



UNIVERSITÉ LIBRE DE BRUXELLES



FACULTÉ DES SCIENCES

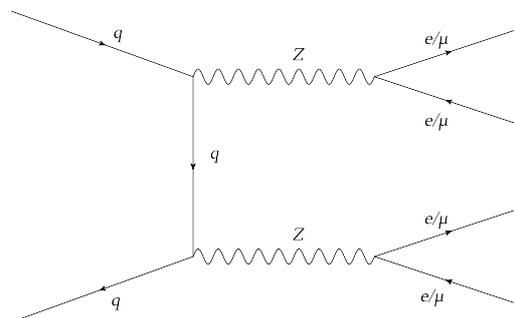
DÉPARTEMENT DE PHYSIQUE

INTER-UNIVERSITY INSTITUTE FOR HIGH ENERGIES

Étude de la production de paires de bosons Z à grande masse invariante dans des collisions proton-proton à $\sqrt{s} = 8$ TeV auprès de l'expérience CMS

Auteur :
Nicolas POSTIAU

Directeur :
Pr. Pascal VANLAER



Année académique 2014-2015

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION DU DIPLÔME DE
MASTER EN SCIENCES PHYSIQUES

Le mépris de la connaissance est un premier pas vers l'intolérance et le fascisme. L'anti-intellectualisme a permis d'amener les foules vers n'importe quelle horreur.

— F. ENGLERT, *Le Soir* du 8 octobre 2013.

Abstract

In this Master Thesis, we study the Z boson pair production in proton-proton collisions at 8 TeV in the center-of-mass collision frame, from the data collected in 2012 by the CMS experiment. This study is performed in the final states containing two electron-antielectron or muon-antimuon pairs. Our goals are to compute the cross section of the process, and to study the electroweak corrections applied to it. These electroweak corrections have only recently been computed, and are for the first time compared with experimental data in these final states. We first identify the sub-processes, at different orders in α_S , contributing to the Z boson pair production in simulated samples from specific generators and we compare these contributions to the published results. We then introduce a parameterization of these electroweak corrections to the dominant sub-process $q\bar{q} \rightarrow ZZ$ and compare the distributions to the compared results. We compare the data and the Monte-Carlo simulations, corrected for NLO electroweak effects, after having applied the necessary cuts in order to isolate the process of interest. After verifying the CMS trigger efficiency, we develop and implement a background estimation method, essentially based on the data samples. We conclude from this background study that the latter can be neglected in the adjustment, and simply subtracted for the cross section determination. The comparison between data and simulations leads to the conclusion that the simulated distributions with or without electroweak corrections are both compatible with the data. Finally, we determine the Z boson pair production cross section, imposing that for each Z boson $|M_Z - 91, 2 \text{ GeV}| < 30 \text{ GeV}$, and that the invariant mass of the two Z bosons system is greater than 200 GeV. This cross section is measured to be $(5.76 \pm 0.68) \text{ pb}$, in agreement with published results.

Keywords : High energy physics, LHC, CMS, data analysis, vector boson pair production, cross section measurement.

Résumé

Dans ce mémoire, nous étudions la production de paires de bosons Z dans des collisions proton-proton à 8 TeV dans le centre de masse de la collision, détectées en 2012 par l'expérience CMS. Cette étude est réalisée dans les états finaux comprenant deux paires électron-antiélectron ou muon-antimuon. Nos objectifs sont de déterminer la section efficace de ce processus, et d'étudier les corrections électrofaibles qui s'y appliquent. Ces corrections électrofaibles ont été calculées récemment et sont pour la première fois comparées avec les données expérimentales dans ces états finaux. Nous identifions tout d'abord les sous-processus à différents ordres en α_S qui contribuent à la production de paires de bosons Z dans des échantillons simulés par des générateurs spécifiques et comparons l'importance des différentes contributions aux résultats publiés dans la littérature. Nous introduisons ensuite une paramétrisation des corrections électrofaibles au sous-processus dominant $q\bar{q} \rightarrow ZZ$ et comparons les distributions aux résultats publiés. Nous comparons données et simulations Monte-Carlo corrigées pour les effets électrofaibles au NLO, après avoir appliqué les coupures nécessaires pour isoler le processus qui nous intéresse. Après avoir vérifié l'efficacité de déclenchement de l'expérience CMS, nous développons et implémentons une méthode d'estimation du bruit de fond basée essentiellement sur les données. Nous concluons de cette étude du bruit de fond que ce dernier peut être négligé lors de l'ajustement, et simplement soustrait lors de la détermination de la section efficace. La comparaison entre données et simulations nous permet de conclure que les spectres simulés avec et sans corrections électrofaibles sont tous deux compatibles avec les données. Finalement, nous déterminons la section efficace de production de paires de bosons Z , en imposant pour chaque boson Z que $|M_Z - 91,2 \text{ GeV}| < 30 \text{ GeV}$, et que la masse invariante du système formé par les deux bosons Z soit supérieure à 200 GeV. Cette section efficace est mesurée à $(5,76 \pm 0,68) \text{ pb}$, en accord avec la littérature.

Mots-clés : Physique des hautes énergies, LHC, CMS, analyse de données, production de paires de bosons vecteurs, mesure de section efficace.

Remerciements

Mes remerciements vont tout d'abord, logiquement, vers mon directeur de mémoire, Pascal Vanlaer. Toujours très disponible et n'hésitant pas à consacrer un temps considérable pour discuter du travail, il a toujours été très ouvert pour répondre à mes questions parfois stupides, et souvent trop nombreuses. Je le remercie de tout cœur pour tout ce qu'il m'a appris cette année, et pour son suivi durant toute cette longue période.

Je remercie mes lecteurs Barbara Clerboux et Michel Tytgat pour leurs commentaires avisés, et leur souhaite une bonne lecture pour les 70 prochaines pages de ce mémoire.

I would like to thank Jian, who first introduced me to the wonderful world of ROOT programming two years ago, and who still helped me at the beginning of this work before flying to CERN.

Merci à Hugo, qui m'a aidé à plusieurs reprises sur des détails parfois fort techniques et sur le plan moral, étant passé par le même chemin que moi l'année précédente. Merci à Patrick qui, avant de partir pour des contrées germaniques, m'a fourni quelques précieux conseils informatiques (il sera heureux de savoir que je me suis mis à Vim). Merci également à Alexandre pour ses dépannages occasionnels.

De manière générale, je remercie tout le service de l'IIHE pour son accueil et sa convivialité.

Je remercie également l'équipe technique, qui a fait preuve d'une grande diligence pour résoudre mes problèmes de réseau au début de l'année.

Merci à mes amis et collègues de bureau, Laurent et Baptiste, qui ont largement contribué à la bonne ambiance dans laquelle s'est passée ce travail.

Finalement, je remercie ma maman pour le soutien moral indispensable lors d'un travail de si longue durée, et pour la correction de l'orthographe dans ce travail.

Table des matières

Abstract	iv
Résumé	v
Remerciements	vi
Table des matières	viii
Introduction	1
1 Production de paires de bosons Z au LHC	3
1.1 Processus étudiés	3
1.1.1 Processus à l'ordre le plus bas	4
1.1.2 Processus d'ordres supérieurs en QCD	5
1.1.3 Processus avec corrections électrofaibles	7
1.2 Valeurs théoriques des sections efficaces	8
2 Étude du processus ZZ au niveau généré	11
2.1 Simulations des événements	11
2.2 Contribution des différents processus	13
2.3 Contribution selon les types de quarks	13
2.4 Étude des corrections électrofaibles	14
2.4.1 Implémentation des corrections électrofaibles	14
2.4.2 Effets des corrections électrofaibles au niveau généré	15
2.4.3 Restriction due à l'activité hadronique	16
3 L'expérience CMS du LHC	21
3.1 Le collisionneur LHC	21
3.1.1 La chaîne d'injecteurs	21
3.1.2 Caractéristiques du LHC	21
3.2 L'expérience CMS	23
3.2.1 Système de coordonnées	23
3.2.2 Composants du détecteur CMS	25
3.2.3 Reconstruction et identification des particules	29
3.3 Échantillons de données et de simulations	31
3.3.1 Échantillons de données	31
3.3.2 Échantillons de simulations	31

4	Sélection des événements	33
4.1	Coupures de sélection	33
4.1.1	Sélection de l'état final	34
4.1.2	Coupures sur la quantité de mouvement transverse des leptons et la couverture du détecteur	34
4.1.3	Identification	35
4.1.4	Isolement	36
4.1.5	Coupures d'analyse	36
4.1.6	Autres coupures	37
4.2	Distributions de contrôle	37
4.3	Pondération des événements simulés selon le nombre de vertex	37
4.4	Vérification de l'efficacité des triggers	39
4.5	Distributions pour l'échantillon final	41
5	Calcul du bruit de fond	47
5.1	Principe	47
5.2	Calcul du taux d'erreur de sélection des leptons	48
5.3	Estimation du bruit de fond	50
5.4	Conclusions concernant le bruit de fond	52
6	Résultats	55
6.1	Description de la méthode d'analyse	55
6.2	Distributions	56
6.3	Ajustement des données et sensibilité aux corrections électrofaibles	61
6.3.1	Test du χ^2	61
6.3.2	Résultats du test avec et sans corrections	61
6.3.3	Sensibilité aux corrections électrofaibles	62
6.4	Incertitudes systématiques	62
6.5	Calcul de $\sigma_{pp \rightarrow ZZ}$	63
6.5.1	Acceptance de l'analyse	64
6.5.2	Détermination de la section efficace	64
	Conclusions et perspectives	67
A	Contributions des différents types de quarks	69
B	Calcul de l'incertitude sur le bruit de fond	71
C	Maximum de vraisemblance pour la section efficace	75
	Bibliographie	78

Introduction

En 2012, le *Large Hadron Collider*, ou LHC, a pour la première fois récolté des données de collisions proton-proton à une énergie de 8 Téra-électronVolts (TeV) dans le centre de masse. L'analyse des données de ces collisions, prises pendant une période de près d'un an par les détecteurs CMS et ATLAS, a permis des progrès majeurs dans de nombreux domaines de la physique des particules. L'une des principales découvertes du LHC fut la découverte du boson scalaire de Brout-Englert-Higgs, ainsi que la mesure de la masse de ce dernier. Le canal de désintégration du scalaire ayant permis la mesure la plus précise de sa masse est celui où le boson se désintègre en une paire de bosons Z , l'un d'eux au moins étant virtuel, et où chaque Z se désintègre ensuite en une paire électron-antiélectron ou muon-antimuon [1].

Même si le boson scalaire prédit par Brout, Englert et Higgs a désormais été découvert et que sa masse est bien connue, des théories au-delà du modèle standard prédisent l'existence d'autres bosons scalaires plus massifs présentant les mêmes canaux de désintégration. En particulier, la recherche d'un boson scalaire massif se désintégrant en deux bosons Z réels est une étude très prometteuse. Le challenge de cette étude est d'estimer avec précision le bruit de fond sur cette production de boson scalaire, provenant principalement de production de paires de bosons Z par des mécanismes n'impliquant pas de boson scalaire.

Récemment, une méthode a été également développée pour mettre une borne supérieure sur la largeur de désintégration du boson scalaire du Modèle Standard [2]. Cette méthode se base sur l'analyse de la "queue de distribution" du scalaire, à de hautes valeurs de masse invariante. Il est donc également crucial dans ce cas de connaître avec précision la section efficace de production de paires de bosons Z , qui forme le bruit de fond dominant pour cette étude. Le processus de paires de bosons Z permet également de rechercher des couplages triples anormaux ZZZ et $ZZ\gamma$, non prédits par le Modèle Standard.

Dans ce mémoire, nous nous intéressons plus en détail à ce processus de production de paires de bosons Z réels. Pour ce faire, nous utilisons les données récoltées lors de l'année 2012 par l'expérience CMS, installée au LHC. Notre principal objectif est de déterminer la section efficace de ce processus à des hautes valeurs de masse invariante pour la paire formée par les deux bosons Z , en regardant uniquement les paires de bosons Z à plus de 200 GeV de masse invariante. La détermination précise de cette section efficace est indispensable pour les recherches décrites plus haut.

La particularité de ce travail est l'étude des corrections électrofaibles au premier ordre sur le processus étudié. Ces corrections ont été calculées très récemment, et aucune analyse expérimentale dans un état final à quatre leptons chargés ne les avait encore incluses au moment de débiter ce travail. Nous étudierons en particulier

l'impact de ces corrections électrofaibles sur la forme des distributions.

Les collaborations CMS et ATLAS ont déjà publié des résultats concernant la section efficace de production de paires de bosons Z à partir d'états finaux à quatre leptons chargés [3, 4], sans toutefois faire intervenir les corrections électrofaibles. CMS a aussi récemment publié une mesure de cette section efficace dans un état final à deux leptons chargés et deux neutrinos [5]. Dans ces trois articles, la région cinématique est cependant différente de notre analyse puisque les auteurs n'imposent pas de coupure sur la masse invariante de la paire de bosons Z .

Ce mémoire est divisé en 6 chapitres.

Le premier présente le processus étudié au niveau théorique, introduit les différentes contributions des processus aux ordres supérieurs en interactions fortes et électrofaibles, et donne les sections efficaces présentées dans la littérature.

Le deuxième chapitre présente notre analyse du processus de production de paires de bosons Z sur base de simulations au niveau généré. Nous y introduisons une paramétrisation des corrections électrofaibles pour le processus $q\bar{q} \rightarrow ZZ$ et nous la validons par une comparaison avec les résultats publiés.

Le troisième chapitre est dédié à la description de l'expérience CMS, sur laquelle nous travaillons, ainsi qu'aux échantillons de simulations utilisés.

Le quatrième chapitre décrit les coupures utilisées pour restreindre notre échantillon au processus que nous étudions, ainsi que les comparaisons entre les données et les simulations. L'efficacité de détection en ligne (triggers) y est aussi étudiée.

Le cinquième chapitre traite du problème du bruit de fond associé à notre processus. En effet, ce bruit de fond est faible et généralement mal décrit par des simulations. Nous avons utilisé et développé une méthode permettant de mesurer ce bruit de fond dans les données et de s'affranchir ainsi des simulations de bruit de fond.

Enfin, le sixième et dernier chapitre présente nos deux résultats principaux : la conclusion sur la sensibilité de nos données aux corrections électrofaibles, et finalement la mesure de la section efficace de production de paires de bosons Z à haute masse invariante lors de collisions proton-proton à 8 TeV.

Chapitre 1

Production de paires de bosons Z au LHC

Dans ce chapitre, nous décrivons les différents processus qui contribuent à la production d'une paire de bosons Z dans des collisions proton-proton à haute énergie. Nous expliquons en particulier l'importance des processus d'ordre supérieur en les constantes de couplage forte α_S et électromagnétique α . Ces corrections seront appelées par la suite respectivement corrections QCD et corrections électrofaibles. Ensuite, nous passons en revue les résultats récents des calculs théoriques des sections efficaces, publiés dans la littérature.

1.1 Processus étudiés

Nous nous intéressons à la production d'une paire de bosons Z réels à haute masse invariante, dans l'état final où chaque boson Z se désintègre soit en une paire e^+e^- , soit en une paire $\mu^+\mu^-$. Pour ce faire, on ne regarde que les événements où les deux bosons ont une masse invariante $M_{l^+l^-}$ comprise dans l'intervalle $|M_{l^+l^-} - 91,2 \text{ GeV}| < 30 \text{ GeV}$, l^+l^- désignant soit e^+e^- , soit $\mu^+\mu^-$, et où la masse invariante du système des deux bosons Z , notée M_{ZZ} , est supérieure à 200 GeV.

Le rapport de branchement du boson Z en une paire de leptons chargés d'une famille de leptons donnée vaut [6] :

$$\text{Br}(Z \rightarrow l^+l^-) = 0,03366 \pm 0,00002. \quad (1.1)$$

En tenant compte des combinaisons possibles, ceci signifie que nous n'analyserons en réalité que $4 \times (0,03366^2) \simeq 0,45\%$ des événements où une paire de bosons Z est produite. Cependant, l'avantage de ce canal de désintégrations est que les électrons et muons sont facilement identifiables dans le détecteur. La masse invariante et le vecteur quadri-impulsion de chaque boson Z sont reconstruits à partir de la mesure des quantités de mouvement des leptons chargés.

Le calcul de la section efficace d'un processus en collisionneur hadronique se décompose en deux éléments. Le premier est le calcul perturbatif de la section efficace d'interaction entre les composantes des protons qui entrent en collision, appelées partons : quarks, antiquarks et gluons. Cette section efficace, dite partonique, dépend

des fractions x_1 et x_2 des quantités de mouvement des protons emportées par les composants en interaction. Le deuxième élément est la convolution de la section efficace partonique avec les distributions de ces fractions. Ces dernières sont décrites par les fonctions de distribution des partons $f(x, Q^2)$, où Q^2 est l'échelle dure du processus, autrement dit l'énergie dans le centre de masse du système partonique. La section efficace totale se calcule comme [7] :

$$\sigma_{pp \rightarrow ZZ+X} = \sum_{a,b \in \{q, \bar{q}, g\}} \int_0^1 dx_1 \int_0^1 dx_2 f_a(x_1, \mu_F) f_b(x_2, \mu_F) \hat{\sigma}_{ab \rightarrow ZZ+X}(x_1, x_2, \mu_F, \alpha_S(\mu_R)), \quad (1.2)$$

où f_a et f_b sont les fonctions de distribution des partons ayant initié le processus dur, $\hat{\sigma}_{ab \rightarrow ZZ+X}$ est la section efficace du processus dur partonique, μ_F est l'échelle de factorisation du calcul et μ_R l'échelle de renormalisation. L'échelle μ_F correspond à la valeur à laquelle la composition des protons en partons a et b est sondée par le processus dur. L'échelle μ_R représente l'échelle à laquelle α_S est évaluée, et indique le fait que le calcul perturbatif de $\hat{\sigma}$ est limité à un certain ordre en α_S . On prend conventionnellement $\mu_F = \mu_R = Q = 2M_Z$, l'échelle dure du processus. Par la suite, une méthode conventionnelle pour estimer l'incertitude liée au choix du jeu de fonctions de distribution et du développement limité de $\hat{\sigma}$ est de regarder l'impact de la différence de section efficace totale en faisant varier μ_F et μ_R par un facteur 2 et $\frac{1}{2}$.

1.1.1 Processus à l'ordre le plus bas

Le processus à l'ordre le plus bas est celui dont le diagramme de Feynman comporte le moins de vertex. Il est initié par une paire quark-antiquark. Son diagramme est représenté à la figure 1.1.

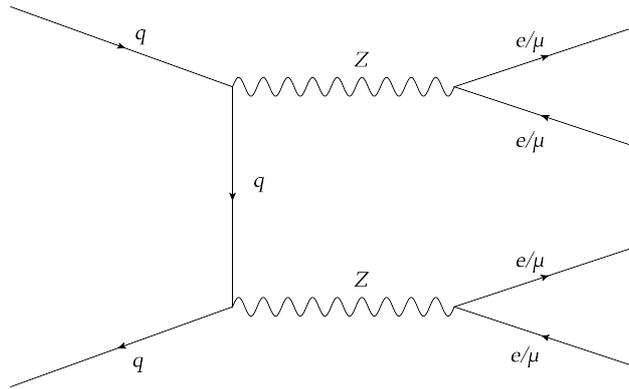


FIGURE 1.1 – Processus $q\bar{q} \rightarrow ZZ \rightarrow l^+l^-l'^+l'^-$ ($l, l' \equiv e, \mu$).

Ce processus implique directement une contribution des quarks de la mer présents dans le proton. En effet, si le quark q peut être un quark de valence, l'antiquark \bar{q} ne le peut pas.

1.1.2 Processus d'ordres supérieurs en QCD

La figure 1.2 montre les diagrammes qui contribuent à l'ordre suivant en α_S . Les diagrammes avec une boucle interfèrent avec le diagramme à l'ordre le plus bas. Ceci produit un terme d'interférence à l'ordre α_S , qui annule la divergence infrarouge due à la radiation d'un gluon de faible quantité de mouvement transverse ou colinéaire avec le faisceau. Ces différentes contributions changent également les spectres en quantité de mouvement des bosons Z dans le plan transverse aux faisceaux, et en masse invariante du système formé par la paire de bosons Z [8].

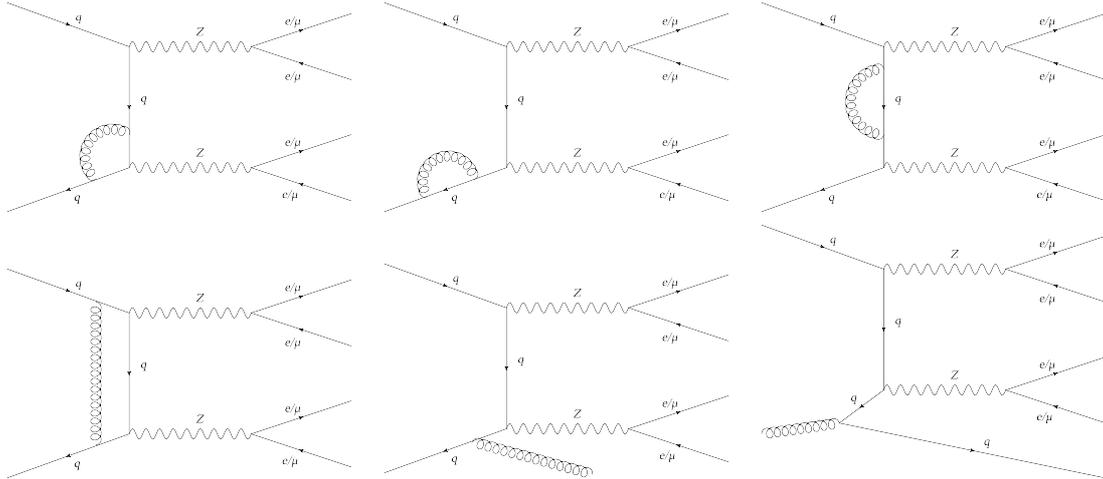


FIGURE 1.2 – Diagrammes qui contribuent à l'ordre suivant en α_S .

Le diagramme en bas à droite, induit par un quark et un gluon, a une contribution particulièrement importante en raison de la grande densité de gluons dans les protons aux valeurs de x et de Q^2 considérées. En effet, dans notre cas, on a $Q = M_{4l} \cong 200$ GeV. Comme $M_{4l} = \sqrt{s x_1 x_2}$, où \sqrt{s} est l'énergie de la collision dans le système du centre de masse ($\sqrt{s} = 8$ TeV), on a $x_1 \simeq x_2 = \frac{M_{4l}}{\sqrt{s}} \simeq 0,025$. On peut voir à la figure 1.4 que la proportion des gluons dans le proton à ces valeurs de x est très importante, même si cet effet est compensé partiellement par la valeur de α_S : on a que $\alpha_S(Q^2 = (200 \text{ GeV})^2) \simeq 0,10$ [6]. La contribution de ce diagramme est donc importante, comme on le verra au chapitre 2.

La figure 1.3 montre un autre diagramme important, celui-ci à l'ordre α_S^2 . Pour les mêmes raisons qu'énoncé plus haut, la contribution de ce diagramme est non négligeable : elle est évaluée à environ 7 % dans [9] pour $\sqrt{s} = 8$ TeV.

D'autres contributions donnant le même état final sont possibles. En particulier, il est possible d'avoir une double radiation de bosons Z par 2 quarks différents (fig. 1.5). Nous avons pu déterminer, sur base de simulations avec le générateur **MadGraph** [11], que la contribution de ce processus est d'environ 2 % des événements attendus dans l'acceptance de notre analyse. Comme nos simulations ne reprennent pas ce processus, une erreur systématique correspondante sera appliquée à notre analyse finale.

La contribution de diagrammes impliquant la production d'un boson scalaire à partir d'un état initial à deux gluons, comme illustré en figure 1.6, devrait également

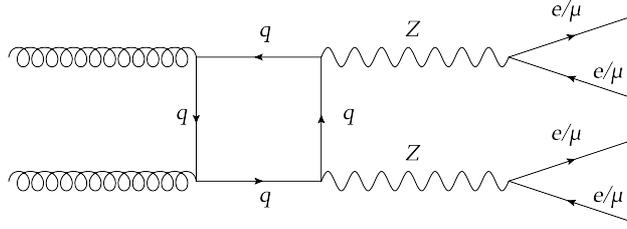


FIGURE 1.3 – Processus $gg \rightarrow ZZ \rightarrow l^+l^-l'^+l'^-$ ($l, l' \equiv e, \mu$).

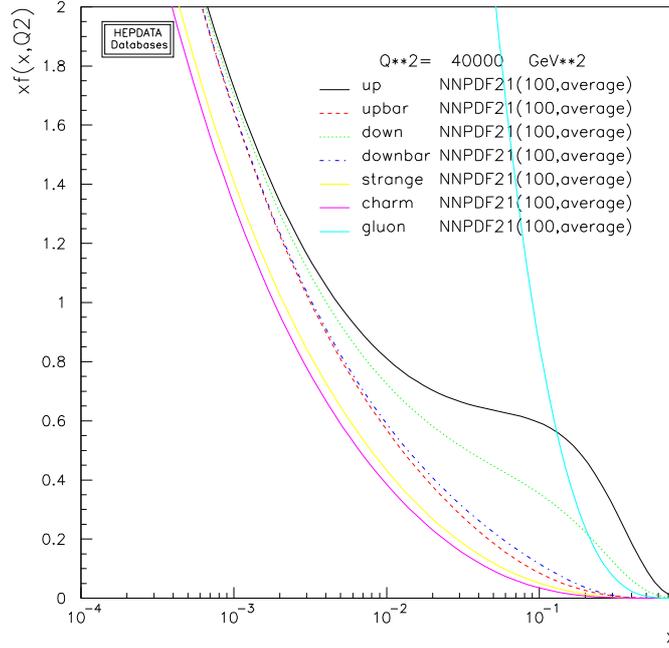


FIGURE 1.4 – Fonctions de distribution de partons à une échelle d'énergie $Q^2 = (200 \text{ GeV})^2$. La figure reprend la paramétrisation NNPDF21. [10]

être envisagée. Ce diagramme, bien que très important pour la mesure des caractéristiques du boson scalaire, contribue très peu à la section efficace dans le domaine cinématique considéré. En effet, selon [12], la section efficace de production du scalaire intégrée hors de son pic de production, à des énergies supérieures à 130 GeV, est de 0,061 fb à $\sqrt{s} = 8 \text{ TeV}$ pour un état final $e^+e^-\mu^+\mu^-$, ce qui correspond à 0,027 pb pour tous les états finaux possibles. Comme la section efficace que l'on veut mesurer est de l'ordre de 5 pb, cette contribution, d'environ 0,5 %, peut être négligée pour notre analyse.

La contribution des diagrammes d'ordre supérieur est, par convention, estimée en testant la stabilité du résultat par rapport au choix des paramètres non-physiques que sont μ_F et μ_R . En variant ces échelles d'un facteur 2 par rapport à leur valeur centrale choisie égale à $2M_Z$, les références [8, 9] trouvent une variation de 3 à 5 %.

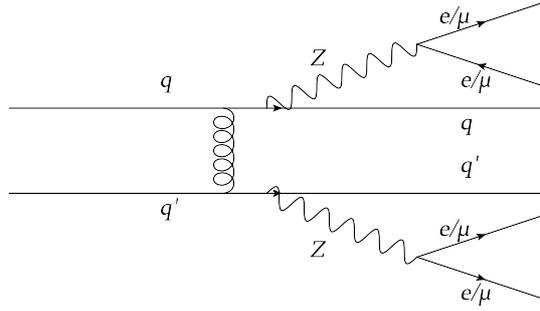


FIGURE 1.5 – Processus $qq' \rightarrow ZZ \rightarrow l^+l^-l'^+l'^-$ ($l, l' \equiv e, \mu$).

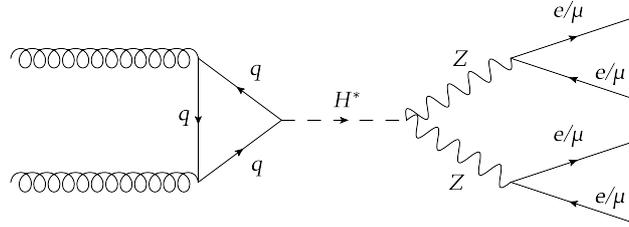


FIGURE 1.6 – Processus $gg \rightarrow H \rightarrow ZZ \rightarrow l^+l^-l'^+l'^-$ ($l, l' \equiv e, \mu$).

1.1.3 Processus avec corrections électrofaibles

Le processus de production de paires de bosons Z peut recevoir différentes corrections électrofaibles. Ces corrections, notées δ_{EW} , ont été calculées récemment [8, 13, 14]. Ce mémoire est l'occasion d'introduire les corrections électrofaibles dans la mesure de la section efficace ZZ pour un état final à quatre leptons chargés, ce qui n'avait pas encore été fait auparavant.

Les diagrammes contribuant à ces corrections se répartissent en trois catégories :

- émission d'un γ ou d'un boson Z ou W réel ;
- interactions induites par un photon, du type $\gamma q \rightarrow ZZq$;
- corrections virtuelles avec un Z , γ ou W dans la boucle (fig. 1.7).

