

Étude des jets associés à la production d'un boson Z
pour les collisions proton-proton à CMS à une énergie
de $\sqrt{s}=8$ TeV

Simoni Rachel

4 octobre 2013



Sous la direction de
Laurent Favart

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION DU DIPLÔME DE MASTER
EN SCIENCES PHYSIQUES
UNIVERSITÉ LIBRE DE BRUXELLES - DÉPARTEMENT DE PHYSIQUE - IIHE



Résumé

Nous présentons les mesures des sections efficaces différentielles pour la production de un ou plusieurs jets associés avec un boson Z se désintégrant en une paire $\mu^+\mu^-$. La section efficace est établie en fonction de l'impulsion transverse des jets, de leur pseudo-rapidité η et en fonction de la multiplicité exclusive de jets. Nous présentons également la section efficace en fonction de l'impulsion et de la pseudo-rapidité η du boson Z. Les données utilisées ont été collectées par l'expérience CMS en 2012 et correspondent à une luminosité intégrée 19.3 fb^{-1} . Dans une première partie, nous décrivons les expériences CMS et LHC qui nous ont permis de prendre les données, ensuite nous exposons les étapes de génération des fichiers MC et de reconstruction. Dans une deuxième partie nous présentons les détails de notre analyse : les différents fichiers utilisés, la sélection de l'échantillon, le calcul de corrections pour le nombre de vertex et les efficacités d'identification et d'isolation des muons et la soustraction du bruit de fond. Nous développons plus particulièrement les procédures liées à l'extraction du "pileup" (empilement de plusieurs événements pour une même collision). Enfin nous expliquons la procédure d'Unfolding permettant de s'affranchir des effets du détecteur et d'extraire les sections efficaces. Les résultats sont présentés en comparaison avec les simulations MC représentant les prédictions théoriques du générateur MadGraph.

Abstract

We describe a measurement of the differential cross section for the production of one or more jets in association with a Z boson decaying in two muons in p-p collisions at a center of mass energy of $\sqrt{s} = 8 \text{ TeV}$. The cross section is presented as a function of the p_T of the nth jet, as a function of η of the nth jet and as a function of the jet exclusif number. We present also the differential cross section as a function of the Z kinematic variables p_T and η . The data have been collected in the year 2012 by the CMS detector and correspond to an integrated luminosity of 19.3 fb^{-1} . In a first part the thesis presents all the steps leading to the data recording at CMS as well as the procedure of event generation and reconstruction. In a second part, the details of the analysis are presented : data and Monte Carlo sets, trigger paths, efficiency determination and signal extraction. The distributions obtained after background and pileup subtraction are then unfolded back to the generator level using inversion matrix obtained from MC samples. The differential cross section are then shown and compared to theoretical predictions from MadGraph MC.

Mots clés : QCD, LHC, CMS, partons, jets, MadGraph, pileup.

Table des matières

Table des matières	i
1 Introduction	1
2 Éléments de théorie	3
2.1 Le Modèle Standard	3
2.1.1 $\text{SU}(2)_{\text{L}} \otimes \text{U}(1)_{\text{Y}}$	4
2.1.2 Propriétés du boson Z	6
2.2 Chromodynamique Quantique	6
2.2.1 Parton distribution function PDF	10
2.3 Processus Drell-Yan $q\bar{q} \rightarrow Z/\gamma^* (\rightarrow l^+l^-) + X$	13
2.3.1 Radiation de gluons associée au processus DY $q\bar{q} \rightarrow Z + jets$. .	14
3 Dispositif expérimental : le LHC et CMS	15
3.1 LHC : Large Hadron Collider	15
3.2 CMS : Compact Muon Solenoid	21
3.2.1 Système de coordonnées	22
3.2.2 Le système de tracker	23
3.2.2.1 Tracker à pixels	23
3.2.2.2 Tracker à micro-pistes de silicium	24
3.2.3 Le calorimètre électronique (ECAL)	24
3.2.4 Le calorimètre hadronique (HCAL)	25
3.2.5 L'aimant solénoïdal supraconducteur	27
3.2.6 Les chambres à muons	29
3.2.7 Le Système de déclenchement (Trigger)	31
3.2.8 CMS Traitement des données	33
4 Chaîne de simulation et reconstruction	37
4.1 Chaîne de Simulation	37
4.1.1 Génération d'événement	38
4.1.1.1 Processus dur - hard process	38
4.1.1.2 Radiation de parton - Parton shower	40
4.1.1.3 Génération de jets associés au Z	41
4.1.1.4 Underlying events	42
4.1.1.5 Hadronisation	44
4.1.1.6 Désintégrations spontanées	44
4.1.2 Modélisation du détecteur et digitalisation	44

4.1.3	Fichiers MC utilisés et calibration par rapport aux données	45
4.2	Reconstruction des événements	46
4.2.1	Stratégie de reconstruction du PF	47
4.2.1.1	Reconstruction itérative de traces	47
4.2.1.2	Agrégation	49
4.2.2	Reconstruction des muons	50
4.2.3	Reconstruction des électrons	51
4.2.4	Reconstruction des hadrons chargés et hadrons neutres	52
4.2.5	Reconstruction des jets	53
4.2.6	Calibration des jets	54
5	Sélection	59
5.1	Fichiers de données	59
5.2	Sélection online	60
5.3	Sélection offline	61
5.3.1	Sélection des muons	61
5.3.1.1	Identification des muons	62
5.3.1.2	Isolation des muons	63
5.3.2	Sélection offline des jets	64
5.3.2.1	Identification des jets	65
5.3.2.2	Isolation des jets	65
5.4	Soustraction du pileup	65
5.4.1	Soustraction du pileup au niveau JEC	66
5.4.1.1	Méthode de "jet area"	67
5.4.1.2	Méthode d'offset moyen	68
5.4.2	Soustraction du pileup par la Charge Hadron Subtraction	70
5.4.3	Soustraction Offline du pileup par l'algorithme PileupJetID	71
5.4.3.1	Principe de Multi Variate Analysis	71
5.4.3.2	Variables utilisées par le PileupJetID	72
5.4.3.3	Étude de la coupure de sélection PuJetID >-0.89	73
6	Traitement de données	77
6.1	Ajustement entre MC et données	77
6.1.1	Correction pour le nombre de vertex	77
6.1.2	Efficacités	78
6.1.2.1	Méthode Tag and Probe	80
6.1.2.2	Efficacités d'identification et d'isolation	80
6.1.2.3	Établissement des erreurs sur les efficacités	81
6.2	Soustraction du bruit de fond	81
6.2.1	Les différentes contributions du bruit de fond	82
6.3	Distributions de contrôle	83
6.3.0.1	Multiplicité de jets	83
6.3.0.2	Distributions en l'impulsion transverse des jets	83
6.3.0.3	Distributions en l'impulsion transverse et en η pour le Z	83
7	Unfolding et résultats	91
7.1	Unfolding	91

7.1.1	Établissement de la matrice de réponse	92
7.1.2	Unfolding par la méthode de Bayes	95
7.1.3	Validation de la procédure à partir du Monte Carlo	97
7.2	Erreurs systématiques	98
7.3	Résultats	100
7.3.1	Section efficace différentielle en la multiplicité exclusive de jets . .	100
7.3.2	Section efficace différentielle en les variables p_T et η du boson Z .	101
7.3.3	Section efficace différentielle en la variable p_T des jets	101
7.3.4	Section efficace différentielle en la variable η des jets	101
7.4	Commentaires sur les résultats	101
8	Conclusion	105
	Table des figures	107
	Liste des tableaux	108
9	Annexe 1	111
9.1	Liste des Acronymes et abréviations	111
	Bibliographie	113

Chapitre 1

Introduction

Le LHC est un collisionneur proton-proton le plus puissant jamais construit. Il a été créé pour la découverte de nouvelles particules et son actualité est certainement la découverte d'une particule pouvant être le boson de Brout-Englert-Higgs annoncée par le CERN, le 4 juillet 2012. Depuis, les collaborations CMS et ATLAS ont poursuivi l'analyse des données de 2011 et 2012 pour la plupart des canaux de désintégration et ont pu confirmer avec une grande précision les caractéristiques du boson recherché.

La découverte du boson de Brout-Englert-Higgs a nécessité des données produites à une énergie de 7 et 8 TeV dans le centre de masse. A une telle énergie les collisions proton-proton présentent des états finaux de grande multiplicité et une production de jets gigantesques. L'étude des mécanismes de production et des propriétés de ces jets est donc toujours plus que nécessaire parce qu'ils constituent le bruit de fond expérimental à la plupart des processus étudiés au LHC.

D'autre part, les jets représentent le principal outil d'accès à une information concernant la dynamique partonique des interactions lors d'une collision proton-proton à haute énergie. Depuis trente ans beaucoup de travail a été mené pour parvenir à comprendre et reproduire les propriétés des jets dans la physique des collisionneurs hadroniques et semi-hadroniques. Cela a mené à la confirmation de la théorie QCD, aux premières mesures de précision (la constante de couplage α_s est actuellement connue avec une précision au pour mille) et par exemple à la compréhension de la production de quarks lourds. Mais de nombreuses manifestations restent incomprises, comme le confinement des partons ou les corrélations de la production des jets. Si la compréhension de la production d'événements multi-jets au LHC est essentielle, l'approche théorique classique (calcul analytique des ordres supérieurs) semble pour l'instant vouée à l'échec. D'autres approches basées sur des générateurs Monte Carlo sont développées depuis une dizaine d'années et obtiennent des résultats bien encourageants.

Une façon d'étudier la production de jets est d'analyser ceux qui sont produits en association avec la production d'un boson Z. En effet, on connaît les propriétés du boson Z avec une grande précision et la statistique de production du Z au LHC est grande. De plus, le canal de désintégration du Z en leptons est la voie d'or expérimentale pour un détecteur comme CMS.

Ce mémoire propose d'étudier la production d'un boson Z associé à la production de jets à partir des données collectées en 2012 par CMS au LHC en considérant uniquement le canal de désintégration du Z en deux muons. Ce processus constitue également un objet

d'étude en soit : la section efficace de production associée de plus de 2 jets n'ayant jamais été mesurée précisément à de telles énergies. Les données ont été prise à une énergie de $\sqrt{s} = 8$ TeV et représentent une luminosité intégrée de 19.3 fb^{-1} . Nous déterminerons les sections différentielles du processus $pp \rightarrow (Z \mapsto \mu^+ \mu^-) + n \text{ jets}$ en fonction de différentes variables : la multiplicité de jets, les impulsions transverses et la pseudo-rapacité η des différents jets, ainsi que l'impulsion transverse et la pseudo rapidité η du Z. Les résultats sont présentés en comparaison avec des simulations MC représentant les prédictions théoriques du générateur MadGraph.

Nous expliquerons toutes les étapes et détaillerons tous les ingrédients de la chaîne conduisant à l'établissement de ces sections efficaces, de la collision de protons au LHC à la procédure d'Unfolding. Nous prendrons soin de bien décrire la portée des différentes procédures utilisées.

Les trois premiers chapitres constituent des chapitres introductifs. Dans le chapitre 2, nous allons énoncer de manière résumée les pré-requis théoriques menant au processus étudié. Au chapitre 3, nous décrirons le collisionneur proton-proton LHC et le détecteur CMS qui ont fourni les données utilisées dans cette analyse. Le chapitre 4 sera consacré à la description de la chaîne de simulation de CMS et des étapes de la reconstruction menée par les algorithmes du Particle Flow. Dans les chapitres suivants, nous décrirons les différentes étapes menées dans notre analyse : la sélection de notre échantillon de données ainsi que les différentes procédures de soustraction du pileup (empilement d'événements pour un même croisement de paquets) seront présentées au chapitre 5, les procédures de calcul d'efficacités et de soustraction du bruit de fond seront développées au chapitre 6. Au chapitre 7, nous discuterons de la procédure d'Unfolding et nous présenterons nos résultats.

Dans ce travail, nous avons pris le parti d'utiliser un certain jargon expérimental sous forme d'acronymes ou d'anglicismes. Nous espérons que ceci ne vous fera pas perdre le fil de la lecture.

Chapitre 2

Éléments de théorie

Avant de décrire les étapes expérimentales liées à notre analyse nous allons énoncer les pré-requis théoriques nécessaires à la compréhension du processus étudié : $pp \rightarrow Z(\mapsto \mu^+\mu^-) + jets$. Le cadre théorique général dans lequel nous travaillons est le Modèle Standard (Standard Model SM). Nous introduirons rapidement ses principales caractéristiques et plus précisément celles des courants neutres. Nous nous concentrerons plus longuement sur les interactions fortes et la Chromodynamique Quantique, sujet de notre étude. Puis nous présenterons les sections efficaces théoriques des processus dits de Drell Yan et Drell Yan + jets.

2.1 Le Modèle Standard

Cette première partie s'inspire de l'ouvrage cité en [1]. La physique des particules a pour but principal de décrire l'univers à son niveau le plus fondamental, c'est-à-dire les particules élémentaires et leurs interactions. Sa théorie la plus poussée et la mieux vérifiée jusqu'à présent est le Modèle Standard. Il s'agit d'une théorie quantique des champs qui traite matière et forces dans un formalisme d'interaction de jauge. La matière du modèle standard est constituée par 12 fermions : 6 leptons et 6 quarks comme représentés dans le tableau 2.1. Il existe trois familles (ou générations) de leptons et trois de quarks.

Le modèle standard décrit 3 interactions : l'interaction électromagnétique, l'interaction faible et l'interaction forte. Chaque interaction est associée à un multiplet bosonique dont les composantes sont les "particules médiatrices" de l'interaction. Le tableau 2.1 donne un résumé des bosons médiateurs, de leur charge et de leur spin.

Chacune des interactions correspond à un courant de jauge associé à un groupe unitaire : U(1) pour l'électromagnétisme, SU(2) pour l'interaction faible, et SU(3) pour l'interaction forte, SU(2) et SU(3) étant des groupes non abéliens.

Le groupe SU(2) agit sur des objets chiraux : on distingue les doublets de chiralité gauche comprenant les doublets de leptons $\begin{pmatrix} \nu_l \\ l \end{pmatrix}_L$ et de quarks $\begin{pmatrix} u \\ d \end{pmatrix}_L$, des singulets de chiralité droite, l_R, u_R, d_R . l représente e, μ et τ et u représente ici les quarks u, c, t , et d les quarks d, s, b .

Chaque courant de jauge est associé à une quantité conservée comme par exemple la charge électrique pour U(1), et la couleur pour SU(3). Pour SU(2) la quantité conservée

Particules		Spin	Charge
Leptons	$\nu_e \ \nu_\mu \ \nu_\tau$	1/2	0
	$e^- \ \mu^- \ \tau^-$		-1
Quarks	u c t	1/2	$\frac{2}{3}$
	d s b		$-\frac{1}{3}$
Interaction	Boson	Spin	Charge
Électromagnétique	γ	1	0
Faible	W^+, W^-, Z	1	1,-1,0
Forte	8 gluons	1	0

TABLE 2.1: Composantes de la matière et des interactions décrites dans le cadre du Modèle Standard. A chaque particule de matière correspond une antiparticule non reprise dans le tableau.

est l'isospin T_3 qui vaut respectivement $\pm 1/2$ pour la partie supérieure ou inférieure d'un doublet left, et 0 pour un singulet droit.

Dans ce formalisme, les équations dynamiques pour le modèle standard sont écrites à partir d'un Lagrangien invariant par :

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y \quad (2.1)$$

Les indices liés aux noms des groupes décrivent les quantités pertinentes pour l'action du groupe sur les multiplets : "c" pour couleur, "L" pour left et "Y" pour l'hypercharge que nous décrirons dans un instant.

2.1.1 $SU(2)_L \otimes U(1)_Y$

La théorie de Glashow, Salam et Weinberg a unifié l'interaction électromagnétique et l'interaction faible pour donner naissance à la théorie électrofaible. Cette théorie repose sur le groupe de symétrie $SU(2)_L \otimes U(1)_Y$ qui comporte quatre générateurs. Le Lagrangien électro-faible invariant sous $SU(2)_L \otimes U(1)_Y$ est donné par :

$$\begin{aligned}
L_{EWK} &= i\bar{\psi}_f^i \gamma_\mu (D^\mu)_{ij} \psi_f^j - \frac{1}{4}(B_{\mu\nu}B^{\mu\nu} - W_{\mu\nu}^\alpha W^{\alpha\mu\nu}) \\
(D_\mu) &= \partial_\mu - ig_1 \frac{Y}{2} B_\mu - ig_2 \frac{\tau_\alpha}{2} W_\mu^\alpha \\
B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\
W_{\mu\nu}^\alpha &= \partial_\mu W_\nu^\alpha - \partial_\nu W_\mu^\alpha + g_2 \varepsilon_{\alpha\beta\gamma} W_{\mu\beta} W_{\nu\gamma}
\end{aligned} \quad (2.2)$$

Où ψ_f^i représente la fonction d'onde d'un fermion f , $\bar{\psi}_f^i$ représente la fonction d'onde de l'anti-fermion, B_ν le champ de jauge associé à $U(1)_Y$, W_ν^α le champ de jauge associé à $SU(2)_L$ ($\alpha = 1, 2, 3$), τ_α les générateurs du groupe $SU(2)$ et g_1 et g_2 les constantes de couplage respectives de $U(1)_Y$ et $SU(2)_L$.

Le modèle inclut une brisure spontanée de symétrie : $SU(2)_L \otimes U(1)_Y \longrightarrow U(1)_{em}$ où le groupe $U(1)_{em}$ décrit l'interaction électromagnétique. La nature spontanée de cette

brisure de symétrie va permettre aux bosons d'acquérir une masse sans détruire l'invariance de jauge du Lagrangien du Modèle Standard. Ceci est possible grâce au mécanisme de Brout-Englert-Higgs qui prédit la présence d'un boson scalaire dont les couplages aux bosons et aux fermions permet de leur conférer une masse. Nous ne développerons pas le mécanisme de Brout-Englert-Higgs dans ce travail.

La relation $SU(2)_L \otimes U(1)_Y \longrightarrow U(1)_{em}$ nous donne également la relation entre isospin T_3 , hypercharge Y et charge électrique Q : $Q = T_3 + Y/2$.

Les hypercharges ne sont pas fixées par la théorie mais sont déterminées par l'observation expérimentale des charges Q .

Il est également possible d'établir les liens entre champ de jauge et bosons médiateurs des interactions électrofaibles, $W_\pm$ et Z .

Les bosons $W^\pm$ dits de "courant chargé" sont reliés aux bosons de jauge W^1, W^2 par la relation :

$$W^\pm = \frac{W^1 \pm iW^2}{2} \quad (2.3)$$

Les composantes Z_μ du boson neutre Z et A_μ du photon sont obtenues par rotation d'un angle θ_W ¹ des champs W_μ^3 et B_μ :

$$\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} \quad (2.4)$$

Des formules 2.3 et 2.4 on peut déduire les relations fondamentales entre les couplages g_1 et g_2 relatifs à $U(1)_Y$ et $SU(2)_L$ et l'angle de Weinberg θ_W :

$$\cos\theta_W = \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \quad (2.5)$$

D'autre part des mesures expérimentales précises ont permis la mesure de la constante de couplage de Fermi qui peut être exprimée en fonction de la constante de couplage g_2 et de la masse du boson W , M_W :

$$G_F = \frac{g_2^2}{4\sqrt{2}M_W^2} \quad (2.6)$$

Et l'on a également la relation liant la masse du boson Z à celle du boson W :

$$M_Z^2 = \frac{M_W^2}{\cos\theta_W} \quad (2.7)$$

De manière générale il existe trois paramètres libres dans le modèle électrofaible : M_W , g_F , θ_W . On utilise plutôt G_F , M_Z et α_{em} qui sont les quantités les mieux définies expérimentalement.

Nous allons rapidement passer en revue les propriétés principales des interactions dites de courant neutre.

1. θ_W est appelé angle de Weinberg ou angle "Weak"

2.1.2 Propriétés du boson Z

Couplage du Z aux fermions :

Le couplage du Z aux fermions peut être déterminé à partir de l'expression de la dérivée covariante décrite en 2.2 et de la rotation décrite en 2.4 et s'exprime de la façon suivante :

$$-\frac{1}{2}g_1 Y \sin\theta_W + g_2 T_3 \cos\theta_W = \frac{e}{\sin\theta_W \cos\theta_W} (T_3 - Q \sin^2\theta_W) \quad (2.8)$$

Où e est la charge d'un électron, et Q la charge de la particule exprimée en fonction de e . Les couplages du boson Z aux fermions présentent donc deux composantes, g_L pour les doublets de chiralité left et g_R pour les singulets droits, définies par :

$$g_L = (T_3 - Q \sin^2\theta_W), \quad g_R = (-Q \sin^2\theta_W). \quad (2.9)$$

On définit à partir de ces composantes les couplages vectoriel et axial :

$$g_V = g_L + g_R = T_3 - 2Q \sin^2\theta_W, \quad g_A = g_L - g_R = T_3 \quad (2.10)$$

Ces couplages vectoriels et axial ont été déterminé expérimentalement avec une grande précision lors de mesures des asymétries Avant-Arrière au LEP.

Masse du Z et temps de vie du Z :

La masse du Z et le temps de vie du Z sont très bien déterminés expérimentalement et ce grâce aux expériences de collision électron - positron menées au LEP et au SLC (SLAC). Les valeurs les plus actuelles de la masse et la largeur totale du boson Z sont [2] :

$$M_Z = 91.1876 \pm 0.0021 \text{ GeV} \cdot c^{-2} \quad (2.11)$$

$$\Gamma_Z = 2.4952 \pm 0.0023 \text{ GeV}. \quad (2.12)$$

Section efficace de $ee \rightarrow Z \rightarrow qq$: La section efficace différentielle au premier ordre du processus $ee \rightarrow Z \rightarrow qq$ peut s'exprimer en fonction de $\cos\theta$ dans le cas ultra-relativiste.

$$\frac{d\sigma}{d\cos\theta} = \frac{\pi\alpha^2 k^2}{2\Gamma_Z^2} \left((1 + \cos^2\theta)(g_{A,e}^2 + g_{V,e}^2)(g_{A,q}^2 + g_{V,q}^2) + 8\cos\theta g_{A,e} g_{V,e} g_{A,q} g_{V,q} \right) \quad (2.13)$$

Où l'indice e réfère à l'électron et l'indice q à la saveur du quark produit, les couplages g_A et g_R sont décrits en 2.10, Γ_Z est la largeur de désintégration du Z et $k = \frac{\sqrt{G_F M_Z^2}}{4\pi\alpha}$.

2.2 Chromodynamique Quantique

La Chromodynamique Quantique (Quantum Chromodynamics, QCD) est la théorie des interactions fortes. Elle décrit une théorie invariante selon le groupe de jauge non abélien $SU(3)_c$.

QCD décrit les interactions de couleur entre les quarks et les gluons aussi désignés sous le nom générique de partons. La quantité conservée par la théorie de jauge $SU(3)_c$

est la couleur. Les partons portent une charge de couleur (les quarks portent un indice de couleur et les gluons une couleur et une anti-couleur) alors que les états liés existants (les hadrons) sont neutres en couleur². La couleur est donc un degré de liberté dégénéré qui n'est pas observé dans la nature.

Le modèle des quarks a été introduit en 1964 par Gell-Man et Zweig qui ont établi des multiplets de hadrons dans des représentations de SU(3) globale. La mise en évidence expérimentale des partons est effectuée en 1968 avec la première expérience de diffusion inélastique.

La QCD est proposée par Politzer, Gross, Wilzek³ en 1973. L'observation d'un jet associé à la radiation d'un gluon en 1978 à l'expérience PETRA (DESY Hambourg) fut une des premières confirmations de la théorie.

L'invariance du Lagrangien de QCD par rapport à SU(3)_c traduit l'invariance de la charge de couleur. Les gluons sont définis comme les champs porteurs de l'interaction de couleur de façon similaire aux photons médiateurs de l'interaction électromagnétique. Une première propriété des gluons est qu'ils portent une charge de couleur et cela constitue une grande différence avec les photons qui ne portent pas de charges électriques. Cela vient du fait que la théorie QCD est non abélienne.

D'un point de vue interaction de jauge, les gluons sont les huit générateurs du groupe dans la représentation fondamentale de SU(3). Une représentation matricielle 3x3 de ces générateurs est donnée par les matrices de Gell-Man λ . Leur règle de commutation est la suivante :

$$\left[\frac{\lambda_a}{2}, \frac{\lambda_b}{2} \right] = i f_{abc} \frac{\lambda_c}{2} \quad (2.14)$$

Et le Lagrangien QCD est donné par :

$$\begin{aligned} L_{QCD} &= i \sum_q \bar{\psi}_q^i \gamma_\mu (D^\mu)_{ij} \psi_q^j - \sum_q m \bar{\psi}_q^i \psi_{qi} - i g_s \bar{\psi} \gamma_\mu \lambda_a \psi G_\mu^a - \frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a \\ G_a^{\mu\nu} &= \partial^\mu A_\nu^a - \partial^\nu A_\mu^a - g_s f_{abc} A_b^\mu A_c^\nu \\ (D_\mu)_{ij} &= \delta_{ij} \partial_\mu + i g_s \sum_a \frac{\lambda_{i,j}^a}{2} A_\mu^a \end{aligned} \quad (2.15)$$

Où ψ_q^i représente la fonction d'onde du quark i de saveur q, $\bar{\psi}_q^i$ la fonction d'onde de l'anti-quark i de saveur q, $A_\mu^{a,b,c}$ les trois composantes du champ de jauge de SU(3). La sommation se fait sur les différentes saveurs q de quarks.

Interactions entre gluons :

Le terme cinétique du champ de jauge $G_a^{\mu\nu} G_{\mu\nu}^a$ présente clairement des termes de couplage entre plusieurs champs de jauge. A partir de ce terme, on peut établir la composante d'interaction du Lagrangien qui présente trois termes : le couplage des gluons à la matière des quarks et deux termes d'interactions entre gluons eux-même couplant 3 gluons ou 4 gluons comme représenté dans l'équation ci-dessous.

2. Les états liés observés sont $q\bar{q}$, ou qqq pour lequel chaque quark porte une couleur différente. A contrario, l'état qq n'étant pas neutre de couleur ne peut pas exister.

3. 't Hooft a largement contribué à l'établissement de cette théorie par son théorème sur les théories renormalisables.

$$L_{int} = -g_s \bar{\psi}_q^c(x) \gamma^\mu A_\mu^a(x) \frac{\lambda^a}{2} \psi_q^c(x) + \frac{g_s}{2} f_{abc} [\partial^\mu A_a^\nu(x) - \partial^\nu A_a^\mu(x)] A_\mu^b(x) A_\nu^c(x) \quad (2.16)$$

$$- \frac{g_s^2}{4} f_{eab} f_{ecd} A_a^\nu(x) A_b^\nu(x) A_\nu^c(x) A_\nu^d(x)$$

Ces termes d'interaction entre gluons est en particulier à l'origine du comportement divergent de la constante de couplage de l'interaction forte comme nous le verrons.

"The running coupling constante" α_s

Dans le Lagrangien QCD, le paramètre g_s intervient comme le couplage mais on définit par commodité $\alpha_s = g_s/4\pi$ comme la constante de couplage de l'interaction forte. Elle présente le comportement suivant :

- A petite distance le couplage est très faible : ce régime s'appelle la liberté asymptotique et permet une approche perturbative dans la résolution des équations de la QCD.
- A longue distance elle est très grande. Ce régime ne peut plus être traité avec la théorie des perturbations. Une conséquence est le confinement des quarks et leur hadronisation.

La figure 2.1 représente le comportement de la constante de couplage en fonction de la distance⁴ à une charge de couleur . La valeur de la constante de couplage α_s va déterminer différentes échelles d'énergie pour lesquelles on peut traiter la QCD de manière perturbative ($\alpha_s \ll 1$) ou non.

Régime perturbatif :

Dans le régime perturbatif l'expression de la constante de couplage forte peut être établie à partir des corrections radiatives au propagateur du gluon. Ces corrections sont des boucles de paires $q\bar{q}$ mais aussi des boucles de gluons qui ajoutent un terme supplémentaire au dénominateur. Ce terme de signe opposé est responsable de la divergence à faible énergie.

Le calcul de α_s nécessite l'introduction d'un cut-off ultra-violet, μ_{UV} . L'expression de la constante de couplage au premier ordre est :

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu_{UV}^2)}{1 - \frac{2N_f - 11N_c}{6\pi} \alpha_s(\mu_{UV}^2) \ln(\frac{Q^2}{\mu_{UV}^2})} \quad (2.17)$$

Le terme en $2N_f$ provient de la contribution des quarks et dépend du nombre de saveurs de la théorie ($N_f = 6$ pour le SM). Le terme $11N_c$ provient de la contribution des gluons et dépend du nombre de couleurs. Pour le modèle standard on a $2N_f - 11N_c = 21$ qui reste positif : le dénominateur devient donc inférieur à 1 et la constante de couplage ne peut donc plus être traitée de façon perturbative.

4. ce qui correspond environ à $1/Q$ si Q est l'échelle d'énergie

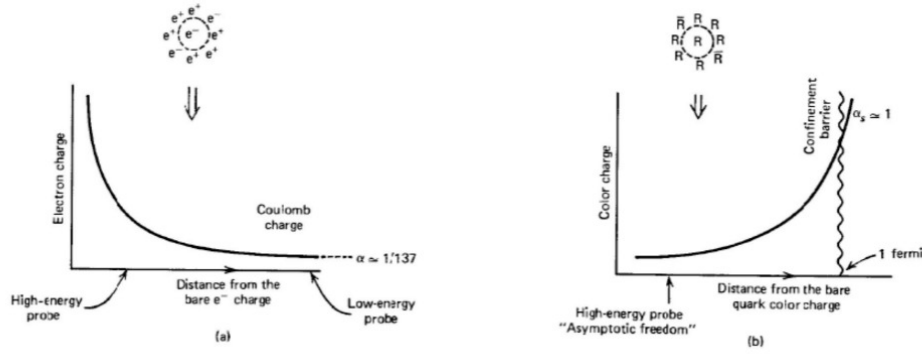


FIGURE 2.1: a) Evolution de la constante de couplage électromagnétique en fonction de la distance à la charge électrique nue b) Évolution de la constante de couplage forte en fonction de la distance à la charge de couleur nue [3].

Le pivot entre les régimes perturbatifs ou non perturbatifs est définie par le paramètre Λ_{QCD} qui correspond à la valeur de Q^2 pour laquelle le dénominateur de 2.17 s'annule. On obtient $\alpha_s \rightarrow \infty$ quand $Q \rightarrow \Lambda_{QCD}$.

La valeur de Λ_{QCD} n'est pas fixée par la théorie. Elle doit correspondre à la masse des hadrons les plus légers et on l'estime à $\Lambda_{QCD} = 210^{+34}_{-30}$ MeV. Dans le cadre de notre travail, le choix de la valeur de Λ_{QCD} et de celle de $\alpha_s(M_Z)$ intervient dans l'établissement des simulations MC.

Régime non perturbatif :

Dans la nature, on se trouve à des échelles d'énergies où le régime non perturbatif est bien dominant. La matière étant constituée de hadrons stables comme les protons. On n'a jamais observé de quarks à l'état libre et l'on suppose qu'ils sont confinés à l'intérieur des hadrons.

Ce phénomène appelé "confinement des quarks" ne peut pas être résolu de manière perturbative, et l'on construit des modèles basés sur les observations. Le potentiel utilisé pour modéliser l'interaction QCD entre un quark et un anti-quark de même couleur est le potentiel de "funel". Dans ce modèle le propagateur des gluons correspond à un champ dipolaire entraînant une interaction de type coulombien entre la paire $q\bar{q}$. A cette interaction coulombienne, on ajoute une composante linéaire en la distance de séparation des quarks [3].

$$V(r) = -\frac{4\alpha_s}{3r} + kr \quad k \simeq 0.9 \text{ GeV} \cdot \text{fm}^{-1} \quad (2.18)$$

Donc, quand la distance entre deux quarks est petite par rapport à la taille typique du confinement ($\sim 1\text{fm}$), le champ de gluon est de forme dipolaire. Quand la distance de séparation devient de l'ordre du fermi, les lignes de champs s'attirent entre elles et forment un tube dont l'énergie potentielle augmente de façon linéaire avec la distance de séparation. Si l'énergie emmagasinée dans le tube atteint par exemple 135 MeV le tube peut se scinder pour donner une nouvelle paire $q\bar{q}$ qui va former un pion. C'est ce

qu'on appelle la "fragmentation". Ce genre de modélisation est utilisée par les logiciels de simulation pour générer l'hadronisation. Nous en parlerons plus longuement dans le chapitre 4.

La figure 2.2 résume les différentes étapes de la création de hadrons dans une annihilation électron-positron : les différents échelles de la QCD concernent les étapes II à IV.

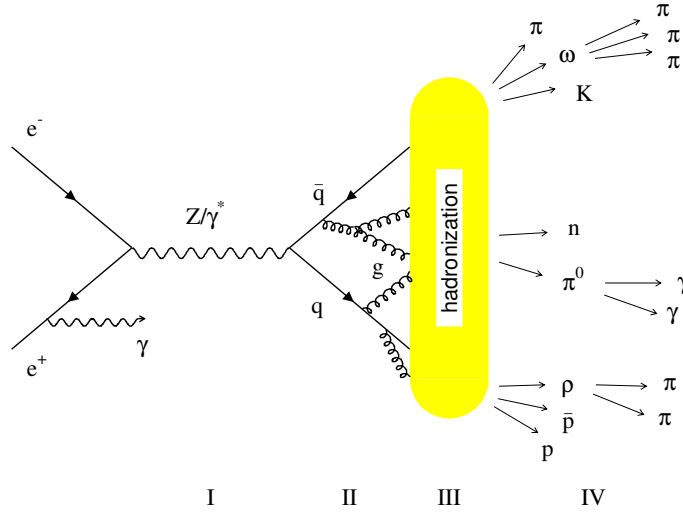


FIGURE 2.2: Schéma représentant les différentes étapes d'une annihilation électron positron : (I) la production d'un Z, (II) la désintégration du Z en une paire $q\bar{q}$ à l'échelle perturbative et ses corrections radiatives, (III) les phénomènes d'hadronisation à l'échelle non perturbative et en (IV) des désintégrations secondaires en hadrons et photons.

2.2.1 Parton distribution function PDF

Le modèle des partons

Lors d'une collision inélastique proton-proton en présence d'une échelle dure l'on estime que seuls deux partons issus de chaque proton interagissent lors de la collision,

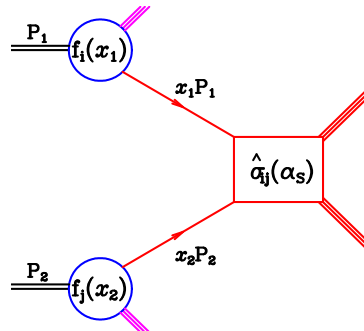


FIGURE 2.3: Schéma représentant le modèle "naïf" des partons.

chaque parton emportant une fraction d'impulsion du proton initial (cf. figure 2.3). Cette approche repose sur une description de la matière hadronique en terme de partons. C'est le modèle dit "naïf des partons".

Théoriquement ce modèle se base sur la propriété de la liberté asymptotique : si l'échelle de l'interaction est suffisamment élevée, la constante de couplage est petite par rapport à 1 et on peut considérer les partons comme quasiment à l'état libre.

L'établissement du lien entre une interaction au niveau hadronique et au niveau partonique nécessite l'intervention de la distribution des partons dans les hadrons ou PDF (Parton Distribution Function).

PDF :

Les Parton Distribution Fonctions (PDF) décrivent la densité de probabilité de trouver un parton dans un hadron. Elles sont étroitement liées à la définition des variables cinématique de Bjorken, qui définissent x par :

$$x = -q^2/2p.q = Q^2/2p.q \quad (2.19)$$

Où l'on a défini l'échelle d'énergie comme $Q^2 = -q^2$. Dans le cas relativiste (où l'on peut négliger les masses) la variable x peut être interprétée comme la fraction d'impulsion du proton emportée par le parton interagissant.

A l'échelle perturbative ($Q^2 \rightarrow \infty$ aussi appelé "domaine asymptotique de Bjorken") l'on peut appliquer le théorème de factorisation : la section efficace d'une diffusion profondément inélastique (Deep Inelastic Scattering DIS) se factorise en une contribution à courte portée calculable en QCD perturbative, et une contribution à longue portée ($\sim Q/x$) extraite des données expérimentales. Cette contribution sera portée par les PDF. Les PDF ne peuvent pas être déterminées par la théorie, la QCD n'étant pas perturbative à l'échelle d'énergie que constituent les hadrons. Les PDF ont été déterminées expérimentalement par des expériences de diffusion profondément élastique comme ZEUS et H1 à HERA (Hambourg) et par des expériences de diffusion dures comme CDF et D0 au TeVatron.

Équation d'évolution de la QCD (équations DGLAP)

Les PDF sont déterminées à une échelle arbitraire Q_0^2 (en général de l'ordre du GeV^2). Les équations d'évolution DGLAP⁵ permettent de déterminer les densités de parton intervenant dans un processus dur à une échelle Q^2 en évoluant les densités de l'échelle Q_0^2 à Q^2 .

Les évolutions de la densité de quark q_i (ou d'anti-quark) et de gluon sont données à l'ordre le plus bas par le système suivant d'équations différentielles partielles :

$$\frac{d}{dQ^2} q_i(x, Q^2) = \frac{\alpha_s(Q)}{2\pi} \left\{ \int_x^1 dy \frac{q_i(y, Q^2)}{y} . P_{qq}\left(\frac{x}{y}\right) + \int_x^1 dy \frac{g(y, Q^2)}{y} . P_{qg}\left(\frac{x}{y}\right) \right\} \quad (2.20)$$

$$\frac{d}{dQ^2} g(x, Q^2) = \frac{\alpha_s(Q)}{2\pi} \left\{ \int_x^1 dy \frac{\sum_i (q_i(y, Q^2) + \bar{q}_i(y, Q^2))}{y} . P_{gq}\left(\frac{x}{y}\right) + \int_x^1 dy \frac{g(y, Q^2)}{y} . P_{gg}\left(\frac{x}{y}\right) \right\} \quad (2.21)$$

5. DGLAP est un acronyme rassemblant les initiales des chercheurs à l'origine de ces équations : Dokshitzer, Gribov, Lipatov, Altarelli, Parisi

Les formes explicites des fonctions de splitting à l'ordre le plus bas (Leading order LO) dérivent directement des vertex de la théorie QCD et ne dépendent pas des PDF. Elles décrivent la dépendance en x de la radiation d'un gluon par un quark (P_{qq}), d'une paire $q\bar{q}$ par un gluon (P_{qg}) et d'un gluon par un gluon (P_{gg}). Altarelli and Parisi (1977) donnent la définition suivante des fonctions de splitting au LO [4] :

$$\begin{aligned}
P_{qq} &= \frac{4}{3} \left[\frac{1+x^2}{(1-x)_+} + \frac{3}{2} \delta(1-x) \right] + \mathcal{O}(\alpha_s) \\
P_{qg} &= \frac{4}{3} \left[\frac{1+(1-x)^2}{x} \right] + \mathcal{O}(\alpha_s) \\
P_{gg} &= \frac{1}{2} \left[x^2 + (1-x)^2 \right] + \mathcal{O}(\alpha_s) \\
P_{gg} &= 6 \left[\frac{x}{(1-x)_+} \frac{(1-x)}{x} x(1-x) \right] + \frac{33-2N_f}{6} \delta(1-x) + \mathcal{O}(\alpha_s)
\end{aligned} \tag{2.22}$$

Où N_f représente le nombre de saveurs, $\delta(1-x)$ correspond aux corrections virtuelles des diagrammes au niveau "tree". Ces corrections disparaissent si l'on applique les règles de conservation de charge et d'impulsion. Les distributions $+$ sont définies⁶ pour régulariser les divergences en $x=1$.

Aux ordres supérieurs la détermination des facteurs de splitting devient rapidement compliquée. Le calcul analytique au NNLO, impliquant un ordre en α_s^3 est un calcul monumental qui a été accompli en 2004 par Moch, Vermaseren and Vogt.

En pratique, les PDF sont évoluées de l'échelle de départ Q_0^2 jusqu'à l'échelle à laquelle on suppose que les partons entrent dans le processus dur que l'on nomme μ_F , échelle de factorisation.

La figure 2.4 représente la PDF CTEQM en fonction de la variable de bjorken x .

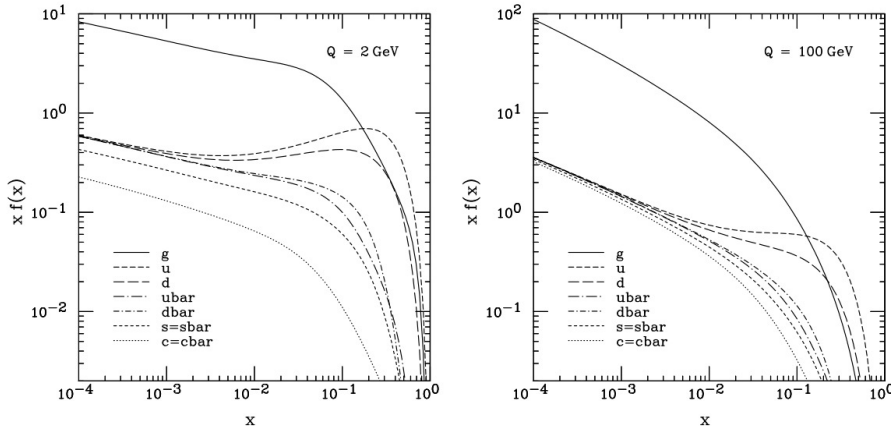
Si on se place dans le cas concret de notre analyse, l'échelle dure est fixée par la masse du Z et la collision p-p se passe à 8 TeV dans le centre de masse. Si l'on considère que le quark et l'anti-quark produisant le Z emportent chacun la même fraction x de l'énergie des deux protons incidents (d'impulsion p), si on néglige la masse des quarks par rapport à leur impulsion, on trouve une valeur de x de l'ordre de 10^{-2} (cf. calcul en 2.24).

$$M_Z^2 = (x_1 p + x_2 p)^2 \simeq (2xp)^2 = x^2 s \quad \Rightarrow x \simeq \frac{90}{8000} \simeq 0.011. \tag{2.24}$$

La figure 2.4 de droite montre que pour une échelle dure $Q = 100 \text{ GeV}$, et une valeur de x proche de 10^{-2} , la production de gluon est dominante et vaut 10 fois celle des quarks.

