échelle.

Le rapport SO/SI peut-être calculé en sélectionnant une région des données riche en événements multijets et en comparant le nombre d'événements SO à celui de SI. Cette région quasiment pure en événements multijets s'obtient en inversant l'isolation du muon (on demande à ce qu'il ne soit pas isolé) et en rendant moins stricte la coupure sur l'isolation du candidat- τ_h . Ce travail a été fait par la référence [63], qui a obtenu le rapport SO/SI

suivant :

$$SO/SI = 1.06 \pm 0.05 \quad (2.15)$$

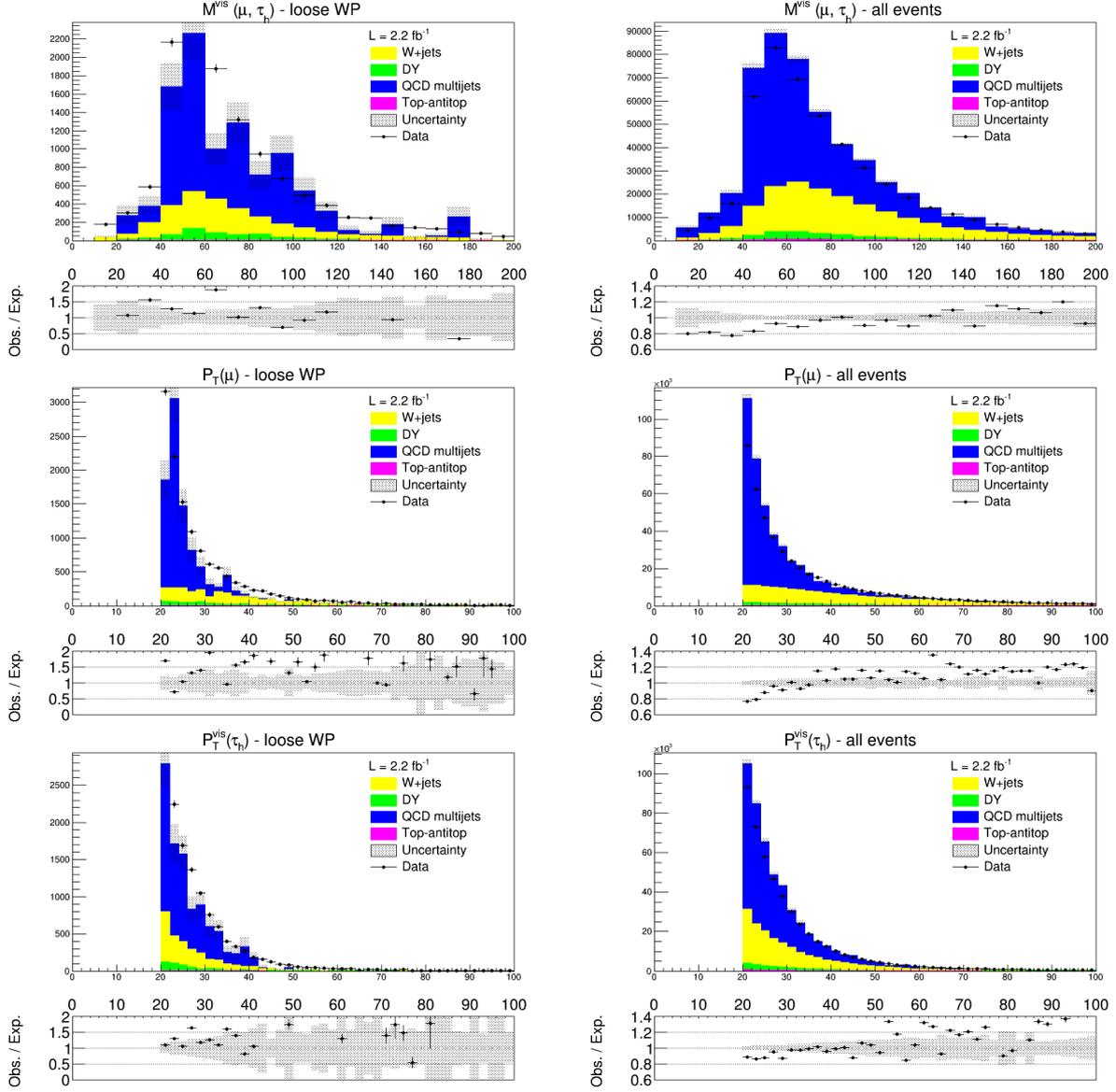


FIGURE 2.4 – Quelques distributions dans la régions de charges identiques pour le muon et le tau hadronique. De haut en bas, les distributions de masse visible M^{vis} du système $\tau_\mu\tau_h$, du p_T du muon et du p_T visible (p_T^{vis}) du tau hadronique sont montrées. À gauche, la sélection *loose* du discriminateur anti-jets de l’algorithme HPS est appliquée, à droite le discriminateur anti-jets n’est pas appliqué. Les données sont sous-estimées par les simulations dans les histogrammes de gauche, alors qu’elles sont surestimées dans les histogrammes de droite. Les incertitudes sur les simulations sont représentées par une zone hachurée grise.

La forme et la normalisation des distributions multijets sont ainsi toutes les deux déterminées à partir des données seulement, sans utiliser les simulations multijets. Des incertitudes sur la normalisation de ces distributions extraites des données proviennent du rapport SO/SI mais aussi de l’incertitude sur la normalisation du processus W+jets (l’incertitude sur le facteur d’échelle calculé ci-dessus). Une incertitude de 5% est retenue pour la normalisation du

processus multijets.

2.2.4 Histogrammes de contrôle dans le canal $\tau_\mu\tau_h$

La normalisation du processus W+jets a été corrigée par un facteur déduit de la région des données avec $M_T > 80$ GeV et sans coupure sur l'isolation du candidat- τ_h . Quant au processus multijets, sa forme et sa normalisation ont été extraites des données. La figure 2.5 permet de comparer la distribution de masse visible dans le canal $\tau_\mu\tau_h$ dans deux scénarios :

- en utilisant seulement les simulations, sans aucune correction.
- avec le processus W+jets corrigé par son facteur d'échelle, et le processus multijets obtenu à partir des données.

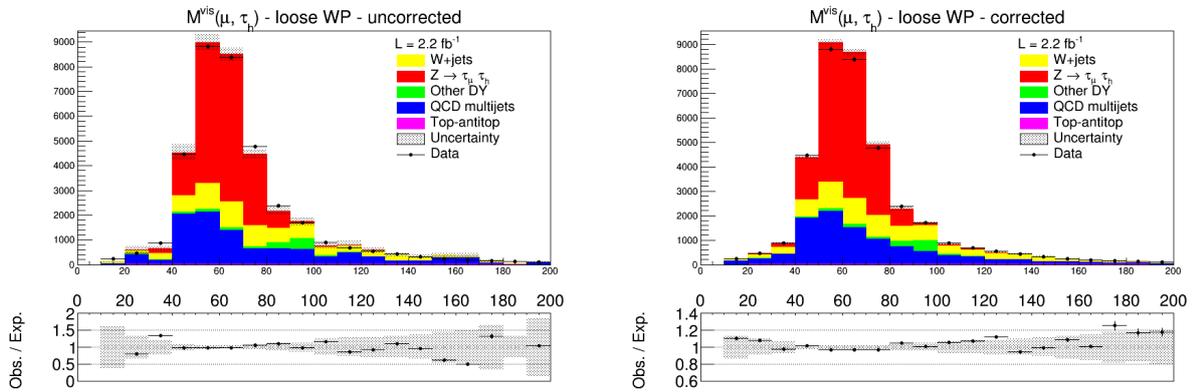


FIGURE 2.5 – Histogrammes de la masse visible de la paire $\tau_\mu\tau_h$, avec les sélections décrites en 2.2.1. Le WP loose est utilisé pour le discriminateur anti-jets. À gauche, les bruits de fond sont issus des simulations, sans correction. À droite, le processus W+jets est renormalisé par son facteur d'échelle et le processus multijets est issu des données. En-dessous de chacun des histogrammes principaux se trouve un histogramme donnant le rapport entre données et simulations. L'accord est meilleur dans l'histogramme de droite. C'est l'histogramme de droite qui est utilisé dans l'ajustement décrit dans la section 2.2.6 ci-dessous.

Il est ainsi possible de s'assurer que l'accord entre données et simulations est meilleur dans le deuxième scénario. Dans l'histogramme de gauche - correspondant au cas sans correction - le rapport données sur simulations s'écarte de plus de 35% de 1 à quatre reprises, alors que dans l'histogramme de droite, une telle variation ne se produit pas.

2.2.5 Sélections dans la région $\mu\mu$

Pour calculer l'efficacité d'identification et de reconstruction du tau hadronique par la méthode du rapport, il est nécessaire de sélectionner des événements dans le canal $\mu\mu$. Ces sélections doivent être aussi proches que possibles que dans la région $\tau_\mu\tau_h$ afin que les paramètres de nuisance soient corrélés dans les deux régions. Les sélections sont les suivantes :

- Le même système de déclenchement que dans la région $\tau_\mu\tau_h$ est utilisé.

- Les deux muons doivent avoir $p_T > 20$ GeV, une isolation relative combinée $I < 0.1$ et une identification muon *medium*. Ces sélections sont les mêmes que pour le muon dans la région $\tau_\mu\tau_h$.
- Le muon de plus grand p_T doit avoir $|\eta| < 2.1$ et l'autre $|\eta| < 2.4$. Ces coupures imitent les coupures de $|\eta| < 2.1$ pour le muon et $|\eta| < 2.3$ pour le tau hadronique dans la région $\tau_\mu\tau_h$ ⁴.
- Autres coupures :
 - Signes opposés, veto anti-électron et veto anti-quarks b, exactement comme dans la région $\tau_\mu\tau_h$.
 - Veto anti-troisième muon, pareil au veto anti-second muon utilisé dans la région $\tau_\mu\tau_h$.
 - $\Delta R(\mu, \mu) > 0.5$. Cette sélection est similaire à la coupure $\Delta R(\mu, \tau_h) > 0.5$ dans la région $\tau_\mu\tau_h$.
 - $M_{\mu\mu} > 50$ GeV, car la simulation du processus DY ne génère que des événements ayant une masse invariante supérieure à 50 GeV.

Les coupures qui ne peuvent s'appliquer qu'au tau hadronique (discriminateurs anti-électrons, anti-muons, et anti-jets) n'ont pas été utilisées. Aucune sélection n'est faite sur M_T non plus. Ci-dessous, dans la figure 2.6, des histogrammes de contrôle sont présentés pour cette région des données. Bien que la normalisation de tous les histogrammes soit en accord entre les données et les simulations, leur forme ne l'est pas toujours. Les histogrammes de la masse invariante et du p_T des muons ont des différences importantes de forme entre les données et les simulations. Cela est dû à certaines erreurs systématiques qui n'ont pas été corrigées en ce qui concerne les muons. Le rapport entre données et simulations de l'efficacité de certaines sélections dépendent du p_T du muon :

- Le système de déclenchement est moins efficace pour des p_T proches du seuil de déclenchement.
- L'isolation combinée relative du muon dépend de son p_T .
- L'efficacité de l'identification du muon par l'algorithme Particle Flow change avec son p_T .

Ces dépendances ne sont pas toujours bien décrites par les simulations. Pour l'ajustement, seule la normalisation de la région $\mu\mu$ sera prise en compte.

2.2.6 Ajustement du rapport entre efficacité dans les données et efficacité dans les simulations

L'ajustement simultané des régions $\tau_\mu\tau_h$ et $\mu\mu$ permet d'obtenir le rapport r de l'efficacité dans les données sur l'efficacité dans les simulation. Cet ajustement a été fait grâce à l'outil statistique *Combine* de CMS [64]. Il faut y introduire un paramètre d'intérêt, dont on cherche

4. La sélection $|\eta| < 2.4$ pour le deuxième muon est une erreur, elle aurait dû être $|\eta| < 2.3$ comme pour le τ_h . Cette erreur n'a été constatée que après la rédaction de ce mémoire, et il était trop tard pour refaire l'ajustement du rapport r avec la bonne sélection. Cependant, un calcul rapide a permis de s'assurer que l'effet de cette erreur sur le résultat final est bien inférieur à 1%.

