



**HAL**  
open science

# Recherche d'un secteur sombre avec le détecteur ATLAS

Nathan Lalloué

► **To cite this version:**

Nathan Lalloué. Recherche d'un secteur sombre avec le détecteur ATLAS. Physique [physics]. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALY054 . tel-03966197

**HAL Id: tel-03966197**

**<https://theses.hal.science/tel-03966197v1>**

Submitted on 31 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

École doctorale : PHYS - Physique

Spécialité : Physique Subatomique et Astroparticules

Unité de recherche : Laboratoire de Physique Subatomique et Cosmologie

**Recherche d'un secteur sombre avec le détecteur ATLAS**

**Search for a Dark Sector with the ATLAS Detector**

Présentée par :

**Nathan LALLOUE**

Direction de thèse :

**Marie-Hélène GENEST**

DIRECTRICE DE RECHERCHE, CNRS, Délégation Alpes

Directrice de thèse

Rapporteurs :

**STEPHANIE BEUCERON**

Chargé de recherche HDR, CNRS DELEGATION RHONE AUVERGNE

**HENRI BACHACOU**

Ingénieur HDR, CEA CENTRE DE PARIS-SACLAY

Thèse soutenue publiquement le **11 octobre 2022**, devant le jury composé de :

**CAROLINE COLLARD**

Directeur de recherche, CNRS DELEGATION ALSACE

Examinatrice

**NIKOLA MAKOVEC**

Chargé de recherche HDR, CNRS DELEGATION ILE-DE-FRANCE  
SUD

Examineur

**CHRISTOPHE FURGET**

Professeur des Universités, UNIVERSITE GRENOBLE ALPES

Président

**STEPHANIE BEUCERON**

Chargé de recherche HDR, CNRS DELEGATION RHONE AUVERGNE

Rapporteuse

**HENRI BACHACOU**

Ingénieur HDR, CEA CENTRE DE PARIS-SACLAY

Rapporteur

**DIRK ZERWAS**

Directeur de recherche, CNRS DELEGATION ILE-DE-FRANCE SUD

Examineur



# Table des matières

<b>Introduction</b>	<b>5</b>
<b>1 Le Modèle Standard de la physique des particules</b>	<b>7</b>
1.1 La Chromodynamique Quantique . . . . .	8
<b>2 La matière noire</b>	<b>14</b>
2.1 Preuves de l'existence de la matière noire . . . . .	14
2.2 Caractéristiques de la matière noire . . . . .	19
2.3 Candidats à la matière noire . . . . .	19
2.4 Méthodes de détection de la matière noire . . . . .	21
2.5 Secteur sombre . . . . .	24
<b>3 Le détecteur ATLAS</b>	<b>29</b>
3.1 Le CERN et le LHC . . . . .	29
3.2 Le détecteur ATLAS . . . . .	33
3.2.1 Systèmes de coordonnées . . . . .	34
3.2.2 Description des sous-détecteurs . . . . .	35
3.2.2.1 Le système d'aimants . . . . .	36
3.2.2.2 Le détecteur interne . . . . .	37
3.2.2.3 Les calorimètres . . . . .	40
3.2.2.4 Le spectromètre à muons . . . . .	44
3.3 La prise de données . . . . .	45
3.4 La qualité des données . . . . .	46
3.5 Les simulations Monte Carlo . . . . .	46
<b>4 La reconstruction des objets</b>	<b>48</b>
4.1 La reconstruction des traces . . . . .	48
4.2 La reconstruction des vertex . . . . .	51
4.3 La reconstruction des jets . . . . .	51
4.3.1 Les constituants des jets . . . . .	51
4.3.1.1 Les topo-clusters . . . . .	51
4.3.1.2 Les Particle Flows . . . . .	56
4.3.1.3 Les Track-CaloClusters . . . . .	59
4.3.1.4 Les vraies particules . . . . .	61
4.3.2 Les algorithmes de regroupement des constituant des jets . . . . .	61
4.3.3 La masse des jets . . . . .	63

4.3.4	Les algorithmes de différentiation de l'empilement s'appliquant aux constituants . . . . .	64
4.3.4.1	Constituent Subtraction . . . . .	65
4.3.4.2	SoftKiller . . . . .	66
4.3.5	Les algorithmes de grooming . . . . .	66
4.3.5.1	Trimming . . . . .	66
4.3.5.2	Soft-drop . . . . .	67
<b>5</b>	<b>L'étalonnage des jets</b>	<b>69</b>
5.1	Les Unified Flow Objects . . . . .	70
5.2	La chaîne d'étalonnage des jets . . . . .	75
5.2.1	La correction d'empilement . . . . .	76
5.2.2	La correction en énergie, masse et pseudo-rapacité . . . . .	78
5.2.3	Global Sequential Calibration . . . . .	82
5.2.4	La correction « <i>in-situ</i> » . . . . .	86
5.3	Performances et étalonnage des UFOs de rayon $R=0,4$ . . . . .	87
5.3.1	Échantillons utilisés . . . . .	87
5.3.2	Sélections . . . . .	88
5.3.2.1	Distributions en impulsion transverse vraie . . . . .	89
5.3.3	Résultats avant étalonnage . . . . .	90
5.3.3.1	Efficacité et pureté des jets . . . . .	90
5.3.3.2	Dépendance de l'impulsion transverse en fonction de variables d'empilement . . . . .	95
5.3.3.3	Échelle en énergie et en masse . . . . .	97
5.3.3.4	Résolution angulaire . . . . .	100
5.3.4	Résultats après correction de l'empilement . . . . .	101
5.3.5	Résultats après correction absolue en énergie, masse et pseudo-rapacité	112
5.3.6	Résultats après GSC . . . . .	117
5.4	Conclusion . . . . .	121
<b>6</b>	<b>La recherche d'un secteur sombre</b>	<b>123</b>
6.1	Description des modèles étudiés . . . . .	124
6.2	Échantillons utilisés . . . . .	129
6.2.1	Données réelles . . . . .	129
6.2.2	Simulations du bruit de fond . . . . .	130
6.2.3	Simulations des signaux . . . . .	130
6.3	Reconstruction des jets . . . . .	132
6.4	Pré-sélections . . . . .	135
6.4.1	Nettoyage des évènements . . . . .	135
6.4.2	Sélection du vertex primaire . . . . .	136
6.4.3	Système de déclenchement . . . . .	136
6.4.4	Nettoyage des jets . . . . .	137
6.4.5	Au niveau de l'analyse . . . . .	137
6.5	Stratégie de l'analyse . . . . .	138
6.6	Optimisation des coupures de sélection . . . . .	141
6.7	Comparaison des données et des simulations . . . . .	156
6.8	Effet de $n_{\text{trk}}^{\epsilon}$ dans les données . . . . .	159



---

6.9	Incertitudes systématiques . . . . .	161
6.9.1	Incertitudes expérimentales . . . . .	162
6.9.1.1	Incertitudes liées à l'étalonnage des jets . . . . .	162
6.9.1.2	Incertitudes liées à la reconstruction des traces au sein des jets . . . . .	168
6.9.1.3	Incertitude sur la luminosité intégrée . . . . .	172
6.9.2	Incertitudes théoriques . . . . .	173
6.9.3	Incertitudes sur les signaux . . . . .	178
6.9.4	Incertitudes sur le bruit de fond . . . . .	178
6.10	Interpolation des signaux . . . . .	181
6.10.1	Ajustement de la section efficace . . . . .	182
6.10.2	Ajustement de l'acceptance $\times$ l'efficacité . . . . .	182
6.10.3	Nombre d'évènements dans la Région de Signal . . . . .	183
6.10.4	Interpolation de la forme des signaux . . . . .	184
6.11	Interprétation statistique . . . . .	186
6.11.1	Statistiques . . . . .	186
6.11.2	Tests de faux signal . . . . .	193
6.11.3	Tests d'injection de signal . . . . .	198
6.11.4	Classement des paramètres de nuisance . . . . .	201
6.11.5	Limites . . . . .	204
6.12	Conclusion . . . . .	204
	<b>Conclusion générale</b>	<b>206</b>

# Introduction

Le Modèle Standard de la physique des particules est une théorie décrivant les particules élémentaires et leurs interactions avec succès, à tel point qu'elle a permis de prédire l'existence du Boson de Higgs, découvert en 2012 par les expériences ATLAS et CMS au Large Hadron Collider (LHC). Cependant, cette théorie présente quelques lacunes, dont le problème de la nature de la matière noire, qui se manifeste à des échelles astronomiques et cosmologiques, et dont une des preuves les plus tangibles réside dans le spectre en anisotropies du Fond Diffus Cosmologique. Plusieurs théories ont été développées afin d'expliquer l'existence de la matière noire, mais aucune n'a jusqu'à présent mené à une découverte permettant de résoudre le problème. Récemment, de nouvelles théories ont été mises au point, et font intervenir l'existence d'un secteur « sombre », en plus de celui du Modèle Standard, dans lequel les particules interagissent fortement entre elles, et qui est relié au Modèle Standard par un médiateur massif.

Dans le Chapitre 1, une description succincte du Modèle Standard est présentée. L'accent est mis sur une partie du Modèle Standard appelée la Chromodynamique Quantique (QCD), qui permet de décrire l'interaction forte, et notamment les *jets* hadroniques, dont l'importance est capitale au sein des collisionneurs de hadrons.

Le Chapitre 2 introduit le problème de la matière noire dans le contexte du Modèle Standard, ainsi que plusieurs hypothèses qui ont été émises afin de résoudre ce problème. Plusieurs méthodes de détection de la matière noire sont ensuite décrites, et en particulier celles au sein des collisionneurs. Enfin, plusieurs modèles de matière noire faisant intervenir un secteur « sombre » de type QCD sont présentés.

Le Chapitre 3 décrit le fonctionnement du Large Hadron Collider, ainsi que celui du détecteur ATLAS.

Ensuite, les méthodes de reconstruction de plusieurs types d'objets, et en particulier les jets, sont exposées dans le Chapitre 4.

Puis, une nouvelle méthode de reconstruction des jets, appelée *Unified Flow Objects* est introduite dans le Chapitre 5. La chaîne d'étalonnage des jets est ensuite détaillée, et les performances des Unified Flow Objects après chacune des étapes de l'étalonnage que j'ai obtenues dans le cadre de cette thèse sont comparées aux définitions de jets jusqu'à présent utilisées dans ATLAS.

Enfin, le Chapitre 6 présente mon travail sur une analyse effectuée sur l'ensemble des

données du Run-2 d'ATLAS cherchant à mettre en évidence un des modèles de secteur sombre, dans lequel un médiateur de type  $Z'$  couple les quarks du Modèle Standard à ceux du secteur sombre. L'analyse recherche une résonance dans le spectre en masse invariante de deux jets après une sélection permettant d'obtenir des données enrichies en jets issus de quarks sombres. Toute la chaîne de l'analyse y est décrite : de la description des modèles étudiés jusqu'à l'interprétation statistique.

# Chapitre 1

## Le Modèle Standard de la physique des particules

Le Modèle Standard de la physique des particules est une théorie permettant de décrire les particules élémentaires, et la façon dont elles interagissent entre elles. Le Modèle Standard permet d'expliquer avec succès de nombreux phénomènes physiques, ce qui en fait la théorie la plus complète développée jusqu'à présent. Une vue d'ensemble des particules du Modèle Standard est représentée Figure 1.1 [1].

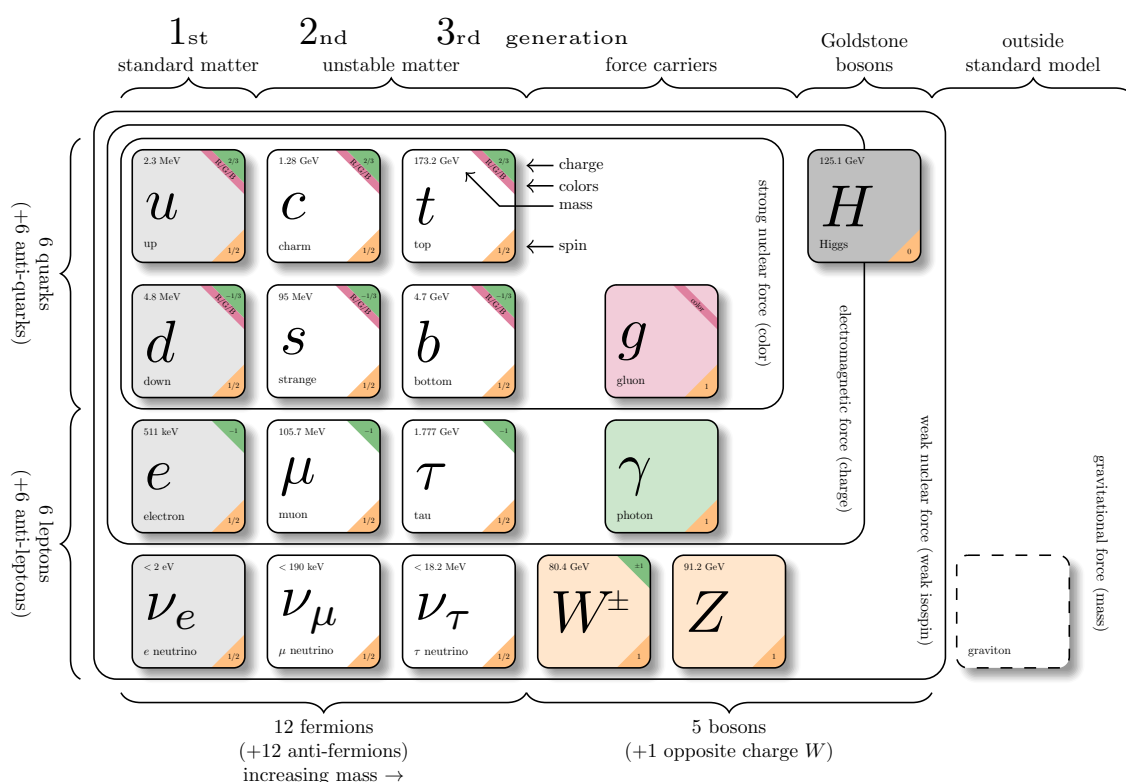


FIGURE 1.1 – Le Modèle Standard de la physique des particules [1].

Le Modèle Standard comprend deux catégories de particules : les fermions, qui com-

posent la matière et sont de spin  $\frac{1}{2}$ , et les bosons, qui sont les médiateurs des interactions entre ces particules et de spin entier.

Trois types d'interactions sont décrites par le Modèle Standard : l'interaction électromagnétique, dont le médiateur est le photon  $\gamma$  de masse nulle, l'interaction forte, dont les médiateurs sont les 8 gluons  $g$  de masse nulle, et l'interaction faible, dont les médiateurs sont les bosons  $W^+$ ,  $W^-$  et  $Z^0$  qui sont massifs. Toutes ces particules sont de spin 1. Le groupe de symétrie de l'interaction forte est  $SU(3)_C$ , et celui de l'interaction électromagnétique et faible est  $SU(2)_L \otimes U(1)_Y$ , où  $SU(2)_L$  est le groupe d'isospin faible agissant uniquement sur les fermions gauches et  $U(1)_Y$  est le groupe d'hypercharge faible.

Les fermions peuvent être décomposés en deux familles de particules : les quarks et les leptons, mais aussi en trois générations.

Les quarks sont composés des quarks *up* :  $u$ ,  $c$  et  $t$  (pour les trois générations de fermions) de charge électrique  $Q = +\frac{2}{3}$ , et des quarks *down* :  $d$ ,  $s$  et  $b$ , de charge électrique  $Q = -\frac{1}{3}$ . Seuls les quarks  $u$  et  $d$  sont stables. Ces quarks forment les protons et les neutrons, qui sont les composants des noyaux atomiques. Les quarks interagissent entre eux par interaction forte, électromagnétique et faible.

Les leptons sont composés de particules de charge électrique  $Q = -1$  : les électrons  $e$ , les muons  $\mu$  et les leptons  $\tau$  (pour les trois générations de fermions), d'un côté, et des neutrinos qui ne sont pas chargés électriquement ( $Q = 0$ ) : les neutrinos électroniques  $\nu_e$ , muoniques  $\nu_\mu$  et tauiques  $\nu_\tau$ , de l'autre. Les leptons chargés interagissent par interaction électromagnétique et faible, tandis que les neutrinos, du fait de leur charge nulle, n'interagissent que par interaction faible, ce qui en fait des particules difficiles à détecter.

Le groupe de symétrie électro-faible  $SU(2)_L \otimes U(1)_Y$  est brisé en  $U(1)_{EM}$  (le groupe de symétrie de l'interaction électromagnétique) par le mécanisme de brisure spontanée de symétrie, qui prédit l'existence d'une nouvelle particule, le boson de Higgs,  $H$ , qui est une particule de spin 0 [2]. Le mécanisme de brisure spontanée de symétrie permet d'expliquer comment les particules massives du Modèle Standard ont acquis leur masse.

Malgré les nombreux succès du Modèle Standard, celui-ci présente quelques lacunes, dont le problème de la matière noire, décrit Section 2.

## 1.1 La Chromodynamique Quantique

Du fait de l'importance capitale de l'interaction forte au sein des collisionneurs de hadrons (et d'autant plus pour les études réalisées dans ce document), l'interaction forte est décrite plus en détail ci-dessous.

La théorie de l'interaction forte est aussi appelée Chromodynamique Quantique (ou *QCD*, pour *Quantum ChromoDynamics* en anglais). Comme décrit auparavant, le groupe de jauge de la chromodynamique quantique est  $SU(3)_C$ , où  $C$  signifie *couleur* (d'où le nom Chromodynamique), ce qui correspond à trois couleurs différentes : *rouge*, *verte* et

*bleue*. Les médiateurs de l'interaction forte sont les 8 gluons de masse nulle.

Le Lagrangien de la QCD est donné par :

$$\mathcal{L}_{\text{QCD}} = \sum_q \bar{q}_i (i\gamma^\mu (D_\mu)_{ij} - m_q \delta_{ij}) q_j - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \quad (1.1)$$

où :

- $q$  représentent les quarks dans la représentation fondamentale de  $\text{SU}(3)$ ,
- $\gamma$  représentent les matrices de Dirac,
- $D$  représente la dérivée covariante,
- $m_q$  représente la masse des quarks,
- $\delta$  représente le delta de Kronecker,
- $G$  représente le tenseur d'intensité de champ des gluons,
- et  $i$  représente l'indice de couleur qui va de 1 jusqu'à  $N_c = 3$ .

La dérivée covariante  $D$  est donnée par :

$$(D_\mu)_{ij} = \partial_\mu \delta_{ij} - ig_s \frac{(\lambda_a)_{ij}}{2} \mathcal{A}_\mu^a \quad (1.2)$$

où :

- $g_s$  représente la constante de couplage de l'interaction forte,
- $\lambda$  représentent les générateurs de  $\text{SU}(3)$  dans leur représentation fondamentale, dont une représentation explicite est donnée par les 8 matrices de Gell-Mann,
- et  $\mathcal{A}$  représente le champs des gluons.

Le tenseur d'intensité de champ des gluons  $G$  est donné par :

$$G_{\mu\nu}^a = \partial_\mu \mathcal{A}_\nu^a - \partial_\nu \mathcal{A}_\mu^a + g_s f^{abc} \mathcal{A}_{\mu b} \mathcal{A}_{\nu c} \quad (1.3)$$

où  $f^{abc}$  représentent les constantes de structure de  $\text{SU}(3)$  ( $[\lambda^a, \lambda^b] = 2if^{abc}\lambda_c$ ).

Afin de garantir l'invariance du Lagrangien sous des transformations de jauge, le champ des gluons doit se transformer selon :

$$\mathcal{A}_\mu^a \rightarrow \mathcal{A}_\mu^a - \frac{1}{g_s} \partial_\mu \theta^a + f^{abc} \theta_b \mathcal{A}_{\mu c} \quad (1.4)$$

Une propriété de la QCD est que, contrairement aux interactions électromagnétique et faible, les gluons portent eux aussi une charge de couleur (plus précisément, ils portent une charge couleur et une charge d'anti-couleur). Cela implique que les gluons peuvent interagir entre eux, et donc qu'il existe des diagrammes de Feynman avec des vertex à 3 ou 4 gluons, comme représenté Figure 1.2 [3].



FIGURE 1.2 – Diagrammes de Feynman avec des vertex à 3 et 4 gluons [3].

Une autre différence majeure entre la QCD d'une part et les interactions électromagnétique et faible d'autre part, est que la constante de couplage de l'interaction forte,  $\alpha_s$  ( $= \frac{g_s}{4\pi}$ ), est grande (de l'ordre de l'unité). Par conséquent les processus de QCD (à basse énergie) ne peuvent pas être calculés avec la théorie des perturbations. Cependant,  $\alpha_s$  n'est pas vraiment une constante, mais sa valeur dépend en fait de l'échelle d'énergie considérée. L'évolution de  $\alpha_s$  en fonction de l'impulsion transférée,  $Q$ , est donnée par :

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \frac{11N_c - 2N_f}{12\pi} \alpha_s(\mu^2) \ln\left(\frac{Q^2}{\mu^2}\right)} \quad (1.5)$$

où :

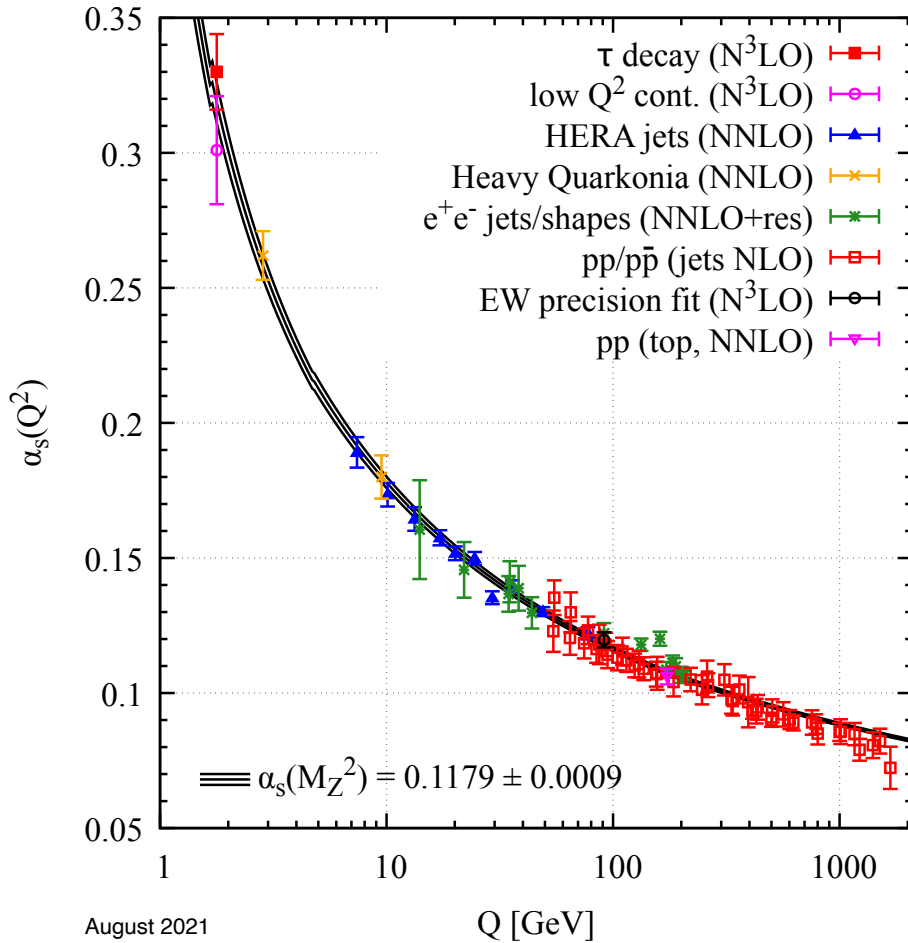
- $\mu$  représente une énergie de référence,
- $N_c$  représente le nombre de couleurs, qui est égal à 3,
- et  $N_f$  représente le nombre de saveurs de quarks, qui est égal à 6.

Ainsi,  $11N_c - 2N_f > 0$ , ce qui implique que  $\alpha_s$  décroît avec  $q^2$ . Cette propriété de la QCD est appelée la *liberté asymptotique*. Par exemple, la valeur de  $\alpha_s$  mesurée à  $|Q^2| = M_Z^2$  est [4] :

$$\alpha_s(M_Z^2) = 0,1179 \pm 0,0009 \quad (1.6)$$

C'est pourquoi à de hautes énergies,  $\alpha_s$  devient suffisamment faible pour que la théorie des perturbations puisse être utilisée. Cependant, même aux échelles d'énergie des grands collisionneurs de hadrons, la valeur de  $\alpha_s(Q^2)$  reste de l'ordre de grandeur de 0,1, ce qui implique que les corrections de second ordre en théorie des perturbations ont un impact significatif en QCD.

L'évolution de  $\alpha_s$ , mesurée dans les données, est représentée Figure 1.3 [4].


 FIGURE 1.3 – Évolution de  $\alpha_s$  en fonction de  $Q$  [4].

En introduisant l'échelle de confinement  $\Lambda$ , définie par :

$$\ln(\Lambda^2) = \ln(\mu^2) - \frac{12\pi}{(11N_c - 2N_f)\alpha_s(\mu^2)} \quad (1.7)$$

$\alpha_s$  peut aussi être formulée selon :

$$\alpha_s(Q^2) = \frac{12\pi}{(11N_c - 2N_f) \ln\left(\frac{Q^2}{\Lambda^2}\right)} \quad (1.8)$$

Les quarks n'ont jamais été observés directement expérimentalement, mais toujours sous forme de hadrons qui ont une charge de couleur nulle, les hadrons correspondant à des particules composées de quarks ou d'anti-quarks. Les hadrons peuvent être classifiés en deux catégories : les mésons, composés d'un quark et d'un anti-quark, et les baryons, composés de trois quarks. Cette observation peut être expliquée par une hypothèse appelée *confinement de couleur*, selon laquelle les objets portant une charge de couleur ne peuvent pas se propager en tant que particules libres, mais qu'ils sont toujours confinés à des états de singlet de couleur. Le confinement de couleur viendrait du fait que les gluons peuvent interagir entre eux, bien qu'il n'y ait jusqu'à présent pas de preuve analytique de ce confinement.



Une conséquence du confinement de couleur et de la liberté asymptotique est que les quarks et les gluons produits au sein des collisionneurs sont observés sous forme de *jets* de particules sans charge de couleur. En effet, les quarks et les gluons émettent des gluons virtuels qui peuvent à leur tour émettre des gluons ou créer des paires quark - anti-quark. Ce processus est appelé la gerbe de partons (*parton shower* en anglais), les partons décrivant les quarks et les gluons. Ensuite, lors de cette gerbe de partons, l'échelle d'interaction diminue et le couplage de l'interaction forte augmente, ce qui déclenche éventuellement le processus d'*hadronisation*, lors duquel les partons se joignent pour former des hadrons de couleur nulle. L'hadronisation se produit à une échelle de distance d'environ  $10^{-15}$  m. Ce sont ces hadrons qui sont enfin enregistrés dans le détecteur. Un exemple de production d'un jet suite à une collision entre hadrons est représenté Figure 1.4 [5].

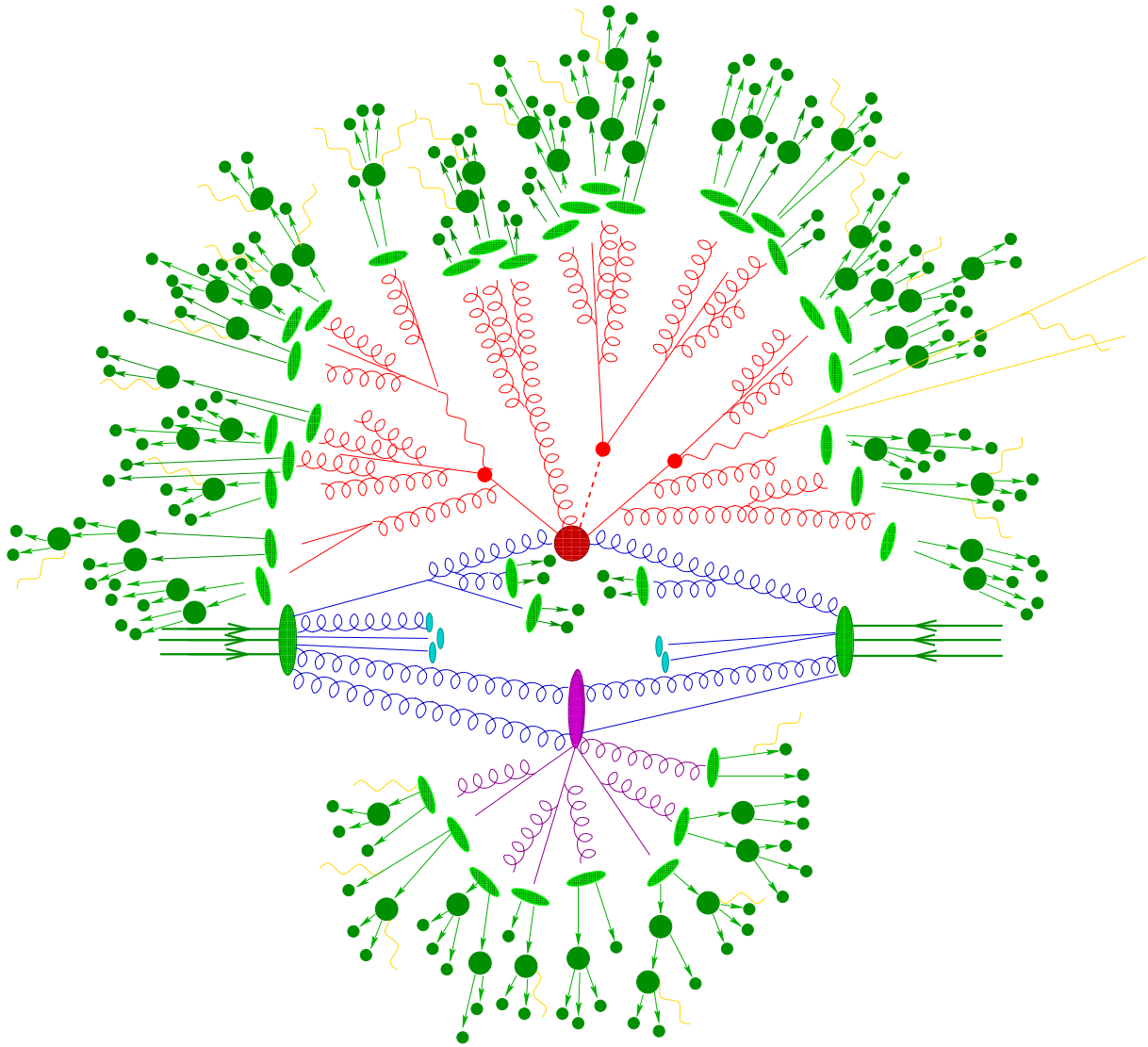


FIGURE 1.4 – Production d’un jet suite à une collision entre hadrons [5]. Le point rouge au centre correspond à la collision principale. La structure rouge en forme d’arbre représente la gerbe de partons. L’hadronisation est représentée en vert. Les hadrons produits peuvent ensuite se désintégrer. Les lignes jaunes représentent des émissions de photons. Le point violet indique une collision secondaire se produisant lors du même évènement (ce qui est appelé l’ « *underlying event* » en anglais).