6. Pour une fonction $f(x)$ non singulière on a :

$$\int_0^1 \frac{f(x)}{(1-x)_+} dx = \int_0^1 \frac{f(x) - f(1)}{(1-x)} dx \tag{2.23}$$

FIGURE 2.4: Description de la PDF CTEQM pour $Q=2$ et $Q=100\text{GeV}$ [5].

2.3 Processus Drell-Yan $q\bar{q} \rightarrow Z/\gamma * (\rightarrow l^+l^-) + X$

Le processus Drell-Yan (DY) décrit la production de deux leptons à partir d'une annihilation $q\bar{q}$.

Drell et Yan sont les premiers à avoir mis en évidence en 1971 que le modèle des partons développé pour la DIS : $lq \rightarrow lq$ tenait pour le processus renversé dans les collisions hadrons-hadrons.

Le processus Drell-Yan constitue un canal privilégié dans la recherche de nouvelles particules. Il a joué un rôle important dans la découverte du J/Ψ (collisionneurs AGS-Brookhaven et SPAR-SLAC en 1974), du Υ (1977), du W et du Z (collisions proton-antiproton au CERN en 1983). Par ailleurs il a permis de tester la théorie des fonctions de distribution et leur dépendance d'échelle.

Estimation de la section efficace à l'ordre le plus bas dans le modèle des partons On peut estimer la section efficace de $q\bar{q} \rightarrow Z/\gamma(\rightarrow l^+l^-) + X$ en sommant les sections efficaces du Drell-Yan au niveau partonique sur toutes les contributions possibles de 2 partons q_1 et q_2 emportant une fraction d'impulsion x_1 et x_2 .

$$\sigma = \sum_{q,\bar{q}} \int dx_1 dx_2 q(x_1) \bar{q}(x_2) \hat{\sigma}(q\bar{q} \rightarrow l^+l^-) \quad (2.25)$$

Avec comme énergie disponible dans le référentiel du centre de masse des deux quarks $\hat{s} = x_1 x_2 s$ où $s = (p_1 + p_2)^2$. Au premier ordre, la section efficace $\hat{\sigma}(q\bar{q} \rightarrow l^+l^-)$ est facilement calculable. Aux énergies du LHC, la section efficace du Drell-Yan présente 3 contributions : une pour le photon, une pour le boson Z et un terme d'interférence entre

photon et Z :

$$\sigma_{(q\bar{q} \rightarrow l^+ l^-)} = \frac{\pi \alpha^2}{3s} \frac{1}{N_c} \left[(e_q^2 - 2e_q g_{V,l} g_{V,q} S_{\gamma/Z} + (g_{A,l}^2 + g_{V,l}^2)(g_{A,q}^2 + g_{V,q}^2) S_Z(s) \right] \quad (2.26)$$

$$S_{\gamma/Z} = k \frac{s(s - M_Z^2)}{(s - M_Z^2)^2 + \Gamma_Z^2 M_Z^2}$$

$$S_Z = k^2 \frac{s^2}{(s - M_Z^2)^2 + \Gamma_Z^2 M_Z^2}$$

Où e_q est la charge fractionnaire du quark q , g_V, g_A et k sont définis en 2.1.2.

2.3.1 Radiation de gluons associée au processus DY

$$q\bar{q} \longrightarrow Z + jets$$

La section efficace prenant en compte les corrections radiatives au premier ordre dépend des échelles de factorisation μ_F et de renormalisation μ_R . On prend de façon arbitraire, $\mu_R = \mu_F = \mu$ et dans notre cas $\mu = M_Z$.

Les contributions supplémentaires en ordre α_s sont représentées dans le tableau 2.5. Du point de vue expérimental, la contribution de ces diagrammes se manifeste par la

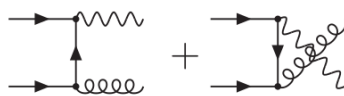
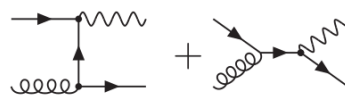
processus d'annihilation	processus QCD
$q + \bar{q} \rightarrow g + \gamma^*$	$q + g \rightarrow q + \gamma^*$
	

FIGURE 2.5: Exemples de diagrammes constituant les corrections au NLO au processus Drell-Yan

présence d'un jet dans l'état final. Ces corrections sont importantes (facteur 2 sur la section efficace) car une nouvelle classe de diagramme est à prendre en compte : ceux faisant intervenir la densité de gluon qui, comme on l'a vu pour notre domaine cinématique est parfois 10 fois plus importante que la densité de quarks.

Pour les ordres supérieurs en QCD, d'autres types de diagrammes vont rentrer en jeu, en faisant intervenir également les contributions des vertex à trois gluons. Le nombre de diagrammes devient vite très grand à mesure qu'on augmente en puissance de α_s .

La section efficace du DY est actuellement déterminée au NNLO mais nous verrons au chapitre 4 comment les logiciels de simulation génèrent les ordres supérieurs associés au processus DY sans pour autant devoir utiliser les sections efficaces du DY aux ordres supérieurs.

Chapitre 3

Dispositif expérimental : le LHC et CMS

Dans ce chapitre nous présentons les principales caractéristiques expériences ayant produit les données utilisées pour notre analyse : l'accélérateur proton-proton LHC et le détecteur de particules CMS.

3.1 LHC : Large Hadron Collider

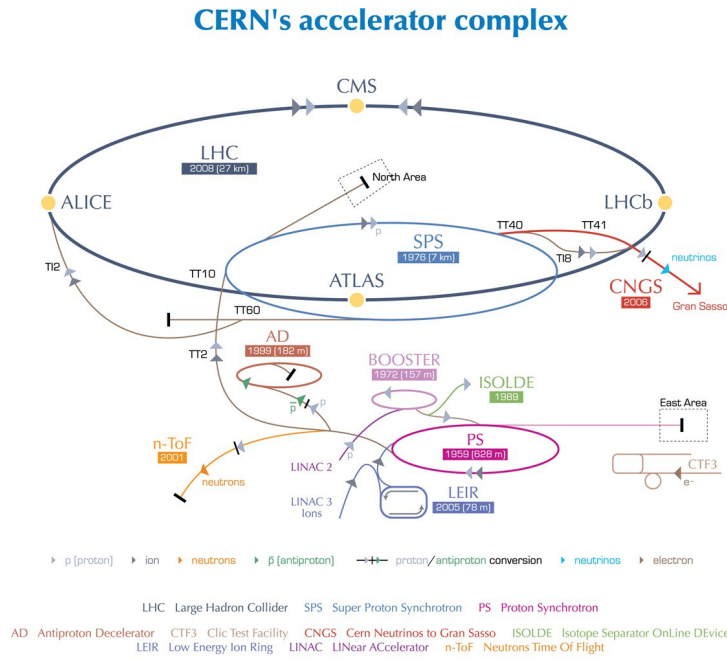


FIGURE 3.1: Chaîne des accélérateurs au CERN [6] : les protons injectés dans le LHC sont accélérés par la chaîne Linac2 $\rightarrow$ Booster $\rightarrow$ PS $\rightarrow$ SPS.

Le LHC a été construit entre 1998 et 2008 grâce à une collaboration mondiale sur le site du Centre Européen de Recherche Nucléaire (CERN) à la frontière franco-suisse.

Il occupe un tunnel de 27 km de circonférence avec une profondeur moyenne de 100 m, précédemment occupé par le LEP (Large Electron Positron Collider).

Le LHC est plus précisément un accélérateur proton-proton (p-p) conçu pour produire des collisions à une énergie nominale de 7 TeV par faisceau (14 TeV dans le centre de masse). Il a la particularité de pouvoir également fonctionner avec des faisceaux d'ions de plomb à une énergie nominale de 1,38 TeV par faisceau.

Son programme est multiple et comprend l'étude du boson de Brout-Englert-Higgs et des mesures de précision dans le cadre du modèle standard ainsi que l'exploration de la nouvelle physique au delà du modèle standard (particules super-symétriques, particules de matière noire, étude de dimensions supplémentaires) et l'étude du plasma quarks-gluons [7].

Le projet de construire un accélérateur pouvant atteindre de telles énergies a été décidé vers 1996. A l'époque, le TeVatron (1987-2011 U.S.[8]), collisionneur proton-antiproton, détenait le record en énergie avec 0.98 TeV par faisceaux. En Europe les principaux accélérateurs précédents sont le LEP (CERN 1989-2000) et HERA (DESY Hambourg 1992-2007). Les expériences qui s'y sont déroulées ont permis de confirmer le modèle standard jusqu'à une énergie d'environ 1 TeV et de mesurer précisément ses paramètres, sans pour autant confirmer l'existence du boson de Brout-Englert-Higgs.

Le démantèlement du LEP et la construction du LHC ont nécessité plus de 8 ans : le LEP a définitivement cessé ses activités en 2000 et les premiers faisceaux sont lancés au LHC le 10 septembre 2008. Malheureusement des incidents techniques le maintiennent immobilisé pendant un peu plus d'un an. L'accélérateur a redémarré le 23 octobre 2009 et le 7 novembre 2009, les premières particules étaient enregistrées dans le détecteur CMS. Le 30 novembre 2009, le CERN annonce avoir fait circuler dans le LHC le faisceau le plus énergétique du monde en ayant conféré aux protons une énergie de 1.18 TeV. En 2010, l'énergie record de 3.5 TeV par faisceaux est atteinte. A cette énergie, de précieuses données vont être enregistrées entre 2010 et 2011 avec une luminosité intégrée de 5 fb^{-1} . Ces données ont permis, entre autres, de réduire la fenêtre de recherche du boson de Brout-Englert-Higgs à 116-126 GeV. En 2012, l'énergie des collisions atteint le nouveau record de 8 TeV pour une luminosité intégrée de 19.3 fb^{-1} (cf. tableau 3.1).

Ces données ont permis la découverte du boson de Brout-Englert-Higgs, à une masse de $125 \text{ GeV}/c^2$ [9]. Ce travail se base également sur ces données.

Description technique

Le LHC est constitué de huit sections droites utilisées pour les collisions et autant de sections courbes équipées de dipôles magnétiques pour maintenir les faisceaux sur leurs trajectoires. Sur les sections droites se trouvent les 4 détecteurs principaux : ATLAS, CMS, ALICE et LHCb. ATLAS et CMS sont diamétralement opposés aux points de plus grande luminosité. Ces sont deux détecteurs dédiés à la recherche en physique des particules des hautes énergies et conçus pour utiliser tout le spectre d'énergie du LHC. ALICE est une expérience d'étude des collisions ion-ion et en particulier du plasma de quarks et gluons, et LHCb travaille à luminosité réduite et étudie principalement les violations de CP à partir de l'étude du quark b.

De nombreuses autres expériences se trouvent sur le site du CERN. Le tableau 3.2 les décrit de manière succincte. Dans ce travail, nous nous intéresserons uniquement à l'expérience CMS.

Accelerator	Experiments	Overview
LINAC2		accelerates proton from the source and sends them to the booster (PSB)
LINAC3		accelerates ions from the source and sends them to LEIR
PSB		accelerates protons and sends to the PS and ISOLDE
PS	EA	East Area - various experiments
	nTOF	neutron time-of-flight
	DIRAC	to observe and then measure lifetimes of Muons and Kaons
	CLOUD	to study possible links between cosmic rays and cloud formation
ISOLDE		to produce a range of isotopes for research
AD	ALPHA	to make, capture and study atoms of antihydrogen and compare these with hydrogen atoms
	ASACUSA	to compare anti-protons and protons using antiprotonic helium
	ATRAP	to compare hydrogen atoms with their antimatter equivalents
LEIR		accelerates ions and sends them to the PS
SPS	NA	North Area - various experiments
	CNGS	to send muon neutrinos to the Gran Sasso National Laboratory in Italy
	COMPASS	to study how elementary quarks and gluons work together to give particles we observe
LHC	CMS	to search for the Higgs boson, extra dimensions, and particles that could make up dark matter
	ATLAS	to search for the Higgs boson, extra dimensions, and particles that could make up dark matter
	LHCb	to understand why we live in a Universe composed almost entirely of matter, but no antimatter
	ALICE	to study a state of matter known as quark-gluon plasma
	TOTEM	to measure the size of the proton and also monitor the LHC's luminosity
	LHCf	to simulate cosmic rays to interpret and calibrate large-scale cosmic-ray experiments

FIGURE 3.2: Liste non-exhaustive des expériences au CERN [6].

Nous allons maintenant nous pencher sur des aspects plus techniques du LHC en tant qu'accélérateur circulaire de particules [10] : la chaîne d'accélération, le transport des faisceaux, et la luminosité.

Chaîne d'accélération

Le LHC est le maillon final d'une chaîne complexe représentée sur la figure 3.1. Des accélérateurs précédemment construits au CERN sont utilisés afin de pré-accélérer les faisceaux pour le LHC.

Les protons sont créés à partir de gaz d'hydrogène ionisé (système de duoplasmatron) et sont ensuite accélérés dans un premier accélérateur linéaire, le LINAC 2. Le faisceau atteint alors une énergie d'environ 50 MeV. Ensuite 3 accélérateurs circulaires vont se succéder : le Proton Synchrotron Booster (PSB) forme des paquets de protons et augmente l'énergie du faisceau jusqu'à 1.4 GeV, le Proton Synchrotron (PS) augmente l'intensité des paquets et les accélère jusqu'à 25 GeV et enfin, le Super Proton Synchrotron (SPS) injecte les faisceaux dans le LHC avec une énergie de 450 GeV.

Par ailleurs, les éléments de la chaîne d'accélération permettent également d'alimenter les expériences nécessitant des faisceaux de différentes énergies et de différentes natures, comme des ions mais également des anti-protons pour *AD*, des neutrons pour *n-ToF* ou des neutrinos pour *CNGS* (cf. figures 3.1 et 3.2).

Transport des faisceaux

En 2012, les faisceaux de protons du LHC possédaient une énergie de 4 TeV. Chaque faisceau contenait 1380 paquets ("bunchs") de 1.6×10^{11} protons espacés de 50 ns. Les faisceaux circulent dans deux tubes à ultravide (10^{-13} atm). Les faisceaux doivent être accélérés, guidés et focalisés : ceci nécessite l'intervention de cavités à radio-fréquence et de pas moins de 9593 aimants distribués sur la circonférence de 26659 m.

Pour guider les faisceaux, 1232 dipôles magnétiques de niobium-titane, d'environ 15m de long, sont placés sur les parties courbées. Les aimants dipolaires fournissent un champ magnétique de 4.5 T (une valeur nominale de 8.33 T est prévue). Ce champ extrêmement fort est atteint grâce à un système de refroidissement à base d'hélium superfluide sous pression qui maintient les dipôles à une température de -271°C . Dans la même enceinte cryogénique 392 quadripôles supraconducteurs sont également utilisés, principalement sur les sections droites. Ils mesurent de 5 à 6 m de long et servent à focaliser le faisceau ainsi qu'à régler la convergence aux points d'interactions.

Enfin, l'accélération du faisceau est réalisée par 16 cavités radiofréquence (8 par faisceau). Ces cavités délivrent chacune un champ accélérateur de 5 MV.m^{-1} . La fréquence des cavités varie de 400.789 MHz à l'injection (450 GeV) à 400.790 MHz (pour la valeur nominale). L'énergie de 4 TeV est atteinte au bout d'environ 20 min après injection des faisceaux dans le LHC.

Design "two-in-one" : Pour élaborer son système de transport des faisceaux, le LHC a bénéficié de l'expérience acquise sur les collisionneurs précédents (LEP, HERA, TeVatron, SLAC). A la différence de ces collisionneurs, le LHC a choisi de transporter les faisceaux côte à côte, dans une même culasse d'environ 50 cm de diamètre et avec un système supra-conducteur sur toute la longueur. On parle de design "two-in-one" ou "twin-bore" (cf. figure 3.3). Cette prouesse technique, en plus d'économiser de l'espace, a permis d'épargner 25% du coût technique.

Une des conséquences est la structure complexe des aimants : dans le cas des aimants dipolaires, par exemple, la séparation entre les deux tubes à faisceau est tellement petite que les deux aimants risquaient de se coupler magnétiquement et mécaniquement. On observe sur la figure 3.3 (à droite) la distribution maîtrisée des champs magnétiques.

Une fois les paquets formés et accélérés, leur étalement longitudinal et transverse varie selon l'action des éléments magnétiques. Cet étalement transverse, et plus généralement le profil des faisceaux, a une conséquence directe sur l'émission, et donc sur la luminosité, comme nous allons le voir. Certaines caractéristiques techniques des faisceaux du LHC sont reprises dans le tableau 3.1.

La luminosité instantanée

La luminosité est une grandeur qui caractérise les faisceaux en un point donné. En général, la luminosité d'un processus donné correspond au rapport du nombre d'événements N et de la section efficace de l'interaction σ_{int} .

$$L = \frac{N}{\sigma_{int}}$$

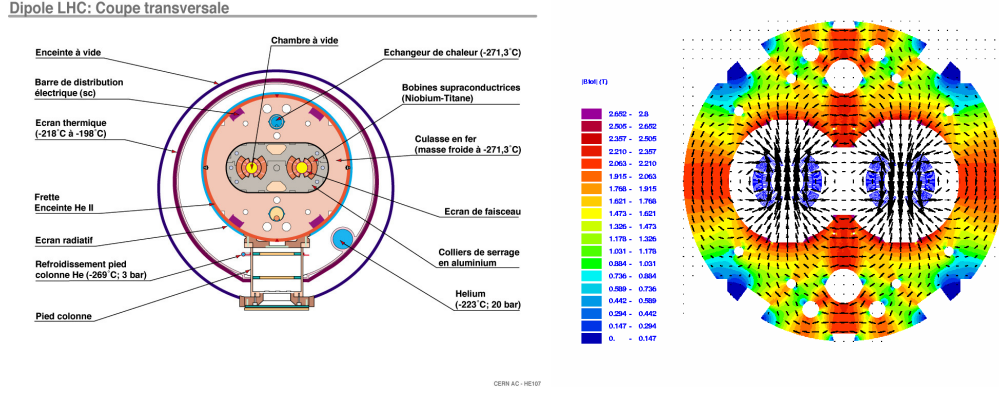


FIGURE 3.3: A gauche, coupe transversale d'un élément dipolaire du LHC. A droite répartition de son champ magnétique.

Dans le cas des accélérateurs circulaires, la luminosité s'exprime soit en fonction des paramètres sections efficaces transverses σ_x , σ_y ou de l'émittance transverse normalisée ϵ_n . Ces deux formules sont équivalentes seulement si l'on considère que les faisceaux ont un profil transverse gaussien. L'émittance s'exprime alors comme $\epsilon_n = \sigma_x \sigma_y \gamma / \beta^*$ [11].

$$L \simeq \frac{N_1 N_2 n_b f_{rev}}{4\pi \sigma_x \sigma_y} \simeq \frac{N_1 N_2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta^*} \quad (3.1)$$

Les autres paramètres apparaissant dans la formule sont : la fréquence de révolution des paquets (f_{rev}), le nombre de proton par paquets (n_b), le nombre de paquets par faisceaux (N), le facteur β^* , et le facteur de Lorentz γ . La luminosité instantanée s'exprime par unité de surface et par unité de temps (e.g : $\text{cm}^{-2} \cdot \text{s}^{-1}$).

β^* est un paramètre de collision utilisé pour caractériser le pouvoir de convergence des quadripôles du point d'interaction des détecteurs. β^* correspond à la distance longitudinale entre le point d'interaction et le point où la section transverse des faisceaux est deux fois plus importante. En 2012, β^* a atteint 0.6 m, presque sa valeur nominal de 0.55 m.

On voit dans la formule 3.1 que la luminosité peut être augmentée de différentes manières : en minimisant le paramètre β^* , ou en maximisant le rapport N_b^2 / ϵ_n ("beam brightness"), ou encore en augmentant le nombre de protons par faisceau $N^2 n_b$... Ainsi, pour les données 2012, on a augmenté N^2 et n_b et baissé ϵ_n et β^* (cf. tableau 3.1). Ceci a permis de conserver une séparation de 50 ns entre chaque bunches, et de diminuer l'empilement dû au chevauchement des collisions entre paquets ("out of time pileup"), tout en tournant à plus haute luminosité qu'en 2011.

La luminosité intégrée

La luminosité intégrée sur une certaine période T est calculée en sommant au cours du temps la luminosité instantanée :

$$L = \int_0^T L(t) dt \quad (3.2)$$

La figure 3.4 présente, pour 2012, les mesures des pics de la luminosité instantanée (gauche) et la luminosité intégrée (droite) pour les quatre expériences CMS, ATLAS, ALICE et LHC-b.

Pour augmenter la statistique des données, on peut donc à même luminosité instantanée, augmenter la durée de prise de données. Toutes ces actions visant à augmenter la luminosité du LHC ont leur limites techniques et dépendent de nombreux paramètres. Mais ce sont des pistes pour les prochaines phases de fonctionnement du LHC [12].

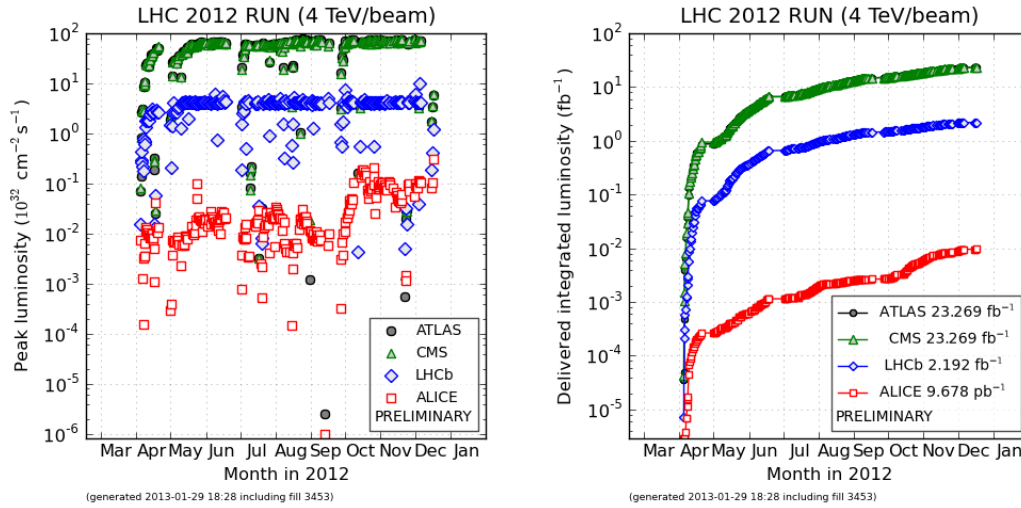


FIGURE 3.4: Pics de luminosité instantanée (gauche) et luminosité intégrée (droite) pour l'année 2012 [13].

Paramètres	Valeur nominale	2011	2012
Energie par faisceaux (TeV)	7	3.5	4
champ magnétique des dipôles (T)	8.33	4.16	4.5
Nombre de paquets par faisceau	2808	1303	1380
Ecart entre paquets (ns)	25	50	50
Nombre de protons par paquets	1.5×10^{11}		1.6×10^{11}
Fréquence de révolution d'un paquet (kHz)	11.246		
Emittance transversale ($\mu m.rad$)	3.75	3.5	2
β^* (m)	0.55	1.0	0.6
Luminosité instantanée max. ($cm^{-2}.s^{-1}$)	1.0×10^{34}	3.6×10^{33}	0.55×10^{34}
Luminosité intégrée ($fb.an$) ⁻¹	100	5.725	19.3

TABLE 3.1: Paramètres des faisceaux de proton et des collisions au LHC. Valeurs nominales comparées aux valeurs atteintes en 2011 [14] et en 2012 [12]

Calendrier et upgrades du LHC

En 2013, Le LHC a tourné en Ion-proton au mois de février et ensuite des tests de maintenance ont été effectués jusqu'en avril, parfois en faisceau. Depuis Pâques, le LHC est à l'arrêt pour une longue période prévue jusqu'à fin 2014 (phase de "Long Shutdown", LS1). Cette pause va permettre, entre autres, d'améliorer la chaîne d'accélération. A l'issue du LS1, les collisions devraient reprendre avec une énergie de 6.5 TeV par faisceau puis atteindre 7 TeV avant 2018. A plus long terme, on prévoit d'atteindre l'énergie optimale de 8 TeV par faisceau après une autre phase d'arrêt, vers 2020. Les phases d'arrêt sont aussi des périodes d'upgrade au niveau des expériences elles-mêmes, les détecteurs devant faire face aux nouvelles performances des faisceaux.

Nous allons maintenant nous pencher en détail sur l'expérience CMS et son détecteur.

3.2 CMS : Compact Muon Solenoid

CMS est un détecteur cylindrique multifonctions au design compact (21,6m de long, 14.6m de diamètre). Il est caractérisé par un solénoïde délivrant un champ magnétique de 4T. Le détecteur présente une partie centrale, le tonneau (barrel en anglais), et deux bouchons (endcaps) qui lui confère une couverture angulaire proche de 4π .

La figure 3.12 (haut) présente une description des différentes composantes de CMS : en blanc, le solénoïde supraconducteur. A l'intérieur du solénoïde, se trouve une structure en couches comprenant les trajectographes (bleus), et les calorimètres (électromagnétiques (cyan) et hadroniques (en beige)). Les détecteurs à muons (en blanc) se trouvent à l'extérieur du solénoïde, disposées de manière alternée avec des structures ferromagnétiques (rouge) assurant le retour du champ magnétique.

La figure 3.12 (bas) résume les interactions des différentes particules dans le détecteur : chaque particule chargée laisse une trace courbée dans le système de tracker. Les leptons et photons déposent leur énergie dans les calorimètres électromagnétiques, les hadrons, neutres ou chargés marquent les calorimètres hadroniques. Enfin, les muons traversant tous les calorimètres presque sans perdre d'énergie, interagissent à grand rayon avec les chambres à muons. D'un point de vue humain, le détecteur est pris en charge par une collaboration de plus de 4000 ingénieurs, physiciens et étudiants, issus de 179 laboratoires et universités de 41 pays différents.

Enjeux expérimentaux et choix de construction :

Les caractéristiques du LHC posent de grands challenges aux expériences de détection : les paquets de protons se succèdent à des intervalles de 50 ns (25 ns prévus) et le taux d'événements moyen par collision est d'environ 25^1 . Le challenge pour CMS consiste donc à enregistrer jusqu'à 10^9 collisions par secondes avec une grande précision. Ceci implique de pouvoir rafraîchir tous les détecteurs suffisamment vite, réduire le volume des données afin de pouvoir les stocker, et prévoir des composants de détection réagissant sans perte de performance à une radiation intense et permanente.

1. $N_{ev} = \frac{\text{nbr d'év.} / \text{s}}{\text{nombre de paquets} / \text{s}} = \frac{10^9}{40 \cdot 10^6} = 25$

La collaboration CMS a répondu à ces exigences en installant des systèmes de détecteurs de haute granularité², de grande robustesse face aux radiations, et d'un hermétisme parfait.

Nous allons passer en revue le système de coordonnées, les différents sous-détecteurs ainsi que le solénoïde et enfin le système de déclenchement de CMS. Par commodité, nous utiliserons certaines dénominations anglaises (tracker pour trajectographe, trigger pour système de déclenchement, pileup pour l'empilement...) et différents acronymes listés en annexe .

3.2.1 Système de coordonnées

Le système de coordonnées de CMS a pour origine le point d'intersection des faisceaux. L'axe z correspond à l'axe des faisceaux et est dirigé dans le sens inverse des aiguilles d'une montre. L'axe des x est dirigé vers le centre du cercle formé par le LHC, et l'axe des y vers le haut. Mais les coordonnées cartésiennes sont rarement utilisées, la structure cylindrique de CMS amenant naturellement à un système de coordonnées polaires (r, θ, ϕ) . Le rayon r est la distance radiale à l'axe z , ϕ est l'angle azimutal par rapport à l'axe x et θ l'angle polaire mesuré à partir de l'axe z .

On utilise encore plus couramment, et ce sera le cas dans ce travail, le système de coordonnées (r, η, ϕ) , qui diffère du système précédant par une redéfinition de la coordonnée polaire. On définit la pseudo-rapacité η comme

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

A $\theta = 0$ correspond $\eta = \infty$, à $\theta = 90^\circ$ correspond $\eta = 0$ et un angle de $\theta = 45^\circ$ donne un η d'environ 0.88.

La variable η est appelé pseudo-rapacité car on peut dériver son expression de la rapidité, dans le cas ultra-relativiste ($mc^2/E \rightarrow 0$), soit :

$$\eta^* = -\frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right)$$

L'intérêt de la variable η est que la quantité $\Delta\eta$ est invariant sous les boost en z .

2. La notion de granularité définit la taille du plus petit élément, de la plus grande finesse d'un système. Elle rend compte de la résolution spatiale et temporelle d'un détecteur.

3.2.2 Le système de tracker

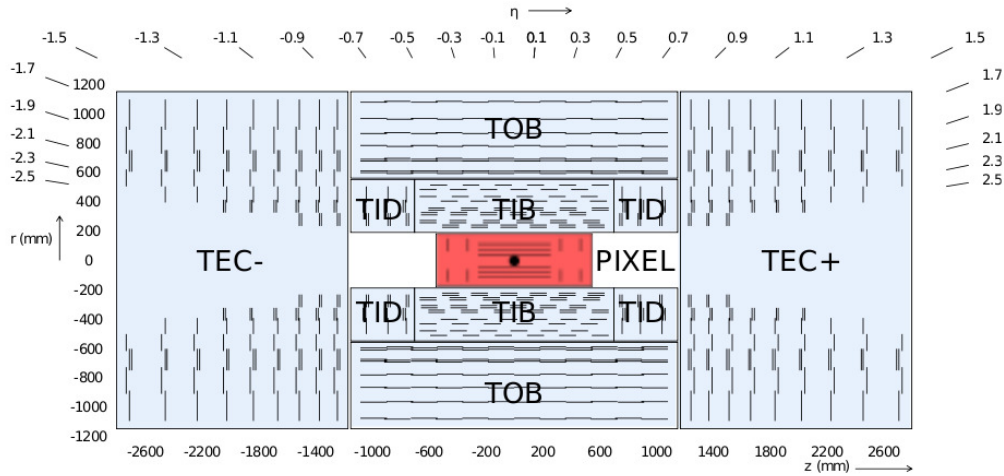


FIGURE 3.5: Le système de tracker de CMS : au centre en rouge le tracker à Pixels, en bleu les différentes composantes du tracker à micro-piste.

Les trackers permettent de reconstituer la trajectoire des particules issues des collisions et les vertex secondaires. La mesure doit être la plus complète et précise possible pour une reconstruction optimale des événements et c'est notamment primordial dans le cas du $b \tau$ - tagging. De plus, le tracker de CMS est positionné au plus près du faisceau, et cette proximité l'expose à une forte radiation. Le choix de CMS s'est porté sur un double tracker utilisant des détecteurs supra-conducteurs au silicium, supportés dans une structure en fibre de carbone. On distingue un tracker de haute résolution à pixels et un tracker à micro-pistes. Dans ces détecteurs, les supraconducteurs sont couplés, de façon hybride (dopés de différentes façons) et fonctionnant sur le principe des jonctions n-p, maintenues en mode de polarisation inverse par le champ magnétique. Le passage d'une particule chargée ionise les zones de dépeuplement et un courant de réponse est enregistré.

Nous allons maintenant passer en revue les deux composantes du tracker.

3.2.2.1 Tracker à pixels

Le tracker à pixels est constitué de trois couches de pixels hybrides, placées dans le tonneau à des rayons de 4.4, 7.3, et 10.2 cm. Deux disques de diamètre 6 et 15 cm sont placés sur chaque côté en guise de bouchon à $|z| = 34.5\text{cm}$ et $|z| = 46.5\text{cm}$. Ainsi le tracker à pixel couvre une surface totale de $\simeq 1 \text{ m}^2$ et une région jusqu'à $|\eta| < 2.5$. Pour éviter les dommages dus à la forte radiation, la couche de pixels se trouvant à l'intérieur est remplacée par une couche $r = 13\text{cm}$ quand le LHC tourne à luminosité élevée.

Dans tous les cas, le tracker à pixel est configuré pour présenter trois points d'impact permettant de reconstituer la trajectoire à haute résolution.

Le tracker à pixels comprend au total 66 millions de pixels : chaque micro-pixel a une surface de $100 \times 150 \mu\text{m}^2$ et une épaisseur de $300 \mu\text{m}$. La résolution du tracker à pixels a été mesurée à environ $10 \mu\text{m}$ dans le plan (r, ϕ) , et environ $20 \mu\text{m}$ dans la direction z .

3.2.2.2 Tracker à micro-pistes de silicium

A plus grande distance du faisceau, le flux de particules est déjà réduit, et CMS a opté pour un tracker à micro-pistes de silicium. Sa résolution est plus basse, entre 23 et $52\mu m$ dans le plan (r, ϕ) , et entre 230 μm (pour TIB) et 530 μm (pour TOB) selon z . La surface totale est bien plus grande que pour le tracker à pixels ($\simeq 200\text{ m}^2$).

Le tracker à micro-pistes entoure le tracker à pixel. Dans la région du tonneau, 10 couches de détecteurs à micro-pistes de silicium sont placés entre $r = 20\text{ cm}$ et $r = 110\text{ cm}$. Du côté des bouchons ce sont 12 couches de micro-pistes qui assurent la couverture angulaire jusqu'à $|\eta| < 2.5$.

Certains modules sont doubles (cf. figure 3.5). En réalité ces doublons font un angle de 100 μrad ce qui permet une mesure bi-dimensionnelle dans les deux plans (r, ϕ) et (r, z) .

Les composantes constituant le tracker à pistes (TIB, TOB, TEC, TID) diffèrent par leur résolution et par l'épaisseur des pistes de silicium utilisées. A mesure que la radiation diminue des pistes plus épaisses peuvent être utilisées tout en gardant un bon rapport signal sur bruit.

Le tracker à micro-pistes est optimisé géométriquement pour que le passage d'une particule laisse au moins 9 points d'impact dont au moins 4 ayant une résolution bi-dimensionnelle.

3.2.3 Le calorimètre électronique (ECAL)

Le Calorimètre électronique entoure le système de tracker : il est constitué d'un tonneau central d'un rayon de 129 cm et une longueur de 628 cm, ainsi que de deux bouchons (cf. figure 3.6). Il mesure l'énergie et la direction des particules interagissant par l'intermédiaire de la force électromagnétique : les électrons, les positrons³ et les photons.

C'est un calorimètre à scintillation constitué de 75848 cristaux de tungstate de plomb PbWO_4 . Ce matériau scintillant a été choisi pour sa densité de 8,28 g.cm^{-3} permettant une courte longueur de radiation ($X_0 = 0,89\text{ cm}$), et un faible rayon de Molière⁴ ($R_m = 22\text{ mm}$).

Ces propriétés confère à l'ECAL une réponse rapide (environ 80% de la lumière émise peut être récoltée en 25ns), une granularité élevée, et une bonne résolution en énergie.

Plus précisément, chaque cristal présente une surface de $22 \times 22\text{ mm}^2$ qui correspond au rayon de Molière, et la longueur des cristaux est de 230mm soit $25.8 X_0$. Spatialement, les cristaux sont décalés de 3° par rapport à leur direction vers le point d'interaction afin d'éviter les imprécisions dues à l'alignement avec la trajectoire de la particule. L'ECAL du tonneau (EB) couvre une surface angulaire jusqu'à $|\eta| < 1.479$. L'ECAL des bouchons (EE) est organisé sous forme de disques, couvrant une surface allant de $1.479 < |\eta| < 3$. Sur les bouchons, des pré-détecteurs à gerbe (Electromagnétique Preshower (ES)) sont positionnés avant les EE. Ils sont composés de 2 détecteurs micro-pistes en silicium et des disques absorbant en plomb. Ces pré-détecteurs permettent d'améliorer la discrimination γ/π_0 en fournissant une mesure plus précise des germes de gerbes électromagnétiques.

3. Dans ce travail, pour la clarté de rédaction, nous oublierons volontairement de citer les positrons associés de fait à l'appellation "électrons".

4. Rayon de Molière = rayon du cylindre contenant 90% des gerbes électromagnétiques déposées dans un cristal.

Notons enfin que l'ECAL présente une zone "morte" généralement exclue de l'analyse : cette zone couvre des pseudo rapidités comprises entre 1.479 et 1.653.

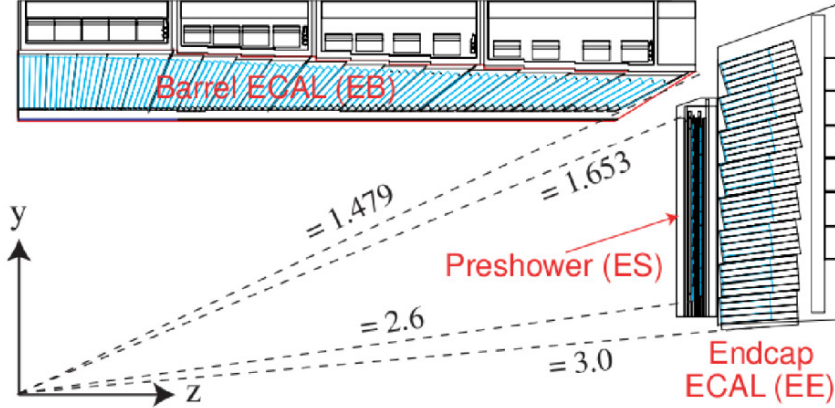


FIGURE 3.6: Coupe transversale de L'ECAL, vue 1/4.

La résolution d'un calorimètre est paramétrée par l'équation 3.3

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (3.3)$$

S est un terme stochastique, N un paramètre de bruit, et C une constante. Le terme stochastique S rend compte des fluctuations de l'étalement de la gerbe électromagnétique due à la photostatistique. N rend compte du bruit induit par l'électronique, la digitalisation et le pileup. Enfin, les dépendances linéaires en énergie, comptés dans C , proviennent de la construction de l'ECAL : la collection longitudinale de la lumière n'est pas uniforme, il existe des erreurs dues à l'intercalibration et une perte de rayonnement par l'arrière des cristaux. Les tests en faisceaux ont permis de déterminer ces paramètres pour l'ECAL de CMS qui présente la résolution suivante pour l'énergie exprimée en GeV :

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2 \quad (3.4)$$

3.2.4 Le calorimètre hadronique (HCAL)

Entre le calorimètre électronique et le solénoïde se trouve le détecteur de hadrons, ou HCAL. Le calorimètre hadronique mesure l'énergie déposée par les hadrons neutres et chargés. Ces dépôts vont permettre de déterminer l'énergie et la direction de jets et du parton les ayant initiés. La production de jets au LHC est massive et le challenge est donc important pour le HCAL.

L'HCAL a aussi pour but de donner une indication sur le flux d'énergie manquante (Missing Energie Transverse MET) et identifier les neutrinos. Pour cela, une herméticité parfaite est nécessaire jusqu'à $|\eta| < 5$. Les détecteurs du HCAL présentent une structure alternant couches d'absorption en acier et en laiton et entre les deux des capteurs à scintillation qui permettent de récolter le rayonnement émis par les hadrons ayant perdu en énergie.

Les couches absorbantes sont définies par leur longueur d'interaction (λ_I) dépendant de l'épaisseur et du métal composant la plaque. On a construit un HCAL en 4 parties, les composantes des tonneaux et bouchons et une composante couvrant la partie avant du faisceau (HF) ($3 < \eta < 5.2$), ainsi qu'une composante positionnée en dehors du solénoïde (HO). La figure 3.7 représente les différentes installations du HCAL dans CMS et les régions en η couvertes.

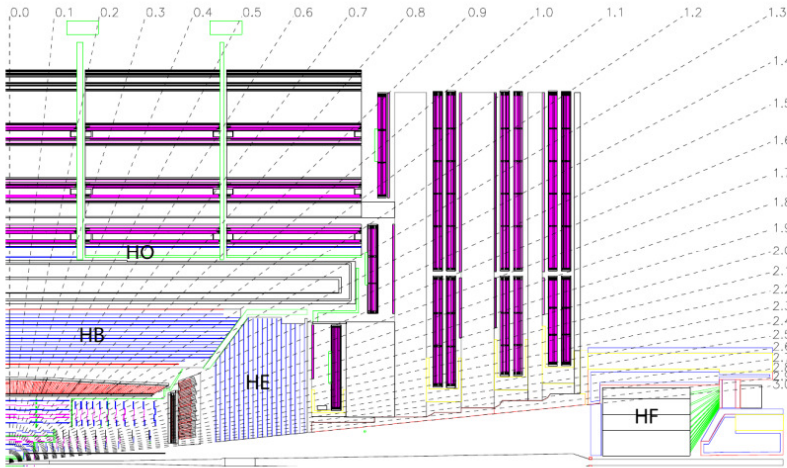


FIGURE 3.7: Coupe transverse et vue 1/4 du HCAL de CMS.