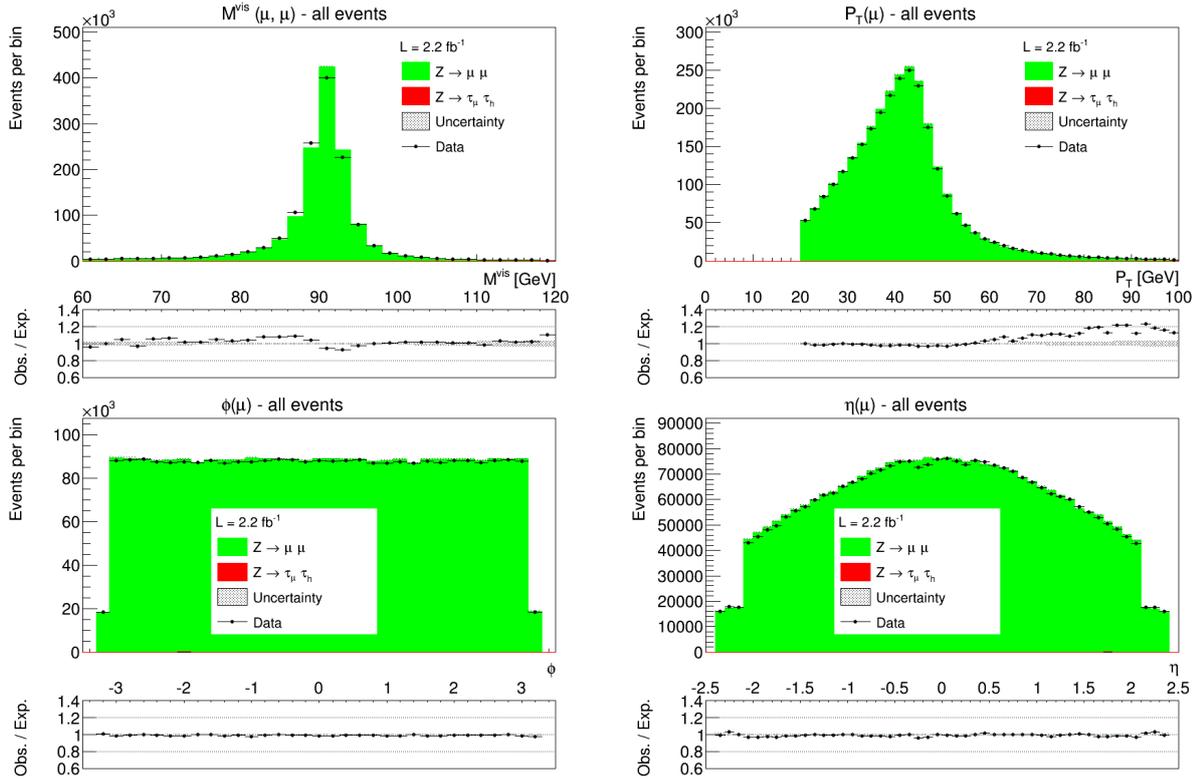


FIGURE 2.6 – Histogrammes de contrôle dans la région $\mu\mu$. De gauche à droite et de haut en bas, les distributions des variables suivantes sont présentées : la masse invariante des deux muons, la quantité de mouvement transverse p_T des muons, l'angle polaire ϕ des muons et leur pseudo-rapidité η .

la valeur, et des paramètres de nuisance, qui interviennent dans le calcul du paramètre d'intérêt mais dont les valeurs ne nous intéressent pas. Le paramètre d'intérêt peut varier librement, alors que les paramètres de nuisance suivent une distribution de probabilité dont la forme est prédéfinie lors de l'ajustement, c'est-à-dire, il existe des valeurs plus probables que d'autres pour les paramètres de nuisance. L'ajustement détermine la configuration des paramètres de nuisance et d'intérêt qui maximise la fonction de vraisemblance \mathcal{L} suivante :

$$\mathcal{L}(r, \mu) = p(\text{data}|r, \mu) \prod_{i=1}^n p(\bar{\mu}_i|\mu_i) \quad (2.16)$$

où r est le paramètre d'intérêt et $\mu = \{\mu_i\}_{i=1}^n$ représente l'ensemble des n paramètres de nuisance. La fonction $p(\text{data}|r, \mu)$ donne la probabilité d'obtenir les données étant donné les paramètres r et μ . La distribution des données est connue, elle peut s'observer dans la figure 2.5 (à droite) pour la région $\tau_\mu\tau_h$ et dans la figure 2.6 (en haut à gauche) pour la région $\mu\mu$. La fonction $p(\text{data}|r, \mu)$ détermine la vraisemblance des données pour chaque valeur des paramètres r et μ . Plus ceux-ci sont compatibles avec les données, plus la vraisemblance est grande. La fonction $p(\bar{\mu}_i|\mu_i)$ détermine quant à elle la vraisemblance de la valeur centrale $\bar{\mu}_i$ pour chaque valeur du paramètre de nuisance μ_i . La valeur centrale $\bar{\mu}_i$ est connue, μ_i peut varier mais plus sa valeur est éloignée de $\bar{\mu}_i$ et plus la probabilité $p(\bar{\mu}_i|\mu_i)$ est petite.

Maximiser la vraisemblance consiste à trouver les valeurs de r et μ telles que la fonction $\mathcal{L}(r, \mu)$ est maximale. L'erreur sur r s'obtient en déterminant les deux valeurs du paramètre

telles que la fonction de log-vraisemblance est de $1/2$ au-dessous de son maximum [58] :

$$\ln \mathcal{L}(r^{\pm 1\sigma}, \mu(r^{\pm 1\sigma})) = \ln \mathcal{L}(r^{obs}, \mu^{obs}) - 1/2 \quad (2.17)$$

où r^{obs} et μ^{obs} désignent les valeurs des paramètres qui maximisent la vraisemblance, $r^{+1\sigma}$ représente l'erreur supérieure sur r^{obs} et $r^{-1\sigma}$ l'erreur inférieure. La valeur de $\mu(r^{\pm 1\sigma})$ dépend de $r^{\pm 1\sigma}$, c'est celle qui maximise la vraisemblance sous la contrainte que $r = r^{\pm 1\sigma}$.

Le paramètre d'intérêt r n'intervient que dans la région $\tau_\mu\tau_h$. Chacun des paramètres de nuisance peut être présent dans l'une, l'autre ou les deux régions. Les paramètres de nuisance qui affectent seulement la normalisation des distributions et pas leur forme sont modélisés par des distributions de probabilité log-normales, ceux qui affectent la forme sont distribués selon des lois gaussiennes [58]. Voici la liste complète des paramètres de nuisance utilisés :

- *Bin by bin*. Ce paramètre de nuisance décrit les fluctuations statistiques du contenu de chaque intervalle (*bin*) des deux histogrammes. Le contenu de chacun des *bins* peut varier selon une loi de Poisson dont la moyenne est donnée par le contenu original (avant l'ajustement) du *bin*.
- Normalisation des événements Drell-Yan. Ce paramètre concerne les deux processus $Z \rightarrow \tau_\mu\tau_h$ et $Z \rightarrow \mu\mu$, il représente l'incertitude sur la section efficace des processus Drell-Yan. Un écart-type de 3% est utilisé pour la fonction de distribution de probabilité.
- Normalisation des événements multijets. Ce paramètre correspond à l'incertitude sur le facteur d'échelle SO/SI utilisé pour normaliser ce processus (cf. l'équation 2.15). L'écart-type utilisé est de 5%.
- Normalisation du processus W+jets. Ce paramètre correspond à l'incertitude sur le facteur d'échelle f calculé pour ce processus (cf. l'équation 2.13). L'écart-type utilisé est de 10%.
- Normalisation du processus $t\bar{t}$. Aucune correction issue des données n'a été appliquée à ce processus, sa normalisation est donc particulièrement incertaine. Un écart-type de 15% est utilisé, provenant de l'incertitude sur le calcul théorique de la section efficace.
- Luminosité. La luminosité intégrée des données n'est connue qu'avec une certaine précision. Une variation de ce paramètre affecte la normalisation de tous les processus dont la normalisation n'a pas été corrigée par les données. Les processus W+jets et multijets sont les seuls à ne pas être concernés. La valeur centrale de la luminosité est 2.2 fb^{-1} et son écart-type 2.6%.
- Efficacité de l'identification des muons. Un écart-type de 3% est pris pour la région $\tau_\mu\tau_h$ et un de 5% est utilisé pour la région $\mu\mu$.
- Échelle d'énergie des taus hadroniques. Ce paramètre est introduit de la manière suivante : l'énergie du tau hadronique est variée de $\pm 3\%$ autour de sa valeur mesurée⁵. Ce paramètre affecte tout aussi bien la forme que la normalisation des histogrammes. La normalisation change car dans le cas où le p_T du tau hadronique est diminué de 3%, certains événements qui étaient acceptés sont rejetés : il s'agit des événements qui passaient les sélections car ils contenaient un tau hadronique qui avait un p_T légèrement au-dessus du seuil de 20 GeV. Inversement, dans le cas où le p_T du tau hadronique est

⁵. Il faut corriger en conséquence la valeur de l'énergie transverse manquante, dont le calcul dépend de toutes les quantités de mouvement mesurées.

augmenté de 3%, certains événements qui ne passaient pas les sélections sont acceptés. La forme de la distribution de masse visible change aussi car pour chaque événement, la valeur de la masse est différente lorsque le p_T du tau hadronique est changé. Lorsque celui-ci est augmenté de 3%, la distribution de masse visible est déplacée vers des plus hautes valeurs de masse, et lorsque le p_T est diminué de 3%, la distribution est déplacée vers des plus petites valeurs. Des distributions différentes sont donc obtenues pour la valeur centrale et pour les variations de $\pm 3\%$, ces dernières représentent des variations de ± 1 écart-type autour de la valeur centrale. Ces trois histogrammes sont fournis au logiciel *Combine*, qui en déduit une distribution de probabilité pour l'échelle d'énergie des taus hadroniques.

- Échelle d'énergie des muons. Tout comme pour la variable précédente, ce paramètre affecte tout aussi bien la forme que la normalisation des distributions. Il est introduit de manière similaire à ce qui a été décrit au point précédent. Des variations de $\pm 1\%$ sur l'énergie du muon sont utilisées.

La table 2.3 résume les informations données ci-dessus sur les paramètres de nuisance.

Paramètres de nuisance	
Nom abrégé du paramètre	Erreur
Bin by bin	dépend du bin
Normalisation DY	3 %
Normalisation multijets	5 %
Normalisation W+jets	10 %
Normalisation $t\bar{t}$	15 %
Normalisation luminosité	2.6 %
Identification des muons	3 % ($\tau_\mu\tau_h$) ou 5 % ($\mu\mu$)
Énergie des taus	3 %
Énergie des muons	1 %

TABLE 2.3 – Paramètres de nuisance utilisés pour l'ajustement. L'erreur est exprimée en pourcentage de la valeur moyenne de chaque paramètre. Pour les paramètres bin by bin, l'erreur est égale à la racine carrée du nombre d'événements contenus dans le bin.

La figure 2.7 montre les distributions de masse visible dans les deux régions des données avant et après l'ajustement. Cette figure sert de contrôle pour vérifier que l'accord entre données et simulations est effectivement meilleur après l'ajustement.

Les valeurs du paramètre d'intérêt r (le rapport de l'efficacité dans les données sur l'efficacité dans les simulations) obtenues après l'ajustement sont données dans la table 2.4 pour chacune des trois sélections standards (WP). Ces valeurs sont comparées à celles obtenues par le groupe de A. K. Kalsi et al., dont les résultats sont présentés dans la référence [61]. Ce groupe a calculé de manière indépendante le rapport entre efficacité dans les données et dans les simulations par une méthode légèrement différente que celle utilisée dans ce mémoire. Il y a un bon accord entre leurs valeurs et celles calculées pour ce mémoire : pour chaque WP la différence entre les deux est inférieure à un cinquième de l'erreur sur le paramètre d'intérêt. La petite différence qui subsiste peut s'expliquer par des différences dans les paramètres de