# Chapitre 2

## La matière noire

### Sommaire du Chapitre

---

2.1	Preuves de l'existence de la matière noire . . . . .	14
2.2	Caractéristiques de la matière noire . . . . .	19
2.3	Candidats à la matière noire . . . . .	19
2.4	Méthodes de détection de la matière noire . . . . .	21
2.5	Secteur sombre . . . . .	24

---

### 2.1 Preuves de l'existence de la matière noire

La première indication de la présence de matière noire a été découverte par l'astronome suisse Fritz Zwicky en 1933 [6]. En comparant deux estimations de la masse de l'amas de la Chevelure de Bérénice (ou amas de Coma), l'une calculée avec la vitesse des galaxies proches de son extérieur, l'autre basée sur sa luminosité, il a remarqué que celles-ci étaient incompatibles. Par conséquent, l'amas de la Chevelure de Bérénice devait contenir une matière invisible aux observations, ou matière noire.

Mais ce n'est qu'à partir des années 1970, grâce au travail de l'astronome américaine Vera Rubin [7], que l'existence potentielle de la matière noire fut prise en considération sérieusement. Pour ce faire, elle a étudié les courbes de rotation de galaxies spirales. Ces courbes de rotation représentent la vitesse  $v$  à laquelle certains objets galactiques visibles tournent autour du centre de la galaxie, en fonction de leur distance  $r$  à celui-ci. Selon la mécanique Newtonienne, la relation entre  $v$  et  $r$  est donnée par :

$$v(r) \sim \sqrt{\frac{M(r)}{r}} \tag{2.1}$$

où  $M(r)$  représente la masse contenue à l'intérieur du rayon  $r$ . Pour un rayon supérieur à la partie visible de la galaxie, la vitesse devrait donc décroître en  $\frac{1}{\sqrt{r}}$ . Or les observations de Vera Rubin montrent que cette vitesse atteint plutôt un plateau à grande distance, comme représenté Figure 2.1 pour la galaxie NGC 6503, qui provient d'études plus récentes [8].

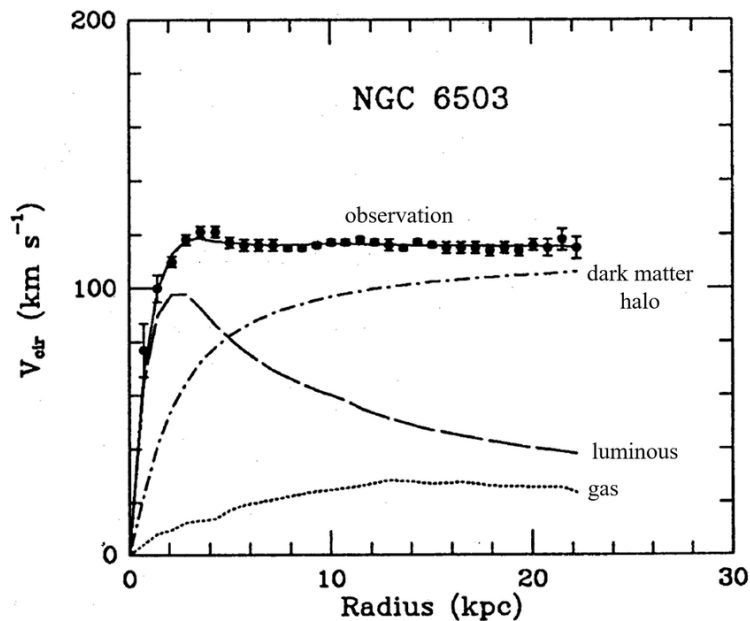


FIGURE 2.1 – Courbe de rotation de la galaxie NGC 6503 [8].

Ces observations impliquent l'existence d'un halo de matière noire qui engloberait les galaxies, s'étendant bien au-delà de leur partie visible.

Une autre preuve de l'existence de la matière noire provient de l'amas de la Balle, qui est un amas formé par la collision de deux autres amas de galaxies, et dont une image est représentée Figure 2.2 [9]. Les étoiles des amas, directement visibles par leur lumière émise, ne sont que très faiblement affectées par la collision. En revanche, les gaz chauds des amas, qui représentent la majeure partie de la matière baryonique, et observés aux rayons X, interagissent électromagnétiquement, ce qui les ralentit beaucoup plus que les étoiles. L'effet de lentille gravitationnelle se produisant sur des objets en arrière plan de l'amas de la Balle permet de mesurer la distribution de la masse à l'intérieur de l'amas. Les observations montrent que la masse est plus forte là où se trouvent les étoiles des galaxies. Cela indique que lorsque les amas se sont croisés, la matière noire contenue dans ceux-ci n'a que très peu interagi avec la matière baryonique autrement que par interaction gravitationnelle. L'amas de la Balle permet aussi de réfuter des théories alternatives à la matière noire, telle que la théorie MOND (pour *Modified Newtonian dynamics*), proposée par Mordehai Milgrom en 1983 [10]. En effet, selon la théorie MOND, la distribution de masse devrait coïncider avec celle du gaz. Enfin, l'amas de la Balle permet de poser des contraintes sur la section efficace de l'auto-interaction de la matière noire. Selon ces contraintes, la section efficace de l'auto-interaction de la particule de matière noire, divisée par sa masse,  $\frac{\sigma}{m}$ , devrait être inférieure à  $1 \text{ cm}^2\text{g}^{-1}$  [11].

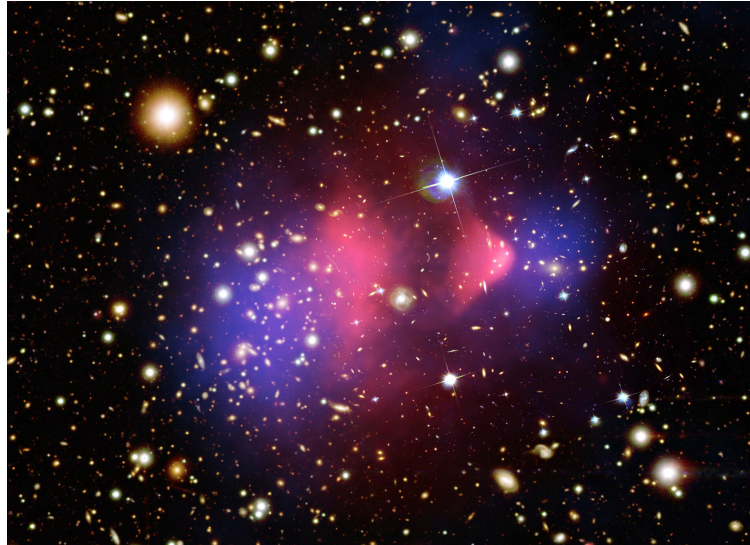


FIGURE 2.2 – Image de l’amas de la Balle [9]. La partie rose correspond au gaz chaud, tel que pris par le télescope Chandra aux rayons X. La partie bleue correspond à la distribution de la masse, détectée par effet de lentille gravitationnelle.

De nos jours, la source d’informations la plus importante pour étudier les propriétés de la matière provient probablement du fond diffus cosmologique (ou *CMB*, pour *Cosmic Microwave Background* en anglais), qui est un rayonnement électromagnétique, dont le pic d’émission est situé dans le domaine des micro-ondes. Ce rayonnement a été émis 380 000 ans après le Big Bang, alors que l’univers refroidit suffisamment pour que les protons et les électrons puissent former des atomes d’hydrogène. Les atomes ne pouvaient alors plus diffuser le rayonnement thermique. C’est ainsi que l’univers est devenu transparent aux photons, produisant le fond diffus cosmologique. Le fond diffus cosmologique a été découvert par accident en 1965 par Arno Penzias et Robert Wilson [12]. Puisque l’univers est en expansion, la radiation du fond diffus cosmologique a subi un décalage vers le rouge (ou *redshift* en anglais). Le profil en énergie des photons observés suit de façon presque parfaite celui d’un rayonnement de corps noir, avec une température  $T = 2,725\text{ K}$ . Alors que le rayonnement du fond diffus cosmologique est presque parfaitement isotrope, de légères anisotropies (de l’ordre de  $\frac{\Delta T}{T} \simeq 10^{-5}$ ) subsistent, comme représenté Figure 2.3 [13]. Ces fluctuations sont causées par des variations de la densité de matière au moment du découplage entre la matière et la radiation. La matière noire n’interagit pas avec la radiation, mais elle affecte le fond diffus cosmologique, car elle interagit gravitationnellement.

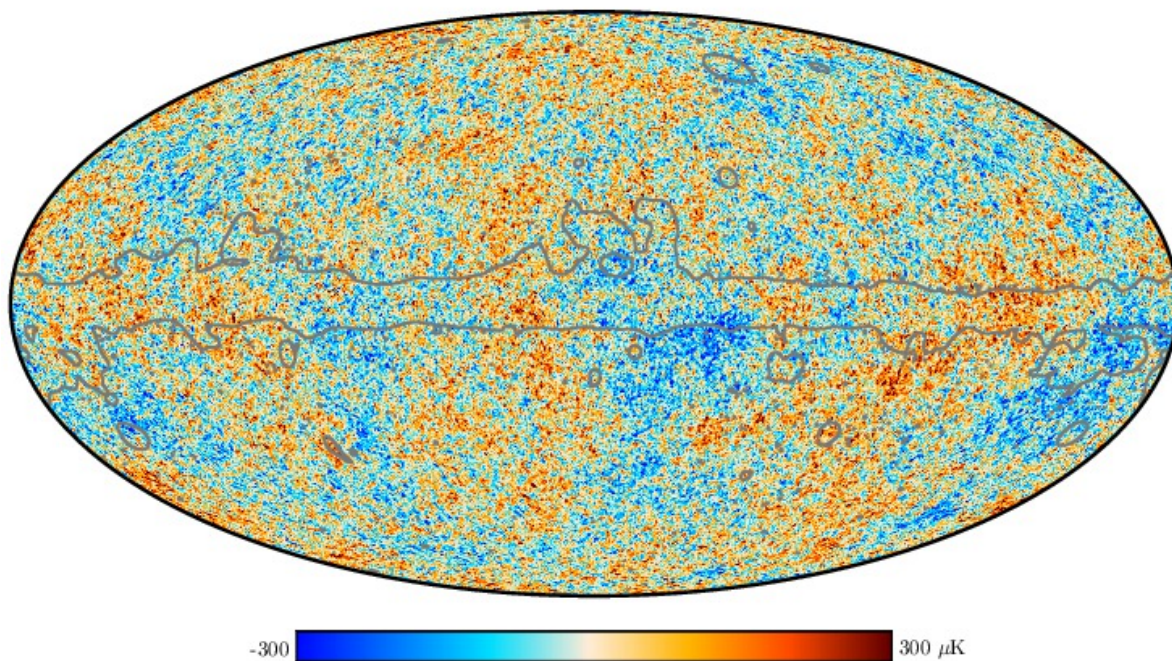


FIGURE 2.3 – Anisotropies en température du fond diffus cosmologique, observées par le satellite Planck [13].

La carte des anisotropies du fond diffus cosmologique peut être décomposée en un spectre de puissance angulaire. La position et la hauteur relative des différents pics observés dans ce spectre en fonction de l'échelle angulaire permet d'obtenir des informations sur la géométrie de l'univers et sur sa composition. Par exemple, le premier pic indique surtout la densité de matière baryonique, tandis que le troisième pic indique surtout la densité de matière noire. Un exemple de spectre de puissance angulaire de la température du fond diffus cosmologique est représenté Figure 2.4 [14].



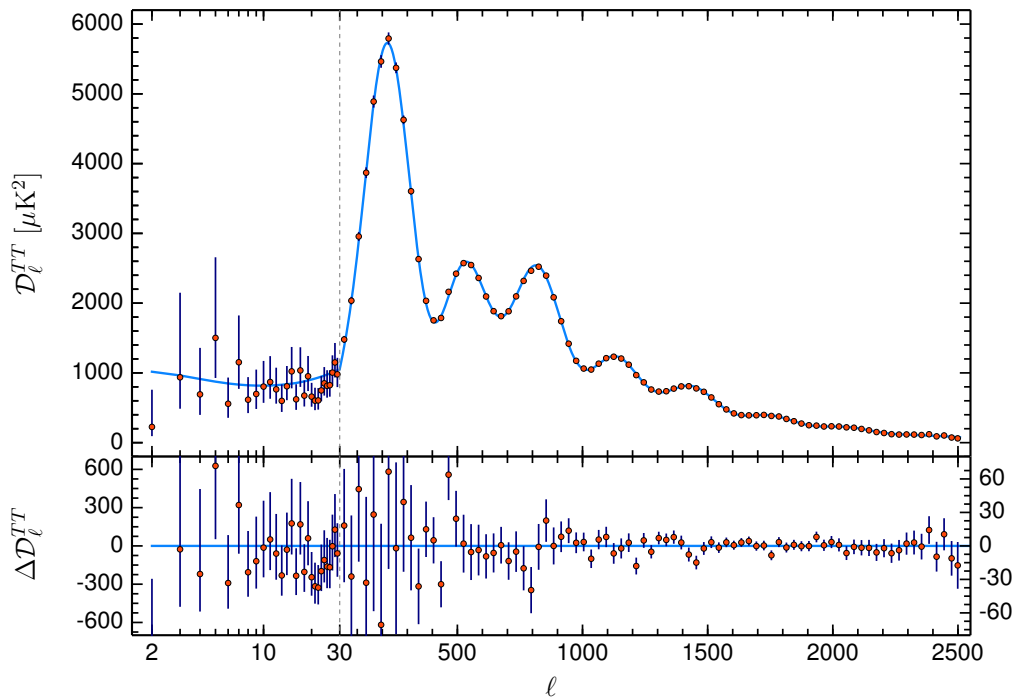


FIGURE 2.4 – Spectre de puissance angulaire de la température du fond diffus cosmologique, mesuré par le satellite Planck [14].

Un ajustement du spectre de puissance angulaire de la température du fond diffus cosmologique avec les paramètres du modèle  $\Lambda$ CDM (un modèle cosmologique avec une constante cosmologique et de la matière noire *froide*) permet d'estimer la densité de matière noire et de matière baryonique. Leurs valeurs, mesurées par le satellite Planck [14], sont :

- $\Omega_c h^2 = 0,120 \pm 0,001$  pour la densité de matière noire *froide*,
- et  $\Omega_b h^2 = 0,0224 \pm 0,0001$  pour la densité de matière baryonique.

En d'autres termes, la matière de l'univers serait composée d'environ 15% de matière ordinaire, et d'environ 85% de matière noire.

D'autres observations permettent de corroborer l'existence de la matière noire, parmi lesquelles :

- la formation des structures : si l'univers peu après le Big Bang n'était composé que de matière baryonique, il aurait été très homogène, car la matière ordinaire est affectée par le rayonnement qui empêche les particules de se regrouper. Ainsi, les structures que nous observons aujourd'hui se seraient formées plus tard qu'observé. La matière noire permet de résoudre ce problème, car elle n'interagit que très faiblement avec le rayonnement, ce qui permet aux perturbations de croître.
- les oscillations acoustiques des baryons : le fond diffus cosmologique permet de mettre en évidence que des ondes acoustiques se propageaient dans le plasma pri-

mordial. La taille de ces ondes permet de déterminer la proportion de matière noire et de matière baryonique.

## 2.2 Caractéristiques de la matière noire

D'après les observations décrites Section 2.1, plusieurs contraintes sur les propriétés de la matière noire (si celle-ci est composée de particules) peuvent être émises :

- elle doit être massive ;
- elle doit être stable ou avec un temps de vie de l'ordre de l'âge de l'univers, car les preuves de son existence concernent tout aussi bien les temps présents que le début de l'univers ;
- elle doit être électriquement neutre, car autrement nous l'observerions ;
- elle doit être non baryonique, d'après les observations du fond diffus cosmologique ;
- elle doit être interagissante avec les particules du Modèle Standard avec une intensité plus faible ou égale à celle de l'interaction faible ;
- et enfin elle doit être froide, c'est-à-dire non relativiste au moment de la formation des structures afin de permettre aux petites structures (galaxies, petits amas de galaxies) de se former en premier.

Cependant, aucune particule du Modèle Standard ne correspond à ces critères, et par conséquent, la nature de la matière noire demeure à ce jour un mystère. La matière noire serait donc une indication de physique au-delà du Modèle Standard.

## 2.3 Candidats à la matière noire

Plusieurs hypothèses ont été émises quant à la nature de la matière noire.

Le premier type d'hypothèse concerne une modification de la théorie de la gravitation, comme c'est le cas de la théorie MOND, déjà mentionnée Section 2.1. Cette théorie repose sur une modification de la gravitation Newtonienne dans la limite des faibles accélérations. Elle permet d'expliquer les courbes de rotation des galaxies spirales, mais peine à décrire les observations de l'amas de la Balle. Plus récemment la théorie *Tensor-vector-scalar gravity* (TeVeS) a été introduite par Jacob Bekenstein en 2004 [15], et représente une généralisation relativiste de la théorie MOND. Cependant, la théorie TeVeS peine elle aussi à décrire les observations de l'amas de la Balle.

D'autres théories émettent l'hypothèse que la matière noire serait composée d'objets macroscopiques, tels que les *MaCHOs* (pour *Massive Compact Halo Objects* en anglais), ou les trous noirs primordiaux. Les MaCHOs seraient des objets célestes qui n'émettraient pas ou peu de lumière, les rendant difficiles à détecter. Les MaCHOs pourraient être des trous noirs, des étoiles à neutron, ou encore des naines brunes. Cependant, les MaCHOs ne permettent pas d'expliquer les observations du fond diffus cosmologique, qui permet de contraindre la quantité de matière noire par rapport à la matière baryonique. De plus, le phénomène de microlentille gravitationnelle permet de poser des contraintes la masse



des MaCHOs. Les trous noirs primordiaux sont quand à eux un type de trous noirs qui se seraient formés très tôt après le Big Bang. Les trous noirs primordiaux ne sont pas composés de matière baryonique, c'est pourquoi ils représentent des candidats plausibles à la matière noire. Cependant, les observations astrophysiques et cosmologiques permettent d'émettre des contraintes strictes quant à leur abondance dans l'univers.

À ce jour, l'hypothèse privilégiée pour la matière noire est qu'elle serait composée de particules élémentaires, dont les propriétés sont décrites Section 2.2. Parmi ces hypothèses, les *WIMPs* (pour *Weakly Interacting Massive Particles* en anglais) sont les particules hypothétiques candidates à la matière noire qui ont été les plus étudiées jusqu'à présent. En effet, sous l'hypothèse qu'au début de l'univers, toutes les particules baryoniques et de matière noire sont en équilibre thermique, il existerait un moment à partir duquel, sous l'effet de l'expansion et du refroidissement de l'univers, cet équilibre aurait été rompu. La production de matière noire serait alors devenue interdite cinématiquement, car les particules baryoniques n'auraient pas eu assez d'énergie pour créer des particules de matière noire. Cependant, l'annihilation des particules et des antiparticules de matière noire se serait poursuivie, menant à une décroissance exponentielle de la densité des particules de matière noire. Finalement, cette densité serait devenue si faible que l'interaction des particules et des antiparticules de matière noire aurait cessé, et que la densité de particules de matière noire serait restée constante, alors que l'univers poursuivait son expansion. L'évolution de la densité de matière noire en fonction de la masse de la particule de matière noire, divisée par sa température (qui est une procuration pour le temps), est représentée Figure 2.5 [16]. La densité de matière noire résultante dépend de la section efficace d'annihilation, ainsi que de la masse de l'hypothétique particule de matière noire. Ainsi, à partir des mesures actuelles de l'abondance de matière noire et en admettant une particule de matière noire avec une masse de l'ordre de l'échelle électrofaible, la section efficace serait de l'ordre de celle de l'interaction faible. Cette coïncidence est appelée *miracle des WIMPs*.

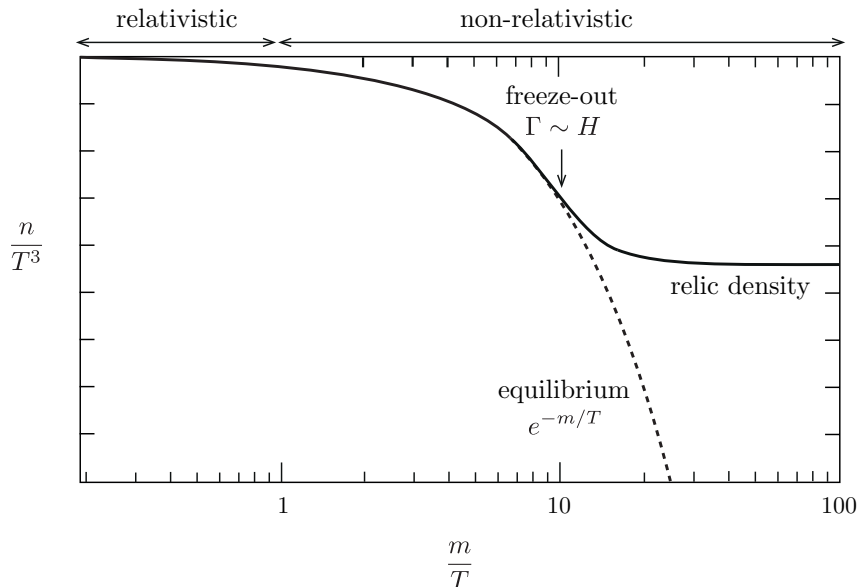


FIGURE 2.5 – Densité de matière noire en fonction du temps [16]. À haute température ( $T \gg m$ ), les particules de matière noire sont en équilibre avec celles du bain. À  $T \simeq m$  les particules de matière noire deviennent non-relativistes, et par conséquent leur densité diminue selon  $e^{-\frac{m}{T}}$ . À basse température ( $T \ll m$ ), la densité de particules de matière noire est *gelée* à cause de l’expansion de l’univers.

Plusieurs théories au-delà du Modèle Standard introduisent naturellement des WIMPs, parmi lesquelles la plus connue est la *supersymétrie* [17], dont la matière noire pourrait par exemple correspondre au *neutralino*.

Cependant, de fortes contraintes sont désormais émises sur la section efficace d’interaction de la matière noire avec les particules du Modèle Standard, ainsi que sur la masse de la particule de matière noire. Bien que ces théories ne soient pas entièrement réfutées, de plus en plus de recherches s’orientent désormais vers de nouvelles théories permettant d’expliquer la matière noire, telles que les axions, les particules de Kaluza-Klein les plus légères, ou encore des modèles faisant intervenir un secteur « caché », dont un exemple est décrit Section 2.5.

## 2.4 Méthodes de détection de la matière noire

Si la matière noire est composée de particules élémentaires pouvant interagir avec celles du Modèle Standard, il existerait alors plusieurs méthodes pour les détecter. Celles-ci sont schématisées Figure 2.6 [18].

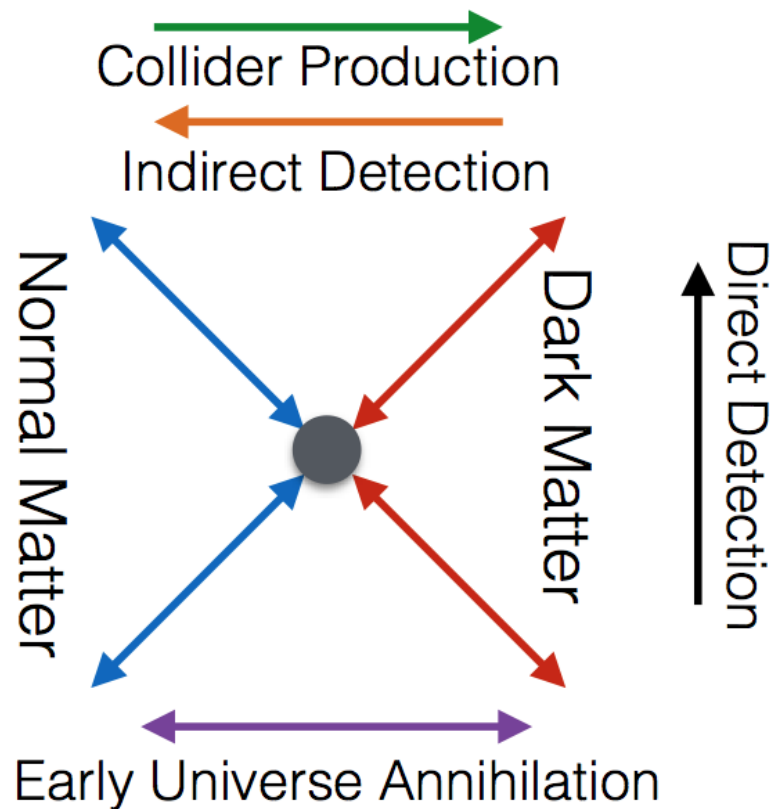


FIGURE 2.6 – Méthodes de détection de la matière noire [18]. La détection directe, correspond à l’interaction d’une particule de matière noire avec une particule du Modèle Standard. La détection indirecte correspond à l’annihilation de particules de matière noire en particules du Modèle Standard, tandis que la production au sein des collisionneurs représente le contraire : des particules de matière noire sont produites par l’interaction de particules du Modèle Standard.

La première méthode, la détection directe, consiste à observer les éventuelles collisions entre une particule de matière noire et un noyau atomique. En effet, si notre galaxie est plongée dans un halo de matière noire sous forme de particules, la Terre devrait être traversée en permanence par un flux de ces particules. Celles-ci pourraient alors entrer en collision avec des noyaux, qui déposeraient alors de l’énergie pouvant être observée par les détecteurs. Pour être efficace, il est essentiel d’obtenir un bruit de fond extrêmement faible. C’est la raison pour laquelle ces expériences sont généralement menées sous terre, où la contamination venant des rayons cosmiques est faible.

Ensuite, la seconde méthode, la détection indirecte, consiste à observer l’éventuelle annihilation de particules de matière noire en photons, neutrinos, protons et anti-protons, électrons et anti-électrons, etc. Cependant, ces annihilations sont rares et il faut alors observer des régions où la matière noire est abondante, telles que les centres de galaxies. La détection indirecte de matière noire consiste à observer ces endroits pour essayer de mettre en évidence un excès de particules détectées, qui aurait un spectre en énergie relié

à la masse de la particule de matière noire.

Enfin, la dernière méthode, la production au sein des collisionneurs, consisterait à produire des particules de matière noire à partir de particules du Modèle Standard. Étant donné les propriétés qu'aurait la matière noire, celle-ci n'interagirait que très peu avec les détecteurs. C'est pourquoi, la signature expérimentale la plus étudiée jusqu'à présent consiste à détecter un excès d'impulsion transverse manquante par rapport aux attentes du Modèle Standard. En effet, dans une configuration telle que celle des grands collisionneurs, les particules initiales se déplacent parallèlement aux faisceaux, et il n'y a donc pas d'impulsion transverse aux faisceaux. Puisque la quantité de mouvement se conserve lors de la collision, l'impulsion transverse du système doit être aussi nulle dans l'état final. Par conséquent, si la somme des impulsions transverse de toutes les particules visibles produites lors de la collision n'est pas nulle, cela pourrait être dû à des particules invisibles produites lors de la collision, et qui transporteraient elles aussi une impulsion. Ce type de signature est représenté Figure 2.7 [19].

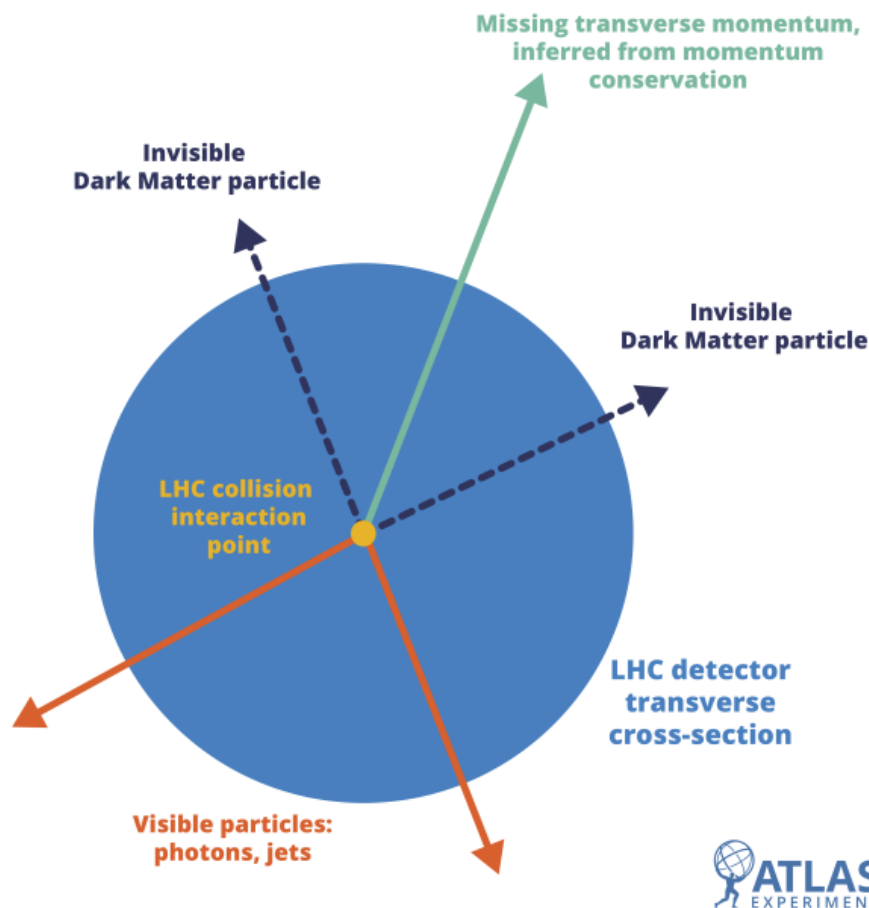


FIGURE 2.7 – Schéma décrivant comment l'impulsion transverse manquante est déterminée dans le plan transversal au détecteur [19].