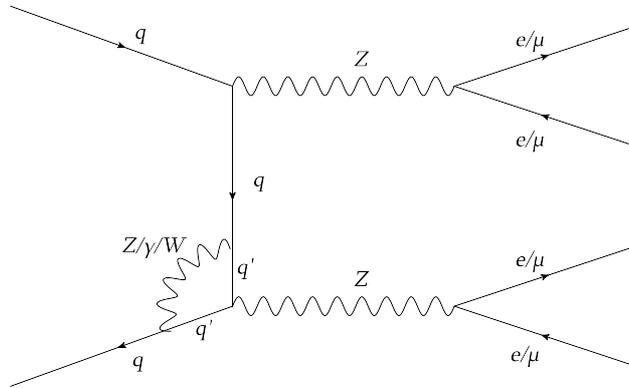


FIGURE 1.7 – Exemple de diagramme au premier ordre de perturbation électrofaible.

Seules les corrections virtuelles sont significatives pour la production d'une paire de bosons Z [8]. Comme on peut le voir à la figure 1.8, elles s'élèvent à environ -5% en moyenne dans la région cinématique $M_{ZZ} > 200$ GeV (où M_{ZZ} désigne la masse invariante du système des deux bosons Z), et atteignent -10% pour $M_{ZZ} \cong 600$ GeV. Elles atteignent également -25% pour $p_{T,Z} = 400$ GeV, où $p_{T,Z}$ est défini comme la quantité de mouvement du boson Z dans le plan transverse à l'axe du faisceau. L'effet de ces corrections est étudié en détail au chapitre 2.

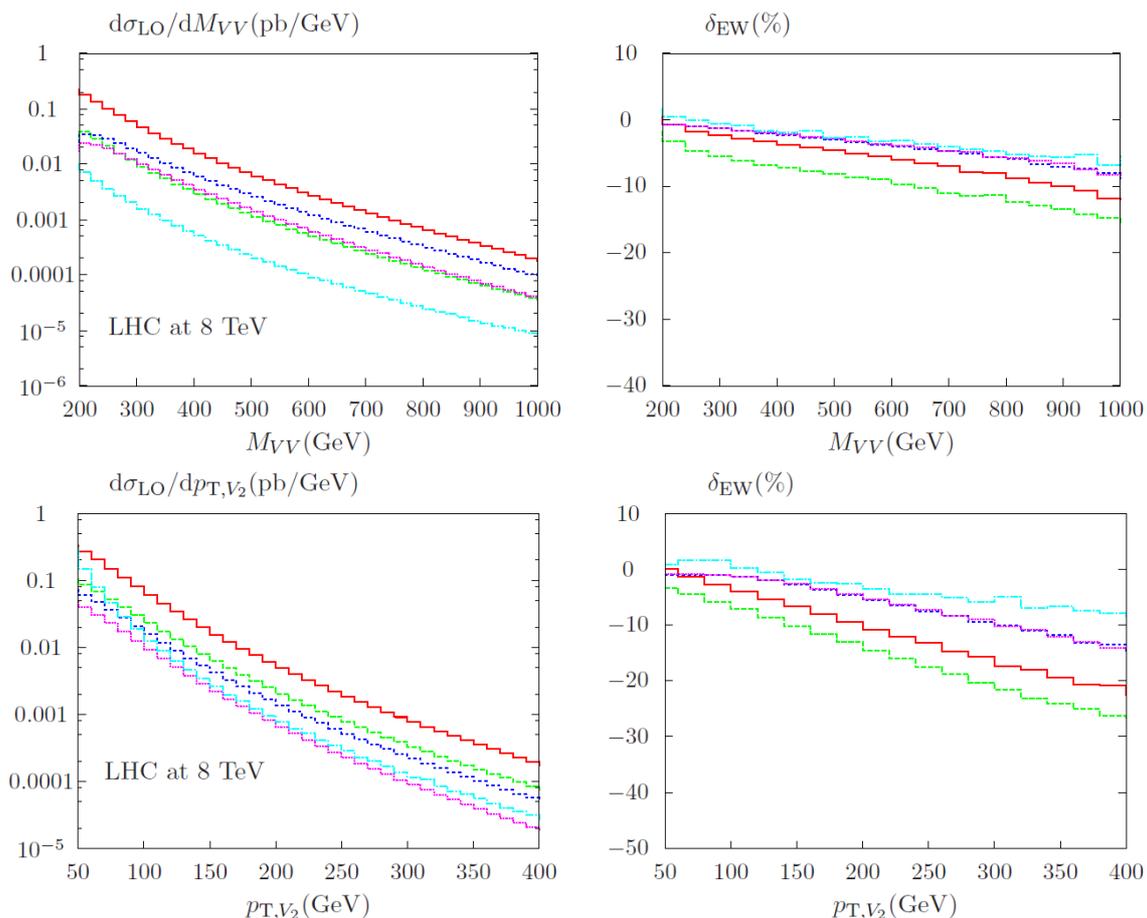


FIGURE 1.8 – À gauche, distributions de masse invariante M_{ZZ} (en haut) et de quantité de mouvement transverse d'un boson Z $p_{T,Z}$ (en bas). À droite, corrections électrofaibles en fonction de ces variables. Ces figures sont issues de [13]. La ligne verte représente le processus ZZ . Les lignes rouge, bleu foncé, violette et bleu clair représentent respectivement les processus W^+W^- , W^+Z , W^-Z et $\gamma\gamma$.

1.2 Valeurs théoriques des sections efficaces

La table 1.1 donne une comparaison entre les différents résultats théoriques parus dans la littérature concernant les sections efficaces avec et sans corrections QCD et électrofaibles. L'observation de ces résultats nous permet de conclure que les cor-

rections QCD sont importantes : elles sont de +57 % selon la référence [9]. En ce qui concerne les corrections électrofaibles, on constate qu'une augmentation en énergie, de 8 TeV à 14 TeV, changera fortement les sections efficaces mais ne modifiera pas significativement la valeur relative des corrections électrofaibles.

Source	$\sigma_{\text{LO}}[\text{pb}]$		$\sigma_{\text{NLO}}[\text{pb}]$		δ_{EW}	
	8 TeV	14 TeV	8 TeV	14 TeV	8 TeV	14 TeV
Réf. [8]			$7,30^{+3,2\%}_{-2,3\%}$ *	$16,24^{+2,6\%}_{-1,7\%}$ *	$\sim -4\%$	$\sim -4\%$ ***
Réf. [13]	5,064	11,02			$-4,1\%$	$-4,2\%$
Réf. [9]	5,06	10,92	$7,92^{+4,7\%}_{-3,0\%}$ **	$17,72^{+3,5\%}_{-2,5\%}$ **		

TABLE 1.1 – Valeurs théoriques des sections efficaces avec et sans corrections QCD, ainsi que des corrections électrofaibles. Ces valeurs sont données pour des collisions proton-proton à 8 et à 14 TeV. L'incertitude sur $\sigma_{\text{NLO,QCD}}$ est estimée en variant les échelles de factorisation et de renormalisation. *Inclut les corrections QCD et électrofaibles. **Inclut uniquement les corrections QCD. ***Le chiffre d'environ -3% , cité dans le texte de [8], est contestable : on voit clairement à la figure 7 que la correction est de -4% .

Les résultats présentés à la table 1.1 sont donnés pour une paire de bosons Z sur leur couche de masse, mais sans restrictions sur la masse invariante du système des deux bosons Z (M_{ZZ}). Dans le cadre de notre étude, nous nous restreignons à des valeurs de masse invariante $M_{ZZ} \geq 200$ GeV. Dans cet intervalle cinématique, la référence [13] donne $\sigma_{\text{LO}} = 3,271$ pb, et $\delta_{\text{EW}} = -4,9\%$. En tenant compte de l'effet des corrections QCD, on peut donc estimer que la section efficace correspondant à notre région cinématique est de 5,1 pb, et que les corrections électrofaibles sont d'environ -5% .

Chapitre 2

Étude du processus ZZ au niveau généré

Dans ce chapitre, nous étudions certaines caractéristiques cinématiques du processus $pp \rightarrow ZZ \rightarrow 4l$ et des corrections électrofaibles qui lui sont associées à partir d'échantillons d'événements simulés. Tout d'abord, nous décrivons les simulations d'événements que nous utilisons pour faire cette étude. Ensuite, nous étudions la contribution relative des différents diagrammes présentés au chapitre 1, ainsi que les proportions relatives des types de quarks qui ont initié l'interaction. Nous poursuivons avec la description des corrections électrofaibles dans les simulations, et avec l'étude de leur impact sur la cinématique du processus afin de valider les résultats parus dans la littérature. Enfin, nous clôturons ce chapitre par une discussion du choix de l'échantillon d'événements le plus approprié pour tester la validité des corrections électrofaibles dans les données.

2.1 Simulations des événements

Nous basons notre analyse sur des simulations Monte-Carlo (MC) du processus de production de paires de bosons Z . Ces simulations proviennent de deux générateurs différents : `POWHEG` et `gg2vv`.

Les simulations des processus au LO (fig. 1.1) et NLO (fig. 1.2) proviennent du générateur `POWHEG` [15, 16, 17], ou Positive Weight Hardest Emission Generator. `POWHEG` est un générateur d'événements simulés au NLO en QCD, où la simulation du reste de l'événement est basée sur le programme `PYTHIA` [15]. La génération de ces échantillons simulés se fait en plusieurs étapes :

- tout d'abord, on génère le processus dur. Cette partie est faite par `POWHEG`, qui ne génère que les processus au LO et au NLO en QCD ;
- ensuite, on génère la *cascade partonique* (ou *parton shower*), composée d'émission de gluons à grand angle par rapport à l'axe du faisceau dans les états initial et/ou final. Cette génération n'est pas calculée de manière exacte, mais est paramétrée par le programme `PYTHIA` [18] ;
- `PYTHIA` gère également la *fragmentation* des hadrons, c'est-à-dire la recombinaison des quarks et des gluons dans l'état final en hadrons, particules observables dans le détecteur. Les jets sont reconstruits à partir de ces hadrons ;

- la simulation de l'événement sous-jacent (*underlying event*) est également prise en charge par PYTHIA. Il s'agit d'une simulation des interactions entre les restes des protons initiaux portant une charge de couleur ;
- la dernière tâche de PYTHIA est de calculer les effets du *pile-up*, c'est-à-dire de l'effet des interactions entre d'autres protons lors du même croisement de paquets de protons des deux faisceaux ;
- enfin, il reste à simuler les interactions des particules produites dans l'état final avec le détecteur CMS. La simulation de la réponse du détecteur est faite par le programme CMSSW, basé lui-même sur GEANT4 [19] pour décrire la géométrie du détecteur et les interactions entre les particules et la matière du détecteur.

Nous avons utilisé des échantillons dans lesquels strictement 2 bosons Z ont été produits, dans les 3 états finaux que nous étudions : $e^+e^-e^+e^-$, $\mu^+\mu^-\mu^+\mu^-$ et $e^+e^-\mu^+\mu^-$. Notons que certaines coupures cinématiques ont déjà été appliquées sur ces échantillons, mais elles sont systématiquement moins strictes que celles que nous appliquerons dans notre analyse (*cf* chapitre 4). Les simulations POWHEG comprennent également des processus de production d'un seul boson Z avec radiation d'un autre Z off-shell, qui donnent également quatre leptons dans l'état final ; ce n'est cependant pas un problème pour notre analyse, car ce processus ne se produit qu'à une masse invariante proche de la masse du Z , qui est bien inférieure à la limite de notre analyse.

POWHEG étant un générateur NLO, le processus de production par une paire de gluons, représenté à la figure 1.3, n'y est pas simulé car il est du second ordre en QCD. Pour simuler ce processus, on se sert d'un autre générateur : **gg2vv** [20], qui génère uniquement des processus de production de paires de bosons vecteurs induits par des gluons et impliquant une boucle de quarks.

La table 2.1 indique les sections efficaces simulées pour chaque échantillon d'événements de signal, ainsi que le nombre d'événements simulés. La connaissance de ces valeurs, ainsi que de la luminosité, permet de normaliser le nombre d'événements simulés afin de le comparer aux données.

Simulation	σ [fb]	nombre d'événements
POWHEG $pp \rightarrow ZZ \rightarrow 4e$	76,91	1499093
POWHEG $pp \rightarrow ZZ \rightarrow 4\mu$	76,91	1499064
POWHEG $pp \rightarrow ZZ \rightarrow 2e2\mu$	176,7	1497445
gg2vv $gg \rightarrow ZZ \rightarrow 4e/4\mu$	4,8	548760
gg2vv $gg \rightarrow ZZ \rightarrow 2e2\mu$	12,03	393774

TABLE 2.1 – Sections efficaces et nombres d'événements pour chaque échantillon de simulations.

Dans la suite de ce chapitre, nous étudierons uniquement les simulations au niveau généré, sans tenir compte de la reconstruction des leptons et des jets dans le détecteur.

2.2 Contribution des différents processus

Le tableau 2.2 reprend les proportions des différents états initiaux simulés pour le processus dur. Ces proportions représentent les fractions d'événements de chaque type attendus dans la région cinématique qui nous intéresse : $M_{ZZ} \geq 200$ GeV et bosons Z_1 et Z_2 sur leur couche de masse : $|M_{Z_{1,2}} - 91,2 \text{ GeV}| < 30 \text{ GeV}$.

$q\bar{q} \rightarrow ZZ(+g)$	$qg \rightarrow ZZ + q$	$gg \rightarrow ZZ$
63,29 %	30,85 %	5,87 %

TABLE 2.2 – Contributions des différents états initiaux au processus ZZ .

On constate qu'une partie importante des événements sont initiés par des gluons. C'est un reflet de l'impact élevé des processus au premier ordre de perturbation QCD, comme discuté au chapitre précédent. On constate également que le processus $gg \rightarrow ZZ$ contribue de façon non négligeable et ne peut pas être ignoré dans notre analyse.

2.3 Contribution selon les types de quarks

La table 2.3 indique, selon la simulation par le programme POWHEG, la saveur des quarks ayant initié l'interaction dure, dans le cas d'événements du type $q\bar{q} \rightarrow ZZ$ dans la région cinématique étudiée.

$u\bar{u}$	$d\bar{d}$	$s\bar{s}$	$c\bar{c}$	$b\bar{b}$
39,4 %	50,4 %	8,6 %	1,1 %	0,5 %

TABLE 2.3 – Proportions des types de quarks ayant initié l'interaction dure dans le cas d'événements $q\bar{q} \rightarrow ZZ$.

On constate que la contribution des quarks $d\bar{d}$ est supérieure à celle des quarks $u\bar{u}$. Ce résultat est contre-intuitif car on a vu à la figure 1.4 que la densité de quarks d est plus faible que celle de quarks u , le proton ayant pour quarks de valence deux quarks u et un seul quark d .

L'explication de ce phénomène provient de la forme générale du vertex d'interaction $Zq\bar{q}$. En effet, dans [21] on voit que le vertex $Zq\bar{q}$ dans un diagramme de Feynman correspond à un facteur dans le calcul de l'élément de matrice dont l'expression est :

$$\frac{-ig}{\cos\theta_W} \frac{1}{2} \gamma^\mu (C_V^q - C_A^q \gamma^5), \quad (2.1)$$

où C_V^q et C_A^q sont respectivement les couplages vectoriel et axial des quarks, définis comme :

$$C_V^q = T_q^3 - 2 \sin^2 \theta_W Q^q \quad (2.2)$$

$$C_A^q = T_q^3, \quad (2.3)$$

où T_q^3 et Q^q sont respectivement l'isospin faible et la charge du quark q . La mesure de l'angle de mélange θ_W donne $\sin^2 \theta_W = 0,231$ [6]. Par conséquent, les couplages vectoriel et axial des quarks u et d sont :

$$C_V^u = +\frac{1}{2} - 2 \cdot \left(\frac{+2}{3}\right) \cdot 0,231 = +0,192, \quad C_A^u = +\frac{1}{2}, \quad (2.4)$$

$$C_V^d = -\frac{1}{2} - 2 \cdot \left(\frac{-1}{3}\right) \cdot 0,231 = -0,346, \quad C_A^d = -\frac{1}{2}. \quad (2.5)$$

Le couplage vectoriel du quark d est supérieur en valeur absolue à celui du quark u . Dans le calcul de la section efficace totale d'un processus similaire, le processus de Drell-Yan, on voit que ces couplages apparaissent au carré (voir annexe A). C'est pour cette raison que la contribution des quarks d est supérieure à celle des quarks u , malgré l'effet des fonctions de distribution partoniques à l'intérieur du proton.

Une preuve de (2.1), ainsi que l'application pour le processus de Drell-Yan, se trouvent à l'annexe A.

2.4 Étude des corrections électrofaibles

Pour notre analyse, nous avons utilisé une table de corrections électrofaibles fournie par l'un des auteurs des références [13] et [14]. Cette table donne le facteur K par lequel il faut corriger le nombre d'événements simulés en fonction de certaines caractéristiques cinématiques détaillées plus loin. K est défini comme :

$$K \equiv \frac{\sigma_{\text{NLO}}}{\sigma_{\text{LO}}} \equiv 1 + \delta_{\text{EW}}. \quad (2.6)$$

Les valeurs de K sont données selon le type de quarks (u , d ou b , en traitant les quarks s comme d et les quarks c comme u), et selon les variables de Mandelstam \hat{s} et \hat{t} calculées dans le référentiel du centre de masse des deux quarks ayant initié le processus.

Le calcul des corrections électrofaibles a été réalisé, pour le processus $q\bar{q} \rightarrow ZZ$, dans l'approximation où la quantité de mouvement transverse de jets supplémentaires est faible par rapport à celle de chaque boson Z . Dans ce cas, la cinématique du processus est complètement décrite par les deux variables \hat{s} et \hat{t} . L'approximation est valable pour la majorité des événements simulés par POWHEG, comme nous le discuterons plus loin. Le calcul des corrections électrofaibles au processus impliquant une boucle de quarks (fig. 1.3) n'est pas disponible. Nous négligeons cet effet car on s'attend à ce qu'il soit de l'ordre de α sur un processus qui ne contribue qu'à hauteur de 6%.

2.4.1 Implémentation des corrections électrofaibles

La variable \hat{s} représente le carré de l'énergie dans le centre de masse : lors d'une interaction du type $1 + 2 \rightarrow 3 + 4$, on a $\hat{s} = (p_1 + p_2)^2 = (p_3 + p_4)^2$, où p_1 et p_2 sont les quadrvecteurs énergie-impulsion des quarks, et p_3 et p_4 sont les quadrvecteurs énergie-impulsion des bosons Z . Dans notre cas, \hat{s} se calcule à partir de la mesure des quadrvecteurs des quatre leptons dans le détecteur : $\hat{s} = \left(\sum_{i=1}^4 p_i\right)^2$.

La variable \hat{t} se calcule, pour une interaction $1+2 \rightarrow 3+4$, comme : $\hat{t} = (p_1 - p_3)^2$ [21]. Dans notre cas, les particules 1 et 2 représentent la paire quark-antiquark et les particules 3 et 4 les bosons Z produits. Si on se place dans le référentiel du centre de masse et que l'on choisit l'axe des particules produites comme l'axe z , on peut sans perte de généralité exprimer les vecteurs énergie-impulsion des 4 particules comme :

$$p_{q_1}^* = \left(\frac{\sqrt{\hat{s}}}{2}, 0, \frac{\sqrt{\hat{s}}}{2} \sin \theta, \frac{\sqrt{\hat{s}}}{2} \cos \theta \right) \quad (2.7)$$

$$p_{q_2}^* = \left(\frac{\sqrt{\hat{s}}}{2}, 0, -\frac{\sqrt{\hat{s}}}{2} \sin \theta, -\frac{\sqrt{\hat{s}}}{2} \cos \theta \right) \quad (2.8)$$

$$p_{Z_1}^* = \left(\frac{\sqrt{\hat{s}}}{2}, 0, 0, \sqrt{\frac{\hat{s}}{4} - m_Z^2} \right) \quad (2.9)$$

$$p_{Z_2}^* = \left(\frac{\sqrt{\hat{s}}}{2}, 0, 0, -\sqrt{\frac{\hat{s}}{4} - m_Z^2} \right) \quad (2.10)$$

où m_Z est la masse au repos des bosons Z : $m_Z = 91,2$ GeV [6]. Notons que nous avons fait l'hypothèse que les bosons Z sont sur leur couche de masse. En effet, avec notre paramétrisation, on a $p_{Z_1}^{*2} = \frac{\hat{s}}{4} - \frac{\hat{s}}{4} + m_Z^2 = m_Z^2$. On obtient dès lors :

$$\hat{t} = (p_{q_1}^* - p_{Z_1}^*)^2 = p_{q_1}^{*2} + p_{Z_1}^{*2} - 2p_{q_1}^* \cdot p_{Z_1}^* \quad (2.11)$$

$$\simeq 0 + m_Z^2 - 2 \left(\frac{\hat{s}}{4} - \frac{\sqrt{\hat{s}}}{2} \cos \theta \sqrt{\frac{\hat{s}}{4} - m_Z^2} \right). \quad (2.12)$$

En effet, on peut négliger la masse du quark en première approximation. On a donc finalement :

$$\hat{t} = m_Z^2 - \frac{\hat{s}}{2} + \cos \theta \sqrt{\frac{\hat{s}^2}{4} - m_Z^2 \hat{s}}. \quad (2.13)$$

L'angle θ est déterminé comme l'angle entre le boson Z considéré et la direction des quarks incidents dans le référentiel du centre de masse des deux Z , autrement dit après avoir effectué la transformation de Lorentz inverse pour passer du système du laboratoire au système du centre de masse. Cependant, pour tenir compte de l'effet éventuel de radiation de gluons émis à petit angle (*cf* section 2.4.3), on prend comme direction des quarks incidents la direction de la différence des quantités de mouvement, normalisées à l'unité, entre les deux quarks, suivant la méthode proposée par [14] :

$$\cos \theta = \frac{\hat{p}_{q_1 b} - \hat{p}_{q_2 b}}{\left| \hat{p}_{q_1 b} - \hat{p}_{q_2 b} \right|} \cdot \hat{p}_{Z_1 b}, \quad (2.14)$$

où $\hat{p}_{q_i/Z_i b}$ représente le vecteur unitaire de quantité de mouvement du quark/boson Z numéro i après la transformation (boost) de Lorentz.

2.4.2 Effets des corrections électrofaibles au niveau généré

La figure 2.1 représente les distributions de masse invariante du système des deux bosons Z au niveau généré et de quantité de mouvement transverse d'un des

bosons Z choisi au hasard parmi les deux de l'événement, avec et sans corrections électrofaibles, sur base de simulations POWHEG d'événements $pp \rightarrow ZZ \rightarrow e^+e^-e^+e^-$, et la contribution relative des corrections électrofaibles pour ces mêmes processus. La forme est similaire pour les deux autres canaux $\mu^+\mu^-\mu^+\mu^-$ et $e^+e^-\mu^+\mu^-$. On constate que cette contribution est systématiquement négative, et qu'elle augmente avec la masse invariante. Les distributions de masse invariante M_{ZZ} et d'impulsion transverse des bosons $p_{T,Z}$ sont donc de bonnes observables pour tester la modélisation des corrections électrofaibles par comparaison avec les données.

On constate un bon accord entre la figures 2.1 et la figure 1.8, tirée de [13], malgré quelques différences mineures attribuées aux incertitudes théoriques à la base de la génération des échantillons d'événements. Le fait que cet accord soit correct permet de nous assurer que les tables de corrections électrofaibles utilisées correspondent bien aux valeurs calculées. La comparaison permet aussi de valider notre implémentation de ces corrections à partir des tables et du calcul de \hat{s} et \hat{t} des événements simulés.

La figure 2.2 présente la distribution du facteur de correction K (cf éq. 2.6) en fonction de la variable cinématique $\sqrt{\hat{s}}$, pour des quarks u , d et b . On constate ainsi que les facteurs de correction fluctuent davantage à de grandes valeurs de $\sqrt{\hat{s}}$, et que ces corrections sont en moyenne plus importantes pour les quarks u que pour les quarks d .

Les corrections sont de l'ordre de -4% à $M_{ZZ} = 200$ GeV, et descendent au-delà de -10% à partir de $M_{ZZ} = 500$ GeV. On peut avancer la raison suivante pour expliquer pourquoi ces corrections sont si importantes, alors qu'elles sont d'ordre 1 ou 2 en $\alpha \sim \frac{1}{137}$: lorsque \hat{s} augmente, la virtualité du quark échangé augmente également. Cela peut créer un pôle (résonance) dans la correction au vertex $Zq\bar{q}$ par une boucle de boson vecteur, ce qui augmente fortement l'espace des phases disponible, et donc l'importance des corrections.

Ce comportement est similaire dans la variable $p_{T,Z}$ et est même plus marqué à des hautes valeurs de cette variable, descendant jusqu'à -25% à plus de 350 GeV de quantité de mouvement transverse. Cependant, il convient de remarquer que la fraction d'événements attendue à ces valeurs élevées de quantité de mouvement transverse est très petite.

Notons que les corrections ne peuvent affecter que les particules dans l'état initial. En effet, il n'existe pas, dans le Modèle Standard, de vertex ZZZ , ZZW ou $ZZ\gamma$. Il est donc impossible d'avoir une émission réelle dans l'état final, contrairement au processus $q\bar{q} \rightarrow WW$ par exemple.

2.4.3 Restriction due à l'activité hadronique

Jusqu'ici, nous avons fait la supposition que les corrections électrofaibles et QCD se factorisaient, c'est-à-dire qu'elles pouvaient se calculer indépendamment. Cette supposition n'est pas absolument valable [14], mais il est pour l'instant impossible de calculer la production de paires de bosons Z à l'ordre $\alpha_S\alpha$. La majorité de ces processus n'implique que des radiations de gluons de faible impulsion ou colinéaires dans l'état initial. Dans ce cas, l'état initial reste décrit par les variables \hat{s} et \hat{t} calculées à partir des bosons Z observés et on peut tout de même appliquer les corrections

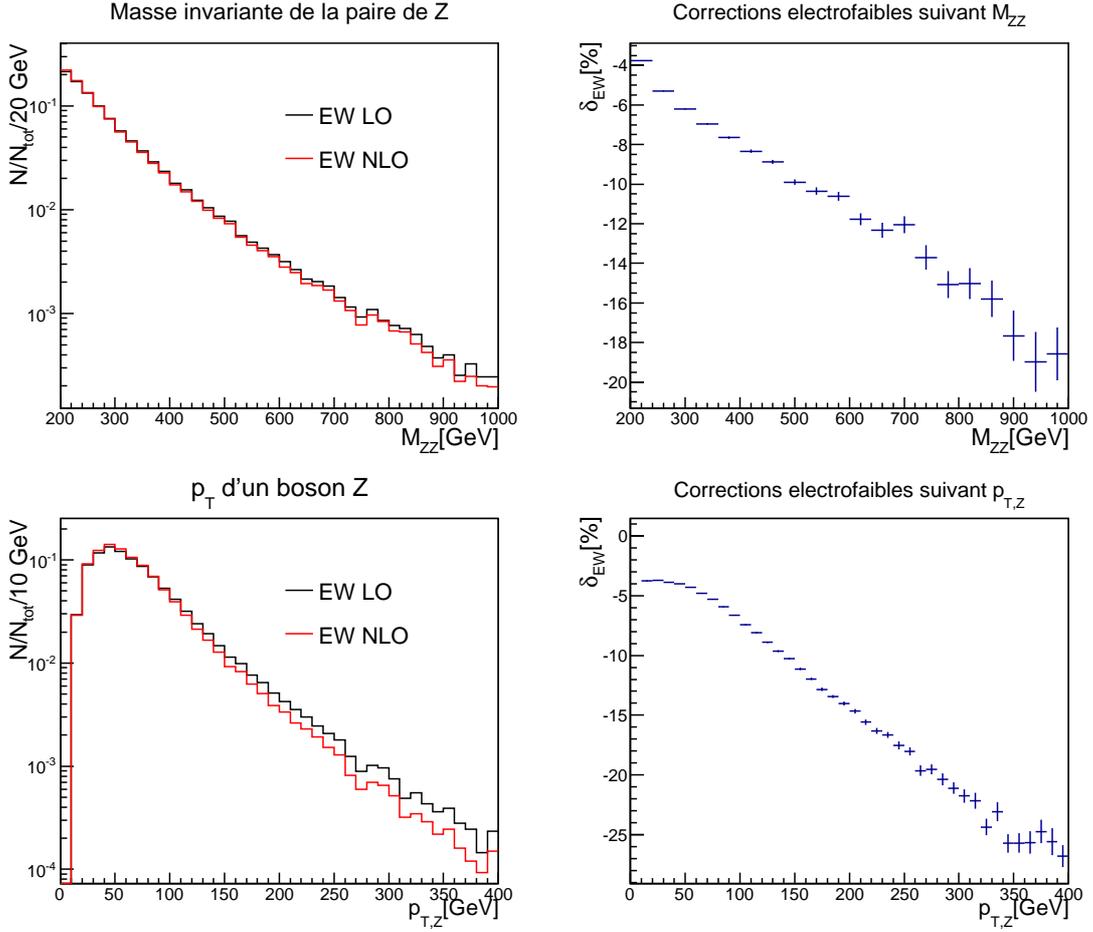


FIGURE 2.1 – Corrections électrofaibles pour des simulations POWHEG $pp \rightarrow ZZ \rightarrow e^+e^-e^+e^-$. À gauche, comparaison des distributions de masse invariante du système des 2 bosons Z , M_{ZZ} (en haut) et de la quantité de mouvement transverse d'un des bosons Z , $p_{T,Z}$ (en bas), au niveau généré, avec et sans corrections électrofaibles. À droite, corrections électrofaibles en fonction de M_{ZZ} (en haut) et de $p_{T,Z}$ (en bas).