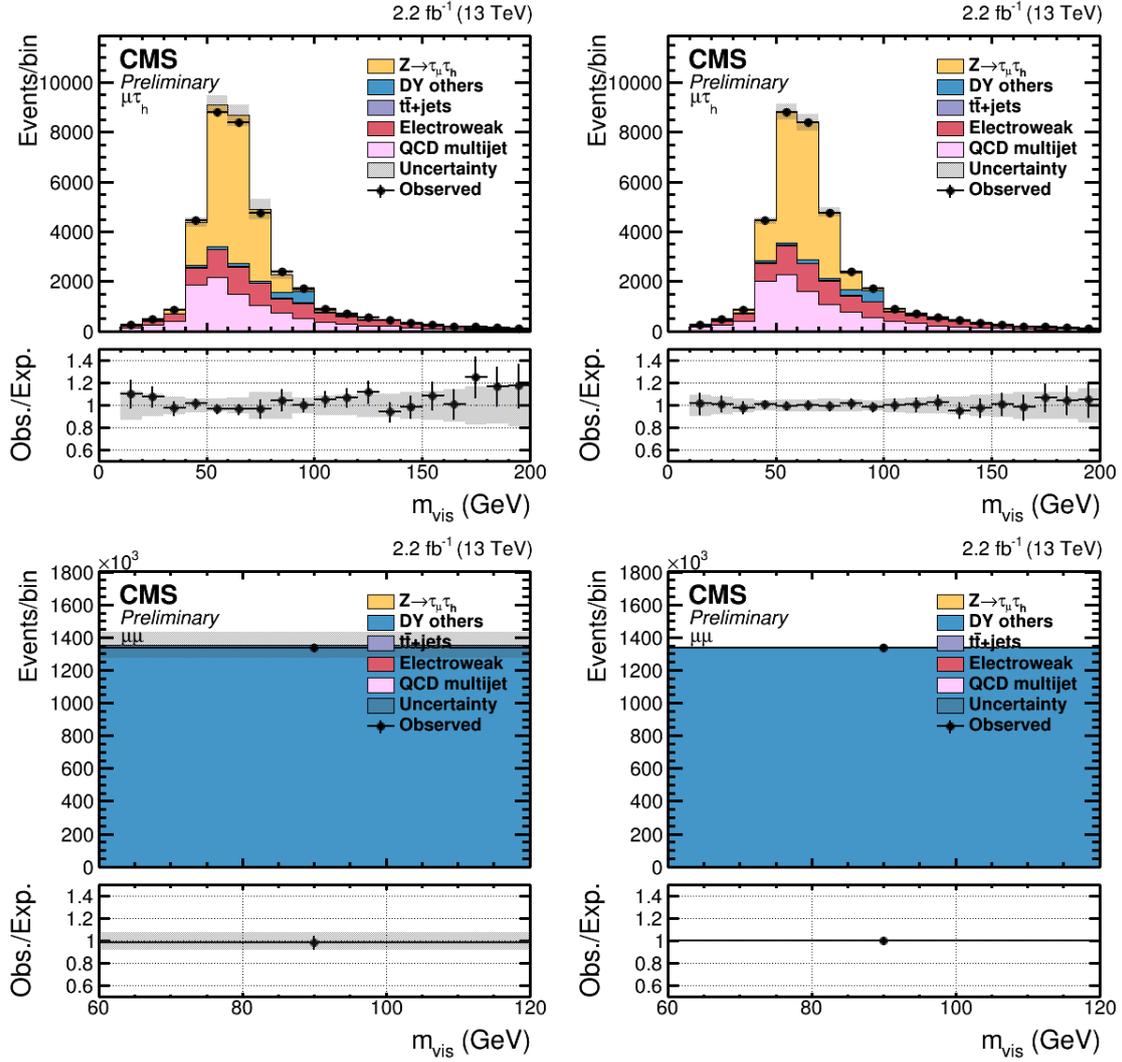


FIGURE 2.7 – Histogrammes de la masse visible, à gauche avant l’ajustement et à droite après. En haut, les histogrammes de la région $\tau_\mu\tau_h$ des données avec le WP loose pour le discriminateur anti-jets, en bas ceux de la région $\mu\mu$. Pour la région $\mu\mu$, un seul bin est utilisé.

nuisance et les simulations utilisés, ainsi que dans la façon d’estimer la normalisation des processus W +jets et multijets.

	Paramètre d’intérêt		
	Loose WP	Medium WP	Tight WP
Valeurs calculées pour ce mémoire	0.929 ± 0.055	0.924 ± 0.058	0.911 ± 0.053
Valeurs obtenues par la réf. [61]	0.933 ± 0.051	0.933 ± 0.056	0.912 ± 0.052

TABLE 2.4 – Valeurs du rapport de l’efficacité dans les données sur l’efficacité dans les simulations pour chacune des trois sélections standards.

Si les simulations décrivaient parfaitement les données, l’efficacité dans les données serait

la même que dans les simulations et leur rapport r vaudrait 1. Cependant l'efficacité pourrait être différente dans les simulations et les données, par exemple s'il y a des particules non-simulées qui contribuent à l'isolation dans les données. Les valeurs obtenues pour r dans la table 2.4 sont de l'ordre de 0.92, avec une incertitude d'environ 5%. Il est possible de quantifier l'incompatibilité des valeurs de r obtenues avec une vraie valeur hypothétique $r = 1$ en calculant la probabilité p qu'une fluctuation statistique produise une valeur plus éloignée de 1 que la valeur mesurée r^{obs} . Une loi gaussienne est utilisée pour calculer cette probabilité :

$$p = \int_{-\infty}^{1-\epsilon} dr N_{(1,\sigma)}(r) + \int_{1+\epsilon}^{+\infty} dr N_{(1,\sigma)}(r), \quad \epsilon = |1 - r^{obs}| \quad (2.18)$$

où $N_{(1,\sigma)}(r)$ est la distribution de probabilité gaussienne de valeur moyenne 1 et d'écart-type σ , évaluée en r , et σ est l'erreur sur le paramètre d'intérêt mesuré. Ce calcul représente une intégration sur les deux queues de distribution de la gaussienne. Les valeurs obtenues pour p sont de 19.7% pour la sélection *loose*, 19.0% pour la sélection *medium* et 9.3% pour la sélection *tight*.

Chapitre 3

Spectre à haute masse invariante du système $\tau_\mu\tau_h$

Ce chapitre est consacré à la recherche d'un boson Z' se désintégrant en une paire $\tau^+\tau^-$ dans l'état final particulier $\tau_\mu\tau_h$. Cette analyse se fait dans le spectre à haute masse invariante, avec les données collectées en 2015 par CMS. Le Modèle Standard séquentiel (MSS) est utilisé comme signal dans cette analyse. Il s'agit d'un modèle où le boson Z' aurait exactement les mêmes couplages que le boson Z du Modèle Standard, la seule différence entre les deux bosons étant leur masse (cf. la section 1.1). Les simulations de signal présentées dans la section 1.4.2 sont utilisées.

La première partie de ce chapitre rappelle les sélections utilisées pour les événements $\tau_\mu\tau_h$. La deuxième partie concerne le choix de la meilleure variable de masse reconstruite à analyser. La troisième explique l'analyse statistique réalisée sur la distribution de cette variable, et ses conséquences pour les modèles contenant un boson Z' . Le calcul de l'efficacité de l'identification du τ_h réalisé dans le chapitre précédent joue un rôle important dans les limites posées sur la section efficace de production du Z' .

3.1 Sélection des événements $\tau_\mu\tau_h$

Dans ce chapitre, les mêmes sélections que celles de la section 2.2 sont appliquées, avec la précision que la sélection standard *loose* est choisie pour le discriminateur anti-jets de l'identification des taus hadroniques. Ces sélections sont résumées dans la table 3.1. Elles sont appliquées aux données et aux simulations.

3.2 Choix de la variable de masse reconstruite

Parmi les produits de désintégration d'un lepton tau, il y a toujours un ou deux neutrinos. Les neutrinos étant inobservables par CMS, il est impossible d'avoir accès à la masse invariante totale du système $\tau^+\tau^-$. Seuls les quantités de mouvement des particules visibles et le vecteur d'énergie transverse manquante sont disponibles. Plusieurs variables de masse peuvent être construites à partir de ces quantités, quatre sont considérées dans le cadre de ce mémoire : la masse visible, la masse totale, la masse transverse totale et la masse colinéaire.

Système de déclenchement	Muon isolé $p_T > 18 \text{ GeV}, \eta < 2.1$
Coupures sur le muon	$p_T > 20 \text{ GeV}$ $ \eta < 2.1$ Identification standard de CMS Isolation relative < 0.1
Coupures sur le tau hadronique	$p_T > 20 \text{ GeV}$ $ \eta < 2.3$ Discriminateur anti-muons <i>tight</i> Discriminateur anti-électrons <i>loose</i> Discriminateur anti-jets <i>loose</i>
Autres coupures	Charges opposées pour le muon et le tau hadronique $\Delta R(\mu, \tau_h) > 0.5$ Veto anti-électron Veto anti-deuxième muon Veto anti-jets de quarks b Masse transverse $M_T(\mu, \text{MET}) < 40 \text{ GeV}$

TABLE 3.1 – Sélections utilisées pour la recherche du boson Z' .

Ces variables sont appelées collectivement *variables de masse reconstruite*. Leurs définitions sont données ci-dessous.

3.2.1 Présentation des variables

Cette section présente chacune des variables de masse reconstruite considérées pour la recherche du boson Z' dans le canal $\tau_\mu\tau_h$.

Masse visible

L'option la plus simple pour le calcul de la masse reconstruite est d'ignorer l'énergie manquante et de ne considérer que les quantités de mouvement des particules visibles. La masse visible est définie comme la masse invariante du système $\mu\text{-}\tau_h$. Cette variable a déjà été utilisée dans le chapitre précédent.

Masse totale

Il est possible de reconstruire différemment la masse invariante de la paire $\tau^+\tau^-$ en utilisant le vecteur d'énergie transverse manquante (MET). La masse totale est définie comme la masse invariante du système $\mu\text{-}\tau_h\text{-MET}$. Contrairement à la masse visible, la masse totale tient compte d'une partie de l'énergie invisible (la partie transverse) portée par les neutrinos issus de la désintégration des leptons taus.

Masse transverse totale

La masse transverse totale (M_T^{tot}) est une variante sur la masse totale. Seules les composantes transverses des impulsions entrent dans le calcul de la masse transverse totale.

Cette variable se calcule de la même façon que la masse transverse du muon et de l'énergie manquante définie dans l'équation 2.9 :

$$M_T^{tot} = \sqrt{2 \cdot p_T^{vis} \cdot E_T^{miss} \cdot (1 - \cos \Delta\phi)} \quad (3.1)$$

où p_T^{vis} est la quantité de mouvement transverse du système visible $\mu\text{-}\tau_h$, E_T^{miss} est l'énergie transverse manquante et $\Delta\phi$ est la différence d'angle azimutal entre le vecteur d'énergie manquante et la quantité de mouvement visible.

Masse colinéaire

La masse colinéaire M_{col} est une tentative de reconstruire la masse totale du système $\tau^+\tau^-$ en faisant l'approximation que les neutrinos sont colinéaires avec les produits visibles de désintégration des leptons taus. L'énergie manquante est alors divisée en deux composantes, l'une colinéaire au muon et l'autre colinéaire au tau hadronique. Les directions de ces deux composantes sont connues, reste à déterminer leurs normes. Cela peut se faire en imposant que toute l'énergie transverse manquante est due à ces composantes colinéaires. Deux équations sont ainsi obtenues dans le plan transverse [65] :

$$x_{MET} = p_1^{inv} \sin \theta_\mu \cos \phi_\mu + p_2^{inv} \sin \theta_{\tau_h} \cos \phi_{\tau_h} \quad (3.2)$$

$$y_{MET} = p_1^{inv} \sin \theta_\mu \sin \phi_\mu + p_2^{inv} \sin \theta_{\tau_h} \sin \phi_{\tau_h} \quad (3.3)$$

où les variables connues sont x_{MET} et y_{MET} (les composantes x et y de l'énergie transverse manquante), θ_μ et ϕ_μ (les angles polaire et azimutal de la quantité de mouvement du muon), θ_{τ_h} et ϕ_{τ_h} (mêmes angles pour le tau hadronique). Les inconnues sont les normes des deux composantes d'énergie invisible : p_1^{inv} pour la composante colinéaire au muon et p_2^{inv} pour la composante colinéaire au tau hadronique. Lorsque p_1^{inv} et p_2^{inv} ont été calculées, les quantités de mouvement du muon et du tau hadronique sont multipliées par des facteurs d'échelle qui tiennent compte de l'énergie invisible dans cette approximation colinéaire (pour le muon, ce facteur est $(p_\mu + p_1^{inv})/p_\mu$, et pour le τ_h il est $(p_{\tau_h} + p_2^{inv})/p_{\tau_h}$). La masse colinéaire est utilisée dans certaines analyses où l'état final est une paire $\tau^+\tau^-$ (comme par exemple dans la référence [66]). L'utilisation de cette variable demande cependant des sélections supplémentaires, notamment la réjection d'événements où le muon et le tau hadroniques sont dos-à-dos. En effet, les équations 3.2 et 3.3 sont dégénérées si $\phi_\mu = \phi_{\tau_h} + \pi$, et les solutions p_1^{inv} et p_2^{inv} divergent lorsque $\phi_\mu \rightarrow \phi_{\tau_h} + \pi$ [65]. Ces sélections supplémentaires ne sont pas souhaitables car elles diminueraient le nombre d'événements disponibles pour l'analyse. La masse colinéaire n'est donc pas utilisée dans ce mémoire.

3.2.2 Comparaison entre les variables

Cette section est consacrée à la comparaison des variables de masse reconstruite présentées ci-dessus (sauf la masse colinéaire, dont l'utilisation est déjà exclue) et à la détermination de la variable la plus adaptée pour la recherche du boson Z' . Une première comparaison est donnée par la figure 3.1 qui contient les distributions de la masse visible M^{vis} , de la masse totale M^{tot} et de la masse transverse totale M_T^{tot} pour quatre hypothèses de masse différentes du boson Z' du Modèle Standard Séquentiel (MSS). Les sélections résumées dans la table 3.1

ont été appliquées. Nous pouvons constater qu'aucune de ces trois variables ne présente un maximum à la masse générée du boson Z' , à cause de la présence de neutrinos non-détectés. Les distributions de la masse totale et de la masse transverse totale sont similaires et ont un maximum à une valeur plus proche de la masse générée que la distribution de masse visible.