- **Barrel Hadronic Calorimeter (HB)** : c'est la partie du HCAL qui couvre le tonneau entre l'ECAL et le solénoïde. Sa couverture en η va jusqu'à 1.4. Le HB est constitué d'une plaque d'acier puis de 8 plaquettes de laiton, et en fermeture 1 couche de laiton et une plaque d'acier. L'épaisseur totale du HB atteint donc 22.25 cm soit $5.82 \lambda_I$. Le système de détection du HB utilise 70000 tuiles de plastique scintillant dotées d'une bonne stabilité sur le long terme. L'HB est relié au système acquisition des données par des fibres scintillantes. Le HB présente une segmentation en $\Delta\eta \times \Delta\phi$ de $0,087 \times 0,087$.
- **Endcap Hadronic Calorimeter (HE)** : le calorimètre hadronique des bouchons est également placé à l'intérieur du solénoïde. Il étend la couverture du HB à $|\eta| = 3$. Pour la composition, l'HE diffère de l'HB par l'épaisseur des couches de laiton (79mm au lieu de 55cm). L'HE a une granularité de 0.087×0.087 pour $|\eta| < 1.6$ et de 0.017×0.017 pour $|\eta| > 1.6$.

- **Outer Hadronic Calorimeter (HO)** : la partie extérieure du calorimètre est accolée à la face extérieure du tonneau du solénoïde. Cette composante extérieure profite de la couche du solénoïde comme couche absorbante supplémentaire à raison de $1.4/(\sin(\theta).\lambda_I)$. Il assure que toutes les particules chargées passant par le tonneau soient absorbées même si la gerbe hadronique s'étend au delà du HB. Il ajoute ainsi une composante à la longueur d'interaction totale dans le tonneau pour atteindre au totale $11.8 \lambda_I$. Le HO est composé de 2 couches de scintillateurs de même granularité que le HB
- **Forward Hadronic Calorimeter (HF)** : le HF constitue la composante du HCAL la plus près du faisceau. Elle est positionnée à 11.2 m du point d'interaction et étend la zone couverte à $|\eta| = 5.2$. Pour le HF, la radiation est beaucoup plus forte⁵, et on a dû changer de matériaux : le scintillateur se compose de fibres de quartz entourées par un absorbeur d'acier.

La résolution en énergie du HCAL de CMS ne présente qu'un terme stochastique et une constante. Le terme de bruit est négligeable.

Pour le HB et le HE, on obtient (pour E en GeV) :

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{90\%}{\sqrt{E}}\right)^2 + (4.5\%)^2 \quad (3.5)$$

Pour le HF, la résolution est donnée par (pour E en GeV) :

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{172\%}{\sqrt{E}}\right)^2 + (9.0\%)^2 \quad (3.6)$$

3.2.5 L'aimant solénoïdal supraconducteur

L'électroaimant supraconducteur mesure 6 m de diamètre 31 cm d'épaisseur, 12,5 m de long, et pèse plus de 200 tonnes. Il délivre un champ magnétique uniforme et stable de 3,8 Teslas à l'intérieur du solénoïde et un champ décroissant à l'extérieur (cf. figure 3.8). C'est la pièce maîtresse de CMS : il permet la compacité de l'expérience et contraint le choix et la disposition des sous-détecteurs. C'est aussi le plus grand aimant solénoïdal du monde.

L'aimant solénoïde est, avec le tracker, un élément essentiel dans la reconstruction des énergies transverses et des directions. Pour se rendre compte de cela, il suffit de regarder la formule donnant le rayon de courbure de la trajectoire d'une particule chargée soumise à un champ magnétique uniforme :

$$R = |\vec{p}_T|/|q||\vec{B}| \quad (3.7)$$

Avec $|\vec{p}_T|$, le module de l'impulsion transverse (en kg.m.s⁻¹), $|\vec{B}|$ et $|q|$, les modules du champ magnétique en Teslas et de la charge en coulombs.

Ainsi, pour une certaine résolution sur la mesure de R, plus le champ magnétique est grand, plus la résolution en $|\vec{p}_T|$ est élevée.

5. Environ 760 GeV par interaction p-p sont déposées dans les HF alors que 100 GeV en moyenne sont enregistrés dans le reste des composantes.

A contrario, à $|\vec{B}|$ fixé, si l'impulsion des particule est trop petite, la trajectoire va se refermer en spirale sur elle-même et ne sort pas du tracker. En ce sens, le solénoïde contribue également à réduire les bruits de fond issus des particules de faible énergie.

Le solénoïde est composé de cinq bobines formées de 20 tresses de niobium-titane, mesurant chacune 2,65 km de long. Les bobines sont baignées dans de l'hélium liquide, permettant d'atteindre une température de 1.8 K nécessaire au régime supra-conducteur. Une enceinte sous vide isole thermiquement le bobinage : elle s'étend de $r=2.9$ m à $r=3.7$ m.

A l'extérieur du solénoïde un ensemble de culasses ferromagnétiques permet de refermer le champ magnétique sur lui-même et de le maintenir le plus uniforme possible. Ces culasses de retour de champ sont constituées de plaques d'acier entre lesquelles sont intercalées les chambres à muons. Ces plaques constituent la majeure partie du poids global de CMS avec 10000 tonnes sur 12500.

Des mesures du champ magnétique du solénoïde ont été effectuées avec des détecteurs à effet Hall et en utilisant les traces laissées par les muons cosmiques. Une carte des champs a pu être établie, sans toutefois obtenir une grande précision de ce qui se passe au niveau des bouchons. La figure 3.8 présente la distribution simulée de champ magnétique dans CMS à partir des données. Cette simulation considère seulement la moitié du solénoïde, l'autre moitié étant obtenue par symétrie, ce qui ne correspond pas vraiment à la réalité.

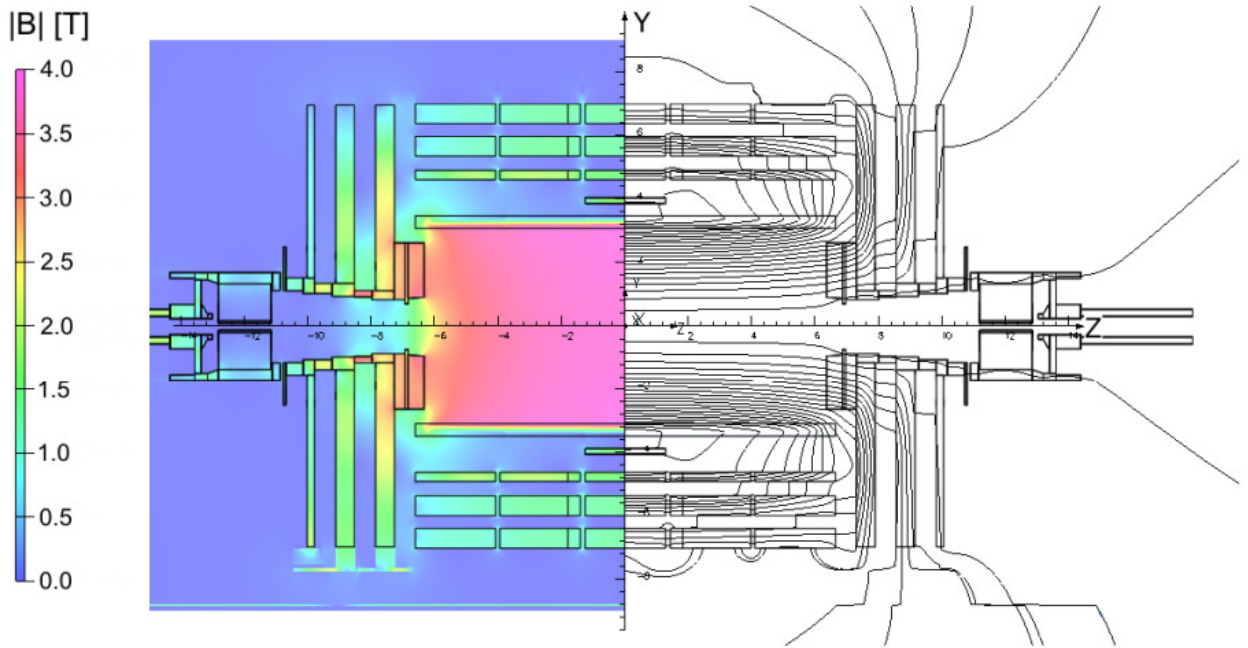


FIGURE 3.8: Simulation de la distribution du champ magnétique dans CMS [16].

3.2.6 Les chambres à muons

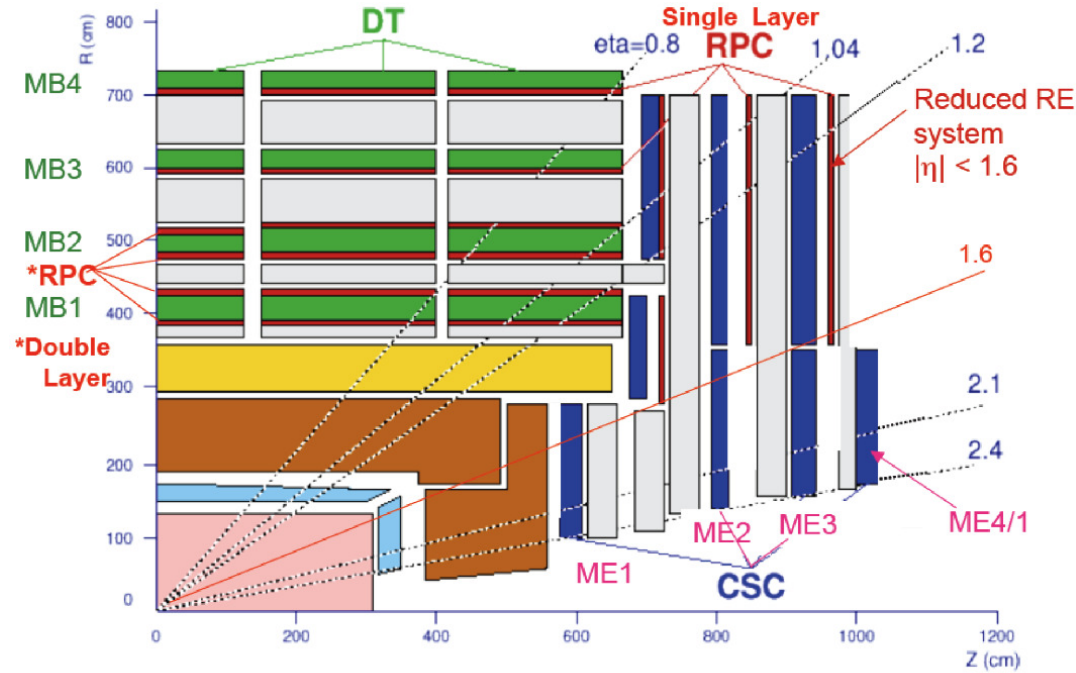


FIGURE 3.9: Coupe transverse d'1/4 du système de détecteurs à muons de CMS.

Les détecteurs à muons sont situés à l'extérieur du solénoïde entre les culasses de retour de champ. Le choix des technologies utilisées était contraint par la grande surface à couvrir et par le fait qu'à ce rayon les différentes zones de radiations sont non homogènes.

Aux débuts du projet CMS, obtenir une mesure précise et robuste des muons était prioritaire. Le détecteur a été imaginé pour qu'à la fois il puisse reconstruire la charge et le moment de muons de grande impulsion transverse (de l'ordre du TeV) et d'en informer suffisamment vite le système de déclenchement. La résolution en masse est prévue pour atteindre 1% à 100 GeV/ c^2 .

Tous les détecteurs utilisés sont des détecteur à gaz ⁶.

Le système de détection des muons (Muon System MS) utilise trois types de détecteurs à gaz : les chambres à dérive (Drift Tube Chamber, DT), les chambres cathodiques à pistes (Cathod Strip Chamber, CSC) et les chambres résistives (Resistive Plate Chamber, RPC).

Drift Tube Chambers, DT :

Les DT sont placées dans le tonneau à $|\eta| < 1.2$. Dans cette région et le champ magnétique résiduel est uniforme valant presque 2T. Les DT sont composées d'une enceinte

6. Les détecteurs à gaz se présentent sous forme d'enceinte renfermant un milieu gazeux facilement ionisable. L'enceinte elle-même ou une structure métallique additionnelle (cf. détecteurs à pistes) joue le rôle de cathode. Un fil métallique à haute conductivité est placé dans l'enceinte en guise d'anode. Le passage du muon ionise le gaz du détecteur et un courant électrique se crée entre l'anode et la cathode.

en aluminium refermant un mélange d'Ar-CO₂ (85%-15%) et équipées d'un fil d'acier inoxydable plaqué or. Une chambre mesure de 2 à 4 m de long pour une section de 40×13 mm² et possède une résolution spatiale de l'ordre de $100 \mu m$. Les signaux issus de l'ionisation des charges sont récoltés par les anodes en moins de 400 ns. Les DT sont réparties sur quatre couches et totalisent 250 modules. Deux dispositions de modules sont présentes : les modules dont les fils sont parallèles aux faisceaux et les modules dont les fils sont perpendiculaires aux faisceaux. Cette disposition permet d'obtenir les trois coordonnées (x,y,z).

Les chambres cathodiques à pistes, CSC :

Les CSC sont utilisées dans les bouchons où le champ magnétique est très intense et surtout non uniforme. Les CSC sont de forme trapézoïdale : chaque chambre est constituée de 6 couches de détection. Chaque couche présente 7 rangées de pistes cathodiques placées perpendiculairement à 6 fils d'anode espacés de 3.21 ou 3.16 mm. Les fils d'anode sont en tungstène plaqué d'or et ont un diamètre de $50 \mu m$. Les pistes cathodiques pointent vers l'axe des faisceaux. Le mélange gazeux des CSC est le Ar-CO₂-CF₄ (40-50-10%). Les CSC ont une réponse temporelle rapide, une bonne segmentation et une bonne résistance à la radiation. Leur résolution spatiale est de l'ordre de $200 \mu m$. Mais la résolution obtenue en combinant offline l'information du système complet de 6 couches de CSC atteint $50 \mu m$.

Sur chaque bouchon sont disposés 234 modules CSC, répartis en 4 stations.

Resistive Plate Chamber, RPC :

Les RPC sont utilisées conjointement aux DT et CSC dans le tonneau et les bouchons jusqu'à $|\eta| < 1.6$. Elles ont une résolution temporelle très performante, inférieure à 2 ns, mais une résolution spatiale moins performante de l'ordre de $200 \mu m$. Leur réponse temporelle rapide et leur fonctionnement en avalanche a motivé leur choix pour le système de déclenchement de premier niveau (L1). Les chambres sont constituées de deux plaques parallèles en bakélite de 2mm d'épaisseur dont la résistivité est grande (10^{10} à $10^{11} \Omega.cm$). Les plaques sont séparées de 2mm de mélange gazeux et recouvertes avec du graphite constituant l'anode.

Le Muon System (MS) est organisé de façon complexe, présentant plusieurs zones : cette configuration permet une grande couverture, mais la résolution n'est pas parfaitement homogène. La figure 3.9, présente la coupe longitudinale d'un quart du système de détection des muons. Un upgrade du MS est prévu durant la phase de long shutdown LS2.

3.2.7 Le Système de déclenchement (Trigger)

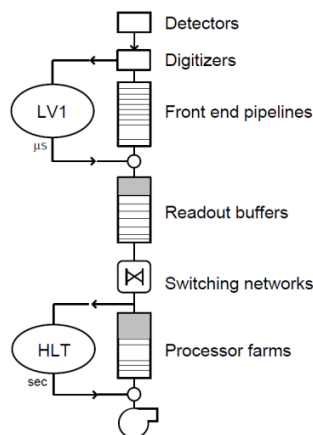


FIGURE 3.10: Le système de trigger de CMS.

Environ 25 collisions p-p inélastiques se produisent toutes les 50 ns dans le LHC. Le détecteur CMS produit en théorie 1 MByte de données par croisement de paquets. C'est un flux beaucoup trop élevé pour le système d'analyse online, et pour les capacités de stockage, qui sont limitées à des fréquences maximales de 300 Hz. La sélection des événements et la réduction du volume de données est effectuée par le système de déclenchement et le DAQ (Data Acquisition System). Pour CMS cette étape s'effectue en deux temps : un premier déclenchement au niveau électronique (Level-1 Trigger (L1)) et un second déclenchement au niveau informatique, le High-Level Trigger (HLT).

Les configurations des triggers sont établies selon les stratégies de recherche et sont basées sur l'identification des muons, électrons/photons, jets ou le missing energy channel.

Level-1 Trigger (L1)

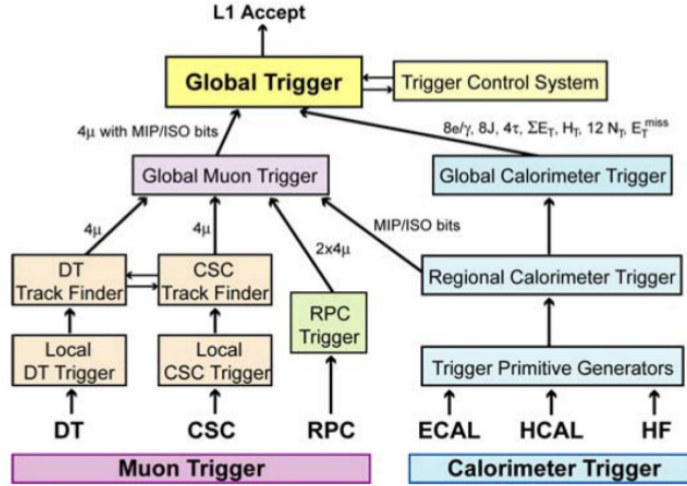


FIGURE 3.11: Schéma représentant les différentes composantes dans la décision de déclenchement au niveau L1.

Le trigger L1 réduit le taux d'événements de 40 MHz à 100kHz, en excluant la majorité des événements produisant uniquement des particules de petit p_T . La décision doit être effectuée toutes les 50ns sans temps mort. Pendant ce laps de temps, la totalité des données est stockée de façon temporaire dans une mémoire tampon de l'électronique front-end. La complexité et la rapidité de décision requise ne permettent pas de considérer les données du tracker. Mais une pré-identification des électrons et des muons va être menée. Le L1 possède trois composantes : local, régional et global (cf.figure 3.11). La composante locale récupère les dépôts d'énergies des calorimètres et les segments de traces des chambres à muons. Ces informations sont fournies à la composante régionale qui détermine et classe les objets physiques comme les électrons ou les muons. Les composantes globales du calorimètre et des chambres à muons sont ensuite combinées pour déterminer si l'événement est sélectionné. Dans le cas où l'événement est à conserver, un signal est envoyé dans le détecteur pour récupérer toute l'information lui étant associée. Le L1 trigger prend sa décision en $3,2 \mu s$ ce qui correspond au nombre maximum d'événements stockables dans les mémoires tampons des modules de détection. Les événements non sélectionnés au niveau L1 sont définitivement perdus. En cas de réponse positive du niveau L1, l'ensemble de l'information du détecteur est lue et soumise au déclenchement du niveau HLT.

HLT

Le HLT est un niveau de déclenchement informatique : les données sont pré-traités dans une ferme de calcul située juste à coté du détecteur. Le niveau HLT utilise une information plus détaillée et des algorithmes de reconstruction plus compliqués. Il réduit le flux de données à quelques 300-400 Hz.

Le HLT fonctionne événement par événement et peut prendre environ 50 ms d'analyse par événement. La prouesse consiste à effectuer, en si peu de temps, une reconstruction partielle des traces et des amas calorimétriques (nous décrirons plus précisément les procédures de reconstruction au chapitre 4). Pour cela le HLT procède par niveaux traitant avec des algorithmes de plus en plus poussés les différentes informations des sous-détecteurs :

- Level-2 est un premier niveau qui utilise uniquement l'information des chambres à muons et des ECAL et HCAL. A ce niveau on reconstruit une première fois les dépôts d'énergie calorimétriques en amas et super-amas pour lesquels on considère le rayonnement par brehmstrahlung.
- Level-2.5 est un niveau pour lequel on considère l'information additionnelle des impacts du tracker à pixel. L'information des calorimètres obtenue au Level-2 est extrapolée pour délimiter domaine d'exploration du tracker. Dans cette fenêtre des segments de traces sont reconstruits.
- Level-3 est un niveau pour lequel on considère toute l'information du tracker à piste et pixels. La reconstruction complète de la trace à partir des pré-traces obtenues au niveau 2.5 et de l'ensemble des données de l'événement est alors effectuée.

A chaque niveau, des événements sont écartés afin d'alléger l'analyse. Par ailleurs, la reconstruction des objets physiques HLT ne se fait qu'après association avec un candidat L1. Cette organisation en étages permet d'éviter d'encombrer le système et des analyses poussées peuvent être menées en peu de temps et moins de coût informatique.

En cas de réponse positive du HLT, l'événement est envoyé au Storage Manager qui l'écrit dans son entièreté sur un disque local. Les événements provenant des différents processeurs du DAQ sont ensuite ordonnés temporellement et enregistrés sur au niveau Tier0 de la grille de calcul du CERN⁷.

Les événements enregistrés sur le Tier0 sont en format brut ("Raw" en anglais) sur des fichiers FEVT(Full EVenT) et chaque événement occupe typiquement 1 à 3 MB.

3.2.8 CMS Traitement des données

Jusqu'à présent, nous avons décrit les étapes conduisant à l'enregistrement de données à la sortie du HLT. Ces données en format brut sont constituées de digits représentant une série d'impacts dans les sous-détecteurs : elles n'ont pas encore de lien clair avec de quelconques particules.

Pour pouvoir étudier la physique des particules élémentaires et de leurs interactions, on doit procéder à deux étapes supplémentaires :

- Associer les digits des différents sous détecteurs pour déterminer des grandeurs physiques : cette étape s'appelle la reconstruction et on parle d'objet reconstruit au "niveau détecteur" ("detector level").
- S'affranchir de la réponse des détecteurs pour parvenir au niveau des particules et pouvoir étudier la physique sous-jacente. On atteint alors le "niveau particule" ("particle level"). Cette étape est effectuée à l'aide d'une méthode dite d'Unfolding

7. La grille de calcul, ou Worldwide LHC Computing Grig regroupe 100000 processeurs et doit faire face à environ 15 petabytes de données par an. Le Tier0 est un des niveaux de la grille basé dans le centre de calcul du CERN. Les autres niveaux Tier1, 2 et 3 sont délocalisés dans des Laboratoires et universités dans le monde.

qui nécessite une maîtrise de la chaîne de simulation. Nous décrivons cette étape au chapitre 7.

La figure 3.13 résume les différentes étapes du traitement de données à CMS, de la sortie du HLT jusqu'à l'établissement de la section efficace du processus désiré, ici $pp \rightarrow Z(\rightarrow \mu^+\mu^-) + jets$.

A partir des fichiers de données brutes (RAW) enregistrés dans les fichiers FEVT, l'on procède à une première reconstruction. Les données sont alors réenregistrées sur les fichiers RECO (RECOⁿstructed data) pour lequel un événement reconstruit occupe typiquement 500 kB et des fichiers plus légers AOD (Analysis Object Data)⁸ où un événement occupe environ 50 kB.

A partir des fichiers AOD et RECO, on opère un nouveau traitement des événements qui consiste à classer les objets reconstruits. Il existe plusieurs sortes de classements comme par exemple le ParticleFlow (PF). Toutes ces opérations sont centralisées par CMSSW, l'interface informatique officielle de CMS : elle est extrêmement complète et permet une gestion multifonctionnelle des données, ainsi qu'une gestion totale des simulations.

Pour mon travail j'ai utilisé des données et fichiers de simulation pré-traités et délivrés sous forme d'un root TTree. Le TTree est un fichier construit pour classer sous forme de tableau un grand nombre d'objets de même nature. Il est optimisé pour réduire la place occupée sur les disques et présente des facilités d'interface avec les logiciels graphiques de root.

Les étapes de reconstruction du Particle Flow et les procédures de calibration seront décrites au chapitre 4.

Les étapes d'analyse que j'ai effectué seront expliquées dans les chapitres 5 et 6 et 7.

8. Les fichiers AOD contiennent toute l'information nécessaire à la reconstruction mais moins d'informations techniques concernant les caractéristiques opérationnelles du détecteur.

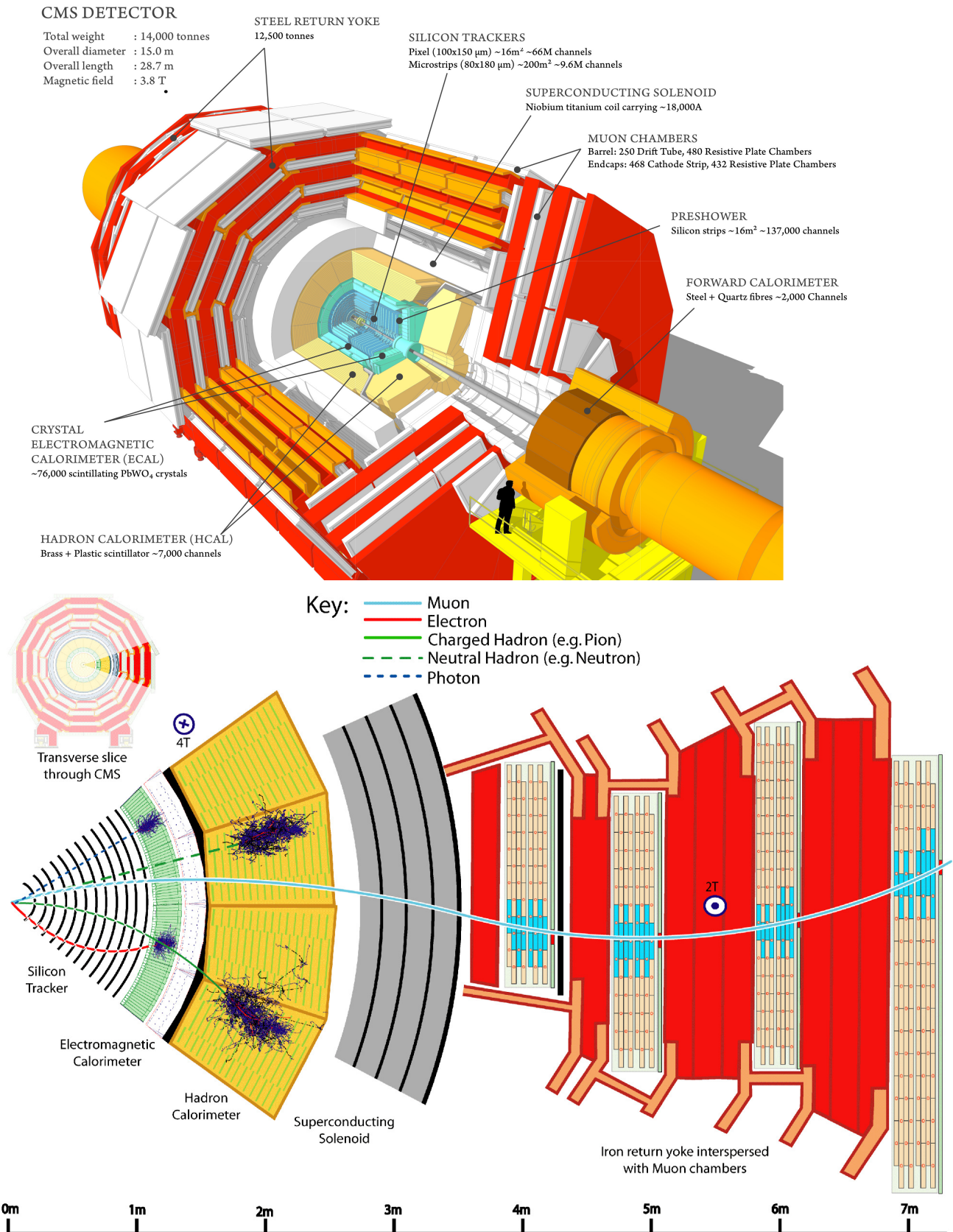


FIGURE 3.12: Le détecteur CMS : description des différents sous-détecteurs (haut), passage d'un muons dans le détecteur (bas)[15].

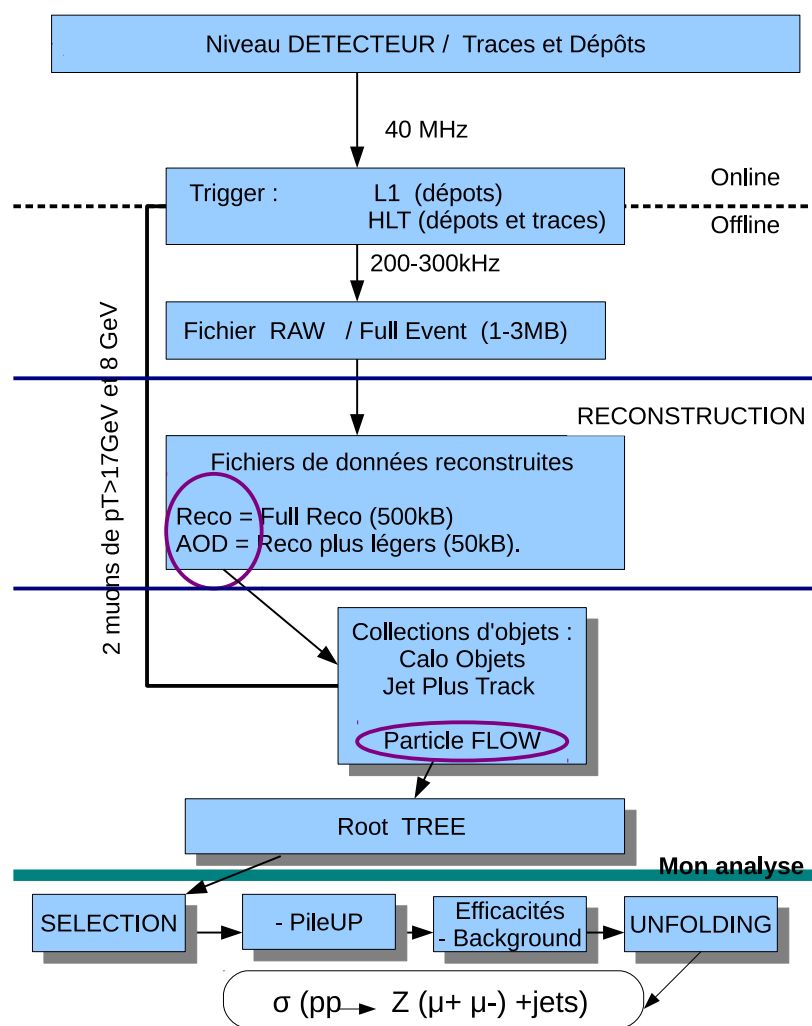


FIGURE 3.13: Shéma représentant la chaîne de traitement des données à CMS appliquée à l'étude de la production d'un boson Z accompagné de jets. Mon travail a consisté en l'élaboration des étapes représentées sous le niveau "Mon Analyse". La ligne reliant les niveaux "collections d'objet" et "Trigger" signifie que l'information "Trigger" est conservée dans les données analysées.

Chapitre 4

Chaîne de simulation et reconstruction

Nous venons de décrire le fonctionnement du détecteur et la chaîne de traitement de données de CMS. Parallèlement au traitement des données, il est indispensable d'établir des fichiers de simulation robustes nécessaires à l'extraction de la section efficace voulue dans ce travail.

Concernant les données, les grandeurs physiques vont être reconstruites à partir des informations du détecteur sous forme digitalisée. Elles peuvent ensuite être analysées. Concernant les fichiers de simulation l'on procède dans l'ordre suivant : les processus qui nous intéressent sont générés au niveau des particules, la réponse des détecteurs est ensuite simulée, puis les grandeurs physiques sont reconstruites de la même façon que pour les données réelles. Les fichiers MC sont formatés dans le but d'être traités de façon similaire aux fichiers de données dans les procédures de reconstruction et d'analyse. On parle de "Full Simulation Chain". C'est l'interface CMSSW qui gère fichiers de données et fichiers de simulation et assure la cohérence des formats.

Pour notre analyse nous avons utilisé des TTree de données et de données simulées de façon officielle par la collaboration et déjà reconstruites. Nous n'avons pas étudié la spécificité des logiciels de simulation utilisés ni effectué les étapes de reconstruction. Nous allons toutefois les passer en revue et décrire les enjeux et la physique liés à chaque étape.

Dans une première partie nous présenterons de manière générale la chaîne de simulation de CMS puis les fichiers de simulation MC que nous avons utilisés. Dans une deuxième partie nous décrirons les étapes techniques de reconstruction au sens du Particle Flow valables à la fois pour les données réelles et simulées.

4.1 Chaîne de Simulation

La maîtrise de la simulation est un élément nécessaire à chaque étape d'analyse de données.

Au LHC, avant même la mise en route du collisionneur, la chaîne de simulation de chaque expérience était développée et déjà testée sur base de données de faisceaux test ou de rayon cosmiques. A chaque étape de fonctionnement du collisionneur et selon les besoins des analyses, les outils de simulation sont sans cesse testés, réévalués et améliorés. Dans le domaine des "générateurs", par exemple, de nombreux logiciels ont été développés

au fil du temps et des expériences de collisions de particules de haute énergie, chacun ayant ses limites et ses spécificités (Pythia, HERWIG, Sherpa, MadGraph, etc..) Les analyses permettent aussi de comparer leur performances.

La simulation d'une collision dans CMS nécessite 3 phases : la génération de données au "niveau particule", la simulation des interactions des particules générées avec la matière que constitue le détecteur, et enfin la simulation du signal électronique de sortie des détecteurs ou digitalisation.

Ces trois étapes nécessitent l'intervention de plusieurs logiciels et l'interface CMSSW se charge de faire le lien entre le choix d'analyse et les différentes chaînes de logiciels et propose des fichiers MC de format similaire à celui obtenu pour les données.

Dans cette partie, nous allons aborder de manière générale la génération d'événements et passer en revue les implications spécifiques des programmes MadGraph et Pythia utilisés pour générer les données MC utilisées. Nous parlerons rapidement de la réponse du détecteur simulée avec Geant4.

4.1.1 Génération d'événement

A l'énergie délivrée par le LHC la multiplicité des états finaux est colossale. Une description analytique de l'entièreté des processus est donc totalement impossible.

La stratégie adoptée par les logiciels de génération¹ est de simuler les états finaux en factorisant en temps les différents processus les constituant : on génère un processus principal ou "processus dur" issu de l'interaction entre des partons constituant les protons incidents, puis la radiation de parton ou "parton shower", l'activité sous-jacente due aux résidus de faisceaux et aux interactions à partons multiples, l'hadronisation et enfin les désintégrations spontanées.

Cette factorisation temporelle est rendue possible dans le cas de la production d'un processus dur, car les échelles de temps des différents sous-processus sont très différentes.

La somme de tous ces processus s'approche ainsi de la complexité de la réalité comme illustré à la figure 4.1 (Nous parlerons des problèmes de pileup au chapitre 5). Nous allons rapidement passer en revue ces différentes composantes et la façon de les générer.

4.1.1.1 Processus dur - hard process

L'interaction entre deux partons emportant une fraction d'énergie de chacun des protons incidents peut constituer un processus comportant des particules d'impulsion transverse importante par rapport à la majorité des processus produits dans la direction du faisceau : c'est ce qu'on appelle le processus dur. Cette présence d'une échelle dure permet d'appliquer une approche perturbative de QCD et donc de pouvoir considérer l'interaction au niveau partonique.

Pour notre analyse nous avons besoin de générer un processus $q\bar{q} \rightarrow Z$ pour lequel l'échelle dure est assurée par la masse du boson Z , M_Z . La désintégration du Z donnera deux leptons de grande impulsion transverse comme représenté sur la figure 4.1. Notre

1. Lire à ce sujet le mode d'emploi de Pythia6.4 [17] ou voir [18]

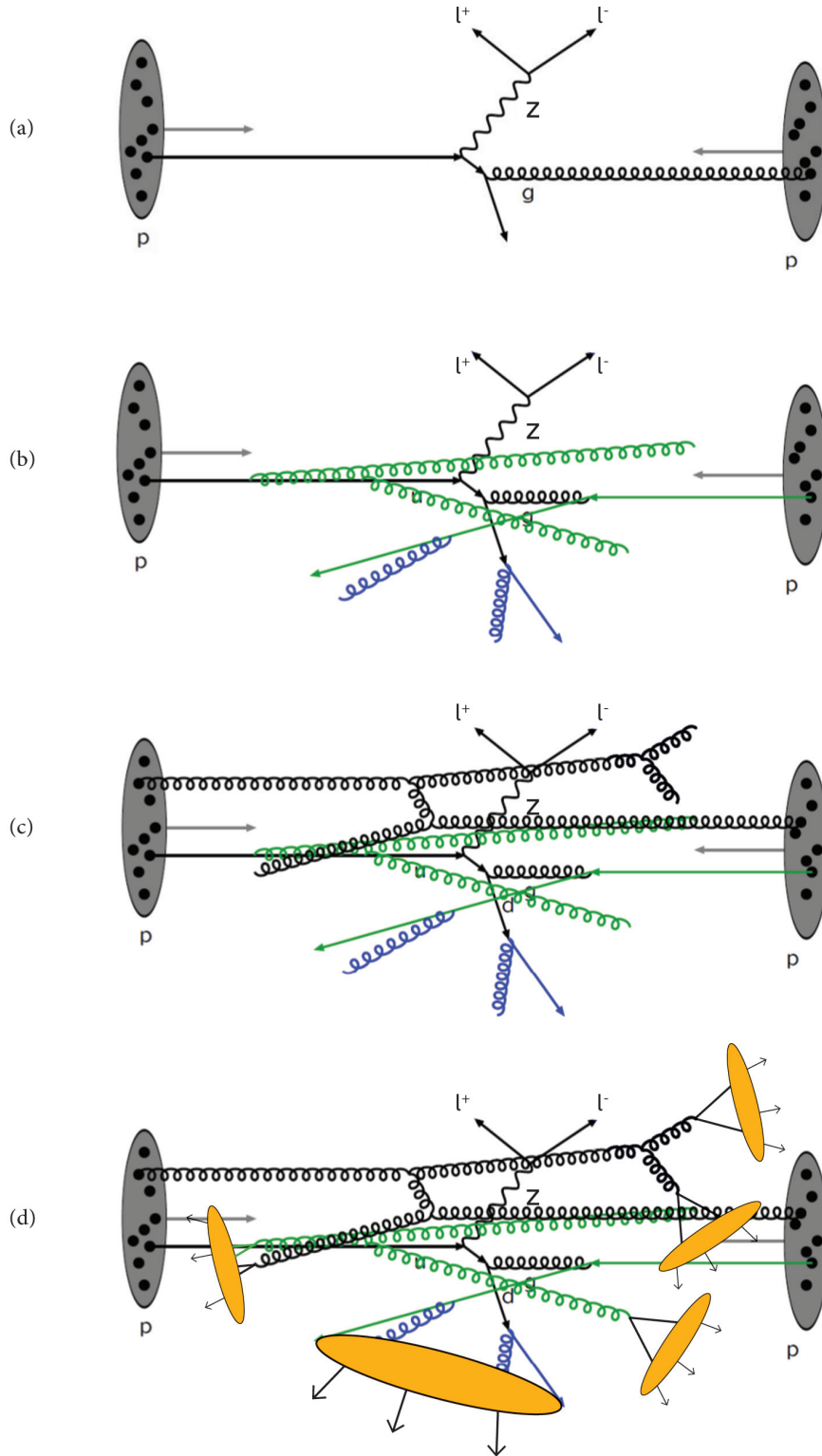


FIGURE 4.1: Vue d'artiste représentant les différents sous-processus générés avec, de haut en bas, (a) le processus dur avec les deux leptons issus de la désintégration du Z, (b) la parton shower, (c) l'underlying event, (d) l'hadronisation.

processus dur est donc un processus Drell-Yan comprenant un Z^2 et dont la section efficace est décrite au chapitre 2.

Considérons pour l'instant ce processus à l'ordre le plus bas sans jet associé. La section efficace du processus $q\bar{q} \rightarrow Z$ peut être décrite dans le modèle des partons :

$$\sigma_{(p\bar{p} \rightarrow Z)} = \sum_{q\bar{q}} \int dx_1 dx_2 f_{q,1}(x_1) f_{\bar{q},2}(x_2) \hat{\sigma}_k(q\bar{q} \rightarrow Z) + q \leftrightarrow \bar{q} \quad (4.1)$$

Où $\hat{\sigma}_k$ est la section efficace du processus dur que l'on peut calculer analytiquement, et les fonctions $f_{i,1}(x_1)$ représentent les PDF. Pour rappel, elles représentent la probabilité que la collision ait engagé un parton i emportant la fraction d'impulsion x_i du proton initial. Elles sont bien déterminées depuis les années 2000 et profitent d'une amélioration continue alimentée par les résultats expérimentaux. Les PDF sont déterminées à une échelle d'énergie basse et extrapolées aux échelles plus élevées par les équations d'évolution de la QCD. Même si les PDF ont gagné en précision, leur simulation génère des erreurs systématiques significatives (cf. section 7.2). Dans notre analyse, la PDF utilisée est la **CTEQ6L** [5].

4.1.1.2 Radiation de parton - Parton shower

Les partons impliqués dans le processus dur sont fortement accélérés (par rapport au référentiel du labo) et peuvent émettre des radiations QCD sous forme de gluons. Les gluons vont ensuite eux-mêmes radier des paires $q\bar{q}$ qui à leur tour sont susceptibles de radier des gluons.

La radiation d'un parton peut être "dure" (si son impulsion transverse par rapport au parton initial est conséquente), ou molle (si l'énergie du gluon radié est faible) ou encore colinéaire (si l'angle de radiation est petit).

La radiation de gluon dite "dure" constitue la correction QCD au NLO pour notre processus de départ : il s'agit de la production de $Z + 1\text{jet}$ dans l'état final partonique. Nous aborderons la génération de radiations dures dans le paragraphe 4.1.1.3 consacré à la génération des jets associés au Z . Ici nous décrivons la génération de radiations colinéaires ou molles appelée parton shower.