électrofaibles. Dans le cas de radiations dures colinéaires au système des quarks, on peut suivre un raisonnement similaire et garder les corrections électrofaibles. Des problèmes surviennent cependant lorsque l'on considère des radiations de gluons à grande quantité de mouvement transverse, qui invalident l'hypothèse à la base du calcul des corrections électrofaibles selon laquelle la cinématique est entièrement déterminée par les quantités de mouvement des bosons Z produits avant l'application des corrections.

La méthode préconisée par [14] est d'établir une coupure afin de supprimer la majorité des éléments de ce type : on définit

$$\rho \equiv \frac{\left| \sum_{i=1}^4 \vec{p}_i^T \right|}{\sum_{i=1}^4 |\vec{p}_i^T|}, \quad (2.15)$$

où \vec{p}_i^T représente la quantité de mouvement transverse, donc dans le plan (x, y) dans

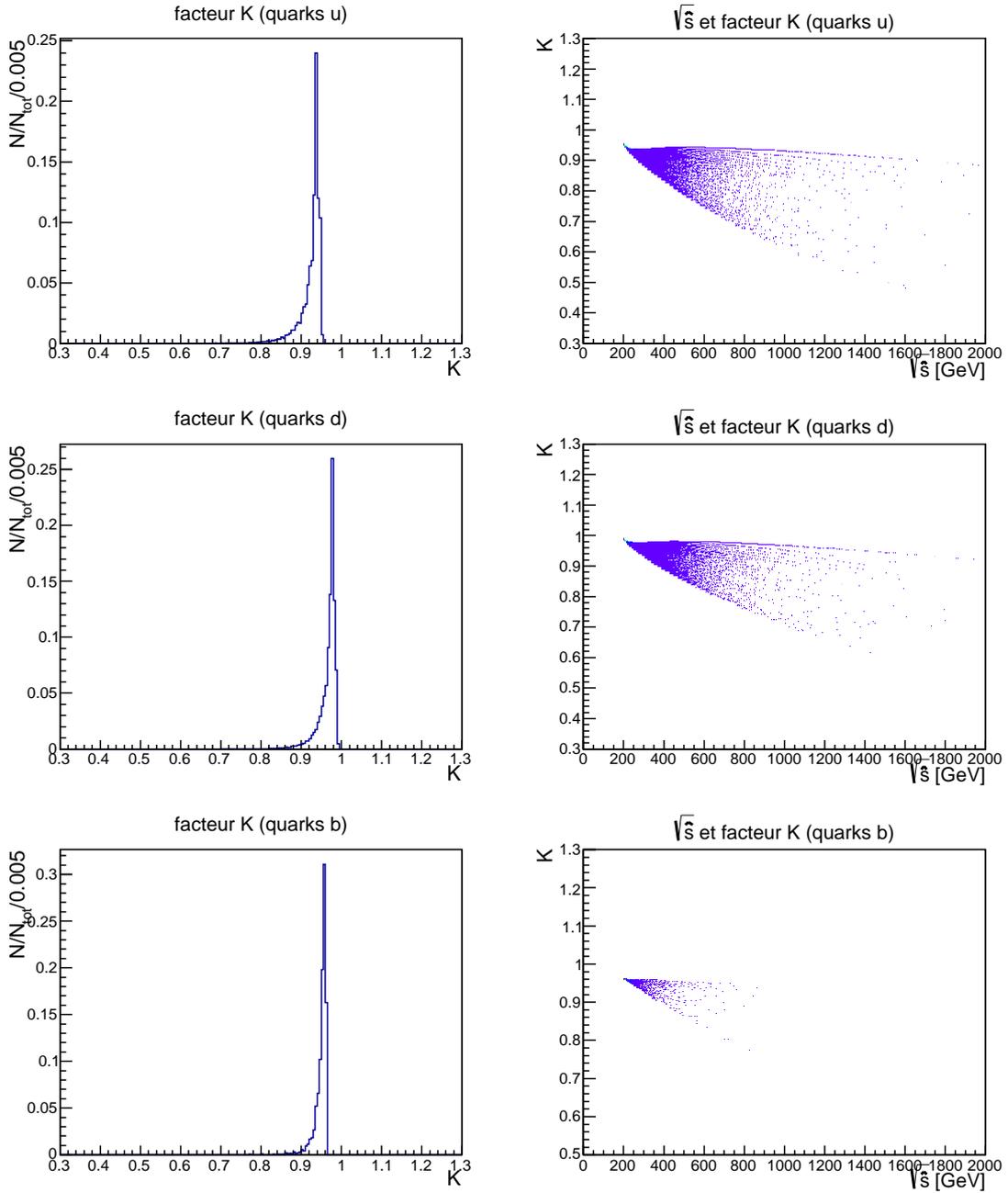


FIGURE 2.2 – Distributions de K , totale (colonne de gauche) et en fonction de \sqrt{s} (colonne de droite), selon le type de quarks qui ont initié l'événement (de haut en bas : quarks u , d et b). Ces graphiques ont été produits à partir de simulations POWHEG $pp \rightarrow ZZ \rightarrow e^+e^-e^+e^-$.

le référentiel du laboratoire, du lepton i dans l'état final et est donc un vecteur à deux composantes. La prescription établie par [14] est de rejeter tous les événements où $\rho > 0,3$. La distribution de cette variable au niveau généré se trouve en figure 2.3. On y constate que la coupure appliquée garde la majorité des événements : on rejette environ 25% de notre échantillon de départ.

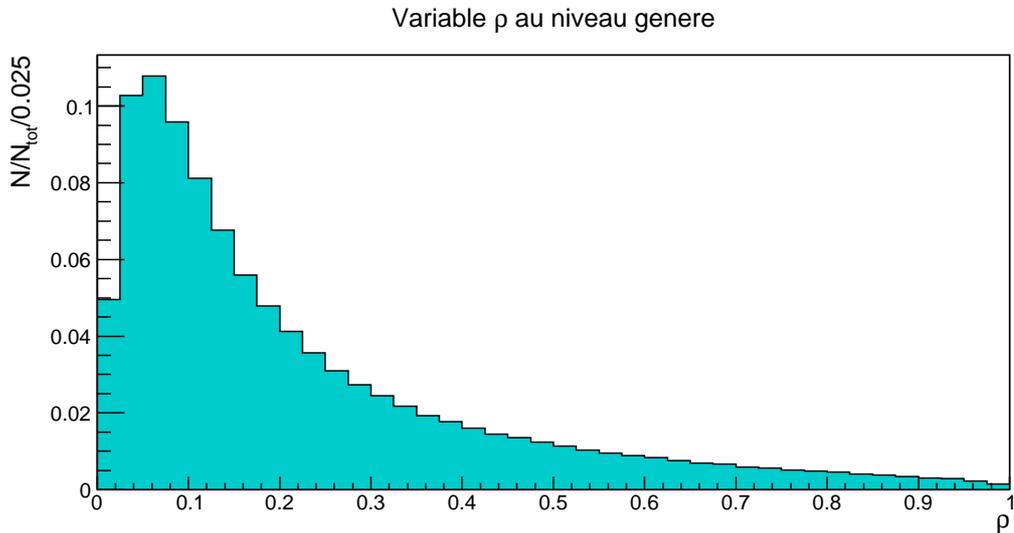


FIGURE 2.3 – Distribution de la variable ρ au niveau généré, pour des simulations POWHEG $pp \rightarrow ZZ \rightarrow e^+e^-e^+e^-$.

Chapitre 3

L'expérience CMS du LHC

Dans ce chapitre, nous présentons l'expérience d'où proviennent nos données : l'expérience CMS installée au LHC. Nous commençons par décrire brièvement le collisionneur LHC et ses caractéristiques principales. Ensuite, nous passons en revue les différents composants de CMS et expliquons plus en détail comment les phénomènes physiques qui nous intéressent sont détectés et enregistrés. Enfin, nous expliquons l'origine des échantillons de données et de simulations que nous utiliserons dans la suite de l'analyse.

3.1 Le collisionneur LHC

Le LHC, pour Large Hadron Collider, est un collisionneur de protons et d'ions lourds à haute énergie situé au CERN (Centre Européen pour la Recherche Nucléaire), près de Genève. Différentes expériences y sont installées pour collecter les données issues des collisions : les plus importantes sont CMS (pour Compact Muon Solenoid) et ATLAS (pour A Toroidal Lhc ApparatuS), ainsi que ALICE et LHCb.

Le LHC est composé de deux anneaux accélérateurs, dans lesquels les faisceaux circulent en des sens opposés. Ces deux anneaux ne se croisent qu'en quatre endroits, autour desquels sont construits les détecteurs.

3.1.1 La chaîne d'injecteurs

L'accélération des particules utilisées dans les collisions au LHC se fait en plusieurs étapes, à l'aide de plusieurs accélérateurs (appelés injecteurs) qui augmentent chacun l'énergie des particules. Le premier étage d'accélération se fait dans un accélérateur linéaire (ou LINAC). Les étapes suivantes sont effectuées par différents accélérateurs circulaires, représentés à la figure 3.1. Comme on peut le voir, différentes expériences ne nécessitant pas des faisceaux de trop grande énergie sont installées sur les injecteurs du LHC.

3.1.2 Caractéristiques du LHC

Les caractéristiques d'un collisionneur peuvent se résumer à deux grandeurs : l'énergie des collisions et la luminosité.

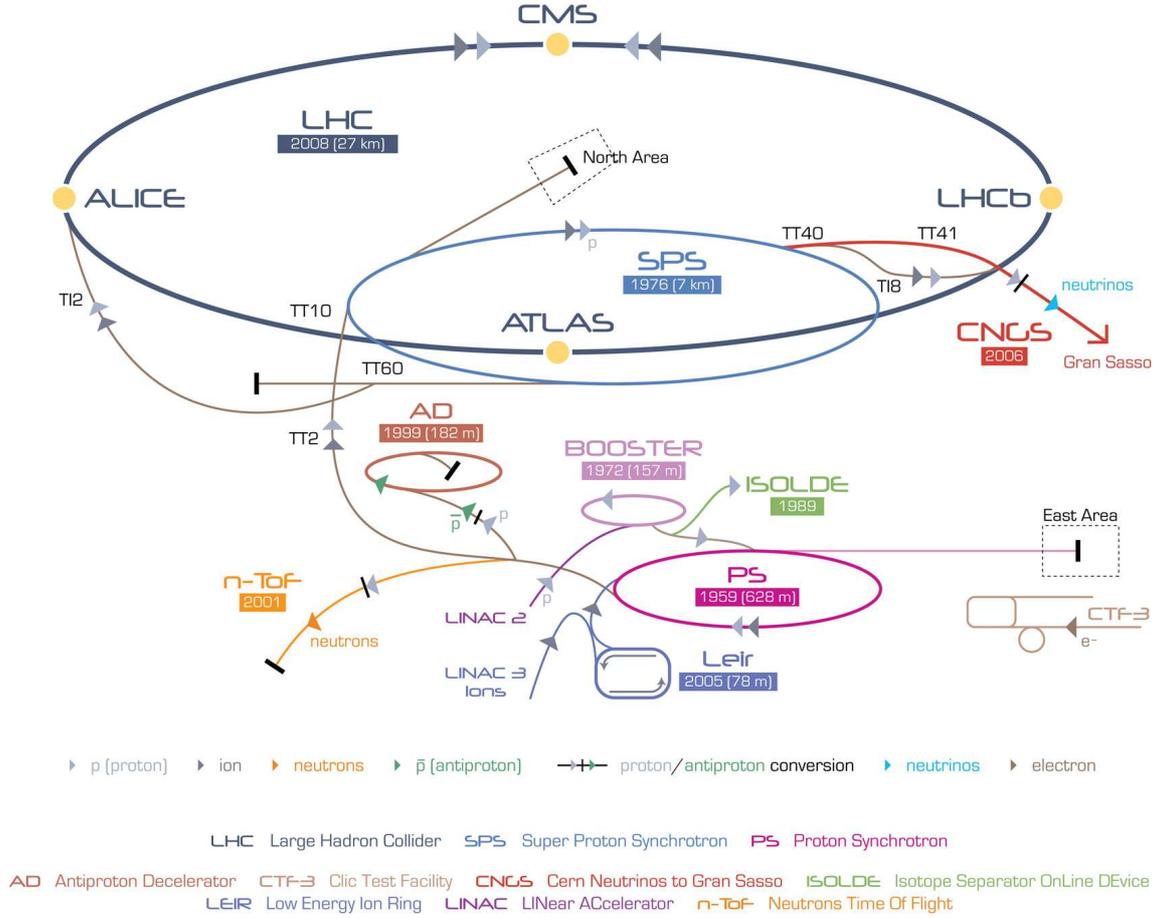


FIGURE 3.1 – Le complexe d’accélérateurs du CERN et les principales expériences. Notons que le schéma n’est pas à l’échelle. [22]

L’énergie des collisions était de 8 TeV dans le référentiel du centre de masse (c’est-à-dire 4 TeV par faisceau) lors de l’année 2012. Lors du run 2, commencé cette année, les collisions se feront à 13 TeV. Il est important de préciser que seule une petite fraction de cette énergie est effectivement utilisée lors du processus d’interaction, comme expliqué à la section 1.1.

La luminosité \mathcal{L} se définit comme [23] :

$$\frac{dN}{dt} = \mathcal{L}\sigma, \quad (3.1)$$

où $\frac{dN}{dt}$ est le nombre d’interactions par unité de temps, et σ est la section efficace d’interaction entre les particules du faisceau. La luminosité est plus élevée si les faisceaux sont de forte intensité et se croisent dans une région la plus ponctuelle possible. Ce qui nous intéresse est la luminosité intégrée $\int dt \mathcal{L}$, où l’intégrale est prise sur toute la période de prise de données. De cette manière, connaissant la section efficace du processus considéré, on peut savoir combien d’événements sont attendus pour ce processus au total durant la durée de l’expérience :

$$N = \sigma \cdot \int dt \mathcal{L}. \quad (3.2)$$

La luminosité intégrée est donc un paramètre essentiel pour améliorer la précision statistique des mesures à CMS. Comme on peut le voir à la figure 3.2, la luminosité intégrée fournie par le LHC à l'expérience CMS pendant l'année 2012 est de $23,3 \text{ fb}^{-1}$, mais la luminosité intégrée des données collectées par le détecteur CMS dans la même année n'est que de $21,8 \text{ fb}^{-1}$ car le détecteur n'est pas toujours opérationnel pendant le fonctionnement de l'accélérateur. De plus, toutes les périodes de prise de données n'ont pas été gardées pour notre analyse à cause d'un fonctionnement temporairement imparfait d'une partie ou l'autre du détecteur. En pratique, la luminosité intégrée utilisée pendant l'année 2012 est de $19,6 \text{ fb}^{-1}$.

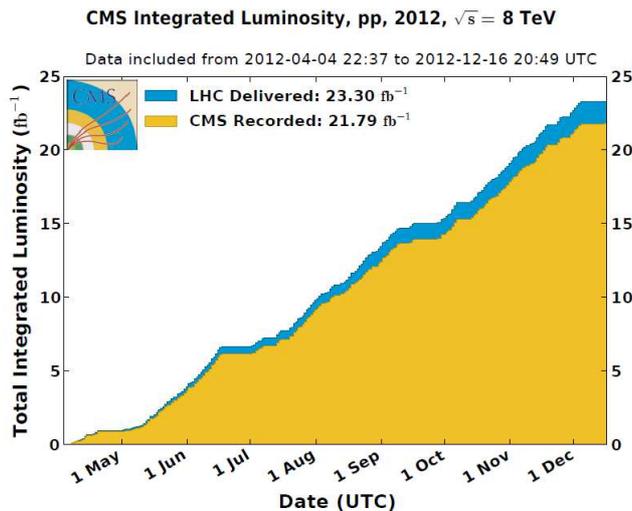


FIGURE 3.2 – Évolution de la luminosité intégrée du LHC et de celle reçue par le détecteur CMS pendant l'année 2012 [24].

3.2 L'expérience CMS

Le détecteur Compact Muon Solenoid [25] est un détecteur cylindrique entourant le point d'interaction des faisceaux. Sa particularité est que le champ magnétique, indispensable pour mesurer la quantité de mouvement des particules chargées, est produit par un grand solénoïde entourant les calorimètres, afin d'optimiser la résolution en énergie de ceux-ci. On distingue deux régions dans le détecteur CMS : le tonneau, de forme cylindrique et d'axe parallèle aux faisceaux, et les bouchons, qui "ferment" le cylindre et servent à mesurer les particules émises dans des directions proches de celle du faisceau. Un schéma du détecteur se trouve en figure 3.3.

3.2.1 Système de coordonnées

On utilise les conventions suivantes pour décrire une position dans le détecteur CMS :

- l'axe z est défini comme l'axe du faisceau. Le plan perpendiculaire à l'axe du faisceau est appelé plan transverse. La séparation de la variable de quantité

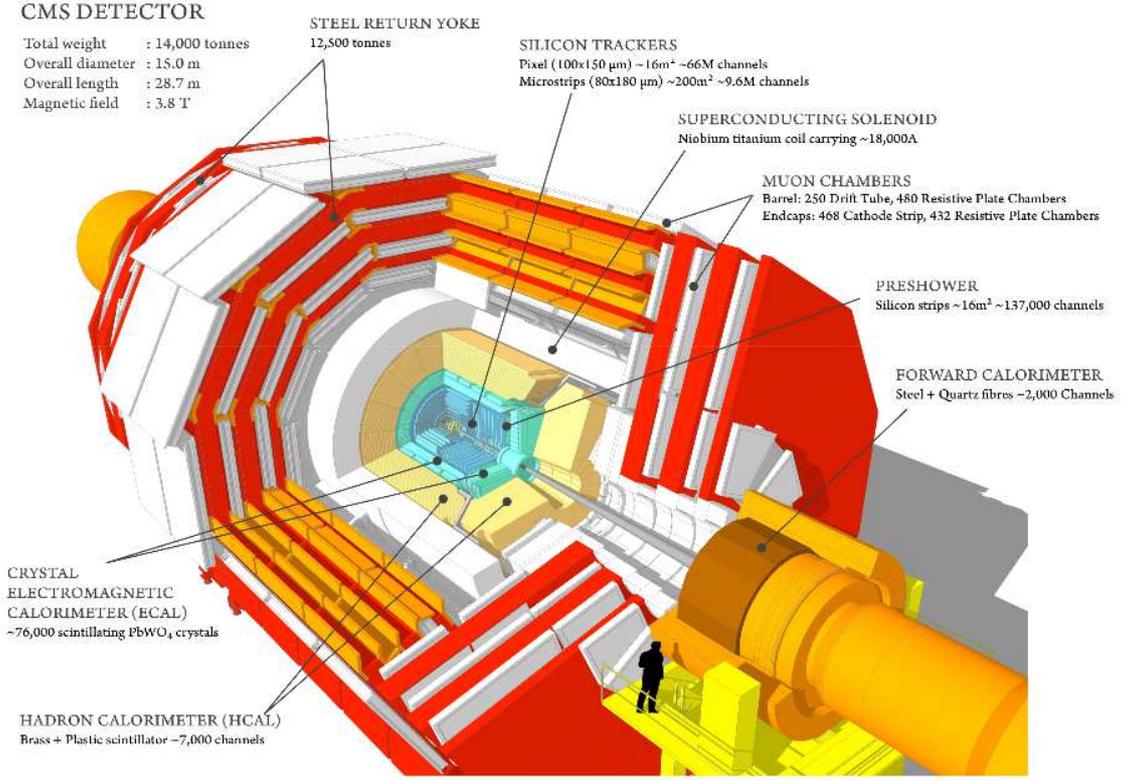


FIGURE 3.3 – Le détecteur CMS et ses différents composants [26].

de mouvement en sa composante parallèle et sa composante transverse est importante : en effet, on ne peut pas faire de bilan en énergie ou en quantité de mouvement sur l'axe z car on ne détecte pas les débris des protons emportés par le tube à vide, où circulent les faisceaux. Les lois de conservation doivent en revanche s'appliquer dans le plan transverse ;

- l'angle θ est défini comme l'angle polaire par rapport à l'axe z ;
- l'angle φ sert à désigner les coordonnées dans le plan transverse. Notons que la géométrie du détecteur est approximativement symétrique en φ .

On définit également la *pseudorapidité* η d'après θ par :

$$\eta \equiv -\ln \left(\tan \frac{\theta}{2} \right). \quad (3.3)$$

La mesure de η permet de déterminer la direction d'une particule par rapport à l'axe du faisceau : lorsque $\eta = 0$, la particule se dirige à la verticale du point d'interaction, et on se rapproche de l'axe du faisceau lorsque $|\eta|$ augmente, l'axe lui-même correspondant à $|\eta| = \infty$. Remarquons que la pseudorapidité est l'équivalent de la *rapidité* $y = \frac{1}{2} \ln \left(\frac{E+p_z}{E-p_z} \right)$ pour des particules ultra-relativistes. L'intérêt de se servir de cette variable est que la différence de rapidité entre deux particules est une grandeur invariante selon un boost suivant l'axe z .

3.2.2 Composants du détecteur CMS

CMS est composé de plusieurs parties, réparties en couches cylindriques (*cf* fig. 3.3). Chaque partie a sa fonction, et c'est en reliant les informations reçues par chaque partie que l'on parvient à identifier les particules ainsi qu'à mesurer leur énergie et leur impulsion. Nous décrivons ici succinctement les différentes parties du détecteur, en partant de la zone de croisement des faisceaux et en remontant vers l'extérieur.

Les détecteurs de traces et de vertex

Le détecteur de traces, ou traceur, représenté en bleu sur la figure 3.3, a pour but de reconstruire la trace des particules chargées près du point d'interaction. La présence du champ magnétique dû au solénoïde courbe les trajectoires et permet ainsi de mesurer la quantité de mouvement transverse des particules d'après

$$p_T[\text{GeV}/c] = 0,3B[\text{T}]R[\text{m}], \quad (3.4)$$

où R est le rayon de courbure de la trajectoire projetée dans le plan transverse et est calculé à partir de la mesure de la flèche (s) de la trajectoire reconstruite, et B est le champ magnétique parallèle à l'axe du faisceau. La mesure de la quantité de mouvement totale, qui est équivalente à l'énergie pour les particules ultra-relativistes que l'on considère, se fait en mesurant θ :

$$p = \frac{p_T}{\sin \theta}. \quad (3.5)$$

On peut montrer [27] que l'erreur relative sur la mesure de la quantité de mouvement dans le détecteur de traces est de la forme :

$$\left| \frac{\sigma_p}{p} \right| \simeq \sqrt{\frac{720}{n+4} \left(\frac{p_T}{0,3BL_T^2} \sigma_x \right)^2 + \left(\frac{0,045 \sin \theta}{\beta BL_T} \right)^2 \frac{x}{X_0}}. \quad (3.6)$$

Dans la formule (3.6), n est le nombre de couches de détection dont on dispose pour reconstruire la trajectoire, et vaut environ 10 pour CMS. L_T est le rayon du détecteur de traces, et vaut 1 m à CMS, sauf pour les particules émises à grand $|\eta|$ dans la région des bouchons, auquel cas cette valeur diminue. σ_x est la résolution de la mesure de la coordonnée transverse sur chaque couche, soit environ 30 μm . B est le champ magnétique dans le traceur et vaut uniformément 3,8 T. La grandeur $\beta = v/c$ vaut presque 1 pour les particules ultra-relativistes que l'on regarde. Le rapport x/X_0 représente l'épaisseur de matière traversée, en nombres de longueurs de radiation du milieu : il s'agit donc d'un indicateur de la quantité de matière traversée par la particule chargée. Ce rapport doit être décrit par une simulation détaillée de la géométrie du détecteur : il prend en compte non seulement les composants actifs de détection, mais aussi la structure, le câblage et les circuits de refroidissement.

La figure 3.4 montre la précision de la mesure de la quantité de mouvement transverse des muons dans CMS, en fonction de p_T , pour deux intervalles en pseudorapacité η . La performance du traceur central est représentée par les marqueurs vides. On reconnaît les deux contributions à la résolution relative de la formule illustrées dans la formule (3.6) : un terme constant à bas p_T , dû à la diffusion multiple,

de l'ordre du pourcent, et un terme linéaire en p_T dû à la précision de la mesure de la flèche dans le traceur, de l'ordre de 1 % par 100 GeV. La précision relative sur la mesure de la quantité de mouvement dans le traceur se détériore donc pour des particules de haute quantité de mouvement transverse.

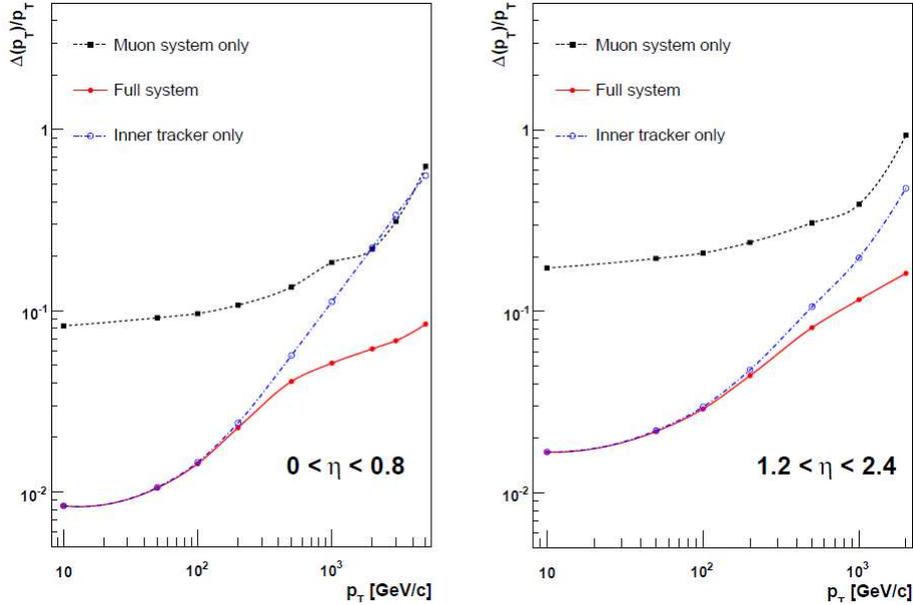


FIGURE 3.4 – Erreur relative sur la mesure de p_T dans les détecteurs de traces (en bleu), dans les chambres à muons (en noir), et pour la mesure combinée des deux détecteurs (en rouge) [25].

Le détecteur de traces de CMS se compose de semi-conducteurs au silicium. Ces traceurs ont une excellente résolution spatiale (de l'ordre de la dizaine de μm) et une assez bonne résolution temporelle pour pouvoir être utilisés à une fréquence très élevée de croisement de faisceaux (20 MHz).

Le traceur de CMS se compose de deux parties. Le détecteur de vertex, très proche du point d'interaction (de 4,4 à 10,2 cm), est constitué d'un détecteur à pixels, qui offre la précision spatiale nécessaire pour distinguer des vertex de désintégration de particules de court temps de vie. Ensuite, jusqu'à un rayon de 1,1 m, ce sont des détecteurs à pistes qui sont utilisés. Ces détecteurs à pistes se composent de 10 couches de détection sur tout l'espace géométrique $|\eta| < 2,5$.

Le calorimètre électromagnétique

Le calorimètre électromagnétique, en turquoise sur la figure 3.3, a pour but de mesurer l'énergie des électrons, qui interagissent avec la matière en émettant du rayonnement de freinage (ou *bremstrahlung*), et des photons, qui interagissent par le processus de production de paires e^+e^- .

Les électrons et photons pénétrant dans le calorimètre forment une cascade d'interactions : un électron radie un photon, qui crée une paire e^+e^- , qui vont pouvoir à leur tour radier un photon, *etc.* À chaque étape de radiation/création de paires, les

particules produites possèdent moins d'énergie. La cascade continue de se développer jusqu'à ce que les électrons atteignent une énergie critique, E_c , sous laquelle ils interagissent plutôt par ionisation. L'énergie est mesurée par la désexcitation du matériau actif du calorimètre, qui émet des photons de scintillation après avoir été excité par les électrons de la gerbe : le nombre de photons émis est proportionnel à l'énergie de la particule de départ.