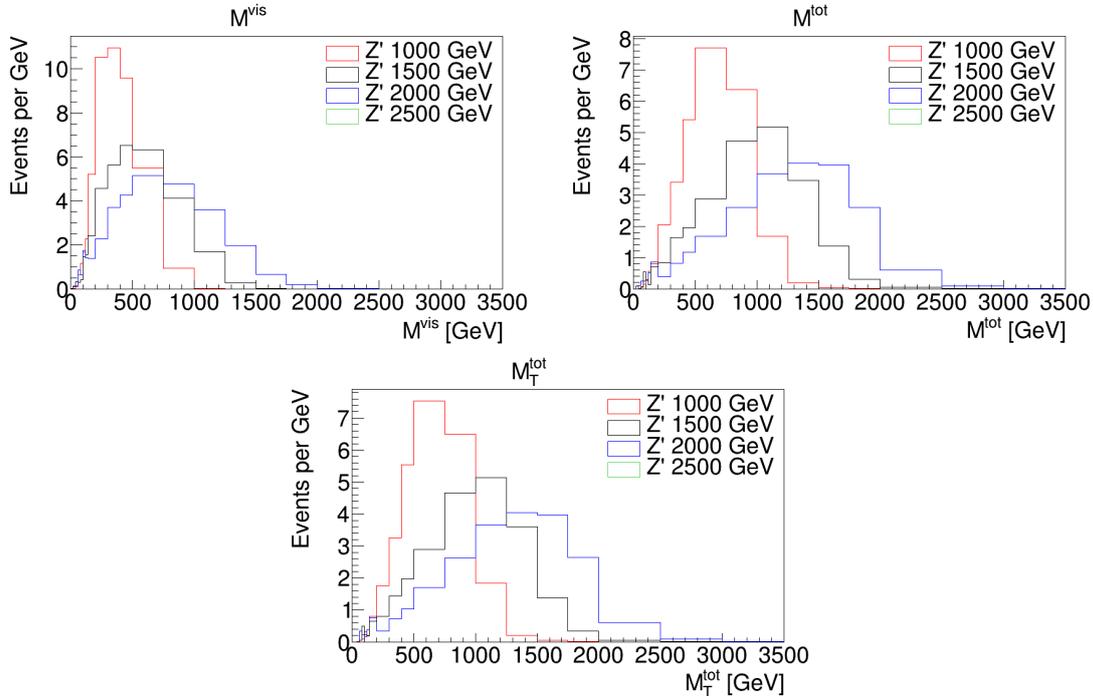


FIGURE 3.1 – Distributions de masse reconstruite pour quatre hypothèses de masse différentes du Z' du MSS : 1000 GeV, 1500 GeV, 2000 GeV et 2500 GeV. Ces distributions n'ont pas été normalisées par leur section efficace. En haut à gauche, la masse visible, à droite la masse totale et en bas la masse transverse totale.

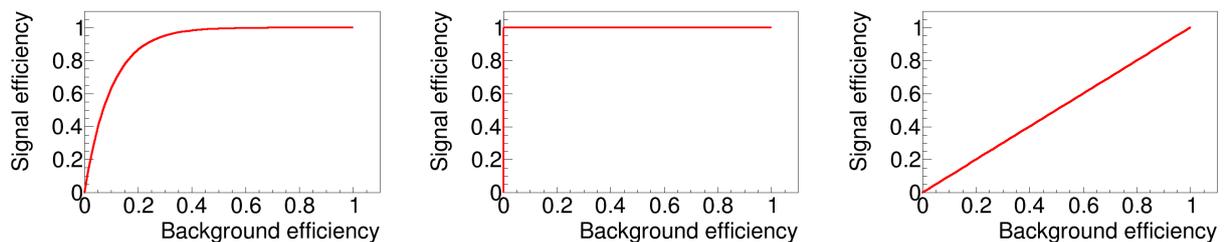


FIGURE 3.2 – Exemples de courbes ROC. À gauche, une courbe typique : des coupures plus strictes permettent de diminuer l'efficacité du bruit de fond sans compromettre l'efficacité du signal, jusqu'à un certain point. Au milieu, la courbe ROC pour une variable permettant de discriminer parfaitement le signal du bruit de fond : il existe une coupure qui accepte 100% du signal tout en acceptant 0% du bruit de fond. À droite, la courbe ROC pour une variable ne présentant aucune différence entre le bruit de fond et le signal : l'efficacité pour le bruit de fond est parfaitement corrélée à l'efficacité pour le signal.

Une comparaison plus sophistiquée entre ces variables est possible, en étudiant les pa-

ramètres suivants : la différence relative entre la masse reconstruite et la masse générée, la courbe ROC, et la signification du signal.

- La **différence relative** se calcule à partir des simulations, en comparant la masse reconstruite M_{reco} à la masse générée M_{gen} . On considère la distribution de la variable suivante :

$$u = \frac{M_{reco} - M_{gen}}{M_{gen}} \quad (3.4)$$

Si M_{reco} est un estimateur gaussien non-biaisé de M_{gen} , alors u devrait suivre une loi normale de moyenne nulle. La valeur moyenne détermine si M_{reco} est un estimateur biaisé : si par exemple la valeur moyenne de u est négative, cela voudrait dire que la masse reconstruite est systématiquement inférieure à la masse générée. L'écart-type de la distribution détermine quant à lui la résolution de M_{reco} en tant qu'estimateur de M_{gen} . Même si la largeur naturelle de la résonance Z' , donnée par une distribution de Breit-Wigner, est étroite, une mauvaise résolution (de l'ordre de 50% par exemple) rend la résonance plus large au niveau reconstruit. Un pic large complique l'analyse car pour avoir une bonne efficacité de signal, il faut une coupure moins stricte sur la masse, ce qui entraîne une efficacité plus élevée pour le bruit de fond. Le rapport signal sur bruit de fond devient plus petit, ce qui n'est pas souhaitable. On cherche donc la variable ayant le moindre biais et la meilleure résolution, c'est-à-dire la moyenne la plus proche de 0 et l'écart-type le plus petit pour la distribution de u .

- Les **courbes ROC** donnent l'efficacité du signal en fonction de l'efficacité du bruit de fond, pour des coupures différentes sur la masse reconstruite, la figure 3.2 donne quelques exemples. Une courbe ROC se construit en appliquant des coupures de plus en plus strictes sur la masse reconstruite, en mesurant les efficacités du signal et du bruit de fond, et plaçant les points correspondant sur le graphique. Le premier point, correspondant à la coupure la moins stricte possible (c'est-à-dire, aucune coupure), est toujours placé au coin supérieur droit : l'efficacité est de 100% tout aussi bien pour le signal que pour le bruit de fond. Ensuite, avec des coupures de plus en plus strictes, les deux efficacités diminuent progressivement, jusqu'à ce que la courbe atteigne le coin inférieur gauche, correspondant à une coupure tellement stricte que l'efficacité est de 0% pour le signal et pour le bruit de fond. L'objectif des coupures est de minimiser l'efficacité du bruit de fond sans compromettre l'efficacité du signal. La variable idéale pour discriminer le signal du bruit de fond serait une variable où une coupure suffisamment stricte permet de réduire l'efficacité du bruit de fond à des valeurs proches de 0% tout en gardant une efficacité de 100% pour le signal. Une mauvaise variable ne permettrait pas de discriminer le bruit de fonds du signal : toute perte d'efficacité du bruit de fond est compensée par une perte égale en efficacité du signal. Ces exemples extrêmes sont illustrés dans la figure 3.2.

- La **significance** du signal caractérise l'importance du nombre d'événements de signal attendus par rapport aux fluctuations statistiques du signal et du bruit de fond. En modélisant ces fluctuations par une loi de Poisson, la signification s se calcule ainsi :

$$s = \frac{N_{sig}}{\sqrt{N_{sig} + N_{bkg}}} \quad (3.5)$$

où N_{sig} est le nombre d'événements de signal attendus et N_{bkg} le nombre d'événements de bruit de fonds attendus. La signification peut être calculée pour plusieurs coupures

différentes sur la masse reconstruite, la coupure maximisant la signification est considérée la meilleure. Une signification petite peut être due à un taux de bruit de fond trop élevé par rapport au signal attendu, dans ce cas les fluctuations du bruit de fond ne vont pas être distinguables d'éventuels événements de signal. Mais même en l'absence de bruit de fond, la signification peut être petite si le nombre d'événements de signal attendus est petit.

Dans la suite de cette section, le boson Z' de 1500 GeV du MSS est utilisé comme référence. La figure 3.3 donne les distributions de différence relative entre masse reconstruite et masse générée pour le Z' de 1500 GeV, et la table 3.2 reprend les valeurs moyennes et écart-types de ces distributions. Cette figure permet de confirmer l'impression donnée par la figure 3.1 que la masse visible a une distribution plus compacte, mais aussi plus éloignée de la masse générée du Z' . La masse visible est en moyenne plus de 50% inférieure à la masse générée, contre environ 30% seulement pour la masse totale et la masse transverse totale.

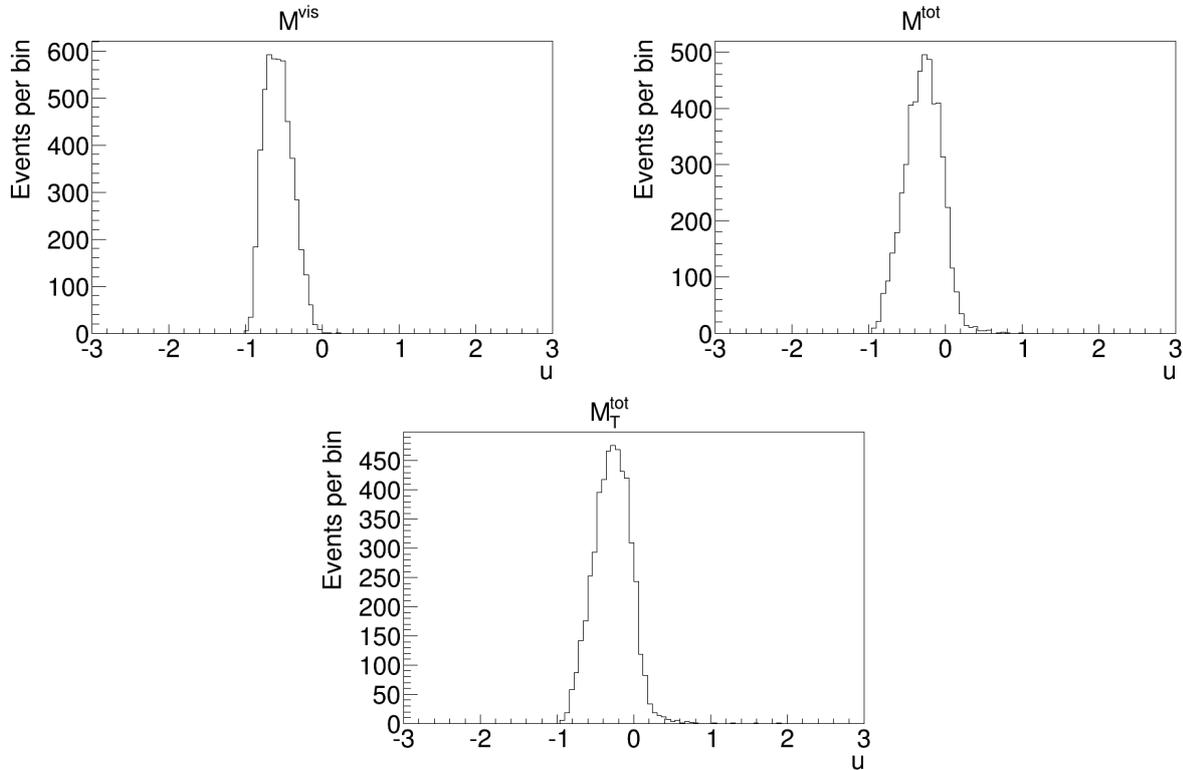


FIGURE 3.3 – Distributions de différence relative u entre masse reconstruite et masse générée pour un boson Z' de 1500 GeV du MSS. En haut à gauche, celle de la masse visible, à droite, celle de masse totale, et en bas, celle de la masse transverse totale.

Une masse reconstruite bien inférieure à la masse générée mènera à une signification plus faible car le nombre d'événements de bruit de fond est plus élevé pour des valeurs plus petite de masse. Cela est illustré dans la figure 3.4, qui montre la distribution de signal superposée à celle du bruit de fond pour les trois variables de masse reconstruite. Les processus ont été normalisés à la luminosité des données, 2.2 fb^{-1} , mais la section efficace du signal a été multipliée par 10. Dans le cas de la masse visible, le signal et le bruit de fonds se chevauchent beaucoup plus que pour les deux autres variables. Cela peut se constater plus précisément

Variable	Valeur moyenne de u	Écart-type
M^{vis}	-0.57	0.18
M^{tot}	-0.29	0.24
M_T^{tot}	-0.28	0.24

TABLE 3.2 – Valeurs moyennes de u et écart-types pour les trois variables de masse reconstruite étudiées. Ces valeurs ont été obtenues pour un boson Z' de 1500 GeV du MSS.