Comment est simulée la parton shower ? La parton shower est simulée en ajoutant un gluon émis à chaque parton participant au processus dur. Cet ajout doit tenir compte des angles de radiation (en général on définit un angle θ relatif à la direction du parton initial), de la fraction d'impulsion du parton initial emportée par la radiation (z) ainsi que de la nature du parton initial (qui va déterminer le facteur de splitting à prendre en compte.)

Une radiation colinéaire $i \rightarrow j+k$ peut être représentée par la formule 4.2. On définit la section efficace différentielle avant la radiation $d\sigma_n$, la radiation après radiation sera alors donnée par $d\sigma_{n+1}$:

$$d\sigma_{n+1} \simeq d\sigma_n \frac{\alpha_s}{2\pi} \frac{dq^2}{q^2} dz P_{ji}(z), \quad \text{avec} \frac{dq^2}{q^2} \sim \frac{d\theta^2}{\theta^2} \quad (4.2)$$

2. la contribution de γ et de l'interférence γ/Z sont également considérées mais ne seront plus mentionnées pour ne pas alourdir les explications.

Où α_s est la constante de couplage de l'interaction forte, P_{ji} est la fonction de splitting qui décrit la dépendance en la fraction d'énergie z que la particule j a emporté de l'énergie totale de i . Les facteurs de splitting correspondent à ceux définis pour les équation DGLAP au chapitre 2 section 2.2.1

Facteur de forme de Sudakov : toutes les radiations colinéaires ne sont pas possibles : leur virtualité sera comprise entre l'échelle d'énergie du processus dur (Q^2) et l'échelle non perturbative de l'hadronisation (Q_0^2).

Pour générer des radiations molles qui aient une cohérence physique, les générateurs utilisent le facteur de forme de Sudakov décrit par la formule 4.3.

$$\Delta_i(q_1^2, q_2^2) = \exp - \left\{ \int_{q_1^2}^{q_2^2} \frac{dq^2}{q^2} \frac{\alpha_s}{2\pi} \int_{\frac{Q_0^2}{q^2}}^{1-\frac{Q_0^2}{q^2}} dz P_{ji}(z) \right\} \quad (4.3)$$

Le facteur de Sudakov représente la probabilité qu'il n'y ait pas de radiation pour un parton ayant une valeur de q^2 comprise entre q_1^2 et q_2^2 .

Pour simuler la parton shower, on utilise donc la distribution $1 - \Delta_i(q_1^2, q_2^2)$ en générant aléatoirement des valeurs de z et de q^2 dans un intervalle cohérent avec les échelles d'énergies Q^2 et Q_0^2 .

Final et Initial State Radiation : La parton shower est appliquée à tous les partons des états initiaux (on parle alors d'Initial State radiation, ISR), mais également sur les partons des états finaux (c'est alors la Final State Radiation, FSR). La modélisation de la radiation dans l'état final découle directement des explications précédentes. On itère le splitting jusqu'à obtenir des échelles d'énergie de l'ordre de Q_0^2 . A chaque étape on procède également à une vérification de conservation des impulsions.

Dans l'état initial les partons ont très peu de virtualité et sont à des échelles d'énergie en deçà du processus dur. La radiation des partons de l'état initial se termine quand les partons collisionnent pour donner le processus dur. Donc, dans le cas des ISR, la difficulté est de modéliser une parton shower avec la contrainte précise de trouver deux partons pouvant créer le processus dur. La solution est de procéder avec la même démarche que pour les FSR mais à rebours. On détermine d'abord les fractions de l'impulsion originale x_1 et x_2 portés par les deux partons créant le processus dur. Ces variables sont liés par la relation : $x_1 x_2 \hat{s} = s$ où $\sqrt{s}$ est l'énergie de collision et $\sqrt{\hat{s}}$ l'énergie du processus dur. A partir de là on procède pour chaque parton à une évolution entre son énergie x_i jusqu'à l'échelle d'énergie initiale $\sqrt{s}$. A chaque étape le parton gagne de l'énergie au lieu d'en perdre comme pour la FSR. Le facteur utilisé est alors le facteur de Sudakov pondéré par les PDF :

$$\Delta' i(q_1^2, q_2^2) = \Delta i(q_1^2, q_2^2) \frac{f_i(x, q_2^2)}{f_i(x, q_1^2)} \quad (4.4)$$

4.1.1.3 Génération de jets associés au Z

La section efficace Drell-Yan est actuellement résolue analytiquement au NNLO mais est implémentée seulement au LO ou au maximum au NLO dans les logiciels de génération utilisés actuellement. Générer des processus aux ordres supérieurs de QCD, c'est à dire

avec 3 ou 4 jets de partons associés requiert donc une stratégie différente. C'est ce que propose les logiciels que nous utilisons : MadGraph/MadEvent combinés avec Pythia.

MadGraph : MadGraph [19] est un générateur matriciel au départ établi pour calculer les diagrammes de Feynman pour n'importe quel processus qui peut être décrit par un Lagrangien. Il utilise des bases de données (HELAS ou FeynRules) et code sous forme matricielle tous les processus durs connus et observés à partir des règles de Feynman et pondérés par les rapports de branchement. L'expression de la section efficace est ensuite établie sous forme totalement différentielle et sera intégrée dans l'espace des phases par la partie MadEvent.

Dans le cas de $Z + jets$, MadGraph/MadEvent va calculer les éléments de matrice de tous les processus comprenant un nombre de partons donné dans l'état final et générer l'échantillon d'événements correspondant. Pour des questions de temps de calcul, les fichiers utilisés dans ce travail considèrent les processus allant jusqu'à quatre partons dans l'état final. Le tableau 4.1 décrit les processus possibles jusqu'à $Z + 2partons$ et rend compte de la complexité rencontrée aux ordres supérieurs.

Chaque section efficace ayant un nombre de partons donné est donc calculée au LO. L'interface avec le logiciel Pythia permet d'ajouter de la parton shower à chaque processus généré au LO. Cette parton shower peut mener à la création de radiations dures qui vont simuler les ordres supérieurs. Les algorithmes évitent les double comptages entre jets générés par MadGraph et jets issus de la parton shower générés par Pythia.

Cette méthode donne de bons résultats et nous verrons que la combinaison de MadGraph avec Pythia modélise très bien les processus $Z + 4jets$, et ce avec une résolution de l'ordre du pour-cent (cf. chapitre 7).

La production de jets associé à un processus DY fut pendant longtemps uniquement modélisée par une "parton shower" ajoutée au processus considéré au LO ou NLO. C'est par exemple ainsi que procède le générateur Pythia.

Une étude similaire à la nôtre [20], utilisant les données CMS à 7 TeV a comparé MadGraph et Pythia et montre que pour la production de $Z + 1jet$ Pythia diffère de 10 % des données (40 % pour $Z + 2jet$).

Dans notre travail, à partir de $pp \rightarrow Z + 5jets$ le cinquième jet est uniquement généré par la parton shower de Pythia. Nous verrons qu'alors la simulation MC s'éloigne des données.

Pythia prend également en charge les étapes suivante de génération de l'"Underlying Event" et de l'hadronisation.

4.1.1.4 Underlying events

Il existe de nombreux processus sous-jacents qui ne peuvent pas être décrits en ajoutant de la parton shower aux partons participant au processus dur. Cette activité sous-jacente est plus importante que le bruit de fond électronique du détecteur et il est nécessaire de bien la modéliser afin d'identifier correctement du bruit de fond (ici au sens général) contaminant notre processus principal.

On désigne ces événements sous-jacents par l'appellation "Underlying Events" (UE).

De façon générale, l'Underlying Event est caractérisé par des processus faisant intervenir des partons ne participant pas directement au processus dur. Soit ces partons

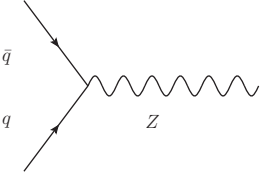
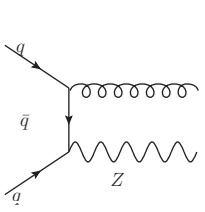
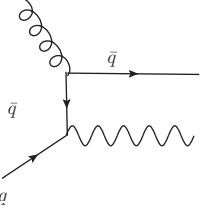
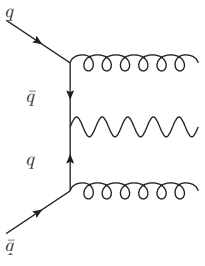
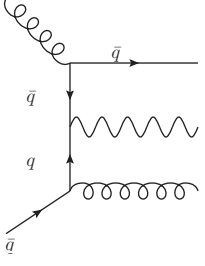
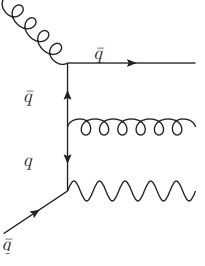
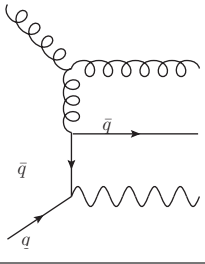
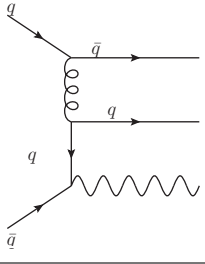
Z + 0 parton		+ $q \Leftrightarrow \bar{q}$
Z + 1 parton	 	+ $q \Leftrightarrow \bar{q}$
Z + 2 parton	    	+ $q \Leftrightarrow \bar{q}$...

TABLE 4.1: Exemples de diagrammes représentant les différentes possibilités d'obtenir un parton supplémentaire dans l'état final.

n'interagissent pas entre eux et constituent alors des résidus de faisceaux (ou "beam remnants"), soit ils interagissent entre eux et cela crée des diffusions à partons multiples (multiple parton interaction ou MPI).

Multiple partons interactions (MPI) : les MPI sont donc des interactions supplémentaires entre les partons résiduels des protons de la collision. Ce sont donc des diffusions gluon-quark, quark-quark ou gluon-gluon. La modélisation des MPI est une phase compliquée de la génération d'événement car le processus n'est pas encore bien compris ni bien mesuré : on sait surtout que dans une collision p-p les MPI sont très probables, et ce d'autant plus que leur impulsion transverse est faible. Pythia propose plusieurs scénarios de génération des MPI. Les scénarios plus sophistiqués tiennent compte d'une certaine corrélation entre le paramètre d'impact de l'interaction et les MPI.

Une fois les événements de l'UE générés au niveau partonique, on peut leur rajouter de la parton shower et on procède à l'hadronisation.

4.1.1.5 Hadronisation

En dernier lieu surviennent les phénomènes d'hadronisation à une échelle de temps à laquelle la QCD devient non perturbative : la simulation doit alors considérer les modèles de fragmentation se basant sur les comportements phénoménologiques généraux de la QCD. On distingue différents modèles de fragmentation : la fragmentation par cordes (String fragmentation), la fragmentation Indépendante, et la fragmentation d'amas (Cluster Fragmentation). Des versions hybrides ont également été implémentées. C'est le cas pour le générateur que nous utilisons, Pythia implémenté pour un modèle de fragmentation à cordes dit de Lund.

Tous ces modèles actuels sont de nature probabiliste et procèdent de façon itérative. Évidemment ces modèles sont incomplets et aucun ne rend totalement compte de la complexité de la réalité, mais chacun dans son domaine représente assez bien les données et est utile pour prédire certaines propriétés qui pourraient se produire à des énergies plus élevées.

Les modèles de fragmentation par cordes considèrent le champ de couleur entre deux quarks comme une corde énergétique. Ce champ de couleur est supposé avoir une longueur de l'ordre du fm, c'est à dire l'échelle d'hadronisation.

L'idée de base est la suivante : si deux quarks $q\bar{q}$ s'éloignent l'un de l'autre à grande vitesse (presque la vitesse de la lumière), à une certaine distance l'énergie emmagasinée dans le champ de couleur permet la création d'une nouvelle paire $q'\bar{q}'$. Deux nouveaux singulets de couleur se forment : $q\bar{q}'$ et $q'\bar{q}$. Si leur masse invariante correspond à celle d'un hadron sur sa couche de masse, l'itération s'arrête et le hadron est généré. Sinon les singulets s'étirent de manière à créer une nouvelle paire $q\bar{q}$ et ce jusqu'à l'obtention d'une série de hadrons. Le modèle est assez sophistiqué pour inclure également la production de baryons.

4.1.1.6 Désintégrations spontanées

Le dernier stade de la génération d'événements est la simulation des désintégrations spontanées des hadrons générés à l'étape précédente. Ces désintégrations constituent une bonne partie des états finaux enregistrés. Leur simulation nécessite une implémentation à plusieurs niveaux et utilise les informations présentes dans les bases de données des désintégrations (cf. "Review of particle Physics" [2]). En général on implémente le phénomène dans le centre de masse de la particule qui se désintègre puis on la booste. Les distributions en impulsions et les corrélations possibles de désintégrations à multiparticules, ainsi que les corrélations de spin doivent être prises en compte. La somme des différents rapports de branchements des canaux de désintégration doit être l'unité : ceci nécessite parfois la simulation de canaux prédits mais pas encore observés (pour la physique des quarks lourds par exemple...).

4.1.2 Modélisation du détecteur et digitalisation

Cette étape consiste à reproduire la réponse des détecteurs au passage des particules générées : on simule d'abord les impacts et dépôts liés au passage des particules dans la matière des détecteurs, puis la réponse de l'électronique des détecteurs (digitalisation).

La simulation du détecteur CMS est basée sur le package de simulation Geant4 (pour GEometry ANd Tracking)[21]. Geant est une plate-forme de logiciels développée de façon générale pour la simulation du passage de particules à travers la matière par la méthode de Monte-Carlo.

Geant4 permet de décrire de façon détaillée la hiérarchie des détecteurs et sous-détecteurs selon leur volume, le matériel qui les constitue, leur positionnement, ainsi que la résolution des capteurs, etc... Une distinction est faite entre parties sensibles (i.e les sous-détecteurs eux même comme décrits dans le chapitre 3) des parties dites "mortes" (les câbles et les circuits de refroidissement par exemple). Geant4 prend en input les particules générées, retrace leur passage à travers les différentes couches de détection ou "mortes" et modélise la physique d'interaction des particules avec la matière. Le résultat est la modélisation des dépôts calorimétriques et impacts dans le tracker et les chambres à muons.

Après la modélisation, la digitalisation des données s'effectue par l'intermédiaire du logiciel officiel de la collaboration CMSSW [22]. La digitalisation consiste à associer dépôts et impacts à des signaux électroniques : c'est un peu l'opération inverse des opérations de reconstruction. Elle simule la réponse de l'électronique frontale des détecteurs. Une procédure de traitement des signaux est appliquée : elle commence par une pré-amplification, puis par une mise en forme et enfin la conversion des signaux analogiques en signaux numériques.

4.1.3 Fichiers MC utilisés et calibration par rapport aux données

Pour notre analyse, nous avons utilisé la version CMSSW_5_3_7, utilisant un système combiné de générateurs : MadGraph 6 et Pythia 6.4 avec le tune Z2.

Les fichiers MC utilisés pour notre analyse sont décrits dans le tableau 4.2. Le fichier principal est donc le processus Drell-Yan, $pp \rightarrow Z(\rightarrow l^+l^-) + jets$ généré par MadGraph et Pythia comme nous venons de le décrire. Ce fichier comprend uniquement les canaux de désintégration du Z en électrons et en muons. La désintégration en leptons τ constitue un fichier à part qui ne sera pas utilisé dans ce travail. Les autres fichiers constituent les bruits de fond présentant deux muons dans l'état final et sont générés par MadGraph et TAUOLA[23].

Les processus participant au bruit de fond seront décrits de façon plus précise au chapitre 6 section 6.2.

Afin de pouvoir comparer les simulations aux données, il est nécessaire d'établir la fraction d'événements du fichier MC qui a contribué au processus final. Pour cela on normalise les fichiers à la luminosité des données, c'est à dire 19.3 fb^{-1} . Les logiciels de génération fournissent la section efficace associée à chaque fichier ainsi que le nombre d'événements compris dans l'échantillon. De ces valeurs on déduit facilement la luminosité L_{MC} pour chaque fichier en utilisant la formule $L_{MC} = N/\sigma$. Les valeurs utilisées sont présentées dans le tableau 4.2.

Fichiers MC	Nombre d'événements	Section efficace (pb)	L_{MC} (pb ⁻¹)
DYJetsToLL	30459503	3503.7	8.64
TTJets_MassiveBinDECAY	6923750	234	29.60
ZZJetsTo2L2Q	1936727	8.258x0.14	1675.20
ZZJetsTo2L2Nu	954911	8.258x0.04	2890.9
ZZJetsTo4L	4807893	8.258x0.01	58221.03
WZJetsTo2L2Q	3215990	33.21x0.07	1383.40
WZJetsTo3LNU	2017979	33.21x0.03	2025.47
WWJetsTo4L	1933235	54.838x0.11	320.50

TABLE 4.2: Fichiers MC utilisés pour notre analyse, avec le nombre d'événements dans l'échantillon, la section efficace et la luminosité correspondantes.

4.2 Reconstruction des événements

La reconstruction est l'étape qui consiste à associer dépôts d'énergie et impacts pour former des objets physiques complexes comme par exemple des traces identifiées de muons, des gerbes électromagnétiques ou des jets de hadrons.

Les algorithmes de reconstruction combinent l'information d'un ou plusieurs sous-détecteurs afin d'identifier les objets physiques et d'évaluer leur cinématique. Les objets ainsi reconstruits sont enregistrés par collections.

Les algorithmes gèrent différemment les informations : certains considèrent uniquement les dépôts calorimétriques, d'autres leur associent les impacts dans le tracker. Par exemple, pour les jets, on distingue plusieurs collections : les Calojets³, les JetPlusTrack (JPT), et les jets du Particle Flow (PF, "flux de particules").

Dans notre analyse nous utiliserons uniquement les objets du Particle Flow.

Le Particle Flow (PF) utilise une combinaison méthodique de toute l'information des sous-détecteurs de CMS pour établir des collections de particules stables, i-e de photons, électrons, muons, hadrons chargés, hadrons neutres. Ces collections de particules sont ensuite utilisées pour reconstruire des jets (pour lesquels on pourra ensuite estimer les directions et énergie du parton initial), pour déterminer l'énergie transverse manquante (E_T^{miss}), pour reconstruire et identifier les leptons taus à partir des produits de leur désintégration, et mener d'autres analyses de haut niveau [24].

La philosophie du PF est de proposer pour chaque événement un fichier facile à utiliser présentant pour chaque objet une estimation de son énergie et de ses variables cinématiques (p_T , η , φ) et différentes variables combinées les caractérisant (variables d'identification, d'isolation, distance de la trace au vertex primaire...). Ce fichier présente les données classées par objet et par événement sous forme de "PATuples"⁴. Les PATuples présentent également des variables caractérisant l'événement lui-même (nombre de vertex, nombre de muons...). Le PF reconstruit les objets en tenant compte des instructions et

3. Les algorithmes de Calojets ne considèrent pas l'info du tracker, ni celle des chambres à muons. Les Calojets sont reconstruits à partir des "tours" calorimétriques rassemblant les dépôts des ECAL et HCAL : 1 "tour" correspond à 1 cellule de HCAL et 5x5 cristaux du ECAL.

4. PATuples pour "Physique Analysis Tuples" : tableau de variables structurées reposant sur le logiciel Root du CERN.

exigences des groupes d'analyses spécialisés dans chaque objet qu'on appelle les POG (Physic Object Group)[25].

Le PF est un algorithme qui reconstruit avec une grande précision et un taux de fausse identification très bas même pour des événements à haute densité. Cette performance est obtenue grâce au développement d'algorithmes avancés de reconstruction de trace et d'agrégation ("clustering") des dépôts calorimétriques. Les traces reconstruite du tracker, les amas de ECAL et HCAL et les traces de muons vont constituer les éléments de base de l'algorithme du PF. Une savante stratégie d'association de ces éléments permet la reconstruction finale des événements.

Nous allons décrire plus précisément les algorithmes de reconstruction du PF et nous pencher sur le cas des muons et des électrons. Nous parlerons ensuite de la reconstruction des jets et de leur calibration.

4.2.1 Stratégie de reconstruction du PF

L'algorithme du PF est séquentiel : pour chaque événement il reconstruit étape par étape les particules stables en considérant les éléments de base suivant [26] :

- Les traces de particules chargées reconstruites dans le tracker,
- Les amas d'énergie déposés dans les calorimètres,
- La pré-identification des muons (selon le muon POG),
- La pré-identification des électrons (selon le E/Gamma POG),

A chaque étape le soucis est de garder une bonne efficacité de reconstruction, mais un faible taux de fausse identification. Pour cela la stratégie générale adoptée est d'appliquer un premier algorithmes avec des critères très sévères (efficacité moyenne mais taux de fausse identification négligeable) et de retirer de l'analyse les impacts correspondant aux objets identifiés lors de cette première étape. On reconduit alors l'analyse sur les objets restant avec des critères plus souples (efficacité meilleure, taux de fausse identification toujours bas, car les combinaisons possibles sont alors réduites).

4.2.1.1 Reconstruction itérative de traces

Le PF utilise la résolution du tracker au maximum en reconstruisant d'abord les objets uniquement avec les impacts du tracker. Les objets ainsi reconstruits ont une bonne résolution mais surtout cela permet de construire d'emblée les informations de distance des traces au vertex primaire.

Pour cette étape, il faut utiliser un algorithme avec une efficacité de reconstruction la plus proche possible des 100%. Pour toutes les particules chargées, la reconstruction de traces est effectuée par des algorithmes séquentiels associant les groupes d'impact aux équations du mouvement d'une particule chargée dans un champ magnétique.

Un exemple d'algorithme est le filtre de Kalman utilisé par exemple dans les radars. C'est un estimateur récursif qui permet de suivre la trajectoire d'un objet en prédisant et corrigeant les impacts appartenant à une même trace. Le PF utilise ce genre d'algorithme pour le tracker mais aussi pour la reconstruction de trace des muons dans les chambres à muons.

Ces algorithmes fonctionnent en 4 étapes :

- La graine : c'est le minimum d'information nécessaire à la détermination d'une trace. En général, c'est un segment de trace constitué par 2 ou 3 impacts.
- Extrapolation de la trajectoire : en utilisant la direction donnée par le segment graine, l'algorithme identifie dans les couches suivantes les impacts compatibles avec les équations du mouvement, dépendant de la nature et de l'impulsion d'une particule. A ce stade plusieurs trajectoires sont créées ayant des impacts en commun. Elles sont générées avec une valeur de χ^2 qui caractérise leur précision.
- Nettoyage : On utilise la valeur du χ^2 généré par les étapes antérieures pour lever une première fois les ambiguïtés entre les différentes trajectoires. On garde encore un certain nombre de candidats-trajectoires.
- Lissage : la dernière étape consiste à utiliser des outils d'ajustement à rebours en partant du dernier impact identifié et en remontant jusqu'au segment graine. Le lissage sert à renforcer l'identification des points d'impacts appartenant réellement à la trace.

La trace reconstruite est une hélice paramétrée par x, y, z, λ, p . λ est l'angle du vecteur tangent à la trajectoire et dépend du champ magnétique, p est le module de l'impulsion de la particule évaluée au point (x, y, z) . Les filtres de Kalman ont le grand avantage de pouvoir réévaluer les paramètres à chaque étape, ce qui donne des trajectoires plus proches de la réalité [27].

Pour constituer la graine, l'algorithme a besoin d'au moins trois impacts ou deux impacts et une "beam constraint". Il utilise donc seulement une partie des impacts du système de tracker. L'on privilégie la construction de la graine à partir des informations du tracker à pixel qui a une meilleure résolution, une plus grande couverture angulaire et propose une bonne mesure bi-dimensionnelle. De plus, certaines particules peuvent subir des interactions destructives en traversant le système de tracker : il est donc important de considérer les informations le plus près possible du point d'interaction.

Reconstruction des vertex Un événement est présélectionné par le HLT s'il possède au moins un vertex primaire : la reconstruction de vertex est donc une opération menée par les algorithmes du système de déclenchement et non du PF. Nous l'expliquerons dans cette partie car elle est liée à la reconstruction de traces et sera utile pour la suite. La reconstruction de vertex se fait en deux étapes : trouver un vertex en associant plusieurs traces entre elles et compter le nombre de traces pour déterminer s'il s'agit d'un vertex primaire [28].

Une fois un vertex trouvé, on détermine sa position longitudinale et transverse par rapport au point d'interaction et l'on regarde le nombre de traces associées et la qualité de leur reconstruction qui dépend du nombre d'impacts et des valeurs de χ^2/nof obtenus après le lissage.

Un vertex primaire (VP) est considéré comme tel s'il respecte les trois conditions suivantes :

- Au moins quatre traces doivent être reliées au vertex,
- La distance dans la direction z , entre le centre du détecteur et le vertex primaire doit être inférieure à 24 cm,
- La distance dans le plan transverse du détecteur entre l'axe des faisceaux et le vertex doit être inférieure à 2 cm.

Dans les étapes de reconstruction des différents objets, nous allons utiliser les positions relatives des particules au VP définies par les variables d_z (distance longitudinale entre la particule et le VP) et D_{xy} définie par :

$$D_{xy} = \frac{(V_y - VP_y)p_x - (V_x - VP_x)p_y}{\sqrt{p_x^2 + p_y^2}} \quad (4.5)$$

p_x est l'impulsion de la particule dans la direction x, V_x et VP_x sont les position en x du vertex associé à la trace de la particule et du VP dans le référentiel de CMS et idem pour les positions en y.

Reconstruction de traces dans le tracker : la procédure de reconstruction de trace dans le tracker est reconduite 5 fois : Les trois premières itérations permettent d'identifier et écarter un certain nombre de traces reliées au vertex primaire (avec une résolution de 99.5% pour les muons isolés et plus de 90% pour les jets) et la 4ème et la 5ème itérations relâchent les contraintes liées au vertex primaire afin de pouvoir reconstruire des traces d'objets secondaires provenant par exemple d'interactions nucléaires ou de désintégration en vol dans le tracker.

Avec cette technique itérative on arrive à reconstruire avec un taux de fausse identification de l'ordre du pour cent des particules chargées de petite impulsion transverse (150 MeV/c) présentant seulement 3 impacts, et issus d'un vertex (secondaire) situé à plus de 50 cm de l'axe des faisceaux [24].

4.2.1.2 Agrégation

Parallèlement à la détermination des traces, le PF effectue une reconstruction d'objets uniquement à partir des dépôts calorimétriques des ECAL et HCAL. Les dépôts individuels sont rassemblés pour former des agrégats ou amas et être associés à des électrons, photons, hadrons chargés ou hadrons neutres.

La procédure d'agrégation dans les calorimètres doit faire face à plusieurs difficultés :

- Reconstruire les électrons en tenant compte des rayonnements brehmstrahlung,
- Reconstruire les particules neutres comme les photons et les hadrons neutres et distinguer leur dépôts de ceux des hadrons chargés,
- Discriminer des dépôts très proches en position,
- Déterminer l'énergie de hadrons chargés pour lesquels les informations de reconstruction du tracker sont imprécises (par exemple pour des traces de grand p_T avec un rayon de courbure très bas).

Les algorithmes d'agrégations sont appliqués de façon indépendante à chacun des sous détecteurs : ECAL tonneau puis bouchons, HCAL tonneau puis bouchons et enfin les modules de Preshower des ECAL.

Nous allons ici décrire la procédure d'agrégation en prenant pour exemple les ECAL. Pour les HCAL la procédure est équivalente, mais leur granularité est beaucoup plus petite.

De façon générale la procédure d'agrégation se fait en 3 étapes :

- La première est d'isoler des cellules "graines" définies comme telles car elles correspondent à des maxima locaux d'énergie dépassant un seuil. Autour de ces graines seront ensuite agrégés d'autres cellules activées dans le plan (η, φ) . Pour l'ECAL

environ 94% de l'énergie est déposée dans 3x3 cristaux, et 97% dans un carré de 5x5 cristaux.

- Lors de la seconde étape, l'amas proprement dit est formé selon un algorithme de type topologique : on associe à la graine les cellules adjacentes dépassant un certain seuil d'énergie et ayant au moins un côté en commun avec la graine. Le seuil représente les déviations standard du bruit de fond : pour l'ECAL on considère 2 déviations standard du bruit de fond, soit 80 MeV dans le tonneau et 300 MeV dans les bouchons. Pour le HCAL, l'énergie seuil d'agrégation atteint 800 MeV. Cette deuxième étape est appliquée de façon itérative en associant ensuite les cellules à l'amas en formation.
- La dernière étape consiste à déterminer la position et l'énergie de l'amas. On applique à nouveau un algorithme itératif qui considère la graine comme position de base puis on pondère par l'information des 9 cellules adjacentes.

La difficulté est ensuite d'associer traces et amas pour identifier les particules. Comme déjà mentionné dans l'introduction, le Particle Flow procède d'abord à la pré-identification des muons et des électrons avant de procéder à l'identification des hadrons.

4.2.2 Reconstruction des muons

Aux énergies du LHC, le temps de vie des muons produits par l'interaction initiale p-p est suffisamment long⁵ pour qu'ils soient observables dans le détecteur avant leur désintégration.

De plus, de toutes les particules, les muons ont la signature la plus claire dans le détecteur. Ils interagissent très peu avec la matière, subissent peu de perte d'énergie par rayonnement brehmstrahlung et leur trajectoire subit une double courbure qui la caractérise. Leur reconstruction profite de la précision de toutes les chambres à muons (RPC, DT, CSC...) ainsi que du système de tracker.

La reconstruction des muons présente 3 configurations selon l'utilisation des traces du tracker ou non :

La reconstruction locale utilise comme "graine" l'information des chambres à DT et des CSC du détecteur à muons. Elle n'utilise pas l'information du tracker. A partir des impacts dans les sous-détecteurs, on évalue un premier segment de trace : ce segment est ajusté linéairement et on ne garde que les situations où le nombre d'impacts est supérieur à 3 et la valeur du χ^2/ndf inférieure à 20. Ces segments sont utilisés comme graines pour remonter vers le vertex primaire en utilisant la procédure de reconstruction de trace décrite ci-dessus.

La reconstruction standalone étend la reconstruction locale aux RPC. Les muons isolés sont reconstruits en alliant les dépôts des RPC à ceux des DT et CSC et en procédant à la reconstruction de trace.

5. $\tau_\mu = 2.197034 \pm 0.000021 \times 10^{-6} s$, $c\tau = 658.654 m$). C'est également le cas pour les électrons ($\tau_e > 4.6 \times 10^{26} s$) mais ils interagissent fortement avec la matière. En revanche le lepton τ possède un temps de vie très court ($\tau_\tau = (290, 6 \pm 1, 0) \times 10^{-15} s$ et $c\tau = 87, 11 \mu m$). Les τ s'hadronisent dans 85% des cas et forment un jet avant même de rencontrer les premières couches du détecteur.

La reconstruction globale étend la reconstruction standalone aux traces du tracker. Lorsqu'il y a coïncidence entre deux traces du tracker et traces standalone, leurs paramètres sont combinés et réajustés. La phase de lissage tient compte maintenant des impacts du tracker et des chambres à muons. Dans ce travail nous utiliserons des muons reconstruits de façon globale.

Calibration des chambres à muons : L'ensemble des impacts enregistrés dans le muon système ne permet pas de déduire directement l'impulsion transverse réelle du muon. Une calibration est nécessaire. La détermination de l'impulsion transverse est très sensible à l'alignement précis entre senseurs de silicium du tracker et des chambres à muons, à la matière composant les détecteurs et leur disposition ainsi que la non-uniformité du champ magnétique.

On détermine ainsi un facteur de correction en fonction de l'impulsion des muons.

L'effet dû au mauvais alignement entraîne le plus de biais entre modélisation MC et données. La géométrie du tracker était modélisée différemment pour les données et le MC et cette différence a tendance à augmenter linéairement avec l'impulsion du muon.

Pour lever cette différence, on construit un facteur de correction dit de Rochester. Pour les muons de bas p_T entre 0 et 100 GeV, on utilise des échantillons avec des muons provenant d'un $J\psi$ ou d'un Z dont les résonances sont bien connues. Pour les muons de haut p_T la correction est obtenue en comparaison à des échantillons maîtrisés de muons cosmiques (voir [29] p24). La correction de Rochester n'est pas appliquée aux données utilisées dans ce travail.

4.2.3 Reconstruction des électrons

Pour les électrons la reconstruction est plus compliquée que pour les muons : leur trace est beaucoup plus courte et sous l'action du champs magnétique, ils rayonnent abondamment de l'énergie par radiation brehmstralung lors de la traversée du tracker. Dans 50% des cas un électron de 10 GeV perd 50% de son énergie en photon de bremsstrahlung et dans 10% des cas, cette perte d'énergie peut atteindre 95% [24].

Cette propriété est utilisée pour reconstruire des amas : le dépôt le plus énergétique est utilisé comme base pour la procédure d'agrégation décrite ci-dessus. On associe ce premier amas à l'électron. Il s'agit ensuite de reconstruire son rayonnement bremsstrahlung. Pour cela, on utilise la reconstruction de trace de l'électron. Pour les électrons on n'utilise pas l'algorithme standard, type Filtre de Kalman, mais l'algorithme Gaussian-Sum Filter (GSF)⁶

On attache l'amas de l'électron à l'extrémité de l'extrapolation de sa trace dans le tracker à pistes, et pour les amas de photons de bremsstrahlung l'on procède à la technique des tangentes comme représenté sur la figure 4.2 : pour chaque point d'impact de la trace de l'électron dans le tracker, une tangente à la trace est propagée jusqu'au calorimètre électromagnétique et si un amas calorimétrique peut lui être associé, il est alors identifié comme provenant d'un photon de bremsstrahlung.

6. L'algorithme GSF exprime les pertes d'énergies en approximant la formule de Bethe-Heitler par une somme de distribution gaussienne alors que le filtre de Kallman a une approche uniquement gaussienne qui ne pas réaliste dans le cas des électrons [30].

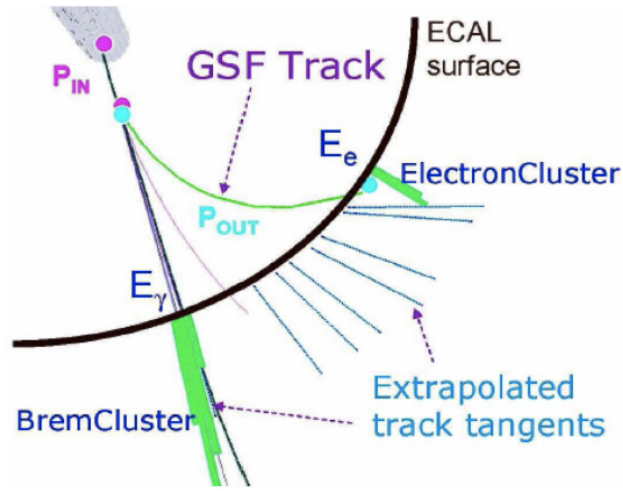


FIGURE 4.2: Illustration de la reconstruction d'un électron par le Particle Flow.

Les photons sont reconstruits avec la procédure d'agrégation dans les ECAL mais à la différence des électrons, on ne peut les relier à aucune trace.

Les calorimètres électromagnétiques sont calibrés pour les électrons et les photons. Ils ne sont pas calibrés pour les dépôts hadroniques.

4.2.4 Reconstruction des hadrons chargés et hadrons neutres

A ce stade, le PF a déjà identifié et reconstruit des muons et des électrons en tenant compte de leur rayonnement électromagnétique. Les impacts et traces correspondantes sont écartées de la suite de l'analyse et l'on procède maintenant à la reconstruction des hadrons [24].

Les hadrons marquent à la fois les ECAL et les HCAL : les associations entre traces et amas seront donc plus élaborées. Les hadrons chargés marquent en plus le tracker, les hadrons neutres eux sont uniquement reconstruits lors de l'agrégation de façon similaire aux photons dans les ECAL. La procédure de reconstruction des hadrons doit donc différencier les amas provenant de hadrons chargés de celle liée à des hadrons neutres ou des photons. La discrimination à partir des traces du tracker ne suffit pas, on procède en plus à des comparaisons entre énergies des amas et impulsion des traces reconstruites. Ceci nécessite la calibration combinée des calorimètres ECAL et HCAL pour les amas hadronique⁷.

La stratégie est de relier d'abord les traces du tracker aux amas de l'ECAL puis de procéder à un lien entre amas des ECAL et HCAL. Les liens privilégiés sont les distances les plus courtes. A chaque étape on demande que l'énergie totale calibrée des amas soit inférieure ou égale à la norme de l'impulsion de la trace ce qui correspond à la physique un hadron chargé. Dans ce cas, les traces sont assignés à des hadrons chargés du PF et l'on détermine leur impulsion et leur énergie à partir de la trace dans l'hypothèse

7. La procédure de calibration combinées des ECAL et HCAL n'est pas décrite ici mais expliquée dans la section 4.2 de [24].

de la masse des pions. Si l'énergie calibrée des amas est supérieure à celle des traces, l'on assigne une partie de ces dépôts à des photons et des hadrons neutres. S'il reste des dépôts reliés à aucune trace on les reconstruit sous forme de hadrons neutres ou photons selon la fraction d'énergie déposée dans les calorimètres. Dans des cas très rares la différence d'énergie est en dessous de trois déviations standard, c'est trop peu pour un hadron chargé, et il s'agit peut être d'un muon oublié. On relance alors une recherche de reconstruction de muon avec des critères plus souples. On écarte à nouveau les traces des nouveaux muons de l'analyse.

4.2.5 Reconstruction des jets

A ce stade, nous nous trouvons donc avec des collections de muons, électrons, photons, hadrons neutres et chargés reconstruits avec toute la précision du PF. Il s'agit maintenant de reconstruire des jets. Les jets sont donc les premiers objets à être observés par CMS, et il est nécessaire de comprendre leurs caractéristiques de façon détaillée : bien reconstruire un jet permet de déterminer l'énergie et la direction du parton qui l'a formé ce qui est essentiel car presque tous les processus souffrent de radiation QCD.

Dans CMS les groupes référents dans l'analyse des jets s'occupent également de reconstruire l'énergie transverse manquante ("Missing Energy transverse" MET), comme par exemple le JetMETPOG [31] ou le "JetMET Algorithms and Reconstruction (JMAR) Group" [32] qui sont en charge de la reconstruction des jets et du MET au niveau PF.

La reconstruction des jets par le PF se base sur les observations suivantes : 65% de l'énergie d'un jet typique est portée par des hadrons chargés (principalement des π^+ et π^-), 25% est emportée par des photons (principalement à partir de $\pi \rightarrow \gamma\gamma$), 10% par des hadrons neutres.

Tous ces objets sont bien reconstruits par le PF, mais pour les hadrons neutres l'on ne peut pas profiter de la résolution du tracker. Le PF va donc pouvoir déterminer 90% de l'énergie et la direction d'un jet avec une bonne résolution et 10% avec une résolution affectée par la faible granularité des HCAL.

Les algorithmes de reconstruction de jets utilisent les quadri-vecteurs (énergie, impulsion) des objets reconstruits par le PF pour établir les "Particle-Flow jets". Nous décrirons brièvement ces algorithmes dans une première partie. Nous parlerons ensuite de la calibration en énergie des jets (JEC).

Les algorithmes de reconstruction de jets En général les algorithmes de reconstruction de jets procèdent par itération d'un dépôt d'énergie à un autre. On peut les séparer en deux groupes selon qu'ils utilisent une technique basée sur des cônes de réjection ou basée sur la distance minimale entre deux points.

La difficulté est d'établir un algorithme à la fois robuste mais rapide afin de pouvoir traiter le volume et la variété d'états finaux produit au LHC. Les algorithmes doivent également être insensibles aux radiations molles de QCD : par exemple, on veut éviter que deux radiations molles émises par une paire $q\bar{q}$ soient combinées entre elles et reconstruites comme un jet alors qu'elles devraient être associées aux jets reconstruits à partir de la paire $q\bar{q}$. On parle d'algorithme ICS (Infrared and Collinear Safe).

CMSSW propose plusieurs algorithmes différents de reconstruction de jets : un algorithme k_T , un algorithme de cône appelé "seedless infrared safe" (SIS), et l'algorithme anti- k_T . Les algorithmes de cônes ont longtemps été privilégiés pour leur rapidité de calcul et leur robustesse. Au LHC on a privilégié une reconstruction délivrant des jets aux limites spatiales stables et régulières par rapport aux radiations.

Pour notre analyse, nous avons utilisé l'algorithme anti- k_T .

L'algorithme anti-kt[33] L'algorithme anti- kt est séquentiel et travaille par paires de particules (ou pseudo-particules). Pour chaque paire i et j l'algorithme calcule les distances relatives (d_{ij}) et les distances entre chaque particule et le faisceau (d_{iB} , d_{jB}).

Il choisit ensuite le minimum de ces distances comparé à un certain cône de rayon R , dont la valeur est choisie par l'utilisateur.

On définit :

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \quad (4.6)$$

$$d_{iB} = k_{ti}^{2p} \quad (4.7)$$

$$\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 .$$

k_{ti} , y_i et ϕ_i représente respectivement l'impulsion transverse de la particule i , sa rapidité et sa coordonnée azimutale.

Tant qu'une paire présente un d_{ij} plus petit que le d_{iB} elle est combinée en un seul dépôt d'énergie. Dès que tous les paramètres d_{ij} possibles pour un dépôt i sont plus grands que le paramètre d_{iB} la somme des dépôts combinés est figée comme étant un jet. Cette procédure continue de manière itérative tant qu'il est possible d'avoir un paramètre d_{ij} plus petit qu'un paramètre d_{iB} .

p est le paramètre qui fixe l'algorithme : sa valeur vaut -1 pour l'algorithme anti- k_T ⁸.