L'incertitude associée à la mesure de l'énergie dans un calorimètre se modélise en 3 termes :

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} + C + \frac{N}{E}. \quad (3.7)$$

Le terme S/\sqrt{E} est un terme stochastique : il provient de la fluctuation statistique sur le nombre de particules secondaires chargées n , à partir duquel on détermine l'énergie. Comme l'incertitude sur n est poissonnienne, celle sur l'énergie l'est aussi. Le terme C est un terme constant, qui provient principalement de l'erreur sur la calibration, ou du facteur de proportionnalité entre n et E . Le terme N/E est indépendant de l'énergie de la particule : il s'agit d'un terme de bruit. Pour le calorimètre électromagnétique de CMS, $S = 0,028 \text{ GeV}^{\frac{1}{2}}$, $N = 0,12 \text{ GeV}$ et $C = 0,03$. La résolution en énergie des calorimètres électromagnétiques s'améliore pour des grandes valeurs d'énergie, et est mauvaise à des valeurs faibles, comme illustré à la figure 3.5. Ce comportement est inverse de celui des détecteurs de traces.

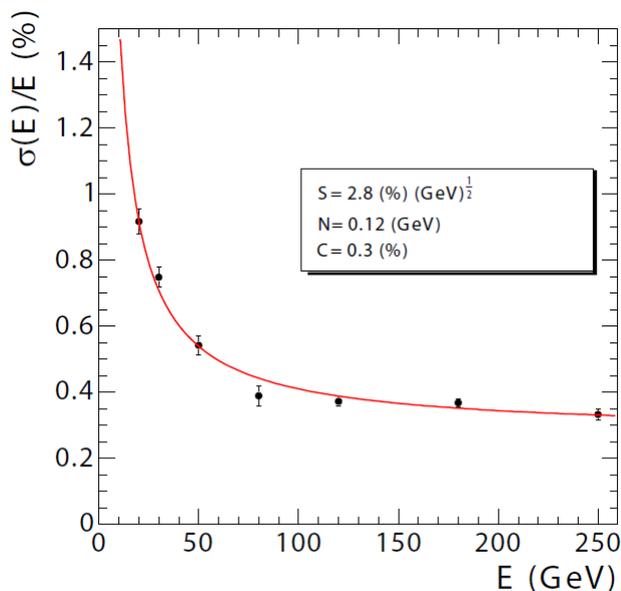


FIGURE 3.5 – Incertitude relative sur la mesure de E dans le calorimètre électromagnétique de CMS [25].

Les calorimètres électromagnétiques de CMS sont composés de cristaux scintillants de tungstate de plomb (PbWO_4), qui servent à la fois d'absorbeurs (qui créent les interactions) et de scintillateurs (qui émettent les photons de scintillation). L'avantage de ce matériau est sa longueur de radiation assez courte (0,89 cm), ce qui permet un calorimètre compact. En effet, il faut environ 25 longueurs de radiation pour contenir entièrement une gerbe électromagnétique d'une centaine de GeV.

Le calorimètre hadronique

Le calorimètre hadronique, en jaune sur la figure 3.3, mesure l'énergie des hadrons chargés et neutres. Le principe est le même que pour le calorimètre électromagnétique, sauf que les particules interagissent ici par des interactions fortes avec les composants du noyau du matériau absorbeur. La cascade hadronique se développe par conséquent jusqu'à une énergie seuil de l'ordre de deux fois la masse du pion. Les gerbes hadroniques comportent également une composante électromagnétique due aux désintégrations de π^0 .

La mesure de l'énergie d'un hadron est beaucoup moins précise que celle d'un électron ou d'un photon dans un calorimètre. Ceci a plusieurs raisons. Premièrement, la réponse du détecteur pour la composante électromagnétique de la cascade est différente de celle pour les autres hadrons, alors que la proportion de π^0 varie d'une gerbe à l'autre. Deuxièmement, une partie de l'énergie n'est pas mesurée car elle est absorbée dans les énergies de liaison des noyaux ou dans des particules qui échappent à la détection. Enfin, le terme stochastique est plus important car l'énergie seuil de la réaction est plus élevée que pour les calorimètres électromagnétiques.

Le calorimètre hadronique de CMS est constitué de couches successives d'absorbeurs (constitués d'un alliage Cu-Zn) et de scintillateurs en plastique. Sa profondeur est beaucoup plus grande que celle du calorimètre électromagnétique car la longueur d'interaction hadronique est plus grande. C'est aussi pour cette raison qu'il est placé derrière le calorimètre électromagnétique : dans le cas contraire, les électrons et photons auraient déjà déposé toute leur énergie avant d'arriver au calorimètre électromagnétique.

Le solénoïde

L'aimant supraconducteur solénoïdal de CMS, en blanc dans la figure 3.3 et entourant le calorimètre hadronique, fournit un champ magnétique uniforme de 3,8 T à l'intérieur du détecteur, c'est-à-dire dans le traceur et dans les calorimètres. Le champ de retour de l'aimant, en rouge dans la figure 3.3, sert également à courber la trajectoire dans l'autre sens, dans les chambres à muons situées à l'extérieur. Le champ n'est alors plus uniforme et prend une structure complexe.

L'aimant de CMS est placé derrière les calorimètres afin d'éviter des interactions de particules avec le matériau du solénoïde.

Les chambres à muons

À l'extérieur du solénoïde se trouve un autre détecteur de traces : il s'agit des chambres à muons, représentées en blanc sur la figure 3.3. En effet, les muons sont les seules particules détectables par un traceur à franchir les calorimètres et le solénoïde sans y perdre toute leur énergie. La résolution en quantité de mouvement transverse du détecteur à muons est nettement moins bonne que pour les détecteurs de traces proches du point d'interaction, comme on peut le constater à la figure 3.4. Les chambres à muons ont trois rôles : identifier et reconstruire les muons, améliorer la précision sur la mesure des muons de haute quantité de mouvement transverse, et contribuer au système de déclenchement (voir point suivant).

Comme les flux de particules à travers ces détecteurs sont beaucoup moins importants que dans les détecteurs de traces, d'autres technologies peuvent y être employées. On distingue trois types de détecteurs à muons dans CMS, tous étant des détecteurs à gaz : tubes à dérive, chambres proportionnelles à pistes et RPC (pour *Resistive Plate Chambers*).

Système de déclenchement et prise de données

Chaque événement à CMS consiste en des données lues par plusieurs millions de canaux de détection. Si on tient en compte la fréquence de croisement des faisceaux, de 20 MHz en 2012, on arrive à plusieurs centaines de TB de données par seconde. Cette quantité est bien trop grande pour pouvoir être stockée et analysée. Par conséquent, il est nécessaire d'appliquer une sélection rapide, dès la prise de données, par un système de déclenchement conçu pour n'enregistrer que des événements potentiellement intéressants. Ce système de déclenchement, ou *trigger*, se déroule en plusieurs niveaux :

- le premier niveau de déclenchement se fait au niveau du câblage, par des modules logiques de type et/ou des tables préprogrammées, et doit être suffisamment rapide pour tenir la cadence élevée du croisement de faisceau. Pour ces raisons, il utilise uniquement les données des calorimètres et des chambres à muons. En effet, le détecteur de traces demande trop de temps de calcul pour appliquer les algorithmes de reconstruction de traces et identifier les particules. Les événements présentant des candidats muons, électrons, photons, et jets de grande impulsion transverse (de l'ordre de la dizaine de GeV) sont gardés. Le premier niveau de trigger fait passer la fréquence de collecte d'événements de 20 MHz à 100 kHz ;
- le déclenchement à haut niveau (HLT, ou High Level Trigger) se base sur une reconstruction des événements dans une ferme de processeurs commerciaux. Le même type d'algorithmes que ceux qui sont appliqués pour la reconstruction détaillée est utilisé. Une sélection plus stricte des candidats détectés au niveau 1 est opérée, pour réduire le taux d'événements de 100 kHz à quelques centaines de Hz.

Chaque événement à CMS doit être reconstruit à partir des différents canaux de lecture. Ceci est fait juste après le premier niveau de trigger, par l'*event builder*. Cette étape demande beaucoup de précision au niveau de la synchronisation temporelle des différents canaux de l'électronique de lecture.

3.2.3 Reconstruction et identification des particules

Les particules sont reconstruites par la combinaison des données des différentes parties du détecteur. Chaque particule ayant un temps de vie suffisant pour avoir un parcours résoluble dans le détecteur a sa signature propre, comme illustré à la figure 3.6. Dans le cadre de ce mémoire, on s'intéresse particulièrement aux électrons et aux muons, qui composent les particules dans l'état final étudié, mais aussi aux jets, toujours présents dans des collisions de hadrons.

Les électrons sont reconstruits par la combinaison d'une trace dans le détecteur de traces et d'un dépôt d'énergie dans le calorimètre électromagnétique. En effet, les

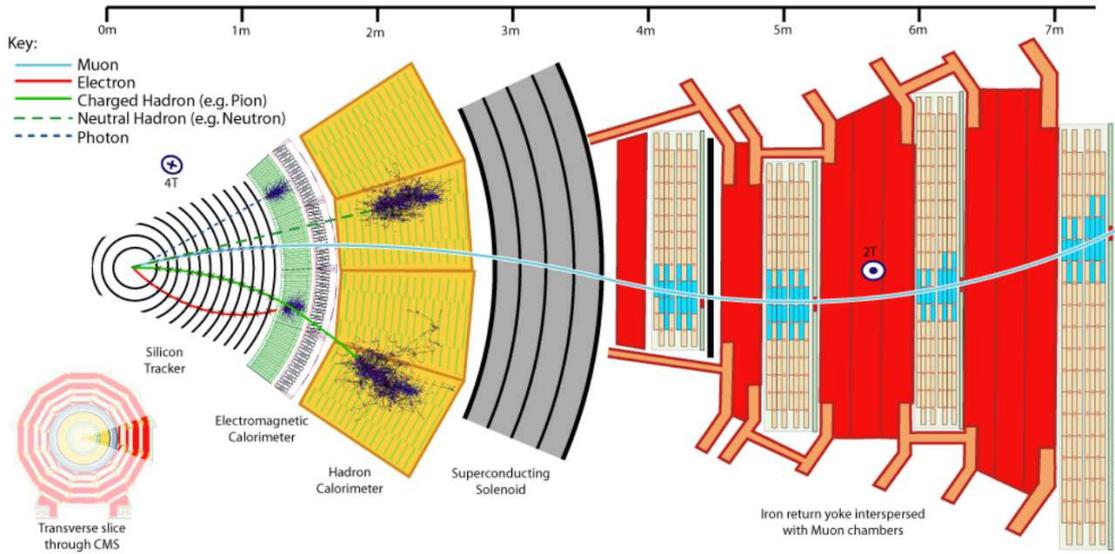


FIGURE 3.6 – Secteur d’une coupe transversale du détecteur CMS. On peut voir les signatures de différents types de particules dans le détecteur : muon, électron, hadron chargé, hadron neutre, photon [28].

électrons déposent complètement leur énergie dans ce calorimètre et ne passent pas plus loin dans le détecteur. La mesure de l’énergie des électrons se fait principalement dans le calorimètre électromagnétique.

Les muons sont reconstruits par la combinaison d’une trace dans le détecteur de traces et dans les chambres à muons. Comme ce sont les seules particules à passer par cette dernière partie du détecteur, cette signature est très nette. La mesure de la quantité de mouvement des muons se fait principalement dans le détecteur de traces.

Les jets, initiés par des quarks ou des gluons, se reconnaissent à la présence de multiples traces dans le traceur, ainsi que par des dépôts d’énergie dans les calorimètres électromagnétique et hadronique. Ces traces et ces dépôts d’énergie sont collimatés, et un algorithme les regroupe en jets, de façon à estimer au mieux l’énergie et la quantité de mouvement des quarks ou gluons initiaux.

Même si les signatures des électrons, des muons et des jets sont différentes, il arrive qu’un jet soit identifié à tort comme un électron ou un muon. C’est cette possibilité d’obtenir des faux électrons ou muons dans nos données qui causera la présence d’un bruit de fond instrumental pour notre analyse, comme on le verra au chapitre 5. Le reste de l’activité hadronique, composée de l’événement sous-jacent ou des interactions des débris du proton ayant initié l’interaction dure, perturbe également la sélection de leptons isolés provenant de la désintégration de bosons Z de notre processus. Les interactions induites par les autres protons lors du même croisement de faisceaux (ou *pile-up*) doivent être prises en compte, mais ne causent que très rarement des faux leptons de p_T suffisante.

3.3 Échantillons de données et de simulations

3.3.1 Échantillons de données

Les échantillons de données de CMS utilisés ont été construits selon différents critères, basés sur le déclenchement de certains triggers. Pour notre analyse, on choisit des échantillons demandant la présence de deux leptons reconstruits au moins. On compte trois échantillons de ce type : les données `DoubleElectron`, `DoubleMuon` et `ElectronMuon`. Ces échantillons requièrent la présence de leptons ayant passé différents critères. En particulier, ils imposent la présence d'un lepton d'une quantité de mouvement transverse $p_T > 17$ GeV, et d'un autre lepton de $p_T > 8$ GeV. Cette catégorisation des événements fait qu'il est probable qu'un même événement contenant 4 leptons soit sélectionné par plusieurs triggers différents, et se retrouve donc dans plusieurs échantillons. Il faudra donc faire attention à éviter le double comptage dans notre analyse finale.

3.3.2 Échantillons de simulations

Afin de comparer nos données à un modèle, nous avons besoin de plusieurs échantillons de simulations. Les échantillons de signal ($pp \rightarrow ZZ \rightarrow l^+l^-l'^+l'^-$) ont déjà été décrits en section 2.1. Afin de vérifier nos coupures, on a utilisé des simulations pour les processus pouvant éventuellement contribuer au bruit de fond. Ces simulations sont reprises dans la table 3.1, avec la section efficace simulée, le nombre d'événements dans l'échantillon et la pondération apportée à chaque événement simulé pour normaliser le nombre d'événements à une luminosité intégrée de $19,6 \text{ fb}^{-1}$.

Processus	Générateur	σ [fb]	Nombre d'événements	Poids pour $\int dt \mathcal{L} = 19,6 \text{ fb}^{-1}$
$ZZ \rightarrow 4e$	POWHEG	76, 91	1499093	0, 0010
$ZZ \rightarrow 4\mu$	POWHEG	76, 91	1499064	0, 0010
$ZZ \rightarrow 2e2\mu$	POWHEG	176, 7	1497445	0, 0023
$gg \rightarrow ZZ \rightarrow 4e/4\mu$	gg2vv	4, 8	548760	0, 00017
$gg \rightarrow ZZ \rightarrow 2e2\mu$	gg2vv	12, 03	393774	0, 00060
$WZ \rightarrow 3l\nu$	MadGraph	1057	1988552	0, 010
$WW \rightarrow 2l2\nu$	MadGraph	$6, 00 \cdot 10^3$	1933235	0, 061
tW	MadGraph	$2, 24 \cdot 10^4$	991118	0, 44
$t\bar{t}$	MadGraph	$2, 39 \cdot 10^4$	6903177	0, 068
$Z(\rightarrow l^+l^-) + \text{jets}$	MadGraph	$3, 533 \cdot 10^6$	30459503	2, 27
$W(\rightarrow l\nu) + \text{jets}$	MadGraph	$3, 1314 \cdot 10^7$	57671176	10, 6

TABLE 3.1 – Processus simulés utilisés dans notre analyse.

Tous les échantillons de bruit de fond sont produits avec le générateur `MadGraph` [11]. Il s'agit d'un générateur à l'arbre, qui n'inclut aucune correction virtuelle et permet de générer jusqu'à 6 partons supplémentaires dans l'état final. Dans le cas de ce générateur, les corrections virtuelles aux processus sont appliqués après la génération, sous forme d'un facteur de correction, en faisant l'hypothèse que les corrections liées aux boucles ne changent pas la forme du signal. La simulation de la cascade partonique, de la fragmentation, de l'événement sous-jacent et de l'empilement sont également gérés par `PYTHIA` dans ce cas.

Les sections efficaces simulées pour les processus ci-dessus servent à transformer un nombre d'événements d'un échantillon de simulations en un nombre d'événements

attendus dans le détecteur pendant la période de prise de données. Ceci se fait en appliquant pour chaque événement d'un même échantillon un poids

$$w = \frac{\sigma \cdot \int \mathcal{L} dt}{N_{\text{sim}}}, \quad (3.8)$$

où σ est la section efficace du processus simulé et N_{sim} le nombre d'événements de l'échantillon. De cette manière, on obtient un nombre d'événements attendus qui peut être comparé aux données.

Chapitre 4

Sélection des événements

Dans le chapitre précédent, nous avons précisé comment les échantillons de données sont sélectionnés au niveau du détecteur CMS. Cependant, la sélection de ces données contient encore de nombreuses contributions de processus différents, et n'est pas encore adaptée pour l'étude de la production de paires de Z se désintégrant en deux électrons ou deux muons. C'est à ce problème que nous nous attaquons dans ce chapitre, en appliquant une série de coupures de sélection. Nous commençons par passer en revue tous les critères appliqués. Ensuite, nous introduisons une correction importante qui doit être appliquée aux simulations, due au nombre d'interactions se produisant lors d'un même croisement de faisceaux. Nous étudions l'efficacité des triggers de sélection appliqués pour sélectionner notre échantillon initial. Enfin, nous comparons quelques distributions obtenues dans les données et les simulations pour des variables de contrôle.

4.1 Coupures de sélection

Les coupures que nous discutons dans cette section sont appliquées de la même manière sur les données et sur les simulations. L'objectif de ces coupures est double :

- premièrement, il s'agit de sélectionner un échantillon où les données et les simulations sont comparables, autrement dit que l'on soit sûr que nos simulations soient fiables. C'est en effet sur cette hypothèse que repose toute l'analyse statistique ;
- deuxièmement, on veut sélectionner un échantillon comprenant le moins possible de bruit de fond et le maximum de signal, afin d'optimiser la sensibilité de notre analyse statistique.

Ce deuxième point est particulièrement important, et la valeur des coupures doit être choisie avec soin : il faut réduire au maximum le bruit de fond tout en ne supprimant pas trop de signal. Un compromis doit en général être trouvé, et il sera de toute façon nécessaire d'estimer le bruit de fond dans l'échantillon final, comme on le verra au chapitre 5.

Les coupures appliquées ici sont fortement inspirées de [3], mais certaines dispositions particulières ont dû être appliquées pour tenir compte des particularités de notre analyse et des échantillons de simulations disponibles.

La sélection de l'échantillon, tout comme le reste de notre analyse, est réalisée avec la bibliothèque ROOT de C++ [29], qui contient de nombreux outils très pratiques pour la physique des particules : vecteurs de Lorentz, histogrammes,... Au final, notre analyse totale aura demandé le développement d'une dizaine de programmes, totalisant plus de 5000 lignes de code.

4.1.1 Sélection de l'état final

La première sélection que nous appliquons est évidente : il s'agit de sélectionner les bons états finaux leptoniques. Les trois états finaux que nous retenons pour notre analyse sont :

1. $e^+e^-e^+e^-$, que nous désignerons par la suite comme $4e$,
2. $\mu^+\mu^-\mu^+\mu^-$, que nous désignerons par la suite comme 4μ , et
3. $e^+e^-\mu^+\mu^-$, que nous désignerons par la suite comme $2e2\mu$.

En pratique, nous demandons donc quatre leptons chargés, pouvant être appariés en deux paires de leptons de même saveur et de charges opposées.

4.1.2 Coupures sur la quantité de mouvement transverse des leptons et la couverture du détecteur

Les leptons de faible quantité de mouvement transverse sont mal reconstruits dans le détecteur et ont également plus de chances d'appartenir à des jets. De plus, nous étudions la production de paires de bosons Z sur leur couche de masse, ce qui mène à une quantité de mouvement transverse (p_T) moyenne de l'ordre de 30 GeV. Par conséquent, il est utile de mettre une coupure sur le p_T de chaque lepton. Nous imposons au moins un lepton avec $p_T > 20$ GeV, et que les 3 autres leptons aient $p_T > 10$ GeV. Ces coupures sont assez strictes, mais améliorent la pureté des candidats leptons tout en assurant une comparaison efficace avec les simulations.

La détection des électrons n'est possible qu'endéans la couverture angulaire combinée du calorimètre électromagnétique et du détecteur de traces, et celle des muons qu'endéans celle des chambres à muons. Afin d'éviter d'avoir des événements simulés hors de la région où la détection est efficace, il faut donc appliquer une coupure sur η pour chaque lepton :

- pour les électrons, la couverture du traceur central est de $|\eta| < 2,5$ (ce qui correspond à un angle d'environ 10° par rapport à l'axe du faisceau). On applique donc la coupure correspondante. De plus, la région de transition entre le tonneau et les bouchons est mal couverte par le calorimètre électromagnétique et a une inefficacité de détection importante. Afin d'éviter ces problèmes, on rejette également les électrons se trouvant dans cette région, correspondant à $1,4442 < |\eta| < 1,566$. Il faut considérer la coordonnée η qui correspond au point d'impact de l'électron dans le calorimètre électromagnétique, et non pas la coordonnée η qui correspond à l'angle polaire de la trace de l'électron. Comme c'est cette dernière coordonnée qui est effectivement mesurée dans nos échantillons, on reconstitue la valeur de η dans les calorimètres par un simple exercice de trigonométrie, en connaissant le déplacement longitudinal du vertex d'interaction par rapport au centre du détecteur ;

- pour les muons, on rejette tous les candidats se trouvant en dehors de la couverture du détecteur en imposant $|\eta| < 2,4$.

4.1.3 Identification

Il faut s'assurer que les quatre leptons reconstruits correspondent bien à des vrais électrons ou à des vrais muons. Pour ce faire, une série de critères doivent être vérifiés. On applique les coupures recommandées par la collaboration CMS, appelées coupures "légères" pour les électrons et "strictes" pour les muons.

Ces coupures tiennent compte, pour les électrons, des critères suivants :

- la différence en η et en ϕ par rapport au vertex primaire d'interaction, selon qu'on reconstitue les variables angulaires à partir du détecteur de traces ou du calorimètre ;
- la largeur de la gerbe dans le calorimètre électromagnétique ;
- l'absence de dépôts dans le calorimètre hadronique derrière la gerbe dans le calorimètre électromagnétique ;
- la différence entre la mesure de la quantité de mouvement dans le détecteur de traces, et celle de l'énergie dans le calorimètre électromagnétique ;
- le déplacement longitudinal et transverse de l'origine de la trace reconstituée par rapport à l'interaction proton-proton ;
- le nombre de points de la trajectoire (impacts) manquants dans le détecteur de pixels, proche du point d'interaction ;
- l'incompatibilité de la reconstruction avec un photon converti en une paire e^+e^- dans la matière du détecteur, ce qu'on vérifie à partir du détecteur de pixels et du détecteur de traces par l'absence d'une trace d'un électron colinéaire de charge opposée.

Pour les muons, on prend en compte les critères suivants :

- l'association de la trajectoire reconstruite dans le détecteur de traces et des impacts dans les chambres à muons ;
- la qualité de l'ajustement de la trajectoire dans le détecteur de traces ;
- le nombre d'impacts dans les chambres à muons ;
- le déplacement longitudinal et transverse de l'origine de la trace reconstituée par rapport à l'interaction proton-proton ;
- la compatibilité avec la trajectoire dans le détecteur à pixels, proche du point d'interaction ;
- le nombre de couches de détection qui ont pris une mesure dans le détecteur de traces.

Une coupure plus stricte sur l'origine de la trace est également appliquée. Pour ce faire, on impose que la signification du paramètre d'impact relatif à l'interaction proton-proton, SIP_{3D} , satisfasse $SIP_{3D} = \left| \frac{IP}{\sigma_{IP}} \right| < 4$ pour chaque lepton, où IP est la distance la plus courte entre la trajectoire du lepton et le vertex primaire, et σ_{IP} est son incertitude associée [3].

4.1.4 Isolement

Une grande différence entre un lepton provenant d'un jet et un lepton issu du processus dur est que le lepton d'un jet est entouré de beaucoup d'autres particules, formant le jet en question. En imposant que la particule considérée soit isolée, on s'assure donc d'avoir de vrais leptons émis lors du processus dans notre échantillon final. Afin de faire cette sélection, on définit pour chaque lepton chargé l la variable suivante, sans dimensions [3] :

$$R_{\text{Iso}}^l \equiv \left(\sum p_T^{\text{chargés}} + \text{MAX} \left[0, \sum p_T^{\text{neutres}} + \sum p_T^\gamma - \rho_{\text{Iso}} \times A_{\text{eff}} \right] \right) / p_T^l, \quad (4.1)$$

où $\sum p_T^{\text{chargés}}$, $\sum p_T^{\text{neutres}}$ et $\sum p_T^\gamma$ représentent respectivement les sommes des impulsions transverses des hadrons chargés, des hadrons neutres et des photons présents dans un cône dont le rayon est donné par : $\sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} = 0,4$, centré sur la direction du lepton. ρ_{Iso} est la densité moyenne d'énergie dans l'événement, due à l'événement sous-jacent de la collision proton-proton et surtout aux autres collisions empilées. Cette quantité est calculée automatiquement d'après les particules détectées dans chaque événement. A_{eff} est l'*aire effective* du cône : dans notre cas, on la prend simplement comme $A_{\text{eff}} = \pi \cdot (0,4)^2$, ce qui est en réalité une simplification.

Pour notre analyse, les leptons sont considérés isolés si $R_{\text{Iso}}^l < 0,4$.

4.1.5 Coupures d'analyse

Pour les événements où les 4 leptons ont passé les coupures ci-dessus, on sélectionne le processus de production de paires de bosons Z en tentant de reconstruire les bosons Z à partir des leptons disponibles. Les leptons sont appariés suivant l'algorithme suivant :

- tout d'abord, il faut que chaque paire contienne deux leptons de même saveur et de charges opposées. Pour le cas du canal $2e2\mu$, ce critère est suffisant pour reconstruire les Z ; pour les canaux $4e$ et 4μ , il reste systématiquement deux possibilités et on doit poursuivre l'algorithme ;
- on apparie les leptons selon la masse invariante : on regarde les deux possibilités d'appariement avec le premier lepton en notant, pour chacune d'elles, la masse invariante de la paire formée. La paire retenue est celle dont la masse invariante est la plus proche de la masse du Z : 91,2 GeV. La deuxième paire est formée par les deux autres leptons dans tous les cas ;
- si la différence entre les masses invariantes des deux paires considérées est inférieure à 10 GeV, le critère précédent n'est pas considéré assez discriminant pour choisir une paire plutôt que l'autre. Dans ce cas, on choisit la paire de leptons émise le plus "dos à dos" possible. Pour ce faire, on calcule le produit scalaire des quantités de mouvement des deux leptons, normalisé au produit des normes, pour les deux paires considérées : la paire retenue est celle pour laquelle ce nombre est le plus petit.

Une fois l'appariement effectué, on peut appliquer les coupures correspondant au domaine cinématique que l'on veut étudier, autrement dit préciser l'*acceptance* de notre analyse. On veut deux bosons Z sur leur couche de masse ; par conséquent, on impose que la masse de chaque boson Z reconstruit se situe dans une fenêtre

de 30 GeV autour de la masse du Z : $|M_{Z_{1,2}} - 91,2 \text{ GeV}| < 30 \text{ GeV}$. De plus, on se restreint aux événements où la masse invariante du système des deux Z est supérieure à 200 GeV.

4.1.6 Autres coupures

Bien qu'il ne s'agisse pas strictement d'une coupure de sélection, il faut également s'assurer que chaque événement de données n'est pris qu'une fois, même s'il a déclenché différents triggers. Chaque événement enregistré par CMS portant un numéro unique, il suffit de vérifier que ce numéro n'est pas dans la liste des événements ayant déjà été acceptés auparavant.

De plus, pour tester le modèle comprenant les corrections électrofaibles, nous appliquons la coupure sur la variable ρ , décrite en 2.4.3. Cette coupure n'est pas appliquée lors de la mesure de la section efficace afin de ne pas trop réduire la statistique, comme on le verra dans le chapitre 6.