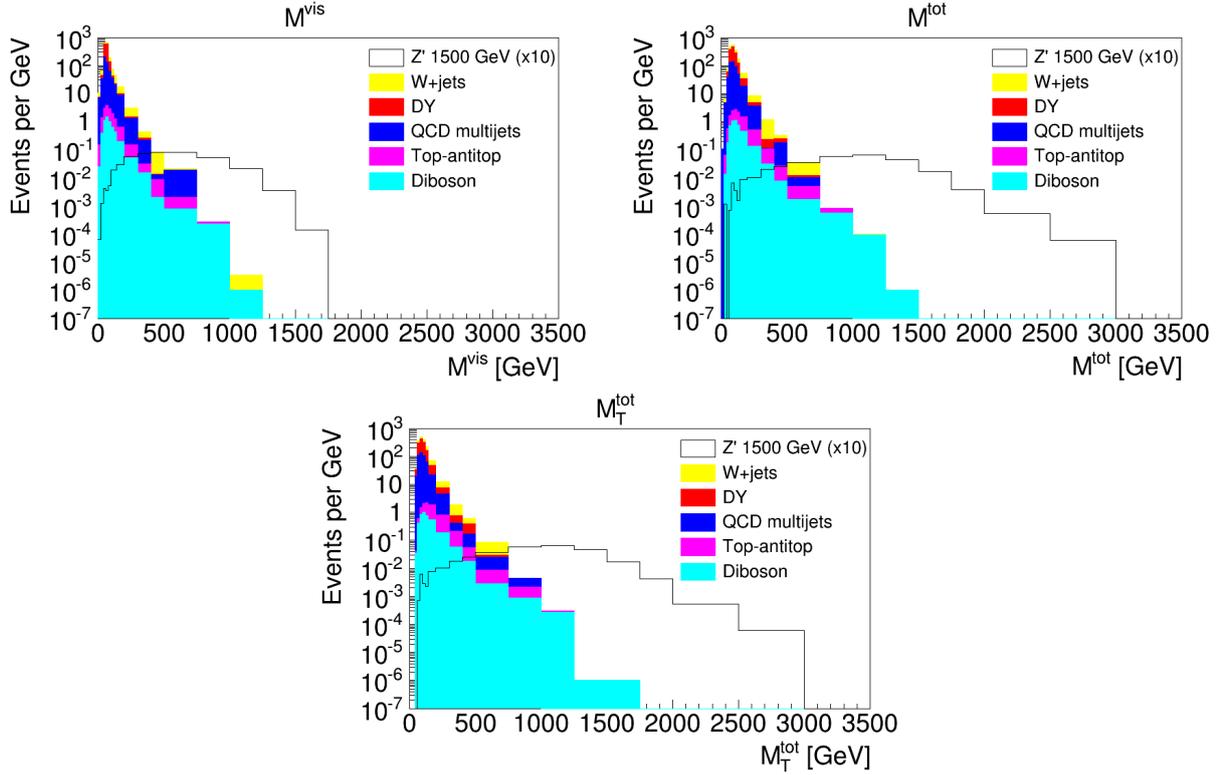


FIGURE 3.4 – Distributions de masse reconstruite pour le bruit de fond et le signal (Z' de 1500 GeV du MSS). La normalisation du signal a été multipliée par 10 afin de faciliter la lecture du graphique. En haut à gauche, la masse visible ; à droite, la masse totale ; en bas, la masse transverse totale.

en considérant les courbes ROC et de signficance du signal données par la figure 3.5. Les courbes ROC sont toutes similaires et ont un bon aspect car la plupart du bruit de fond se trouve à des petites valeurs de la masse reconstruite. Une coupure $M_{reco} > 200$ GeV élimine déjà la quasi-totalité du bruit de fond sans perte d'efficacité notable pour le signal. La signficance présente cependant des différences pour les trois variables : elle atteint un maximum d'environ 1.35 pour la masse visible, 2.1 pour la masse totale et 2.0 pour la masse transverse totale. La masse totale et la masse transverse totale permettent donc d'atteindre les meilleures significances (c'est-à-dire le meilleur rapport entre nombre d'événements de signal et fluctuations statistiques du signal et du bruit de fond) par rapport à la masse visible. Cette caractéristique se maintient pour les autres masses du boson Z' (1000 GeV, 2000 GeV, 2500 GeV), avec un avantage de la masse totale sur la masse transverse totale pour le Z' de 1000 GeV (les significances sont respectivement de 4.8 et 4.1), cf. la figure 3.6.

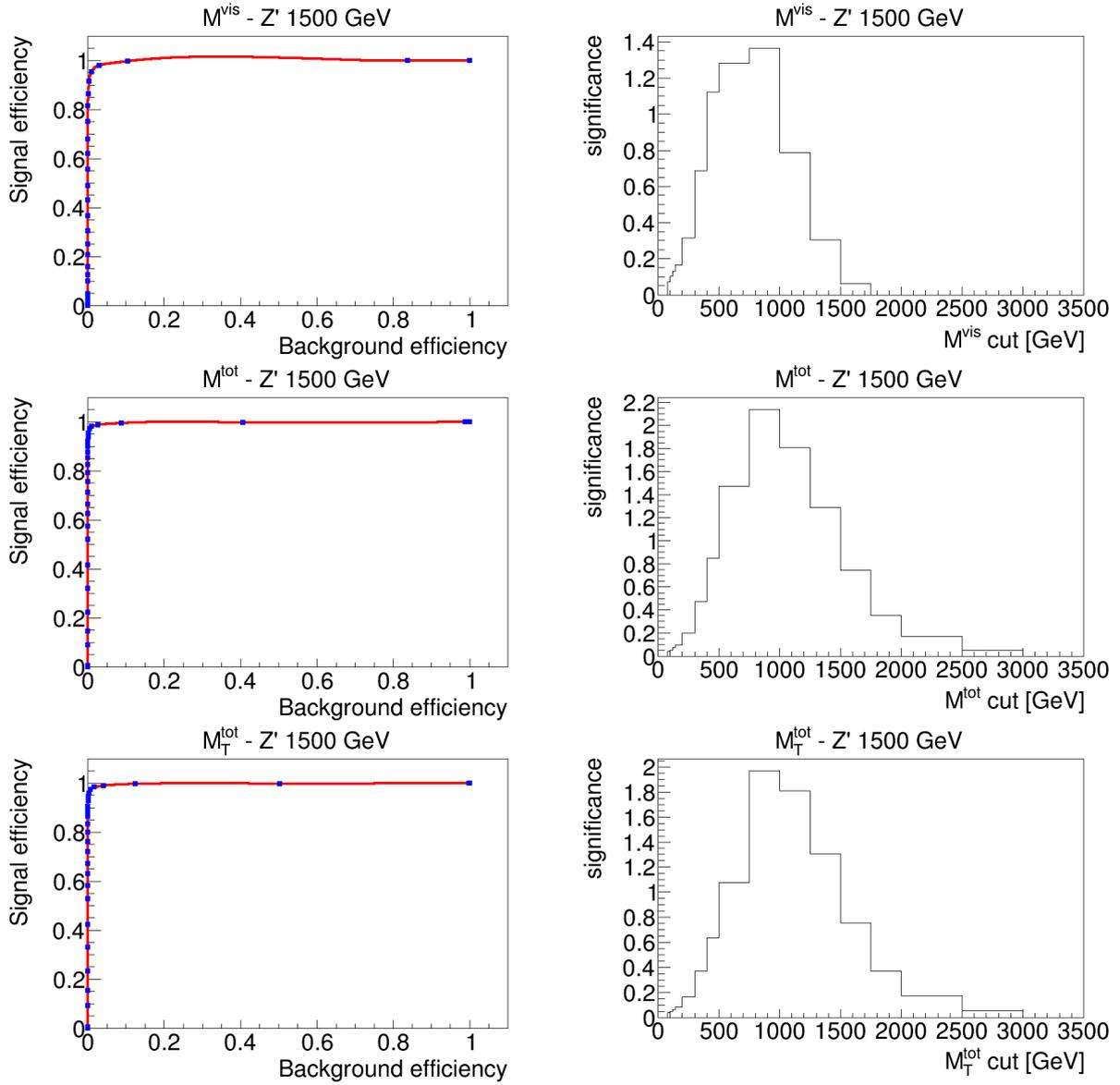


FIGURE 3.5 – Du côté gauche, les courbes ROC, représentant l’efficacité du signal (Z' de 1500 GeV du MSS) en fonction de l’efficacité du bruit de fond, pour différentes coupures sur la masse reconstruite. Les coupures sont de la forme $M_{reco} > x$. x commence à 0 au coin supérieur droit et augmente de 50 en 50 GeV jusqu’à atteindre 1000 GeV. À partir de là les augmentations se font de 100 en 100 GeV. Du côté droit, les distributions de la signficance du signal en fonction de la coupure utilisée sur la masse reconstruite. La coupure correspond à la valeur inférieure de chaque bin. De haut en bas, les graphiques concernent la masse visible, la masse totale et la masse transverse totale.

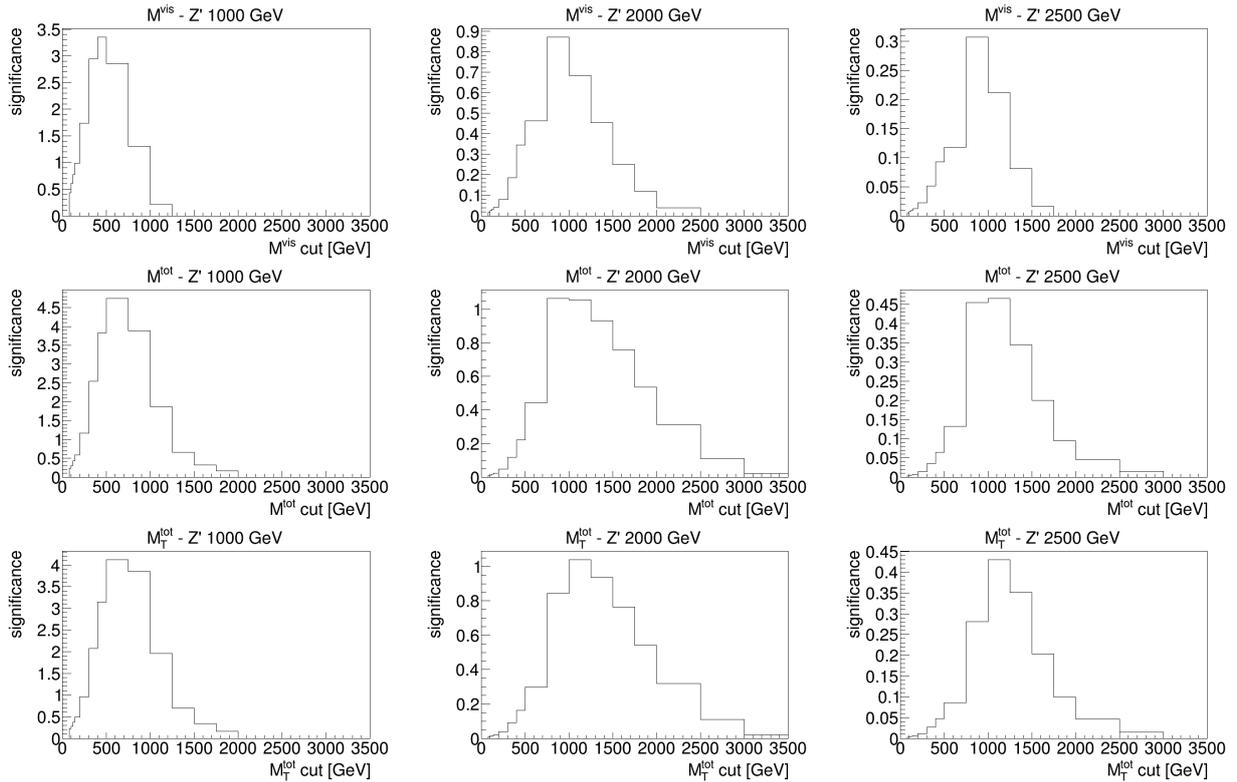


FIGURE 3.6 – Significances pour les bosons Z' du MSS en fonction de la coupure sur la masse reconstruite. De gauche à droite des bosons Z' avec une masse générée de 1000 GeV, 2000 GeV et 2500 GeV. De haut en bas pour la masse visible, la masse totale et la masse transverse totale.

La masse totale est donc la variable de masse reconstruite la plus adaptée pour la recherche d'un boson Z' du Modèle Standard Séquentiel dans le spectre à haute masse du canal $\tau_\mu\tau_h$. Les coupures qui permettent d'atteindre la signification la plus élevée sont différentes pour les différentes masses possibles du Z' : 500 GeV pour le boson de 1000 GeV, 750 GeV pour celui de 1500 GeV, et 1000 GeV pour ceux de 2000 et 2500 GeV. Des significances plus élevées peuvent être atteintes pour les plus petites masses du boson Z' , qui ont une section efficace plus importante (cf. la table 1.2).