Cependant, les recherches d'excès en impulsion transverse manquante au sein des grands collisionneurs n'ont jusqu'à présent pas mené à de nouvelles découvertes.

C'est pourquoi les recherches de matière noire au sein des grands collisionneurs se mettent à explorer de nouveaux modèles, tout en continuant les recherches pour les modèles précédents.

## 2.5 Secteur sombre

Un autre scénario pour la matière noire est celui de la matière noire asymétrique. Ce scénario est motivé par la coïncidence entre la quantité de matière baryonique et de matière noire ( $\Omega_c \simeq 5\Omega_b$ ). Selon le scénario de la matière noire asymétrique, la matière noire et son antiparticule de matière noire ne sont pas produites en quantités égales au début de l'univers. Ainsi, l'annihilation entre la matière noire et son anti-particule élimine l'anti-particule, et les particules de matière noire restantes composent la densité relique actuelle [20].

Afin de stabiliser la matière noire et d'annihiler efficacement l'anti-matière noire, un groupe de jauge caché est généralement introduit dans les modèles de matière noire asymétrique.

Si le groupe de jauge caché dans le secteur sombre est  $SU(N_d)$ , ce qui provoque un confinement à une certaine échelle  $\Lambda_d$ , alors un parton sombre énergétique, qui est la particule chargée sous  $SU(N_d)$ , pourrait produire une gerbe de partons sombres qui s'hadroniserait en un amas de hadrons sombres (la plupart d'entre eux correspondraient au méson sombre le plus léger). Les hadrons sombres se désintégreraient ensuite à travers certains portails en particules du Modèle Standard, et produiraient un signal de type jet, tel que décrit par la chromodynamique quantique. Ces jets sont appelés « jets sombres ». La matière noire correspondrait alors aux particules stables de ce secteur caché, les « protons sombres ».

Les propriétés des jets sombres dépendent des paramètres du secteur sombre et du portail entre le secteur sombre et le Modèle Standard. Ainsi, selon ces paramètres, les jets de QCD sombre peuvent produire différentes signatures dans les détecteurs, comme représenté Figure 2.8 [20].

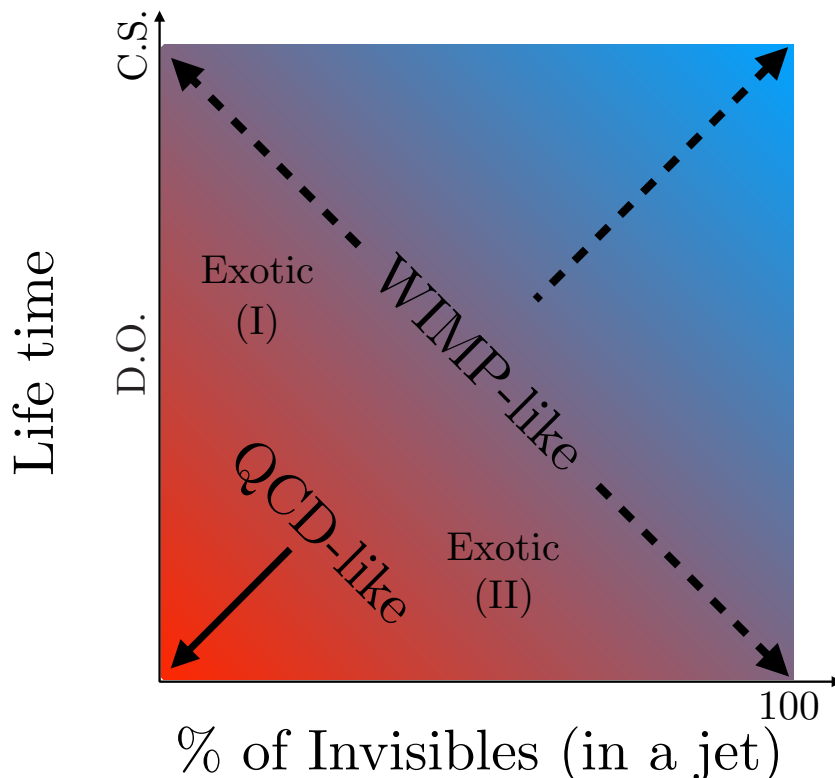


FIGURE 2.8 – Signatures attendues pour les jets de QCD sombre en fonction du pourcentage de hadrons sombres stables (invisibles) dans un jet (axe des abscisses) et du temps de vie des mésons sombres (axe des ordonnées) [20]. Seuls les mésons sombres (et non les baryons sombres) sont considérés pour l’axe des ordonnées, car dans le cadre de  $SU(3)$ , le nombre de baryons produits lors du processus d’hadronisation est négligeable par rapports au nombre de mésons. De plus, le candidat à la matière noire correspondrait au baryon sombre le plus léger (c’est-à-dire le « proton sombre »), et donc il est stable. « *C.S.* » signifie un temps de vie suffisant pour être stable aux collisionneurs (*Collider Stable* en anglais) et « *D.O.* » signifie une durée de vie suffisante pour être identifié comme « objet déplacé » (*Displaced Object* en anglais).

Par exemple, si tous ou la plupart des mésons sombres légers se désintègrent rapidement en particules du Modèle Standard, les jets sombres ressemblent aux jets du Modèle Standard car, dans ce cas, il n’y aurait pas d’objets déplacés ou d’impulsion manquante à l’intérieur des jets sombres.

Au contraire, si les mésons sombres ont un long temps de vie (mais toutefois se désintègrent dans le détecteur), les jets sombres émergeraient dans le détecteur, comme représenté Figure 2.9 [21]. Ces jets sont appelés jets *émergents*.

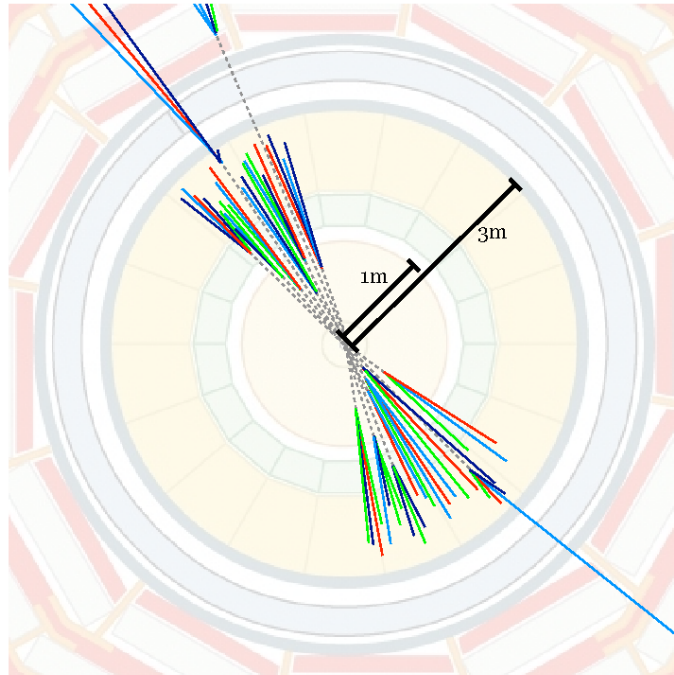


FIGURE 2.9 – Schéma d'une paire de quarks sombres formant deux jets émergents dans le détecteur [21].

Enfin, si une fraction non-négligeable des hadrons sombres ne se désintègrent pas, et restent invisibles, comme représenté Figure 2.10 [22], la signature est celle d'une impulsion transverse manquante alignée avec la direction d'un des deux jets. Ces jets sont appelés jets *semi-visibles*.

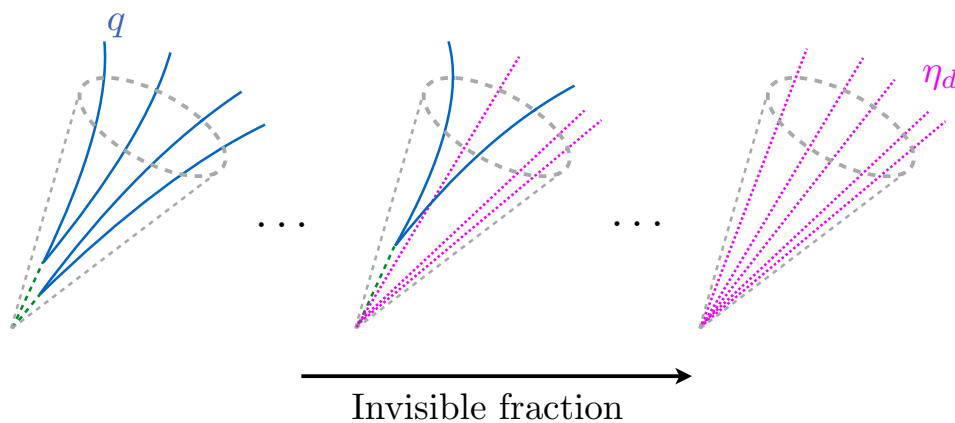


FIGURE 2.10 – Gerbe de partons du secteur sombre. Les hadrons sombres instables (en pointillés verts) se décomposent en quarks du Modèle Standard (en bleu plein). Les hadrons sombres stables aux collisionneurs sont représentés en pointillés roses [22].

Par la suite, nous nous intéresseront plus particulièrement au premier scénario, dans lequel la plupart des mésons sombres se désintègrent rapidement en particules du Modèle Standard, produisant des jets qui ressemblent à ceux du Modèle Standard.

La dynamique dans le secteur sombre est décrite par le groupe de jauge non abélien  $SU(N_d)$ . Plusieurs quarks sombres légers dans la représentation fondamentale de  $SU(N_d)$  sont également nécessaires à la constitution d'un hadron sombre. Un quark sombre léger signifie un quark sombre qui est plus léger que l'échelle de confinement sombre  $\Lambda_d$ . Pour que  $SU(N_d)$  confine, le nombre de saveurs de quarks sombres légers  $N_f$  doit être plus petit que  $\frac{11}{2}N_d$  (voir Équation 1.5). Le Lagrangien du secteur sombre peut être écrit comme suit :

$$\mathcal{L}_d = \bar{q}'_i (i\not{D} - m_{q'_i}) q'_i - \frac{1}{4} G'^{\mu\nu} G'_{\mu\nu} \quad (2.2)$$

où :

- $q'$  représente les quarks sombres,
- $G'^{\mu\nu}$  représentent les gluons sombres,
- $D$  représente la dérivée covariante de  $SU(N_d)$ ,
- et  $i$  représente l'indice de saveur des quarks sombres.

Afin de produire aux collisionneurs des partons sombres qui se désintègrent en hadrons sombres, un médiateur entre le secteur sombre et le Modèle Standard est nécessaire. Il pourrait par exemple s'agir d'un boson vecteur lourd,  $Z'$ , médiateur entre la paire de quarks sombres et la paire de quarks du Modèle Standard :

$$\mathcal{L}_{\text{med}} = -\frac{1}{4} Z'^{\mu\nu} Z'_{\mu\nu} - \frac{1}{2} M_{Z'}^2 Z'^\mu Z'_\mu + Z'_\mu (\bar{q}'_i \gamma^\mu q'_i + \bar{q}_j \gamma^\mu q_j) \quad (2.3)$$

où :

- $q$  représentent les quarks du Modèle Standard,
- et  $i$  et  $j$  sont les indices de saveur des quarks sombres et des quarks du Modèle Standard.

Le diagramme de Feynman de la production de quarks sombres à partir de la désintégration du  $Z'$  est représenté Figure 2.11, adaptée de [23].



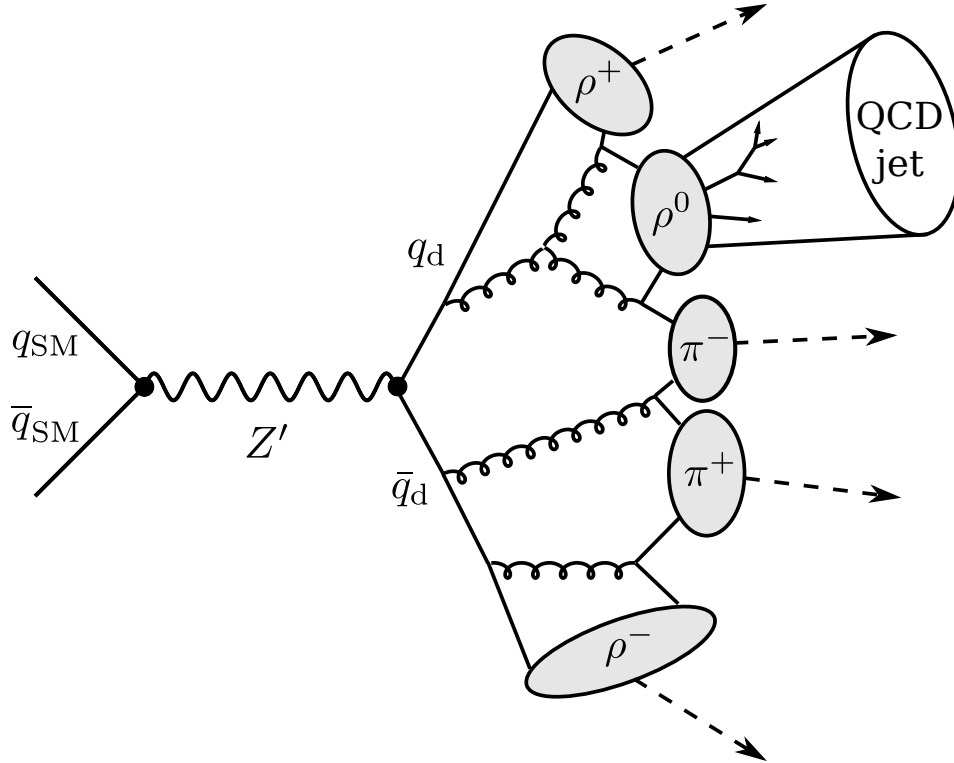


FIGURE 2.11 – Diagramme de Feynman de quarks du Modèle Standard formant le médiateur  $Z'$ , qui se désintègre ensuite en quarks du secteur sombre. Adapté de [23]. Les  $\rho^+$ ,  $\rho^-$ ,  $\rho^0$ ,  $\pi^+$  et  $\pi^-$  sont ceux du secteur sombre. Leur charge ne sont pas celles du Modèle Standard.

Pour que les mésons sombres se désintègrent rapidement, le groupe de jauge caché peut être étendu de  $SU(N_d)$  à  $SU(N_d) \times U(1)'$  (dont le boson de jauge serait le photon sombre  $\gamma'$ ), avec les quarks sombres également chargés sous  $U(1)'$ . Ainsi, les mésons sombres légers pourraient se désintégrer en paire de photons sombres  $\gamma'$  immédiatement, et les  $\gamma'$  se désintégreraient en particules du Modèle Standard. Ainsi, tous les mésons sombres se désintègrent rapidement, c'est-à-dire que la fraction de particules stables est supposée être négligeable.

# Chapitre 3

## Le détecteur ATLAS

### Sommaire du Chapitre

---

<b>3.1</b>	<b>Le CERN et le LHC</b> . . . . .	<b>29</b>
<b>3.2</b>	<b>Le détecteur ATLAS</b> . . . . .	<b>33</b>
3.2.1	Systèmes de coordonnées . . . . .	34
3.2.2	Description des sous-détecteurs . . . . .	35
<b>3.3</b>	<b>La prise de données</b> . . . . .	<b>45</b>
<b>3.4</b>	<b>La qualité des données</b> . . . . .	<b>46</b>
<b>3.5</b>	<b>Les simulations Monte Carlo</b> . . . . .	<b>46</b>

---

### 3.1 Le CERN et le LHC

Le *CERN* (pour *Conseil européen pour la recherche nucléaire*, désormais *Organisation européenne pour la recherche nucléaire*) est une organisation de recherche qui exploite le plus grand laboratoire de physique des particules au monde. Créée en 1954, l'organisation se situe à quelques kilomètres de Genève, sur la frontière franco-suisse.

La principale fonction du CERN est de fournir les accélérateurs de particules et autres infrastructures nécessaires à la recherche en physique des hautes énergies. Le CERN est le site du grand collisionneur de hadrons (ou *LHC*, pour *Large Hadron Collider* en anglais), le plus grand et le plus puissant collisionneur de particules au monde. Le LHC est un collisionneur circulaire proton-proton (*pp*) mais aussi d'ions lourds. D'une circonférence de 27 km, il est installé à une profondeur moyenne de 100 m sous terre.

Avant d'être injectées dans le LHC, les particules sont accélérées par une série de systèmes qui augmentent successivement leur énergie. La chaîne des accélérateurs est représentée Figure 3.1 [24]. Dans un premier temps, les atomes d'hydrogène sont ionisés et les protons sont accélérés jusqu'à 50 MeV par l'accélérateur linéaire LINAC2. Ils sont ensuite injectés dans le *Proton Synchrotron Booster* (ou plus brièvement *PSB*) et sont accélérés jusqu'à 1,4 GeV avant d'entrer dans le synchrotron à protons (ou *PS*, pour *Proton Synchrotron* en anglais). Ils quittent le PS avec une énergie de 26 GeV et sont

injectés dans le super synchrotron à protons (ou *SPS*, pour *Super Proton Synchrotron* en anglais), qui constitue la dernière étape avant le LHC. Le SPS peut accélérer les protons jusqu'à une énergie de 450 GeV. Enfin, les protons sont accélérés au sein même du LHC, jusqu'à atteindre une énergie de 6,5 TeV par faisceau durant le Run 2 du LHC (c'est-à-dire entre 2015 et 2018) et 6,8 TeV en 2022.

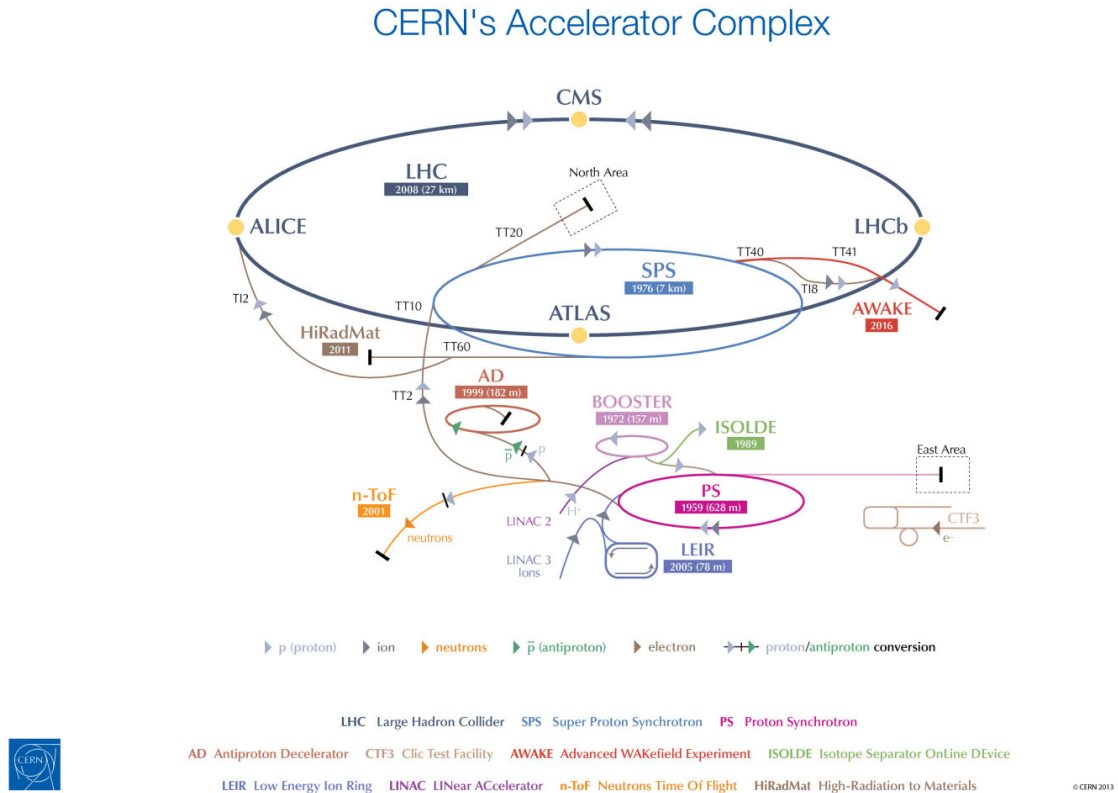


FIGURE 3.1 – La chaîne des accélérateurs du CERN [24].

Les collisions se produisent à 4 différents emplacements sur l'anneau du LHC, qui correspondent à 4 détecteurs de particules : *ALICE* [25], *ATLAS* [26], *CMS* [27] et *LHCb* [28]. *ATLAS* et *CMS* sont des détecteurs polyvalents, optimisés pour les interactions proton-proton, afin d'effectuer des mesures de précision du Modèle Standard et de rechercher une nouvelle physique, au-delà du Modèle Standard. Plus d'informations sur le détecteur *ATLAS* sont données Section 3.2. Le détecteur *ALICE* est optimisé pour l'étude des interactions entre ions lourds, tandis que *LHCb* est destiné à l'étude particulière des désintégrations des quarks *b*. Une collision se produit environ toutes les 25 ns à ces points d'interaction.

Le nombre d'évènements se produisant par seconde pour un processus donné est égal au produit de la section efficace du processus considéré,  $\sigma$ , multipliée par une quantité, appelée la luminosité instantanée,  $\mathcal{L}_{\text{inst}}$ , comme décrit par la formule suivante :

$$\frac{dN_{\text{events}}}{dt} = \mathcal{L}_{\text{inst}} \times \sigma \quad (3.1)$$

Ainsi, plus la luminosité instantanée est élevée, plus les processus étudiés se produisent. C'est pourquoi il est important de maximiser la luminosité instantanée.

La luminosité instantanée est une quantité qui dépend des paramètres des faisceaux de protons et qui est donnée par :

$$\mathcal{L}_{\text{inst}} = \frac{N_b^2 f F}{4\pi\sigma_x\sigma_y} \quad (3.2)$$

où :

- $N_b$  représente le nombre de protons par paquet de protons, qui est environ égal à  $1,2 \times 10^{11}$ ,
- $f$  représente la fréquence de croisement des paquets, qui, en théorie, est égale à 40 MHz.  $f$  dépend du nombre de paquets de protons dans un faisceau, qui, en théorie, est égal à 3600. En réalité, certains paquets de protons sont vides, et seulement 2808 sont remplis. Par conséquent, la fréquence réelle obtenue est plus faible que 40 MHz,
- $F$  représente le facteur de réduction géométrique de la luminosité (entre 0,3 et 1),
- et  $\sigma_x$  et  $\sigma_y$  représentent la taille des faisceaux aux points d'interaction dans les directions transversales au plan de collision (entre 10 et 20  $\mu\text{m}$ ).

La luminosité instantanée était environ de  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$  pour le Run 2.

Une autre quantité d'importance est la luminosité intégrée,  $\mathcal{L}_{\text{int}}$ , qui représente la quantité de données délivrée par le LHC et qui est donnée par l'intégrale sur une période de temps donnée de la luminosité instantanée :

$$\mathcal{L}_{\text{int}} = \int \mathcal{L}_{\text{inst}} dt \quad (3.3)$$

La luminosité intégrée est exprimée en inverse *barns* ( $\text{b}^{-1}$ ), le *barn* étant une unité de section efficace ( $1 \text{ b} = 10^{-24} \text{ cm}^2 = 10^{-28} \text{ m}^2$ ). Habituellement la luminosité est exprimée en inverse *femto-barns*,  $\text{fb}^{-1}$ , afin de correspondre aux quantités de données délivrées par le LHC.

L'évolution de la luminosité intégrée en fonction du temps entre 2015 et 2018, ce qui correspond au Run 2 du LHC, est représentée Figure 3.2 [29]. La courbe de luminosité présente plusieurs phases : les périodes d'arrêt du LHC, qui correspondent aux plateaux et lors desquelles les données ne sont pas collectées, et des périodes lors desquelles le LHC fonctionne, ce qui permet la prise de données et ce qui mène à un accroissement de la luminosité. Ces périodes correspondent aux années 2015-2016, 2017 et 2018. Au total, la luminosité intégrée prise dans des conditions optimales sur la période 2015-2018 représente  $139 \text{ fb}^{-1}$  de données.

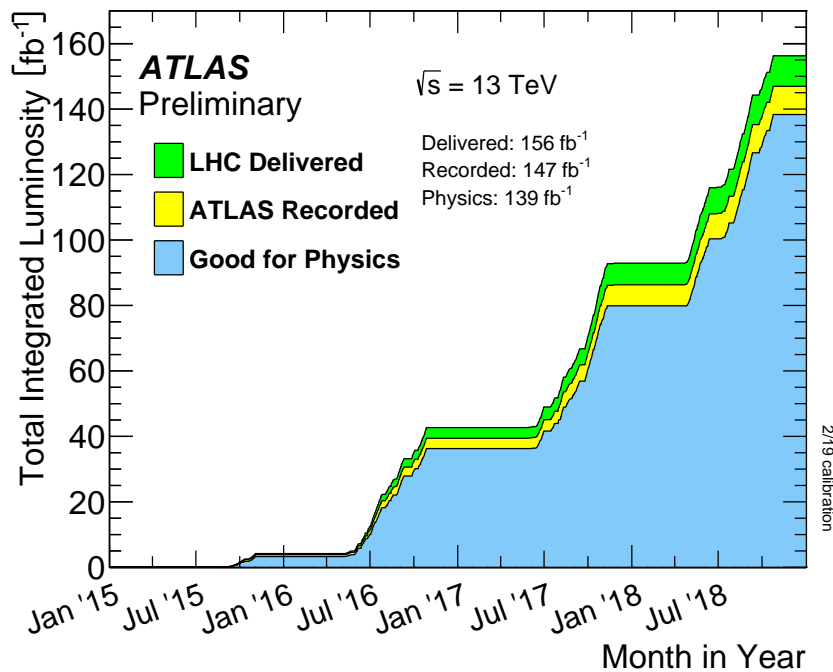


FIGURE 3.2 – Luminosité intégrée totale entre 2015 et 2018 [29].

Sur la Figure 3.2 [29], la luminosité délivrée tient compte de la luminosité délivrée dès lors que les faisceaux sont stables et jusqu'à ce que le LHC demande à ATLAS de mettre le détecteur en mode d'attente sécurisé afin de permettre un vidage des faisceaux ou d'étudier les faisceaux. La luminosité enregistrée reflète l'inefficacité du système d'acquisition des données, l'inefficacité du démarrage à chaud (lorsque le faisceau est stable, le détecteur interne subit une rampe de haute tension), ainsi que l'inefficacité due à la période d'attente. L'évaluation de la qualité des données présentée correspond à l'efficacité *All Good*. Les critères de qualité des données *All Good* exigent que le détecteur soit en bon état de fonctionnement.

Avec l'accroissement de la luminosité instantanée vient un autre phénomène : l'empilement (ou *pile-up* en anglais). L'empilement correspond au fait que d'autres collisions entre protons se produisent, en plus de la collision d'intérêt. Ces collisions se font généralement à basses énergies. Elles se superposent dans le détecteur avec la collision d'intérêt, ce qui crée du bruit, car cela affecte la reconstruction des objets. Deux formes d'empilement sont différenciées : l'empilement hors-temps et l'empilement dans le temps. L'empilement dans le temps correspond aux collisions supplémentaires qui se produisent au sein du même croisement de paquets que la collision d'intérêt. L'empilement dans le temps est mesuré par la variable  $N_{PV}$ , qui correspond au nombre de vertex primaires (*primary vertices* en anglais). Le vertex primaire principal correspond au vertex où la collision d'intérêt a lieu. Les autres vertex, s'il y en a, correspondent aux collisions venant de l'empilement. Le vertex primaire principal est souvent appelé *hard-scatter vertex* en anglais, et noté  $PV_0$ . Il sera appelé « le » vertex primaire par la suite, bien qu'il y ait plusieurs vertex primaires. Les autres vertex seront appelés « vertex provenant de l'empilement » par la suite. L'empilement hors-temps correspond aux collisions provenant des croisements des

paquets précédents ou suivants. L'empilement hors-temps est mesuré par la variable  $\mu$ , qui correspond au nombre moyen d'interactions par croisement de paquets, c'est-à-dire à la moyenne de la distribution de Poisson du nombre d'interactions par croisement calculée pour chaque paquet. Ce nombre peut être moyenné sur tous les paquets en collision, ce qui résulte en une nouvelle variable, dénotée  $\langle\mu\rangle$ .  $N_{PV}$  est aussi corrélé avec l'empilement hors-temps. Les distributions de  $\langle\mu\rangle$  pour les années 2015, 2016, 2017 et 2018 sont représentées Figure 3.3 [29]. La valeur moyenne de  $\langle\mu\rangle$  augmente avec les années. La valeur moyenne de  $\langle\mu\rangle$  sur les 4 années est égale à 33,7 interactions par croisement de paquets.