4.2 Distributions de contrôle

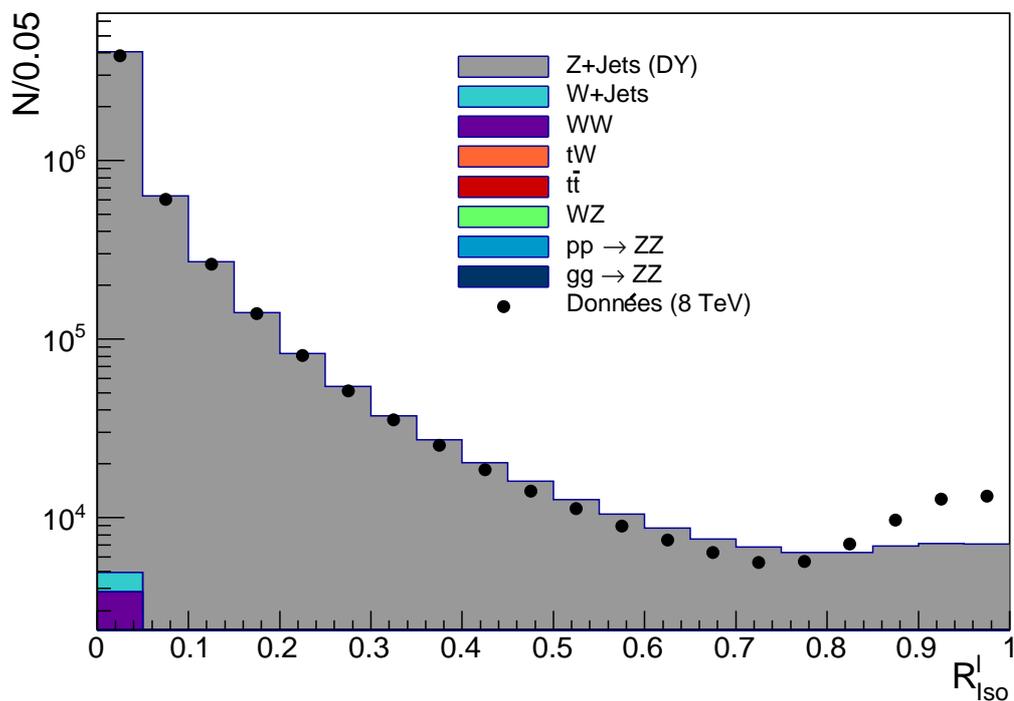
La figure 4.1 représente un exemple de distribution de contrôle que l'on peut obtenir afin de vérifier nos coupures. Pour faire cette distribution, on a sélectionné un échantillon d'événements constitué d'au moins deux leptons d'une masse invariante compatible avec celle du Z , où au moins l'un des deux leptons passe les coupures finales d'identification, d'isolement, cinématiques et de couverture du détecteur. Le deuxième lepton doit passer également tous les critères, sauf le critère d'isolement. C'est l'isolement de ce deuxième lepton qu'on a représenté à la figure 4.1. La sélection de cet échantillon fait que l'on prendra principalement des événements de production d'un seul Z (appelés *Drell-Yan*), mais d'autres processus, dont notre signal, sont également présents.

On observe un bon accord entre les simulations et les données pour les faibles valeurs d'isolement. Au-delà de notre valeur de coupure (0,4), on constate que l'accord se détériore, surtout pour les électrons. Ceci nous indique que notre coupure est bien placée : en effet, avec un critère moins strict, on garderait des événements manifestement mal simulés. On constate également que l'intégrale au-delà d'une valeur de 0,4 est petite par rapport au total : ceci indique que notre coupure ne fait pas perdre trop de statistique. Enfin, comme on pouvait s'y attendre d'après les sections efficaces des processus (table 3.1), le processus $Z + \text{Jets}$ domine à ce niveau de sélection.

4.3 Pondération des événements simulés selon le nombre de vertex

Le nombre de vertex d'interaction est un effet de l'interaction des autres protons des faisceaux. La distribution de cette variable pour les données se trouve en figure 4.2. Cette variable est difficile à simuler, ce qui résulte en une incompatibilité pour les spectres de cette variable entre les données et les simulations et des différences entre

Variable d'isolement pour des électrons



Variable d'isolement pour des muons

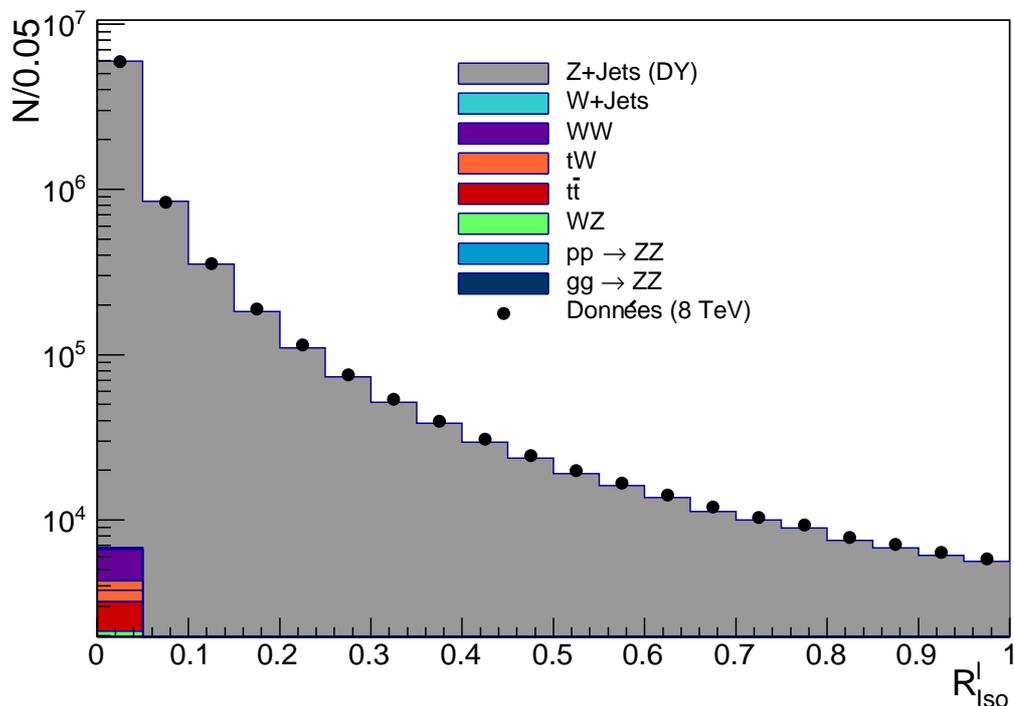


FIGURE 4.1 – Distributions de contrôle pour R_{Iso}^l , pour des événements avec au moins un boson Z se désintégrant en deux leptons chargés. Les distributions sont montrées pour des électrons (en haut) et pour des muons (en bas).

les variables d'isolement dans les données et les simulations. Afin de pouvoir faire une comparaison, on applique un facteur de pondération pour chaque événement simulé.

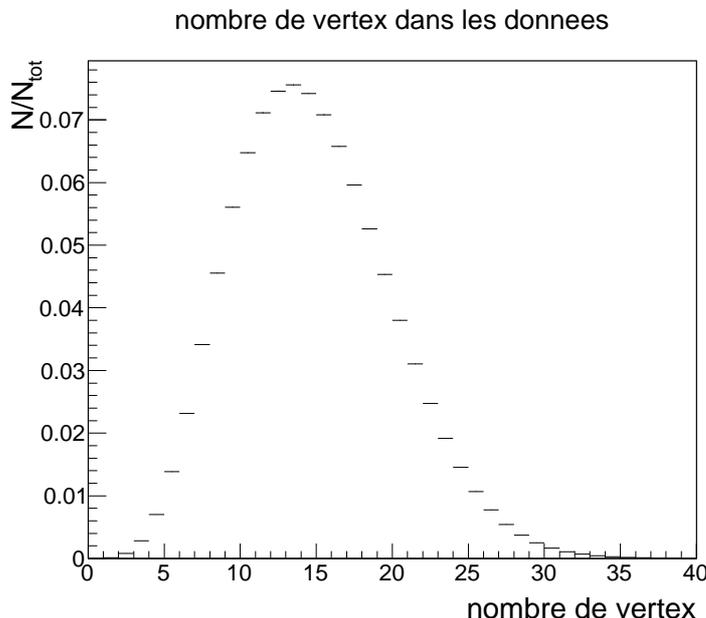


FIGURE 4.2 – Distribution du nombre de vertex pour les données, après normalisation de l'intégrale à l'unité.

Ce facteur de pondération se calcule comme suit. On sélectionne un échantillon d'événements comprenant au moins un Z se désintégrant en 2 leptons chargés (le choix du type d'événements importe peu, cette variable représentant l'activité des faisceaux), pour les données et pour chaque générateur de simulations : **POWHEG**, **gg2vv** et **MadGraph**. On normalise l'intégrale de chacune de ces distributions à l'unité. Ensuite, pour chaque générateur, on définit le facteur de pondération à donner à l'événement simulé comme le rapport entre le nombre d'événements dans les distributions normalisées de données et de simulations, à cette valeur du nombre de vertex.

Les distributions du facteur de correction selon le nombre de vertex sont très différentes d'un générateur à l'autre, comme on peut le voir à la figure 4.3.

4.4 Vérification de l'efficacité des triggers

Avant de poursuivre notre analyse, il est nécessaire de vérifier si les triggers utilisés sont efficaces pour notre sélection d'événements. En effet, afin de pouvoir comparer données et simulations, il faut d'une part que tous les événements intéressants soient déclenchés par au moins un trigger correspondant à nos lots de données, et d'autre part que les simulations satisfassent les critères d'au moins un trigger (sinon, le vrai événement correspondant n'aurait pas été enregistré). Si ce n'est pas le cas, il faut appliquer un facteur de correction.

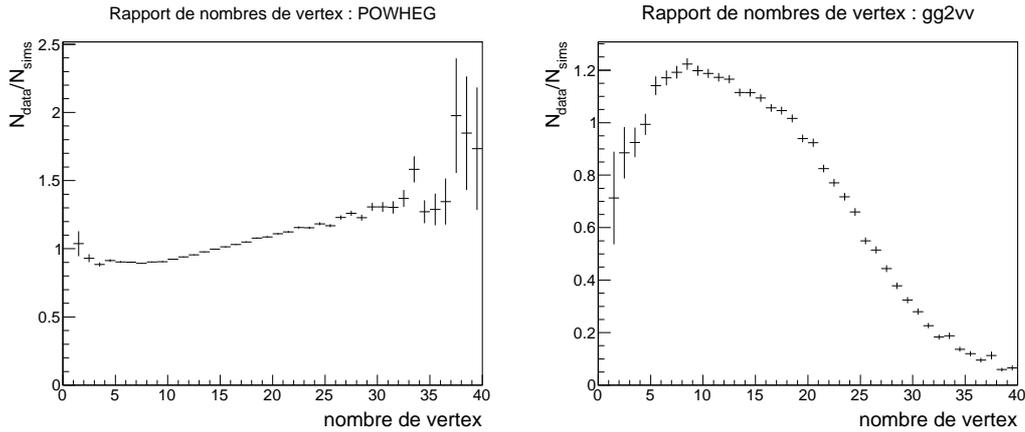


FIGURE 4.3 – Correction selon le nombre de vertex, pour des simulations POWHEG (à gauche) et gg2vv (à droite).

Dans notre cas, les triggers d'intérêt sont les suivants :

- le trigger $e(17)e(8)$ demande au moins 1 électron de $p_T > 17$ GeV et 1 autre électron de $p_T > 8$ GeV. Ces deux électrons doivent également remplir des critères de reconstruction au niveau du détecteur de traces et du calorimètre électromagnétique ;
- le trigger $\mu(17)\mu(8)$ demande au moins 1 muon de $p_T > 17$ GeV et 1 autre muon de $p_T > 8$ GeV ;
- les triggers $e(17)\mu(8)$ et $\mu(17)e(8)$ demandent au moins un électron et un muon, suivant le même principe.

Pour notre échantillon avec les coupures finales décrites plus haut, les taux de déclenchement pour ces 4 triggers sont donnés dans la table 4.1.

Canal	Trigger	$e(17)e(8)$	$\mu(17)\mu(8)$	$e(17)\mu(8)$	$\mu(17)e(8)$
	4e		100%	0%	0%
4 μ		0%	100%	0%	0%
2e2 μ		92%	88%	100%	100%

TABLE 4.1 – Taux de déclenchement des triggers pour les événements de notre échantillon final de données.

Montrons que l'efficacité des deux triggers $e\mu$ est de 100 % pour notre analyse. En effet, 88 % des événements du type $2e2\mu$ ont déclenché le trigger $\mu(17)\mu(8)$. Prenons cet échantillon comme un échantillon de contrôle. Le nombre d'événements dans cet échantillon est :

$$N_{\mu\mu}^* \equiv 0,88 \cdot N_{2e2\mu}, \quad (4.2)$$

où $N_{2e2\mu}$ est le nombre d'événements dans l'état final $2e2\mu$. Parmi ces événements, tous ont déclenché les deux triggers $e\mu$ car on a l'information que tous les événements $2e2\mu$ ont déclenché ces triggers. Par conséquent, la probabilité conditionnelle de déclencher le trigger $e\mu$ sachant que l'événement a déclenché le trigger $\mu\mu$ est de

100 % :

$$N_{e\mu|\mu\mu} = N_{\mu\mu}^* \cdot \mathcal{P}(e\mu|\mu\mu) \stackrel{!}{=} N_{\mu\mu}^*. \quad (4.3)$$

Or, $\mathcal{P}(e\mu|\mu\mu)$ peut être interprété comme l'efficacité $\mathcal{E}(e)$ de la "jambe e " du trigger $e\mu$: en effet, la "jambe μ " du trigger $e\mu$ ne joue pas de rôle ici, car l'événement a déjà déclenché le trigger $\mu\mu$. Donc,

$$\mathcal{P}(e\mu|\mu\mu) = \mathcal{E}(e) = 100 \%. \quad (4.4)$$

On a donc montré que l'efficacité de la jambe e du trigger $e\mu$ est de 100 %. Le raisonnement est identique pour chacun des deux triggers $e\mu$. On peut ensuite répéter le même raisonnement pour la jambe μ des deux triggers $e\mu$, en se basant sur l'échantillon $2e2\mu$ ayant déclenché le trigger $e(17)e(8)$. On arrive à la même conclusion, que l'efficacité de la jambe μ du trigger $e\mu$ est de 100 %. Par conséquent, les efficacités des triggers $e(17)\mu(8)$ et $\mu(17)e(8)$ sont de 100 %.

Nous pouvons extrapoler ces résultats pour estimer que l'efficacité du trigger $e(17)e(8)$ pour l'échantillon $4e$ et celle du trigger $\mu(17)\mu(8)$ pour l'échantillon 4μ sont également de 100 %, puisque les critères appliqués pour les électrons et les muons sont équivalents à ceux des triggers $e\mu$.

L'incertitude systématique liée à la détermination de l'efficacité des jambes e et μ des triggers peut être estimée. En effet, notre raisonnement se base sur le fait qu'il n'y a aucun événement qui passe un trigger $e\mu$, tout en passant le trigger $\mu\mu$. Si l'on compte que le nombre d'événements dans ce cas suit une distribution de Poisson, d'un paramètre inconnu, nous savons que le fait d'avoir observé zéro événement permet d'exclure les valeurs de ce paramètre supérieures à 1,15 à 67,3 % de niveau de confiance (ou 1σ) : en effet, la probabilité d'observer zéro événement pour une loi de Poisson de paramètre supérieur à 1,15 est inférieure à 32,7 %, comme on peut le voir par exemple dans [30]. L'incertitude relative totale pour l'efficacité de la jambe e , conventionnellement à 1σ , peut donc être estimée comme :

$$\Delta_{\text{triggers}} \simeq \frac{1,15}{N_{\mu\mu}^*}, \quad (4.5)$$

et similairement pour la jambe μ . Comme on le voit au tableau 4.2, on a que le nombre d'événements observés dans l'état final $2e2\mu$ est de 73. On peut donc estimer l'incertitude systématique due à l'efficacité des triggers comme $\frac{1,15}{0,88 \cdot 73} \approx 2 \%$.

Il faut également s'assurer que les simulations sont comparables aux données en regardant le taux de déclenchement que les événements de signal simulés auraient eu dans le détecteur. Pour ce faire, nous regardons, pour chaque événement dans nos simulations de signal passant les critères finaux mentionnés ci-dessus, la proportion n'ayant passé aucun des quatre triggers ci-dessus. Le résultat est que tous les événements sélectionnés par les coupures d'analyse dans les simulations déclenchent au moins un trigger simulé. Nous pouvons donc supposer que nos simulations peuvent être comparées aux données sans appliquer un facteur de correction dû aux triggers.

4.5 Distributions pour l'échantillon final

La table 4.2 reprend la composition finale de notre échantillon après application des coupures, pour les données, les simulations de signal et les simulations de bruit

de fond WZ .

État final	Données	ZZ	WZ
$4e$	16	21, 16	0, 05
4μ	45	45, 25	0, 01
$2e2\mu$	73	65, 02	0, 05
Total	134	131, 43	0, 11

TABLE 4.2 – Composition de l'échantillon final, dans les données et dans les simulations des processus ZZ et WZ .

Les figures 4.4 à 4.6 présentent les distributions pour différentes variables de sélection discutées plus haut, dans notre échantillon final, après application de toutes les coupures sauf celle sur ρ .

La figure 4.4 présente les distributions en quantité de mouvement transverse des quatre leptons, triés par ordre décroissant de quantité de mouvement transverse. On peut ainsi constater que la coupure à 10 GeV sur tous les leptons est efficace même pour les leptons de plus bas p_T . Le bon accord entre les données et les simulations est aussi dû aux coupures de sélection (isolement et identification).

La figure 4.5 montre les variables η et R_{iso}^l , pour les électrons et pour les muons séparément. On constate également un bon accord entre données et simulations.

La figure 4.6 montre les distributions des masses des paires de leptons identifiées comme des bosons Z . On constate que ces masses sont, pour les deux paires de leptons, regroupées autour de la masse du Z , 91,2 GeV, ce qui nous indique que notre sélection des paires est correcte : pour chaque événement, il est raisonnable de penser que l'on a deux bosons Z réels. Cette figure montre également la distribution de la variable ρ au niveau reconstruit. On perd 16 % des événements dans les données et 18 % des événements de simulations en rejetant les événements où cette variable est supérieure à 0,3.

Sur les figures 4.4 à 4.6, le bruit de fond simulé n'a été représenté que par le canal WZ . En effet, les autres canaux de simulations de bruit de fond ne donnent aucun événement qui passe les coupures finales. Même le bruit de fond WZ est très faible et uniquement visible en échelle logarithmique. Cependant, par expérience, il faut s'attendre à ce que les bruits de fond $Z + \text{Jets}$ et $WZ + \text{Jets}$ soient sous-estimés par les simulations. En effet, le taux de jets erronément identifiés comme des leptons isolés est très difficile à simuler de façon fiable. C'est pour cette raison qu'une autre méthode d'estimation du bruit de fond a été employée pour notre analyse. Cette méthode est décrite en détail dans le chapitre suivant.

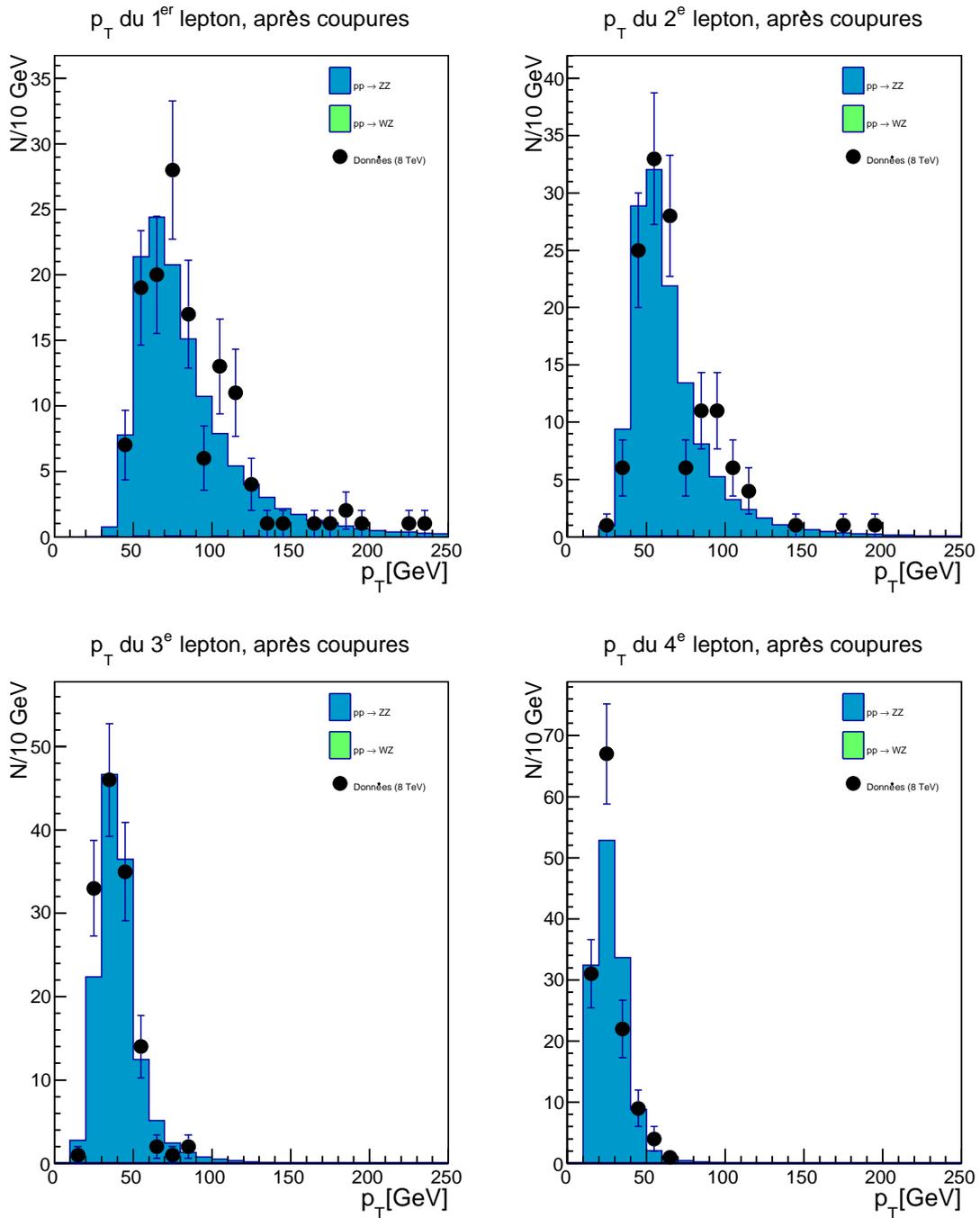


FIGURE 4.4 – Distributions de contrôle pour la quantité de mouvement transverse (p_T) des quatre leptons de chaque événement, dans l'échantillon final, pour les données et les simulations. Les leptons sont triés, dans un même événement, par ordre décroissant de quantité de mouvement transverse.

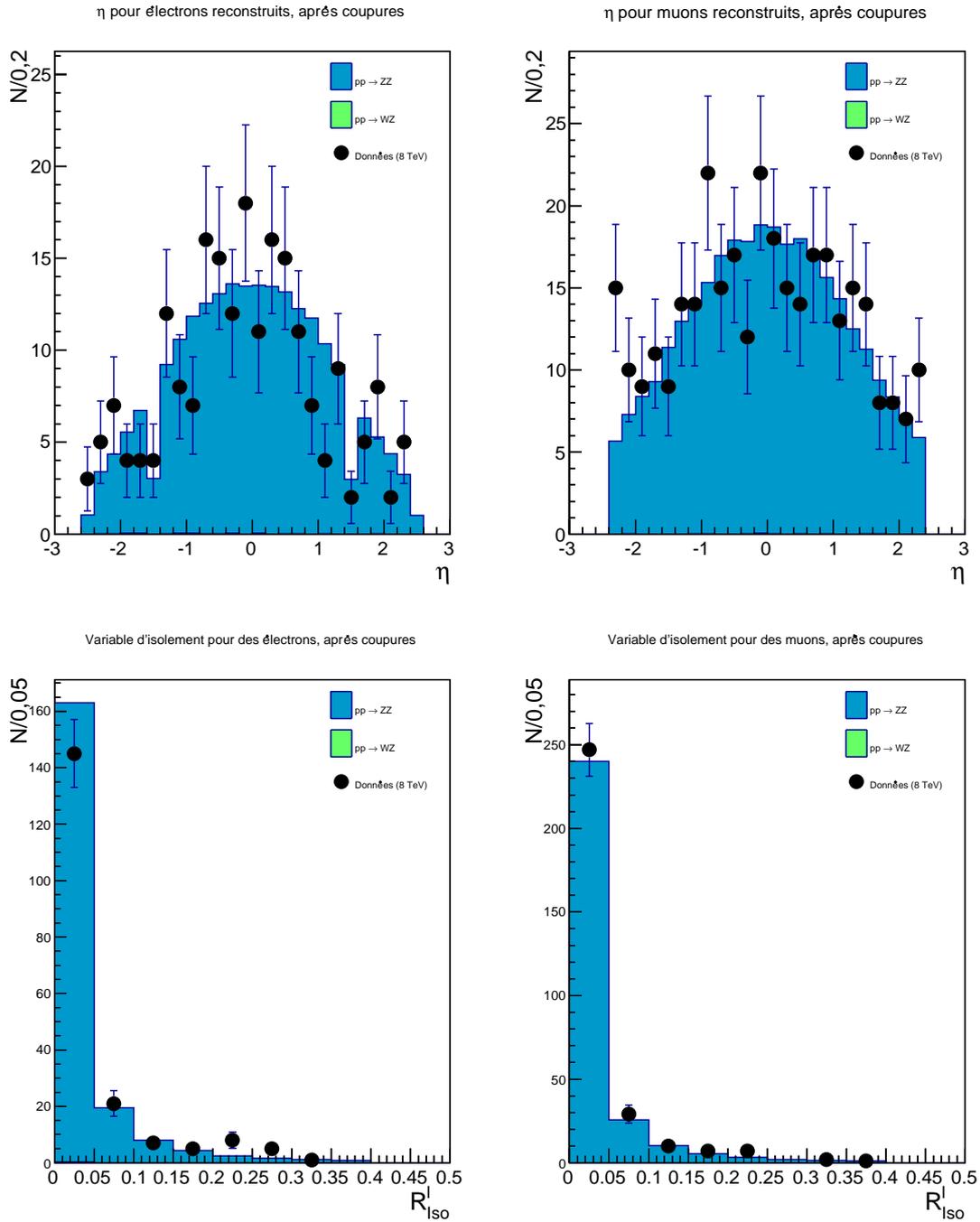


FIGURE 4.5 – En haut, coordonnée η , pour les électrons (à gauche) et les muons (à droite), dans l'échantillon final, pour les données et les simulations. En bas, variable d'isolation R_{Iso}^l pour les électrons (à gauche) et les muons (à droite).

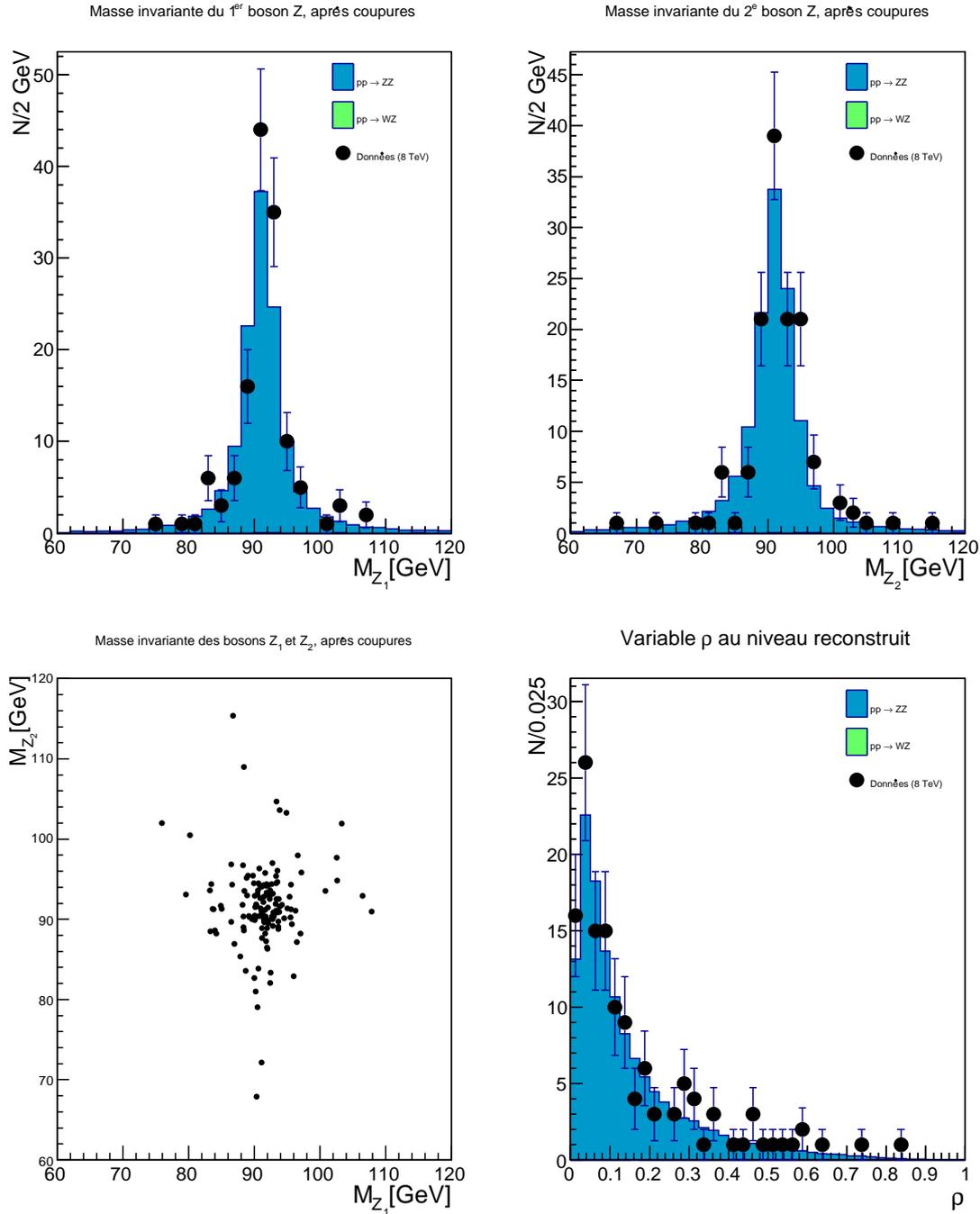


FIGURE 4.6 – En haut, masses invariantes des paires de leptons sélectionnées pour constituer les deux bosons Z, dans l'échantillon final, pour les données et les simulations. En bas à gauche, masses invariantes des deux paires de leptons constituant les bosons Z pour un même événement. En bas à droite, variable ρ au niveau reconstruit pour les événements de notre échantillon final.