3.3 Analyse des données

Nous cherchons un boson Z' dans l'état final $\tau_\mu\tau_h$. Les sélections appliquées sont celles que nous avons présentées dans la table 3.1. Les données utilisées sont les mêmes que pour le chapitre précédent, ce sont les données des collisions proton-proton du LHC collectées en 2015 par CMS à une énergie de centre de masse $\sqrt{s} = 13$ TeV correspondant à une luminosité intégrée de 2.2 fb^{-1} . Les événements de bruit de fond sont tous estimés par les simulations décrites dans la section 1.4.2, sauf pour les processus W +jets et multijets, qui sont estimés de la même façon que dans le chapitre 2 : le processus W +jets est corrigé par le facteur d'échelle donné dans l'équation 2.13 ; quant au processus multijets, sa forme est extraite de la région de charges identiques pour le muon et le tau hadronique et sa normalisation est

	$M^{tot} > 200$ GeV	$M^{tot} > 400$ GeV
Données	1092	52
Bruit de fond	995 ± 39	43.2 ± 7.4
Signal (1500 GeV)	62.53 ± 9.0	59.4 ± 8.7

TABLE 3.3 – Nombre d'événements dans les données et nombre d'événements attendus pour le bruit de fond et le signal (un boson Z' de 1500 GeV du MSS), pour deux coupures différentes sur la masse totale M^{tot} . Les erreurs données ici pour le nombre d'événements de bruit de fond et de signal sont purement statistiques.

corrigée par le facteur SO/SI exprimé dans l'équation 2.15.

La figure 3.7 permet de comparer la distribution de masse totale dans les données à celle du bruit de fond attendu. La distribution pour un boson Z' du MSS de 1500 GeV est aussi représentée. Pour les grandes masses, on observe un léger excès de données par rapport au bruit de fond attendu. La table 3.3 permet de comparer plus précisément le nombre de données aux nombres d'événements de bruit de fond et de signal attendus. L'excès de données est supérieur de deux écarts-types du bruit de fond attendu pour la coupure $M^{tot} > 200$ GeV, mais l'erreur présentée dans le tableau est purement statistique et ne tient pas compte des effets systématiques. Pour la coupure $M^{tot} > 400$ GeV, l'excès de données est moindre et est davantage compatible avec l'incertitude sur le bruit de fond. À première vue, il semble

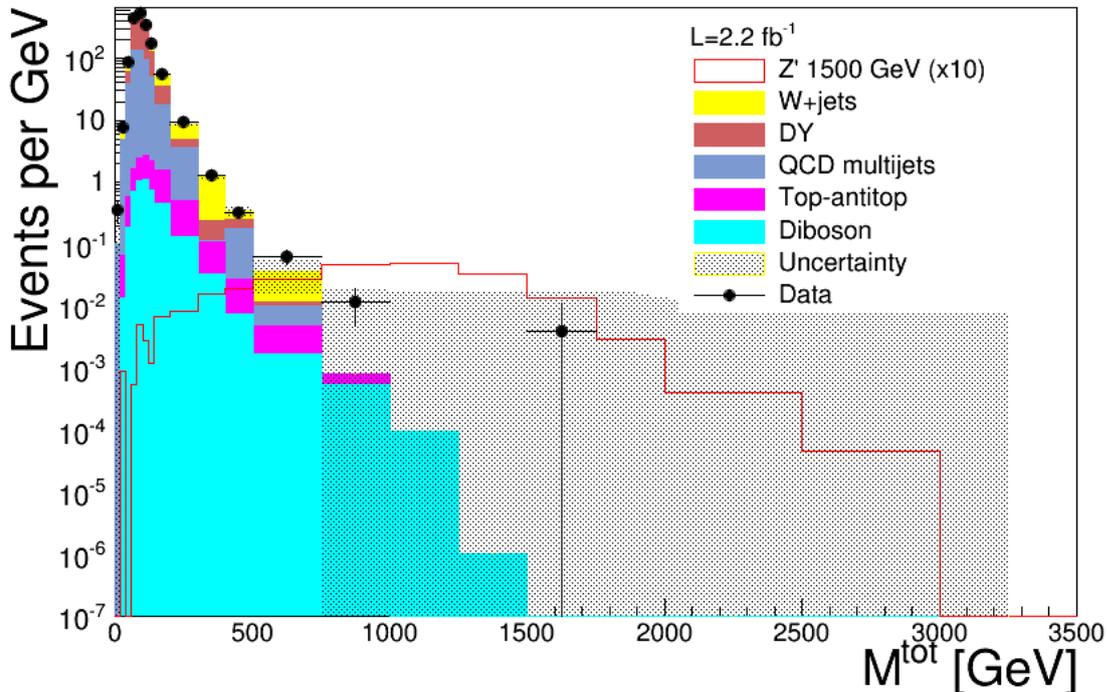


FIGURE 3.7 – Distribution de masse totale, comparaison entre données, bruit de fond et signal (Z' de 1500 GeV du MSS, normalisation multipliée par 10). L'incertitude sur le bruit de fond est illustrée par la bande grise.

donc ne pas y avoir de signe de nouvelle physique dans les données, le nombre d'événements observés étant compatible avec les prédictions du Modèle Standard.

Cette analyse qualitative doit être rendue plus précise par une analyse quantitative : quelles limites peuvent être posées sur les modèles de nouvelle physique prédisant l'existence d'un boson Z' ? Quels modèles sont exclus, et quels modèles sont compatibles avec ces données ? Clairement, tout modèle prédisant un nombre d'événements de signal beaucoup plus élevé que le nombre de données observées est exclu. Une analyse statistique des données est nécessaire afin d'obtenir une limite à 95% de niveau de confiance sur le nombre d'événements de signal compatible avec les données.

3.3.1 Méthode statistique utilisée

Nous cherchons une limite supérieure sur le nombre d'événements du processus $pp \rightarrow Z' + X \rightarrow \tau\tau + X$ (où X correspond aux particules provenant de l'hadronisation des résidus des protons) compatibles avec les données. Une méthode dite *cut-and-count* est utilisée pour poser cette limite. Le principe de la méthode est le suivant : une coupure sur la masse totale est choisie, le nombre de données et le nombre d'événements de bruit de fond supérieurs à cette coupure sont comptés. Ensuite, on détermine le nombre maximal d'événements de signal qui vérifie l'inégalité suivante :

$$\sum_{n=0}^{N_{data}} P(n | N_{bkg} + N_{sig}) \geq 5\% \quad (3.6)$$

où N_{data} est le nombre de données, N_{bkg} est le nombre d'événements de bruit de fond attendus et N_{sig} est le paramètre libre dont nous cherchons à déterminer la limite supérieure. La distribution de probabilité P suit une loi de Poisson de paramètre $N_{bkg} + N_{sig}$, $P(n | N_{bkg} + N_{sig})$ représente la probabilité d'observer n événements dans les données si le nombre d'événements attendus est $N_{bkg} + N_{sig}$. La somme $\sum_{n=0}^{N_{data}} P(n | N_{bkg} + N_{sig})$ représente donc la probabilité que le nombre total d'événements soit inférieur ou égal à celui qui a effectivement été observé dans les données, si le nombre d'événements attendus est $N_{bkg} + N_{sig}$. Le nombre maximal d'événements de signal qui satisfait à cette inégalité correspond à la limite à 95% de niveau de confiance sur le nombre d'événements de signal compatible avec les données. Si le nombre moyen d'événements de signal attendus est plus grand que cette limite, la probabilité que le nombre d'événements dans les données soit aussi petit que le nombre d'événements observés est inférieure à 5%.

En plus de calculer cette *limite observée* sur le nombre d'événements de signal, une *limite attendue* peut être déterminée. La limite attendue correspond à la limite sur le nombre d'événements de signal si le nombre de données observées était égal au bruit de fond attendu. Une distribution de limites attendues peut s'obtenir en réalisant un grand nombre de pseudo-expériences. Pour chacune de ces pseudo-expériences, un nombre aléatoire est généré selon une loi de Poisson dont la valeur moyenne est le nombre d'événements de bruit de fond attendus N_{bkg} . Ce nombre aléatoire représente le nombre de "données" observées lors de cette pseudo-expérience. Ensuite ce nombre fictif de données observées est utilisé dans l'équation 3.6 pour poser une limite supérieure sur le nombre d'événements de signal (précisons que N_{bkg} reste inchangé et égal au nombre d'événements de bruit de fond attendus). Une distribution de limites supérieures attendues est ainsi obtenue, avec une certaine moyenne et

un certain écart-type. Cette distribution peut être comparée à la limite observée. Si aucun événement de signal n'a été produit, les deux limites devraient être compatibles. Dans ce mémoire, l'erreur utilisée pour la limite attendue est purement statistique, et correspond à l'écart-type de la distribution. L'utilisation d'erreurs systématiques demanderait un travail plus poussé.

La méthode décrite ci-dessus a été implémentée avec la coupure $M^{tot} > 400$ GeV, pour les raisons expliquées dans la section 3.2.2. Dans un premier temps, seules les limites sur le nombre d'événements de signal sont calculées, les limites sur la section efficace demandent la considération d'un modèle particulier. La table 3.4 donne un résumé des paramètres utilisés et des résultats obtenus. La limite observée est différente de la limite attendue d'un facteur d'environ 1.4 fois l'écart-type. Les deux limites sont donc compatibles.

Luminosité des données	2.2 fb ⁻¹
Bruit de fond attendu	43.2 ± 7.4
Limite attendue (95%) sur le nombre d'événements de signal	12.5 ± 7.2
Nombre de données	52
Limite observée (95%) sur le nombre d'événements de signal	22.3

TABLE 3.4 – Paramètres de la méthode *cut-and-count* pour $M^{tot} > 400$ GeV et limites (95%) attendue et observée sur le nombre d'événements de signal.

Des limites supérieures sur le nombre d'événement de signal ont été posées, nous souhaitons en déduire des limites sur sa section efficace. Le nombre d'événements d'un processus observé dans les données est proportionnel à la luminosité de l'échantillon et au produit de l'acceptance α du détecteur et de l'efficacité ϵ des sélections :

$$N_{obs} = N_{tot} \cdot \alpha \cdot \epsilon \quad (3.7)$$

où N_{obs} est le nombre d'événements du processus observé dans les données et N_{tot} est le nombre total d'événements de ce processus produits lors des collisions proton-proton. Or comme nous l'avons vu dans l'équation 1.10, le nombre total d'événements produits lors d'une prise de données est égal au produit de la luminosité L et de la section efficace σ du processus. L'équation suivante est donc vraie :

$$N_{obs} = L \cdot \sigma \cdot \alpha \cdot \epsilon \quad (3.8)$$

En isolant la section efficace, on obtient :

$$\sigma = \frac{N_{obs}}{L \cdot \alpha \cdot \epsilon} \quad (3.9)$$

Une limite maximale sur le nombre d'événements de signal peut donc se traduire en une limite maximale sur la section efficace du processus par la formule ci-dessus, si l'on connaît

l'acceptance et l'efficacité.

3.3.2 Acceptance et efficacité

Le produit de l'acceptance par l'efficacité $\alpha \cdot \epsilon$ s'obtient à partir d'une simulation du signal : le nombre d'événements reconstruits passant toutes les sélections est divisé par le nombre total d'événements générés. Remarquons que ce produit $\alpha \cdot \epsilon$ dépend du modèle de boson Z' considéré, et en particulier de la masse et des couplages du boson Z' . Les simulations utilisées contiennent uniquement des événements $Z' \rightarrow \tau\tau$, pour calculer l'acceptance fois l'efficacité il suffit donc de diviser le nombre d'événements passant les sélections données dans la table 3.1 et la coupure $M^{tot} > 400$ GeV par le nombre total d'événements générés.

Des facteurs d'échelle peuvent être appliqués afin de corriger le produit $\alpha \cdot \epsilon$ par des informations extraites des données. Dans notre cas, nous utilisons le facteur $r = 0.929 \pm 0.055$ obtenu dans le chapitre 2 pour la reconstruction du tau hadronique et la sélection *loose* du discriminateur anti-jets. La table 3.5 résume les résultats obtenus. Le produit de l'acceptance par l'efficacité est proche de 3% pour toutes les masses considérées pour le boson Z' du Modèle Standard Séquentiel (MSS).

Masse du Z'	Événements générés	Événements acceptés	r (loose WP)	Acceptance*Efficacité
1000 GeV	149543	4297	0.929	0.0267
1500 GeV	148116	4691		0.0294
2000 GeV	149769	4769		0.0296
2500 GeV	147912	4994		0.0314

TABLE 3.5 – Calcul du produit de l'acceptance par l'efficacité pour chacune des quatre masses considérées pour le boson Z' du MSS.