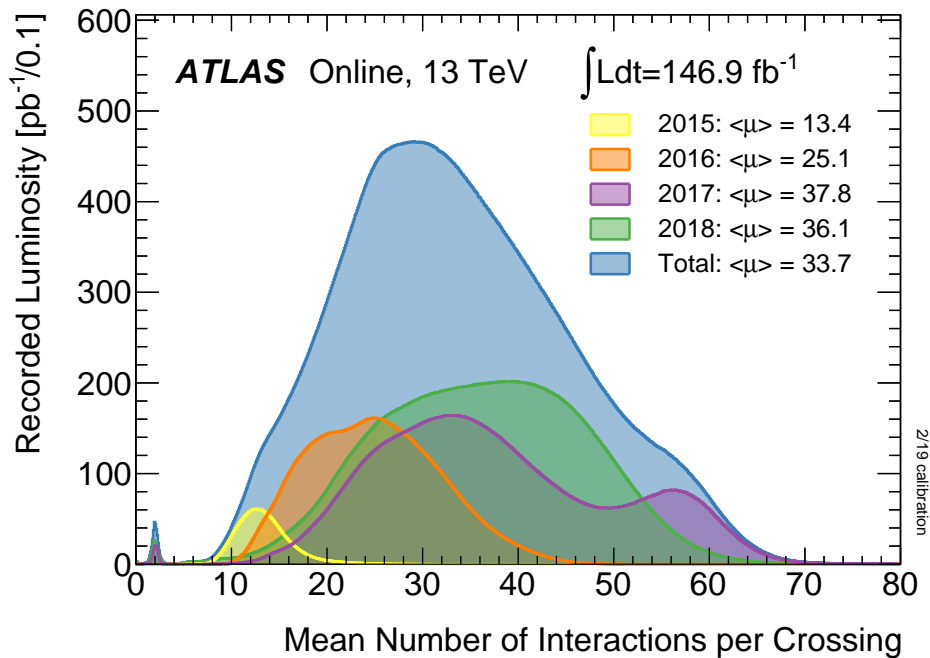


FIGURE 3.3 – Nombre d'interactions par croisement de paquets [29].

## 3.2 Le détecteur ATLAS

Le détecteur *ATLAS* (pour *A Toroidal LHC ApparatuS* en anglais) est l'un des 4 grands détecteurs du LHC au CERN. La collaboration ATLAS regroupe environ 3000 physiciens et physiciennes du monde entier, afin de collecter des données et de les analyser.

Le détecteur ATLAS, représenté Figure 3.4 [26], est un détecteur cylindrique de 25 mètres de diamètre pour environ 44 mètres de longueur, et un poids de 7000 tonnes. Le détecteur couvre presque tout l'angle solide autour du point de collision.

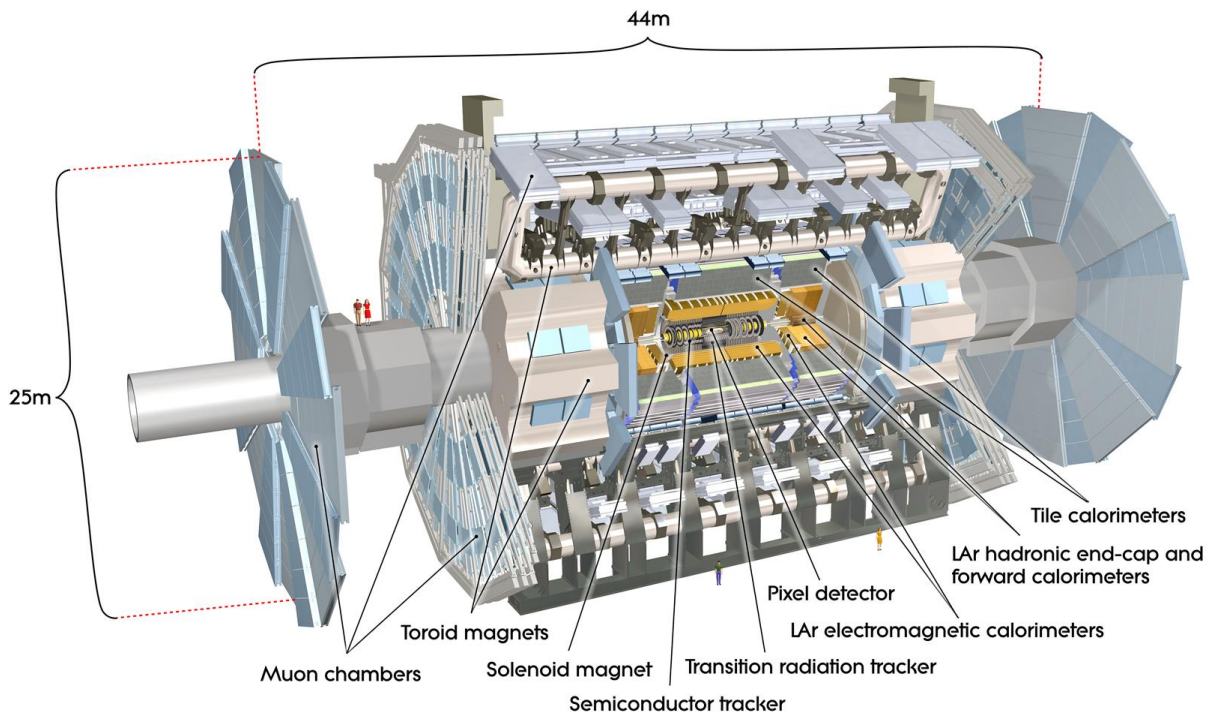


FIGURE 3.4 – Le détecteur ATLAS [26].

### 3.2.1 Systèmes de coordonnées

ATLAS utilise un système de coordonnées droites dont l'origine est le point d'interaction au centre du détecteur. L'axe  $x$  positif est défini par la direction du point d'interaction vers le centre de l'anneau du LHC, l'axe  $y$  positif pointe vers le haut, tandis que la direction du faisceau définit l'axe  $z$ . Ce système de coordonnées est représenté Figure 3.5 [30].

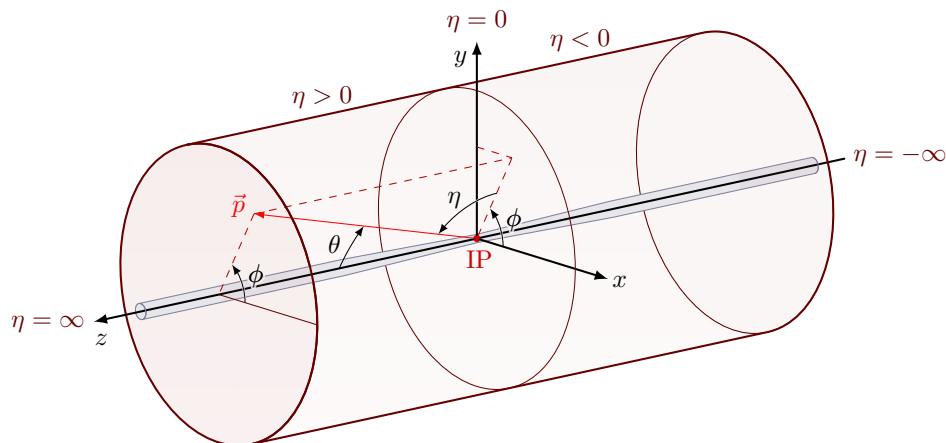


FIGURE 3.5 – Système de coordonnées du détecteur ATLAS [30].

Les coordonnées cylindriques  $(r, \phi)$  sont utilisées dans le plan transversal,  $\phi$  étant

l'angle azimutal autour de l'axe  $z$ . La pseudo-rapacité  $\eta$  est définie en fonction de l'angle polaire  $\theta$  par :

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \quad (3.4)$$

La rapidité  $y$  est définie par :

$$y = -0,5 \times \ln\left(\frac{E + p_z}{E - p_z}\right) \quad (3.5)$$

où  $E$  désigne l'énergie et  $p_z$  désigne la composante de l'impulsion dans la direction du faisceau. Dans la limite d'objets avec une faible masse  $m$ , comparée à l'énergie :  $\eta \simeq y$ .

L'impulsion dans le plan transverse,  $p_T$ , est définie par :

$$p_T = \sqrt{p_x^2 + p_y^2} = |\vec{p}| \sin(\theta) \quad (3.6)$$

de façon à ce que  $E$  et  $p_T$  soient reliées par :

$$p_T = \frac{\sqrt{E^2 - m^2}}{\cosh(\eta)} \quad (3.7)$$

Enfin, la distance angulaire,  $\Delta R$ , entre deux objets est donnée par :

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (3.8)$$

Cette quantité est invariante par boosts de Lorentz.

#### 3.2.2 Description des sous-détecteurs

Le détecteur ATLAS est organisé en différents sous-détecteurs successifs, afin de mesurer et d'identifier au mieux les différents types de particules et leurs propriétés. La partie la plus interne correspond au trajectographe (ou détecteur interne). Ensuite, plus à l'extérieur, se trouvent le calorimètre électromagnétique et le calorimètre hadronique. Enfin, tout à l'extérieur se trouve le spectromètre à muons. Chacun de ces sous-détecteurs permet de mesurer la trajectoire ou l'énergie de différents types de particules. Ils sont représentés Figure 3.6 [31], dans une vue transverse du détecteur ATLAS. Tous ces composants sont décrits plus en détail ci-dessous.



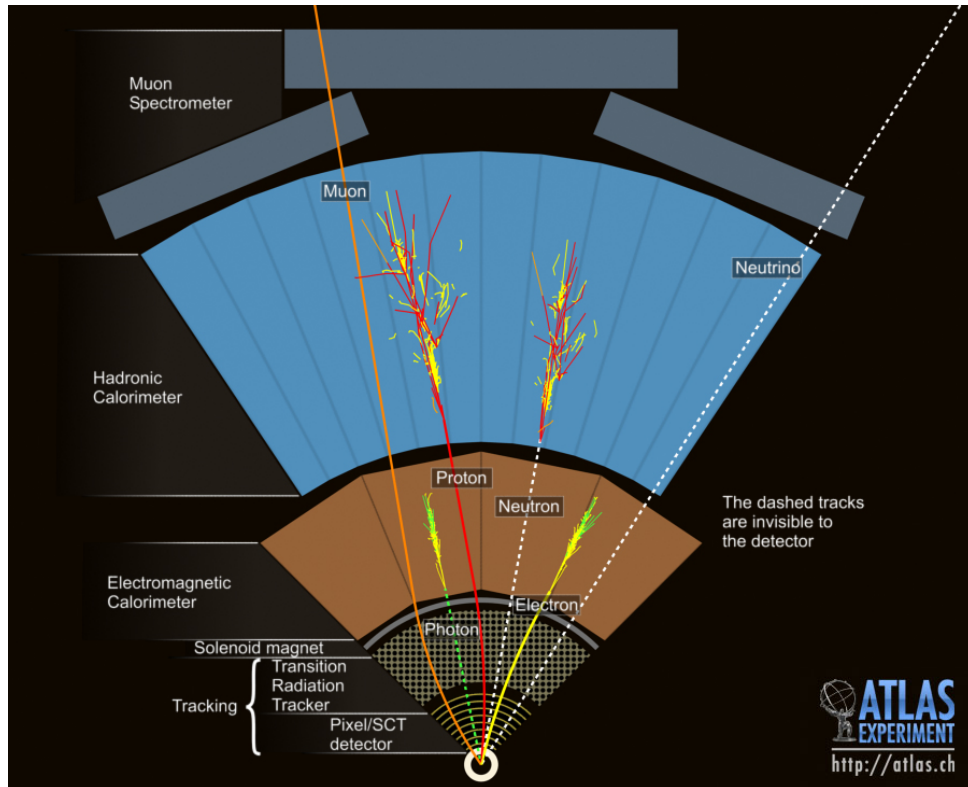


FIGURE 3.6 – Vue transverse du détecteur ATLAS [31].

### 3.2.2.1 Le système d'aimants

Le détecteur interne et le spectromètre à muons d'ATLAS sont immergés dans un champ magnétique, ce qui permet de mesurer l'impulsion et la charge des particules chargées qui y laissent une trace. En effet, le champ magnétique permet de courber la trajectoire des particules chargées. Le sens de la courbure permet d'obtenir la charge de la particule et le rayon de courbure, sa quantité de mouvement. Le système d'aimants est représenté Figure 3.7 [32]. Le détecteur interne est entouré d'un solénoïde (*Central Solenoid* sur la Figure 3.7) permettant de fournir un champ magnétique de 2 T. Le spectromètre à muons est entouré de trois tores, un couvrant la région centrale (*Barrel Toroid* sur la Figure 3.7), et deux à chaque extrémité (*Endcap Toroid* sur la Figure 3.7).

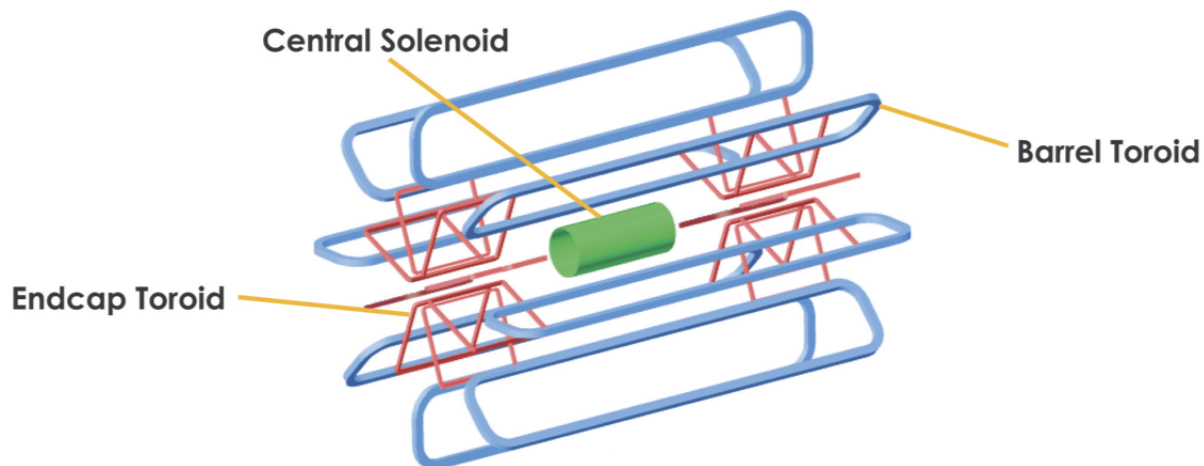
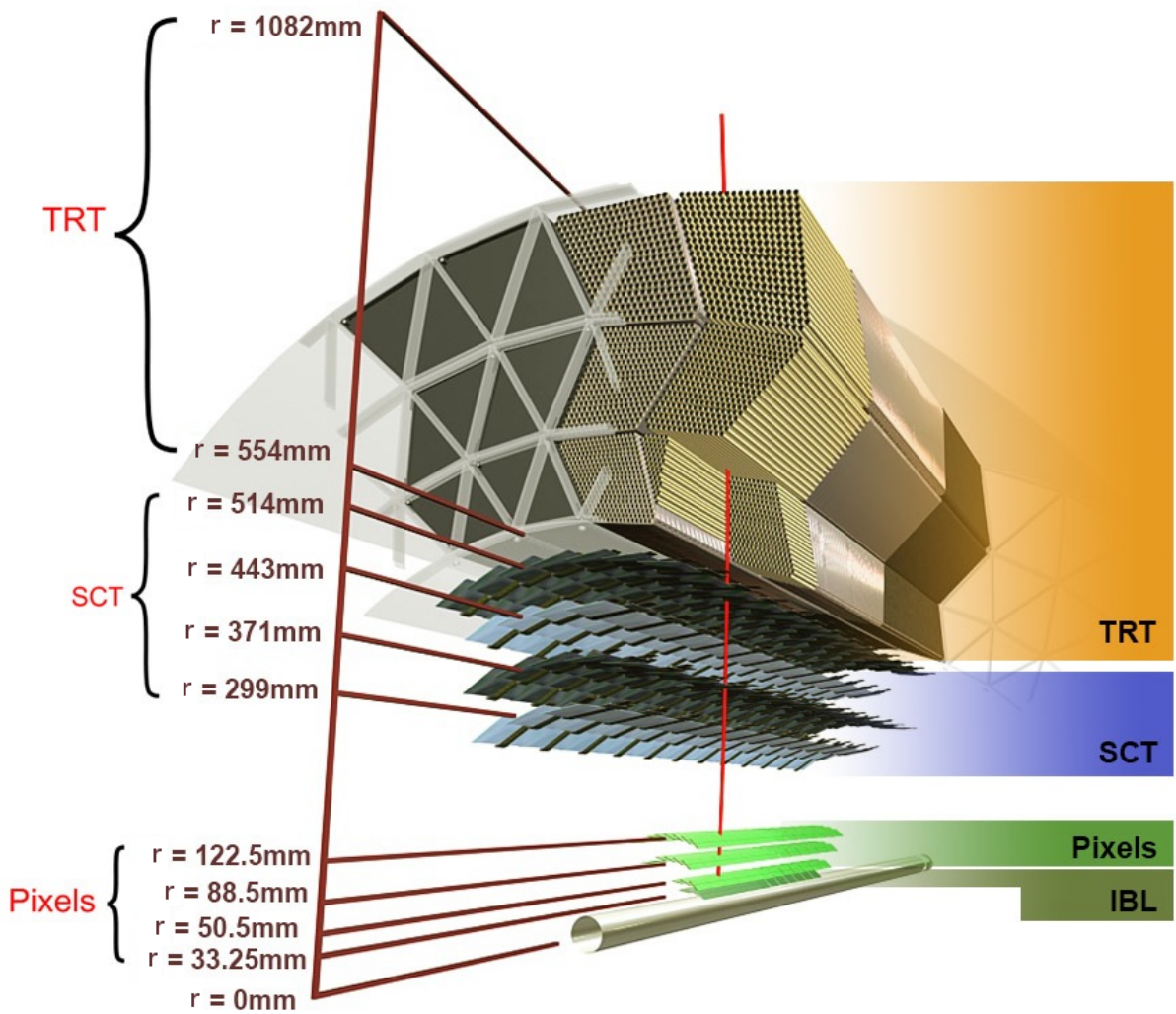


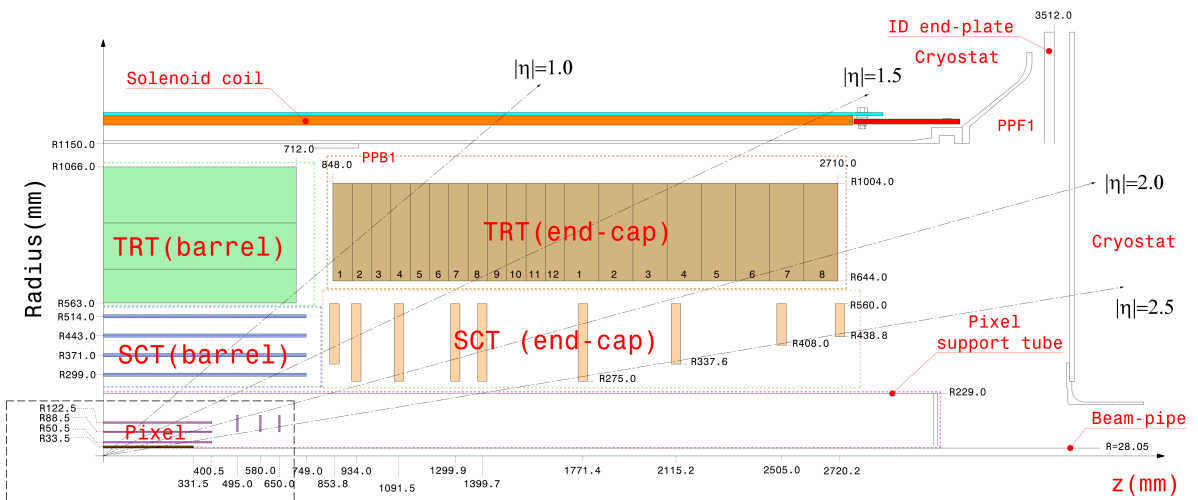
FIGURE 3.7 – Système d'aimants du détecteur ATLAS [33].

### 3.2.2.2 Le détecteur interne

Comme décrit dans la Section 3.2.2.1, le détecteur interne est immergé dans un champ magnétique de 2 T, ce qui permet de mesurer la trajectoire des particules chargées. Le détecteur interne couvre la région  $|\eta| < 2,5$ . Le détecteur interne est lui même subdivisé en trois sous-détecteurs qui utilisent trois technologies différentes : d'un détecteur à pixels en silicium, d'un détecteur à micro-pistes en silicium et d'un détecteur à rayonnements de transition. Un schéma du détecteur interne est représenté Figure 3.8 [34, 35].



(a) Schéma du détecteur interne du détecteur ATLAS [34].



(b) Disposition des sous-détecteurs du détecteur interne, avec les valeurs de  $\eta$  marquées pour référence [35].

FIGURE 3.8 – Schémas du détecteur interne du détecteur ATLAS.

La partie la plus proche du point d'interaction correspond à un détecteur à pixels en silicium de haute granularité, et comprend la couche B insérable [36] (ou *IBL*, pour *Inserable B-Layer* en anglais), qui a été installée avant le Run 2 afin d'améliorer la qualité de reconstruction des traces. L'*IBL* est la partie du détecteur à pixels la plus proche des faisceaux. Le rayon de l'*IBL* est de 33 mm. La taille typique d'un pixel de l'*IBL* est de 50  $\mu\text{m}$  dans la direction transverse et de 250  $\mu\text{m}$  dans la direction longitudinale. Leur épaisseur est de 200  $\mu\text{m}$ . Les trois autres couches du détecteur à pixels sont localisées à des rayons de 50,5, 88,5, and 122,5 mm respectivement. La taille typique d'un pixel dans ces couches est de 50  $\mu\text{m}$  dans la direction transverse et de 400  $\mu\text{m}$  dans la direction longitudinale. Leur épaisseur est de 250  $\mu\text{m}$ . La couverture aux extrémités est augmentée grâce à trois disques de chaque côté. Le détecteur à pixels fournit généralement quatre impacts par trace.

Le détecteur à pixels en silicium est suivi du détecteur à micro-pistes en silicium (ou *SCT*, pour *SemiConductor Tracker* en anglais). Le *SCT* consiste en quatre couches à double piste situées à des rayons entre 299 et 514 mm. Le *SCT* est complété par neuf disques de chaque côté. La taille typique d'une piste du *SCT* est de 80  $\mu\text{m}$  dans la direction transverse et de 126  $\mu\text{m}$  dans la direction longitudinale. Sur chaque couche, les pistes sont parallèles à la direction du faisceau d'un côté et à un angle de 40 mrad de l'autre. Les informations provenant des deux côtés de chaque couche peuvent être combinées pour fournir en moyenne huit impacts par trace.

Lorsqu'ils sont combinés, les détecteurs en silicium s'étendent jusqu'à  $|z| < 805$  mm dans la région centrale, et  $|z| < 2797$  mm aux extrémités.

Enfin, à l'extérieur du détecteur interne se trouve le détecteur à rayonnements de transition (ou *TRT*, pour *Transition Radiation Tracker* en anglais), qui s'étend jusqu'à un rayon de 1082 mm et jusqu'à  $|z| < 780$  mm dans la région centrale, et jusqu'à un rayon de 1106 mm et jusqu'à  $827 < |z| < 2744$  mm aux extrémités. Cela permet au *TRT* de couvrir la région  $|\eta| < 2,0$ . Le *TRT* est composé de tubes à dérive, ce qu'il fait qu'il ne peut mesurer que les coordonnées  $r$  et  $\phi$  dans la région centrale, et  $z$  et  $\phi$  aux extrémités. Le *TRT* fournit généralement 30 impacts par trace.

La résolution relative en impulsion transverse des traces peut être formulée selon [37] :

$$\sigma\left(\frac{1}{p_T}\right) \times p_T = 0,036\% \times p_T \oplus 1,3\% \quad (3.9)$$

Dans cette formule,  $p_T$  est exprimée en GeV et le symbole  $\oplus$  signifie que les différents termes sont ajoutés en quadrature.

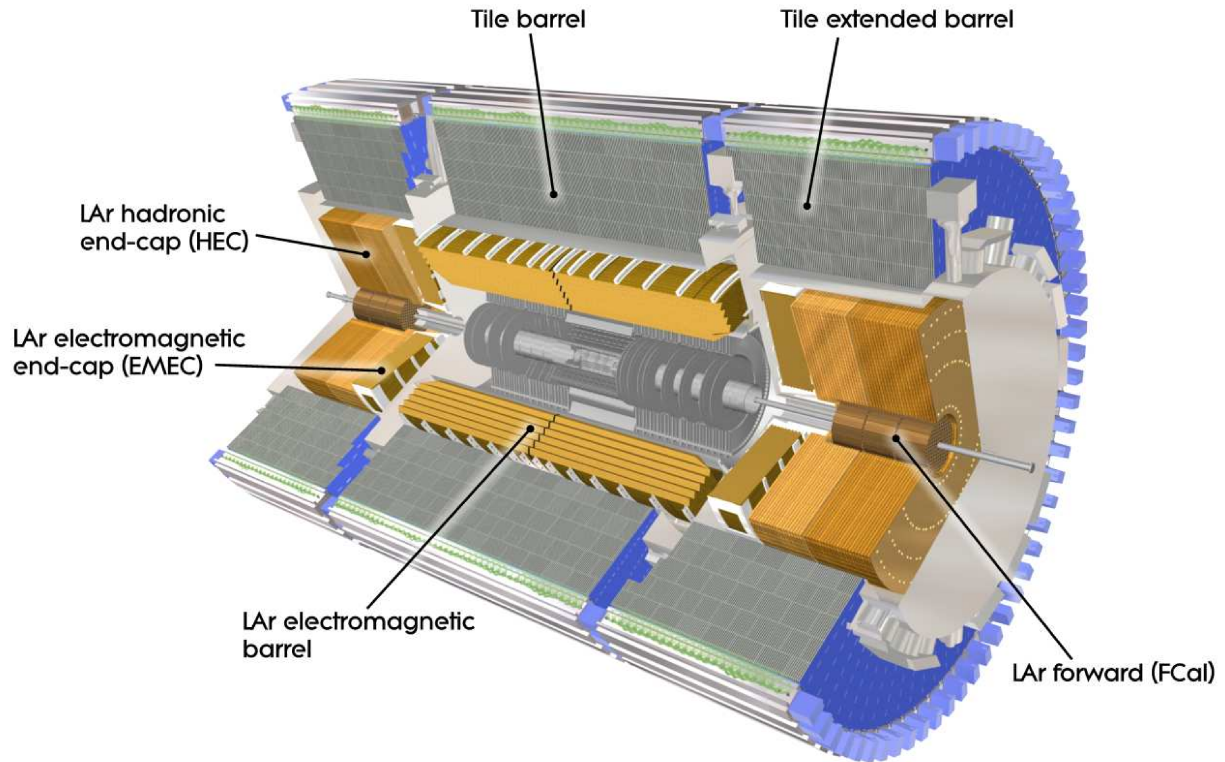
Ainsi, l'estimation de l'impulsion transverse de traces se dégrade avec l'impulsion transverse, ce qui peut s'expliquer par le fait qu'elles sont moins courbées par le champ magnétique, et donc que le rayon de courbure est plus difficile à déterminer.

### 3.2.2.3 Les calorimètres

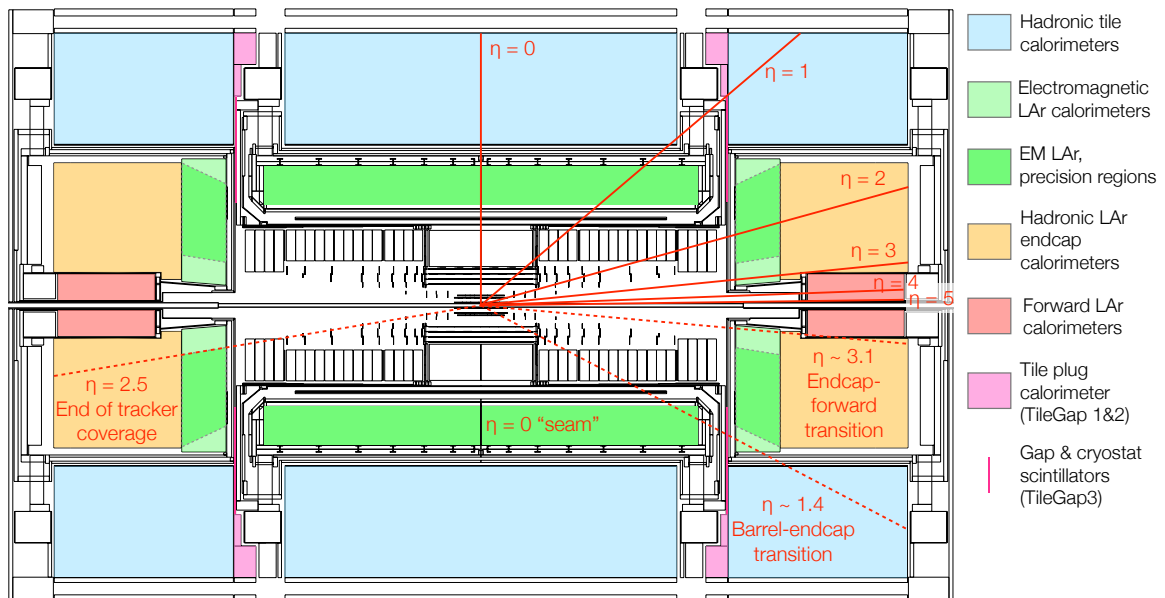
Plus à l'extérieur, les calorimètres permettent de mesurer l'énergie des particules. Les calorimètres ont pour objectif d'arrêter entièrement les particules qui interagissent avec eux, et de mesurer leur énergie qui est déposée. Pour ce faire, les calorimètres sont constitués successivement de couches d'un matériel absorbant et de couches actives de détection, qui servent à induire les gerbes de particules et à mesurer leur énergie respectivement.