L'avantage majeur de cet algorithme est la faible contamination des jets par l'UE le pileup. Il permet également de délimiter de façon conique la frontière entre deux jets : dans le cas des superpositions, l'algorithme construit une frontière et pondérée par leurs impulsions transverses (cf. figure 4.3).

4.2.6 Calibration des jets

Une fois les jets reconstruits on cherche à déterminer leur direction et leur énergie.

L'énergie des jets ne peut pas directement être déduite de la somme des énergies de ses constituants, les calorimètres hadroniques n'ayant pas une réponse linéaire en l'énergie des jets. On doit donc tenir compte de la calibration des calorimètres en appliquant un facteur de correction à l'énergie des jets.

CMS a développé une série de corrections pour l'énergie des jets (Jet Energy Corrections, JEC) organisés en trois niveaux. Le premier niveau tient compte du pileup et du bruit de fond (UE). Deux autres niveaux calibrent la réponse en énergie pour qu'elle soit plate en p_T et en η . Un niveau corrige la simulation MC par rapport aux données.

Les corrections des différents niveaux sont appliquées les unes à la suite des autres en pondérant avec le nouveau facteur le quadri-vecteur déjà corrigé par les corrections précédentes. On parle d'approche "factorisée".

8. la valeur $p=1$ correspond à l'algorithme k_T , $p=0$ à l'algorithme de Cambridge/Aachen.

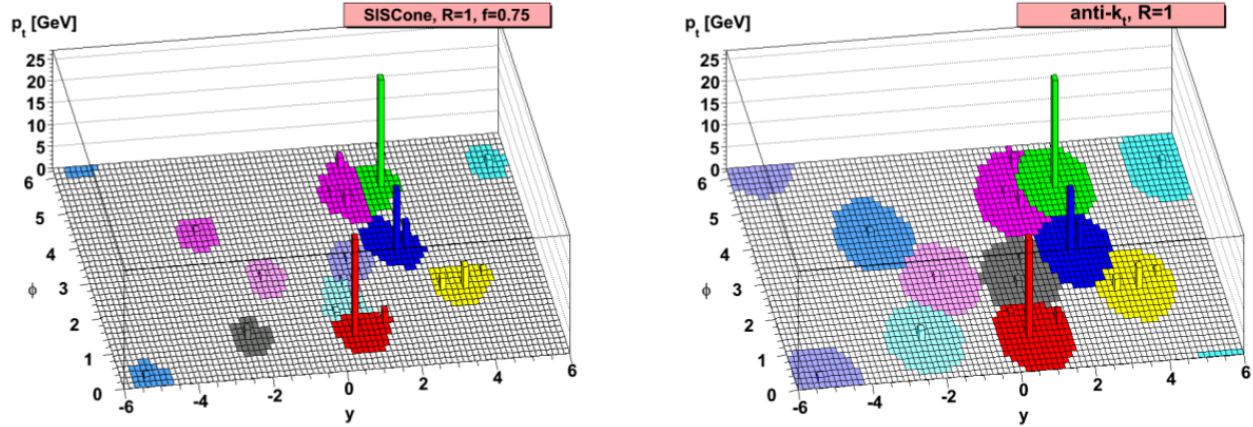


FIGURE 4.3: Exemple de reconstruction de jets par différents algorithmes. L'échantillon utilisé est constitué d'événements simulés au niveau partonique par HERWIG. A chaque processus dur est ajouté de la parton-shower pour tester les algorithmes. Sur les schémas les tours sont proportionnelles à l'énergie des partons, les radiations molles n'excitant qu'une seule cellule. Chaque jet reconstruit a une couleur différente [33].

La calibration des jets à CMS est gérée par le JEC Group"[34] qui tient à jour les différentes corrections à appliquer.

Les corrections sont appliquées jet par jet, en considérant chacune des composantes du quadri-vecteur impulsion du jet identifié et en les multipliant par un facteur de correction [35]. Soit j_μ le quadri-vecteur énergie impulsion d'un jet et C le facteur de JEC : On applique la JEC de la façon suivante :

$$j_{\mu}^{corr} = C \cdot j_{\mu} \quad (4.8)$$

et le facteur C est calculé de la façon suivante :

$$C = C_{offset}(p_T) \cdot C_{MC}(p'_T) \cdot C_{rel}(\eta) \cdot C_{abs}(p''_T) \quad (4.9)$$

p_T est l'impulsion transverse du jet, p'_T l'impulsion transverse corrigée par le offset, p''_T l'impulsion transverse après toutes les corrections.

Les corrections sont calculées en utilisant des échantillons de référence dont on connaît très bien les propriétés.

Tous ces échantillons présentent des caractéristiques communes garantissant une bonne représentation des collision p-p enregistrées par CMS :

- Avoir un vertex primaire avec au moins 4 traces identifiées. Ce vertex doit être situé à moins de 24cm du point d'interaction selon l'axe des z et à moins de 2 cm selon la direction radiale.
- Les critères d'identification des jets développés par le PF doivent être appliqués (cf.section5.3.2.1).

Nous allons décrire brièvement ces échantillons.

Échantillon de zéro bias et minimum bias Comme son nom l'indique, l'échantillon du zero bias doit rendre compte d'événements qui n'ont subis aucune contrainte de sélection. Il est constitué par un nombre d'événements déclenché de façon aléatoire au niveau du trigger lors d'un croisement de faisceau.

L'échantillon du minimum bias comporte une petite composante biaisée : on demande une coïncidence entre les mesures des compteurs de faisceau⁹ placés de chaque côté du détecteur CMS. Les échantillons de zéro et minimum bias vont permettre d'établir la JEC "offset" relative au bruit de fond et au pileup. Nous développerons ces corrections dans le chapitre 5, dans la section 5.4 consacrée au pileup.

Echantillon Dijet Il s'agit d'un échantillon qui comprend uniquement des événements avec au moins 2 jets reconstruits. On demande au HLT de garder les jets de p_T non calibrés supérieurs à 15 GeV. Cet échantillon est utilisé pour établir la correction relative en η ainsi que la résolution en énergie des jets en fonction de leur variable η .

Pour calculer la correction on utilise la méthode de "dijet balancing". On sélectionne un jet dans la région du tonneau ($|\eta| < 1.3$), et on regarde un autre jet de η arbitraire. On demande ensuite que les jets soient dos à dos dans le plan (r, ϕ) soit $\Delta\phi > 2.5rad$.

On définit alors une variable de "balancing" :

$$\mathcal{B} = \frac{p_T^{prob} - p_T^{barrel}}{p_T^{moy}} \quad \text{avec } p_T^{moy} = (p_T^{prob} - p_T^{barrel})/2 \quad (4.10)$$

On détermine les $\mathcal{B}$ en fonction des variable η du jets prob.

La correction est alors déterminée par

$$R_{rel}(\eta_{prob}) = (2 + \langle \mathcal{B} \rangle) / (2 - \langle \mathcal{B} \rangle) \quad (4.11)$$

Echantillon $\gamma + jet$ On établit cet échantillon en déclenchant le HLT si l'événement comprend au moins un γ reconstruit de $p_T > 15\text{GeV}$. On procède ensuite à des coupures offline pour le photon qui assure son isolation et son énergie (l'amas dans l'ECAL doit dépasser un certain seuil). On demande enfin $|\eta|_{photon} < 1.3$. Cet échantillon est utilisé pour établir la correction absolue en p_T ainsi que la résolution en p_T des jet en comparant l'énergie du jet à celle du photon.

Echantillon $Z(\mu + \mu-) + jets$ Cet échantillon correspond plus ou moins à celui que nous étudions. Il est d'un grand intérêt pour établir les calibrations et estimer l'énergie manquante car si les événements sont parfaitement reconstruits, les quadri-vecteurs des muons et des jets se compensent parfaitement dans le plan (r, ϕ) . S'il n'y a pas de jets, les quadri-vecteurs de muons sont dos à dos et leur énergie vaut $M_Z/2$. Pour déterminer cet échantillon on déclenche au niveau HLT les événements ayant deux muons de charges opposés reconstruits de façon globale, avec $p_T > 15\text{GeV}$ et $\eta_\mu < 2.3$. On demande également au moins un jet reconstruit avec $|\eta| < 2.3$. On procède ensuite à une coupure offline sur la masse invariante du système composé des deux muons : $70\text{GeV} < M_{\mu\mu} < 110\text{GeV}$.

9. Il s'agit de détecteurs à scintillation qui ne participent pas à l'analyse de données mais servent à l'établissement de la luminosité à CMS et d'autres développements techniques.

Enfin on demande que le Z reconstruit soit bien dos à dos dans le plan transverse par rapport au jet de plus haut p_T . Pour cela, on impose $|\Delta\phi(Z, jet)| > 2.7rad$

Un échantillon $Z(e+e-)$ + jets est établi de façon similaire à celui ci. Son utilité est la même.

Ces échantillons sont utilisés pour établir la "JEC Absolue", c'est à dire indépendante en p_T .

Pour calculer cette correction, on se base sur le fait que la résolution d'un γ ou un Z est bien meilleure que celle d'un jet. On calcule $R_{abs} = p_T^{jet} / p_T^{\gamma, Z}$

Chapitre 5

Sélection

Une fois les objets physiques reconstruits (particules et jets) il reste quatre grandes étapes préliminaires à l'établissement d'une section efficace du processus $pp \rightarrow Z(\rightarrow \mu^+\mu^-) + jets$:

1. Sélectionner l'échantillon à analyser : dans notre cas un boson Z se désintégrant en une paire de muons et les jets associés,
2. Nettoyer l'échantillon obtenu en identifiant et soustrayant pileup et bruit de fond,
3. Ajuster la simulation MC en procédant à des corrections liées aux procédures de déclenchement et de sélection,
4. Passer du niveau "détecteur" au niveau "particule" en appliquant une procédure d'Unfolding.

Dans ce chapitre nous décrirons brièvement les fichiers de données que nous avons utilisés, puis nous parlerons des étapes de sélection et de soustraction du pileup. L'établissement des facteurs de correction pour les efficacités et la soustraction du bruit de fond seront présentés au chapitre 6, la procédure d'Unfolding sera présentée au chapitre 7.

5.1 Fichiers de données

Les fichiers de données à partir desquels nous avons travaillé sont présentés dans le tableau 5.1. La prise de données correspondante été effectuée en 2012 à une énergie de 8 TeV dans le centre de masse et pour une luminosité intégrée totale de 19.3 fb^{-1} .

Les runs utilisés correspondent à ceux repris dans les fichiers de certification présentés dans le tableau 5.2. Ces fichiers en format JSON¹ sont générés par CMS lorsqu'un run (ou une partie du run) est validé. En effet, pendant une prise de données il se peut qu'une partie du détecteur (soit un des sous-détecteurs, soit l'aimant ou le système de trigger, etc...) n'ait pas fonctionné correctement : le run n'est alors pas validé et les données sont rejetées. Les fichiers JSON "Prompt" correspondent aux fichiers RECO (cf. section 3.2.8) traités une seule fois. Ce sont les premiers fichiers délivrés par CMS après un run. Les fichiers ReReco sont en fait des fichiers RECO ayant été vérifiés. Par exemple, si une partie du run particulièrement bruyante était dans un premier temps écartée du Prompt, il se peut qu'après une deuxième analyse on décide de la conserver parce qu'on a identifié

1. Java Script Object Notation

la source du bruit. La luminosité peut donc varier entre fichiers Prompt et ReReco pour un même run.

Nom du fichier	# Events	Lumi (fb ⁻¹)
/DoubleMu/Run2012A-13Jul2012-v1	4143940	0.81
/DoubleMu/Run2012B-13Jul2012-v4	20027612	4.37
/DoubleMu/Run2012C-24Aug2012-v1	2069195	0.49
/DoubleMu/Run2012C-PromptReco-v2	24707287	6.38
/DoubleMu/Run2012D-PromptReco-v1	28520349	7.24
Luminosité totale		19.3

TABLE 5.1: Description des fichiers de données utilisés pour notre analyse.

Cert_190456-196531_8TeV_13Jul2012ReReco_Collisions12_JSON.txt
Cert_198022-198523_8TeV_24Aug2012ReReco_Collisions12_JSON.txt
Cert_190456-208686_8TeV_PromptReco_Collisions12_JSON.txt

TABLE 5.2: Fichiers JSON utilisés pour notre analyse.

5.2 Sélection online

Une première sélection de données est opérée par le choix du lot de données analysées. Pour notre analyse nous avons utilisé le lot (Data Stream) DoubleMu correspondant à l'ensemble des événements sauvés par au moins un des trigger HLT "double muons". Ensuite nous nous sommes limités aux événements comportant au moins 2 muons avec un premier muon d'impulsion transverse supérieure à 17 GeV et un second muon d'impulsion transverse supérieure à 8 GeV. Soit les événements passant la sélection trigger suivante : HLT_Mu17_Mu8_v*. Ces critères sont établis afin d'obtenir un échantillon précis tout en conservant une statistique presque maximale.

Cette première sélection au niveau trigger entraîne une légère baisse d'efficacité que nous décrirons plus précisément au prochain chapitre.

Une deuxième étape consiste à effectuer une correspondance ("matching") entre la paire de muons des fichiers Particle Flow et les muons qui ont déclenché le HLT². Pour chaque fichier PF de notre échantillon on ne garde que les muons qui ont déclenché le trigger en demandant que ces muons coïncident à l'intérieur d'un cône de $\Delta R^3 < 0.3$.

L'échantillon ainsi sélectionné contient donc des événements possédant au moins 2 muons. En réalité la proportion d'événements comportant plus de 2 muons est très faible : pour notre échantillon, sur 10210451 événements, on compte 927 événements comprenant plus de 2 muons et 38 événements comprenant plus de 3 muons.

2. Ce double matching est nécessaire pour appliquer les corrections liées à l'efficacité de trigger établies en fonction des domaines en η des muons reconstruits au niveau trigger (cf. plus loin, chapitre 6).

3. $\Delta R = \sqrt{(\eta_\mu)^2 + (\varphi_\mu)^2}$.

La sélection de l'échantillon est ensuite affinée par une série de coupures effectuées offline sur le fichier TTree des données reconstruites.

5.3 Sélection offline

Dans cette partie nous allons passer en revue les coupures de sélection effectuées pour isoler l'échantillon le plus pur possible de nos objets d'étude : c'est-à-dire comprenant des événements avec deux muons issus de la désintégration d'un boson Z et les jets associés pour lesquels nous considérerons les 4 premiers jets classés par ordre de p_T décroissant.

La stratégie de notre recherche se base sur une sélection de type "coupure" pour laquelle le détecteur donne de très bons résultats (l'analyse pourrait être menée sur des sélections de type MVA (Multi Variate Analysis) par exemple). Les coupures concernent principalement les variables combinées d'identification et d'isolation des objets du PF ainsi que des coupures cinématiques qui assurent un bon domaine de résolution et d'acceptance pour notre échantillon et réduisent les contaminations par le bruit de fond ou le pileup.

La sélection concerne d'abord une paire de muons issue de la désintégration d'un boson Z, nous traiterons cette démarche dans une première partie. Nous décrirons la sélection des jets associés dans une seconde partie.

5.3.1 Sélection des muons

La sélection online a déjà sélectionné des événements comprenant au moins deux muons ayant déclenché le trigger asymétrique décrit en 5.2. Ces muons sont reconstruits au sens du PF comme décrit au chapitre 4.

Pour l'instant, les muons de cet échantillon ne proviennent pas forcément de la désintégration d'un Z. Ils peuvent également provenir de la désintégration de W, de la désintégration de quarks lourds, de désintégrations en vol de pions ou de kaons ou et, dans certains cas appartenir aux états finaux d'un jet ("hadronic punchthrough").

A CMS, des observations et des études à partir de logiciels de simulation ont pu déterminer les principales tendances de production de muons selon les domaines cinématiques [29].

Aux petites valeurs de p_T , la plupart des muons provient de processus QCD de quarks légers comme les désintégrations de kaons et pions. Pour des p_T entre 10 et 30 GeV ce sont les muons des quarks c et b et t qui sont les sources principales, alors que pour $p_T > 30$ GeV, ce sont les muons issus des Z et W qui dominent. Nous verrons dans la partie consacrée au bruit de fond la contribution de ces différents processus sur le pic du Z (cf. section 6.2). La question est donc d'identifier les muons provenant de la désintégration d'un Z par rapport aux autres processus. Les contributions les plus difficiles à discerner sont les muons issus d'une désintégration en vol ou des muons produits par l'interaction des hadrons dans le détecteur.

Les coupures appliquées sont résumées dans le tableau 5.3. Elles concernent les variables combinées d'isolation et d'identification que nous allons décrire plus longuement. Les coupures cinématiques permettent de rester dans des bons domaine de résolution du détecteur.

Coupures d'identification et d'isolation
Muon PFID Tight $PF\Delta\beta_{Iso} < 0.2$
Coupures cinématiques
$p_{T,\mu} > 20 \text{ GeV}/c$ $ \eta < 2.4$

TABLE 5.3: Liste des coupures de sélection appliquées aux muons. Les variables utilisées sont décrites dans le texte.

Une fois ces coupures appliquées à chaque muon, événement par événement, l'on vérifie que l'événement possède au minimum une paire de muons. Puis on vérifie que les deux muons de plus haut p_T sont de charges opposées. Si ce n'est pas le cas, on vérifie la même chose pour le premier muon et un éventuel troisième muon. Si ces derniers sont également de même charge, l'on rejette l'événement. Enfin, l'on procède à une coupure sur la masse invariante de la paire de muons sélectionnés pour bien isoler le pic du Z. On se place à 20 GeV de part et d'autre de la masse du Z soit : $71 < M_{\mu\mu} < 111 \text{ GeV}$.

Le tableau 5.4 présente un décompte des événements relatif à chaque coupure appliquée à notre échantillon de données.

Données		$\frac{\#ev.selection}{ev.total}(\%)$
# événements	10210451	100
Au moins deux muons avec :		
$ \eta < 2.4$ et $p_T > 20 \text{ GeV}$	8379488	82.06
PFID Tight	7789057	76.285
$PF\Delta\beta_{Iso}$	7395608	72.431
Charge opposée	7395416	72.429
$ M_{\mu\mu} - M_Z < 20$	6962431	68.189

TABLE 5.4: Compteur pour la sélection des muons de notre échantillon de données. Les coupures sont appliquées l'une après l'autre dans l'ordre du tableau. La colonne de droite représente la fraction d'événements préservés par la coupure, les coupures précédentes étant déjà appliquées.

5.3.1.1 Identification des muons

La variable appelée ID du PF est une variable combinée d'identification des muons. Pour les muons nous utilisons une identification dite Tight du PF comme le conseille le Muon POG [36] pour les muons prompt associés à un W ou Z.

Cette sélection demande que les muons identifiés comme Tight aient les caractéristiques suivantes :

Un muon qui a passé l'identification Tight a donc sa trace bien définie, du tracker aux chambres à muons. Son lien au vertex primaire est également assuré par les variables d_Z

Muon reconstruit comme muon global
Muon du Particle-Flow
Le fit de la trace du muon global a $\chi^2/\text{ndof} < 10$
Le fit de la trace du muon global contient au moins un impact dans une chambre à muons
Le segment du muon traverse au moins deux stations différentes dans les chambres à muons
On sélectionne uniquement les couches de tracker avec au moins 5 impacts
Nombre de pixels excités > 0
La distance transverse de la trace reconstruite au Vertex primaire est $D_{xy} < 2 \text{ mm}$
La distance transverse de la trace reconstruite au Vertex primaire est $dz < 5 \text{ mm}$

et D_{xy} qui rendent compte de la distance longitudinale et transverse au vertex. Cette coupure d'identification réduit de façon significative le taux de muons qui proviennent d'une désintégration en vol. Le prix à payer est une perte d'efficacité de l'ordre du pourcent comme nous le verrons à la section 6.1.2.

5.3.1.2 Isolation des muons

Le Particle Flow applique une procédure d'isolations aux muons reconstruits : c'est l'Isolation Particle Flow (IPF). Cette variable est construite dans le but de vérifier qu'il y a peu ou pas de dépôts calorimétriques dans un cône d'un certain rayon autour de la trace reconstruite du muon.

On établit le cône d'isolation de rayon $\Delta R < 0.4$ autour du vecteur tangent à la trace du muon au point d'interaction. On rejette la trace du muon en établissant un cône veto de $\Delta R < 0.1$ dont le centre se déplace suivant la trace du muon (cf.figure 5.1). Dans le volume restant l'on considère les impulsions transverses de tous les hadrons chargés, photons et hadrons neutres et on l'on compare l'ensemble de ces impulsions au p_T du muon. Si le rapport est supérieur à un certain seuil, l'on classe le muon comme n'étant pas isolé.

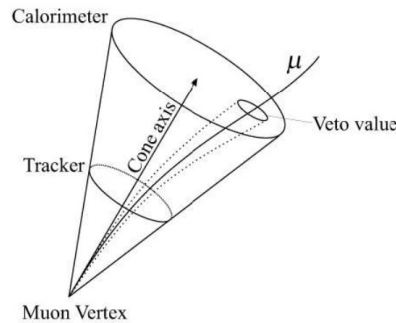


FIGURE 5.1: Illustration schématique de la procédure d'isolation des muons[37].

La variable d'isolation que nous utilisons est de type $\Delta\beta^4$. C'est une variable combinée

4. Les variables β sont souvent utilisées par le PF : elles représentent des moyennes d'impulsion transverse.

basée sur les infos contenues dans le Particle Flow et présentée ci-dessous.

$$\Delta\beta_{Iso}(p_T) = \left(\sum_{Ch.had} p_T + \max(0, \sum_{N.had} p_T + \sum_{\gamma} p_T - 0.5 \sum_{PU} p_T) \right) / p_T^{muon} \quad (5.1)$$

Le pileup est soustrait avec un facteur 0.5 qui correspond à une moyenne naïve du rapport des hadrons neutres sur hadrons chargés comme déterminé dans l'étude [38].

La variable d'isolation réduit de façon significative le taux de muons qui proviennent de jets. Cette coupure va également entraîner une perte d'efficacité de l'ordre du pour-cent.

5.3.2 Sélection offline des jets

Notre échantillon contient donc des muons bien sélectionnées comme décrit ci-dessus. Ces muons proviennent principalement de la désintégration d'un Z et ont une masse invariante comprise entre 71 GeV et 111 GeV.

Il est maintenant question de sélectionner les jets associés à notre processus $pp \rightarrow Z$. Le fichier TTree présente le nombre de jets associé à chaque événement : ces jets sont reconstruits par le PF et leurs directions et énergies sont déterminées. Les jets sont triés dans le TTree par p_T décroissants.

Mais ces jets ne sont pas forcément associés à la production d'un Z : ils peuvent être issus du pileup ou ce sont peut-être des jets mal reconstruits. Un jet peut, par exemple, être confondu avec un lepton τ qui s'est hadronisé ou avec des photons ou des électrons et muons mais avec une plus faible probabilité. Il s'agit donc de sélectionner des "bons jets", identifiés avec des critères qui réduisent le taux de mauvaise identification et dans un domaine cinématique peu contaminé par le pileup. Dans cette partie, nous allons passer en revue les choix offline d'identification et d'isolation des jets au sens du PF.

Coupures cinématiques sur les jets $p_{T,jet} > 30 \text{ GeV}/c$ $ \eta < 2.4$
Coupures d'identification et d'isolation Jet Loose Id Le jet n'est pas un muon, soit : $\Delta R = \sqrt{(\eta_{\mu 0,1} - \eta_{jet})^2 + (\phi_{\mu 0,1} - \phi_{jet})^2} \leq 0.5$
Sélection concernant le pile-up $PUJetID > -0.89$

TABLE 5.5: Liste des coupures effectuées sur les jets, les critères sont développés dans ce qui suit.

Une coupure de sélection du pileup est également effectuée. Elle est particulièrement efficace pour s'affranchir d'une partie du pileup résiduel. Elle sera discutée dans la section 5.4 qui reprend toutes les coupures liées à ce problème.

5.3.2.1 Identification des jets

Cette variable est utilisée pour valider l'identité du "jet". Il s'agit d'une variable combinée du PF basée sur l'analyse des traces dans les calorimètres. Des études (basée sur des simulations MC) ont montré qu'un bon critère de détermination d'un jet est la fraction d'énergie du jet déposée dans le calorimètre électromagnétique (EMF). La variable JetLooseId se base sur l'EMF et prend en compte d'autre critères comme le nombre de hadrons chargés composants le jet et les multiplicités. Les critères de sélection sont repris dans le tableau 5.6.

Pour mon analyse j'ai sélectionné des jets correspondant à l'identité Loose comme recommandé par le JetMET POG [39]. La contamination due aux faux jets étant très faible, des critères plus sévères ne feraient que diminuer l'efficacité de la sélection.

PF Jet ID $ \eta > 2.4$	Loose (Recommandé)	Medium	Tight
Fraction d'hadrons neutres	<0.99	<0.95	< 0.90
Fraction EM neutre	<0.99	<0.95	< 0.90
Nombre de constituants	>1	>1	> 1
Pour $ \eta < 2.4$ on doit appliquer les coupures supplémentaires suivantes			
Fraction d'hadrons chargés	>0	>0	>0
Multiplicité chargée	>0	>0	>0
Fraction d'EM chargée	<0.99	<0.99	<0.99

TABLE 5.6: Critères d'identification des jets [39].

5.3.2.2 Isolation des jets

Cette sélection s'applique au cas précis de notre analyse : une fois les deux muons sélectionnés on demande que notre jet ne soit pas un de ces muons. La procédure choisie est de déterminer des cônes de sélection de rayon ΔR associés aux deux muons et de vérifier pour chaque jet qu'il se trouve hors de ces cônes de sélection.

$$\Delta R = \sqrt{(\eta_\mu - \eta_{jet})^2 + (\phi_\mu - \phi_{jet})^2}$$

$$\Delta R \leq 0.5$$

Pour notre échantillon comprenant 10210451 événements, on obtient 1144443 événements avec au moins un jet, 238173 événements avec au moins 2 jets, 47446 événements avec au moins 3 jets et 9407 événements avec au moins 4 jets. Nous présenterons les distributions pour les multiplicités de jet aux chapitres 6 et 7.

5.4 Soustraction du pileup

Notre stratégie vise à sélectionner uniquement les jets associés au vertex primaire du Z et pas ceux associés à des vertex secondaires. Or au LHC, pour chaque collision il se passe bien plus qu'une seule interaction proton-proton : c'est ce qu'on appelle le pileup. Par exemple pour notre échantillon de données 2012, il se produit en moyenne 15 collisions

simultanées par croisement de faisceaux et le nombre de vertex reconstruits par événements dépasse souvent 30 comme le montre la figure 5.2. Les différentes collisions peuvent provenir d'un même croisement de paquets ou de résidus des collisions précédentes : on parle alors de pileup "out of time".

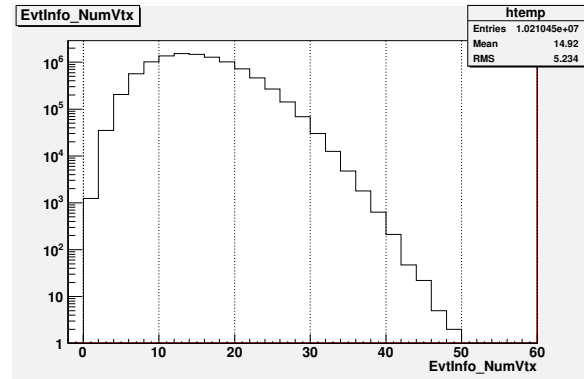


FIGURE 5.2: Distribution du nombre de vertex pour les données 2012. En abscisse est représenté le nombre de vertex, et en ordonnées le nombre d'événements.

Le pileup est sans doute la manifestation expérimentale la plus difficile à gérer dans une expérience comme le LHC : il est massif et il est difficile à simuler. Maîtriser le pileup est donc un énorme challenge et c'est une opération sans cesse améliorée par la collaboration : c'est une problématique gérée par plusieurs groupes de travail (les JMetPOG, les groupes du tracker, de PUJetID, etc...).

De plus une définition des jets associés au pileup est difficile à établir : le critère de lien au vertex primaire ne suffit pas et l'on doit considérer en plus d'autres propriétés.

Jusqu'à présent la procédure d'identification soustraction du pileup qui présente les meilleures performances procède en trois étapes :

- Une première étape se passe pendant la reconstruction lors de la calibration des jets. C'est d'ailleurs la première correction appliquée par le JECGroup.
- Une seconde étape est la "charge fraction subtraction", elle concerne principalement le lien des objets du PF au vertex reconstruit. Elle est appliquée après la JEC par les algorithmes du PF.
- Une troisième étape se passe lors des sélections online. Une fois l'échantillon de dimuons constitué l'on applique une coupure sur une variable combinée qui concerne les jets appartenant au pileup résiduel. Cette coupure est appelée PUJetID.

Nous avons mené une étude concernant cette dernière variable qui permet d'illustrer les problèmes expérimentaux liés au pileup et la conséquence directe sur le choix des domaines cinématiques de notre analyse. Dans la suite de cette section nous présenterons les trois étapes de sélection du pileup et leur principe de fonctionnement puis nos résultats concernant la coupure PUJetID.

5.4.1 Soustraction du pileup au niveau JEC

La correction au niveau JEC est la première étape dans la chaîne des calibrations : son but est d'estimer le bruit de fond dû entre autre au pileup contaminant les diffusions

de haut p_T , et d'en déduire une correction à appliquer à l'énergie des jets. Deux méthodes sont proposées par CMS : une méthode des moyennes ("average offset") qui considère le lien des jets au vertex primaire et une méthode qui considère l'aire efficace des jets par des algorithmes de "jet area".

La méthode d'offset moyen donne lieu à une correction appelée L1. La méthode de jet area donne lieu à la correction L1FastJet, gérée par le même logiciel que l'algorithme anti-kT.

Il existe une méthode hybride calculant une correction à partir des deux méthodes. C'est cette correction hybride qui est appliquée aux jets du PF.

Nous parlerons de façon générale des méthodes de jet area et d'offset moyen.

5.4.1.1 Méthode de "jet area"

La méthode de jet area vise à séparer de façon topologique les jets issus du hard-process de ceux du pileup : on se base sur une discrimination concernant l'étalement des gerbes hadroniques dans le plans η, ϕ . Cette discrimination est basée sur les observations des jets reconstruits avec les algorithmes de reconstruction de jets : des jets de bas p_T issus du pileup peuvent être agrégés lors de la reconstruction en faux jets de haut p_T . Ces jets issus du pileup contiennent plus de radiations et leur forme est plus étalée dans le plan transverse.

A partir de cette identification, une correction est calculée et appliquée à l'impulsion des jets lors de la procédure de calibration.

La méthode de jet area est basée sur la définition des aire actives [40], A_i , associées au jet i . A partir des aires actives on définit une variable ρ représentant la densité moyenne en p_T par unité de surface associée au jet. La méthode peut être utilisée avec n'importe quel algorithme de reconstruction de jets pour lequel on peut définir A_i et ne dépend pas de comparaisons avec des logiciels de simulation. A_i et ρ ne sont pas des données triviales : pour déterminer l'aire active d'un jet l'on ajoute un grand nombre de radiations (infiniment) molles et on compte combien ont été agrégées au jet par l'algorithme de reconstruction utilisé. L'aire active mesure donc la susceptibilité moyenne d'un jet à une contamination dense et uniforme venant de l'Underlying event ou du pileup [41].

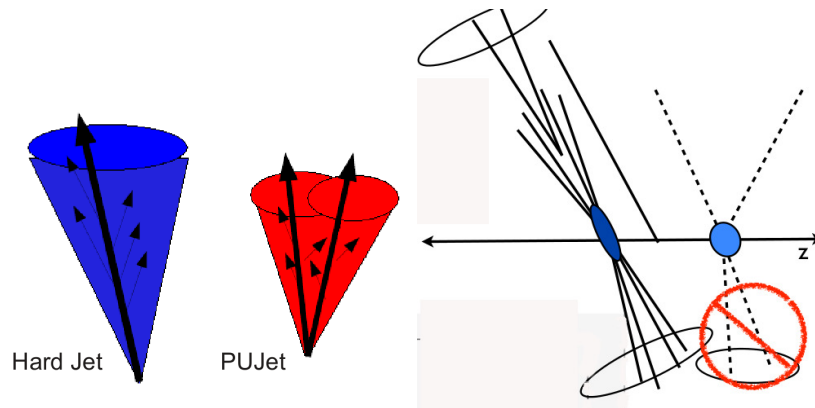


FIGURE 5.3: Schémas représentant les principes inspirant les méthodes de jet area (à gauche) et d'offset moyen (à droite) [42].

Une fois la variable A_i déterminée, on calcule la variable ρ en considérant pour chaque jet le rapport $p_{T,j}/A_j$. Si l'algorithme anti-kT est utilisé, on détermine ρ par la médiane de la distribution $p_{T,j}/A_j$, où j parcourt tous les jets de l'événement qui ont $|\eta| < 5$. ρ représente la convolution de l'activité molle sous-jacente (pileup, underlying event) avec la réponse du détecteur aux différents types de particules.

La figure 5.4 représente la détermination de ρ pour un échantillon généré par Pythia et comprenant des événements avec deux jets et présentant un pile up de 22 interactions. On voit que les jets du pileup se trouvent sur la médiane et que le pileup sous-jacent est bien uniforme en η comme c'est le cas sur la figure de droite.

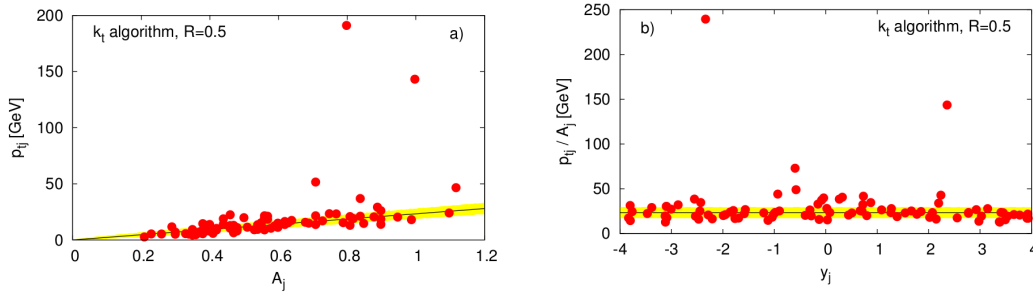


FIGURE 5.4: Exemple illustrant la détermination de la variable ρ à partir d'un échantillon MC (voir texte). A gauche, le p_T des jets en fonction de leur aire active : la bande jaune représente $\rho.A_j \pm \sigma.\sqrt{A_j}$. A droite, le rapport $p_{T,j}/A_j$ en fonction de la rapidité : la bande jaune représente $\rho \pm \sigma/\sqrt{\langle A \rangle}$ [40].

Le facteur de correction L1FastJet utilisé pour la JEC est ensuite calculé de la façon suivante :

$$\mathcal{C}_{area}(p_T^{RAW}, A_j, \rho) = 1 - \frac{(\rho - \langle \rho_{UE} \rangle) \cdot A_j}{p_T^{RAW}} \quad (5.2)$$

p_T^{RAW} est l'impulsion transverse du jet non calibré. $\langle \rho_{UE} \rangle$ est l'activité moyenne due à l'underlying event et au bruit électronique : elle est calculée en utilisant un échantillon avec uniquement des événements à 1 vertex. Il faut garder à l'esprit que la définition de ρ concerne toutes les contaminations par radiation que ce soit du pileup ou de l'underlying event. La correction L1FastJet est appliquée pour chaque événement, jet par jet.

5.4.1.2 Méthode d'offset moyen

La méthode d'offset moyen mesure l'énergie moyenne due à la contamination du pileup et du bruit de fond. Pour cette méthode la détermination du pileup est directement lié à la détermination des vertex primaire. C'est une correction dite de "offset" puisque la correction calculée va décaler de manière générale les distributions [35].

La méthode utilise les échantillons du zéro et minimum bias comme décrits dans la section 4.2.6. On applique un veto aux événements du zéro bias qui ont passé le trigger du minimum bias. Pour les événements restant on somme les dépôts d'énergie à l'intérieur d'un cône défini par un rayon $R = 0.5$ dans le plan (η, φ) . Cette somme est établie en fonction d'une direction fixe en η et en moyennant dans la direction φ . On détermine

ensuite la correction moyenne sur l'entièreté de l'échantillon et on l'exprime en fonction du nombre de vertex pour les jets du minimum bias.

La correction pour l'offset moyen est donnée par :

$$\mathcal{C}_{offset}(p_T^{RAW}, \eta, N_{PV}) = 1 - \frac{(N_{PV} - 1) \cdot \mathcal{O}(\eta)}{p_T^{RAW}} \quad (5.3)$$

Ici p_T^{RAW} représente l'impulsion transverse du jet non calibré, $\mathcal{O}(\eta)$ est la pente de la moyenne des p_T sur le nombre de vertex (en GeV/N_{PV}) calculée en fonction de η , N_{PV} est le nombre de vertex primaires reconstruits. On définit également la quantité $\langle p_{T,offset} \rangle = (N_{PV} - 1) \mathcal{O}(\eta)$

La figure 5.5 représente la fonction $\langle p_{T,offset} \rangle$ de la correction moyenne offset pour des événements associés à un premier vertex du pileup. Les différents objets du PF y sont également représentés. On peut voir que la distribution est plus ou moins uniforme en η sauf pour les leptons qui présentent des "pointes" aux alentours de $\eta = 3.5$.

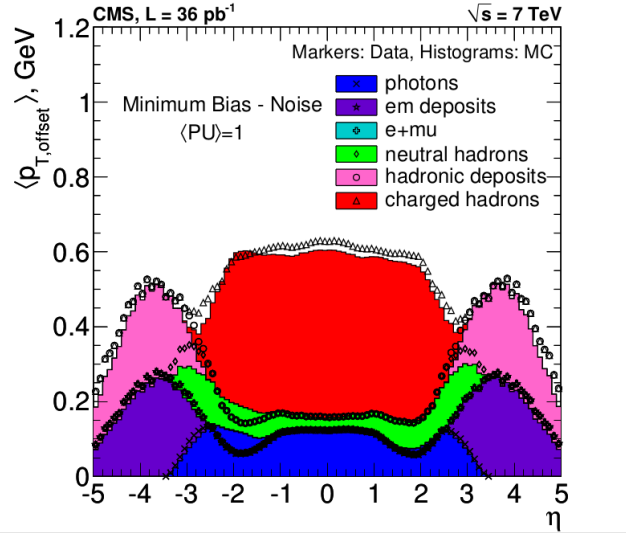


FIGURE 5.5: Correction moyenne offset en fonction de η et des objets du PF. Les données sont représentées par des points, le MC par des histogrammes pleins [35].

La figure 5.6 représente les corrections de jet area, d'offset moyen et la méthode hybride en fonction du nombre de vertex et pour 6 domaines différents en η . La méthode hybride de calibration allie les principes de jet area et d'offset moyen, c'est celle qui est appliquée aux jets du PF. Pour les CaloJet et JPT on applique uniquement la correction obtenue avec la méthode d'offset moyen. Sur cette figure l'on peut observer que la correction hybride correspond plutôt à la correction average offset, pour $|\eta| < 1.5$ la correction de jet area est également en accord avec la correction average offset, même si on observe des écarts pour un grand nombre de vertex. A partir de $|\eta| > 1.5$ ces écarts deviennent plus importants.

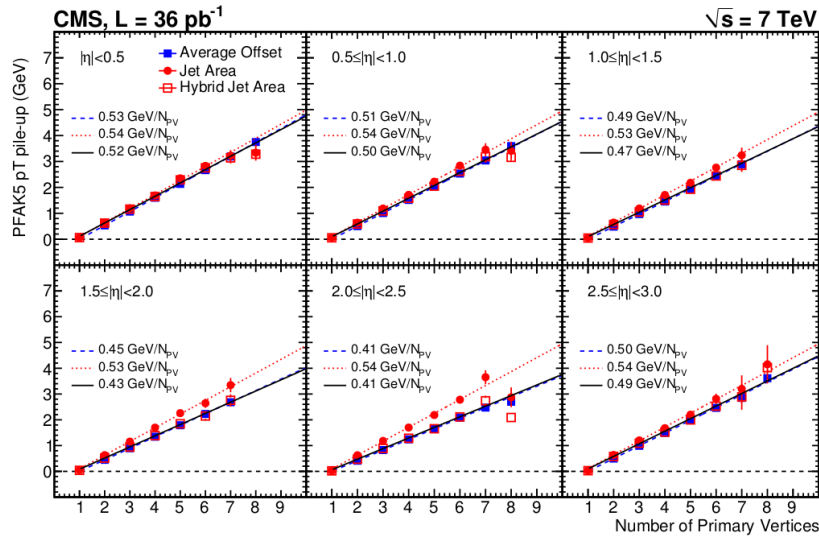


FIGURE 5.6: Corrections pour la méthode d'offset moyen ($\mathcal{O}(\eta)$), de jet area et méthode hybride en fonction du nombre de vertex. En ordonnée PFAK5 pT désigne les corrections pour les jets du PF reconstruits avec l'algorithme anti-kT et $R=0.5$ [35].

5.4.2 Soustraction du pileup par la Charge Hadron Substraction

Une deuxième étape consiste à éliminer de notre échantillon tous les hadrons chargés n'étant pas associés à notre vertex primaire : c'est la "Charge Hadron Substraction" (CHS) ou soustraction des hadrons chargés [43]. Cette étape est une option proposée dans les algorithmes de reconstruction du PF et appliquée après les JEC à l'échantillon PF calibré.

La CHS consiste à éliminer de l'échantillon tous les hadrons chargés associés à un vertex secondaire identifié comme PU. Par exemple, dans la figure 5.5 cela correspond à la partie centrale en rouge.