3.3.3 Limites sur la section efficace et conséquences pour les modèles prédisant un boson Z'

Nous cherchons une limite supérieure à la section efficace du signal, c'est-à-dire du processus $pp \rightarrow Z' + X \rightarrow \tau\tau + X$. La formule 3.9 permet de calculer cette limite à partir de celle sur le nombre d'événements de signal, étant donné l'acceptance du détecteur et l'efficacité des sélections.

La limite obtenue sur la section efficace peut être comparée à la limite attendue et à la section efficace prédite par le MSS. C'est ce qui est fait dans la table 3.6, pour des masses de 1000, 1500, 2000 et 2500 GeV du boson Z' . Pour les quatre points de masse étudiés, seul le boson Z' de 1000 GeV est exclu à 95% de confiance par l'analyse, car la section efficace prédite est supérieure à la limite observée posée par les données. Pour les autres hypothèses de masse, les sections efficaces prédites sont inférieures à la limite observée et ne peuvent donc pas être exclues.

Masse du Z'	Section efficace (MSS) [pb]	Limite attendue (95%) [pb]	Limite observée (95%) [pb]
1000 GeV	0.502 ± 0.011	0.21 ± 0.12	0.380
1500 GeV	0.08525 ± 0.0018	0.19 ± 0.11	0.345
2000 GeV	0.01990 ± 0.00041	0.19 ± 0.11	0.345
2500 GeV	$(6.13 \pm 0.13) \cdot 10^{-3}$	0.18 ± 0.10	0.323

TABLE 3.6 – Sections efficaces prédites par le Modèle Standard Séquentiel (MSS) et limites supérieures attendue et observée sur la section efficace, en fonction de la masse du boson Z' . Les sections efficaces ont été calculées à l'ordre dominant et multipliées par un facteur 1.3 pour représenter l'effet des ordres plus élevés (cf. la table 1.2)

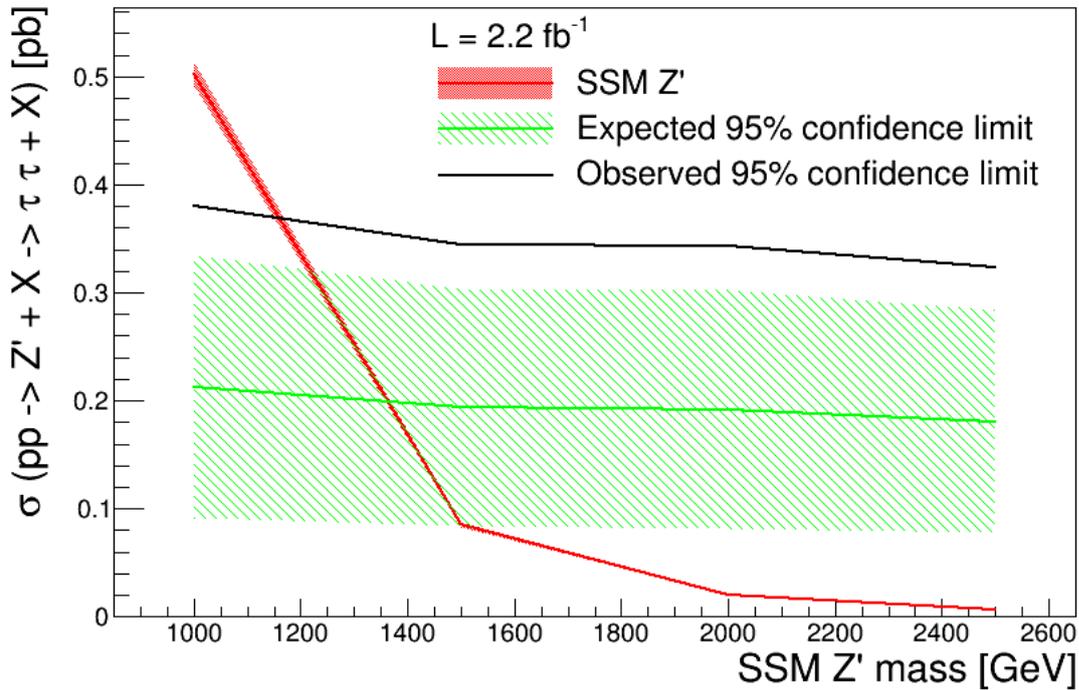


FIGURE 3.8 – Graphique d'exclusion pour la section efficace du boson Z' du MSS en fonction de sa masse. La section efficace prédite du boson Z' du MSS est donnée en rouge, la limite attendue à 95% de confiance est donnée en vert et la limite observée est donnée en noir. La limite inférieure attendue pour la masse du Z' est de 1.37 ± 0.13 TeV, la limite observée est de 1.18 TeV.

La figure 3.8 représente graphiquement les données de la table. La limite inférieure attendue sur la masse du boson Z' du MSS est de 1.37 ± 0.13 TeV, alors que la limite observée est de 1.18 TeV, si l'on suppose une extrapolation linéaire entre les points. Si l'on considère un modèle plus exotique, comme le modèle $U(1)_{B-3L_\tau}$ présenté dans le chapitre 1, des masses plus lourdes du Z' peuvent être exclues. Le rapport de branchement du Z' en paire $\tau^+\tau^-$ est d'environ 58% dans ce modèle contre 3.3% dans le MSS (même rapport de branchement que $Z \rightarrow \tau\tau$). Si l'on suppose que le boson Z' de ce modèle a le même couplage aux quarks que le Z' du MSS, les sections efficaces prédites pour le processus $pp \rightarrow Z' + X \rightarrow \tau\tau + X$ sont multipliées par 17.6. Un boson Z' de 2000 GeV de ce modèle est alors exclu.

Le boson Z' du MSS est cependant celui qui est utilisé comme référence dans les articles de recherche. Les limites actuelles de CMS sur la masse du Z' du MSS dans le canal $\tau\tau$, sont celles obtenues par les références [47] et [67] :

- La réf. [47] a analysé les données prises en 2011 par CMS à une énergie du centre de masse $\sqrt{s} = 7$ TeV, qui correspondent à une luminosité intégrée de 4.9 fb^{-1} . Quatre états finaux de la paire $\tau^+\tau^-$ ont été considérés, $\tau_e\tau_\mu$, $\tau_e\tau_h$, $\tau_\mu\tau_h$ et $\tau_h\tau_h$, qui représentent ensemble 93.8% du rapport de branchement de la paire $\tau^+\tau^-$. La limite attendue sur la masse générée du boson Z' est de 1.1 TeV, et la limite observée est de 1.35 TeV¹. La différence entre les deux est due à un léger déficit d'événements dans les données.
- La réf. [67] a analysé les données collectées en 2012 par CMS à $\sqrt{s} = 8$ TeV, qui correspondent à une luminosité de 19.7 fb^{-1} . Seul l'état final $\tau_e\tau_\mu$ a été considéré, il représente un rapport de branchement de 6.2% de la paire $\tau^+\tau^-$. Les limites attendue et observées sont de 1.3 TeV.

La limite attendue obtenue dans ce mémoire est meilleure que celles des analyses précédentes de CMS, et la limite observée est moins stricte. Le fait que la limite observée soit moins stricte est dû au léger excès de données par rapport au nombre d'événements de bruit de fond attendus : la réf. [47] avait observé un petit déficit de données et la réf. [67] avait observé un bon accord entre données et simulations. Les différences entre les résultats des trois analyses s'expliquent par les facteurs suivants :

- L'énergie du centre de masse \sqrt{s} affecte les sections efficaces de production du boson Z' . Plus \sqrt{s} est élevée, plus les sections efficaces sont élevées et plus le nombre d'événements de signal attendu sera élevé. L'analyse est alors plus sensible aux masses plus élevées du boson Z' , qui ont une section efficace plus faible. Les deux articles de CMS ont utilisé des données avec une valeur de \sqrt{s} plus petite que celle des données de ce mémoire.
- La luminosité intégrée L affecte le nombre d'événements observés pour tous les processus. Une luminosité plus grande augmente le nombre d'événements de signal attendu et donc la sensibilité de l'analyse à des masses plus élevées du boson Z' . Les échantillons de données utilisés dans les deux articles de CMS avaient des luminosités plus élevées que celle des données de ce mémoire.
- Le choix des états finaux considérés pour la paire $\tau^+\tau^-$ affecte le nombre d'événements $\tau^+\tau^-$ observés. Si une plus grande proportion du rapport de branchement total de la paire $\tau^+\tau^-$ est étudiée, la sensibilité de l'analyse augmente. La réf. [47] a inclus dans son analyse 4 états finaux, correspondant à un rapport de branchement combiné de 93.8%, la réf. [67] n'a considéré qu'un état final avec un rapport de branchement de 6.2%. L'état final considéré dans ce mémoire, $\tau_\mu\tau_h$, a un rapport de branchement intermédiaire, de 22.5%.

En conclusion, les résultats obtenus dans ce mémoire sont très compétitifs comparés à ceux de CMS obtenus avec les données du Run 1, ceci grâce à l'augmentation de l'énergie du centre de masse.

1. Il y a une erreur dans la réf. [47], le texte n'est pas cohérent avec la figure présentant le graphique d'exclusion. Le texte annonce une limite observée de 1.4 TeV mais la figure suggère une limite de 1.35 TeV.

Conclusion et perspectives

Deux analyses de données ont été réalisées dans ce mémoire : l'étude de l'efficacité de la reconstruction et de l'identification des taus hadroniques et la recherche d'une nouvelle résonance neutre massive (un boson Z') dans le canal $\tau^+\tau^-$, dans l'état final particulier $\tau_\mu\tau_h$. Les données utilisées ont été récoltées en 2015 par le détecteur CMS au LHC, elles correspondent à des collisions proton-proton avec une énergie du centre de masse $\sqrt{s} = 13$ TeV et une luminosité de 2.2 fb^{-1} .

La première analyse réalisée a été l'étude de l'efficacité de l'algorithme *Hadrons Plus Strips* (HPS) de CMS, qui reconstruit et identifie les leptons taus hadroniques à partir de leurs produits de désintégration. Cet algorithme est complexe et nécessite une étude approfondie de ses performances, en particulier pour les nouvelles données de haute énergie récoltées en 2015. L'étude de cet algorithme constitue le coeur du travail réalisé dans ce mémoire. L'efficacité de l'algorithme a d'abord été déterminée dans les simulations, ensuite les données ont été analysées afin de trouver le rapport r de l'efficacité de l'algorithme dans les données sur son efficacité dans les simulations. Ce paramètre r a été calculé par la méthode du rapport $Z \rightarrow \tau_\mu\tau_h / Z \rightarrow \mu\mu$ et pour cela il a fallu sélectionner deux régions des données, l'une riche en événements du processus $Z \rightarrow \tau_\mu\tau_h$ et l'autre riche en événements $Z \rightarrow \mu\mu$. La région $\mu\mu$ est composée quasi-exclusivement d'événements qui nous intéressent ($Z \rightarrow \mu\mu$) et n'a pas subi de traitement particulier. La région $\tau_\mu\tau_h$ est par contre contaminée par un nombre important d'événements de bruit de fond, de l'ordre de 50%. Deux processus de bruit de fond ont été estimés à partir des données : un facteur d'échelle a été calculé pour corriger la simulation du processus W +jets, quant au processus multijets, tout aussi bien sa normalisation que sa forme ont été extraites des données. Ensuite, un ajustement combiné des régions $\mu\mu$ et $\tau_\mu\tau_h$ a été réalisé avec l'outil statistique *Combine* de CMS, afin de déterminer la valeur du paramètre r pour chacune des sélections standards de l'algorithme HPS. Des valeurs $r = 0.929 \pm 0.055$, $r = 0.924 \pm 0.058$ et $r = 0.911 \pm 0.053$ ont été trouvées respectivement pour les sélections *loose*, *medium* et *tight*. Ces valeurs sont compatibles avec celles calculées indépendamment par une équipe faisant partie du groupe d'étude des leptons taus de CMS. Ces résultats sont utilisés dans toutes les analyses de CMS qui étudient des états finaux avec un ou plusieurs leptons taus.