Chapitre 5

Calcul du bruit de fond

Ce chapitre est consacré à l'étude du bruit de fond associé au processus de production de paires de bosons Z se désintégrant en leptons chargés. La détermination de ce bruit de fond est très importante pour la suite de notre analyse : il est impératif de savoir quelle proportion des événements observés correspond à notre signal, et quelle proportion constitue du bruit de fond. Nous commençons par décrire le principe de notre méthode d'estimation du bruit de fond. Elle repose sur l'estimation du taux de jets erronément identifiés comme des leptons isolés, noté f . Nous regardons en détail les deux parties du calcul du bruit de fond : l'estimation de f et l'estimation du bruit de fond à proprement parler.

5.1 Principe

Le bruit de fond associé à un processus peut toujours se décomposer en deux catégories :

1. le bruit de fond *irréductible* ou *résonant*. Ce bruit de fond a les mêmes états initiaux et finaux que le processus étudié, et ne peut donc pas être supprimé par des coupures sur la sélection de l'état final ;
2. le bruit de fond *réductible*, qui se compose de particules dans l'état final différentes de celles du signal. Ce bruit de fond est dit *instrumental*, car il provient nécessairement d'une mauvaise identification de certaines particules dans le détecteur.

La particularité de notre processus est de ne comprendre que des bruits de fond d'origine instrumentale. En effet, il n'y a pas d'autre processus du Modèle Standard qui produit un état final à quatre leptons chargés¹. Les bruits de fond les plus importants viennent des processus suivants :

1. $W^\pm(\rightarrow l^\pm\nu)Z(\rightarrow l'^+l'^-)$, où un jet présent dans l'événement est identifié à tort comme un lepton l^\mp ;

1. Cette affirmation n'est pas strictement exacte : on pourrait avoir, par exemple, des processus du type $ZWW \rightarrow 4l2\nu$. Cependant, la section efficace très faible de ces processus les rend négligeables.

2. $Z(\rightarrow l^+l^-) + \text{jets}$, ou processus Drell-Yan, où deux jets sont identifiés comme des leptons. Cette occurrence assez rare peut être contrebalancée par la forte section efficace de ce processus, comme on l'a vu à la table 3.1.

Le bruit de fond dans nos données dépend de la sensibilité et de la précision du détecteur, dans des régions cinématiques très particulières. On ne peut donc pas se baser sur des simulations Monte-Carlo pour chaque processus de bruit de fond, car les queues de distribution de la fragmentation des jets et la réponse du détecteur y sont mal connues pour une cinématique correspondant à notre échantillon final : le taux de jets erronément identifiés comme des leptons y est souvent mal modélisé.

La méthode que l'on emploie est inspirée de la note d'analyse de [3], mais a été adaptée pour tenir compte des particularités de notre analyse, puisqu'on étudie des événements à haute masse invariante et impliquant des bosons Z sur leur couche de masse. Elle se base principalement sur les données et les simulations de signal. Comme on va le remarquer, il a été nécessaire d'utiliser également des simulations du processus WZ , mais sans s'en servir directement pour estimer le bruit de fond.

Cette méthode consiste en deux parties distinctes. Premièrement, on calcule la probabilité d'un "faux" électron de passer tout de même les critères finaux de sélection de notre processus. Ensuite, sur base d'échantillons contenant respectivement trois "vrais" et un "faux" lepton, et deux "vrais" et deux "faux" leptons et en les pondérant par la probabilité que le "faux" lepton passe tout de même les critères finaux, on pourra estimer la contribution d'événements de bruit de fond dans notre échantillon final.

5.2 Calcul du taux d'erreur de sélection des leptons

La première étape est de calculer le taux de jets erronément identifiés comme des leptons, que l'on appelle f . Ce taux se définit comme la proportion de leptons passant uniquement des critères "légers" d'identification, qui passent également nos coupures finales d'identification et d'isolement. Ces critères "légers" se basent sur les mêmes variables que celles décrites dans la section 4.1.3, mais avec des coupures moins strictes sur chacune d'entre elles : par rapport aux standards de la collaboration CMS, on prend les critères "légers" pour les muons et "veto" pour les électrons.

L'étape délicate pour estimer ce taux est de sélectionner un échantillon d'événements où l'on sait que les leptons passant les critères "légers" ne sont effectivement pas des vrais leptons.

Nous choisissons un état final comprenant une paire de leptons chargés pouvant être associée à un boson Z en imposant $|M_{l+l-} - 91,2 \text{ GeV}| < 10 \text{ GeV}$, et exactement un lepton supplémentaire. Les deux leptons appariés doivent satisfaire les critères finaux décrits pour notre analyse. Le lepton supplémentaire requis doit uniquement passer les critères "légers" d'identification, décrits plus haut, et aucun critère d'isolement. De plus, nous imposons que l'événement ait une énergie transverse manquante de 25 GeV au maximum. L'énergie transverse manquante est définie comme l'opposé du module de la somme vectorielle des quantités de mouvement transverses de toutes les particules dans l'état final détectées dans l'événement. Une grande

énergie transverse manquante indique généralement la présence d'un neutrino : en effet, ces particules n'interagissent que très rarement avec la matière, et traversent toute l'épaisseur du détecteur sans interagir avec celui-ci. Cette coupure sur l'énergie transverse manquante permet de réduire la contribution d'événements du type $WZ \rightarrow 3l\nu$.

Cet état final ne correspond à aucun processus physique avec trois vrais leptons. On fait l'hypothèse que ces événements sont tous des processus de Drell-Yan où le lepton supplémentaire est un jet erroné identifié comme un lepton. On estime f en faisant le rapport, dans une région cinématique donnée, du nombre d'événements de cet échantillon où le lepton supplémentaire passe les coupures finales d'identification et d'isolement sur le nombre total d'événements de l'échantillon. Cependant, malgré les précautions prises pour éviter la présence de trop d'énergie transverse manquante dans nos événements, des événements du type $WZ \rightarrow 3l\nu$ restent tout de même présents dans notre échantillon. Ces événements tirent f vers le haut, car le lepton présent dans ces événements est par nature un vrai lepton. De même, des événements du type $ZZ \rightarrow 4l$, où l'un des leptons n'est pas identifié, contribuent également à l'échantillon. De tels événements invalident également l'hypothèse faite plus haut, car le troisième lepton produit est un vrai lepton du processus. Il faut donc soustraire ces contributions, estimées à partir des simulations décrites à la section 3.3.2 lors du calcul de f . L'estimation de f se fait donc comme :

$$f = \frac{T_{\text{données}} - T_{WZ} - T_{ZZ}}{L_{\text{données}} - L_{WZ} - L_{ZZ}}, \quad (5.1)$$

où $T_{\text{données}}$ est le nombre de leptons passant les coupures "strictes" dans l'échantillon de données constitué. T_{WZ} et T_{ZZ} sont le nombre d'événements passant les coupures "strictes" dans les échantillons simulés des processus WZ et ZZ , après renormalisation à la luminosité intégrée. $L_{\text{données}}$, L_{WZ} et L_{ZZ} représentent le nombre de leptons passant les coupures "légères", respectivement dans les données, les simulations WZ et les simulations ZZ .

Pour tenir compte des effets dus au détecteur, f est calculé dans différentes régions en η . De plus, pour tenir compte des effets liés à la cinématique des jets, f est aussi calculé en fonction du p_T du candidat lepton.

On calcule le taux en prenant en compte les variables suivantes :

- la saveur du lepton : e ou μ ;
- la région du détecteur où est reconstruit le lepton : tonneau ($|\eta| < 1,4442$) ou bouchons ($|\eta| > 1,4442$) ;
- la quantité de mouvement transverse du lepton, par boîtes de 10 GeV jusqu'à 80 GeV. On place tous les leptons de plus de 80 GeV dans une autre boîte (hors échelle).

Les résultats du calcul de l'estimation de f sont donnés en figure 5.1. Les incertitudes sont la somme quadratique des erreurs statistiques sur les données, suivant une loi binomiale, et de l'incertitude sur les sections efficaces WZ et ZZ , où une incertitude conservatrice de 10 % a été prise.

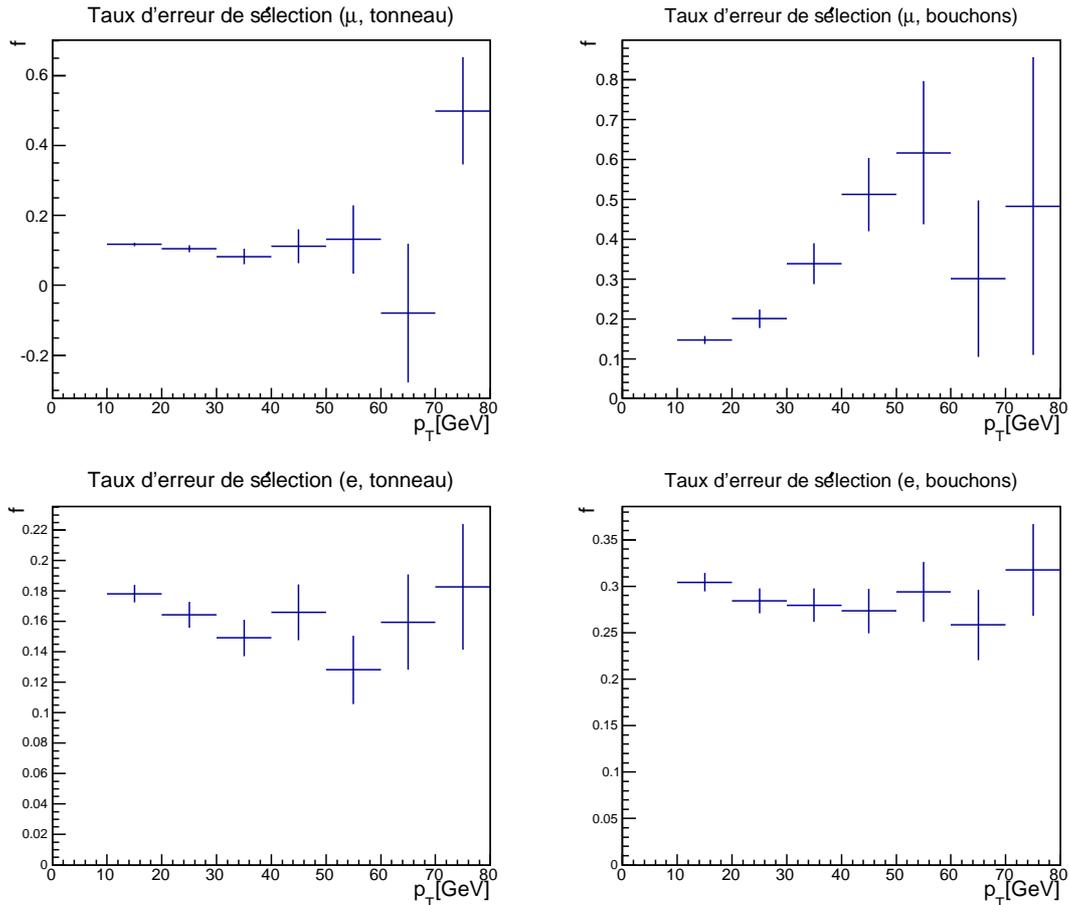


FIGURE 5.1 – Taux d'erreur de sélection des leptons f , calculé pour les muons (en haut) et les électrons (en bas), dans le tonneau (à gauche) et dans les bouchons (à droite).

5.3 Estimation du bruit de fond

Connaissant le taux d'erreur de sélection des leptons f , nous pouvons estimer le bruit de fond. Pour ce faire, nous commençons par construire deux échantillons. Un premier échantillon est construit en prenant des événements comprenant exactement trois leptons passant les critères finaux d'identification et d'isolement, ainsi que les coupures cinématiques sur la quantité de mouvement, et un lepton ne passant que les critères "légers" d'identification et aucun critère d'isolement, tout en ne passant *pas* les critères finaux. De plus, les leptons doivent satisfaire la cinématique du processus de production de paires de bosons Z , comme décrit en 4.1.5. Cet échantillon est désigné "3P1F", pour "3 *Pass*, 1 *Fail*". Ensuite, un deuxième échantillon est construit selon le même principe, mais en prenant cette fois deux leptons passant les critères finaux et deux autres leptons ne passant que les critères "légers". Cet échantillon est désigné "2P2F".

Les événements composant notre échantillon final de données sont du type "4P", les quatre leptons ayant effectivement tous passé les coupures "strictes" d'identifica-

tion et d'isolement. Connaissant f , l'estimation du bruit de fond consiste à déduire les contributions des processus physiques contenant 3 leptons et 1 jet, désignés "3l1j", ou processus WZ , et des processus contenant 2 leptons et 2 jets ("2l2j"), ou processus de Drell-Yan, à la catégorie 4P constituant notre échantillon final. Pour estimer ces contributions, on regarde les nombres d'événements dans les catégories 3P1F et 2P2F. Un schéma explicatif se trouve en figure 5.2.

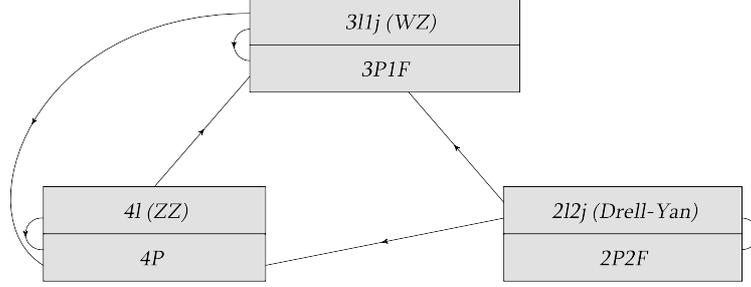


FIGURE 5.2 – Schéma récapitulatif des états finaux physiques ($4l$, $3l1j$ et $2l2j$) et de leurs contributions aux différents états finaux observés ($4P$, $3P1F$, $2P2F$).

Si l'on appelle f_i le taux d'erreur de sélection associé au lepton numéro i , le bruit de fond est alors donné par :

$$N^{\text{bkg}} = \sum \frac{f_i}{1-f_i} (N_{3P1F} - N_{3P1F}^{\text{bkg}} - N_{3P1F}^{ZZ}) + \sum \frac{f_i}{1-f_i} \frac{f_j}{1-f_j} N_{2P2F}. \quad (5.2)$$

Dans la formule (5.2), les sommes doivent se comprendre comme portant sur tous les événements des échantillons. Les N doivent être compris comme valant 1 si l'événement en question appartient à cette catégorie, 0 sinon. Les indices i et j se réfèrent, le cas échéant, aux leptons ne passant pas les critères finaux (c'est-à-dire les "fail") dans l'événement considéré. Expliquons l'origine des différents termes de cette formule, un par un.

Le terme $\frac{f_i}{1-f_i} N_{3P1F}$ représente l'estimation de la contribution des événements $3l1j$ dans l'état final, calculée à partir de l'échantillon 3P1F. Montrons pourquoi ce terme est de cette forme. En négligeant la contribution des autres états finaux (qu'on traitera plus tard), le nombre d'événements de bruit de fond est donné par le nombre d'événements comprenant trois leptons et un jet, où le jet est identifié comme un vrai lepton (T comme "tight") :

$$N^{\text{bkg}} = N_{3l1j} \cdot \mathcal{P}(j \rightsquigarrow l_T) \quad (5.3)$$

$$= N_{3l1j} \cdot \mathcal{P}(j \rightsquigarrow l_L) \cdot \mathcal{P}(l_L \rightsquigarrow l_T) \quad (5.4)$$

$$= N_{3l1j} \cdot \mathcal{P}(j \rightsquigarrow l_L) \cdot f_i, \quad (5.5)$$

où l_L désigne un lepton passant uniquement les critères "légers". En effet, par définition du taux d'erreur de sélection, on a $f_i = \mathcal{P}(l_L \rightsquigarrow l_T)$. D'autre part, le nombre

d'événements dans l'échantillon 3P1F est donné par le nombre d'événements comprenant trois leptons et un jet, où le jet est identifié comme un lepton "léger" :

$$N_{3P1F} = N_{3l1j} \cdot \mathcal{P}(j \rightsquigarrow l_L) \cdot (1 - \mathcal{P}(l_L \rightsquigarrow l_T)) \quad (5.6)$$

$$= N_{3l1j} \cdot \mathcal{P}(j \rightsquigarrow l_L) \cdot (1 - f_i). \quad (5.7)$$

En effet, il faut que le lepton provenant du jet ne passe que les coupures "légères" et pas les coupures "strictes". À partir de ce système de deux équations à deux inconnues, on peut éliminer le produit $N_{3l1j} \cdot \mathcal{P}(j \rightsquigarrow l_L)$. On obtient ainsi :

$$N^{\text{bkg}} = N_{3P1F} \cdot \frac{f_i}{1 - f_i}, \quad (5.8)$$

ce qui est bien le premier terme de (5.2).

Le terme $\frac{f_i}{1-f_i} \frac{f_j}{1-f_j} N_{2P2F}$ suit la même logique, pour la contribution des événements issus de l'échantillon 2P2F.

Le terme en N_{3P1F}^{bkg} représente des événements issus de processus $2l2j$, mais qui contribuent dans la région 3P1F sous forme de bruit de fond. Afin d'éviter de compter deux fois ces événements (c'est-à-dire de les estimer une fois en 2P2F et de les compter une deuxième fois en 3P1F), il faut soustraire cette contribution. Ce terme se calcule d'après l'échantillon 2P2F par la formule :

$$N_{3P1F}^{\text{bkg}} = \sum \left(\frac{f_i}{1 - f_i} + \frac{f_j}{1 - f_j} \right) N_{2P2F}. \quad (5.9)$$

Le terme en N_{3P1F}^{ZZ} représente la contribution du signal ("4l") dans la région 3P1F : en effet, il est possible qu'un des leptons d'un vrai événement de signal ne passe pas les coupures finales. Cet événement se retrouve donc dans l'échantillon 3P1F, alors qu'il ne représente pas un vrai bruit de fond. Il faut donc soustraire cette contribution. Ce terme est le seul à être estimé à partir des simulations de signal, tous les autres étant estimés à partir des données. Notons que la contribution de ce terme est importante dans le calcul du bruit de fond. Nous devons donc supposer que les simulations décrivent correctement l'efficacité du détecteur pour les vrais leptons.

Les résultats finaux pour le calcul du bruit de fond sont présentés dans la table 5.1. Les incertitudes tiennent compte de la statistique limitée des échantillons 3P1F et 2P2F pour les données, et de l'incertitude sur f . Les incertitudes sur les différents termes de (5.2) sont additionnées quadratiquement. Une explication plus détaillée sur les détails du calcul de l'incertitude pour le bruit de fond se trouve à l'annexe B.

Les distributions de bruit de fond pour la masse invariante des quatre leptons se trouvent en figure 5.3. La statistique limitée de l'échantillon 3P1F ne permet pas de prendre des intervalles de masse invariante plus fins.

5.4 Conclusions concernant le bruit de fond

La contribution du bruit de fond est très faible, de l'ordre de quelques événements sur l'ensemble du spectre observé. Cette conclusion est compatible avec notre

Canal	N^{bkg}
$4e$	$-0,05 \pm 1,31$
4μ	$1,03 \pm 2,54$
$2e2\mu$	$0,94 \pm 3,48$
Total	$1,92 \pm 4,50$

TABLE 5.1 – Nombre d'événements de bruit de fond estimés dans les différents canaux.

intuition provenant de l'étude comprenant les simulations **MadGraph**. Le nombre d'événements total est compatible avec zéro, endéans les incertitudes associées.

La contribution légèrement négative dans l'état final $4e$ ne devrait pas choquer : en effet, elle provient probablement d'une légère surestimation de la contribution des événements de signal, et donc d'une petite différence entre données et simulations.

L'incertitude associée au bruit de fond est raisonnable par rapport à l'incertitude statistique qui est dominante pour notre analyse, comme nous le verrons au chapitre suivant. Nous pointons cependant deux éléments principaux qui contribuent à cette incertitude :

- la statistique sur les échantillons 3P1F et 2P2F est très limitée : on a moins de 20 événements par état final dans les données utilisées pour l'échantillon 3P1F, qui représente la contribution principale à (5.2). Cela donne évidemment une incertitude statistique importante ;
- l'incertitude sur le f est élevée, à cause de la statistique peu importante dans l'échantillon de contrôle utilisé, pour certaines régions cinématiques. En particulier, on dispose de peu de statistique à des hautes valeurs de quantité de mouvement transverse, alors que plusieurs événements 3P1F correspondent à cette cinématique. Les contributions à l'incertitude totale de ces événements peuvent devenir importantes.

Pour l'analyse de [3], le bruit de fond total obtenu était légèrement plus important en raison de la gamme cinématique plus étendue pour cette étude. Cependant, ce nombre d'événements restait très faible par rapport au nombre d'événements de signal. Nos résultats semblent donc cohérents avec les leurs. Notons que les valeurs de f dépendent fortement de la définition des coupures d'identification : d'une analyse à l'autre, cette valeur peut sensiblement changer.

Afin de vérifier le calcul de l'estimation du bruit de fond et d'éventuellement améliorer la précision de notre estimation, il pourrait être utile de trouver une méthode alternative pour estimer le bruit de fond, afin de combiner les résultats des deux méthodes. C'est ce qui est fait pour l'analyse de [3]. On pourrait également réaliser un test de cohérence pour notre résultat. Ces points n'ont cependant pas pu être abordés, à cause du temps limité dont nous disposons.

En guise de conclusion, le bruit de fond semble suffisamment faible pour pouvoir être simplement soustrait de la contribution des données lors du calcul de la section efficace, sans avoir besoin de passer par un ajustement paramétrique du bruit de fond et des données. Comme on constate que la contribution de chaque boîte de la figure 5.3 est compatible avec zéro, on supposera également que le bruit de fond ne

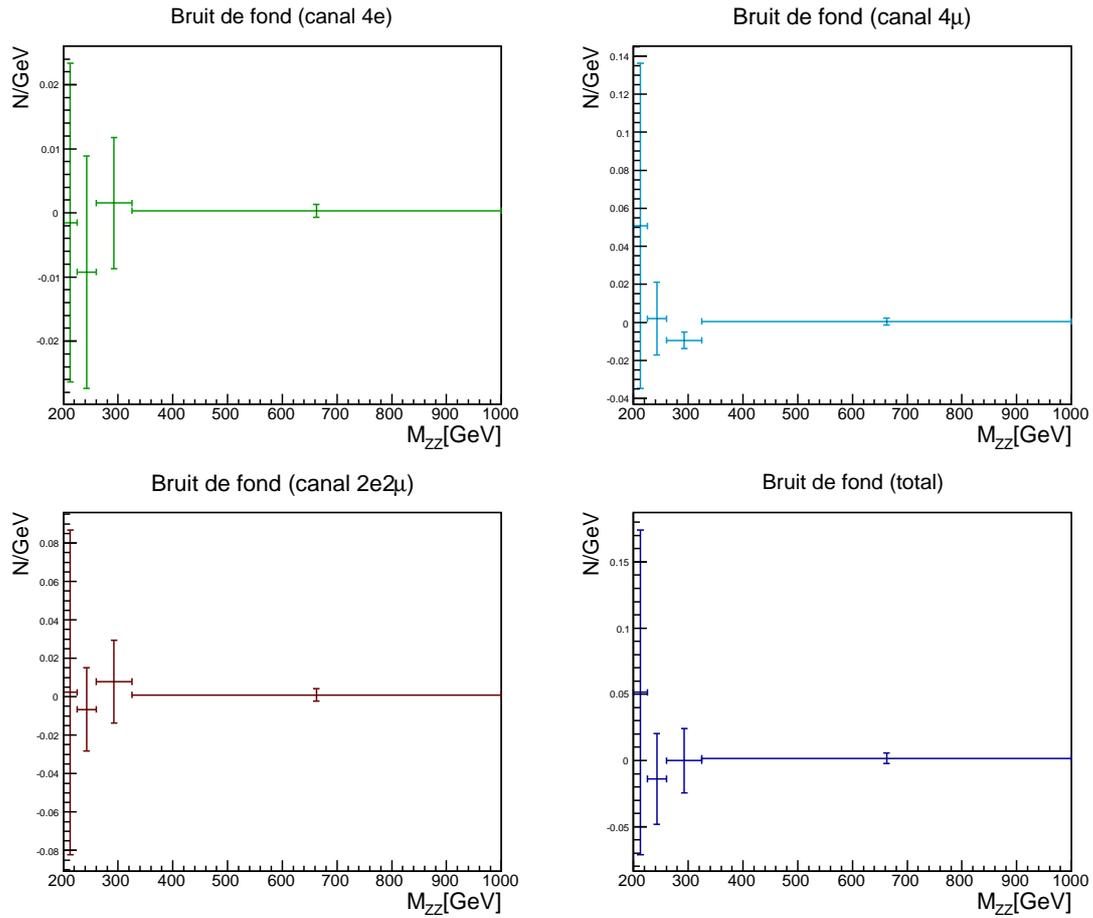


FIGURE 5.3 – Distributions en masse invariante M_{ZZ} de bruit de fond estimé, par canal : en haut à gauche, $4e$; en haut à droite, 4μ ; en bas à gauche, $2e2\mu$; en bas à droite, total sur les trois canaux précédents.

change pas la forme de la distribution, et que tous les tests d'ajustement par la suite pourront se faire sans tenir compte du bruit de fond. L'incertitude totale associée au bruit de fond sera traitée comme une source d'incertitude lors de l'estimation finale de la section efficace.

Chapitre 6

Résultats

Dans ce dernier chapitre, nous présentons enfin les résultats de notre analyse. Nous commençons par décrire le principe de notre méthode d'analyse. Ensuite, nous présentons nos distributions finales pour les variables étudiées. Ceci fait, nous ajustons les données aux simulations et vérifions la qualité de cet ajustement, afin de pouvoir tirer une conclusion sur la sensibilité de nos données aux corrections électrofaibles et de valider notre méthode de mesure de la section efficace. Nous discutons les différentes erreurs systématiques que nous avons dû prendre en compte dans notre calcul de section efficace. Enfin, nous présentons notre calcul de la section efficace du processus de production de paires de bosons Z à haute masse invariante.

6.1 Description de la méthode d'analyse

Au cours de notre analyse, nous avons deux objectifs principaux :

1. étudier la sensibilité des données aux corrections électrofaibles ;
2. calculer la section efficace inclusive de production de paires de bosons Z à plus de 200 GeV de masse invariante et sur leur couche de masse ($|M_{Z_{1,2}} - 91,2 \text{ GeV}| < 30 \text{ GeV}$), que nous appellerons par la suite $\sigma_{pp \rightarrow ZZ}$.

En l'absence de bruit de fond, l'estimateur du nombre d'événements de signal est immédiatement le nombre d'événements observés. C'est également vrai quand les événements observés sont répartis dans un histogramme, comme nous le montrons dans l'annexe C. Il suffit donc de normaliser nos simulations au nombre total d'événements observés pour faire notre ajustement. Ensuite, on peut comparer les spectres simulés, normalisés au nombre d'événements observés, aux spectres observés. Cette comparaison se fera sur base des distributions de deux variables qui nous semblent particulièrement discriminantes pour l'étude des corrections électrofaibles, comme montré au chapitre 2 : la masse invariante du système des deux bosons Z , M_{ZZ} , et la quantité de mouvement transverse d'un boson Z de l'événement : $p_{T,Z}$. Pour réaliser une comparaison objective des deux modèles, on doit se baser sur un test statistique pour calculer la bonté de l'ajustement. Les tests statistiques les plus fréquemment utilisés sont le test du χ^2 et celui du maximum de vraisemblance. Pour notre analyse, nous avons décidé d'implémenter le test du χ^2 , qui a des propriétés asymptotiques bien connues. Nous appliquerons ce test simultanément pour les trois

états finaux possibles : $4e$, 4μ et $2e2\mu$. Il permettra de tester les deux hypothèses, avec ou sans corrections électrofaibles, et nous permettra de tirer une conclusion sur la sensibilité aux corrections électrofaibles.