La CHS est donc directement liée à la reconstruction des vertex menée par les algorithmes de reconstruction de trace. Selon la CHS, pour les hadrons on considère que si $\chi^2/dof < 20$ pour un certain vertex alors la trace est associée à ce vertex et uniquement à celui-ci. Ensuite la distinction entre vertex est déterminée en calculant la somme des carrés des traces transverses, $(\sum |p_T^{TRK}|^2)$ pointant vers ce vertex. Le vertex primaire est associé à la plus grande valeur de cette somme, les vertex secondaires sont assignés au pileup. Le CHS enlève également une partie des leptons et ceci en considérant en plus le paramètre d'impact longitudinal dz : on demande $dz < 5cm$. Si le vertex associé est identifié comme vertex pileup, le lepton reconstruit est écarté. Cette sélection est désignée par "agressive CHS" car pour les hadrons on ne considère pas la distance longitudinale dz .

La CHS est appliquée dans un domaine $|\eta| < 2.5$ pour lequel le tracker donne une bonne résolution. C'est une opération prise en charge par CMSSW et conseillée par le JECGroupe.

Après la CHS, on obtient typiquement des distributions où, pour des jets d'impulsion 25 GeV, la moitié des jets sont encore issus du pileup et l'autre moitié est associée au

vertex du Z.

5.4.3 Soustraction Offline du pileup par l'algorithme PileupJetID

Après les opérations de calibration et de CHS, il reste un certain pileup résiduel : soit parce que les traces n'ont pas pu être associées à un vertex PU, soit en l'absence de traces (particules neutres ou chargées mais de traces non reconstruites...). Ce pileup va être soustrait par une dernière coupure menée offline et calculée par l'algorithme PUJetID [44]. Cette coupure est issue d'une analyse à variables multiples ("Multi Variate Analysis", MVA) qui prend en compte la topologie des jets et le lien entre traces et vertex. Dans ce travail nous avons étudié l'action de cette variable. La sélection effectuée est **PUJetID** > -0.89 comme conseillé par le JetMETPOG [44].

5.4.3.1 Principe de Multi Variate Analysis

L'expérience CMS fournit une interface centralisée permettant d'appliquer des méthodes d'analyse à variables multiples. Cette interface s'appelle le CMS MVA Package. L'analyse MVA s'applique à toutes sortes d'analyses comme par exemple l'association jet/parton ou la différenciation signal/background...

Pour chaque événement on choisit un certain nombre de variables et le package MVA applique à ces variables différents algorithmes : les différentes étapes consistent à normaliser les variables, décorréler les variables en établissant des matrices de rotation et enfin, procéder à une interpolation pour déterminer une variable unique, le discriminant.

Les algorithmes d'interpolation fonctionnent soit de façon linéaire, soit avec l'interpolation du maximum de vraisemblance (ProcLikelihood) ou encore avec le BDT (Boost decision Tree) qui permet de traiter toutes les variables en même temps et d'étudier différentes configuration afin de déterminer la combinaison optimale [45].

Dans le cas du pileup, le discriminant est établi à partir d'une analyse utilisant un BDT et 9 variables. 4 variables représentent les propriétés liées au vertex associé au jet, 5 variables sont liées à la topologie du jet. Les variables associées au vertex ou à la trace du jet servent à explicitement cibler le vertex primaire : leur domaine d'action est limitée à celui du tracker ($|\eta| < 2.5$) et leur efficacité est faible dans le cas de pileup élevé.

Les variables associées à la forme du jet prennent le relais pour les domaines d'acceptance des bouchons (où ECAL et HCAL sont toujours présents et enregistrent la plupart des jets), et pour $|\eta| > 3$ (où seul le HF est valide).

L'algorithme de PUJetID détermine également un discriminant à partir de simples coupures ("cut-based") concernant les mêmes variables et qui est plus robuste. Une procédure est également appliquée à l'échantillon n'ayant pas subi de CHS. Pour toutes ces procédures, on détermine trois points de fonctionnement, Loose, Medium et Tight.

Le discriminant va permettre de sélectionner dans notre cas l'association jet/Pileup en demandant que sa valeur soit supérieure à -0.89. Cela correspond à la sélection dite Loose de PUJetID.

5.4.3.2 Variables utilisées par le PileupJetID

Dans le cas de d'identification d'un jet du pileup, l'analyse se base sur deux types de propriétés du jet :

1. au niveau du tracker, l'association ou non du jet avec le vertex primaire. Les variables concernées sont β , β^* , d_Z , et N_V le nombre de vertex.
2. au niveau des calorimètres, la forme des jets permet de déterminer si le jet est diffus ou s'il est très collimaté. Ces variables permettent de séparer les "vrais" jets durs de ceux correspondant à des radiations molles de gluons faussement reconstruits en jet durs : ce sont les variables $\langle \Delta R^2 \rangle$, $\sum_{\Delta R_i} p_T^i$, p_T^D . On considère également les multiplicités $N_{charges}$, $N_{neutral}$.

Variables de vertex Les variables de vertex sont principalement des variables β dépendent de la définition des vertex donnée par la CHS et de l'association du jet à ces vertex. On utilise aussi d_Z , qui représente la distance longitudinale de la trace de plus haut p_T du jet au vertex primaire (VP).

La variable β représente la fraction des candidats chargés du PF associés au VP sur tous les candidats chargés. Pour rappel, la définition d'association au VP implique une distance entre la trace et le VP de $d_Z < 0.2cm$.

$$\beta = \frac{\sum_{VP} p_T^{cand}}{\sum_{tot} p_T^{cand}} \quad (5.4)$$

La variable β^* représente la fraction des candidats chargés du PF associés aux vertex du PU sur tous les candidats chargés. $\beta^* = 1 - \beta$. C'est β^* la variable dominante de la variable combinée PUJetID.

Enfin, la multiplicité de vertex N_V est utilisée par le BDT pour définir un seuil au delà duquel les variables de forme deviennent plus discriminantes que les variables de vertex.

Variables de forme Pour les variables de forme la démarche est un peu similaire à celle décrite pour la méthode jet area (cf. section 5.4.1.1). Ces variables caractérisent l'étalement des jets dans le plan transverse afin de discriminer les jets du pileup caractérisés par un étalement plus important.

La variable la plus pertinente est le $\langle \Delta R^2 \rangle$ qui représente la moyenne du ΔR du jet pondéré par les p_T^2 , soit :

$$\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R^2 p_T^2(i)}{\sum_i p_T^2(i)} \quad (5.5)$$

La somme correspond à toutes les particules PF constituant le jet.

D'autres variables viennent préciser l'action de $\langle \Delta R^2 \rangle$, comme par exemple $p_T^A(\Delta R) = \frac{1}{p_T^{jet}} \sum_{\Delta R} p_T^i$. Cette dernière variable considère la moyenne des p_T individuels dans 5 cônes différents autour de l'axe du Jet. Enfin, la variable p_T^D est définie par $p_T^D = \frac{\sum_i p_T^2}{\sum_{p_T}}$ et concerne la discrimination quark gluon. Une étude détaillée de l'action de ces variables sur les données de 2012 vient d'être soumise à la collaboration CMS dans [46] mais cet article n'est pas encore approuvé.

5.4.3.3 Étude de la coupure de sélection PuJetID >-0.89

Dans cette section, nous présenterons quelques distributions rendant compte de l'effet de la coupure PuJetID >-0.89 sur notre échantillon. Les données représentées sont issues de notre TTree et les procédures de soustraction du pileup au niveau JEC et CHS ont été appliquées. Notons également que les corrections pour le nombre de vertex et les efficacités sont déjà appliquées au MC (cf. plus loin, chapitre 6).

Pour chaque distribution nous avons représenté les jets sélectionnés par la coupure PUJetID >-0.89 mais aussi ceux non sélectionnés, et ce pour les données réelles et les données MC. Nous parlerons de façon simplifiée de jets associés au boson Z ou au pileup. Pour chaque distribution nous avons également représenté le rapport Pileup/Total pour les données réelles. Les coupures en $|\eta|$ effectuées pour notre échantillon est $|\eta| < 2.4$. Ici nous considérerons également des domaines en $|\eta| < 7$.

Les figures 5.7 représentent les distributions de l'impulsion transverse du jet de plus haut p_T , pour tous les jets de l'échantillon. La fraction des jets identifiés comme appartenant au pileup y est représentée : elle décroît beaucoup plus vite que la distribution des jets associés au boson Z. Les jets du pileup (selon PUJetID) sont donc en majorité des jets de bas p_T . A droite, une coupure de 30 GeV est effectuée sur le p_T des jets. Cette coupure entraîne une réduction de la proportion de jets du pileup. Typiquement pour la proportion du pileup à 30 GeV est inférieure à 8% du total (5% à 40 GeV). Pour ces distributions le MC est en assez bonne adéquation avec les données.

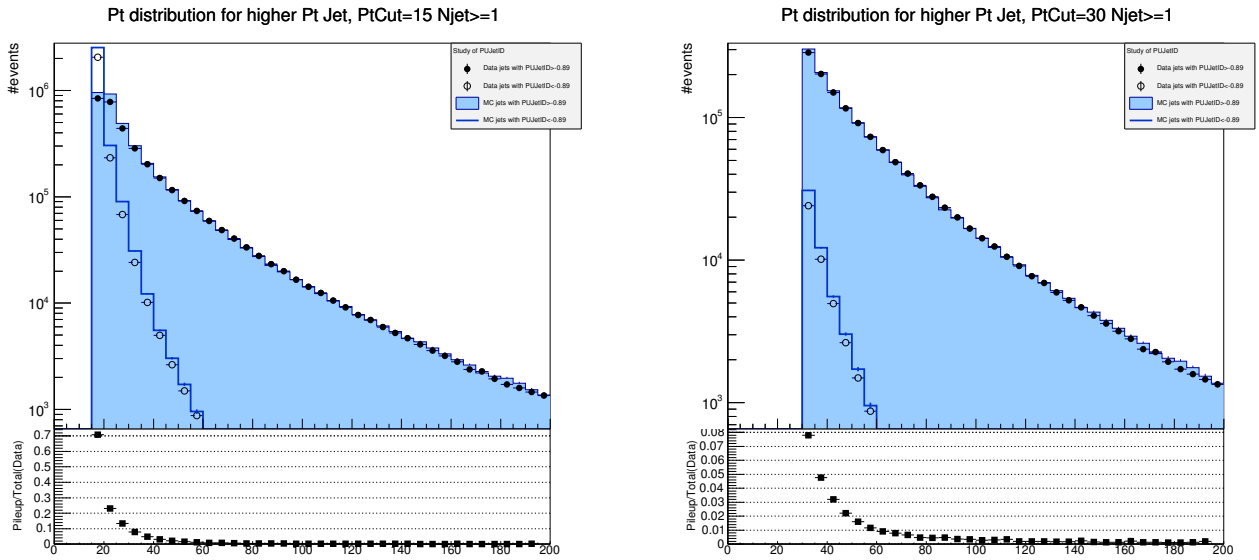


FIGURE 5.7: Distributions en p_T du jet de plus haut p_T associé au boson Z et ou pileup avec $N_{jets} \geq 1$. $p_T > 15$ GeV à gauche, $p_T > 30$ GeV à droite.

Pour les figures représentées en 5.8 nous avons levé la coupure en η ($|\eta| < 7$) et nous considérons l'effet d'une coupure en p_T en considérant des jets de p_T 15 GeV à gauche et 30 GeV à droite. Pour la figure de gauche, la première observation est le phénomène de

"pointes". Ces pointes sont observables sur les données mais sont beaucoup plus fortes sur le MC qui les surestime de presque le double de la valeur des données, et ceci autant pour les jets associé au pileup qu'au Z. Sur la figure de droite la coupure en p_T à 30 GeV diminue considérablement ces effets de pointes (Le MC les surestime toujours mais seulement par un facteur d'environ 2/3). Mais surtout la coupure en p_T diminue considérablement le pileup dont la distribution est devenue parfaitement indépendante en η .

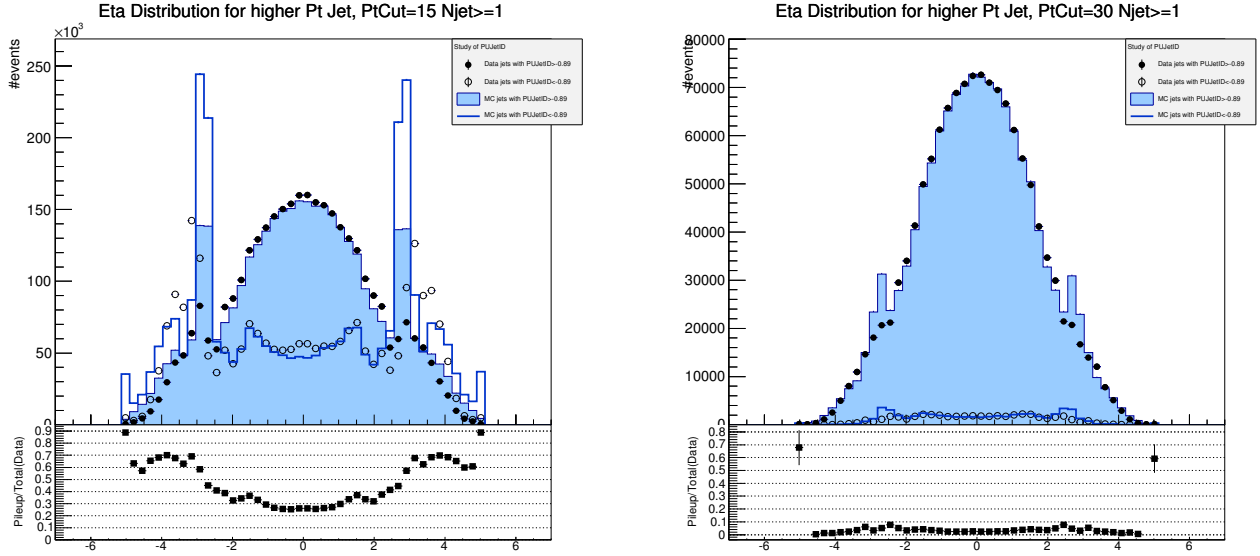


FIGURE 5.8: Distributions en η du jet de plus haut p_T associé au boson Z et au pileup avec $N_{jets} \geq 1$. $p_T > 15$ GeV à gauche, $p_T > 30$ GeV à droite.

Sur les figures représentées en 5.9 on demande que l'impulsion transverse du Z soit inférieure à 10 GeV (à droite) ou 5 GeV (à gauche). On contraint ainsi les jets associés au boson Z à être plus longitudinaux. On observe que la fraction des jets issus du pileup sélectionnée avec PUJetID devient effectivement plus importante. L'effet de pointes est également augmenté. On observe clairement deux zones selon η : une zone centrale ($|\eta| < 2.4$) où la distribution varie peu selon η , mais surtout où les données MC et réelles correspondent assez bien (avec de plus grands écarts sur les bords pour le pileup). Et les zones $\eta > 2.4$ et $\eta < -2.4$ pour lesquelles le MC simule très mal les données et notamment le pileup.

Cette étude précise le domaine cinématique choisit pour notre analyse ($|\eta| < 2.4$ et $p_T > 30$ GeV) : dans ce domaine, le MC représente bien les données et l'échantillon de jet associés au pileup (sélectionnée par la variable PuJetID) est bien uniforme en η et représente une très faible fraction du total des jets ($< 10\%$).

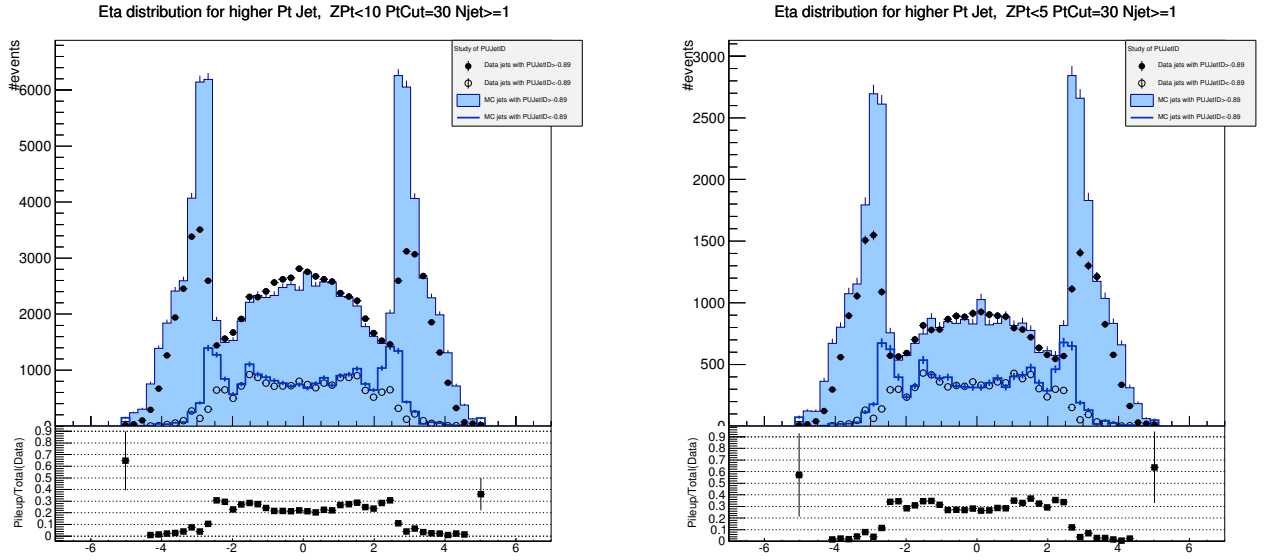


FIGURE 5.9: Distributions en η du jet de plus haut p_T , et tels que $p_{T,Z} < 10$ GeV (gauche) et $p_{T,Z} < 5$ GeV (droite) ($N_{jets} \geq 1$)

Chapitre 6

Traitement de données

Notre échantillon a été sélectionné, le pileup est identifié et soustrait. La prochaine étape consiste à traiter le bruit de fond, mais avant cela il faut vérifier la bonne description des effets expérimentaux par le MC.

Les mêmes procédures de reconstruction, de sélection et de calibration ont été appliquées aux fichiers MC et aux fichiers de données. Ces procédures ont chacune leur efficacité propre qui sera prise en compte de façon globale par la procédure d'Unfolding pour corriger les effets de perte d'événements correspondants. Cependant les efficacités concernant les fichiers MC ne sont pas équivalentes à celles des fichiers de données. Avant de pouvoir soustraire le bruit de fond, il faut procéder à un ajustement entre MC et données.

Dans ce chapitre nous expliquerons de manière générale les facteurs de correction appliqués au MC. Une première correction concerne la modélisation du nombre de vertex par événement, les autres concernent les efficacités de sélection concernant les variables d'identification et d'isolement des muons. Nous présenterons les facteurs de corrections que nous avons utilisé. Dans une seconde partie nous expliquerons la procédure de soustraction du bruit de fond.

6.1 Ajustement entre MC et données

6.1.1 Correction pour le nombre de vertex

Notre mesure du pileup se fait par le comptage du nombre de vertex reconstruits (des méthodes plus sophistiquées sont parfois utilisés dans CMS). Les échantillons MC ont été générés avec un pileup proche de ce que l'on a pu observer dans les données. Mais ils surestiment la fraction de luminosité produit à haut pileup. La figure 6.1 de gauche montre le décalage entre MC et données qui varie fortement selon le nombre de vertex. Pour les événements ayant un nombre de vertex entre 1 et 26 le taux relatif Data/MC varie entre 0.6 et 1.2, mais partir de 27 vertex ce rapport décroît fortement. Le MC surestime le nombre d'événements par rapport aux données et ce jusqu'à 100 fois pour les événements possédant plus de 40 vertex. Dans ce travail nous avons re-pondéré le MC bin par bin par le rapport data/MC présenté sur la figure 6.1 de gauche (le MC est ensuite ajusté à la luminosité des données). Après correction le rapport donne 1 par construction

comme représenté sur la figure 6.1 de droite. Pour un nombre de vertex supérieur à 49 nous pondérons le MC par un facteur égal à 0.

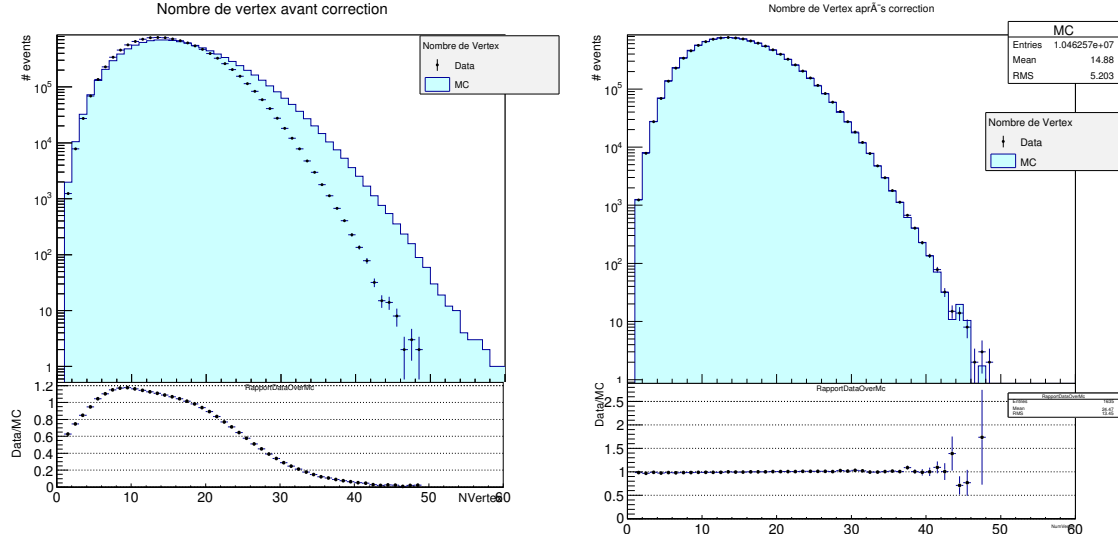


FIGURE 6.1: Distribution du nombre de vertex pour le MC et les données. À gauche, établissement du rapport qui est utilisé pour la correction, à droite test de fermeture après correction du MC par rapport aux données.

La correction pour le nombre de vertex est la première correction que nous appliquons au MC.

6.1.2 Efficacités

De manière générale une certaine efficacité est associée à une sélection appliquée à un échantillon. Cette efficacité rend compte du taux de perte de bons événements par l'application de cette coupure. Par exemple, il existe une petite probabilité qu'une paire de muons provenant de la désintégration d'un Z n'ai pas déclenché le trigger : l'efficacité de l'application du critère de trigger est donc légèrement inférieure à 1. Pour la sélection, il existe un faible pourcentage de muons mal sélectionnés soit parce qu'ils ont finalement été mal identifiés ou mal isolés.

On mesure ces efficacités en établissant un rapport entre les objets ayant passé la sélection et les objets de l'échantillon de départ. La méthode que nous avons utilisée pour calculer ce rapport est le Tag and Probe (cibler et sonder) pour une paire de muons sur le pic du Z .

Dans notre travail, les trois contributions à l'efficacité globale que nous considérons sont :

- l'efficacité de trigger ϵ_{trig}
- l'efficacité d'identification des muons ϵ_{ID}
- l'efficacité d'isolation des muons ϵ_{Iso}

Ces efficacités ne sont pas corrélées. L'efficacité de trigger concerne la paire de muons et est calculée selon leur valeur en η , les efficacités d'identification et d'isolation concernent

chacun des muons et sont calculées en fonction de leur domaine cinématique. L'efficacité totale calculée est donc :

$$\epsilon_{event} = \epsilon_{trig}(\eta_1, \eta_2) \cdot \epsilon_{ID}(\eta_1, p_{T,1}) \cdot \epsilon_{Iso}(\eta_1, p_{T,1}) \cdot \epsilon_{ID}(\eta_2, p_{T,2}) \cdot \epsilon_{Iso}(\eta_2, p_{T,2}) \quad (6.1)$$

Nous avons besoin de déterminer ces efficacités à partir des données et de les comparer à celles des fichiers MC afin de calculer la correction d'ajustement (SF, le "scale factor") :

$$SF_{MC} = \frac{\epsilon_{data}}{\epsilon_{MC}} \quad (6.2)$$

Les sélections concernant les muons et électrons sont standards et gérées par le POG, les efficacités sont également calculées par le POG avec la méthode Tag and Probe [47]. L'interface présente différentes recommandations selon les analyses et fournit les efficacités et les facteurs de correction ainsi que leur erreurs systématiques. Pour le trigger, nous avons utilisé les facteurs de correction SF_{MC}^{trig} délivrés par le POG [48] et dont les valeurs sont reprises dans le tableau 6.1. Pour l'identification et l'isolation, nous avons calculé les facteurs de correction de manière indépendante.

$\leq \eta_1 \leq$		$\leq \eta_2 \leq$		Trigger correction
0.0	0.9	0.0	0.9	0.975 ± 0.002
0.0	0.9	0.9	1.2	0.975 ± 0.002
0.0	0.9	1.2	2.1	0.972 ± 0.002
0.0	0.9	2.1	2.4	0.973 ± 0.002
0.9	1.2	0.0	0.9	0.977 ± 0.002
0.9	1.2	0.9	1.2	0.980 ± 0.002
0.9	1.2	1.2	2.1	0.974 ± 0.002
0.9	1.2	2.1	2.4	0.979 ± 0.002
1.2	2.1	0.0	0.9	0.970 ± 0.002
1.2	2.1	0.9	1.2	0.971 ± 0.002
1.2	2.1	1.2	2.1	0.967 ± 0.002
1.2	2.1	2.1	2.4	0.970 ± 0.002
2.1	2.4	0.0	0.9	0.970 ± 0.002
2.1	2.4	0.9	1.2	0.975 ± 0.002
2.1	2.4	1.2	2.1	0.950 ± 0.002
2.1	2.4	2.1	2.4	0.974 ± 0.002

TABLE 6.1: Corrections pour l'efficacité de trigger des muons selon les différents domaines en η de chacun des muons. Les erreurs représentées correspondent au 0.2% conseillés par le POG [47].

Dans un premier temps, nous expliquerons le principe de la méthode Tag and Probe puis nous présenterons les facteurs de correction calculés ainsi que les distributions des efficacités d'identification et d'isolation pour les données et le MC.

6.1.2.1 Méthode Tag and Probe

Le principe de la méthode Tag and Probe est assez simple : à partir d'un échantillon, on cible ("tag") une partie des objets en leur imposant un certain nombre de critères très sévères. Le taux de mauvaise identification de ces critères doit être très bas ($< 1\%$).

D'autres objets de l'échantillon vont servir à sonder les critères appliqués aux tags. On les appelle les probes. Les probes sont des objets similaires au tags mais sélectionnés de manière plus souple : ce sont uniquement les critères pour lesquels on mesure l'efficacité qui différencient les tags des probes. Dans notre cas on demande que la paire "tag and probe" soit deux muons dont la masse invariante se trouve sur le pic du Z.

L'efficacité d'un critère de sélection est ensuite calculée en comptant le nombre de probes qui passent cette sélection, soit :

$$\epsilon = \frac{N_{pass}}{N_{all}} = \frac{N_{pass}}{N_{pass} + N_{fail}} \quad (6.3)$$

N_{pass} est le nombre de probes qui ont passé la sélection, N_{all} est le nombre initial de probes, N_{fail} le nombre de probes n'ayant pas passé la sélection.

Nous allons présenter la mesure de l'efficacité de reconstruction des muons obtenue par simple comptage des quantités N_{pass} , N_{fail} , N_{all} définies dans la formule 6.3 en ne soustrayant pas le bruit de fond. Dans le cas des muons, le bruit de fond est largement dominé par des processus produisant des muons isolés et identifiés : ce bruit de fond n'altère donc pas le calcul des efficacités.

Dans certains cas, la présence de bruit de fond se révèle trop importante et nécessite l'application de coupures sur le probe ce qui limite la mesure car l'efficacité est alors mesurée sur une sous-partie. Une technique d'ajustement modélisant les distributions pour le signal et le bruit de fond peut alors être utilisée afin d'en extraire la composante lors de la mesure de l'efficacité [49].

6.1.2.2 Efficacités d'identification et d'isolation

Pour les variables d'identification et d'isolation des muons, nous avons appliqué la méthode Tag and Probe à notre échantillon (trigger asymétrique HLT_Mu17_Mu8_v*) comprenant au moins 2 muons du PF de charges opposées et de masse invariante comprise entre 61 GeV et 121 GeV. Nous avons appliqué les coupures cinématiques $|\eta| < 2.4$, $p_T > 20$ GeV à chaque muon.

Nous avons assigné les muons de plus haut p_T aux tags. Les deux critères d'identification et d'isolation leurs ont été appliqués, soit : PFID = Tight et PFIso $\Delta\beta < 0.2$

Les seconds muons de plus haut p_T sont les probes. Nous leur avons appliqué dans un premier temps la sélection d'identification. Les efficacités d'identification ont alors été établies pour l'entièreté du domaine cinématique comme représenté à la figure 6.2. A partir de ces distributions, nous avons déterminé différents domaines pour l'établissement des facteurs de correction d'identification (cf. tableau 6.2). Nous avons appliqué les facteurs de correction d'identification avant de passer à une procédure similaire pour la variable d'isolation.

Pour la variable d'isolation, les distributions et les efficacités sont représentées sur la figure 6.3 et les facteurs de correction sont repris dans le tableau 6.3. Sur les graphiques du haut sont représentées les différentes distributions des second muons (de plus haut

p_T dans les échantillons seulement taggés (MC all, Data all), dans les échantillons des muons probes (MC pass, Data pass), pour les données réelles ainsi que pour les données MC. Les graphiques du bas représentent les distributions des efficacités d'identification et d'isolation en fonction des variables η et p_T avec leur erreurs asymétriques. Nous allons brièvement commenter les graphiques du bas concernant des efficacités :

- De manière générale, les efficacités sont proches de 1 et données réelles et MC sont en bon accord.
- Pour l'identification : l'efficacité varie très peu en fonction de p_T : ce sont donc les domaines en η qui ont compté pour le calcul des facteurs de correction (nous avons quand même considéré différents domaines en p_T afin de gagner en précision).
- Pour l'isolation : ici l'efficacité n'est pas indépendante de p_T . Pour des $p_T < 45 GeV$, l'on perd presque 10% des données : ceci est dû au fait que la variable d'isolation ait une dépendance en $1/p_T^{muon}$. Si le muon a une petite impulsion transverse $PF_{Iso}\Delta\beta$ devient grand et ne passe plus la coupure < 0.2 .

Les figures 6.2 et 6.3 nous ont permis d'établir les domaines de calcul du SF, en p_T et en η . Pour des $p_T > 60 GeV$ nous avons effectué un fit linéaire sur les distributions des efficacités.

6.1.2.3 Établissement des erreurs sur les efficacités

Calculer l'erreur statistique sur une efficacité n'est pas trivial : l'efficacité est comprise entre 0 et 1 et l'erreur devient donc asymétrique pour les efficacités proches des limites, comme représenté sur les résultats (cf. figure 6.2). Ceci est dû au fait que les populations N_{all} et N_{fail} sont statistiquement fortement corrélées, N_{fail} constituant une sous-partie de N_{all} si $\varepsilon \simeq 1$.

En général on détermine les erreurs de distribution de type $(A/A + B)$ avec la loi binomiale ou la loi de Poisson. Mais dans le cas des efficacités les valeurs sont limitées, très proches de 1, et les lois pré-citées présentent des problèmes de comportement comme expliqué dans l'étude [50]. L'erreur sur les efficacités affichée sur les figures est établie en utilisant le théorème de Bayes et avec un intervalle de confiance de 0.68 %. Nous présenterons le théorème de Bayes au chapitre 7 pour la procédure d'Unfolding.

Pour la détermination des erreurs sur les facteurs de correction ($= \frac{\varepsilon_{data}}{\varepsilon_{MC}}$), la procédure pour combiner les erreurs asymétriques n'est pas établie, nous avons donc utilisé les erreurs systématiques conseillées par le POG soit 0.2% pour le trigger, 0.5% pour l'identification, 0.2% pour l'isolation.

6.2 Soustraction du bruit de fond

Les facteurs de correction appliqués au MC permettent de le rendre cohérent par rapport aux données. Une fois les fichiers MC pondérés par les corrections pour le nombre de vertex et pour les efficacités, on peut les utiliser pour soustraire le bruit de fond des données.

$p_T[GeV] \setminus \eta $	0-0.9	0.9-1.2	1.2-1.8	1.8-2.4
20-25	0.993±0.004	0.987±0.0081	1.002±0.0061	0.995±0.006
25-30	0.992±0.003	0.987±0.007	1.003±0.005	0.993±0.005
30-35	0.991±0.003	0.987±0.006	1.002±0.005	0.993±0.005
35-40	0.993±0.002	0.987±0.005	1.003±0.004	0.994±0.004
40-45	0.994±0.003	0.989±0.005	1.003±0.004	0.994±0.004
45-50	0.993±0.006	0.987±0.011	1.003±0.008	1.003±0.008
50-55	0.994±0.015	0.987±0.0268	0.997±0.0210	0.987±0.025
55-60	0.991±0.0244	0.982±0.0439	1.002±0.035	0.989±0.043
> 60	0.999 ± 0.020	1.001 ± 0.038	1.003 ± 0.031	1.009 ± 0.047

TABLE 6.2: Facteurs de corrections pour l'efficacité d'identification des muons. Le facteur représente le rapport $\frac{\epsilon_{data}}{\epsilon_{MC}}$. Les erreurs représentées sont statistiques et calculées par root.

$p_T[GeV] \setminus \eta $	0-0.9	0.9-1.2	1.2-1.8	1.8-2.2	2.2-2.4
20-25	0.988±0.004	0.997±0.008	0.998±0.006	1.011 ± 0.007	1.06±0.012
25-30	0.992±0.003	0.998±0.007	1.000±0.005	1.010 ± 0.007	1.051±0.011
30-35	0.996±0.003	0.999±0.006	1.002±0.004	1.010 ± 0.007	1.036±0.010
35-40	0.998±0.003	0.999±0.005	1.002±0.004	1.007 ± 0.006	1.028±0.009
40-45	0.998±0.006	1.0000±0.005	1.000±0.003	1.004 ± 0.005	1.018±0.009
45-50	0.999±0.015	0.999±0.011	1.001±0.008	1.002 ± 0.011	1.007±0.020
50-55	1.002±0.025	0.994±0.027	1.002±0.021	1.003 ± 0.029	1.007±0.050
55-60	1.002±0.025	1.000±0.044	1.003±0.035	1.005 ± 0.050	1.002±0.087
> 60	0.999 ± 0.0207	1.001± 0.0387	1.004 ± 0.0312	1.009 ± 0.047	0.984 ± 0.085

TABLE 6.3: Facteurs de corrections pour l'efficacité d'isolation des muons. Le facteur représente le rapport $\frac{\epsilon_{data}}{\epsilon_{MC}}$. Les erreurs représentées sont statistiques et calculées par root.

6.2.1 Les différentes contributions du bruit de fond

Les différents processus constituant un bruit de fond pour la production d'un paire de muons sur le pic du Z que nous considéreront sont les suivant :

- $t\bar{t} \rightarrow \mu^+\mu^-$ par désintégration faible en quark b,
- $ZZ \rightarrow \mu^+\mu^- + q\bar{q}$
- $ZZ \rightarrow \mu^+\mu^- + \nu_\mu\nu_\mu$
- $ZW \rightarrow \mu^+\mu^- + q_i\bar{q}_j$
- $ZW \rightarrow \mu^+\mu^- + l\nu_l$
- $WW \rightarrow \mu\nu_\mu + \mu\nu_\mu$

Dans notre analyse, nous n'utiliserons pas de fichiers simulant les processus QCD impliquant les quarks légers ni le fichier impliquant une paire de leptons $\tau^+\tau^-$. Ce bruit de fond est tout à fait négligeable selon les analyses menées à 7 TeV et à 8 TeV [51].

Ces différents processus se différencient selon qu'ils présentent une résonance (WZ, ZZ) ou non ($t\bar{t}$, WW, DY τ , τ au QCD). Et selon qu'ils présentent des états finaux comprenant des neutrinos ou non. Les contributions présentant des résonances sont bien mises en évidence sur la figure 6.7 représentant la masse invariante du Z en échelle logarithmique.

Les différentes contributions vont intervenir différemment selon les domaines cinématiques et selon les distributions mais nous verrons de façon générale leur contribution ne dépasse pas le pour-cent. Le détail de ces contributions est visible sur les distributions de contrôle présentées en 6.3.

De manière générale, l'on observe une augmentation de la proportion de bruit de fond quand on augmente en multiplicité de jet. Le processus constituant la majorité de notre bruit de fond est le processus impliquant une paire $t\bar{t}$.

Le bruit de fond est soustrait des fichiers de données en retranchant bin par bin les différentes contributions MC (pondérées à la luminosité des données) des distributions de données.

6.3 Distributions de contrôle

La contribution du bruit de fond est représentée sur les distributions de contrôles ci-dessous. La cohérence entre données et fichiers MC est également représentée par le rapport entre données et MC total.

6.3.0.1 Multiplicité de jets

Les distributions représentées en 6.4 représentent le décompte des jets associés aux événements : ce sont les distributions de multiplicité de jets. La multiplicité "exclusive" considère dans chaque bin les événements associés à un nombre exacte de jet. La multiplicité "inclusive" considère dans chaque bin tous les événements dont le nombre de jets est supérieurs ou égal à un certain nombre de jets.

Pour ces distributions, l'adéquation entre MC et données est très bon jusqu'à 4 jets. La contribution du bruit de fond est principalement due au $t\bar{t}$ et représente au maximum une fraction 10^{-3} du données. On remarque une certaine contribution présentant des neutrinos (en vert) jusqu'à 4 jets.

6.3.0.2 Distributions en l'impulsion transverse des jets

En 6.5 sont représentées les distributions de l'impulsion transverse des différents jets. Pour rappel, nous considérons les jets par classement en p_T . Les distributions sont inclusives : on demande que l'événement contienne au moins N jets mais pour chaque événement l'on représente la variable p_T du jet de plus haut p_T .

La proportion du bruit de fond augmente à mesure qu'on augmente le nombre de jet.

6.3.0.3 Distributions en l'impulsion transverse et en η pour le Z

Enfin, en 6.6 sont représentées les distributions de l'impulsion transverse et de la variable η du Z. On remarque que la contribution du bruit de fond est beaucoup moins importante que pour la distribution en p_T du premier jet par exemple.

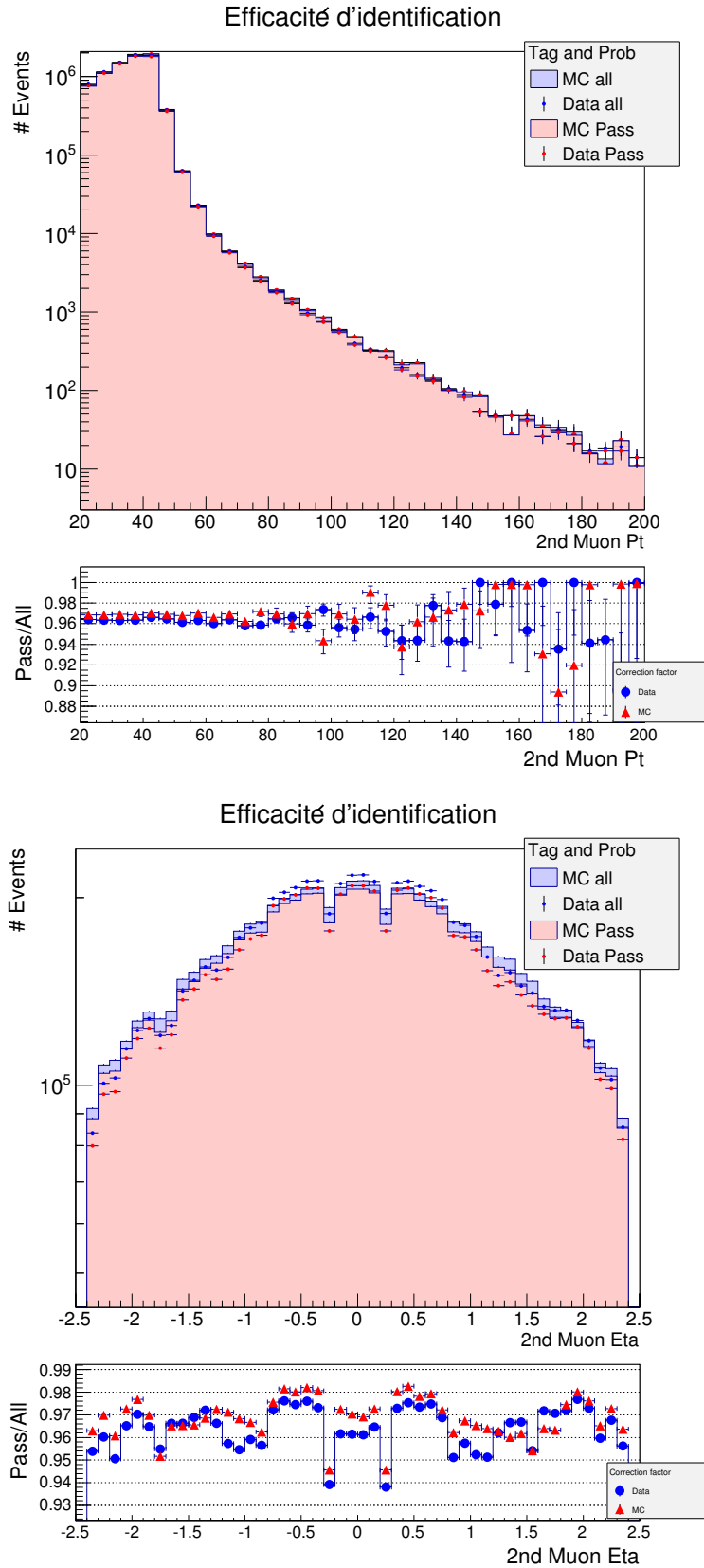


FIGURE 6.2: Distributions pour l'identification du second muon de plus haut p_T et pour l'efficacité d'identification selon les variables p_T (haut) et η (bas).