Les résultats de cette première analyse ont été utilisés dans la deuxième, qui concerne la recherche d'un boson Z' dans le spectre de haute masse de la paire $\tau^+\tau^-$, dans l'état final particulier $\tau_\mu\tau_h$. Le boson Z' du Modèle Standard Séquentiel (MSS) a été choisi comme signal, avec quatre hypothèses de masse : 1000, 1500, 2000 et 2500 GeV. Les sélections appliquées dans la première analyse pour la région $\tau_\mu\tau_h$ ont été reprises pour cette deuxième analyse. La sélection *loose* de l'algorithme HPS a été choisie, afin d'avoir la meilleure efficacité possible pour l'identification des taus hadroniques. Trois variables de masse reconstruite différentes ont été considérées pour la recherche du boson Z' : la masse visible, la masse totale et la masse

transverse totale. Pour ces trois variables, leurs différences relatives par rapport à la masse générée du boson Z' , leurs courbes ROC et les significances du signal ont été comparées, et la masse totale s'est avérée être la meilleure variable discriminante pour cette analyse. Un bon accord entre les données et les prédictions du Modèle Standard a été trouvé. Une méthode statistique *cut-and-count* a été utilisée afin de déterminer les limites supérieures attendue et observée, à 95% de niveau de confiance, pour le nombre d'événements de signal exclus. Afin de transformer ce nombre en section efficace exclue, le produit de l'acceptance par l'efficacité des sélections appliquées a été calculé pour les quatre hypothèses de masse du boson Z' du MSS, des valeurs d'environ 3% ont été trouvées, en tenant compte de la valeur de r calculée précédemment pour la sélection *loose* de l'algorithme HPS. Un graphique d'exclusion de la section efficace du boson Z' du MSS en fonction de sa masse a été construit. Une limite inférieure attendue de 1.37 ± 0.13 TeV et une limite inférieure observée de 1.18 TeV ont été obtenues sur la masse du boson Z' , la différence entre les deux étant due à un léger excès de données observées par rapport au bruit de fond attendu. La limite observée est légèrement moins stricte que celle obtenue par des analyses précédentes de CMS, mais la limite attendue est meilleure. Les résultats obtenus dans ce mémoire sont donc très compétitifs par rapport à ceux de CMS obtenus avec les données du Run 1 du LHC, ceci grâce au gain d'énergie du centre de masse dans les données du Run 2.

L'étude de l'efficacité de l'algorithme HPS de reconstruction et d'identification des taus hadroniques a été particulièrement poussée. Il reste quelques détails à améliorer, comme l'utilisation de simulations plus précises pour le processus Drell-Yan, mais les résultats obtenus dans ce mémoire sont déjà très bons. Ils ont été présentés au groupe de travail sur les leptons taus de CMS, et ils sont en accord avec d'autres résultats obtenus indépendamment. Le rapport trouvé pour l'efficacité dans les données sur l'efficacité dans les simulations peut être utilisé par toutes les analyses de CMS travaillant avec des taus hadroniques : les recherches de nouvelle physique dans le canal $\tau^+\tau^-$ mais aussi la détermination du couplage du boson H aux leptons taus. La deuxième analyse, concernant la recherche d'un boson Z' dans le canal $\tau^+\tau^-$ a obtenu des résultats comparables que les analyses précédentes de CMS. Elle pourrait être améliorée en utilisant des simulations de bruit de fond plus appropriées pour les hautes masses invariantes, en considérant plus de points de masse pour le signal, surtout entre 1000 et 1500 GeV, en optimisant la sélection du canal $\tau_\mu\tau_h$ pour les hautes masses, et en analysant aussi d'autres états finaux de la paire $\tau^+\tau^-$. Une étude des erreurs systématiques est nécessaire, même si elles sont dominées par les erreurs statistiques dans cette analyse. La collection de plus de données lors du restant du Run 2 du LHC devrait permettre soit de découvrir une nouvelle résonance soit de poser des limites plus strictes. Pour l'année 2016, la luminosité attendue est de 40 fb^{-1} , et 150 fb^{-1} sont attendus d'ici la fin 2018 [34].

Bibliographie

- [1] A. Salam, J. C. Ward, *Electromagnetic and Weak interactions*, Phys. Rev. Lett. 13 (1964) 168
- [2] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. 13 (1964) 321
- [3] P. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. 13 (1964) 508
- [4] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. 19 (1967) 1264
- [5] M.Y. Han et Y. Nambu, *Three-Triplet Model with Double $SU(3)$ Symmetry*, Phys. Rev. Lett. 139 (1965) 1006
- [6] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B716 (2012) 1
- [7] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B716 (2012) 30
- [8] L. Susskind, *Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory*, Phys. Rev. D20 (1979) 2619
- [9] P. Grangé, J.-F. Mathiot, B. Mutet, et E. Werner, *Aspects of Fine-tuning of the Higgs Mass within Finite Field Theories*, Phys. Rev. D88 (2013) 125015
- [10] W. Hu and S. Dodelson, *Cosmic Microwave Background Anisotropies*, Ann. Rev. Astron. Astrophys. 40 (2002) 171
- [11] V.A. Kuzmin, V.A. Rubakov et M.E. Shaposhnikov, *On anomalous electroweak baryon-number non-conservation in the early universe*, Phys. Lett. B155 (1985) 36
- [12] J. Baez, *How Many Fundamental Constants Are There ?*, sur son site Internet <http://math.ucr.edu/home/baez/constants.html> (consulté le premier mars 2016)
- [13] J.M. Frère, *Physics beyond the standard model*, cours de MA2 de l'ULB
- [14] B. Grinstein, *TASI-2013 Lectures on Flavor Physics*, arxiv :1501.05283 (2015)
- [15] L. Randall, R. Sundrum, *A Large Mass Hierarchy from a Small Extra Dimension*, Phys. Rev. Lett. 83 (1999) 3370
- [16] T. Han, J. Lykken, R.-J. Zhang, *On Kaluza-Klein States from Large Extra Dimensions*, Phys. Rev. D59 (1999) 105006
- [17] M. Guchait, F. Mahmoudi et K. Sridhar, *Associated production of a Kaluza-Klein excitation of a gluon with a $t \bar{t}$ pair at the LHC*, Phys. Lett. B666 (2008) 347
- [18] R. Argurio, *Introduction to Supersymmetry*, cours de MA2 de l'ULB
- [19] E. Ma, *Gauged $B - 3L_\tau$ and Radiative Neutrino Masses*, Phys. Lett. B433 (1998) 74
- [20] T. G. Rizzo, *Z' Phenomenology and the LHC*, arxiv :hep-ph/0610104 (2006)

- [21] K.A. Olive et al. (Particle Data Group), *The Review of Particle Physics*, Chin. Phys. C38 (2014) 090001
- [22] G. Aad et al., *Measurement of hard double-parton interactions in $W(\rightarrow l\nu)+2$ jet events at $\sqrt{s}=7$ TeV with the ATLAS detector*, New J. Phys. 15 (2013) 033038
- [23] Site de l'UC Dublin, *Electroweak physics from Drell-Yan production of di-muons*, <http://www.ucd.ie/physics/lhcb/dy/dy.html> (consulté le 8 mars 2016)
- [24] Site du CERN : <http://home.cern/fr/about/engineering/radiofrequency-cavities> (consulté le 8 mars 2016)
- [25] Site du CERN : <http://home.cern/fr/about/engineering/pulling-together-superconducting-electromagnets> (consulté le 8 mars 2016)
- [26] Site du CERN : <https://cds.cern.ch/record/1621583> (consulté le 31 mars 2016)
- [27] R. Garoby, *Bunch merging and splitting techniques in the injectors for high energy hadron colliders*, rapport technique du CERN, CERN-PS-98-048-RF (1998)
- [28] D. Manglunki, *Calcul, technique et réalisation des accélérateurs de particules*, cours de MA2 de l'ULB donné au CERN
- [29] Site du CERN : <http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/beam.htm> (consulté le 31 mars 2016)
- [30] Twiki publique de ATLAS : <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BeamSpotPublicResults> (consultée le 31 mars 2016)
- [31] S. Bartolomé-Jiménez, G. Trinquart, *Installation of the LHC long straight sections (LSS)*, Proceedings of 2005 Particle Accelerator Conference, Knoxville, Tennessee
- [32] LHC Programme Coordination, *Luminosity plots pp (2015)*, http://lpc.web.cern.ch/lpc/lumiplots_2015_pp.htm (consulté le 12 avril 2016)
- [33] Site du CERN, *Longer term LHC schedule*, <https://lhc-commissioning.web.cern.ch/lhc-commissioning/schedule/LHC-long-term.htm> (consulté le 8 mars 2016)
- [34] M. Neubauer, *LHC (and ATLAS) Status and Plans*, présentation au Chicagoland LHC Workshop (2015) disponible à l'adresse web https://indico.cern.ch/event/372883/contribution/0/attachments/742159/1018109/Chicagoland_LHCstatus.pdf (consultée le 31 mars 2016)
- [35] CMS collaboration, *The CMS experiment at the CERN LHC*, JINST 3 (2008) S08004
- [36] C. Caillol, *Recherche du boson scalaire de Brout-Englert-Higgs dans le canal $H \rightarrow ZZ \rightarrow 2l2\nu$: étude des données 2011 du détecteur CMS*, mémoire de master ULB (2012)
- [37] G. de Lentdecker, P. Van Laer, *Détection des particules, acquisition et analyse de données*, cours de MA2 de l'ULB
- [38] Site de l'UCLA, www-collider.physics.ucla.edu/cms (consulté le 6 avril 2016)
- [39] D. Barney (2011) pour la collaboration CMS, <http://cms.web.cern.ch/org/cms-presentations-public> (consulté le 4 avril 2016)
- [40] V. Halyo, P. LeGresley, P. Lujan, *Massively Parallel Computing and the Search for Jets and Black Holes at the LHC*, arxiv :1309.6275 [physics.comp-ph] (2013)

- [41] C. Eggel, https://people.phys.ethz.ch/~ceggel/presentations/blhc_th/pdf/ (consulté le 11 avril 2016)
- [42] S. Schramm, *The ATLAS detector : status and performance in Run-II* (2016), disponible sur <https://cds.cern.ch/record/2131867> (consulté le 12 avril 2016)
- [43] A. Dabrowski, *CMS Status Report - Run II Performance and Recent Physics Highlights* (2015), présentation disponible sur <https://indico.cern.ch/event/460278/> (consulté le 12 avril 2016)
- [44] A. Buckley et al., *General-purpose event generators for LHC physics*, CERN-PH-TH-2010-298 (2011)
- [45] J. Allison et al., *GEANT4 - a simulation toolkit*, Phys. Research A506 (2003) 250
- [46] T. Seva, H. Y.-S. Tong, *Standard Model Cross Sections for CMS at 13 TeV*, TWiki interne de CMS (consulté le 15 avril 2016)
- [47] CMS collaboration, *Search for high mass resonances decaying into τ lepton pairs in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Lett. B716 (2012) 82
- [48] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer et Tim Stelzer, *MadGraph 5 : Going Beyond*, arxiv :1106.0522 [hep-ph] (2011)
- [49] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comp. Phys. Comm. 191 (2015) 159
- [50] J. AlWall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, arxiv :1405.0301 [hep-ph] (2014)
- [51] CMS collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus and E_T^{miss}* , CMS Physics Analysis Summary CMS-PAS-PFT-09-001 (2009)
- [52] CMS Collaboration, *Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV*, CMS PAS MUO-10-004 (2013)
- [53] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuc. Instr. and Meth. in Phys. Res. A262 (1987) 444
- [54] P. Verwilligen, *Muons in data analysis*, présentation disponible sur le site http://w3.iihe.ac.be/~jdhondt/Website/CSHEP_MuonDetectors_L4.pdf (consulté le 11 avril 2016)
- [55] W. Adam, R. Frühwirth, *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*, CMS Note 2005/001 (2005)
- [56] M. Cacciari, G. P. Salam et G. Soyez, *The anti- k_t jet clustering algorithm*, JHEP 0804 (2008) 063
- [57] CMS collaboration, *Performance of τ -lepton reconstruction and identification in CMS*, JINST 7 (2012).
- [58] CMS collaboration, *Tau reconstruction and identification in CMS during LHC run 1*, CMS Physics Analysis Summary CMS PAS TAU-14-001 (2015).
- [59] C. Böser, S. Fink et Steffen Röcker, *Introduction to Boosted Decision Trees*, diapositives disponibles sur le site web <https://indico.scc.kit.edu/indico/event/48/session/4/contribution/35/material/slides/0.pdf> (consulté le 6 mars 2016)
- [60] S. Arora, *MetResolution*, TWiki interne de CMS (consulté le 15 avril 2016)
- [61] Présentation de A. K. Kalsi lors d'une réunion interne du Tau Particle Object Group de CMS

- [62] CMS collaboration, *Performance of CMS Muon Reconstruction in Cosmic-Ray Events*, CMS Paper CFT-09-014 (2010), arxiv :0911.4994v2
- [63] CMS collaboration, *Evidence for the 125 GeV Higgs boson decaying to a pair of tau leptons*, JHEP 05 (2014) 104
- [64] N. Wardle, *Documentation of the RooStats-based statistics tools for Higgs PAG*, TWiki interne de CMS (consulté le 10 mai 2016)
- [65] A. Elagin, P. Murat, A. Pranko, A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to $\tau\tau$* , Nuc. Instr. and Meth. in Phys. Res. A654 (2011) 481
- [66] ATLAS collaboration, *Search for the Standard Model Higgs boson in the $H \rightarrow \tau^+\tau^-$ decay mode in $\sqrt{s} = 7$ TeV pp collisions with ATLAS*, JHEP 09 (2012) 070
- [67] CMS collaboration, *Search for high mass resonances and large extra-dimensions with τ -lepton pairs decaying to final states with one electron and one muon at $\sqrt{s} = 8$ TeV*, CMS-PAS-EXO-12-046 (2012)