Le calorimètre est subdivisé en deux sous-systèmes : le calorimètre électromagnétique en premier, et le calorimètre hadronique en second. Le calorimètre électromagnétique permet de mesurer l'énergie des particules qui interagissent par interaction électromagnétique, telles que les électrons ou les photons. Le calorimètre hadronique permet de mesurer l'énergie des particules qui interagissent par interaction forte, c'est-à-dire les hadrons.

Ce système permet de mesurer l'énergie des particules jusqu'à  $|\eta| < 4,9$ . Les calorimètres du détecteur ATLAS sont représentés Figure 3.9 [38, 39].



(a) Vue en coupe des calorimètres du détecteur ATLAS [38].



(b) Disposition des calorimètres, avec les valeurs de  $\eta$  marquées pour référence [39].

FIGURE 3.9 – Schémas des calorimètres du détecteur ATLAS.

Le calorimètre électromagnétique sert à mesurer l'énergie déposée par les particules chargées et les photons. Les particules chargées légères telles que les électrons perdent de l'énergie principalement par *bremstrahlung*. Lors de ce processus, des photons sont créés.

Ensuite, ces photons se convertissent par création de paires électron-positron. Ces électrons et positrons perdent à leur tour leur énergie par bremsstrahlung, et c'est ainsi que se forment des *gerbes* électromagnétiques dans le détecteur. Les photons perdent directement leur énergie par production de paires électron-positron, ce qui induit aussi une gerbe électromagnétique. L'objectif du calorimètre électromagnétique est de mesurer l'énergie de ces gerbes. Pour ce faire, ATLAS utilise un calorimètre en argon liquide (ou *LAr*) et en plomb, où le plomb sert d'absorbeur et l'argon liquide de matériel actif de détection.

Le calorimètre électromagnétique, aussi appelé *EMCal* (pour *ElectroMagnetic Calorimeter* en anglais), est composé d'une partie *barrel*, qui couvre la région  $|\eta| < 1,475$  et de deux parties *end-cap* qui couvrent les régions  $1,375 < |\eta| < 3,2$ , chacune des régions end-cap étant divisée en une roue interne qui couvre la région  $1,375 < |\eta| < 2,5$  et une roue externe qui couvre la région  $2,5 < |\eta| < 3,2$ . Dans la région  $|\eta| < 2,5$ , le calorimètre est segmenté en trois sections le long de sa profondeur. La roue externe de l'end-cap est segmentée en deux sections selon la profondeur. La région barrel est aussi appelée *EMB* (pour *ElectroMagnetic Barrel* en anglais). La région end-cap est aussi appelée *EMEC* (pour *ElectroMagnetic End-Cap* en anglais).

La partie barrel est elle-même divisée en deux parties, une couvrant la région  $z < 0$ , l'autre couvrant la région  $z > 0$ , et chacune d'une épaisseur de 1,2 m, ce qui correspond à 22 longueurs de radiation ( $X_0$ ) à  $\eta = 0$ ,  $X_0$  augmentant avec  $\eta$ . La granularité de la première couche de la région barrel est  $\Delta\eta \times \Delta\phi = 0,025/8 \times 0,1$  pour  $|\eta| < 1,40$ . La granularité de la seconde couche de la région barrel est  $\Delta\eta \times \Delta\phi = 0,025 \times 0,025$  pour  $|\eta| < 1,40$ . Enfin, la granularité de la troisième couche de la région barrel est  $\Delta\eta \times \Delta\phi = 0,050 \times 0,025$  pour  $|\eta| < 1,35$ .

Chacune des deux roues de la région end-cap ont une épaisseur de 63 cm, ce qui correspond à au moins 24  $X_0$ , ce qui augmente avec  $\eta$ . La granularité de la première couche de la roue interne de la région end-cap est  $\Delta\eta \times \Delta\phi = 0,050 \times 0,1$  pour  $1,375 < |\eta| < 1,425$ . La granularité de la seconde couche de la roue interne de la région end-cap est  $\Delta\eta \times \Delta\phi = 0,050 \times 0,025$  pour  $1,375 < |\eta| < 1,425$ . Enfin, la granularité de la troisième couche de la roue interne de la région end-cap est  $\Delta\eta \times \Delta\phi = 0,050 \times 0,025$  pour  $1,5 < |\eta| < 2,5$ . La granularité de la première et de la deuxième couche de la roue externe de la région end-cap est  $\Delta\eta \times \Delta\phi = 0,1 \times 0,1$  pour  $2,5 < |\eta| < 3,2$ .

Dans la région  $|\eta| < 1,8$ , un *pré-échantillonneur* est utilisé afin de corriger la perte d'énergie des photons et des électrons en amont du calorimètre. Le pré-échantillonneur est constitué d'une couche active de LAr d'une épaisseur de 1,1 cm (0,5 cm) dans la région du barrel (end-cap).

Le calorimètre hadronique, aussi appelé *HCal* (pour *Hadronic Calorimeter* en anglais), se situe à l'extérieur du calorimètre électromagnétique. Il est conçu de façon à contenir les gerbes produites par interaction forte, et c'est pourquoi il est plus épais que le calorimètre électromagnétique. Le calorimètre hadronique est composé de trois parties : un calorimètre à tuiles scintillatrices dans la région  $|\eta| < 1,7$ , un calorimètre en argon liquide dans la région  $1,5 < |\eta| < 3,2$ , et un autre calorimètre en argon liquide dans la région



$3,1 < |\eta| < 4,9$ . La région  $1,5 < |\eta| < 3,2$  et la région  $3,1 < |\eta| < 4,9$  sont appelées région *end-cap* et région *avant* respectivement.

Le calorimètre en tuiles scintillatrices, aussi appelé *TileCal*, est divisé en deux parties : une partie barrel qui couvre la région  $|\eta| < 1,0$  et deux extensions qui couvrent les régions  $0,8 < |\eta| < 1,7$ . Dans ce calorimètre, l'absorbeur est composé d'acier. Il est d'une épaisseur de 1,97 m, ce qui correspond à 9,7 longueurs d'interaction ( $\lambda$ ) à  $\eta = 0$ ,  $\lambda$  augmentant avec  $\eta$ . Le calorimètre à tuiles scintillatrices est segmenté en trois sections le long de sa profondeur, qui correspondent à 1,5, 4,1 et 1,8  $\lambda$  dans la région barrel, et 1,5, 2,6, and 3,3  $\lambda$  dans la région étendue. La granularité de la première et de la seconde couche est  $\Delta\eta \times \Delta\phi = 0,1 \times 0,1$  pour la région barrel et la région étendue. La granularité de la troisième couche est  $\Delta\eta \times \Delta\phi = 0,2 \times 0,1$  pour la région barrel et la région étendue.

Le calorimètre en argon-liquide de la région end-cap, aussi appelé *HEC* (pour *Hadronic End-cap Calorimeter* en anglais), est divisé en quatre roues indépendants, deux de chaque côté. Chaque roue est composée de deux sections le long de la profondeur, ce qui fait donc quatre sections par région end-cap. Ces roues utilisent du cuivre comme absorbeur. La granularité des couches est  $\Delta\eta \times \Delta\phi = 0,1 \times 0,1$  dans la région  $1,5 < |\eta| < 2,5$ . La granularité des couches est  $\Delta\eta \times \Delta\phi = 0,2 \times 0,2$  dans la région  $2,5 < |\eta| < 3,2$ .

Enfin un calorimètre en argon liquide est utilisé pour les mesures électromagnétiques et hadroniques dans la région  $3,1 < |\eta| < 4,9$ . Ce calorimètre est appelé *FCal* (pour *Forward Calorimeter* en anglais). Afin de réduire le nombre de neutrons réfléchis dans le détecteur interne, ce calorimètre est placé 1,2 m plus loin du point d'interaction par rapport au calorimètre électromagnétique. Le calorimètre à argon liquide de la région avant a une profondeur d'environ 10  $\lambda$  et se compose de trois modules de chaque côté : le premier, en cuivre, est optimisé pour les mesures électromagnétiques, tandis que les deux autres, en tungstène, mesurent principalement l'énergie des particules qui interagissent par interaction forte. La granularité de la première couche est  $\Delta x \times \Delta y = 3,0 \text{ cm} \times 2,6 \text{ cm}$  dans la région  $3,15 < |\eta| < 4,30$ . La granularité de la seconde couche est  $\Delta x \times \Delta y = 3,3 \text{ cm} \times 4,2 \text{ cm}$  dans la région  $3,24 < |\eta| < 4,50$ . Enfin, la granularité de la troisième couche est  $\Delta x \times \Delta y = 5,4 \text{ cm} \times 4,7 \text{ cm}$  dans la région  $3,32 < |\eta| < 4,50$ .

Les matériaux utilisés dans les calorimètres sont résumés Table 3.1 [40].

Calorimètre	Barrel	End-cap	Avant
EMCal	Argon liquide Plomb	Argon Liquide Plomb	
HCal	Tuiles scintillatrices Acier	Argon Liquide Cuivre	
FCal			Argon Liquide Cuivre/tungstène

TABLE 3.1 – Matériaux utilisés dans calorimètres du détecteur ATLAS. La première ligne correspond au matériel actif, et la seconde à l'absorbeur.



La couverture en  $|\eta|$  de chacun des calorimètres est résumée Table 3.2 [40].

Calorimètre	Barrel	End-cap	Avant
EMCal	$ \eta  < 1,5$	$1,4 <  \eta  < 3,2$	
TileCal	$ \eta  < 1,0$	$0,9 <  \eta  < 1,7$	
HEC		$1,5 <  \eta  < 3,2$	
FCal			$3,1 <  \eta  < 4,9$

TABLE 3.2 – Couverture en  $|\eta|$  des calorimètres du détecteur ATLAS [40].

La résolution relative en énergie des calorimètres est paramétrée par :

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (3.10)$$

Le premier terme,  $\frac{a}{\sqrt{E}}$ , représente les fluctuations dans la gerbe de particules secondaires créées lors de l'interaction avec les calorimètres. Le second terme,  $\frac{b}{E}$ , représente le bruit électronique des calorimètres. Le troisième terme,  $c$ , représente les non-uniformités dans le détecteur. Pour des collisions à haute énergie, le deuxième terme est négligeable, et la résolution peut être paramétrée par :  $\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus c$ .

Les valeurs des paramètres dépendent du calorimètre considéré. Ces valeurs sont représentées Table 3.3 [26].

Calorimètre	Résolution
Électromagnétique	$\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E}} \oplus 0,7\%$
Hadronique :	
Régions <i>barrel</i> et <i>end-cap</i>	$\frac{\sigma(E)}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$
Région <i>avant</i>	$\frac{\sigma(E)}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\%$

TABLE 3.3 – Résolution en énergie des calorimètres du détecteur ATLAS [26]. L'énergie est exprimée en GeV.

Ainsi, la résolution en énergie s'améliore avec l'énergie des particules, contrairement à la résolution en impulsion transverse du détecteur interne.

#### 3.2.2.4 Le spectromètre à muons

Enfin, le spectromètre à muons est un système de détection très volumineux, ce qui est nécessaire pour identifier et mesurer avec précision la quantité de mouvement des muons. Ces particules sont presque les seules à traverser tous les autres détecteurs (du fait de leur masse importante, elles n'émettent que très peu de rayonnement de freinage) et à laisser une trace dans le spectromètre à muons. Il fonctionne sur le même principe que le détecteur interne, les muons étant courbés par le champ magnétique de sorte que leur

impulsion puisse être mesurée. Il permet d'identifier et de mesurer l'impulsion des muons en combinant la trace du spectromètre avec celle du détecteur interne. Le spectromètre à muons permet de couvrir la région  $|\eta| < 2,7$ .

Les seules particules stables qui ne peuvent pas être détectées directement sont les neutrinos. Les neutrinos peuvent être détectés indirectement, en mesurant une impulsion transverse manquante, comme décrit Section 2.4.

## 3.3 La prise de données

Aux 4 points d'interaction du LHC, une collision se produit environ toutes les 25 ns, ce qui correspond à un taux de 40 MHz. La taille d'un évènement étant d'environ 1,4 MB, sauvegarder toutes ces données est techniquement impossible. De plus, la plupart de ces collisions se produisent à basse énergie et ne présentent pas d'intérêt. C'est pourquoi ATLAS a mis en place un système de déclenchement à deux niveaux, permettant de sélectionner les collisions intéressantes, et par conséquent de réduire la taille des données à sauvegarder. Ces deux niveaux sont appelés *Level 1* (ou plus brièvement *L1*) et *High Level Trigger* (ou plus brièvement *HLT*).

Le premier niveau, le système de déclenchement L1, est simple mais rapide. Il est basé essentiellement sur des cartes électroniques, et non sur des logiciels. L'objectif du système de déclenchement L1 est de sélectionner des objets physiques, tels que des muons, des électrons ou des jets, de haute impulsion. Pour ce faire, il utilise les informations qui proviennent des calorimètres et du spectromètre à muons. Le système de déclenchement L1 permet de réduire le taux de 40 MHz à 100 kHz. De plus, le L1 permet de diminuer la taille des évènements à considérer pour le second système de déclenchement en définissant des *régions d'intérêt*, ou *Regions Of Interest* (plus brièvement *ROI*) en anglais, dans le détecteur afin d'optimiser le temps de réponse du système.

Le second niveau, le système de déclenchement HLT, est quand à lui essentiellement basé sur des logiciels. Les objets physiques tels que les muons, les électrons ou les jets, sont reconstruits, et des sélections sont appliquées afin de réduire le taux jusqu'à 1 kHz. La reconstruction au niveau HLT diffère de la reconstruction *hors-ligne* (c'est-à-dire au niveau des analyses), mais a pour objectif de s'en rapprocher le plus possible. C'est pourquoi une coupure sur un objet reconstruit hors-ligne peut avoir un effet différent que la même coupure appliquée sur un objet reconstruit au niveau HLT, ce qui peut entraîner une perte d'efficacité due au système de déclenchement.

Le taux des différents systèmes de déclenchement dépend des seuils en impulsion utilisés. Certains systèmes de déclenchement dont le seuil est trop bas enregistrent trop d'évènements. C'est pourquoi l'acceptance de ces systèmes de déclenchement est réduite, par exemple en ne sauvegardant qu'un évènement sur deux. Ces systèmes de déclenchement sont dits *pré-échelonnés*. Les analyses utilisent typiquement des systèmes de déclenchement qui ne sont pas pré-échelonnés, c'est-à-dire qui sauvegardent tous les évènements qui passent la sélection définie par ce système de déclenchement.

### 3.4 La qualité des données

Comme décrit précédemment, un *Run du LHC* correspond à une longue période de prise de données. Par exemple, le Run 2 correspond aux années 2015, 2016, 2017 et 2018.

Chaque Run est divisé en environ 20 ou 30 *périodes*, chaque période pouvant être divisée en environ 5 ou 10 *sous-périodes* avec des conditions similaires (par exemple des incréments après des arrêts techniques). Les sous-périodes peuvent elles mêmes être divisées en environ 5 ou 10 *Runs* (à ne pas confondre avec les *Run du LHC*) avec des conditions similaires (par exemple en terme de système de déclenchement). Ces Runs correspondent à une période continue de prise de données de la part d'ATLAS, typiquement plusieurs heures. Les Runs peuvent enfin être divisés en environ 1000 *blocs de luminosité*, qui correspondent à environ 1 minute de prise de données, c'est-à-dire environ 60 000 événements enregistrés par le système de déclenchement.

Certains blocs de luminosité sont signalés comme ayant une mauvaise qualité, et ne doivent donc pas être utilisés par les analyses. Techniquement, les blocs de luminosité ayant une bonne qualité sont listés dans des *Good Run Lists* (ou plus brièvement *GRL*). La luminosité intégrée des blocs listés dans les Good Run Lists correspond à la courbe bleue sur la Figure 3.2 [29].

### 3.5 Les simulations Monte Carlo

La plupart des analyses dans ATLAS font usage de simulations, que ce soit pour optimiser les sélections, ou encore pour étudier la réponse du détecteur.

La première étape de production d'événements consiste à simuler des particules produites au sein de collisions *pp* d'intérêt avec des générateurs tels que *PYTHIA* [41], *SHERPA* [42] ou encore *HERWIG* [43]. Cette étape est appelée *génération*.

Des collisions correspondant à de l'empilement sont ajoutées à cette étape, afin de reproduire les conditions réelles lors desquelles sont prises les données. Cependant, ces conditions ne sont pas connues lorsque la production des événements simulés commence. C'est pourquoi les événements simulés sont associés à un poids, qui permet de tenir compte des différences de condition entre les données et les simulations (voir ci-dessous). Puisque les conditions diffèrent entre 2015-2016, 2017 et 2018, trois différentes collections de simulations sont utilisées pour reproduire les données. Ces collections sont généralement appelées *MC16a*, correspondant aux conditions de prise de données des années 2015 et 2016, *MC16d*, correspondant aux conditions de prise de données de l'année 2017, et *MC16e*, correspondant aux conditions de prise de données de l'année 2018. Pour les simulations seulement, tous les paquets voisins dans un événement simulé ont le même  $\mu$  que l'événement central, ainsi :  $\mu = \langle \mu \rangle$ . Par conséquent, ces variables seront utilisées interchangeablement par la suite, lorsque seulement des simulations sont impliquées.

Ensuite, les interactions de ces particules avec le détecteur ATLAS sont simulées avec *GEANT4* [44]. Cette étape est appelée *simulation*.

Puis, les dépôts d'énergie sont convertis en signaux électroniques. Cette étape est appelée *digitisation*.

Enfin, les algorithmes de reconstruction d'ATLAS sont appliqués aux données après digitisation. Cette étape est appelée *reconstruction*, et utilise les mêmes logiciels que ceux utilisés pour reconstruire les vraies données.

Les informations pour les vraies données et les simulations sont les mêmes, si ce n'est que les informations *vraies*, c'est-à-dire provenant directement du générateur, sont aussi disponibles pour les simulations.

Afin de correspondre aux données, les événements simulés sont associés à des poids individuels, dont la formule est donnée par :

$$w_{\text{tot}} = \frac{w_{\text{MC}} \times \sigma \times \mathcal{L}_{\text{int}} \times \varepsilon \times k \times w_{\text{pile-up}}}{\sum w_{\text{MC}}} \quad (3.11)$$

où :

- $w_{\text{tot}}$  représente le poids total, qui est utilisé dans les analyses,
- $w_{\text{MC}}$  représente le poids interne de l'évènement, déterminé par le générateur,
- $\sigma$  représente la section efficace du processus considéré,
- $\mathcal{L}_{\text{int}}$  représente la luminosité intégrée des données avec lesquelles les simulations sont comparées,
- $\varepsilon$  représente l'efficacité du filtre appliqué aux événements au niveau de la génération, afin de ne sélectionner que les événements d'intérêt pour les analyses, et par conséquent, de rendre le générateur plus efficace. Les événements sélectionnés doivent correspondre à certains critères, par exemple reliés aux propriétés en  $p_T$  et en  $\eta$  des électrons ou des muons, aux saveurs de désintégration de bosons  $W$  ou  $Z$ , ou de quarks  $t$ , etc,
- $k$  représente le *facteur*  $k$ , qui est égal à 1 par défaut, et sert à prendre en compte les termes d'ordre supérieur lors du calcul de la section efficace,
- $w_{\text{pile-up}}$  représente un poids qui est appliqué de façon à ce que les distributions de  $\mu$  correspondent entre les données et les événements simulés,
- et  $\sum w_{\text{MC}}$  représente la somme des poids de tous les événements générés (sans coupure), et sert à normaliser les distributions, afin que la somme des  $w_{\text{tot}}$  ne dépende pas du nombre d'évènements générés (qui représente juste la « statistique » de la simulation).

Selon le processus étudié, il peut arriver que les événements soient générés en tranches d'impulsion transverse, afin de générer suffisamment d'évènements à basse et à haute impulsion transverse.

# Chapitre 4

## La reconstruction des objets

### Sommaire du Chapitre

---

<b>4.1</b>	<b>La reconstruction des traces . . . . .</b>	<b>48</b>
<b>4.2</b>	<b>La reconstruction des vertex . . . . .</b>	<b>51</b>
<b>4.3</b>	<b>La reconstruction des jets . . . . .</b>	<b>51</b>
4.3.1	Les constituants des jets . . . . .	51
4.3.2	Les algorithmes de regroupement des constituant des jets . . .	61
4.3.3	La masse des jets . . . . .	63
4.3.4	Les algorithmes de différentiation de l'empilement s'appliquant aux constituants . . . . .	64
4.3.5	Les algorithmes de grooming . . . . .	66

---

À partir des dépôts d'énergie dans les calorimètres et des impacts dans le détecteur interne et dans le spectromètre à muons, plusieurs types d'objets peuvent être identifiés et reconstruits : des traces, des vertex, des électrons, des photons, des muons, des leptons  $\tau$ , des jets ou encore de l'impulsion transverse manquante. Ces objets sont ensuite utilisés dans les analyses physiques. La reconstruction des traces, des vertex et des jets est détaillée ci-dessous.

### 4.1 La reconstruction des traces

Les traces sont reconstruites à partir des impacts de particules chargées dans le détecteur interne. Plusieurs algorithmes sont appliqués successivement afin de reconstruire les traces à partir de ces impacts. La reconstruction des traces commence par une procédure qui va de l'intérieur du détecteur vers l'extérieur, suivie d'une reconstruction qui va de l'extérieur vers l'intérieur.

La première étape de la reconstruction des traces, qui va de l'intérieur du détecteur vers l'extérieur, est de former des *graines* (*seeds* en anglais) de traces à partir de trois impacts dans le détecteur à pixel, ou de trois impacts dans le SCT, ou d'une combinaison d'impacts dans les deux détecteurs. Si la graine de trace satisfait certains critères de qualité, une fenêtre de recherche est construite à partir de celle-ci et des candidats aux traces

sont formés à l'aide d'un filtre de Kalman combinatoire [45]. Les candidats aux traces sont ensuite traités par des algorithmes de résolution d'ambiguïtés [46]. Puis, les candidats aux traces restants sont extrapolés dans le TRT [47] : les impacts dans le TRT compatibles avec l'extrapolation de la trace sont associés à celle-ci. La qualité de l'extrapolation de la trace dans le TRT est mesurée avec un ajustement avec un  $\chi^2$ . Cette étape a pour objectif d'améliorer la qualité de la trace, si une extrapolation dans le TRT est possible.

Ensuite, une seconde étape de reconstruction des traces, qui va de l'extérieur vers l'intérieur du détecteur commence. Premièrement, de nouvelles potentielles traces sont formées à partir du TRT seulement. Ensuite, ces traces sont étendues aux détecteurs en silicium, en y associant les impacts qui sont pas déjà utilisés lors de l'étape de reconstruction qui va de l'intérieur du détecteur vers l'extérieur. Cette étape a pour objectif de reconstruire les traces qui proviendraient de vertex secondaires, tels que les produits d'une conversion de photon par exemple.

Les traces résultantes, issues de la reconstruction qui va de l'intérieur du détecteur vers l'extérieur et de celle qui va de l'extérieur vers l'intérieur, constituent la collection finale des traces. Ces traces peuvent être décrites à l'aide de cinq paramètres :

- le paramètre d'impact transverse de la trace,  $d_0$ , défini comme le point le plus proche dans le plan  $(r, \phi)$  entre la trajectoire de la trace et l'axe des  $z$ ,
- le paramètre d'impact longitudinal de la trace,  $z_0$ , défini par la distance entre le vertex primaire et le point où  $d_0$  est défini,
- l'angle azimutal,  $\phi$ , de la trace,
- l'angle polaire,  $\theta$ , de l'impulsion de la trace au point de référence,
- et le rapport entre la charge de la trace divisée par la magnitude de son impulsion,  $\frac{q}{p}$ .

Ces paramètres sont représentés Figure 4.1 [48].

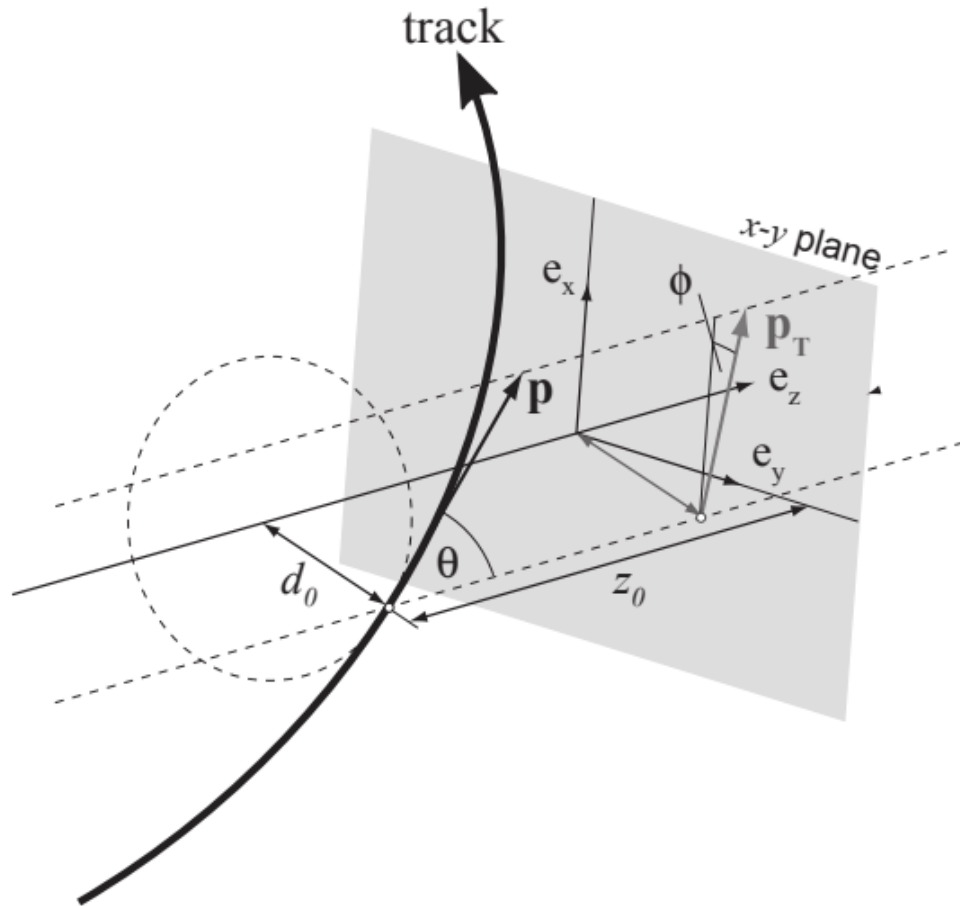


FIGURE 4.1 – Paramètres des traces [48].

Enfin, afin de garantir que seules les traces bien reconstruites sont utilisées, des critères de sélection additionnels sur la qualité de ces traces sont appliqués. Ces critères de sélection correspondent au point de travail *Loose* [49] :

- $p_T > 500$  MeV, car les traces de trop basse impulsion sont trop courbées par le champ magnétique,
- $|\eta| < 2,5$ , ce qui correspond à l'acceptance du détecteur interne,
- au moins 7 impacts dans le détecteur à pixels et dans le SCT (12 sont attendus),
- pas plus d'un impact partagé dans le détecteur à pixels ou de deux impacts partagés dans le SCT (les impacts partagés sont ceux qui sont utilisés dans la reconstruction de deux traces différentes),
- pas plus de deux trous dans le détecteur à pixel et dans le SCT combinés (les trous sont définis comme les intersections de la trajectoire reconstruite de la trace avec un élément sensible du détecteur qui ne contient pas d'impact),
- pas plus d'un trou dans le détecteur à pixels,
- et association avec le vertex primaire, c'est-à-dire :
  - $d_0 < 2$  mm,
  - $|z_0 \sin(\theta)| < 2$  mm.

Ainsi, comme l'indiquent les deux derniers critères de sélection, les traces peuvent être associées au vertex primaire, ce qui fournit un moyen de différencier les traces venant de la collision d'intérêt, de celles venant de l'empilement.

## 4.2 La reconstruction des vertex

Le vertex primaire est sélectionné comme étant le vertex avec la somme des impulsions transverses des traces qui lui sont associées,  $\sum_{\text{tracks}} (p_{\text{T}}^{\text{track}})^2$ , la plus élevée. Au moins deux traces sont requises pour qu'un vertex puisse former un vertex primaire. Seules les traces ayant une impulsion transverse plus grande que 500 MeV sont utilisées.

## 4.3 La reconstruction des jets

Du point de vue expérimental, la reconstruction des jets est un vrai défi, tant ces objets sont complexes. Il existe plusieurs moyens de définir et de reconstruire ces jets, qui dépendent de la méthode de regroupement des constituants des jets, ainsi que de la nature de ces constituants. Dans un premier temps, plusieurs types de constituants sont détaillés. Ensuite, plusieurs méthodes de regroupement de ces constituants sont exposées. Puis la masse des jets est définie. Enfin, différents algorithmes servant à retirer l'empilement sont détaillés. Ces algorithmes peuvent s'appliquer sur les constituants avant que les jets ne soient construits, ou une fois que le jet a été construit.

### 4.3.1 Les constituants des jets

Plusieurs constituants des jets ont été utilisés par ATLAS. Dans un premier temps, ces constituants étaient reconstruits à partir des calorimètres seulement. Ces constituants sont appelés les *topo-clusters*. Ensuite, afin d'exploiter au mieux les informations complémentaires des calorimètres et du détecteur interne, plusieurs méthodes de combinaison de traces et des dépôts d'énergie dans les calorimètres ont été développées par ATLAS : les *Particle Flows* et les *Track-CaloClusters*. Les topo-clusters, les Particle Flows et les Track-CaloClusters sont détaillés ci-dessous.