Pour le second point, si le test établi au premier point permet de conclure à un ajustement correct, nous procédons simplement par comptage du nombre d'événements total dans les données et dans les simulations. En effet, on peut montrer que pour le cas d'un signal sans bruit de fond (ce qui est le cas en première approximation), il n'est pas nécessaire de procéder à un ajustement entre les données et les simulations pour déterminer le facteur de normalisation et donc la section efficace. Cette preuve est donnée dans l'annexe C. Ensuite, en connaissant la section efficace correspondant à nos événements simulés et en calculant l'acceptance, ou la fraction de cette section efficace que l'on regarde pour notre analyse, on obtient la mesure de notre section efficace.

6.2 Distributions

Nous présentons ici les distributions obtenues pour notre échantillon final, après application de toutes les coupures décrites au chapitre 4. Ces distributions sont faites en supposant le bruit de fond nul, d'après les conclusions du chapitre 5 : les seules simulations employées sont celles du signal $ZZ \rightarrow 2l2l'$, y compris les contributions induites par une boucle de quarks. Nous regardons deux variables différentes : la masse invariante des quatre leptons formant les deux bosons Z , M_{ZZ} , et la quantité de mouvement $p_{T,Z}$ du système de deux leptons formant l'un des deux Z de l'événement, cette paire étant choisie au hasard parmi les deux bosons Z reconstruits afin d'éviter des biais systématiques. Les figures 6.1 et 6.2 présentent les distributions de ces deux variables, en regardant le modèle sans inclure les corrections électrofaibles ni la coupure sur ρ .

On constate un accord correct entre les données et les simulations dans les deux modèles. Remarquons que nous avons très peu d'événements à haute masse invariante et à haute quantité de mouvement transverse, c'est-à-dire dans les régions où les corrections électrofaibles sont importantes. Ce point indique l'importance d'augmenter la statistique d'événements dans ces régions, ce qui sera fait lors des runs suivants du LHC grâce à une augmentation de la luminosité instantanée.

La paramétrisation que nous avons choisie pour les figures 6.1 et 6.2 est utile car elle permet de visualiser la forme de la distribution, mais elle contient également de nombreux bins sans données, ce qui est problématique pour réaliser une étude statistique : en effet, l'approximation de prendre des erreurs poissoniennes en \sqrt{n} n'est pas valable pour de très petits nombres d'événements. Par conséquent, pour la suite de notre analyse, nous choisissons un découpage en boîtes différent, conçu pour avoir suffisamment d'événements dans chaque boîte pour faire un test statistique et comparer les modèles avec et sans corrections électrofaibles. Pour les deux variables étudiées, nous répartissons ainsi les événements en quatre boîtes. Ces distributions sont présentées à la figure 6.3 pour la variable M_{ZZ} et à la figure 6.4 pour la variable $p_{T,Z}$. Cette fois, on a appliqué la coupure sur ρ afin de pouvoir étudier les corrections électrofaibles : en effet, la supposition à la base de l'implémentation des corrections est l'absence de jets de gluons non colinéaires de grande quantité de mouvement

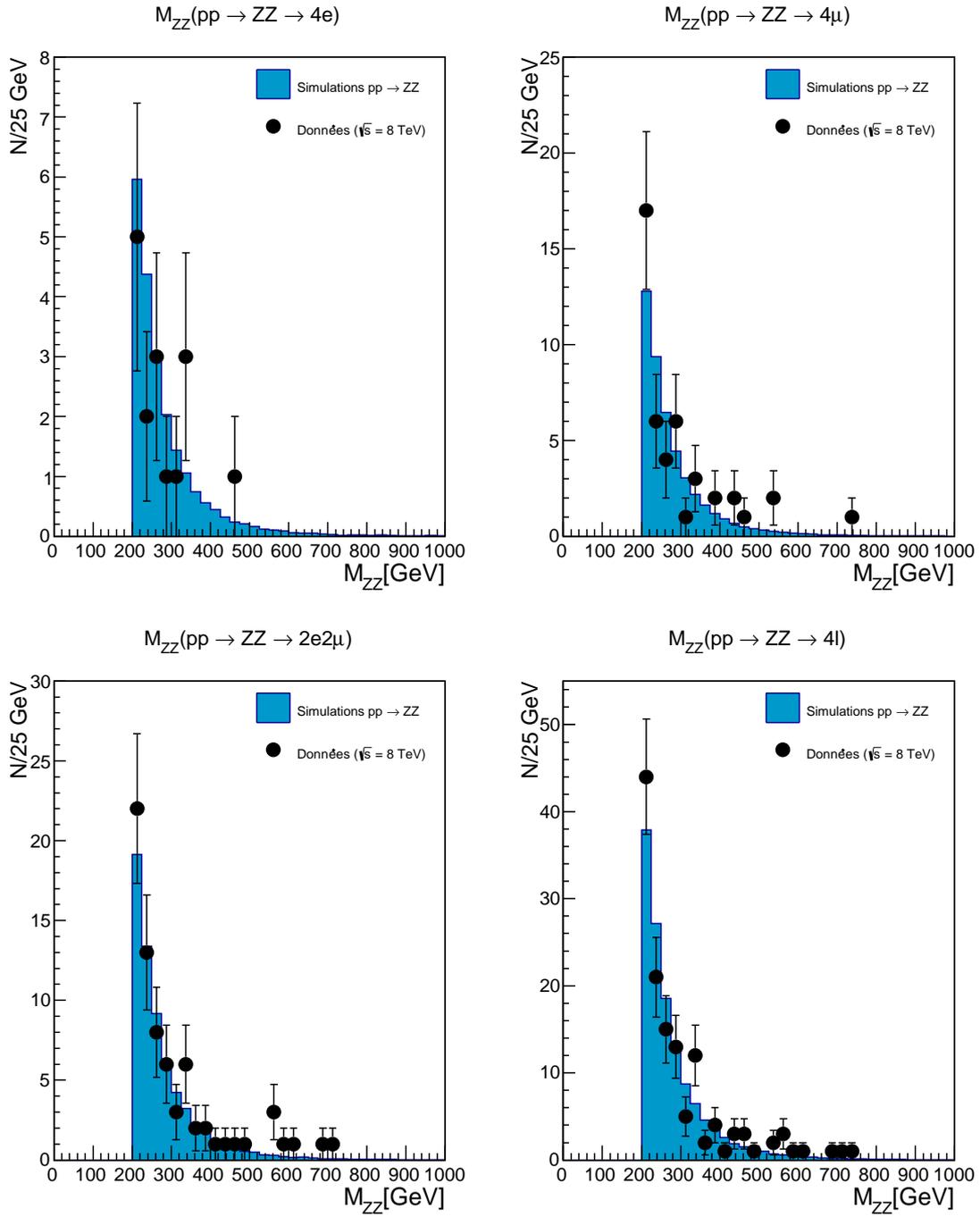


FIGURE 6.1 – Distributions de masse invariante M_{ZZ} du système des deux bosons Z , dans les trois états finaux $4e$, 4μ et $2e2\mu$ ainsi qu'au total, sans appliquer les corrections électrofaibles.

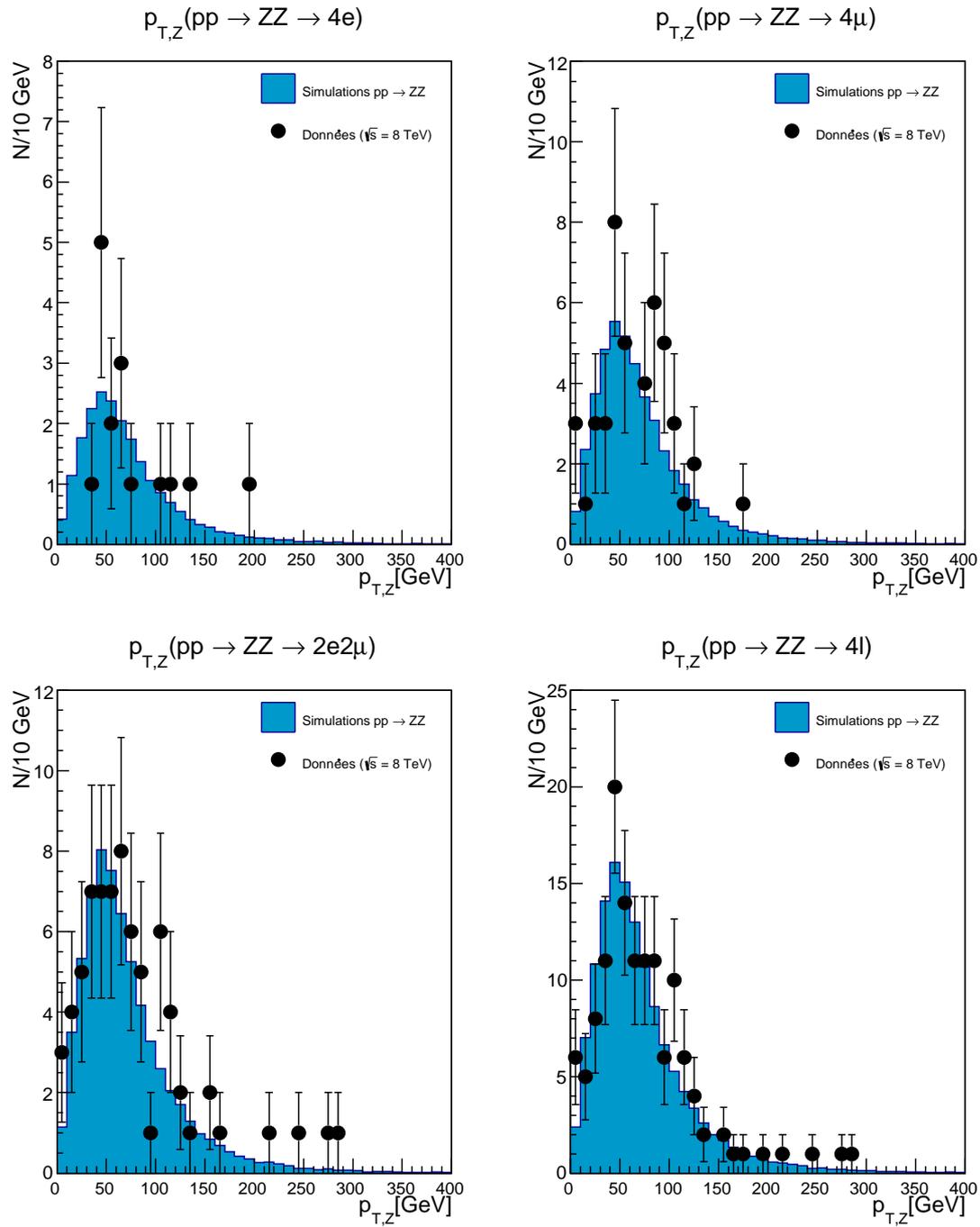


FIGURE 6.2 – Distributions de quantité de mouvement transverse $p_{T,Z}$ d'un boson Z reconstruit, dans les trois états finaux $4e$, 4μ et $2e2\mu$ ainsi qu'au total, sans appliquer les corrections électrofaibles.

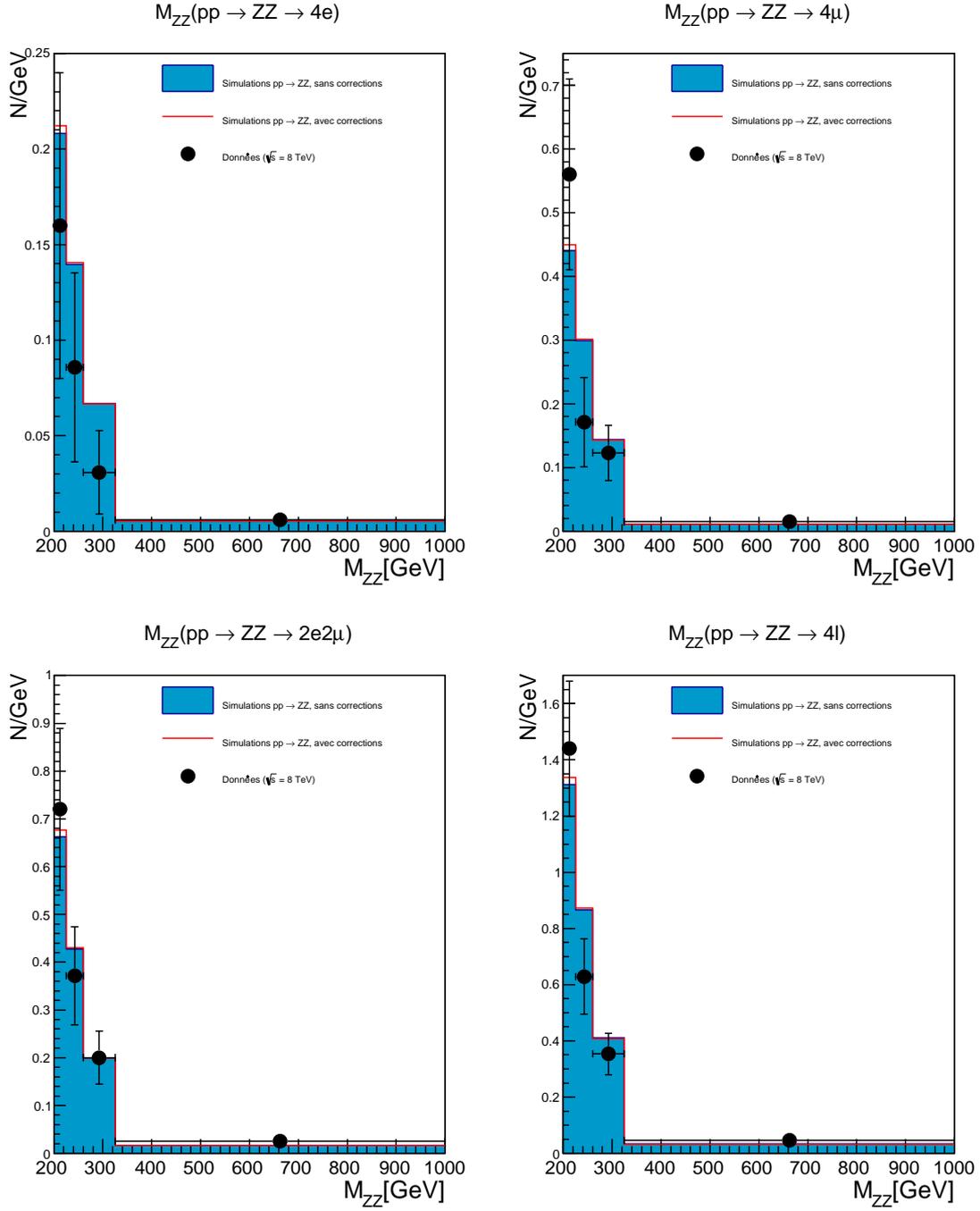


FIGURE 6.3 – Distributions de masse invariante M_{ZZ} du système des deux bosons Z , dans les trois états finaux $4e$, 4μ et $2e2\mu$ ainsi qu'au total, avec coupure sur la variable ρ , avec (ligne rouge) et sans corrections électrofaibles. Les distributions simulées sont normalisées au nombre total d'événements observés.

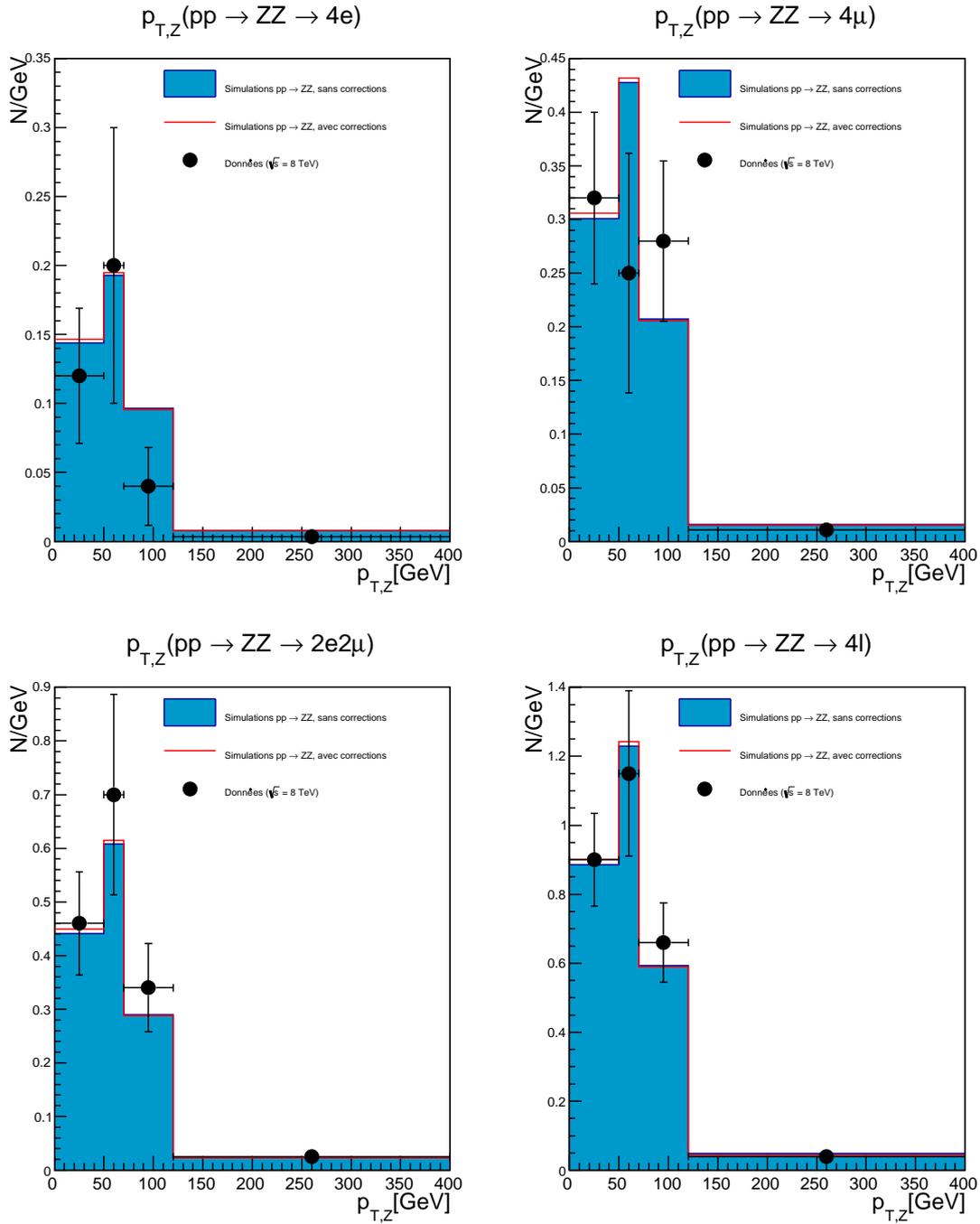


FIGURE 6.4 – Distributions de quantité de mouvement transverse $p_{T,Z}$ d'un boson Z reconstruit, dans les trois états finaux $4e$, 4μ et $2e2\mu$ ainsi qu'au total, avec coupure sur la variable ρ , avec (ligne rouge) et sans corrections électrofaibles. Les distributions simulées sont normalisées au nombre total d'événements observés.

transverse, et seule une application de la coupure sur la variable ρ permet de s'assurer que cette hypothèse est satisfaite. De plus, sur les figures 6.3 et 6.4, on a renormalisé les distributions de simulations, avec et sans corrections électrofaibles, au nombre total d'événements observés. En effet, l'ajustement est ici simplement réalisé en faisant cette normalisation.

6.3 Ajustement des données et sensibilité aux corrections électrofaibles

Nous calculons ici le χ^2 correspondant à l'ajustement des données sur les simulations tenant compte ou pas des corrections électrofaibles, en ayant appliqué la coupure sur ρ .

6.3.1 Test du χ^2

La définition du χ^2 est donnée par [30] :

$$\chi^2 \equiv \sum_{i=1}^k \left[\frac{N_i^{\text{obs}} - N_i^{\text{sim}}}{(\Delta N_i^{\text{obs}})} \right]^2, \quad (6.1)$$

où k est le nombre de boîtes utilisées dans nos histogrammes, N_i^{obs} le nombre d'événements observés dans la boîte i , N_i^{sim} le nombre d'événements simulés, attendus dans le bin i pour nos simulations et ΔN_i^{obs} l'erreur associée au nombre d'événements observés. Dans notre cas, nous supposons que les incertitudes sont purement poissonniennes : nous négligeons toute incertitude systématique affectant la forme de la distribution. Nous prenons donc $\Delta N_i^{\text{obs}} = \sqrt{N_i^{\text{obs}}}$. On ne tient pas compte ici du bruit de fond, en accord avec les conclusions avancées à la fin du chapitre 5.

Nous calculons le χ^2 pour les trois états finaux *simultanément*. En effet, nous disposons de simulations détaillées pour la cinématique et la sélection des trois états finaux, et nous savons que le paramètre d'intérêt, la section efficace, est le même pour les trois états finaux, la seule différence résidant dans les rapports de branchements qui sont très bien connus. Nous avons donc trois termes pour le calcul du χ^2 , chacun correspondant à un état final, et nous additionnons les contributions de chaque boîte des trois histogrammes. Comme nos histogrammes font 4 bins, nous avons $k = 12$ dans (6.1). Comme nous avons ajusté un paramètre, la normalisation des simulations, nous avons donc $12 - 1 = 11$ degrés de liberté.

6.3.2 Résultats du test avec et sans corrections

Les résultats du test de χ^2 , pour les deux variables étudiées et avec notre choix de mise en boîtes, sont repris dans la table 6.1.

Signalons que l'on n'a pas suffisamment de statistique pour que la variable soit vraiment distribuée comme une distribution χ^2 à onze degrés de liberté. Cependant, en allant consulter les tables de χ^2 pour 11 degrés de liberté (voir [30], p.130 par

χ^2	M_{ZZ}	$p_{T,Z}$
Sans corrections	11,65	10,74
Avec corrections	12,16	10,26

TABLE 6.1 – Résultats du test de χ^2 , en M_{ZZ} et en $p_{T,Z}$, avec et sans corrections électrofaibles.

exemple), on trouve :

$$\mathcal{P}(\chi^2 > 5,58) = 0,9 \quad (6.2)$$

$$\mathcal{P}(\chi^2 > 10,34) = 0,5 \quad (6.3)$$

$$\mathcal{P}(\chi^2 > 12,90) = 0,3. \quad (6.4)$$

Donc, toutes les valeurs obtenues, pour les deux variables, correspondent à des valeurs plausibles dans les deux modèles. Cela indique que l’ajustement est correct, et qu’il y a un sens à se baser sur les simulations pour calculer la section efficace.

6.3.3 Sensibilité aux corrections électrofaibles

La table 6.1 indique que les résultats du test sont très proches quel que soit le modèle utilisé, l’ajustement étant bon dans les deux cas. Pour la variable $p_{T,Z}$, l’accord est légèrement meilleur en appliquant les corrections électrofaibles que sans. Pour la variable M_{ZZ} , c’est le contraire, et l’accord est moins bon pour le modèle avec les corrections électrofaibles. On en conclut que l’on n’est pas sensible aux corrections électrofaibles avec les données utilisées pour notre analyse. En effet, les deux modèles présentent un accord comparable : ni l’un ni l’autre ne peuvent être rejetés.

Les mêmes conclusions peuvent être tirées de l’étude selon M_{ZZ} que de l’étude selon $p_{T,Z}$. Cela indique que les deux variables sont également discriminantes pour étudier les corrections électrofaibles. Le manque de statistique actuellement disponible pour des hautes valeurs de $p_{T,Z}$ nous empêche vraisemblablement d’être sensibles aux corrections importantes qui y sont présentes. Précisons qu’il n’y a aucun sens à comparer, de manière absolue, les valeurs de χ^2 obtenues en M_{ZZ} et en $p_{T,Z}$: les distributions ne sont pas basées sur la même distribution en boîtes, et il peut y avoir des biais systématiques différents.

En guise de conclusion, puisque les deux modèles donnent des résultats similaires, on pourra prendre indifféremment l’un ou l’autre pour le calcul de la section efficace. Par facilité, nous prendrons le modèle sans inclure les corrections électrofaibles. Ce choix nous permet également de nous affranchir de la coupure sur ρ par la suite, puisqu’on ne tiendra plus du tout compte des corrections électrofaibles.

6.4 Incertitudes systématiques

Avant de passer au calcul de la section efficace, il est tout d’abord nécessaire d’estimer les différentes contributions d’erreurs systématiques qui s’y appliquent.

Ces erreurs systématiques comprennent tous les effets qui ne sont pas encodés dans le nombre d'événements observés, et peuvent avoir des sources très diverses. Nous en faisons ici la liste.

Les différentes incertitudes systématiques considérées sont :

- l'incertitude sur la luminosité reçue par le détecteur CMS lors de la prise de données en 2012. Cette incertitude a été mesurée par la collaboration CMS [31] et est la même pour toutes les analyses. Elle est de 2,6 % ;
- l'incertitude sur l'efficacité de sélection ε des leptons. L'erreur relative sur cette efficacité de sélection est estimée à 1 % par lepton par la collaboration CMS. Comme la section efficace est inversement proportionnelle à la quatrième puissance de ε ,

$$\sigma \propto \frac{N^{\text{obs}}}{\varepsilon^4}, \quad (6.5)$$

car on a quatre leptons sélectionnés, l'erreur relative due à l'efficacité totale de sélection pour les quatre leptons est de

$$\left. \frac{\Delta\sigma}{\sigma} \right|_{\varepsilon} = \left| \frac{\partial\sigma}{\partial\varepsilon} \right| \frac{\Delta\varepsilon}{\sigma} = 4 \cdot \frac{\Delta\varepsilon}{\varepsilon}. \quad (6.6)$$

On prend donc cette incertitude égale à 4 % ;

- comme annoncé au chapitre 1, le processus $qq' \rightarrow ZZ + \text{jets}$ n'étant pas simulé pour notre analyse, nous mettons une incertitude conservative de 2 % pour tenir compte de cette contribution manquante ;
- l'incertitude sur l'efficacité de sélection des triggers. Elle est prise à 2 %, comme établi au chapitre 4 ;
- les incertitudes théoriques. Les contributions dominantes à ces incertitudes sont :
 1. le choix de la paramétrisation des fonctions de densité des partons dans le proton ;
 2. les ordres supérieurs en α_S manquants dans la simulation du processus de production de paires de bosons Z .

Ces incertitudes sont délicates à évaluer, et les inclure dans un test statistique n'est pas évident. Nous nous sommes donc résolus à utiliser les incertitudes données par [3], en faisant l'hypothèse que ces incertitudes ne changent pas la forme de la distribution. Ces incertitudes sont prises à 4 % ;

- l'incertitude sur le bruit de fond. Estimée au chapitre 5, elle sera également propagée dans le calcul final, comme nous le verrons à la section suivante.

6.5 Calcul de $\sigma_{pp \rightarrow ZZ}$

Nous pouvons maintenant passer au calcul de la section efficace de production de paires de bosons Z . Ce calcul se fera en deux étapes : d'abord, on calculera l'acceptance de notre analyse afin d'avoir les sections efficaces simulées, puis on extraira la section efficace sur base du rapport entre données et simulations.

6.5.1 Acceptance de l'analyse

Dans notre cas, nous ne voulons calculer la section efficace que dans la région cinématique qui nous intéresse, à savoir des bosons Z tels que $|M_{Z_{1,2}} - 91,2 \text{ GeV}| < 30 \text{ GeV}$ et $M_{ZZ} < 200 \text{ GeV}$. On ne regardera donc de toute façon qu'une fraction des événements simulés par nos échantillons Monte-Carlo. Cette fraction est appelée l'*acceptance*. Pour notre mesure, c'est cette fraction qui sera utilisée.

Pour calculer l'acceptance, on compte la fraction d'événements qui passent les deux coupures ci-dessus au niveau généré, dans les cinq échantillons de simulations dont on dispose. La table 6.2 reprend les acceptances pour chaque échantillon de simulations, ainsi que la section efficace corrigée pour ce canal.

Simulation	Acceptance	Section efficace dans l'acceptance [fb]
POWHEG $pp \rightarrow ZZ \rightarrow 4e$	0,079	6,05
POWHEG $pp \rightarrow ZZ \rightarrow 4\mu$	0,078	6,09
POWHEG $pp \rightarrow ZZ \rightarrow 2e2\mu$	0,070	12,32
gg2vv $gg \rightarrow ZZ \rightarrow 4e/4\mu$	0,137	0,66
gg2vv $gg \rightarrow ZZ \rightarrow 2e2\mu$	0,072	0,86

TABLE 6.2 – Acceptance et sections efficaces corrigées pour chaque échantillon de simulations.