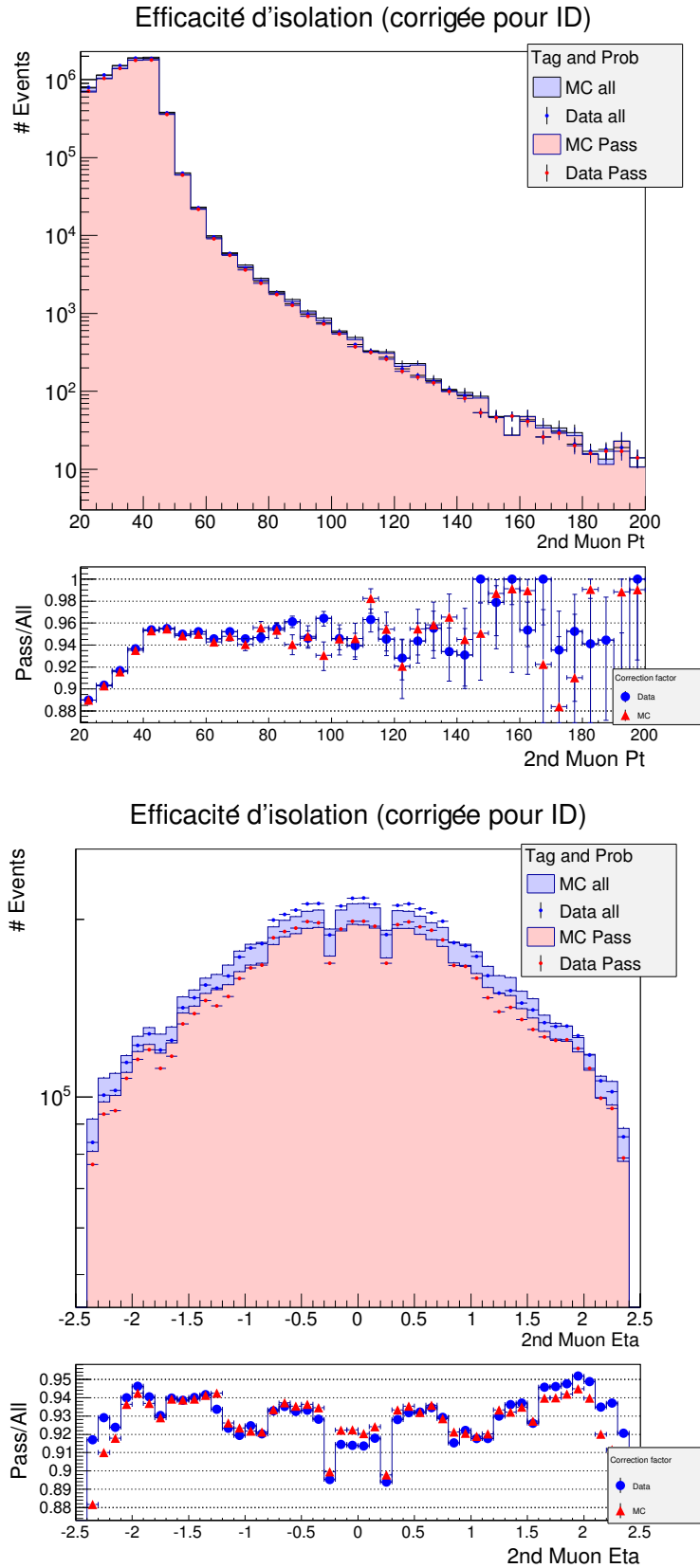


FIGURE 6.3: Distributions pour l'isolation du second muon de plus haut p_T et pour l'efficacité d'isolation selon les variables p_T (haut) et η (bas).

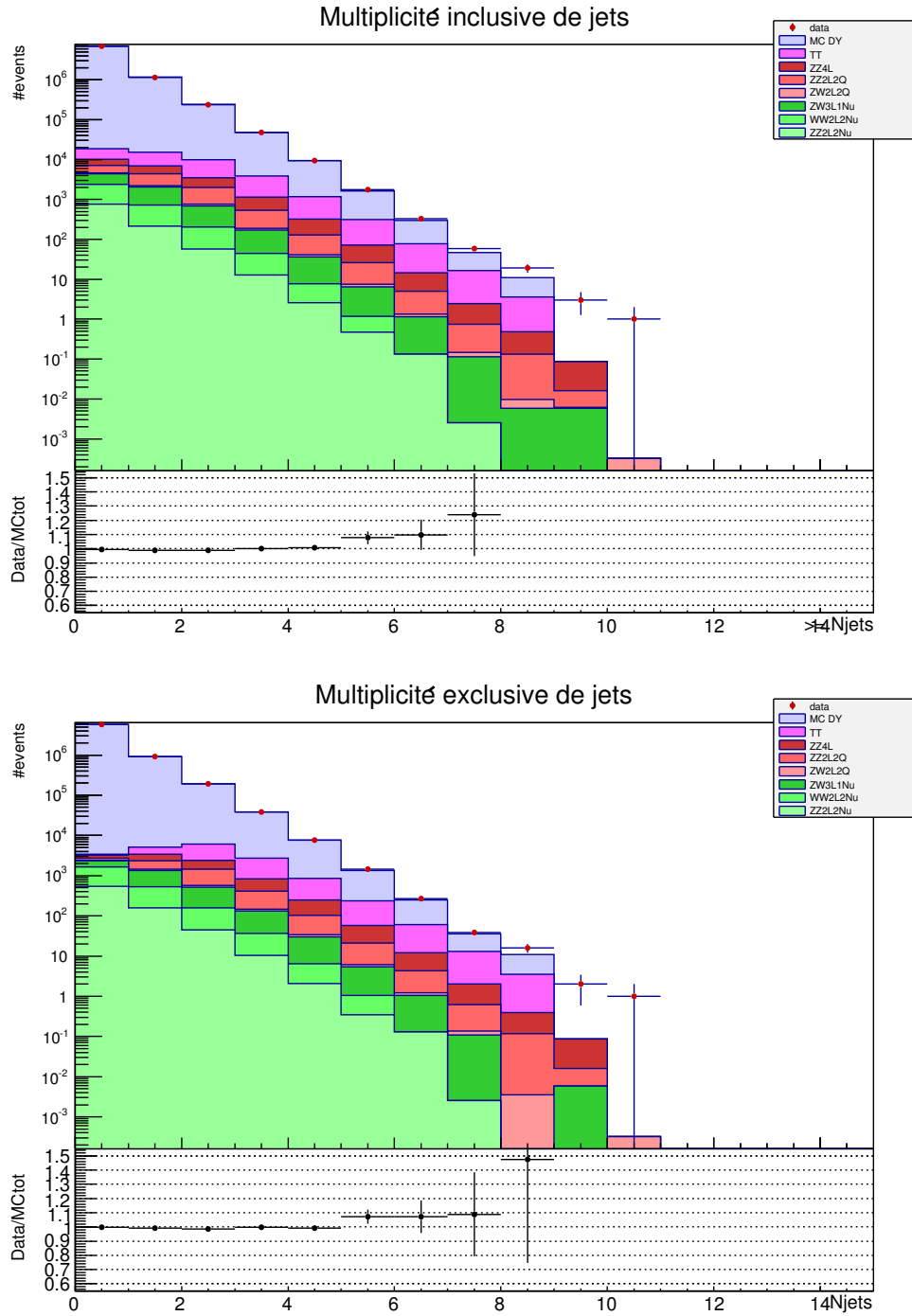


FIGURE 6.4: Distributions de contrôles pour la multiplicité de jets inclusive (haut) et exclusive (bas).

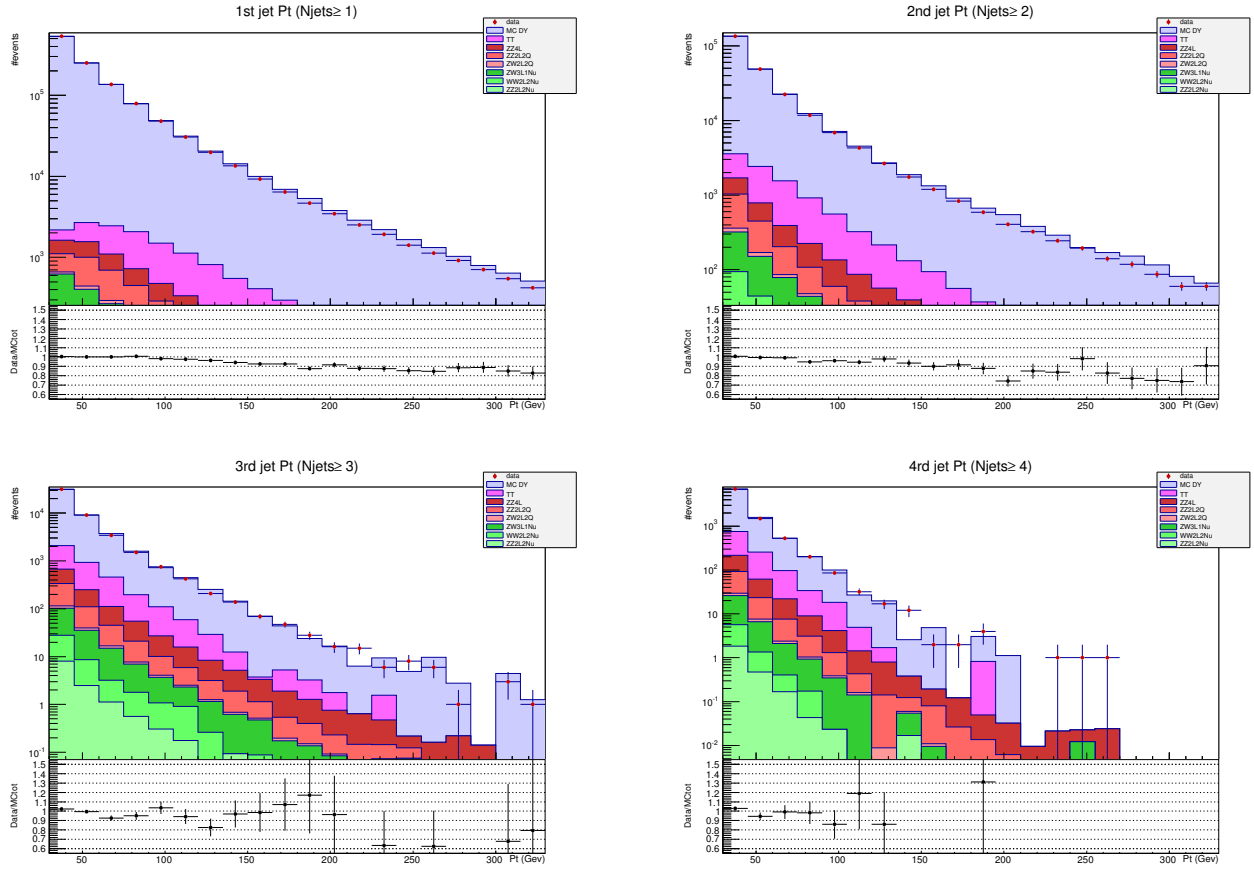


FIGURE 6.5: Distributions de contrôle pour l'impulsion transverse des jets associés au Z. Ces distributions sont inclusives.

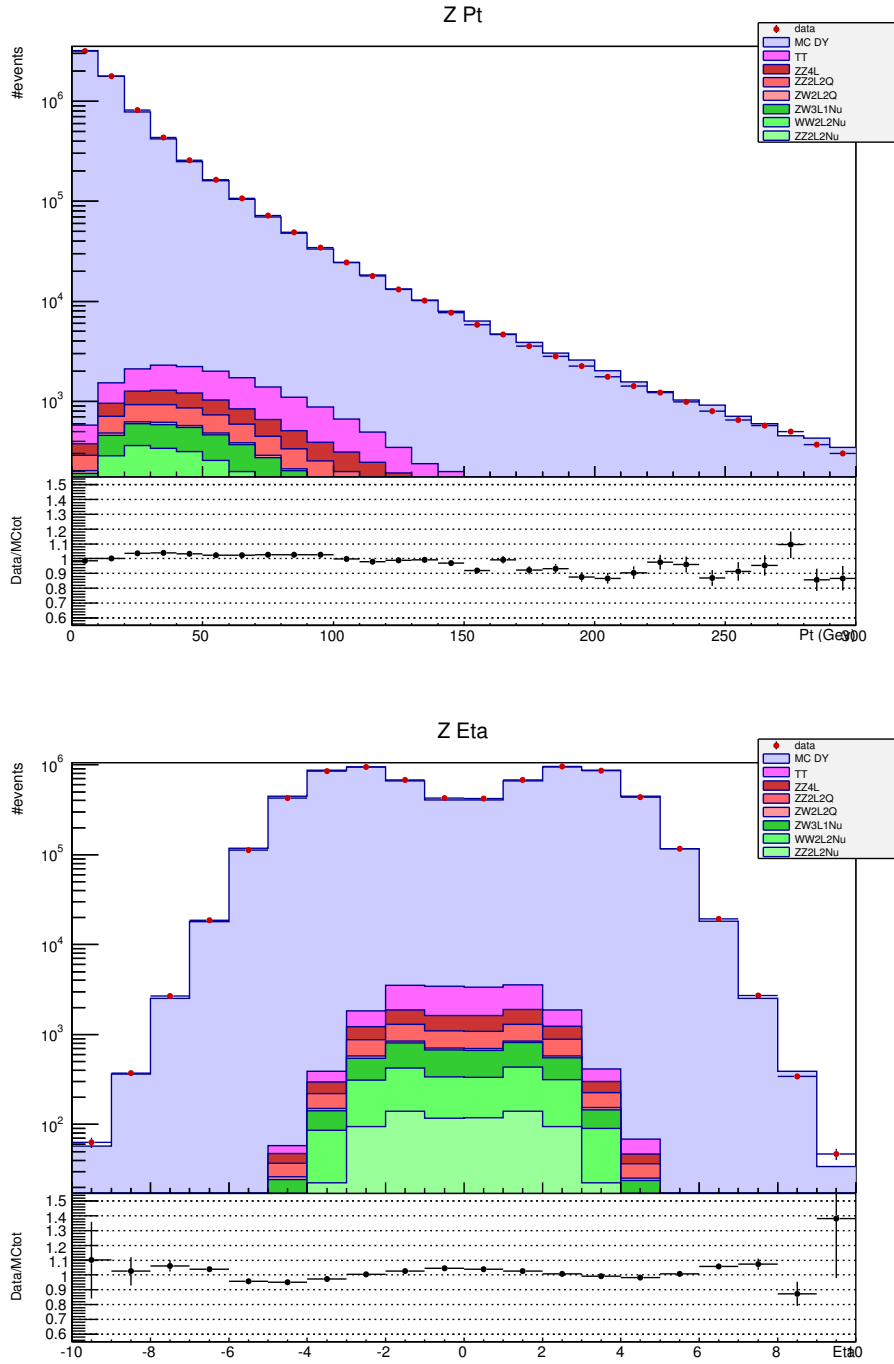


FIGURE 6.6: Distributions de contrôle pour l'impulsion transverse (haut) et la variable η (bas) du boson Z reconstruit.

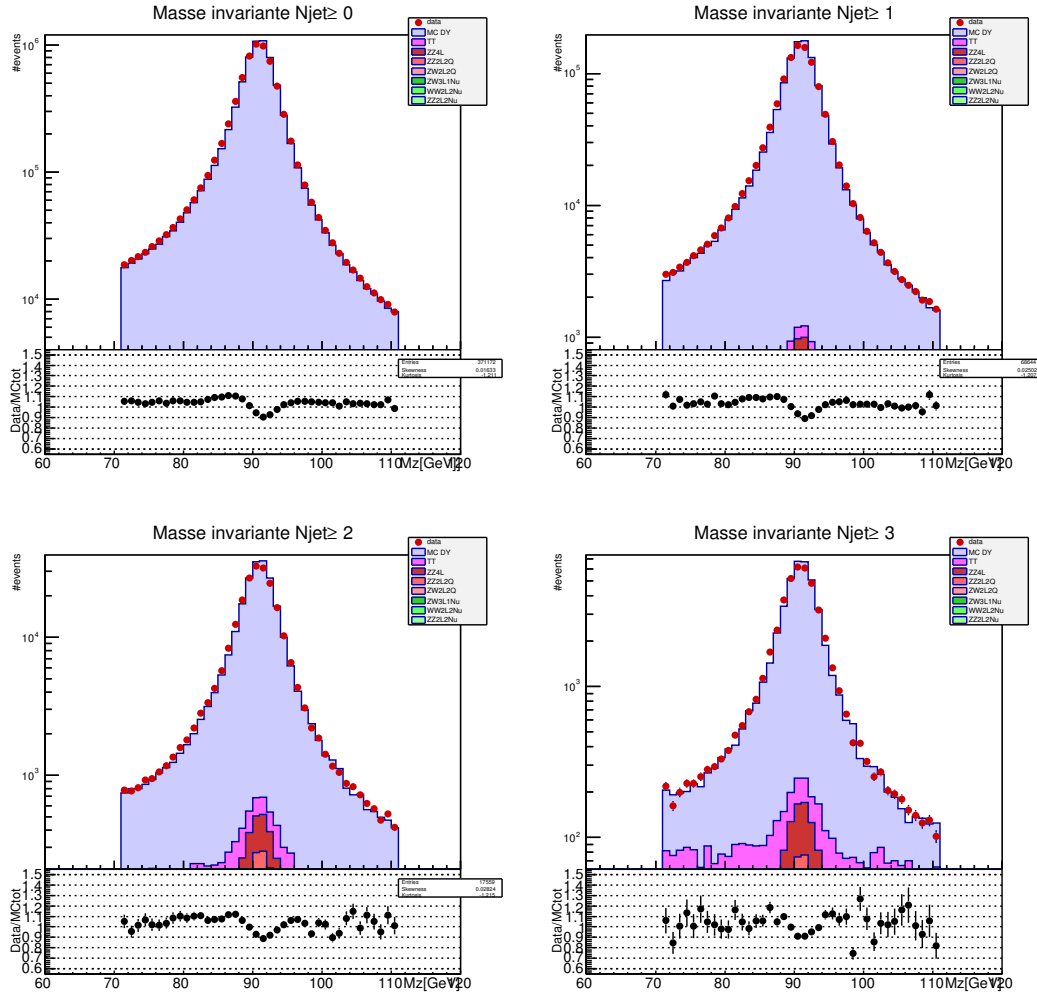


FIGURE 6.7: Distribution de contrôle pour la masse invariante du Z. Sont représentés les échantillons comprenant 0 jets, au moins 1 jet, au moins 2 jets et au moins 3 jets.

Chapitre 7

Unfolding et résultats

Au stade où nous en sommes de notre analyse, nous possédons un échantillon presque pur d'événements $pp \rightarrow Z + jets$: le pileup et le bruit de fond ont été soustraits et le fichier MC est ajusté aux données pour les corrections du nombre de vertex et pour les efficacités. Les fichiers de données et le MC Drell-Yan sont déjà en bon accord comme le montrent les distributions de contrôle.

Nous voulons extraire la section efficace au niveau hadronique pour pouvoir la comparer à différents modèles théoriques dans le futur, or les distributions établies jusqu'ici représentent des données reconstruites : elles sont dépendantes des réactions du détecteur qui ne sont évidemment pas idéales. La résolution des différents éléments n'est pas parfaite, leur couverture angulaire n'est pas totale, des pixels ou des capteurs peuvent être cassés, et l'électronique de transmission peut également être défaillante... Tout cela constitue ce qu'on appelle "les effets détecteurs".

La seule façon de connaître ce que la nature a produit est de s'affranchir des "effets du détecteur". Pour cela on utilise les fichiers MC sélectionnés et ajustés. Pour ces fichiers on connaît à la fois les processus au niveau des particules (c'est le niveau généré) et au niveau reconstruit, qui représente l'action des effets détecteurs sur les processus générés.

Dans ce chapitre, nous présenterons la procédure d'Unfolding et nos résultats pour les sections efficaces du processus $pp \rightarrow Z + jets$.

7.1 Unfolding

L'Unfolding (le dépliage, pour le traduire littéralement) est motivé par le fait que pour toute expérience les observables mesurées diffèrent des observables présumées. Un des rôles de l'expérimentateur est de démêler les distributions observées de l'effet général dû aux distorsions induites par les détecteurs mais aussi par des effets physiques (radiations QED ou QCD, etc...).

La simulation MC permet cette opération. Pour cela on établit la matrice permettant de passer des données MC générées aux données MC reconstruites et on l'applique ensuite de manière inversée à l'échantillon de données mesurées. Cette matrice est appelée "matrice de réponse".

Appliquer l'Unfolding à une distribution est donc un processus en deux étapes :

- Une première étape consiste à établir la matrice de réponse.

- Une seconde étape consiste à appliquer de manière itérative cette matrice inversée à l'échantillon de données. L'itération est nécessaire à la façon d'une interpolation afin d'arriver au point où l'erreur statistique de la mesure est physiquement significative.

L'enjeu des différents algorithmes d'Unfolding est d'appliquer la matrice de réponse de façon inversée en tenant compte des différentes migrations entre variables générées et reconstruites. Il faut également fixer le bon nombre d'itérations. Nous développerons ces différents objectifs dans le cas pratique de notre étude.

L'outil que nous avons utilisé est RooUnfold [52]. Il s'agit d'une interface développée depuis 2007 à partir des outils de root, par une équipe indépendante travaillant à l'origine sur l'expérience Babar. RooUnfold propose plusieurs algorithmes d'Unfolding : un algorithme développé autour du théorème de Bayes, un algorithme utilisant la Single Value Decomposition et d'autres algorithmes plus simples (bin par bin...). Elle propose également une série de tests pour ces algorithmes ainsi que des calculs d'erreurs systématiques.

Ces différents algorithmes ont été testés dans des études similaires à la nôtre [20] [51].

Dans ce travail, nous avons utilisé la méthode d'Unfolding basée sur le théorème de Bayes. Nous l'avons appliquée à nos quantités mesurées : la multiplicité exclusive des jets, les variables η et p_T du Z produit, les variables p_T et η des jets associés à la production d'un Z (jusqu'à 4 jets).

7.1.1 Établissement de la matrice de réponse

La première étape est la création d'une matrice de réponse cartographiant pour chacune des variables le passage des valeurs générées aux valeurs reconstruites. Soit R cette matrice, on la construit en établissant la distribution relative du MC généré en fonction du MC reconstruit normalisée à la valeur du MC généré .

La matrice de réponse dépend donc du lien entre une valeur générée et une valeur reconstruite.

La matrice de réponse est remplie de la manière suivante : pour un événement donné et une variable donnée, si T_i la valeur générée correspondant au bin (i), et M_j la valeur reconstruite correspondant au bin (j), l'élément de matrice R_{ij} sera la fraction d'événements de T_i mesurées dans M_j . La correspondance entre les jets générés et reconstruits est faite suivant leur ordonnement en impulsion transverse, i.e. on fait correspondre le jet de plus grand p_T du niveau généré au jet de plus grand p_T du niveau reconstruit et ainsi de suite pour les autres jets.

De façon pratique, nous avons établi nos matrices de réponse à partir des fichiers MCDY (cf. tableau 4.2 chapitre 4). Pour un événement donné, si la variable est générée et reconstruite, nous avons rempli la matrice de réponse. Si la variable est générée mais non reconstruite nous avons rempli une distribution appelée "Miss" proposé par l'interface RooUnfold. Si la variable est reconstruite mais non générée nous avons rempli les distributions dites "fake".

L'algorithme d'Unfolding va ensuite utiliser ces différentes distributions afin de déterminer les distributions finales ("unfoldées") pondérées par l'efficacité globale $\frac{N_{Reco}}{N_{gen}}$.

Les matrices de réponse pour certaines de nos variables d'intérêt sont représentées dans les figures 7.1 à 7.4. Ces distributions mettent en évidence les migrations d'un bin à l'autre. Pour la matrice de la multiplicité de jets par exemple (cf. figure 7.1), si l'on

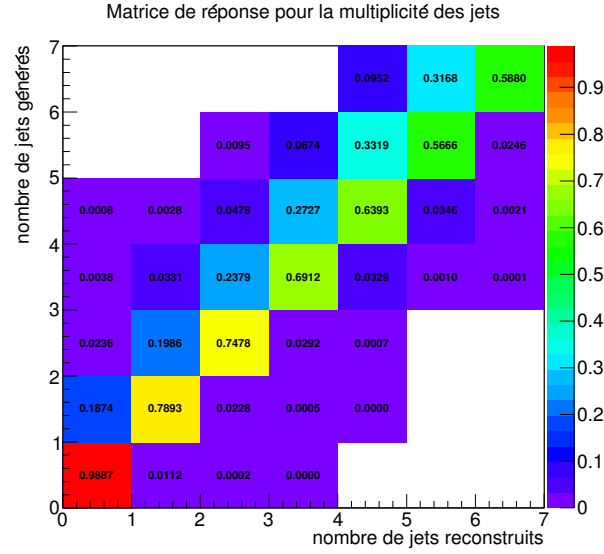


FIGURE 7.1: Matrice de réponse pour la multiplicité exclusive de jets.

considère la ligne pour laquelle le nombre de jets générés vaut 1 alors on peut lire de gauche à droite que 18.8% de ces événements sont reconstruits avec 0 jet, 78.8% avec 1 jet, 2.28% avec 2 jets, 0.04% avec 3 jets et quelques $10^{-4}\%$ avec 4 jets... La somme des valeurs exactes sur une ligne vaut 1 car pour ces distributions les valeurs sont normalisées à la valeur générée.

Selon les variables les migrations sont plus ou moins importantes. Les distributions des variables η et p_T du Z présentent très peu de migrations. Ceci est dû à la bonne résolution des mesures angulaires et de l'impulsion du Z. L'algorithme d'Unfolding doit tenir compte de ces migrations quelque soit leur ordre de grandeur.

Choix du binning :

Pour établir les matrices de réponse nous avons établi le binning des distributions de manière à conserver une majorité des événements provenant des bins diagonaux (qui représentent les bins pour lequel les valeurs générées et reconstruites sont équivalentes). C'est important car si, par exemple, seulement 40% est représenté dans un élément diagonale, cela signifie que la majorité des événements observés dans ce bin ne vient pas d'éléments générés dans ce bin et la mesure de la section efficace dans ce bin serait donc basée sur une majorité d'événements n'appartenant pas à ce bin. Pour les distributions en la variable p_T des jets, cela a conduit au binning variable que l'on peut observer sur les matrices de réponse : il a été établi en essayant de conserver un pourcentage minimum de 60% sur la diagonale. Ce binning variable permet aussi de garder une bonne statistique pour chaque valeur des impulsion transverse des jets, ce qui est important pour la procédure d'Unfolding avec le théorème de Bayes.

Limites de distributions :

Pour les bins correspondant aux limites des distributions l'on doit quand même considérer les migrations sur les bins adjacentes même si elles ne sont pas représentées.

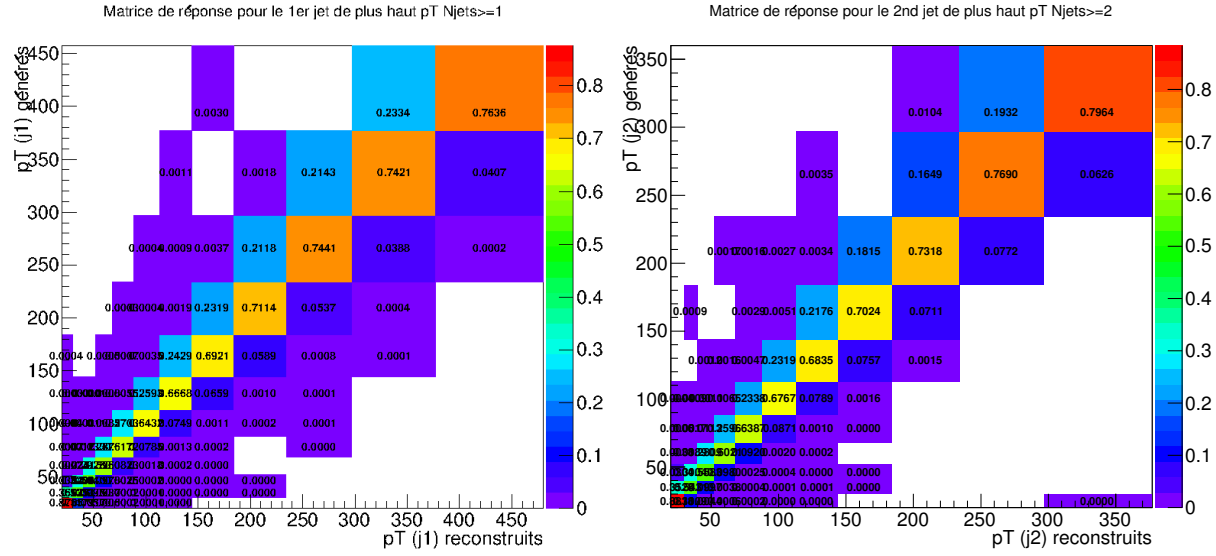


FIGURE 7.2: Matrices de réponses pour l'impulsion transverse du premier jet de plus haut p_T (à gauche) et du second jet de plus haut p_T (à droite).

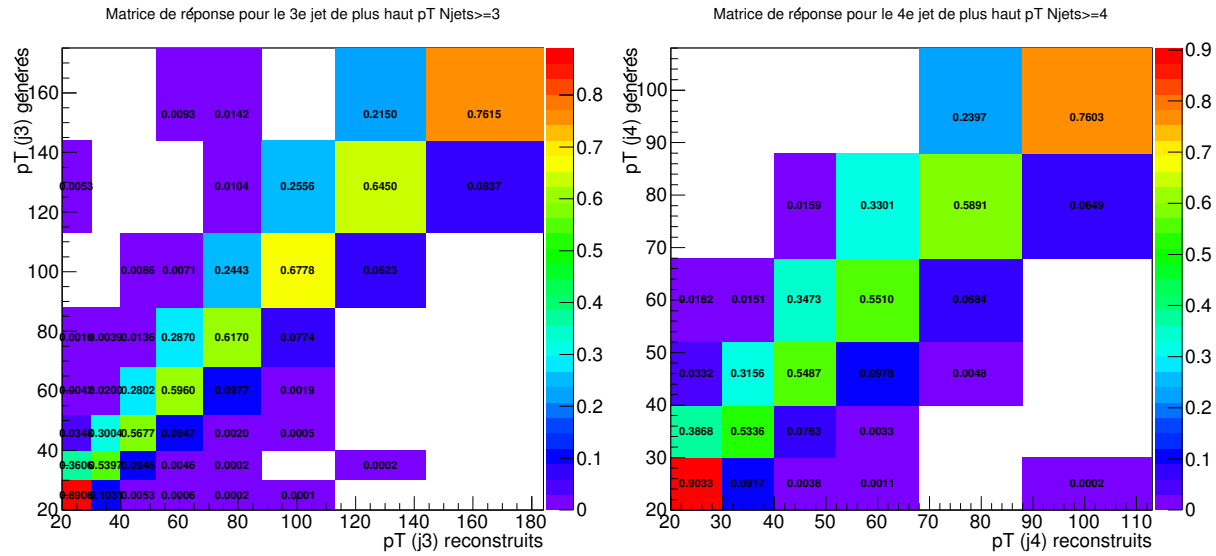


FIGURE 7.3: Matrices de réponses pour l'impulsion transverse du troisième jet (à gauche) et du quatrième jet (à droite).

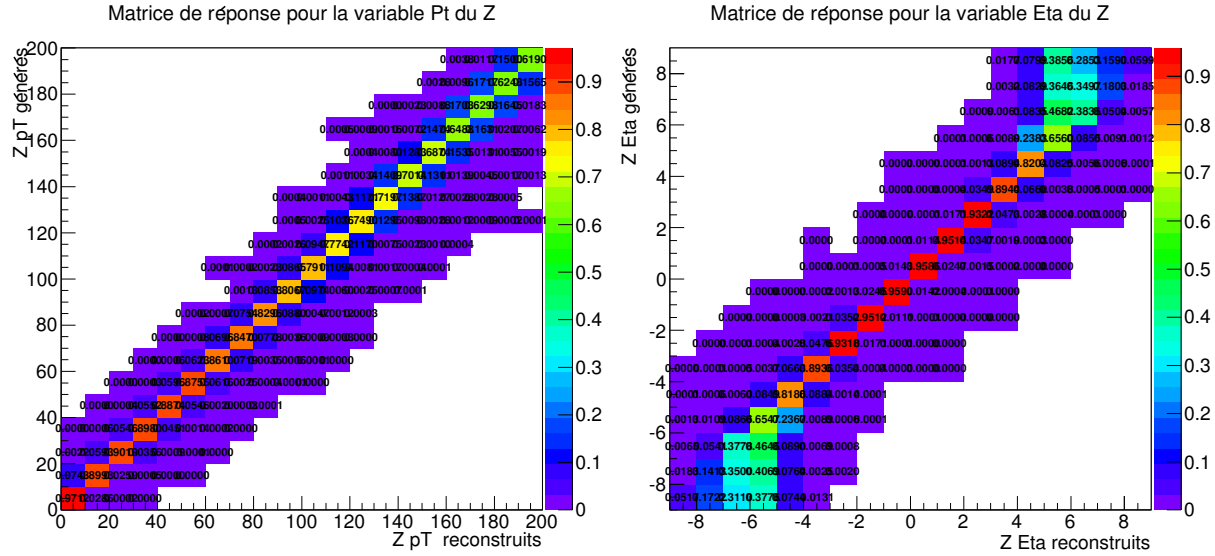


FIGURE 7.4: Matrices de réponses pour les variables cinématiques du boson Z : p_T à gauche et η à droite.

Pour les distributions des variables η , le p_T du Z et de la multiplicité de jets, nous avons utilisé les valeurs underflow stockées dans les histogrammes construits par root. Pour les distributions en p_T des jets, nous avons rempli les histogrammes en prenant une coupure en p_T de 20 GeV mais en gardant l'échelle des histogrammes présentés à 30 GeV (ce qui est équivalent à garder un bin d'underflow de 20 à 30 GeV).

Nombre d'itérations :

La procédure d'Unfolding est dépendante du type de générateur utilisé, de la méthode d'interpolation et dans la plupart des méthodes du nombre d'itérations représentées par un paramètre de régularisation (k_{term}) que l'on entre en paramètre de la méthode choisie dans RooUnfold.

Il s'agit d'effectuer ni trop, ni trop peu d'itérations : on peut optimiser le k_{term} en comparant à chaque étape l'erreur statistique générée à celle de l'étape précédente. Lorsque la nouvelle itération est compatible avec le précédente on décide d'arrêter l'itération. Nous verrons plus en détail le choix du k_{term} pour l'Unfolding avec le théorème de Bayes.

7.1.2 Unfolding par la méthode de Bayes

Nous avons utilisé la méthode d'Unfolding dite Bayesian, proposée par RooUnfold et basée sur le théorème de Bayes. Le théorème de Bayes est un outil statistique général qui décrit les distributions de densité de probabilité soumises à des contraintes et permet de tenir compte d'une certaine corrélation entre ces probabilités. Nous l'avons déjà mentionné dans le cadre de l'établissement des erreurs sur les efficacités (cf.chapitre 6.1.2.3)

La méthode de Bayes pour l'Unfolding est décrite dans [53]. C'est à partir de cet

article que l'implémentation RooUnfold a été menée¹. Nous allons ici en résumer les arguments et les étapes logiques en utilisant la nomenclature de l'article et en faisant le lien avec notre application.

La méthode de Bayes pour l'Unfolding a le grand avantage de ne pas procéder à une inversion de matrice. Elle évite ainsi les problèmes dus aux singularités. De plus, elle est basée sur un théorème bien ancré théoriquement, et permet de prendre en compte n'importe quel type de migrations.

Les limites de la méthode concernent les cas où les distributions de la matrice de réponse sont trop éloignées de la diagonale, i.e. quand la distribution de données est trop éloignée de la distribution MC. De manière générale, la méthode ne gère pas bien les grandes fluctuations statistiques mais dans notre cas, les distributions de données à unfold sont déjà très proches du niveau généré recherché et le choix de la taille des bins permet de garder une bonne statistique par bin. En principe cette approche requiert de connaître la distribution attendue. On contourne cette difficulté en procédant de façon itérative.

Principes mathématiques de la méthode :

La méthode considère les probabilités $P(C_i)$ de n causes indépendantes C_i , $i=1,2,\dots,n$ et de probabilité initiale $P_0(C_i)$.

Par exemple dans notre cas les C_i sont les valeurs générées de nos variables d'intérêt. On définit également un nombre d'effets E_j , avec $j=1,2,\dots,n_E$. Pour nous, il s'agit des valeurs reconstruites de nos variables d'intérêt.

$P(E_j/C_i)$ est ainsi la probabilité conditionnelle qu'un effet E_j soit causé par la contrainte C_i . Cette probabilité correspond à un élément R_{ij} de notre matrice de réponse. Elle est entièrement déterminée par des données MC.

Le théorème de Bayes nous donne l'expression de la probabilité inverse $P(C_i/E_j)$ en fonction des éléments de la matrice de réponse et des probabilités initiales $P_0(C_i)$:

$$P(C_i/E_j) = \frac{P(E_j/C_i)P_0(C_i)}{\sum_{i=1,\dots,n} P(E_j/C_i)P_0(C_i)} \quad (7.1)$$

On peut le comprendre de la façon suivante : si l'on observe un effet E_j particulier, la probabilité qu'il ait été obtenu par la i ème cause est proportionnelle à la probabilité de la cause elle-même fois la probabilité que la cause produise l'effet E .

Les $P(C_i/E_j)$ dépendent au départ des distributions de probabilité initiales $P_0(C_i)$. A la première itération de la formule une nouvelle probabilité $P(C_i)$ va être déterminée à partir des $P(C_i/E_j)$ calculées par le théorème de Bayes en utilisant la relation : $P(C_i) = \sum_{j=1}^{n_E} P(C_i/E_j)$. A l'itération suivante, on remplace les $P_0(C_i)$ par les nouvelles $P(C_i)$.

La puissance du théorème est donc d'augmenter notre connaissance de $P(C_i)$ à partir du nombre d'observations $P(E/C_i) = R_{ij}$.

1. Il faut noter toutefois que des erreurs se trouvent dans l'article mais ont été ensuite corrigées dans l'implémentation RooUnfold. Ces erreurs ont été mises en évidence dans [54].

En pratique, on cherche à déterminer des distributions unfoldées $\hat{n}(C_i)$ à partir de distributions de données reconstruites $n(Ej)$:

$$\hat{n}(C_i) = \frac{1}{\epsilon_i} \sum_{j=1}^{n_E} n(Ej)P(C_i/Ej), \quad \text{avec } \epsilon_i = \sum_{j=1}^{n_E} P(Ej/C_i), \epsilon_i \neq 0 \quad (7.2)$$

Les ϵ_i représentent les efficacités de détecter une certaine cause quel que soit son effet : si $\epsilon_i = 0$ alors $\hat{n}(C_i)$ sera mis à 0 puisque l'expérience est insensible à la cause C_i .

De plus, à partir des éléments $\hat{n}(C_i)$ nous pouvons déterminer le nombre total d'événements unfoldés $\hat{N}_{true}$ et déterminer ainsi des distributions de probabilité $\hat{P}(C_i)$ telles que :

$$\hat{N}_{true} = \sum_{i=1}^{n_C} \hat{n}(C_i) \quad (7.3)$$

$$\hat{P}(C_i) = \frac{\hat{n}(C_i)}{\hat{N}_{true}} \quad (7.4)$$

Pour déterminer les distributions unfoldées $\hat{n}(C_i)$ l'on procède de façon itérative en trois étapes :

1. On choisit les distributions initiales $P_0(C_i)$: dans notre cas on utilise nos distributions de données MC générées mais l'on pourrait commencer par une distribution uniforme,
2. on calcule $\hat{n}(C_i)$ en utilisant les formules 7.1 et 7.3 et $P_0(C_i)$,
3. on remplace $P_0(C_i)$ par $\hat{P}(C_i)$ et $n_0(C_i)$ par $\hat{n}(C_i)$, et l'on procède à une nouvelle itération.

A chaque itération l'on établit un χ^2 de comparaison entre les distributions $\hat{n}(C_i)$ et les distributions de l'itération précédente $\hat{n}_0(C_i)$. Si le χ^2 obtenu est assez petit on peut arrêter l'itération. Sinon l'on continue. Dans notre cas la distribution générée est proche de ce que l'on cherche : déjà à partir de quatre itérations il y a convergence. On peut continuer les itérations mais chaque distribution sera compatible avec la précédente.

7.1.3 Validation de la procédure à partir du Monte Carlo

Un des tests de la procédure d'Unfolding est d'appliquer la matrice aux échantillons MC eux même. Trois tests différents peuvent être effectués :

- test sur des distributions où le même échantillon MC est utilisé pour calculer la matrice de réponse et effectuer le test. Dans ce cas on sépare l'échantillon en deux parties égales : l'une qui sert pour calculer la matrice de réponse, l'autre sur laquelle on applique l'Unfolding. Pour ces tests la statistique est divisée par deux.
- test sur des distributions où des générateurs différents sont utilisés pour l'échantillon généré, et l'échantillon sur lequel on applique l'Unfolding. Ce test permet de garder la statistique complète.
- test de l'effet de petites variations dans le paramètre de régularisation qui permet de vérifier la stabilité de l'Unfolding.

Ces test dit "de fermeture" permettent de valider la procédure et de déterminer les comportements statistiques. Le rapport entre MC "unfoldé" et MC généré est supposé revenir à 1.

Dans ce travail nous avons effectué uniquement le premier test cité, et ce pour l'ensemble de nos distributions. Nous présentons ici deux exemples dans la figure 7.5.