#### 4.3.1.1 Les topo-clusters

Lorsque les particules contenues dans les jets interagissent avec les calorimètres, leurs dépôts d'énergie sont enregistrés sous forme de cellules individuelles. Ces cellules individuelles sont ensuite regroupées en amas de dépôts d'énergie en trois dimensions appelés *topological cell clusters* en anglais, ou plus brièvement *topo-clusters* [38]. Le regroupement des cellules individuelles se fait de façon à ce que les bruits des calorimètres ou provenant de l'empilement soient supprimés, en enlevant les cellules avec un dépôt d'énergie non significatif qui ne sont pas proches de cellules avec un dépôt d'énergie significatif. Les dépôts d'énergie des particules individuelles à l'intérieur des jets ne sont en général pas contenus dans un seul et même topo-cluster. Au contraire, selon le type de ces particules, de leur énergie, ou encore de leur séparation spatiale, les topo-clusters peuvent contenir



une fraction complète ou partielle de l'énergie de la particule, ou encore l'énergie provenant de plusieurs particules.

Afin de former les topo-clusters, une variable est introduite : la signifiante d'une cellule,  $\zeta_{\text{cell}}^{\text{EM}}$ , qui est définie par le rapport entre l'énergie déposée dans la cellule,  $E_{\text{cell}}^{\text{EM}}$ , et le bruit moyen (attendu) dans cette même cellule,  $\sigma_{\text{noise,cell}}^{\text{EM}}$  :

$$\zeta_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise,cell}}^{\text{EM}}} \quad (4.1)$$

$\sigma_{\text{noise,cell}}^{\text{EM}}$  comprend le bruit électronique provenant du calorimètre, ainsi que le bruit provenant de l'empilement :

$$\sigma_{\text{noise,cell}}^{\text{EM}} = \sqrt{(\sigma_{\text{noise}}^{\text{electronic}})^2 + (\sigma_{\text{noise}}^{\text{pile-up}})^2} \quad (4.2)$$

Puisque le bruit correspondant à l'empilement dépend de la quantité d'empilement,  $\sigma_{\text{noise}}^{\text{pile-up}}$  dépend de celle-ci. Les valeurs optimales de  $\sigma_{\text{noise}}^{\text{pile-up}}$  sont donc estimées en fonction de la quantité d'empilement, mesurée par la valeur de  $\mu$ .

Dans les formules précédentes, *EM* signifie que les quantités sont mesurées à « l'échelle d'énergie électromagnétique », c'est-à-dire que les corrections pour la perte d'énergie des hadrons due au caractère non compensatoire des calorimètres du détecteur ATLAS ne sont pas incluses. En effet, lorsque des hadrons interagissent avec les calorimètres hadroniques, des hadrons secondaires de plus faible énergie sont créés lors de la gerbe hadronique. Ces hadrons sont capturés par les noyaux des particules qui composent le calorimètre hadronique, et par conséquent cette énergie reste invisible. Cela signifie que les gerbes hadroniques ont une échelle d'énergie beaucoup plus basse que les gerbes électromagnétiques. Ainsi, l'échelle d'énergie électromagnétique reconstitue correctement l'énergie déposée par les électrons et les photons mais n'inclut aucune correction pour la perte de signal des hadrons. Elle correspond donc à l'énergie brute des cellules.

Les topo-clusters sont formés par un algorithme de volume croissant à partir d'une cellule du calorimètre avec un dépôt d'énergie significatif. Cet algorithme comprend trois étapes : une étape d'ensemencement, une étape de croissance, et une étape de frontière. Ces trois étapes sont contrôlées respectivement par trois paramètres :  $S$ ,  $N$  et  $P$ , qui correspondent à différents seuils pour l'Équation 4.1 [38] :

— Ensemencement, avec  $S = 4$  :

$$|E_{\text{cell}}^{\text{EM}}| > S\sigma_{\text{noise,cell}}^{\text{EM}} \Leftrightarrow |\zeta_{\text{cell}}^{\text{EM}}| > S \quad (4.3)$$

— Croissance, avec  $N = 2$  :

$$|E_{\text{cell}}^{\text{EM}}| > N\sigma_{\text{noise,cell}}^{\text{EM}} \Leftrightarrow |\zeta_{\text{cell}}^{\text{EM}}| > N \quad (4.4)$$

— Frontière, avec  $P = 0$  :

$$|E_{\text{cell}}^{\text{EM}}| > P\sigma_{\text{noise,cell}}^{\text{EM}} \Leftrightarrow |\zeta_{\text{cell}}^{\text{EM}}| > P \quad (4.5)$$

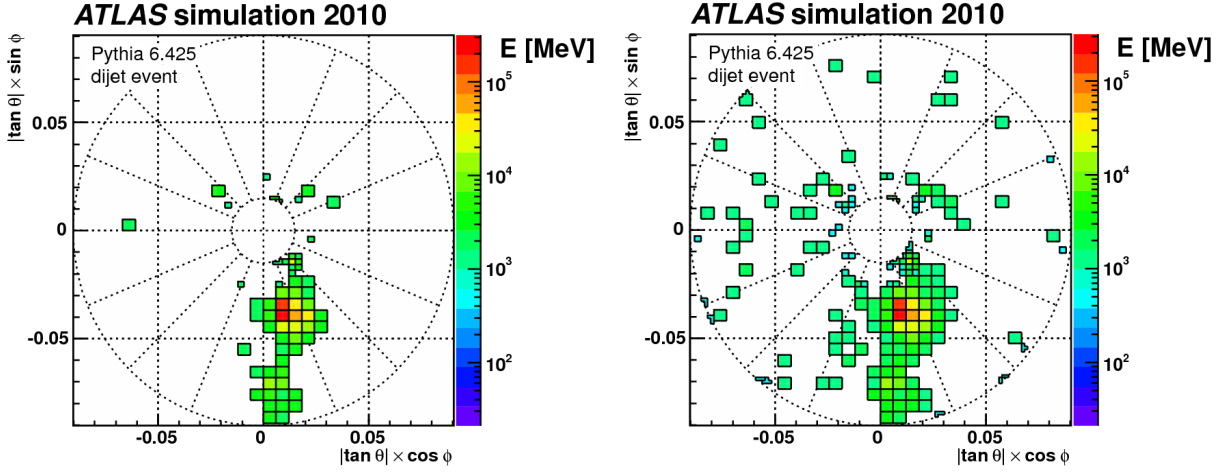
L'algorithme de formation des topo-cluster commence par sélectionner les cellules qui satisfont la condition de l'Équation 4.3. Ces cellules sont ensuite ordonnées par ordre décroissant en  $\zeta_{\text{cell}}$ . Ces cellules sont appelées des *graines*. Cette étape correspond à l'ensemencement.

Ensuite, pour chacune de ces graines, leurs cellules adjacentes qui satisfont l'Équation 4.4 ou 4.5 leur sont associées. Cette étape correspond à l'accroissement. Si une cellule adjacente passe le critère de sélection de l'Équation 4.3, la cellule et la graine sont fusionnées. Si une cellule adjacente est associée à deux graines différentes et passe le critère de sélection de l'Équation 4.4, les deux graines sont fusionnées.

Enfin, si une cellule adjacente passe le critère de sélection de l'Équation 4.4 et dispose elle-même de cellules adjacentes passant le critère de sélection 4.5, toutes ces cellules sont elles-aussi associées à la graine. Cette étape correspond à la frontière.

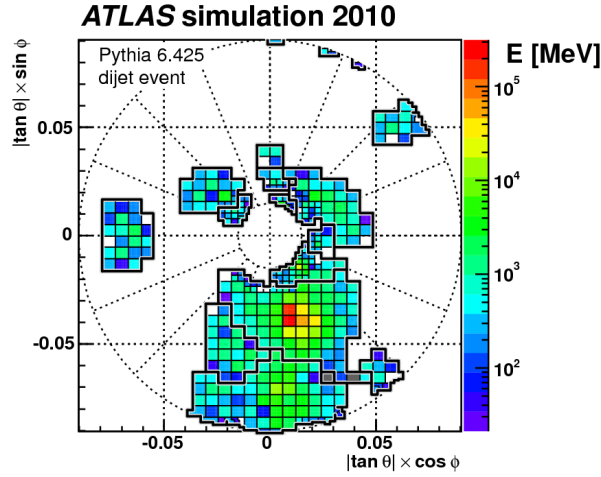
Cette procédure est appliquée itérativement à toutes les cellules adjacentes, de façon à ce que toutes celles qui passent le critère de sélection de l'Équation 4.5, mais pas celui de l'Équation 4.4, soient collectées.

Un exemple de formation d'un topo-cluster est représenté Figure 4.2 [38], en utilisant des données simulées dans les conditions de 2010.



(a) Cellules qui passent la sélection de l'Équation 4.3.

(b) Cellules qui passent la sélection de l'Équation 4.4.



(c) Toutes les cellules formant les topo-clusters.

 FIGURE 4.2 – Différentes étapes de la formation des topo-clusters dans le plan  $(|\tan(\theta)| \times \cos(\phi), |\tan(\theta)| \times \sin(\phi))$  [38]. L'empilement n'est pas inclus.

À cause des fluctuations dues à l'empilement ou au bruit électronique, il peut arriver que certaines cellules du calorimètre en argon liquide aient une énergie négative. En effet, la forme de l'impulsion électrique dans le calorimètre en argon liquide a été conçue de façon à ce que son intégrale prise sur le temps soit approximativement nulle, afin d'annuler en moyenne les contributions de l'empilement dans le temps et de l'empilement hors-temps. C'est pourquoi l'impulsion électrique devient négative à  $t \simeq 150$  ns, comme représenté sur la Figure 4.3 [38].

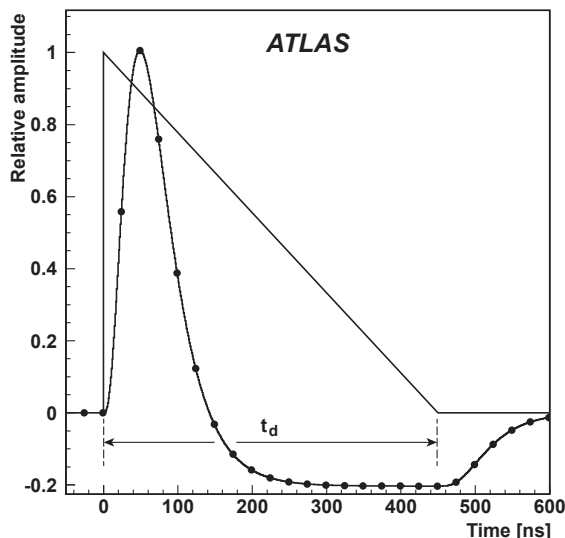


FIGURE 4.3 – Forme de l’impulsion électrique dans les calorimètres en argon liquide en fonction du temps [38].  $t_d$  représente le temps de collection de charge et est environ égal à 450 ns. Les points noirs sont espacés de 25 ns.

Ainsi, les contributions des signaux provenant des interactions proton-proton se produisant lors de croisement de paquets précédents se manifestent généralement par des dépôts d’énergie négative. Ces cellules d’énergie négative peuvent participer à la formation des topo-clusters, puisque ce sont les valeurs absolues des énergies des cellules qui sont utilisées dans les Équations 4.3, 4.4 et 4.5.

Comme décrit précédemment, les topo-clusters utilisent l’échelle d’énergie électromagnétique. Or la reconstruction d’énergie pour un hadron est plus petite que celle d’un électron ou d’un photon qui déposerait la même énergie. C’est pourquoi ATLAS a développé une technique d’étalonnage des topo-clusters, appelée *Local Cell Weighting*, ou plus brièvement *LCW* [38]. Cette technique comporte deux étapes. Dans un premier temps, l’origine des topo-clusters est identifiée comme étant électromagnétique ou hadronique en utilisant des informations sur les topo-clusters telles que l’énergie du topo-cluster,  $\eta$  du topo-cluster, sa profondeur longitudinale, ou encore la densité moyenne du signal dans chaque cellule. Ensuite, une correction est appliquée aux topo-clusters. Cette correction dépend principalement de l’énergie et de la localisation des topo-clusters, ainsi que de l’origine des topo-clusters établie lors de la première étape. La technique LCW tient aussi compte des cellules qui ne seraient pas incluses lors de la reconstruction des topo-clusters, et de la perte d’énergie en dehors du détecteur. Elle permet d’améliorer la résolution à haut énergie, mais peine à classifier et à étalonner les topo-clusters à basse énergie sans augmenter le bruit.

Ainsi, deux types de constituants peuvent être donnés en entrée aux algorithmes de regroupement des constituants des jets : les topo-clusters à l’échelle d’énergie électromagnétique, auquel cas les jets sont appelés *EMTopo jets*, ou les topo-clusters auxquels la technique d’étalonnage LCW est appliquée, auquel cas les jets sont appelés *LCTopo jets*.

Contrairement à la résolution angulaire des traces, celle des topo-clusters n'est pas suffisamment bonne pour permettre de déterminer si un dépôt d'énergie vient du vertex primaire ou non. D'autres techniques sont donc nécessaires pour différencier les dépôts d'énergie qui viennent de la collision d'intérêt, de ceux qui viennent de l'empilement.

#### 4.3.1.2 Les Particle Flows

Les topo-clusters n'utilisent que des informations provenant des calorimètres seulement. Cependant les calorimètres et le détecteur interne ont des performances complémentaires, qui sont résumées ci-dessous :

- Les calorimètres permettent de mesurer l'énergie de particules neutres et de particules chargées, alors que le détecteur interne ne permet de mesurer que l'impulsion de particules chargées.
- La résolution angulaire d'une particule chargée est meilleure lorsque celle-ci est mesurée avec le détecteur interne que lorsqu'elle est mesurée avec les calorimètres.
- Cette excellente résolution angulaire permet de déterminer si une trace provient du vertex primaire, ou de ceux associés à l'empilement, ce qui fournit une méthode pour rejeter les traces dues à l'empilement.
- Les particules chargées de très basse impulsion sont éjectées par le champ magnétique en dehors du jet avant qu'elles n'atteignent les calorimètres.
- Comme mentionné dans les Sections 3.2.2.2 et 3.2.2.3, pour des particules chargées de basse énergie, la résolution en impulsion transverse du détecteur interne est meilleure que la résolution en énergie des calorimètres. Au contraire, pour des particules chargées de haute énergie, la résolution des calorimètres est meilleure.
- Les calorimètres permettent de couvrir approximativement tout l'angle solide, alors que le détecteur interne ne couvre que la région  $|\eta| < 2,5$ .

C'est pourquoi ATLAS a développé un nouveau type de constituants, permettant d'exploiter au mieux les informations des calorimètres et du détecteur interne, en les combinant. Ces nouveaux objets sont appelés les *Particle Flow Objects* [37] (ou juste *Particle Flows*). L'algorithme des Particle Flows est représenté Figure 4.4 [37].

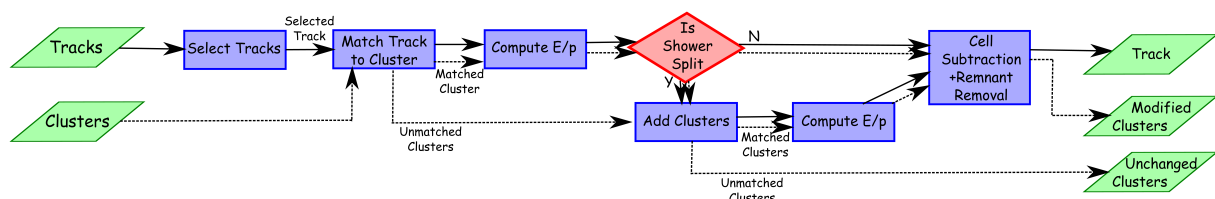


FIGURE 4.4 – Schéma de l'algorithme des Particle Flows [37].

La première étape de l'algorithme des Particle Flow Objects est de sélectionner les traces et les topo-clusters. Pour ce faire, les topo-clusters sont utilisés à l'échelle d'énergie électromagnétique, et les traces sont sélectionnées avec le point de travail *Tight*. Des topo-clusters corrigés avec la technique LCW pourraient être utilisés pour cet algorithme, mais des développements supplémentaires sont nécessaires pour combiner ces deux méthodes, d'autant plus que, comme décrit dans la Section 4.3.1.1, la technique LCW ne fonctionne

pas très bien à basse énergie. Le point de travail *Tight* pour sélectionner les traces est plus strict que le point de travail *Loose*, car il requiert qu'il y ait au moins 9 impacts dans le détecteur à silicium, et qu'il n'y ait pas de trou associé à la trace dans le détecteur à pixels. Autrement, les autres critères de qualité sont les mêmes que pour le point de travail *Loose*, si ce n'est qu'il n'y a pas de sélection sur le vertex avec lequel sont associées les traces pour le moment. Cette sélection sur les traces permet de minimiser le nombre de traces mal reconstruites.

Ensuite, la seconde étape consiste à essayer de faire correspondre un topo-cluster à chaque trace. Pour ce faire, la distance entre la trajectoire de la trace extrapolée dans les calorimètres et les topo-clusters est calculée avec :

$$\Delta R' = \sqrt{\left(\frac{\Delta\phi}{\sigma_\phi}\right)^2 + \left(\frac{\Delta\eta}{\sigma_\eta}\right)^2} \quad (4.6)$$

où  $\sigma_\phi$  and  $\sigma_\eta$  représentent la largeur angulaire des topo-clusters. Ces valeurs sont calculées comme étant l'écart-type des déplacements des cellules constitutives des topo-clusters en  $\phi$  et en  $\eta$  respectivement par rapport aux barycentres des topo-clusters [38] :

$$\sigma_\eta \simeq \sigma_\phi \simeq \arctan\left(\frac{\sqrt{\langle r^2 \rangle}}{|\vec{c}|}\right) \times \cosh(\eta) \quad (4.7)$$

où  $\sqrt{\langle r^2 \rangle}$  représente l'extension radiale du topo-cluster et  $\vec{c}$  représente le centre de gravité du topo-cluster, mesuré à partir du vertex primaire. Le topo-cluster avec le plus petit  $\Delta R'$  est associé à la trace, si  $\Delta R' < 1,64$ . Il peut arriver qu'il n'y ait aucun topo-cluster à une distance de moins de  $\Delta R' = 1,64$ , auquel cas il est estimé que la particule qui a formé la trace n'a pas déposé d'énergie dans le calorimètre. La valeur limite  $\Delta R' = 1,64$  permet de presque toujours inclure le topo-cluster qui correspond correctement à une trace donnée, tout en rejetant la majeure partie des clusters incorrects, comme décrit dans [37].

Puis, l'énergie déposée dans le calorimètre par la particule qui a créé la trace,  $\langle E_{\text{dep}}(p_{\text{T}}^{\text{trk}}) \rangle$ , est estimée, selon la position du topo-cluster et l'impulsion de la trace,  $p_{\text{T}}^{\text{trk}}$ . Pour ce faire, il est nécessaire de déterminer qu'elle est l'énergie déposée en moyenne dans les calorimètres par une particule chargée avec une impulsion transverse  $p_{\text{T}}^{\text{trk}}$ . Cette valeur est donnée par :

$$\langle E_{\text{dep}}(p_{\text{T}}^{\text{trk}}) \rangle = p_{\text{T}}^{\text{trk}} \times \left\langle \frac{E_{\text{ref}}^{\text{clus}}}{p_{\text{T,ref}}^{\text{trk}}} \right\rangle \quad (4.8)$$

où  $E^{\text{clus}}$  représente l'énergie du topo-cluster et  $p_{\text{T}}^{\text{trk}}$  représente l'impulsion transverse de la trace. La valeur attendue  $\left\langle \frac{E_{\text{ref}}^{\text{clus}}}{p_{\text{T,ref}}^{\text{trk}}} \right\rangle$  est déterminée à partir d'échantillons de particules uniques sans empilement en additionnant les énergies des topo-clusters dans un cône de taille  $\Delta R = 0,4$  autour de la position de la trace.

Il n'est pas rare qu'une particule puisse déposer de l'énergie dans plusieurs topo-clusters. C'est pourquoi, pour chaque paire de trace et de topo-cluster, l'algorithme évalue la probabilité que l'énergie de la particule ait été déposée dans plus d'un topo-cluster. Si

tel est le cas, des topo-clusters supplémentaires sont associés à la trace, afin de reconstruire l'énergie complète de la gerbe.

Enfin, l'énergie estimée déposée dans le calorimètre par la particule qui a produit la trace est soustraite aux topo-clusters qui lui sont associés. Si l'énergie dans le calorimètre après cette soustraction est cohérente avec les fluctuations attendues de la gerbe d'une seule particule, l'énergie restante dans le calorimètre est supprimée. Cette procédure a pour objectif de ne pas compter deux fois l'énergie venant des particules chargées qui laissent une trace dans le détecteur interne et un dépôt d'énergie dans les calorimètres.

Les topo-clusters qui ne sont associés à aucune trace sont estimés provenir de particules neutres, et ne sont par conséquent pas modifiés.

Dans le coeur des jets, les particules sont souvent produites avec de hautes impulsions et dans un environnement dense, ce qui réduit les performances du détecteur interne pour reconstruire les traces. Auparavant, toutes les traces avec  $p_T^{\text{trk}} > 40$  GeV n'étaient pas utilisées par l'algorithme. Désormais, pour toutes les traces jusqu'à  $p_T^{\text{trk}} = 100$  GeV, si l'énergie  $E_{\text{clus}}$  dans un cône de taille  $\Delta R = 0,15$  autour de la trajectoire extrapolée de la trace satisfait [39] :

$$\frac{E_{\text{clus}} - \langle E_{\text{dep}}(p_T^{\text{trk}}) \rangle}{\sigma(E_{\text{dep}})} > 33,2 \times \log_{10}\left(\frac{40 \text{ GeV}}{p_T^{\text{trk}}}\right) \quad (4.9)$$

où  $\langle E_{\text{dep}}(p_T^{\text{trk}}) \rangle$  représente l'énergie estimée déposée dans le calorimètre par un pion chargé, alors la soustraction n'est pas effectuée. Avec cette paramétrisation, la soustraction est effectuée pour des impulsions de traces plus faibles, sauf si l'activité calorimétrique mesurée par  $E_{\text{clus}}$  est très élevée, comme dans les environnements très denses où la précision de la soustraction est dégradée. La soustraction est complètement désactivée pour des traces avec  $p_T > 100$  GeV, et l'algorithme devient équivalent à celui des EMTopo. Le coefficient 33,2 dans l'Équation 4.9 contrôle l'impulsion maximale jusqu'à laquelle la soustraction est encore autorisée.

Puisque le détecteur interne ne couvre que la région  $|\eta| < 2,5$ , les topo-clusters et les Particle Flows Objects sont équivalents dans la région avant du détecteur.

À la fin de l'algorithme des Particle Flows, trois types d'objets subsistent : les traces, qui constituent les Particle Flow Objects *chargés*, les topo-clusters pour lesquels l'énergie des traces a été soustraite, et les topo-clusters non modifiés, qui constituent les Particle Flow Objects *neutres*. Les Particle Flow Objects *chargés* qui ne sont pas associés au vertex primaire sont enlevés avant que le jet ne soit reconstruit, afin de réduire la contribution venant de l'empilement. Cette procédure est appelée *Charged Hadron Subtraction* [50, 51].

La procédure des Particle Flows améliore la résolution angulaire et en  $p_T$  des jets, et réduit la contribution de l'empilement.

### 4.3.1.3 Les Track-CaloClusters

Les Track-CaloClusters [52] sont un autre type de constituants développés par ATLAS, qui utilisent l'information du détecteur interne et des calorimètres, mais d'une façon différente de celle des Particle Flows. En effet, la procédure des Track-CaloClusters est optimisée pour des jets de très haute impulsion transverse. Or, la résolution en impulsion des traces se dégrade avec leur impulsion. Cependant, il est toujours possible d'utiliser l'information angulaire de ces traces, dont la résolution angulaire s'améliore avec l'impulsion transverse. C'est pourquoi l'algorithme des Track-CaloClusters utilise l'information énergétique des topo-clusters et l'information angulaire des traces.

Pour ce faire, l'algorithme originel des Track-CaloClusters utilise des topo-clusters sur lesquels la technique d'étalonnage LCW est appliquée. De plus, les traces sont sélectionnées selon le point de travail *Loose*, mais sans condition sur le vertex (primaire ou d'empilement) d'origine. Ainsi, même les traces venant de l'empilement sont utilisées pour aborder le partage de l'énergie.

La première étape de l'algorithme des Track-CaloClusters consiste à faire correspondre un topo-cluster à chaque trace. Si l'incertitude sur l'extrapolation de la trace dans le calorimètre est plus grande que la largeur d'un topo-cluster, la trace n'est pas utilisée. Si ce n'est pas le cas, des paires trace - topo-cluster sont formées si leur séparation angulaire  $\Delta R$  est plus petite que  $\sqrt{(\sigma_{\text{cluster}})^2 + (\sigma_{\text{track}})^2}$ , où  $\sigma_{\text{cluster}}$  représente la largeur d'un topo-cluster et  $\sigma_{\text{track}}$  représente l'incertitude sur l'extrapolation de la trace. Puisque toutes les traces sont utilisées (même celles qui proviennent des vertex de l'empilement), il est possible qu'à un topo-cluster soient associées plusieurs traces, même venant de l'empilement. Toutes les paires trace - topo-cluster sont utilisées pour former les Track-CaloClusters, qui peuvent être répartis en 3 catégories. La première catégorie correspond aux traces associées à un topo-cluster. Ces traces sont appelées Track-CaloClusters *combinés*. La seconde et la troisième catégories correspondent aux topo-clusters auxquels ne sont pas associées de traces, et les traces qui ne sont pas associées à un topo-cluster. Ils sont appelés respectivement Track-CaloClusters *neutres* et Track-CaloClusters *chargés*. Même si les traces provenant des vertex d'empilement sont utilisées par l'algorithme des Track-CaloClusters, les Track-CaloClusters associés ne seront pas utilisés au moment de reconstruire les jets, afin de limiter l'impact de l'empilement.

Une fois que les paires trace - topo-cluster sont formées, les quadri-vecteurs des Track-CaloClusters individuels sont construits. Plusieurs cas, représentés Figure 4.5 [52], peuvent alors se présenter :

- Il y a une unique association entre une trace et un topo-cluster. Ce cas correspond au cas ① sur la Figure 4.5. Le quadri-vecteur du Track-CaloCluster est alors formé par l'information angulaire de la trace et l'information énergétique du topo-cluster (les constituants sont supposés être de masse nulle) :

$$\text{TTC}_{\textcircled{1}} = (p_{\text{T}}^{c_1}, \eta^{t_1}, \phi^{t_1}, m^{c_1} = 0) \quad (4.10)$$

- Un topo-cluster n'est associé à aucune trace. Ce cas correspond au cas ② sur la Figure 4.5. Le quadri-vecteur du Track-CaloCluster est alors formé directement



par celui du topo-cluster :

$$\text{TTC}_{\textcircled{2}} = (p_{\text{T}}^{c_7}, \eta^{c_7}, \phi^{c_7}, m^{c_7} = 0) \quad (4.11)$$

- Une trace n'est associée à aucun topo-cluster. Ce cas correspond au cas  $\textcircled{3}$  sur la Figure 4.5. Le quadri-vecteur du Track-CaloCluster est alors formé directement par celui de la trace :

$$\text{TTC}_{\textcircled{3}} = (p_{\text{T}}^{t_6}, \eta^{t_6}, \phi^{t_6}, m^{t_6} = 0) \quad (4.12)$$

- Deux traces sont associées au même topo-cluster. Ce cas correspond aux cas  $\textcircled{4}$  et  $\textcircled{5}$  sur la Figure 4.5. Deux Track-CaloClusters (un pour chaque trace) sont alors formés par l'information angulaire des traces et en répartissant l'énergie du topo-cluster selon la fraction d'impulsion transportée par chaque trace par rapport à la somme des impulsions de toutes les traces associées à ce topo-cluster ( $p_{\text{T}}[\mathbf{p}^a + \mathbf{p}^b]$  signifie l'impulsion transverse de la somme des quadri-vecteurs de  $a$  et  $b$ ) :

$$\text{TTC}_{\textcircled{4}} = (p_{\text{T}}^{c_2} \frac{p_{\text{T}}^{t_2}}{p_{\text{T}}[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}]}, \eta^{t_2}, \phi^{t_2}, 0) \quad (4.13)$$

$$\text{TTC}_{\textcircled{5}} = (p_{\text{T}}^{c_2} \frac{p_{\text{T}}^{t_3}}{p_{\text{T}}[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}]}, \eta^{t_3}, \phi^{t_3}, 0) \quad (4.14)$$

- Les cas plus complexes faisant intervenir plusieurs traces et plusieurs topo-clusters, tels que les cas  $\textcircled{6}$  et  $\textcircled{7}$  sur la Figure 4.5, ne sont pas détaillés ici.

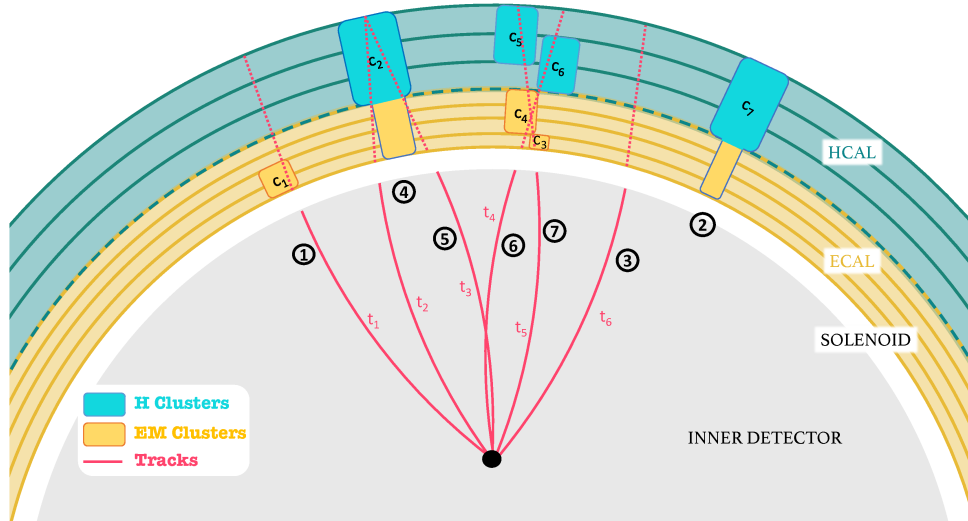


FIGURE 4.5 – Schéma de l'algorithme des Track-CaloClusters [52].