La section efficace simulée totale pour tous les états finaux est simplement prise comme la somme sur tous les échantillons. Elle vaut donc 25,97 fb. Cependant, nous voulons calculer une section efficace inclusive pour tous les états finaux, et pas uniquement pour les états finaux à quatre leptons chargés (électrons ou muons). Il faut donc diviser ce nombre par le rapport de branchement de chaque boson Z en deux leptons chargés (e^+e^- ou $\mu^+\mu^-$) : ce rapport vaut 0,0045 comme montré en section 1.1. La section efficace simulée, correspondant à notre modèle, est donc :

$$\sigma_{pp \rightarrow ZZ}^{\text{sim}} = 5,731 \text{ pb.} \quad (6.7)$$

6.5.2 Détermination de la section efficace

La section efficace totale de production de paires de bosons Z , dans notre acceptance, se calcule comme :

$$\sigma_{pp \rightarrow ZZ}^{\text{mes}} = \mu \cdot \sigma_{pp \rightarrow ZZ}^{\text{sim}}, \quad (6.8)$$

où μ est la force relative du signal : il s'agit simplement du rapport entre le nombre d'événements observés, auquel on a soustrait le bruit de fond, et le nombre d'événements attendus d'après nos simulations :

$$\mu = \frac{N^{\text{obs}} - N^{\text{bkg}}}{N^{\text{sim}}}. \quad (6.9)$$

Une démonstration rigoureuse de (6.8), sur base du maximum de vraisemblance, se trouve dans l'annexe C.

L'incertitude statistique sur μ est prise comme poissonnienne sur N^{obs} :

$$\Delta\mu|_{\text{stat}} = \frac{\sqrt{N^{\text{obs}}}}{N^{\text{sim}}}. \quad (6.10)$$

L'incertitude liée au bruit de fond est simplement prise d'après notre estimation du bruit de fond, présentée dans la table 5.1.

Les nombres d'événements totaux dans les données et les simulations, ainsi que le bruit de fond, pour les trois états finaux, sont donnés dans la table 6.3.

État final	N^{sim}	N^{bkg}	N^{obs}
$4e$	21,16	$-0,05 \pm 1,31$	16
4μ	45,25	$1,03 \pm 2,54$	45
$2e2\mu$	65,02	$0,94 \pm 3,48$	73
Total	131,43	$1,92 \pm 4,50$	134

TABLE 6.3 – Nombres d'événements après coupures de sélection finales, dans les données, simulations et par le calcul du bruit de fond.

La valeur estimée de μ est donc :

$$\mu = 1,005 \pm 0,088 \text{ (stat)} \pm 0,034 \text{ (bkg)}. \quad (6.11)$$

En tenant compte de toutes les erreurs systématiques listées en section 6.4, on obtient ainsi finalement notre mesure de section efficace dans notre région d'acceptance :

$$\sigma_{pp \rightarrow ZZ} = (5,76 \pm 0,51 \text{ (stat)} \pm 0,20 \text{ (bkg)} \pm 0,15 \text{ (lumi)} \pm 0,26 \text{ (eff)} \pm 0,26 \text{ (theo)}) \text{ pb}, \quad (6.12)$$

où par "stat" on désigne l'incertitude statistique, par "bkg" l'incertitude sur le bruit de fond, par "lumi" l'incertitude sur la luminosité, par "eff" l'incertitude liée à l'efficacité de sélection et sur l'efficacité des triggers, et par "theo" l'incertitude sur le choix de fonctions de distribution partoniques, les ordres manquants du calcul théorique et le processus $qq' \rightarrow ZZ$ non simulé.

En combinant toutes les incertitudes en une seule, quadratiquement, on obtient :

$$\sigma_{pp \rightarrow ZZ} = (5,76 \pm 0,68) \text{ pb}, \quad (6.13)$$

soit une précision d'environ 11 %.

Cette valeur peut être comparée aux valeurs théoriques données en section 1.2 : nous avons estimé que la section efficace prédite dans notre région d'acceptance était d'environ 5,1 pb. Notre valeur mesurée est en accord avec cette prédiction, endéans un écart de 1σ .

L'incertitude dominante est statistique. De plus, on peut également considérer l'incertitude liée au bruit de fond comme d'origine statistique, d'après notre méthode d'estimation. Ceci nous conduit à penser que les prochains runs de prise de données, avec augmentation de la luminosité et donc de la statistique disponible, permettront d'améliorer notablement la précision sur cette mesure.

Conclusions et perspectives

Dans ce mémoire, nous avons étudié la production de paires de bosons Z à haute masse invariante dans des collisions proton-proton à 8 TeV auprès de l'expérience CMS, en étudiant plus en détail l'impact des corrections électrofaibles à ce processus.

Une étude des corrections électrofaibles au niveau généré nous a permis de constater que ces corrections sont négatives, et d'environ -4% sur l'ensemble des événements. Cependant, nous avons également remarqué que l'étude de deux variables particulières, la masse invariante de la paire de bosons Z et la quantité de mouvement transverse d'un boson Z de l'événement, permet de distinguer des zones cinématiques où ces effets peuvent devenir importants. En appliquant une paramétrisation de ces corrections, nous avons reproduit les résultats de la littérature. En particulier, nous avons observé qu'à 500 GeV de masse invariante M_{ZZ} , ces corrections devenaient de l'ordre de -10% , et qu'elles s'élevaient même à près de -25% à 350 GeV de quantité de mouvement transverse $p_{T,Z}$.

Nous avons vérifié l'accord entre les données et les simulations, après avoir appliqué nos coupures de sélection. Ensuite, nous avons estimé l'efficacité de déclenchement de l'expérience CMS pour notre état final avec $M_{ZZ} > 200$ GeV et conclu qu'elle était compatible avec 100% .

Ensuite, nous avons pu estimer le bruit de fond associé à notre processus. Le principe de la méthode est connu : il s'agit de mesurer le taux d'erreur d'identification de jets identifiés comme des leptons dans un échantillon de contrôle, puis d'utiliser ce taux pour estimer le bruit de fond dû aux processus avec seulement deux ou trois leptons. Nous avons cependant dû tenir compte d'effets non documentés, comme la contamination des échantillons de contrôle pour l'étude du taux d'erreur d'identification par des vrais leptons provenant des processus WZ et ZZ . Le bruit de fond est compatible avec zéro endéans les incertitudes de mesure. Il a été soustrait globalement lors du calcul de la section efficace. En raison de sa faible contribution, le spectre de bruit de fond a été négligé lors du test de compatibilité.

Une comparaison des spectres en M_{ZZ} et en $p_{T,Z}$ entre les données et les simulations de signal nous ont permis de conclure à un accord correct entre données et simulations, rendant possible un calcul de section efficace. Cependant, cette analyse nous a aussi révélé que les données actuelles ne nous permettent pas de trancher entre le modèle comprenant les corrections électrofaibles et celui ne les incluant pas. Par simplicité, nous avons donc réalisé le calcul de la section efficace sur le modèle sans corrections électrofaibles.

La mesure de la section efficace obtenue est en accord avec la littérature. La source d'incertitude principale sur cette section efficace est d'origine statistique. Ce point nous conforte dans l'idée que cette mesure s'améliorera lors des prochains runs

du LHC, au cours desquels davantage de données seront prises.

De nombreuses extensions et perspectives sont possibles pour notre travail.

Premièrement, il est clair que la même analyse pourra être répétée avec les données à 13 TeV, qui devraient commencer à être récoltées dans peu de temps. L'augmentation de l'énergie et de la luminosité instantanée pour ce deuxième run devraient donner davantage de statistique à haute masse invariante et à haute quantité de mouvement transverse, et pourraient peut-être permettre une sensibilité aux corrections électrofaibles, ce qui n'était pas le cas pour les données à 8 TeV.

Cependant, même en restant avec les données à 8 TeV, il est possible d'améliorer cette analyse.

Concernant le bruit de fond, il pourrait être utile d'implémenter une autre méthode indépendante d'estimation puis de combiner les résultats avec ceux obtenus par notre méthode, afin de réduire l'erreur sur ce bruit de fond qui est élevée dans notre analyse. Différents tests de cohérence pourraient également être mis en place pour vérifier notre méthode. Cela n'a pas pu être réalisé par manque de temps.

L'incertitude sur la forme des distributions en M_{ZZ} et en $p_{T,Z}$ pourrait être complétée. En particulier, les corrections QCD différentielles au-delà du NLO viennent d'être publiées [32]. Elles peuvent changer les spectres, et devraient donc être étudiées et éventuellement incluses dans le modèle.

Finalement, l'ouverture principale de ce travail devrait être de se servir de l'estimation de cette section efficace et des distributions en M_{ZZ} et en $p_{T,Z}$ pour l'étude de la largeur du scalaire ou pour tenter de trouver de la nouvelle physique dans un canal avec production de paires de bosons Z .

Annexe A

Contributions des différents types de quarks

Dans cette annexe, nous démontrons explicitement l'expression (2.1), et nous donnons une application numérique de l'effet des couplages vectoriels sur les contributions entre quarks d et quarks u pour le cas du processus de Drell-Yan.

Suite au mécanisme pour l'unification électrofaible, on sait que le lagrangien contient les termes de courants neutres :

$$\mathcal{L} \ni -igJ_\mu^3 W^{3\mu} - i\frac{g'}{2}j_\mu^Y B^\mu, \quad (\text{A.1})$$

où

$$J_\mu^3 \equiv \frac{1}{2}\bar{\psi}\gamma_\mu\tau^3\psi = \bar{\psi}\gamma_\mu\frac{1-\gamma_5}{2}T^3\psi \quad (\text{A.2})$$

(où T^3 est l'isospin faible) et

$$j_\mu^Y \equiv Q^Y\bar{\psi}\gamma_\mu\psi \quad (\text{A.3})$$

sont les courants associés respectivement aux groupes SU(2) et U(1). Suite à l'unification électrofaible, on introduit le mélange suivant :

$$A_\mu = B_\mu \cos \theta_W + W_\mu^3 \sin \theta_W \quad (\text{A.4})$$

$$Z_\mu = -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W, \quad (\text{A.5})$$

en termes du paramètre θ_W . On peut dès lors réécrire (A.1) comme :

$$\mathcal{L} \ni -i \left(g \sin \theta_W J_\mu^3 + g' \cos \theta_W \frac{j_\mu^Y}{2} \right) A^\mu - i \left(g \cos \theta_W J_\mu^3 - g' \sin \theta_W \frac{j_\mu^Y}{2} \right) Z^\mu. \quad (\text{A.6})$$

En identifiant l'expression devant A^μ avec les interactions électromagnétiques, en sachant que $j^{em} = J_\mu^3 + \frac{1}{2}j_\mu^Y$, on impose

$$g \sin \theta_W = g' \cos \theta_W = e, \quad (\text{A.7})$$

donc

$$\tan \theta_W = \frac{g'}{g}. \quad (\text{A.8})$$

La deuxième partie de (A.6) concernant les courants neutres peut donc se réécrire :

$$\mathcal{L}_{CN} = -ig \cos \theta_W J_\mu^3 Z^\mu + ig \tan \theta_W \sin \theta_W (j_\mu^{em} - J_\mu^3) Z^\mu \quad (\text{A.9})$$

$$= \frac{-ig}{\cos \theta_W} (J_\mu^3 - \sin^2 \theta_W j_\mu^{em}) Z^\mu, \quad (\text{A.10})$$

ou encore par (A.2) et en se souvenant que $j_\mu^{em} = Q\bar{\psi}\gamma_\mu\psi$:

$$\mathcal{L}_{CN} = \frac{-ig}{\cos \theta_W} \bar{\psi}\gamma^\mu \left[\frac{1}{2} (1 - \gamma_5) T^3 - \sin^2 \theta_W Q \right] \psi Z_\mu. \quad (\text{A.11})$$

On constate donc que l'on peut écrire le vertex d'interaction $Zq\bar{q}$ comme :

$$\frac{-ig}{\cos \theta_W} \gamma^\mu \frac{1}{2} (C_V^q - C_A^q \gamma^5), \quad (\text{A.12})$$

en définissant les couplages vectoriel et axial :

$$C_V^q = T_q^3 - 2 \sin^2 \theta_W Q^q \quad (\text{A.13})$$

$$C_A^q = T_q^3. \quad (\text{A.14})$$

La mesure de l'angle de mélange θ_W donne $\sin^2 \theta_W = 0,231$ [6] ; par conséquent, on peut calculer les couplages vectoriel et axial associés aux quarks u et d :

$$C_V^u = +\frac{1}{2} - 2 \cdot \left(\frac{+2}{3}\right) \cdot 0,231 = +0,192, \quad C_A^u = +\frac{1}{2} \quad (\text{A.15})$$

$$C_V^d = -\frac{1}{2} - 2 \cdot \left(\frac{-1}{3}\right) \cdot 0,231 = -0,346, \quad C_A^d = -\frac{1}{2}. \quad (\text{A.16})$$

On constate donc que le couplage vectoriel du quark d est supérieur en valeur absolue à celui du quark u .

Le calcul de la section efficace du processus $q\bar{q} \rightarrow ZZ$, sort du cadre de ce mémoire ; mais on peut avoir une idée du comportement de ce processus en les variables C_V et C_A en regardant le processus de Drell-Yan : $q\bar{q} \rightarrow Z/\gamma \rightarrow e^+e^-$ près de la masse du Z . Dans ce cas, la section efficace différentielle selon l'angle entre les particules initiales est connue analytiquement [7] :

$$\frac{d\sigma}{d\cos\theta} = \frac{\pi\alpha^2 k^2}{2\Gamma_Z^2} \left[(1 + \cos^2\theta)(C_A^{e^2} + C_V^{e^2})(C_A^{q^2} + C_V^{q^2}) + 8\cos\theta C_A^e C_V^e C_A^q C_V^q \right], \quad (\text{A.17})$$

où $k = \frac{\sqrt{2}G_F M_Z^2}{4\pi\alpha} \approx 1,4$. Comme les facteurs $(C_A^{q^2} + C_V^{q^2})$ et $(C_A^q C_V^q)$ sont plus grands pour les quarks d que pour les quarks u , la contribution totale est plus importante pour les quarks d que pour les quarks u , même si cette contribution est contrebalancée par les fonctions de densité partoniques qui favorisent les quarks u .

Par ailleurs, lorsqu'on intègre (A.17) pour $\cos\theta \in [-1,1]$, le dernier terme s'annule et les couplages apparaissent au carré.

Annexe B

Calcul de l'incertitude sur le bruit de fond

L'incertitude associée au bruit de fond, comme calculé au chapitre 5, demande un traitement particulièrement soigneux. En effet, le calcul d'erreur sert à tirer une conclusion finale sur l'estimation du bruit de fond. La difficulté est que l'incertitude associée à chaque événement est différente, à cause de la distribution du taux d'erreur de sélection des leptons f : on doit donc mélanger des erreurs systématiques et statistiques.

Tout d'abord, l'erreur sur f a elle-même deux contributions : la première vient de la statistique sur le nombre d'événements présents dans l'échantillon $Z + 1l$ utilisé pour calculer f , et la deuxième de l'incertitudes sur les sections efficaces WZ et ZZ , utilisées pour corriger le calcul.

L'erreur statistique sur T , due aux données, est prise pour une loi binomiale de probabilité f et de nombre d'essais L . Par les propriétés de la binomiale, et en se souvenant que $f = \frac{T}{L}$, on a :

$$\text{Var } T = L \cdot f \cdot (1 - f) \quad (\text{B.1})$$

$$\Rightarrow \Delta T = \sqrt{L \cdot f \cdot (1 - f)} \quad (\text{B.2})$$

$$= \sqrt{T \cdot (1 - f)} \quad (\text{B.3})$$

$$\Rightarrow \Delta_{\text{stat}} f = \frac{1}{L} \cdot \sqrt{T \cdot (1 - f)}. \quad (\text{B.4})$$

L'erreur due aux contributions WZ et ZZ est prise à 10 % du nombre d'événements pour ces deux canaux. Cette erreur est additionnée quadratiquement à l'erreur statistique. Cela nous donne l'erreur Δf .

Pour déterminer le bruit de fond, en se basant sur la formule (5.2), chaque événement k de l'échantillon 3P1F reçoit une pondération, que nous appellerons α_k , égale à

$$\alpha_{k,3\text{P1F}} = \frac{f_i}{1 - f_i} \cdot n, \quad (\text{B.5})$$

où f_i est le taux d'erreur de sélection associé au lepton "fail" de l'événement, et $n = 1$. L'erreur associée à α a deux contributions : une erreur statistique, liée au nombre d'événements, et une erreur systématique liée à f_i . L'incertitude sur le

nombre total d'événements pondéré par f dans la région 3P1F est prise comme la somme quadratique des incertitudes associées à chaque événement :

$$\delta N_{(3P1F)}^{\text{bkg}} = \sqrt{\sum_{k=1}^{N_{\text{tot},3P1F}} (\alpha_k)^2}. \quad (\text{B.6})$$

La contribution statistique pour chaque événement vient de l'incertitude sur la variable n de (B.5), qui vaut 1 ± 1 . En effet, si on ne prenait pas en compte f , on aurait bien une incertitude totale qui vaudrait $\sqrt{\sum_1^{N_{\text{tot},3P1F}} (\delta n)^2} = \sqrt{N_{\text{tot},3P1F}}$, ce qui est bien l'incertitude d'une loi de Poisson. On a donc "décomposé" notre statistique de Poisson pour chaque événement. La contribution de f est connue indépendamment, et a été tabulée selon la saveur et la quantité de mouvement transverse du lepton, ainsi que par la région du détecteur. L'incertitude sur α_k se calcule en additionnant quadratiquement les erreurs dues aux différents facteurs :

$$\Delta \alpha_{k,3P1F} = \sqrt{\left(\frac{\partial \alpha_{k,3P1F}}{\partial f_i} \Delta f_i\right)^2 + \left(\frac{\partial \alpha_{k,3P1F}}{\partial n} \Delta n\right)^2} \quad (\text{B.7})$$

$$= \sqrt{\left(\frac{1}{1-f_i} + \frac{f_i}{(1-f_i)^2}\right)^2 \Delta f_i^2 + \frac{f_i^2}{(1-f_i)^2} \cdot 1} \quad (\text{B.8})$$

$$= \sqrt{\frac{\Delta f_i^2}{(1-f_i)^4} + \frac{f_i^2}{(1-f_i)^2}} \quad (\text{B.9})$$

$$= \frac{\sqrt{\Delta f_i^2 + f_i^2 (1-f_i)^2}}{(1-f_i)^2}. \quad (\text{B.10})$$

L'incertitude dans la région 2P2F suit la même ligne de raisonnement. Il y a cependant une différence importante dans le calcul. En effet, dans la pondération de chaque événement de l'échantillon 2P2F, il faut tenir compte non seulement du terme N_{2P2F} de (5.2), mais aussi de N_{3P1F}^{bkg} , qui se calcule sur le même échantillon. Par conséquent, le facteur de pondération pour chaque événement de l'échantillon 2P2F est, en tenant compte de ces deux contributions :

$$\alpha_{k,2P2F} = \frac{f_i}{1-f_i} \cdot \frac{f_j}{1-f_j} \cdot n - \frac{f_j}{1-f_j} \cdot \left(\frac{f_i}{1-f_i} + \frac{f_j}{1-f_j}\right) \cdot n \quad (\text{B.11})$$

$$= -\frac{f_j^2}{(1-f_j)^2} \cdot n, \quad (\text{B.12})$$

où f_i et f_j sont les valeurs du taux d'erreur de sélection associées respectivement au premier et au deuxième lepton "fail", l'ordre étant fait par valeurs décroissantes de p_T . Remarquons qu'on a fait le choix (arbitraire) d'utiliser le lepton de plus bas p_T pour calculer le taux utilisé dans N_{3P1F}^{bkg} . Le calcul de l'incertitude associée devient :

$$\Delta \alpha_{k,2P2F} = \sqrt{\left(\frac{\partial \alpha_{k,2P2F}}{\partial f_i} \Delta f_i\right)^2 + \left(\frac{\partial \alpha_{k,2P2F}}{\partial f_j} \Delta f_j\right)^2 + \left(\frac{\partial \alpha_{k,2P2F}}{\partial n} \Delta n\right)^2} \quad (\text{B.13})$$

$$= \sqrt{0 + \left(\frac{\partial \alpha_{k,2P2F}}{\partial f_j} \Delta f_j\right)^2 + (\alpha_{k,2P2F})^2}, \quad (\text{B.14})$$

puisque l'expression finale de $\Delta\alpha_{k,2P2F}$ ne dépend plus de f_i . Un simple calcul permet dès lors de montrer que l'erreur sur le facteur de pondération pour chaque événement issu de la région 2P2F vaut :

$$\Delta\alpha_{k,2P2F} = \frac{f_j}{(1-f_j)^2} \sqrt{\frac{4\Delta f_j^2}{(1-f_j)^2} + f_j^2}. \quad (\text{B.15})$$

Notons finalement que nous devons encore multiplier ces facteurs de pondération et les erreurs qui leur sont associées par la correction selon le nombre de vertex discutée à la section 4.3. Nous négligeons dans ce cas l'incertitude sur cette correction.

Annexe C

Maximum de vraisemblance pour la section efficace

Nous montrons ici que le paramètre μ , défini comme la force relative du signal que l'on veut estimer, peut s'obtenir directement par comptage des nombres d'événements totaux dans les données et les simulations, en première approximation, c'est-à-dire en supposant le bruit de fond très petit.

Nous faisons ici l'hypothèse que nous avons un signal seul, *sans bruit de fond* pour commencer, et que la force de ce signal est μ . Nous cherchons à estimer le paramètre μ . Pour ce faire, nous utilisons la méthode du maximum de vraisemblance, décrite par exemple dans [30]. La définition de la *vraisemblance* L est :

$$L(\text{data}|H_\mu) = \prod_{i=1}^N f(n_i|H_\mu), \quad (\text{C.1})$$

où H_μ désigne l'hypothèse, *composée* dans notre cas, que le signal est distribué selon une force μ . N est le nombre d'expériences indépendantes : dans notre cas, c'est le nombre de boîtes dans nos distributions (on suppose que les distributions d'événements dans chaque boîte sont indépendantes, ce qui est en principe le cas). La fonction de distribution f est la densité de probabilité d'obtenir n_i événements dans la boîte i en supposant le signal d'une force μ .

Nous nous plaçons dans un certain découpage en boîtes. Le nombre d'événements dans chaque boîte i est décrit par une distribution de probabilité poissonnienne, dont le paramètre est donné par le produit du nombre d'événements de signal attendus dans cette boîte, déterminé d'après les simulations, et du paramètre μ que l'on cherche à estimer :

$$f(n_i|H_\mu) = e^{-\mu s_i} \frac{(\mu s_i)^{n_i}}{n_i!}. \quad (\text{C.2})$$

La fonction de vraisemblance vaut donc :

$$L(\mu) = \prod_{i=1}^N e^{-\mu s_i} \frac{(\mu s_i)^{n_i}}{n_i!} \quad (\text{C.3})$$

$$= e^{-\mu \sum_{i=1}^N s_i} \cdot \mu^{\sum_{i=1}^N n_i} \cdot \prod_{i=1}^N \frac{s_i^{n_i}}{n_i!} \quad (\text{C.4})$$

$$= e^{-\mu s} \cdot \mu^n \cdot \prod_{i=1}^N \frac{s_i^{n_i}}{n_i!}, \quad (\text{C.5})$$

en appelant $n \equiv \sum_{i=1}^N n_i$ le nombre total d'événements observés et $s \equiv \sum_{i=1}^N s_i$ le nombre total d'événements simulés. Par simplicité de calcul, on prend le logarithme de la vraisemblance :

$$\log L(\mu) = -\mu s + n \log \mu + \sum_{i=1}^N \log \left(\frac{s_i^{n_i}}{n_i!} \right). \quad (\text{C.6})$$

Notons que le terme avec la somme est indépendant de μ . La meilleure estimation du paramètre est donnée par le maximum de la vraisemblance, ou de manière équivalente par le maximum de log-vraisemblance (la fonction logarithme étant croissante). Nous devons donc prendre le ou les extrema de $L(\mu)$:

$$\frac{\partial(\log L)}{\partial \mu} = -s + \frac{n}{\mu} \stackrel{!}{=} 0 \quad (\text{C.7})$$

$$\Rightarrow \frac{n}{\mu} = s \quad (\text{C.8})$$

$$\Rightarrow \mu = \frac{n}{s}. \quad (\text{C.9})$$

Nous constatons donc que l'estimation de μ est indépendante du découpage en boîtes si on suppose le bruit de fond nul : il suffit de compter le nombre d'événements dans les données et dans les simulations. Notons cependant que ce résultat n'indique en rien si l'accord entre les données et les simulations est correct, autrement dit si on a raison de prendre un modèle en particulier pour calculer s : ceci doit être vérifié indépendamment.

Dans le cas général incluant du bruit de fond, il n'est pas possible de s'affranchir du découpage en boîtes. Cependant, dans notre cas, comme notre bruit de fond est très petit, nous ferons l'approximation que nous pouvons simplement le soustraire du nombre d'événements de données :

$$\mu = \frac{n - b}{s}, \quad (\text{C.10})$$

où b est le nombre total d'événements de bruit de fond.

Bibliographie

- [1] CMS Collaboration. Measurement of the properties of a Higgs boson in the four-lepton final state. *Phys. Rev. D*, 89 :092007, May 2014.
- [2] CMS Collaboration. Constraints on the Higgs boson width from off-shell production and decay to Z -boson pairs. *Phys.Lett.*, B736 :64, 2014.
- [3] CMS Collaboration. Measurement of the $pp \rightarrow ZZ$ production cross section and constraints on anomalous triple gauge couplings in four-lepton final states at $\sqrt{s} = 8$ TeV. *Phys.Lett.*, B740 :250–272, 2015.
- [4] ATLAS Collaboration. Measurement of the total ZZ production cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV in 20 fb^{-1} with the ATLAS detector. Technical Report ATLAS-CONF-2013-020, CERN, Geneva, Mar 2013.
- [5] CMS Collaboration. Measurements of the ZZ production cross sections in the $2\ell 2\nu$ channel in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV and combined constraints on triple gauge couplings. *ArXiv*, 1503.05467, 2015.
- [6] J. Beringer et al. Review of Particle Physics. *Phys. Rev. D*, 86 :010001, Jul 2012.
- [7] Laurent Favart. Physique auprès des collisionneurs, 2013. Cours suivi en MA1.
- [8] Julien Baglio, Le Duc Ninh, and Marcus M. Weber. Massive gauge boson pair production at the LHC : A next-to-leading order story. *Phys. Rev. D*, 88 :113005, Dec 2013.
- [9] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 1107 :018, 2011.
- [10] Université de Durham. Site du projet HepData. <http://hepdata.cedar.ac.uk/pdf/pdf3.html>. Dernière consultation le 14 mai 2015.
- [11] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 1407 :079, 2014.
- [12] Fabrizio Caola and Kirill Melnikov. Constraining the Higgs boson width with ZZ production at the LHC. *Phys.Rev.*, D88 :054024, 2013.
- [13] Anastasiya Bierweiler, Tobias Kasprzik, and Johann H. Kühn. Vector-boson pair production at the LHC to $\mathcal{O}(\alpha^3)$ accuracy. *JHEP*, 1312 :071, 2013.
- [14] Stefan Gieseke, Tobias Kasprzik, and Johann H. Kühn. Vector-boson pair production and electroweak corrections in HERWIG++. *Eur.Phys.J.*, C74(8) :2988, 2014.

- [15] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 0411 :040, 2004.
- [16] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 0709 :126, 2007.
- [17] Tom Melia, Paolo Nason, Raoul Rontsch, and Giulia Zanderighi. W^+W^- , WZ and ZZ production in the POWHEG BOX. *JHEP*, 1111 :078, 2011.
- [18] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05) :026, 2006.
- [19] S. Agostinelli et al. GEANT4 : A Simulation toolkit. *Nucl.Instrum.Meth.*, A506 :250–303, 2003.
- [20] Site du générateur gg2vv. <https://gg2vv.hepforge.org/>. Dernière consultation le 14 mai 2015.
- [21] Francis Halzen and Alan D. Martin. *Quarks & Leptons : An Introductory Course in Modern Particle Physics*. John Wiley & Sons, 1984.
- [22] Site des accélérateurs du CERN. <http://te-dep-epc.web.cern.ch/te-dep-epc/machines/general.stm>. Dernière consultation le 14 mai 2015.
- [23] Pierre Germain, Daniel Dekkers, and Django Manglunki. Introduction aux accélérateurs de particules. Cours suivi au CERN en MA2.
- [24] Site de la collaboration CMS. http://cms-service-lumi.web.cern.ch/cms-service-lumi/publicplots/int_lumi_per_day_cumulative_pp_2012.pdf. Dernière consultation le 14 mai 2015.
- [25] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3 :S08004, 2008.
- [26] Site de la collaboration CMS. <http://cms.web.cern.ch/news/cms-detector-design>. Dernière consultation le 14 mai 2015.
- [27] Gilles De Lentdecker and Pascal Vanlaer. Méthodes expérimentales de la physique des particules. Cours suivi en MA2.
- [28] Olga Kodolova and Michael Murray. Heavy Ion Physics Program in CMS Experiment. *Nucl.Phys.*, A830 :97C–104C, 2009.
- [29] R. Brun and F. Rademakers. ROOT : An object oriented data analysis framework. *Nucl.Instrum.Meth.*, A389 :81–86, 1997.
- [30] Escoubès B. *Probabilités et statistiques à l'usage des physiciens*. Ellipses, 1998.
- [31] CMS Collaboration. CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, 2013.
- [32] Francisco Campanario, Michael Rauch, and Sebastian Sapeta. ZZ production at high transverse momenta beyond NLO QCD. *ArXiv*, 1504.05588, 2015.