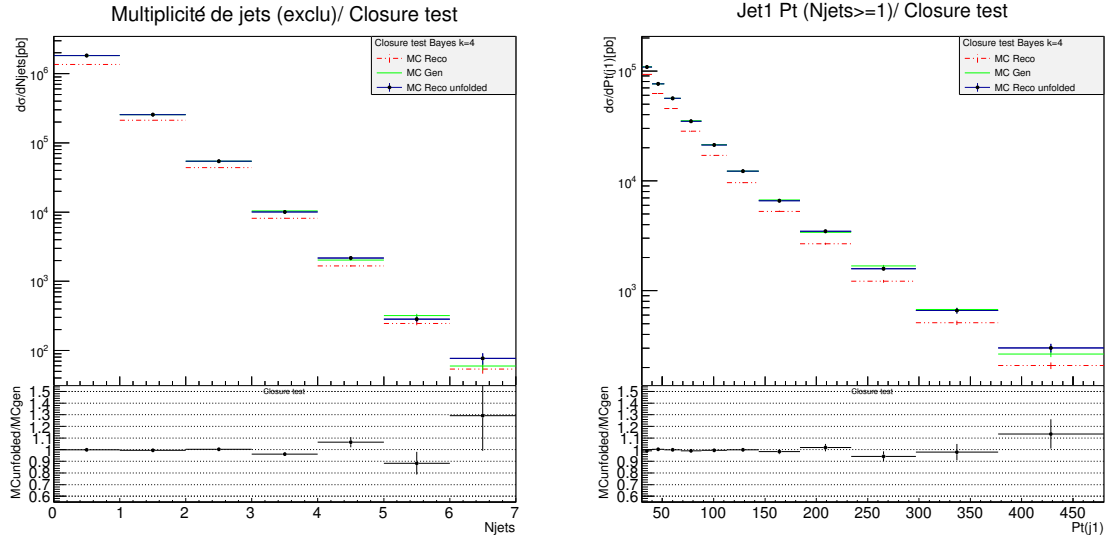


FIGURE 7.5: Test de fermeture pour la procédure d'unfolding pour les distributions de multiplicité exclusive de jets (gauche) et de la variable p_T du premier jet (droite) et comparaison avec le MC généré.

Pour les deux distributions l'on observe une bonne adéquation entre valeurs unfoldées et valeurs générées (graphiques du bas). Pour la multiplicité de jet, le désaccord estimé par le rapport $MC_{unfolded}/Gen$ est inférieure à 1 % jusqu'à 3 jets, vaut environ 4 % pour 4 jets, et environ 10 % pour 5 jets. Pour la distribution en p_T du premier jet le désaccord est inférieur à 2 % jusqu'à $p_T = 80$ GeV et inférieur à 10 % jusqu'à $p_T = 350$ GeV. Pour la dernière bin avec moins de statistique la procédure est moins précise et la valeur du rapport $MC_{unfolded}/Gen$ dépasse les 10 %.

Ces distributions représentées sur les graphiques du haut permettent d'observer le comportement entre données MC Reco et données MC unfoldées : cela rend compte de la gestion des migrations d'une bin à l'autre par l'algorithme d'Unfolding. Typiquement pour une distribution en fonction de la variable p_T la pente ne varie pas beaucoup entre données "reco" et données "unfoldées" et ce grâce à la bonne résolution de reconstruction de jets.

7.2 Erreurs systématiques

Nous avons déjà mentionné à plusieurs reprises que l'analyse effectuée comportait des sources d'erreurs à prendre en compte lors de l'établissement des résultats. Ces erreurs peuvent provenir de l'analyse mais aussi du dispositif expérimental, des procédures de reconstruction et également des approximations théoriques. Ces erreurs sont regroupées dans l'établissement général des erreurs systématiques.

Dans ce travail, faute de temps nous n'avons pas calculé les erreurs systématiques associées aux sections efficaces différentielles présentées dans nos résultats. Néanmoins, nous présenterons ici les différentes sources d'erreurs systématiques intervenant dans notre analyse ainsi qu'un ordre de grandeur de leur action [51].

De façon générale, la détermination des erreurs systématiques se fait en incluant les pourcentages d'erreurs dans les algorithmes construits pour notre échantillon, et ce dans les deux sens : une fois vers les valeurs inférieures, une fois vers les valeurs supérieures. On applique alors les algorithmes modifiés aux données et l'on obtient pour chaque distribution les valeurs limites correspondant aux erreurs considérés.

Les erreurs systématiques intervenant dans ce travail proviennent de différentes sources que nous énumérerons ici par ordre d'importance.

- Jet Energie Scale (JES) : la détermination de l'énergie des jets entraîne une proportion d'erreur d'environ 5% sur les valeurs d'énergies déterminées. Pour prendre en compte cette erreur, on fait tourner la même procédure en appliquant aux p_T des jets une variation de 5%. Les erreurs correspondantes à la JES sont dominantes pour les distributions considérées.
- Luminosité : la détermination de la luminosité à CMS est une opération technique qui a une certaine incertitude. Elle entraîne 4.4% d'erreur à prendre en compte sur la normalisation des sections efficaces.
- Jet Energie Resolution (JER) : cette erreur provient principalement de la mauvaise description de la résolution des jets par la simulation MC qui surestime légèrement l'énergie des jets par rapport aux données. Les systématiques correspondant à la JER ne dépassent pas 2%.
- Estimation du bruit de fond : la modélisation du bruit de fond ne correspond pas exactement à la réalité, chaque contribution étant susceptible de participer un peu plus ou un peu moins au bruit de fond de notre signal. Pour déterminer cette incertitude l'on fait varier de façon indépendante les sections efficaces correspondant à chaque processus ainsi que les luminosités correspondantes.
- Pileup : la modélisation du pileup est également une source d'incertitude et cela se manifeste dans la façon de pondérer le fichier MC pour le nombre de vertex. CMS recommande de tenir compte d'une variation de 5% pour cette correction, ce qui entraîne des erreurs de l'ordre de 0.5% sur nos distributions.
- Efficacités : Comme nous l'avons déjà mentionné, les erreurs dues à la détermination des facteurs de corrections entraînent des erreurs systématiques d'un maximum de 1 ou 2 % dans certains bins.
- Unfolding : les erreurs provenant de la procédure d'Unfolding dépendent de la méthode utilisée (Bayesian ou SVD) et sont fortement corrélées aux distributions concernées. Elle sont déterminées en faisant varier des paramètres dans les procédures d'Unfolding. La procédure d'Unfolding entraîne des erreurs de l'ordre de 0.5%.
- Parton distribution function PDF : il s'agit ici d'une erreur théorique qui rend compte des imprécisions de description des PDF. Cette erreur est minime par rapport aux erreurs précédentes. On la détermine en appliquant des procédures qui varient les PDF utilisées pour générer le MC.

7.3 Résultats

Les résultats présentés dans cette section ont été obtenus en appliquant les différentes procédures décrites précédemment : l'échantillon de données est sélectionné comme présenté au chapitre 5, les fichiers MC sont corrigés pour le pileup et les efficacités (cf. chapitre 6), et enfin la matrice de réponse est construite à partir du fichier MCDY comme nous venons de le décrire dans ce chapitre. Cette matrice de réponse est maintenant appliquée aux données auxquelles le bruit de fond a été soustrait bin par bin.

Sur chaque distribution sont représentées les données au niveau reconstruit (données "reco") et les données après Unfolding ainsi que les valeurs du MC généré. A chaque distribution est également associée un graphique représentant le rapport entre données unfoldées et données MC générées. Ainsi MC généré et données "reco" servent de repères pour se rendre compte de l'action de la procédure d'Unfolding sur chaque distribution.

Les erreurs représentées sur les distributions sont les erreurs statistiques. Les erreurs systématiques ne sont donc pas représentées (leur contribution serait en général beaucoup plus importante).

7.3.1 Section efficace différentielle en la multiplicité exclusive de jets

La section efficace différentielle en la multiplicité exclusive de jets est représentée sur la figure 7.6.

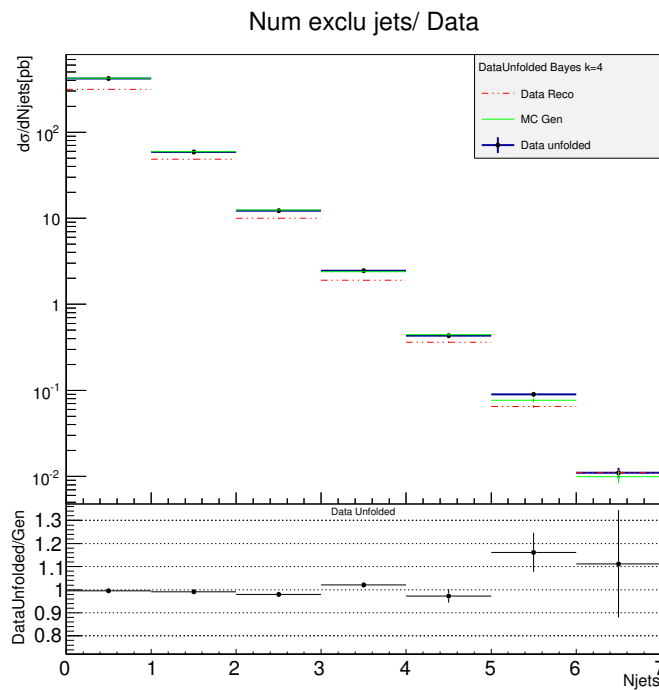


FIGURE 7.6: Section efficace différentielle en fonction de la multiplicité exclusive de jets : le nombre de jets considéré est strictement inférieur à 7.

7.3.2 Section efficace différentielle en les variables p_T et η du boson Z

Ici nous nous intéressons aux variables associées au boson Z. Nous avons vu que le pic du Z était décrit avec une bonne précision par le MC. La résolution de sélection et reconstruction d'un Z est très bonne et ceci a permis un binning plus petit.

Les résultats pour la section efficace différentielle en l'impulsion transverse et la variable η du Z sont représentés sur la figure 7.7.

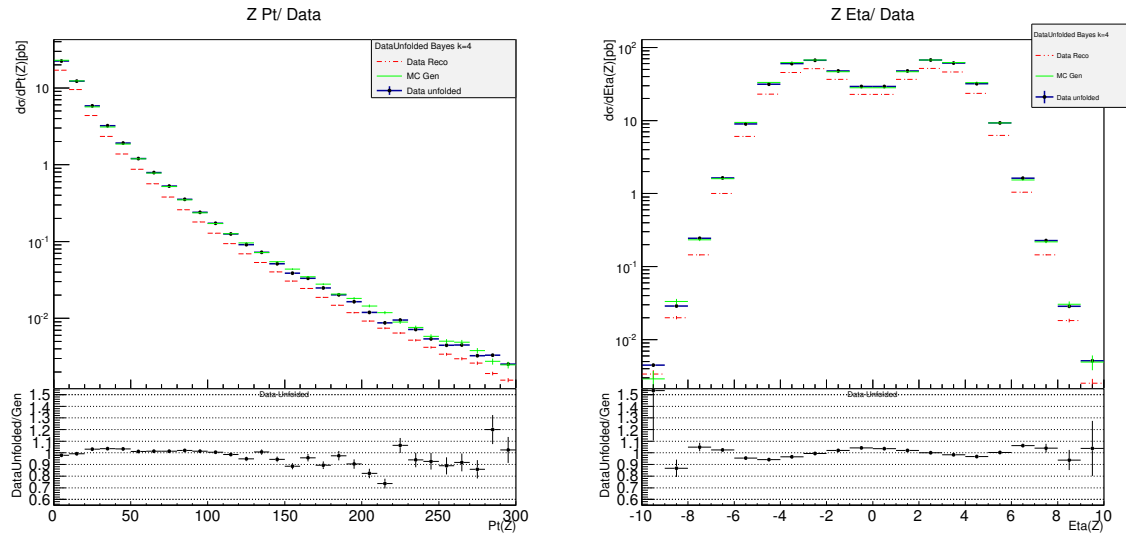


FIGURE 7.7: Section efficace différentielle en fonction de l'impulsion transverse du Z (gauche) et de variable η du Z (droite).

7.3.3 Section efficace différentielle en la variable p_T des jets

Les résultats pour la section efficace différentielle en l'impulsion transverse pour chaque jet sont représentés sur les figures 7.8 et 7.9 .

7.3.4 Section efficace différentielle en la variable η des jets

Les résultats pour la section efficace différentielle en la variable η des jets sont représentés sur les figures 7.10 et 7.11.

7.4 Commentaires sur les résultats

Sans estimation rigoureuse des erreurs systématiques il est difficile de discuter les résultats. Cependant si nous supposons qu'elle est de l'ordre de 8 à 10% pour la plupart des bins, nous pouvons conclure que l'accord entre les données expérimentales et la prédiction de la simulation MC est remarquable pour la multiplicité de jets, les variables du Z et pour les variables cinématiques des deux premiers jets : pour la plupart des bins l'accord est inférieur ou de l'ordre de 10%.

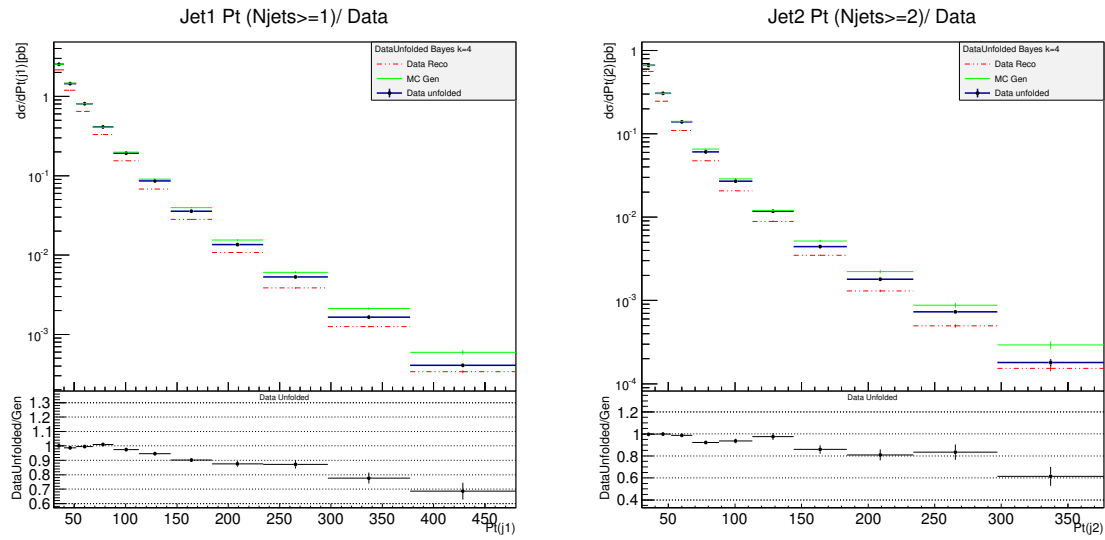


FIGURE 7.8: Section efficace différentielle en fonction de l'impulsion transverse du premier jet (gauche) et du second jet (droite).

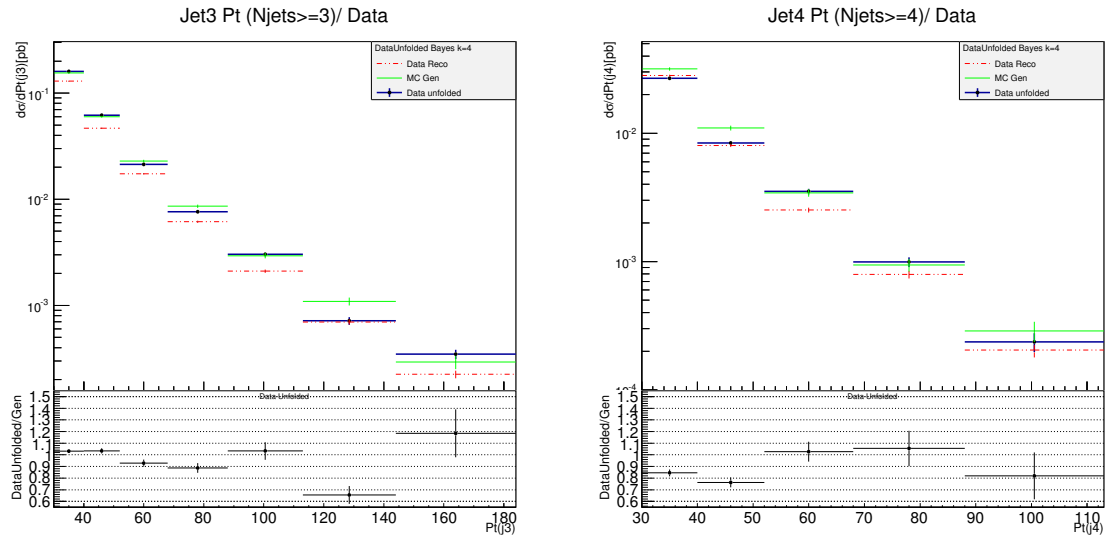


FIGURE 7.9: Section efficace différentielle en fonction de l'impulsion transverse du troisième jet (gauche) et du quatrième jet (droite).

Dans le détail nous pouvons cependant remarquer que :

- la distribution de la multiplicité des jets présente une adéquation avec les valeurs du MC généré. Pour les deux premiers jets, le désaccord représenté par "Dataunfolded/MC" est inférieur à 1%, pour le troisième et quatrième jet il reste inférieur à 2 %. Pour le 5ème jet, le désaccord atteint environ 18% : le MC sous-estime alors l'impulsion transverse du jet qui est dans ce cas uniquement généré par la "parton shower" de Pythia.
- la distribution en impulsion transverse des jets décroît plus rapidement dans les données que dans le MC, cela est plus clair plus le 1er et le 2ème jets. Plus de

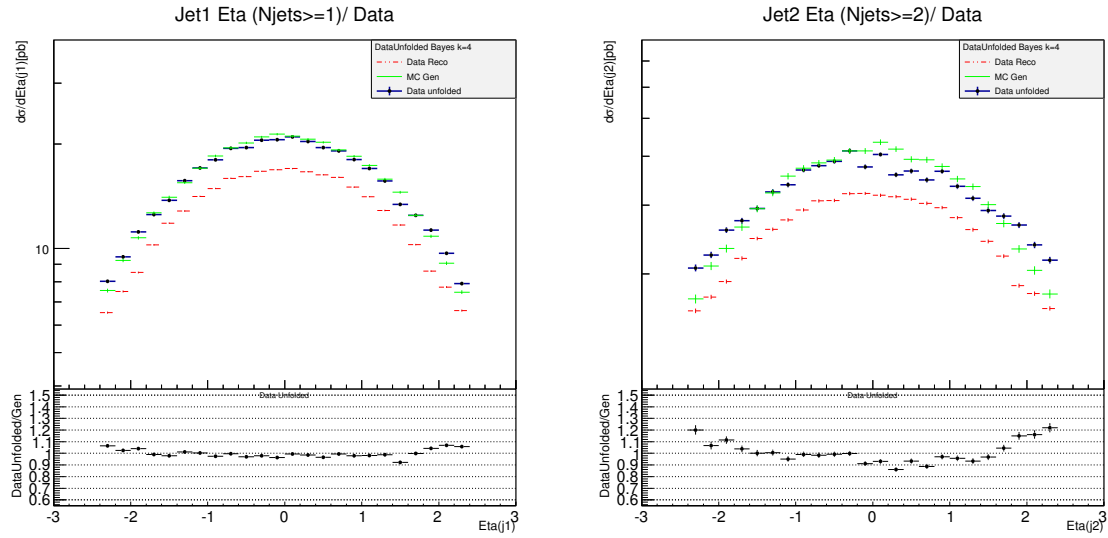


FIGURE 7.10: Section efficace différentielle en fonction de la variable η du premier jet (gauche) et la variable η du second jet (droite).

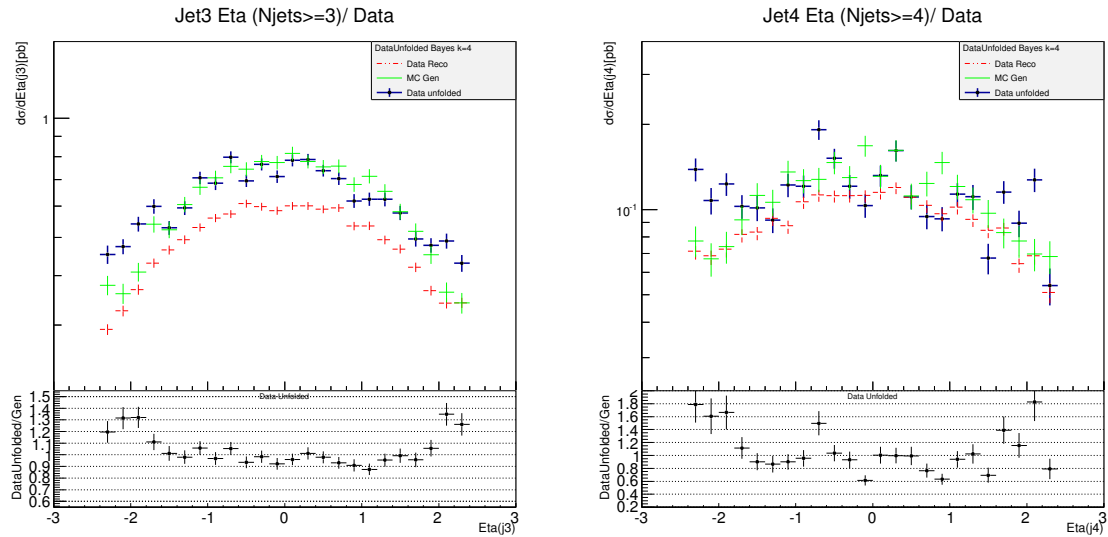


FIGURE 7.11: Section efficace différentielle en fonction de la variable η du troisième jet (gauche) et la variable η du quatrième jet (droite).

statistiques sont nécessaires pour pouvoir confirmer cette tendance pour les 3eme et 4eme jets

- la distribution en p_T du Z tombe subitement plus vite dans les données que dans le MC pour des valeurs proches de 210 GeV et ensuite remonte un peu. Une étude détaillée des erreurs systématiques est ici nécessaire pour savoir s'il s'agit d'un effet physique réel ou d'un effet expérimental.
- les distributions en η des jets sont systématiquement plus hautes dans les données que dans le MC au grandes valeurs de $|\eta|$. Cette distribution est sensible à la dépendance en x des PDF, une étude se basant sur différentes PDF serait nécessaire

pour voir si ces données permettraient de mieux contraindre les PDFs (après une prise en compte des erreurs systématiques bien entendu).

Chapitre 8

Conclusion

Nous venons de décrire les différentes étapes de l'analyse complète du processus $pp \rightarrow Z(\rightarrow \mu_+\mu_-) + jets$ à une énergie de 8 TeV dans le centre de masse.

Les résultats présentés sont de façon générale en bonne adéquation avec les prédictions théoriques représentées par la simulation MC de MadGraph. Toutefois, l'établissement des erreurs systématiques doit être mené afin de préciser certains comportements.

La multiplicité exclusive de jets associée à la production d'un Z est remarquablement bien décrite par le MC généré par MadGraph/Pythia : jusqu'à 4 jets, données et MC présentent un accord au pour-cent près. A de telles énergies c'est un véritable accomplissement et ces résultats confirment donc la pertinence du modèle QCD à une telle énergie.

Pour les distributions associées au 2 premiers jets produits avec un boson Z, le rapport entre données et MC reste inférieur à 10% pour des valeurs de p_T inférieures à 120 GeV. Ces résultats resteraient donc en adéquation avec une erreur systématique globale estimée à 10%. Pour les troisième et quatrième jets associés à la production d'un Z, les fluctuations statistiques deviennent importantes. Des données prises à plus haute luminosité vont pouvoir augmenter la statistique et l'étude de ces jets sera plus précise. Ces résultats confirment d'une part la validité du modèle standard à la plus haute énergie jamais atteinte en laboratoire, et d'autre part valident l'approche du générateur MC MadGraph dans la description d'événements allant jusqu'à 4 jets.

A partir des résultats obtenus et au vue de leur précision, l'on peut envisager l'étude d'autres propriétés liées à la production des jets comme les corrélations angulaires entre jets.

Les propriétés du Z sont également décrites avec une bonne précision à une énergie de 8 TeV. Comme nous l'avons vu, le processus de production d'un Z peut servir d'étalon pour des futures calibrations à plus hautes énergies.

Pour décrire de façon plus complète la production d'un Z au LHC, il conviendrait également d'étendre notre analyse au canal de désintégration en électron-positron. Les étapes sont les mêmes, seule la reconstruction et le calcul des efficacités diffèrent : notre analyse pourrait donc facilement être élargie à ce canal et cela permettrait de doubler la statistique des données.

Pour la partie pileup, nous avons bien établi les enjeux sous-jacents : dans les prochaines phases du LHC ce problème sera amplifié par l'augmentation de la luminosité et la diminution de l'espacement entre paquets et les outils décrits devront être réévalués.

L'établissement des erreurs systématiques permettrait également de pouvoir comparer les résultats à des analyses similaires. Par exemple, nous avons mené exactement les mêmes étapes d'analyse que l'étude présentée en [51] sauf que nous avons déterminé nos propres calculs pour les corrections liées aux efficacités. Les échantillons utilisés concernent les mêmes données mais n'ont pas exactement les mêmes niveaux de reconstruction. Les résultats des deux analyses présentent les mêmes comportements et les valeurs numériques semblent correspondre dans les limites des erreurs systématiques établies en [51].

Table des figures

2.1	Schémas représentant les évolutions de α_{em} et α_s en fonction de la distance .	9
2.2	Schéma représentant les différentes étapes du processus $e^+e^- \rightarrow Z \rightarrow q\bar{q}$. . .	10
2.3	Schéma représentant le modèle "naïf" des partons.	10
2.4	Description de la PDF CTEQM pour $Q=2$ et $Q=100\text{GeV}$	13
2.5	Exemples de diagrammes constituant les correction au NLO au processus Drell-Yan	14
3.1	Chaîne des accélérateurs au CERN	15
3.2	Liste non-exhaustive des expériences au CERN.	17
3.3	coupe transverse du LHC -	19
3.5	Shéma du système de tracker de CMS	23
3.6	Coupe transversale de L'ECAL, vue 1/4.	25
3.7	Coupe transverse et vue 1/4 du HCAL de CMS.	26
3.8	Simulation de la distribution du champ magnétique dans CMS.	28
3.9	Coupe transverse d'1/4 du système de détecteurs à muons de CMS.	29
3.10	Description du système de trigger de CMS	31
3.11	Description du niveau de trigger L1 de CMS	32
3.12	Descriptions du détecteur CMS	35
3.13	Shéma représentant la chaîne de traitement des données à CMS.	36
4.1	Schéma représentant les différents sous-processus générés	39
4.3	Exemple de reconstruction de jets par différents algorithmes	55
5.1	Illustration schématique de la procédure d'isolation des muons.	63
5.2	Distribution du nombre de vertex pour les données 2012.	66
5.3	Schémas représentant les principes inspirant les méthodes de jet area et d'offset moyen.	67
5.4	Exemple illustrant la détermination de la variable ρ	68
5.5	Correction moyenne offset en fonction de η et des objets du PF.	69
5.6	Comparaison pour les corrections d'offset moyen, jet area et méthode hybride.	70
5.7	Distributions en p_T du jet de plus haut p_T associé au boson Z et au pileup. .	73
5.8	Distributions en η du jet de plus haut p_T associé au boson Z ou au pileup. .	74
5.9	Distributions en η du jet de plus haut p_T avec contrainte sur l'impulsion transverse du Z.	75
6.1	Distribution du nombre de vertex pour le MC et les données	78

6.2	Distributions en pour l'identification du second muon de plus haut p_T et pour l'efficacité d'identification	84
6.3	Distributions pour l'isolation du second muon de plus haut p_T et pour l'efficacité d'isolation	85
6.4	Distributions de contrôles pour la multiplicité de jets inclusive (haut) et exclusive (bas).	86
6.5	Distributions de contrôle pour l'impulsion transverse des jets associés au Z. Ces distributions sont inclusives.	87
6.6	Distributions de contrôle pour l'impulsion transverse (haut) et la variable η (bas) du boson Z reconstruit.	88
6.7	Distribution de contrôle pour la masse invariante du Z. Sont représentés les échantillons comprenant 0 jets, au moins 1 jet, au moins 2 jets et au moins 3 jets.	89
7.1	Matrice de réponse pour la multiplicité exclusive de jets.	93
7.2	Matrices de réponses pour l'impulsion transverse du premier et du second jet.	94
7.3	Matrices de réponses pour l'impulsion transverse du troisième et du quatrième jet.	94
7.4	Matrices de réponses pour les variables cinématiques du boson Z.	95
7.5	Test de fermeture pour la procédure d'unfolding.	98
7.6	Section efficace différentielle en fonction de la multiplicité exclusive de jets.	100
7.7	Section efficace différentielle en fonction de l'impulsion transverse et la variable η du Z.	101
7.8	Section efficace différentielle en fonction de l'impulsion transverse du premier jet (gauche) et du second jet (droite).	102
7.9	Section efficace différentielle en fonction de l'impulsion transverse du troisième jet (gauche) et du quatrième jet (droite).	102
7.10	Section efficace différentielle en fonction de la variable η du premier jet (gauche) et la variable η du second jet (droite).	103
7.11	Section efficace différentielle en fonction de la variable η du troisième jet (gauche) et la variable η du quatrième jet (droite).	103

Liste des tableaux

2.1	Composantes de la matière et des interactions décrites dans le cadre du Modèle Standard	4
3.1	Paramètres des faisceaux de proton et de collisions du LHC	20

4.1	Exemples de diagrammes représentant les différentes possibilités d'obtenir un parton supplémentaire dans l'état final.	43
4.2	Fichiers MC utilisés pour notre analyse, avec le nombre d'événements dans l'échantillon, la section efficace et la luminosité correspondantes.	46
5.1	Description des fichiers de données utilisés pour notre analyse.	60
5.2	Fichiers JSON utilisés pour notre analyse.	60
5.3	Liste des coupures de sélection appliquées aux muons.	62
5.4	Compteur pour la sélection des muons de notre échantillon de données. . . .	62
5.5	Liste des coupures effectuées sur les jets.	64
5.6	Critères d'identification des jets	65
6.1	Corrections pour l'efficacité de trigger des muons selon les différents domaines en η de chacun des muons	79
6.2	Facteurs de corrections pour l'efficacité d'identification des muons.	82
6.3	Facteurs de corrections pour l'efficacité d'isolation des muons.	82

Chapitre 9

Annexe 1

9.1 Liste des Acronymes et abréviations

AOD		Analysis Object Data
CHS		Charge Hadron Subtraction
CMS		Compact Muon Solenoid
CMSSW		CMS Software
CSC		Cathodic Strip Chambers
DIS		Deep Inelastic Scattering
DT		Drift Tube Chambers
DY		Drell-Yan
ECAL		Electronic Calorimeter
EMF		Electromagnetical Fraction
FSR		Final State Radiation
GEANT		GEometry AND Tracking
GSF		Gaussian Sum Filter
HB		Barel Hadronic Calorimeter
HCAL		Hadronic Calorimeter
HE		Endcap Hadronic Calorimeter
HF		Hadronic Forward (Calorimeter)
HLT		High-Level Trigger
HO		Hadronic Outer Calorimeter
ICS		Infrared and Colinear Safe
IP		Interaction Point
IPF		Isolation Particle Flow
ISR		Initial State Radiation
JEC		Jet Energy Correction
JMAR		JetMET Algorithms and Reconstruction

JMetPOG	Jet and Missing energy transverse POG
JPT	Jets Plus Track
JSON	JavaScript Object Notation
L1	Level-1 Trigger
LEP	Large Electron Positron
LHC	Large Hadron Collider
LO	Leading Order
LS1	Long Shutdown 1
MC	Monte Carlo
MET	Missing Energie Transverse
MPI	Multiple Parton Interaction
MVA	Multi Variate Analysis
MS	Muon System
SM	Standard Model
NLO	Next Leading Order
PDF	Parton Distribution Function
PF	Particle Flow
POG	Physic Object Group
PS	Proton Synchrotron
PSB	Proton Synchrotron Booster
PU	Pileup
QCD	Quantum Chromodynamics
RECO	Reconstructed Data
RPC	Resistive Plate Chamber
sc	supraconducteur
SF	Scale Factor
SPS	Super Proton Synchrotron
SVD	Single Value Decomposition
UE	Underlying Event
VP	Vertex Primaire

Bibliographie

- [1] P.N. Ratoff MG Green, S.L. Lloyd and D.R.Ward. *Electron-Positron Physics at the Z*. IOP Publishing, ISBN-10 : 0750303832, 1998.
- [2] Particle Data Book. <http://pdg.lbl.gov/2013/tables/rpp2013-sum-gauge-higgs-bosons.pdf>. *Review of Particle Physics*, 2013.
- [3] Laurent Favart. *Physique auprès des collisionneurs*. Cours à option de MA1 ou MA2 en Physique, ULB <http://homepages.ulb.ac.be/~lfavart/>, 2012.
- [4] G. Altarelli. QCD evolution equations for parton densities, http://www.scholarpedia.org/article/QCD_evolution_equations_for_parton_densities. *Scholarpedia*, 4(1) :7124, 2009.
- [5] D.R. Stump J.Pumplin and al. New generation of parton Distributions with Uncertainties from Global QCD Analysis. *Journal of High Energy Physics*, 2002(07) :012, 2002.
- [6] CERN Beam Departement Opération Group. <https://espace.cern.ch/be-dep/OP/default.aspx>, Mars 2012.
- [7] CERN Homepage. <http://home.web.cern.ch/about/experiments>.
- [8] FERMILAB Website US Dept of Energy. <http://www.fnal.gov/pub/science/experiments/energy/tevatron/>, Mars 2011.
- [9] CERN Press Center. <http://press.web.cern.ch/fr/press-releases/2013/03/de-nouveaux-resultats-indiquent-que-la-particule-decouverte-au-cern-est-un>, 14 mars 2013.
- [10] Lyndon R Evans and Philip Bryant. LHC machine. *J. Instrum.*, 3 :S08001. 164 p, 2008. This report is an abridged version of the LHC Design Report (CERN-2004-003).
- [11] Particle Data Book. <http://pdg.lbl.gov/2011/reviews/rpp2011-rev-accel-phys-colliders.pdf>. *Review of Particle Physics*, 2012.
- [12] R.J.Steinhausen for LHC Team. LHC status and future upgrades, 07-04-2012. Talk at ICHEP Melbourne.
- [13] LHC Programme Coordination LPC. <http://lpc.web.cern.ch/lpc/>, 2013.
- [14] LHC. <http://lhc.web.cern.ch/lhc/LHC-DesignReport.html>. Technical report, CERN, 2013.

- [15] CMS Public documentation. CMS slice image with transverse, longitudinal 3-D views <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=5697>, février 2012.
- [16] CMS Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *Journal of Instrumentation*, 5(03) :T03021, 2010.
- [17] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605 :026, 2006.
- [18] B. Webber. Parton shower Monte Carlo event generators, http://www.scholarpedia.org/article/Parton_shower_Monte_Carlo_event_generators. 6(12) :10662, 2011.
- [19] M. Maltoni Alwall, J. Herquet and all. Madgraph 5 : Going beyond. Technical Report arXiv :1106.0522, Jun 2011. Cite op p 4-12.
- [20] Lazaridis Christos. Jets produced in association with Zbosons in CMS at the LHC. 2011. under the supervision of Wesley H. Smith.
- [21] Geant4 Collaboration. Geant4 developments and applications. *Nuclear Science, IEEE Transactions on*, 53(1) :270–278, 2006.
- [22] CMS Web SWGuideWorkBook : Simulation et Digitalisation. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookSimDigi>, 26-Feb-2010.
- [23] Z. Was. TAUOLA the library for tau lepton decay, and KKMC KORALB KORALZ ... status report. *Nucl.Phys.Proc.Suppl.*, 98 :96–102, 2001.
- [24] CMS Collaboration. Particle-flow event reconstruction in CMS and performance for Jets, Taus, and MET. *CMS Physics Analysis Summary, CMS-PAS-PFT-09-001*, 21(22) :9, 2009.
- [25] CMS Web Particle Flow and Tau Identification Physics Object Group. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/EFLOW>, 12-Jan-2011.
- [26] CMS Web Particle Flow Group. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideParticleFlow>, 10-Oct-2011.
- [27] Thomas Lenzi. *Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors*. PhD thesis, Brussels U., IIHE, May 2013. Presented 28 May 2013.
- [28] L Barbone, N De Filippis, O Buchmueller, Schilling, and al. Impact of CMS silicon tracker misalignment on track and vertex reconstruction. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, 566(1) :45–49, 2006.
- [29] CMS Collaboration. Performance of muon identification in 2010 data. *CMS Physics Analysis Summary CMS-PAS-MUO-10-004*, 2011.
- [30] David Sabes and Claude Charlot. *Reconstruction des électrons et recherche de couplages anormaux dans le canal $pp \rightarrow ZZ \rightarrow 4l$ avec le détecteur CMS auprès du LHC*. PhD thesis, Ecole Polytechnique, 2012. presented 19 Oct 2012.

- [31] CMS Web JetMet POG. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetMET>, 07-Apr-2013.
- [32] CMS Web JetMET Algorithms and Reconstruction (JMAR) Group. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetMETAlgorithmsReconstruction>, 7-March-2013.
- [33] M. Cacciari, G. P Salam, and G. Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04) :063, 2008.
- [34] CMS Web JEC Group. Jec group "intro to jec" <https://twiki.cern.ch/twiki/bin/viewauth/CMS/IntroToJEC>, 22-Dec-2011. voir aussi le Wikibooks du même group.
- [35] CMS Collaboration. Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. *J. Instrum.*, 6(arXiv :1107.4277. CMS-JME-10-011. CERN-PH-EP-2011-102) :P11002. 67 p, Jul 2011.
- [36] CMS Web SWGuideMuons : Baseline muon selections. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId>, 19-Aug-2013.
- [37] Iorio, Crisostomo Sciacca, and Luca Lista. *Study of the $pp \rightarrow Zjets \rightarrow \mu^+\mu^- + jets$ channel with the CMS detector at LHC*. PhD thesis, Naples U., Napoli, 2008. Presented on 16 Jul 2008.
- [38] CMS Collaboration. Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 TeV. *CMS Physics Analysis Summary CMS-PAS-PFT-10-002*, 23, 2010.
- [39] CMS Web JMAR Group. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID>, 02-Apr-2012.
- [40] Matteo Cacciari and Gavin P Salam. Pileup subtraction using jet areas. *Physics Letters B*, 659(1) :119–126, 2008.
- [41] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Catchment Area of Jets. *JHEP*, 0804 :005, 2008.
- [42] S. Rappoccio for CMS Jet Reco Group. Status and Plans for Pile Up Subtraction in Jet Reconstruction , CMS internal talk 2012.
- [43] CMS Web WorkBookJetEnergyCorrections. Official software tools for applying jec corrections.<https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookJetEnergyCorrections>, 05-Feb-2013.
- [44] CMS Web JMAR Group. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJetID>, 01-May-2013.
- [45] CMS Web MVA Algorithms and Tools. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/MVATools>, 17-Jan-2012.
- [46] M.Malberty J.Steggemann P.Harris, P.Musella. Pileup Jet Identification in 8 TeV pp data. *CMS Physics Analysis Summary. CMS PAS JME-13-005*, Draft submitted on 12 Juillet 2013.

- [47] CMS Web Muon POG. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonTagAndProbe>, 14-Aug-2013.
- [48] CMS Web Muon POG. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffs>, 16-Aug-2013.
- [49] CMS Web PhysicsTools. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TagAndProbe>. 24-May-2012.
- [50] Marc Paterno. *Calculating efficiencies and their uncertainties*. United States. Department of Energy, 2004.
- [51] T.Seva, A.Leonard, and L.Favart. Differential crosssection of jets associated to Z boson in proton proton collisions at $\sqrt{s} = 8\text{ TeV}$. *CMS Physique Analysis Summary*, To be Published.
- [52] RooUnfold. ROOT Unfolding Framework <http://hepunix.rl.ac.uk/~adye/software/unfold/RooUnfold.html>., 2013.
- [53] Giulio D’Agostini. A multidimensional unfolding method based on Bayes’ theorem. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, 362(2) :487–498, 1995.
- [54] Tim Adye. Unfolding algorithms and tests using RooUnfold. *arXiv preprint arXiv :1105.1160*, 2011.

Remerciements

Un grand merci à Laurent Favart pour ce qu'il m'a enseigné tout le long de ce travail, et également pour le soutien qu'il m'a apporté depuis ma reprise d'études en 2010. Cela a beaucoup compté pour moi.

Je tiens également à remercier particulièrement Alexandre Léonard ses explications parfaites, pour sa grande disponibilité, sa patience et sa bonne humeur et Tomislav Seva pour son expertise, et ses références qui m'ont été d'une grande aide. Vous constituez vraiment une « dream team » et vous allez certainement me manquer !

Je remercie également Barbara Clerbaux pour sa relecture active et ses conseils.

Merci également au service de l'IIHE et à tous mes collègues de classe, particulièrement mon collègue de bureau Thomas Lenzi qui a répondu imperturbablement à mes questions, à Laure-Anne Douchamps, Cedric Bamps, Francois Postic, Ismael Zekhnini.

Merci à Sarah Simoni, Gwenaël Breës et Marie Vela qui ont eu le courage de relire cette littérature, à mes amis Ben, Mouss, Maia, David, Manu qui m'ont supportés au quotidien dans cette période de stress, à Cécile et Gaëtane pour leur belle maison, à Romuald Bianchi, à ma mère et mes sœurs chéries, et ma nièce adorable Julia, et bien sûr à Federico, carissimo mio.

Enfin, spéciale dédicace à Marianne qui m'a convaincue de reprendre ces études. Je t'ai maudite parfois !