Pour des jets de haute impulsion, les Track-CaloClusters ont de meilleures performances en terme de résolution des variables de sous-structure des jets, telles que la masse (définie après) ou  $D_2$  [53, 54].

Un autre type de constituants, combinant les algorithmes des Particle Flows et des Track-CaloClusters a récemment été développé par ATLAS. Ces constituants, appelés *Unified Flow Objects* (ou plus brièvement *UFOs*) sont détaillés Section 5.1.

### 4.3.1.4 Les vraies particules

Les jets peuvent aussi être reconstruits à partir des *vraies* particules, c'est-à-dire des particules simulées au niveau du générateur. Toutes les particules stables dans le détecteur issues de la collision d'intérêt et ayant une durée de vie  $\tau$  telles que  $c\tau > 10$  mm sont utilisées. Les particules qui sont censées ne laisser que des dépôts d'énergie négligeables dans le calorimètre, c'est-à-dire les muons et les neutrinos, sont exclues. Les jets reconstitués à partir de ces particules sont appelés jets *vrais*, et peuvent par exemple servir de référence pour l'étalonnage des jets reconstitués à partir des détecteurs, comme décrit Section 5.2.

## 4.3.2 Les algorithmes de regroupement des constituant des jets

Une fois que les constituants sont définis, ils servent d'entrée à des algorithmes de regroupement, afin de définir des *jets*. Il existe plusieurs algorithmes pour regrouper ces constituants. Ces algorithmes doivent remplir des conditions de sécurité infrarouge et de sécurité collinéaire, afin qu'ils puissent être reproduits dans les calculs analytiques, et d'avoir des prédictions théoriques avec lesquelles comparer les données. La sécurité infrarouge signifie que le résultat de l'algorithme doit être le même si l'on ajoute des constituants de basse impulsion. La sécurité collinéaire signifie que le résultat de l'algorithme doit être le même si des séparations avec de petits angles se produisent lors de la gerbe de partons.

Les algorithmes séquentiels de regroupement sont une classe d'algorithmes de regroupement qui remplissent les conditions de sécurité infrarouge et de sécurité collinéaire. Ces algorithmes introduisent une distance entre les constituants  $i$  et  $j$ ,  $d_{ij}$ , qui est donnée par [55] :

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (4.15)$$

où :

- $p$  représente un paramètre libre,
- et  $R$  représente aussi un paramètre libre qui contrôle la taille du jet formé.

Ces algorithmes introduisent aussi une autre distance entre le constituant  $i$  et le faisceau  $B$ ,  $d_{iB}$ , donnée par :

$$d_{iB} = p_{T,i}^{2p} \quad (4.16)$$

Les algorithmes séquentiels de regroupement procèdent comme suit :

- Les distances  $d_{ij}$  sont calculées entre toutes les paires de constituants et les distances  $d_{iB}$  sont calculées pour tous les constituants.
- Si la plus petite distance est l'une des  $d_{ij}$ , les deux constituants  $i$  et  $j$  sont regroupés en un nouveau constituant.
- Sinon le constituant  $i$  forme un jet et est retiré de la liste des constituants.
- Après chaque étape, les distances  $d_{ij}$  et  $d_{iB}$  sont recalculées à partir de la nouvelle liste de constituants.

- La procédure est répétée jusqu'à ce qu'il n'y ait plus de constituants.

Plusieurs choix sont possibles pour les paramètres  $p$  et  $R$  qui interviennent dans la distance  $d_{ij}$ . Les trois choix les plus courants pour  $p$  sont :

- $p = 0$ , auquel cas l'algorithme est dit de *Cambridge/Aachen*. Cette configuration implique que l'algorithme est indépendant de l'impulsion transverse des constituants et est seulement déterminé par leur séparation angulaire. Les paires de constituants avec une faible séparation angulaire sont regroupés en premier. Un exemple de formation de jet à partir de l'algorithme de Cambridge/Aachen est représenté Figure 4.6a [55].
- $p = 1$ , auquel cas l'algorithme est appelé  $k_t$ . Cette configuration implique que les constituants de plus basse impulsion transverse sont regroupés en premier. Un exemple de formation de jet à partir de l'algorithme  $k_t$  est représenté Figure 4.6b [55].
- $p = -1$ , auquel cas l'algorithme est appelé *anti- $k_t$*  [55]. Cette configuration implique que les constituants de plus haute impulsion transverse sont regroupés en premier. Un exemple de formation de jet à partir de l'algorithme anti- $k_t$  est représenté Figure 4.6c [55].

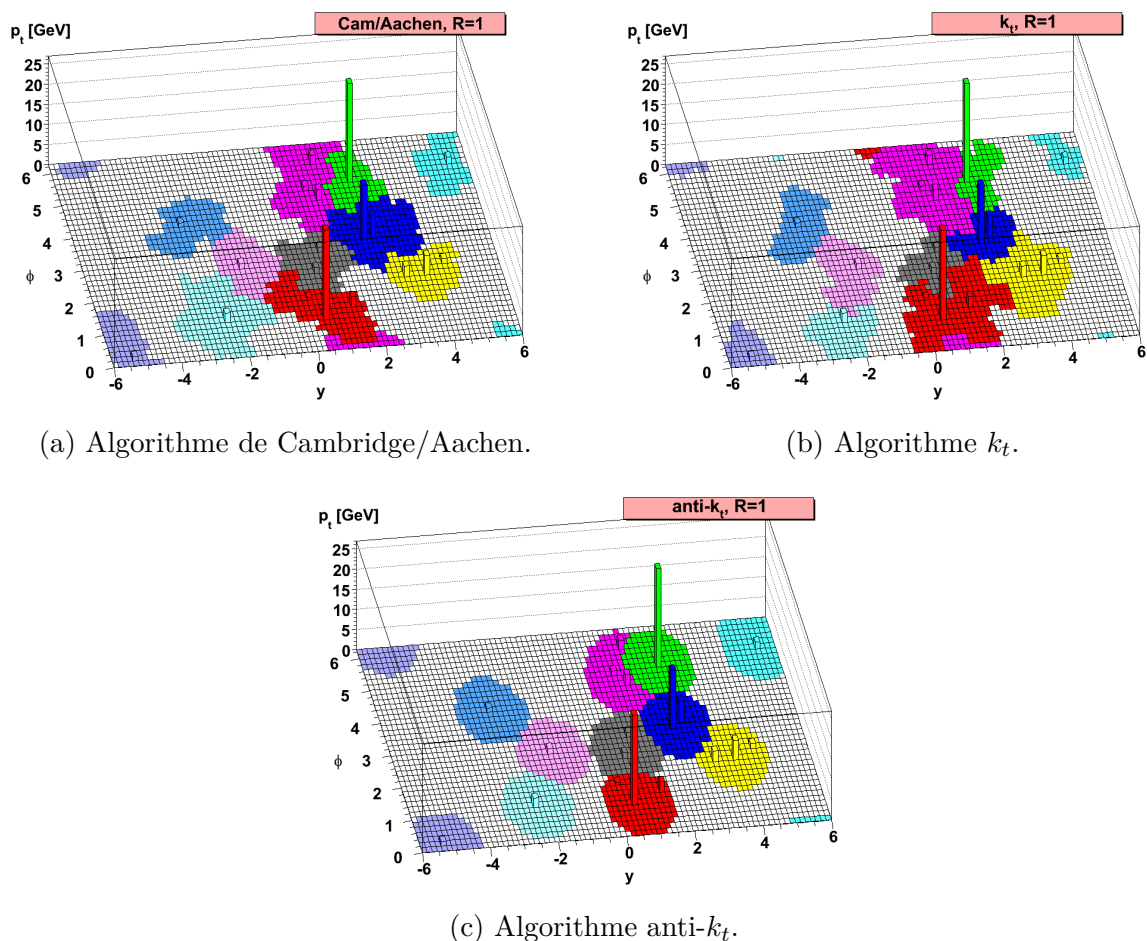


FIGURE 4.6 – Différents algorithmes de regroupement des constituants des jets [55].

L'algorithme adopté par ATLAS est l'algorithme anti- $k_t$ , car il est moins sensible aux radiations de faible impulsion (comme l'empilement), les constituants de plus haute impulsion transverse étant regroupés en premier. De plus, la forme des jets reconstitués avec l'algorithme anti- $k_t$  est approximativement circulaire, comme représenté Figure 4.6c. L'implémentation du logiciel *FASTJET* [56] est utilisée par ATLAS afin de former les jets avec l'algorithme anti- $k_t$ .

Deux choix sont utilisés pour le paramètre  $R$  par ATLAS. Le premier,  $R = 0,4$  est suffisant pour contenir les radiations de jets créés à partir de quarks ou de gluons. Le second,  $R = 1,0$ , sert à capturer les radiations émises par les désintégrations hadroniques de particules massives ou de particules boostées, telles que les bosons  $W$ ,  $Z$ , ou encore les quarks  $t$ . En effet, la taille de ces jets est environ donnée par :  $R \sim \frac{2m}{p_T}$ .

### 4.3.3 La masse des jets

La masse d'une particule est l'une de ses propriétés les plus distinctives. Puisque les jets de rayon  $R = 1,0$  sont utilisés pour reconstruire les désintégrations d'objets massifs, la masse associée à un jet a pour objectif de correspondre à la masse de la particule à son

origine (par exemple un boson  $W$  ou un quark  $t$ ).

La masse d'un jet est reconstituée selon la formule suivante :

$$m = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2} \quad (4.17)$$

où la somme est effectuée sur tous les constituants associés à un jet.

Les constituants provenant des calorimètres sont supposés être de masse nulle, de façon à ce que  $|\vec{p}_i| = E_i$ . Ainsi, pour ces constituants, le vecteur  $\vec{p}$  peut être déterminé en utilisant les relations :

$$p_T = \frac{E}{\cosh(\eta)} \quad (4.18)$$

et ensuite :

$$p_x = p_T \times \cos(\phi) \quad (4.19)$$

$$p_y = p_T \times \sin(\phi) \quad (4.20)$$

$$p_z = p_T \times \sinh(\eta) \quad (4.21)$$

Au contraire, pour les constituants provenant du détecteur interne, leur énergie n'est pas connue, mais leur impulsion transverse l'est. Leur énergie peut être déterminée en utilisant la relation suivante, généralement en supposant la masse d'un pion chargé ( $m \simeq 139$  MeV) :

$$E = \sqrt{(p_T \times \cosh(\eta))^2 + m^2} \quad (4.22)$$

La masse d'un jet est extrêmement sensible à la répartition des constituants à l'intérieur du jet et à la précision avec laquelle leur position est estimée, et donc à l'empilement.

La masse d'un jet *vrai* est obtenue en appliquant l'Équation 4.17 aux particules simulées au niveau du générateur, telles que décrites dans la Section 4.3.1.4, et dont l'énergie, l'impulsion, et la direction sont connues. La masse des jets *vrais* ainsi obtenue est notée  $m_{\text{true}}$ .

### 4.3.4 Les algorithmes de différentiation de l'empilement s'appliquant aux constituants

Alors qu'il est possible de différencier les traces qui viennent de la collision d'intérêt par rapport à celles qui proviennent de l'empilement grâce au vertex duquel elles proviennent, ce n'est pas le cas pour les topo-clusters. C'est pourquoi des algorithmes permettant de réduire la contribution provenant de l'empilement sur les topo-clusters ont été développés. Ces algorithmes s'appliquent à tous les constituants neutres de l'évènement (comme par exemple les Particle Flows neutres ou les Track-CaloClusters neutres) avant que l'algorithme de regroupement des constituants des jets ne soit appliqué.

### 4.3.4.1 Constituent Subtraction

Le premier algorithme détaillé est appelé *Constituent Subtraction* [57] (ou plus brièvement *CS*). La première étape de cet algorithme consiste à recouvrir uniformément chaque évènement par des particules dites « fantômes ». Ces particules fantômes ont une impulsion transverse,  $p_T^g$ , donnée par :

$$p_T^g = A_g \times \rho \quad (4.23)$$

où :

- $A_g$  représente l'aire d'une particule fantôme, qui est fixée à  $\Delta\eta \times \Delta\phi = 0,1 \times 0,1 = 0,01$ ,
- et  $\rho$  représente la densité en impulsion transverse provenant de l'empilement. Afin de calculer cette quantité, pour chaque évènement indépendamment, des jets sont reconstruits avec l'algorithme  $k_t$  avec paramètre  $R = 0,4$ . Seuls les jets qui satisfont  $|\eta| < 2,0$  sont utilisés. Ensuite, l'aire  $A$  de chacun de ces jets ( $\simeq \pi R^2$ ) est calculée. Pour ce faire, chaque évènement est recouvert par un second ensemble de particules fantômes, mais cette fois-ci d'impulsion transverse extrêmement faible, de façon à ce qu'elles ne modifient pas la direction ou l'impulsion des jets reconstruits. Chaque jet est alors re-reconstruit avec l'algorithme  $k_t$  en incluant les constituants initiaux et le second ensemble de particules fantômes, et leur aire est déterminée à partir du nombre relatif de particules fantômes qui leur sont associées après regroupement. Enfin,  $\rho$  est donnée par la médiane de la distribution de  $\frac{p_T}{A}$  pour chacun des jets ainsi reconstruits à partir de cet évènement :  $\rho = \text{median}\left\{\frac{p_{T,i}^{\text{jet}}}{A_i^{\text{jet}}}\right\}$ , où l'indice  $i$  parcourt tous les jets de l'évènement. Ainsi,  $\rho$  est une quantité calculée évènement par évènement. L'algorithme  $k_t$  est utilisé pour regrouper les jets dans le calcul de  $\rho$  car il regroupe les constituants de plus basse impulsion (c'est-à-dire provenant de l'empilement) en premier. Le critère de sélection  $|\eta| < 2,0$  est utilisé car la valeur de  $\rho$  devient proche de 0 dans la région au-delà de  $|\eta| = 2,0$ . Cette baisse est due à la plus faible occupation dans la région  $|\eta| > 2,0$  par rapport à la région  $|\eta| < 2,0$ , qui résulte d'une segmentation plus grossière dans la région  $|\eta| > 2,0$  [58]. La même méthode utilisant  $\rho$  est utilisée pour étalonner les jets de rayon  $R = 0,4$ , comme décrit Section 5.2.

Ainsi,  $p_T^g$  correspond à la contribution attendue provenant de l'empilement dans une aire de 0,01, et la somme des  $p_T^g$  représente une estimation de la contribution moyenne due à l'empilement.

Une fois que l'évènement est recouvert par les particules fantômes, la distance  $\Delta R_{i,k}$  est calculée entre chaque constituant  $i$  et chaque particule fantôme  $k$ . Les paires constituant - particule fantôme sont ensuite ordonnées par ordre croissant en  $\Delta R_{i,k}$ , et l'algorithme procède de façon itérative à travers chaque paire, en modifiant le  $p_T$  de chaque constituant par :

- Si  $p_{T,i} > p_{T,k}$  :  $p_{T,i} \rightarrow p_{T,i} - p_{T,k}$  et  $p_{T,k} \rightarrow 0$

— Sinon :  $p_{T,k} \rightarrow p_{T,k} - p_{T,i}$  et  $p_{T,i} \rightarrow 0$

jusqu'à ce que  $\Delta R_{i,k} > \Delta R_{\max}$ . La valeur utilisée par ATLAS est  $\Delta R_{\max} = 0,25$ , basé sur des études de performances de jets de rayon  $R = 0,4$  [59].

Les particules fantômes sont enfin éliminées de l'évènement.

#### 4.3.4.2 SoftKiller

Le deuxième algorithme est appelé *SoftKiller* [60] (ou plus brièvement *SK*). Cet algorithme applique une coupure en impulsion transverse sur les constituants des jets, car l'empilement correspond à des émissions de basse impulsion transverse. La coupure est choisie de façon à ce que la valeur de  $\rho$  soit approximativement zéro après sélection. Pour ce faire, l'évènement est divisé en une grille dans le plan  $(\eta, \phi)$ . ATLAS utilise une échelle de longueur (c'est-à-dire l'espacement de la grille)  $\ell = 0,6$ , basé sur des études de performances de jets de rayon  $R = 0,4$  [59]. La coupure en impulsion transverse est ensuite choisie de façon à ce que la moitié des cellules de la grille soient vides après sélection.

Les meilleures performances sont obtenues en appliquant une forme de correction telle que Constituent Subtraction aux constituants avant d'appliquer Soft Killer : puisque SK effectue une coupure basée sur le  $p_T$  des topo-clusters, la soustraction effectuée par CS doit être appliquée aux constituants avant d'appliquer SK. Cela permet de corriger les quadri-vecteurs des topo-clusters avant de décider de les supprimer ou non. C'est pourquoi, les algorithmes de Constituent Subtraction et de SoftKiller sont appliqués successivement (dans cet ordre). La combinaison des deux est notée *CS+SK*.

### 4.3.5 Les algorithmes de grooming

Les algorithmes de *grooming* ont pour objectif de supprimer les radiations de basse impulsion transverse ou avec un grand angle d'émission des jets, car ces radiations proviennent très probablement de l'empilement ou de l'évènement sous-jacent (interactions secondaires qui ne proviennent pas de la collision principale. L'évènement sous-jacent inclut parfois l'empilement, mais aussi des radiations dans l'état initial ou final, ou des interactions multiples de partons). Les algorithmes de grooming, contrairement à ceux décrits dans la Section 4.3.4, s'appliquent aux jets une fois que ceux-ci ont formés avec un algorithme de regroupement des constituants. Les algorithmes de grooming ne sont appliqués qu'aux jets de rayon  $R = 1,0$ , car, du fait de leur plus grande taille, ils contiennent plus d'empilement, et de ce fait les méthodes d'étalonnage décrites dans la Section 5.2 pour les jets de rayon  $R = 0,4$  ne sont pas optimales. De plus les variables de sous-structure des jets, très sensibles aux radiations venant de l'empilement, sont très fréquemment utilisées pour les jets de rayon  $R = 1,0$ , alors qu'elles ne sont pas fréquemment utilisées pour les jets de rayon  $R = 0,4$ .

#### 4.3.5.1 Trimming

L'algorithme de grooming utilisé jusqu'à présent par ATLAS est appelé *trimming* [61]. La première étape de cet algorithme est de re-regrouper les constituants des jets en des

*sous-jets* de rayon  $R_{\text{sub}}$  en utilisant l'algorithme  $k_t$ , qui regroupe les constituants de plus basse impulsion (c'est-à-dire provenant de l'empilement) en premier. Ensuite les sous-jets avec une impulsion transverse inférieure à une fraction  $f_{\text{cut}}$  de l'impulsion transverse du jet originel sont enlevés. ATLAS utilise les paramètres  $R_{\text{sub}} = 0,2$  et  $f_{\text{cut}} = 5\%$ . Cette procédure est schématisée Figure 4.7 [62].

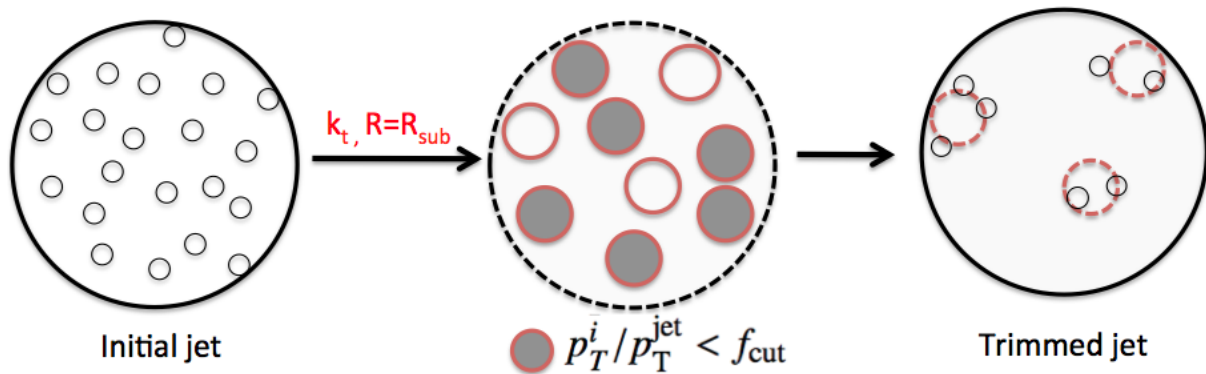


FIGURE 4.7 – Schéma de l'algorithme de trimming [62].

#### 4.3.5.2 Soft-drop

Plus récemment, ATLAS a étudié un autre algorithme de grooming appelé *Soft-drop* [63]. Cet algorithme a pour objectif de retirer les radiations de faible impulsion et avec un grand angle d'un jet. La première étape de cet algorithme consiste à re-regrouper les constituants en utilisant l'algorithme de Cambridge/Aachen, qui regroupe les constituants selon leur séparation angulaire seulement. Ensuite, la séquence de regroupement est parcourue en sens inverse, c'est-à-dire en partant de la radiation avec le plus grand angle d'émission, et en itérant vers le coeur du jet. À chaque étape, la condition suivante est testée :

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} < z_{\text{cut}} \left( \frac{\Delta R_{1,2}}{R} \right)^\beta \quad (4.24)$$

où les indices 1 et 2 désignent respectivement les branches de plus haute impulsion et de plus basse impulsion, et les paramètres  $z_{\text{cut}}$  et  $\beta$  déterminent la quantité de radiations de basse impulsion et à grand angle d'émission qui est éliminée. ATLAS utilise les paramètres  $z_{\text{cut}} = 0,1$  et  $\beta = 1,0$ . Si l'étape remplit cette condition, la branche de plus faible impulsion est supprimée, et le processus est répété sur la branche de plus grande impulsion. Si la condition n'est pas satisfaite, le processus s'achève et les constituants restants forment le jet final. Cette procédure est schématisée Figure 4.8 [64].



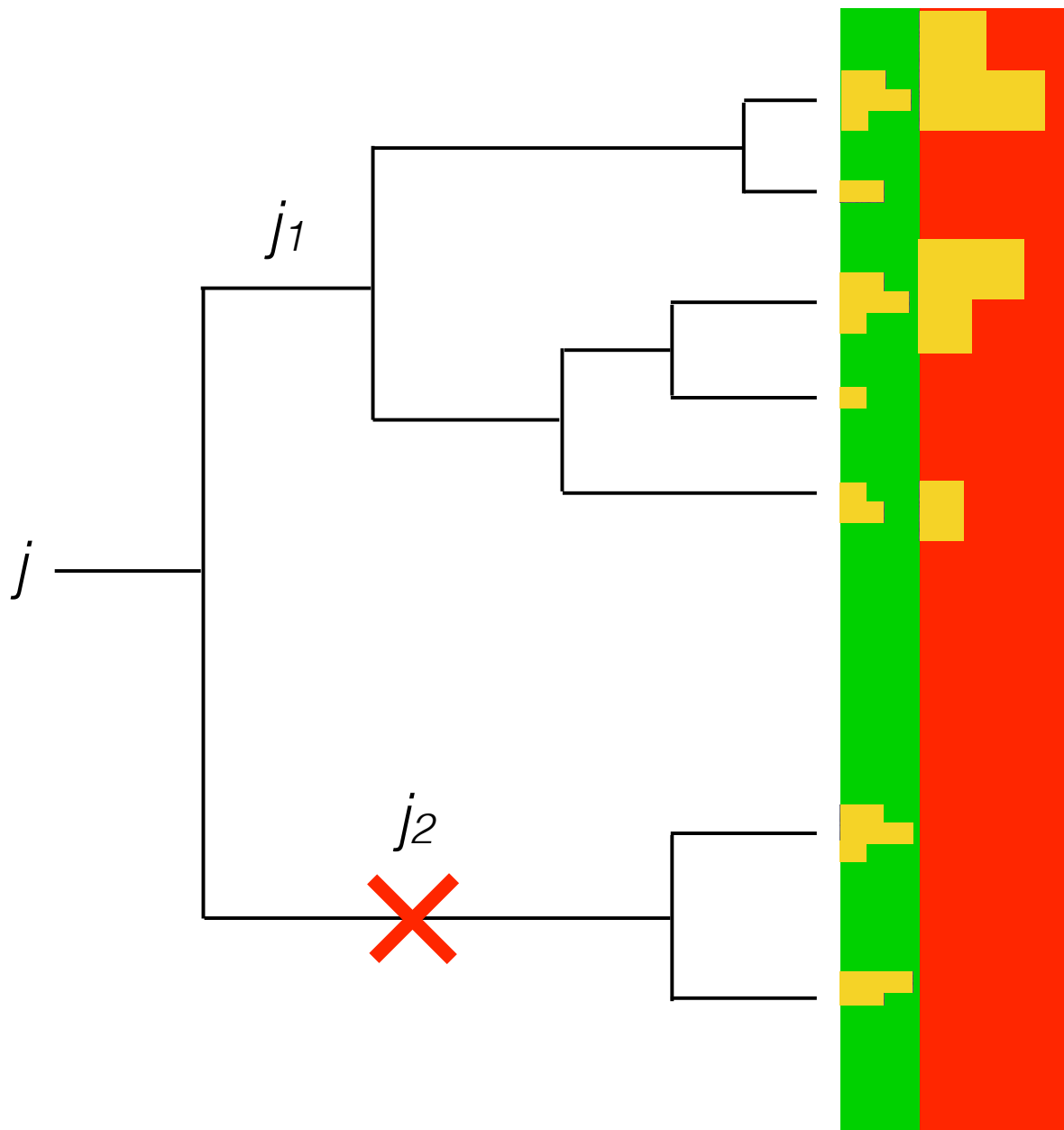


FIGURE 4.8 – Schéma de la première itération de l’algorithme Soft-drop appliqué lors du regroupement avec l’algorithme de Cambridge/Aachen d’un jet [64]. La partie verte et rouge représentent le calorimètre électromagnétique et hadronique respectivement.

Les rectangles jaunes représentent des dépôts d’énergie. Les sous-jets  $j_1$  et  $j_2$  sont évalués en utilisant l’Équation 4.24. Si la condition est remplie, la branche de plus basse impulsion (ici  $j_2$ ) est enlevée, et la procédure continue avec la branche  $j_1$ .

# Chapitre 5

## L'étalonnage des jets

### Sommaire du Chapitre

---

<b>5.1</b>	<b>Les Unified Flow Objects . . . . .</b>	<b>70</b>
<b>5.2</b>	<b>La chaîne d'étalonnage des jets . . . . .</b>	<b>75</b>
5.2.1	La correction d'empilement . . . . .	76
5.2.2	La correction en énergie, masse et pseudo-rapacité . . . . .	78
5.2.3	Global Sequential Calibration . . . . .	82
5.2.4	La correction « <i>in-situ</i> » . . . . .	86
<b>5.3</b>	<b>Performances et étalonnage des UFOs de rayon <math>R=0,4</math> . . . . .</b>	<b>87</b>
5.3.1	Échantillons utilisés . . . . .	87
5.3.2	Sélections . . . . .	88
5.3.3	Résultats avant étalonnage . . . . .	90
5.3.4	Résultats après correction de l'empilement . . . . .	101
5.3.5	Résultats après correction absolue en énergie, masse et pseudo-rapacité . . . . .	112
5.3.6	Résultats après GSC . . . . .	117
<b>5.4</b>	<b>Conclusion . . . . .</b>	<b>121</b>

---

J'ai contribué à l'étalonnage et à l'étude des performances d'un nouveau type de constituants développé par ATLAS : les *Unified Flow Objects*, et ce pour des jets de rayon  $R = 0,4$ . Ce travail a été effectué en collaboration avec Pierre-Antoine Delsart, du Laboratoire de Physique Subatomique et de Cosmologie de Grenoble. Ce travail a donné lieu à l'élaboration d'une note publique [65]. Dans un premier temps, les motivations pour les Unified Flow Objects sont décrites et leur performance dans le contexte des jets de rayon  $R = 1,0$  sont exposées dans la Section 5.1. Ensuite la chaîne d'étalonnage des jets de rayon  $R = 0,4$  est détaillée dans la Section 5.2. Enfin les performances et l'étalonnage des jets de rayon  $R = 0,4$  reconstitués à partir des Unified Flow Objects sont décrits dans la Section 5.3.

## 5.1 Les Unified Flow Objects

Comme décrit dans la Section 4.3.1, ATLAS a développé deux types de constituants des jets permettant de combiner les informations complémentaires des calorimètres et du détecteur interne : les *Particle Flows*, optimisés pour des jets de basse impulsion transverse, et les *Track-CaloClusters*, optimisés pour des jets de haute impulsion transverse.

En effet, à haute impulsion transverse, les performances des Particle Flows sont moindres que celles des Track-CaloClusters car la résolution en impulsion transverse du détecteur interne se dégrade avec celle-ci et car l'environnement à l'intérieur des jets est très dense, ce qui détériore aussi la résolution en impulsion transverse du détecteur interne. Au contraire, à basse impulsion transverse, les performances des Track-CaloClusters sont moindres que celles des Particle Flows car l'algorithme n'utilise pas une compréhension détaillée de la corrélation entre les traces et les topo-clusters, et par conséquent, la redistribution de l'énergie des topo-clusters peut être incorrecte.

C'est pourquoi ATLAS a développé un nouveau type de constituants, appelés les *Unified Flow Objects* (ou plus brièvement *UFOs*) [66], qui combine les aspects bénéfiques des Particle Flows et des Track-CaloClusters, afin d'obtenir des performances optimales à basse et à haute impulsion transverse.

Le schéma de l'algorithme des Unified Flow Objects est représenté Figure 5.1 [66]. La première étape de l'algorithme consiste à appliquer l'algorithme des Particle Flows. Les Particle Flows chargés qui ne sont pas associés au vertex primaire sont enlevés (ce qui correspond à la procédure de *Charged Hadron Subtraction*, ou plus brièvement *CHS*). Les Particle Flows restants sont classifiés en deux catégories : les Particle Flows *chargés* et les Particle Flows *neutres*. Des algorithmes de suppression de l'empilement, tels que *CS+SK*, peuvent éventuellement être appliqués aux Particle Flows neutres. Par la suite et par convention, lorsque les algorithmes *CS+SK* sont appliqués aux constituants neutres, l'algorithme *CHS* est aussi appliqué en parallèle aux constituants chargés. Ensuite, une version modifiée de l'algorithme des Track-CaloClusters est appliquée aux Particle Flows restants : seules les traces provenant du vertex primaire sont utilisées en entrée de l'algorithme, afin d'éviter les instabilités dues à l'empilement. Les traces qui ont déjà été utilisées pour l'algorithme des Particle Flows ne sont pas utilisées, car elles ont déjà été associées et leurs contributions dans les calorimètres ont déjà été soustraites. Enfin, l'algorithme des Track-CaloClusters procède comme décrit dans la Section 4.3.1.3, en utilisant la collection de traces modifiée et les Particle Flows chargés non soustraits afin de répartir l'énergie des Particle Flows neutres à la place des topo-clusters. Cette approche permet de bénéficier au maximum de la procédure de soustraction des Particle Flows à basse impulsion transverse, et de la procédure de répartition de l'énergie des topo-clusters provenant de l'algorithme des Track-CaloClusters à haute impulsion transverse.

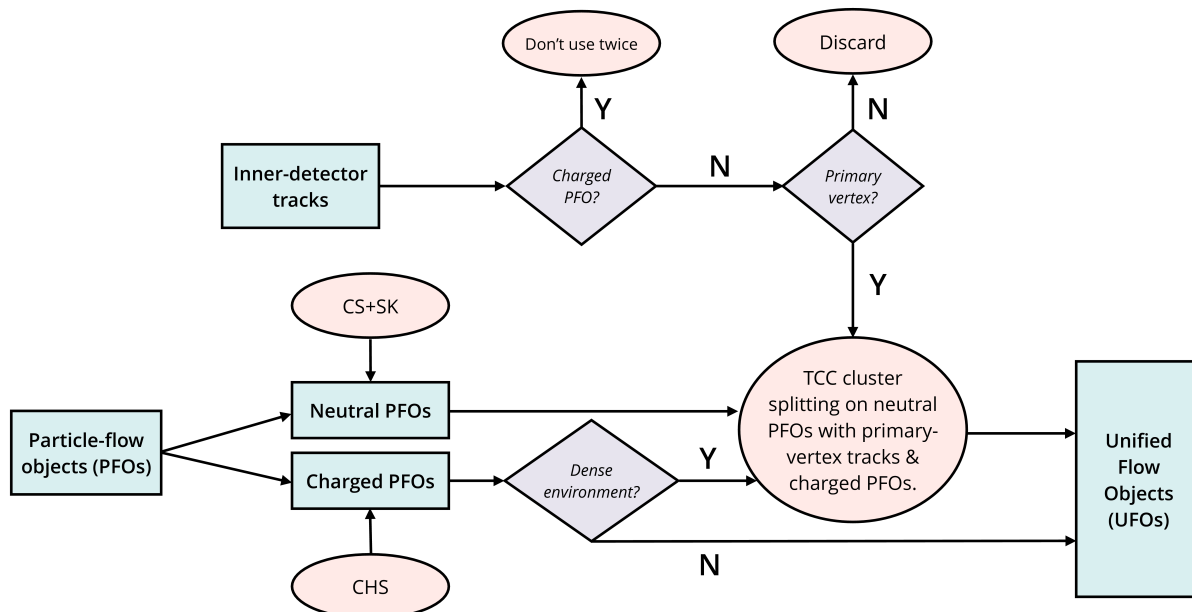


FIGURE 5.1 – Schéma de l’algorithme des Unified Flow Objects [66].

Les performances des Unified Flow Objects ont été étudiées par ATLAS dans le contexte des jets de rayon  $R = 1,0$  [66], et comparées avec la référence actuelle d’ATLAS : les LC Topos avec *trimming*. Par exemple, la Figure 5.2 [66] représente une comparaison en terme d’identification de bosons  $W$  entre 4 différentes définitions de jets avec l’algorithme de *trimming* appliqué aux constituants : *LC Topo*, *TCC*, *EM PFlow* et *UFO*. Sur la Figure 5.2, l’axe des  $y$  représente le rejet du bruit de fond, défini comme l’inverse de l’efficacité de sélection lié à l’identification des jets provenant du bruit de fond QCD ( $\frac{1}{\varepsilon_{\text{bkg}}}$ ). L’efficacité de sélection est définie comme le nombre de jets qui passent une coupure de sélection, divisé par le nombre de jets initiaux :  $\varepsilon = \frac{N_{\text{jets after selection}}}{N_{\text{jets before selection}}}$ . L’axe des  $x$  représente l’efficacité de sélection liée à l’identification des jets provenant de la désintégration d’un boson  $W$  en deux quarks ( $\varepsilon_{\text{sig}}$ ). La courbe est obtenue en faisant varier la coupure de sélection. Ainsi, les meilleures performances sont obtenues pour les courbes qui se rapprochent le plus du coin supérieur droit de la figure. Les Figures 5.2a et 5.2b montrent les performances pour deux intervalles différents en impulsion transverse *vraie* des jets :  $300 \text{ GeV} < p_{\text{T}}^{\text{true}} < 500 \text{ GeV}$  et  $1000 \text{ GeV} < p_{\text{T}}^{\text{true}} < 1500 \text{ GeV}$  respectivement. Comme le montre la Figure 5.2, à basse impulsion transverse, les EM PFlow ont de meilleures performances que les TCC. Au contraire, à haute impulsion transverse, les TCC ont de meilleures performances que les EM PFlow. Dans les deux cas, les UFO montrent des performances supérieures ou égales aux EM PFlows et aux TCCs.

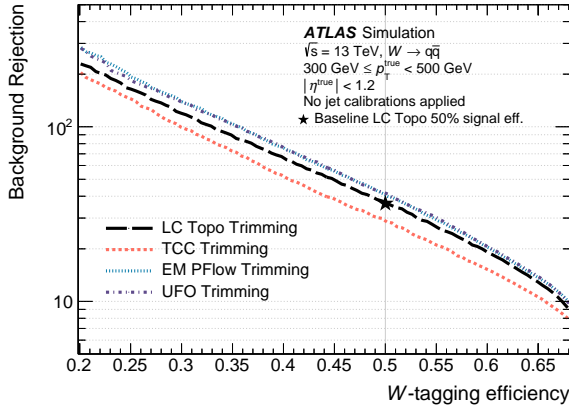
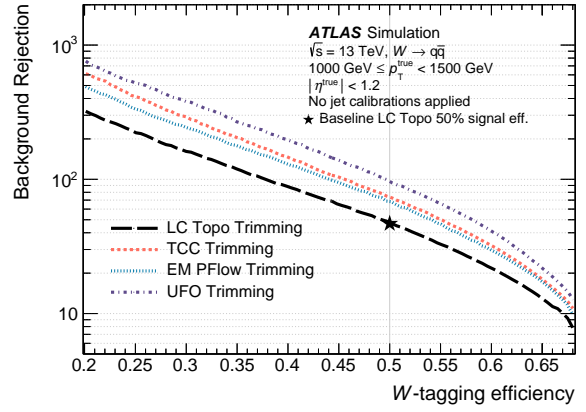

 (a)  $300 \text{ GeV} < p_T^{\text{true}} < 500 \text{ GeV}$ 

 (b)  $1000 \text{ GeV} < p_T^{\text{true}} < 1500 \text{ GeV}$ 

 FIGURE 5.2 – Rejet du bruit de fond en fonction de l'efficacité du signal pour l'identification de jets provenant de la désintégration d'un boson  $W$  [66].

La Figure 5.3 [66] représente la même quantité que la Figure 5.2, mais cette fois-ci en terme de performances d'identification de jets issus de la désintégration d'un quark  $t$  en trois quarks. Les Figures 5.3a et 5.3b montrent les performances pour deux intervalles différents en impulsion transverse *vraie* des jets :  $500 \text{ GeV} < p_T^{\text{true}} < 1000 \text{ GeV}$  et  $1000 \text{ GeV} < p_T^{\text{true}} < 1500 \text{ GeV}$  respectivement. Encore une fois, les UFO montrent des performances supérieures ou égales aux EM PFlow et aux TCC, que ce soit à basse impulsion transverse ou à haute impulsion transverse.

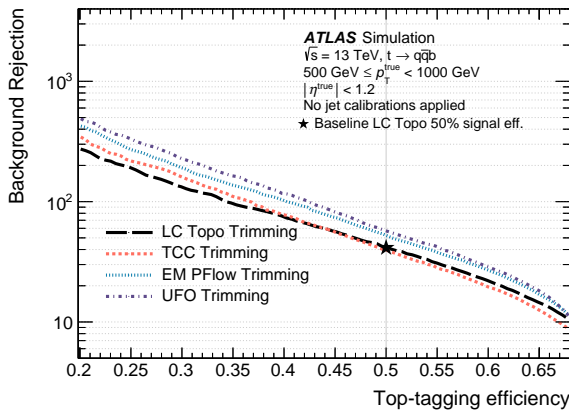
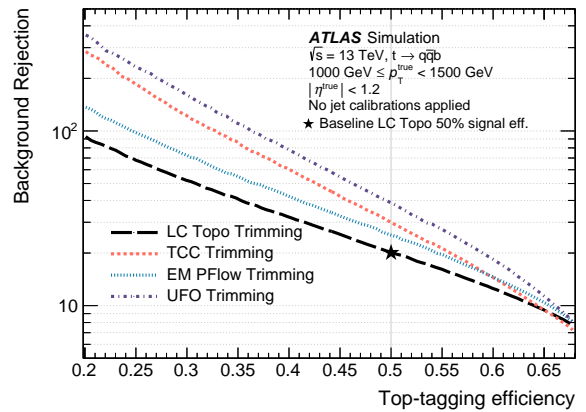

 (a)  $500 \text{ GeV} < p_T^{\text{true}} < 1000 \text{ GeV}$ 

 (b)  $1000 \text{ GeV} < p_T^{\text{true}} < 1500 \text{ GeV}$ 

 FIGURE 5.3 – Rejet du bruit de fond en fonction de l'efficacité du signal pour l'identification de jets provenant de la désintégration d'un quark  $t$  [66].

La Figure 5.4 [66] représente une comparaison en terme de résolution en masse (définie dans la Section 5.2) en fonction de l'impulsion transverse *vraie* entre 4 différentes définitions de jets avec l'algorithme de *trimming* appliqué aux constituants : *LC Topo*, *TCC*, *EM PFlow* et *UFO*. Les meilleures performances sont obtenues pour les courbes les plus basses. Les Figures 5.4a et 5.4b montrent les performances pour des jets issus de la

désintégration de bosons  $W$  et de quarks  $t$  respectivement. Les jets sont étalonnés avec une procédure telle que décrite dans la Section 5.2 (avec des différences entre les jets de rayon  $R = 0,4$  et  $R = 1,0$  qui ne sont pas décrites ici), ce qui est représenté par l'inscription «  $JES + JMS$  » sur la Figure 5.4. Les UFO montrent des performances supérieures ou égales aux EM PFlow et aux TCC, que ce soit à basse impulsion transverse ou à haute impulsion transverse, mais aussi pour des jets issus de la désintégration de bosons  $W$  ou de quarks  $t$ .

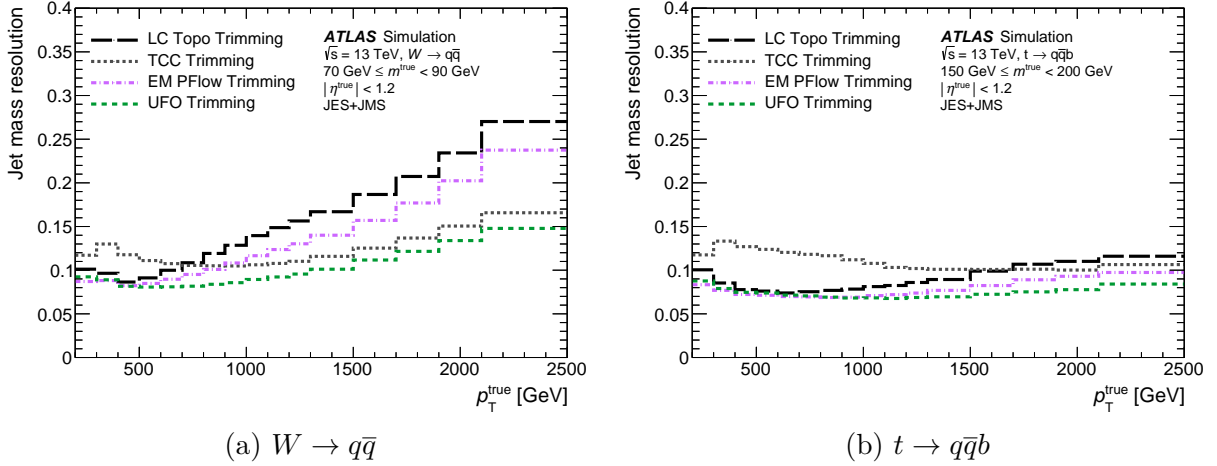


FIGURE 5.4 – Résolution en masse en fonction de l'impulsion transverse *vraie* [66].

La Figure 5.5 [66] représente une comparaison en terme de résolution en impulsion (définie dans la Section 5.2) en fonction de l'impulsion transverse *vraie* entre 4 différentes définitions de jets avec l'algorithme de *trimming* appliqué aux constituants : *LC Topo*, *TCC*, *EM PFlow* et *UFO*. Les jets sont issus d'un quark ou d'un gluon seulement. Les meilleures performances sont obtenues pour les courbes les plus basses. Les meilleures performances sont donc obtenues pour les TCC et les LC Topo. Ceci est dû au fait que les UFO utilisent des topo-clusters qui sont à l'échelle d'énergie électromagnétique, alors que les TCC et les LC Topo utilisent des topo-clusters qui sont étalonnés avec la technique *LCW*, décrite dans la Section 4.3.1.1. Les UFO montrent des performances égales aux EM PFlow, que ce soit à basse impulsion transverse ou à haute impulsion transverse.

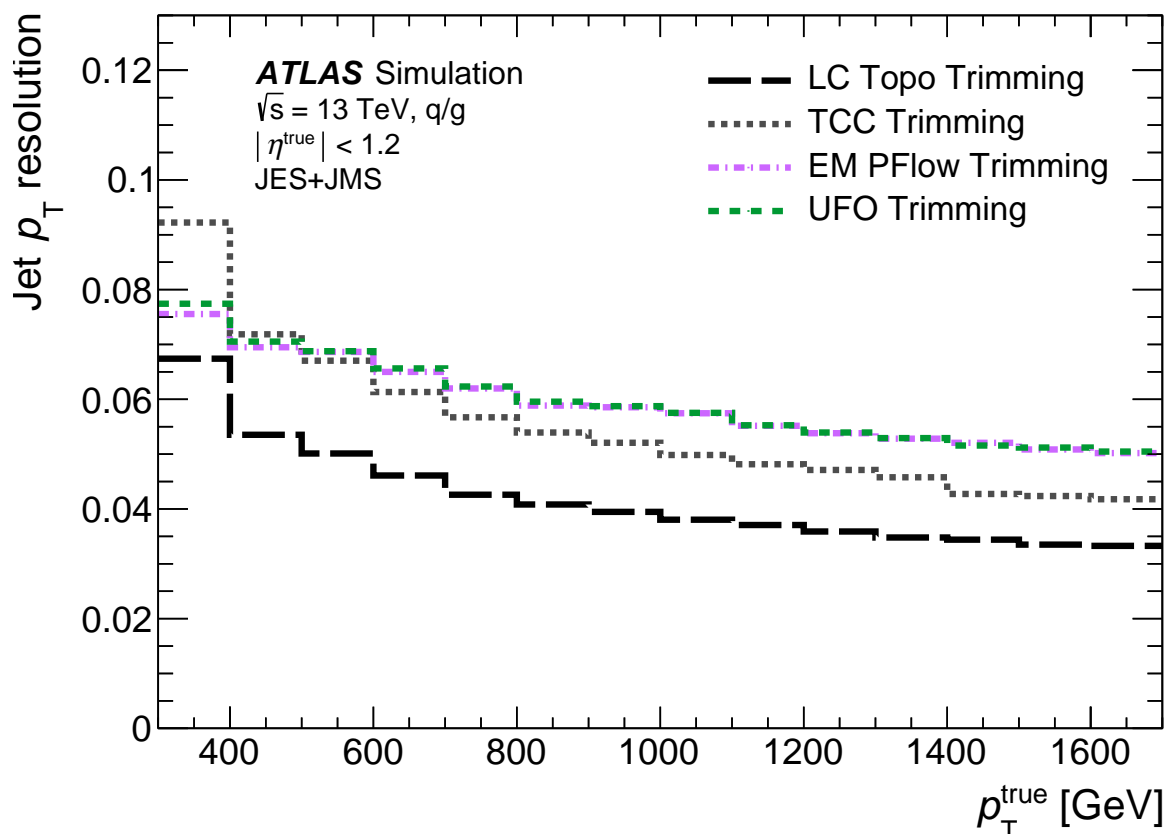


FIGURE 5.5 – Résolution en impulsion transverse en fonction de l'impulsion transverse *vraie* [66].

Ainsi, pour des jets de rayon  $R = 1,0$ , les UFOs ont de meilleures performances en terme d'identifications de jet issus de bosons  $W$  ou de quarks  $t$  et en terme de résolution en masse, que les jets utilisés jusqu'à présent par ATLAS. De plus, comme démontré dans [66], l'algorithme de Soft-drop présente aussi de meilleures performances que celui de *trimming* utilisé jusqu'à présent par ATLAS. Enfin, comme démontré dans [66], l'algorithme de  $CS+SK$  induit une meilleure stabilité quant à l'empilement. Ceci peut être vu dans la Figure 5.6 [66], qui représente la dépendance de la masse reconstruite des jets en fonction de  $N_{PV}$ , pour différentes définitions de jets :  $LC\ Topo$ ,  $PFlow$ ,  $TCC$  et  $UFO$ , avec et sans les algorithmes  $CS+SK$  et avec différents algorithmes de grooming. On peut remarquer que la dépendance est plus faible lorsque les algorithmes  $CS+SK$  sont appliqués, et avec l'algorithme Soft-drop pour un type de constituants donné.

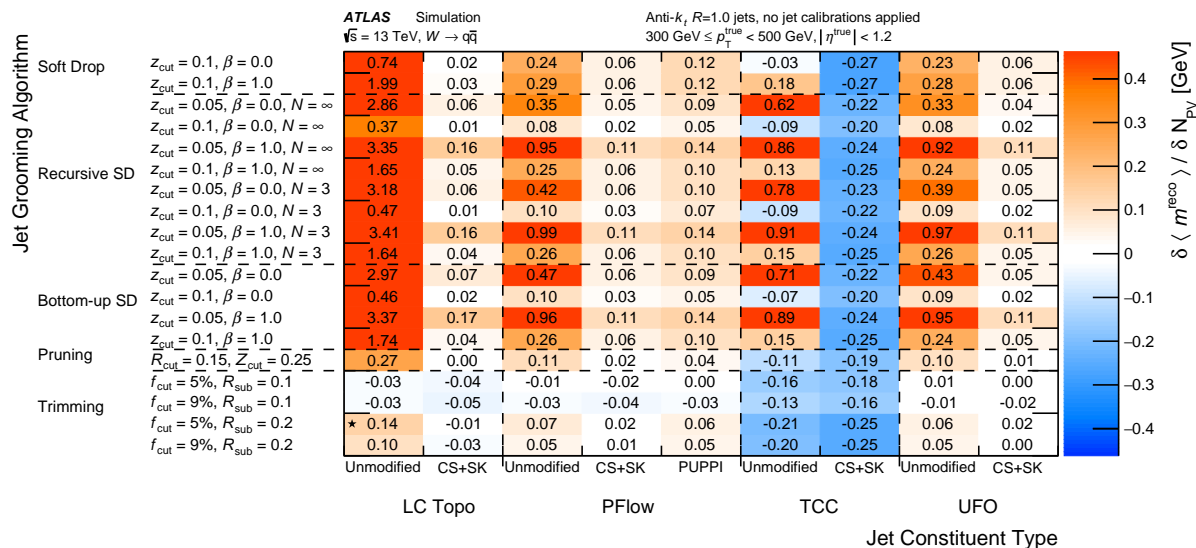


FIGURE 5.6 – Dépendance de la valeur de la masse reconstruite des jets en fonction de  $N_{\text{PV}}$  [66].

C'est pourquoi ATLAS recommande désormais d'utiliser des jets de rayon  $R = 1,0$  définis avec les *Unified Flow Objects*, avec les algorithmes de *CS+SK* et de *Soft-drop* appliqués aux constituants des jets, à la place des jets définis à partir des *LC Topos* avec *trimming*.

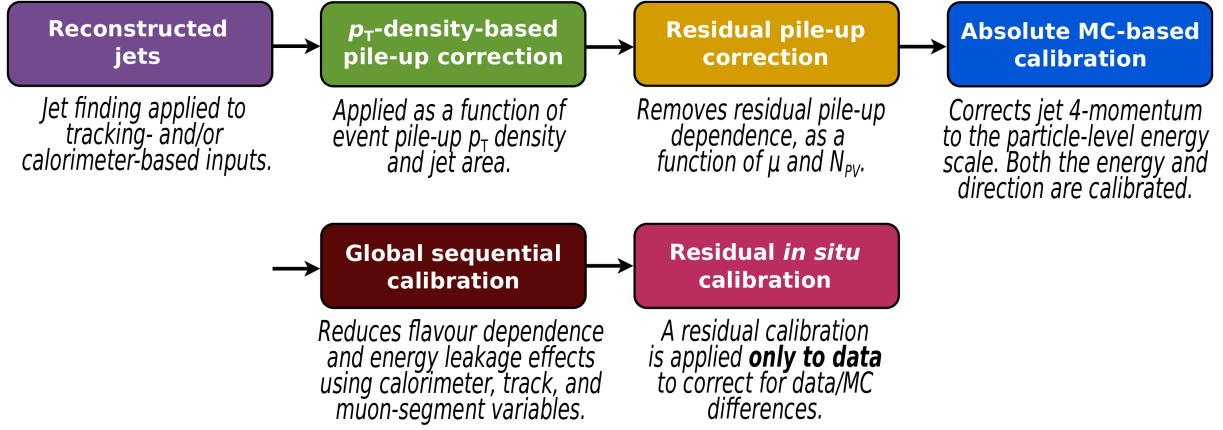
Les jets de rayon  $R = 0,4$  ont des métriques différentes que les jets de rayon  $R = 1,0$ . C'est pourquoi il est intéressant d'étudier l'usage des UFO pour des jets rayon  $R = 0,4$ , afin d'en améliorer potentiellement les performances, mais aussi d'harmoniser la définition des jets utilisée par ATLAS.

Cependant, cette comparaison doit être effectuée une fois que les jets sont étalonnés, dont la procédure est décrite dans la Section suivante.

## 5.2 La chaîne d'étalonnage des jets

À cause des effets du détecteur, des méthodes de reconstruction imparfaites et des fluctuations inhérentes à la mécanique quantique, les quantités reconstruites à partir du détecteur d'un jet sont généralement différentes des quantités *vraies*, c'est-à-dire reconstruites au niveau du générateur. C'est pourquoi l'énergie, la masse ou encore la direction d'un jet ont besoin d'être étalonnées, afin de correspondre aux quantités *vraies*. Pour ce faire, une chaîne d'étalonnage comportant plusieurs étapes a été développée par ATLAS. Cette chaîne d'étalonnage pour les jets de rayon  $R = 0,4$  est représentée Figure 5.7 [39] et sera détaillée dans cette Section.




 FIGURE 5.7 – Schéma de la chaîne d'étalonnage des jets de rayon  $R = 0,4$  [39].

Toutes les étapes, sauf la dernière, sont dérivées à partir des simulations seulement. Pour chacune d'elles, les jets reconstruits sont associés au jet *vrai* le plus proche géométriquement, via une procédure qui est décrite plus en détail dans la Section 5.3.2. Les jets *vrais* associés peuvent ainsi servir de référence pour l'étalonnage.

### 5.2.1 La correction d'empilement

La première étape sert à enlever l'excès ou le déficit d'énergie dû à l'empilement. Cette étape comporte elle-même deux corrections : une correction basée sur l'aire du jet et sur la densité en impulsion transverse due à l'empilement (appelée « correction d'aire »), et une correction résiduelle.

La correction d'aire peut être formulée selon :

$$p_T^{\text{area}} = p_T^{\text{reco}} - \rho_{\text{PUSB}} \times A \quad (5.1)$$

où :

- $p_T^{\text{area}}$  représente l'impulsion transverse du jet après correction d'aire,
- $p_T^{\text{reco}}$  représente l'impulsion transverse du jet reconstitué à partir du détecteur,
- $A$  représente l'aire du jet. Cette quantité est calculée en recouvrant uniformément chaque évènement par des particules fantômes d'impulsion transverse extrêmement faible. Ensuite, l'algorithme de regroupement avec l'algorithme Anti- $k_t$  des jets est appliqué à tous les constituants de l'évènement, c'est-à-dire en incluant les particules fantômes. L'aire du jet est donnée par le nombre relatif de particules fantômes associées au jet après ce regroupement. Cette quantité est environ égale à  $\pi \times R^2 = \pi \times 0,4^2$ ,
- et  $\rho_{\text{PUSB}}$  est une quantité calculée évènement par évènement qui représente la densité en impulsion transverse des jets dans le plan  $(\eta, \phi)$  et est donnée par :  $\rho_{\text{PUSB}} = \text{median}\left\{\frac{p_{T,i}^{\text{jet}}}{A_i^{\text{jet}}}\right\}$ , où l'indice  $i$  parcourt tous les jets de  $R = 0,4$  formés avec l'algorithme  $k_t$  à partir d'un même évènement. Cette procédure est similaire à celle décrite dans la Section 4.3.4.1, sauf qu'elle s'applique aux jets, et non pas à leurs constituants. Cependant, une modification a récemment été apportée à la façon

dont  $\rho$  est calculée pour les Particle Flows : afin d'être le moins biaisé possible par la topologie de la collision d'intérêt, seuls les Particle Flows *chargés* satisfaisant  $2 \text{ mm} < |z_0 \sin(\theta)| < 4 \text{ mm}$  sont utilisés (pour rappel l'association avec le vertex primaire est donnée par  $|z_0 \sin(\theta)| < 2 \text{ mm}$ ). C'est pourquoi cette variable est notée avec l'indice « *PUSB* », qui signifie « pile-up side-bands » en anglais. Seuls les Particle Flows *chargés* associés avec le vertex primaire sont utilisés en entrée à l'algorithme des Unified Flow Objects (voir Figure 5.1). Par conséquent, il n'existe pas de Unified Flow Objects *chargés* qui satisfont  $2 \text{ mm} < |z_0 \sin(\theta)| < 4 \text{ mm}$ , et donc si  $\rho_{\text{PUSB}}$  était calculé à partir des Unified Flow Objects, la composante chargée due à l'empilement manquerait. Cependant, comme décrit dans la Section 5.3, l'impulsion transverse des Unified Flow Objects et des Particle Flows est très similaire. C'est pourquoi le  $\rho_{\text{PUSB}}$  calculé à partir des Particle Flows est utilisé pour les Unified Flow Objects.

Enfin, l'algorithme de *SoftKiller*, décrit dans la Section 4.3.4.2, a pour objectif d'obtenir une valeur de  $\rho$  approximativement zéro. C'est pourquoi, pour les jets auxquels les algorithmes de *Constituent Subtraction* et *SoftKiller* sont appliqués, cette correction n'est pas utilisée, et la correction d'empilement comprend seulement la correction résiduelle décrite ci-dessous. La comparaison entre la valeur de  $\rho_{\text{PUSB}}$  avec et sans CS+SK est représentée dans la Section 5.3.

Après la correction d'aire, il a été observé qu'il persiste une dépendance de l'impulsion transverse des jets en fonction de  $N_{\text{PV}}$  et de  $\mu$ , et en particulier dans la région  $|\eta| > 2, 0$ , puisque  $\rho_{\text{PUSB}}$  est calculée dans la région  $|\eta| < 2, 0$ . C'est pourquoi une correction résiduelle est appliquée, en plus de la correction d'aire. Originellement, cette correction résiduelle pouvait être formulée selon :

$$p_{\text{T}}^{\text{PU}} = p_{\text{T}}^{\text{area}} - \alpha \times (N_{\text{PV}} - 1) - \beta \times \mu \quad (5.2)$$

où :

- $p_{\text{T}}^{\text{PU}}$  représente l'impulsion transverse finale du jet,
- $p_{\text{T}}^{\text{area}}$  représente l'impulsion transverse du jet après correction d'aire,
- et  $\alpha$  et  $\beta$  représentent la dépendance résiduelle de l'impulsion transverse après correction d'aire en fonction de  $N_{\text{PV}}$  et de  $\mu$ .

Cependant, cette correction ne prenait pas en compte la corrélation entre  $N_{\text{PV}}$  et  $\mu$ , et était dérivée avec des jets ayant une impulsion transverse *vraie*,  $p_{\text{T}}^{\text{true}}$ , égale à 25 GeV, et par conséquent ne considérait pas une éventuelle dépendance en impulsion transverse. C'est pourquoi une amélioration de cette correction a récemment été développée par ATLAS dans le cadre des Particle Flows. La correction résiduelle peut désormais être formulée selon :

$$p_{\text{T}}^{\text{PU}} = p_{\text{T}}^{\text{area}} + \Delta p_{\text{T}}^{\text{area-true}}(N_{\text{PV}}, \mu, \eta, p_{\text{T}}^{\text{area}}) \quad (5.3)$$

où le terme  $\Delta p_{\text{T}}^{\text{area-true}}(N_{\text{PV}}, \mu, \eta, p_{\text{T}}^{\text{area}})$  est calculé séparément en intervalles de  $(N_{\text{PV}}, \mu, \eta)$ . Dans chaque intervalle en  $(N_{\text{PV}}, \mu, \eta)$ , la valeur moyenne de la différence entre  $p_{\text{T}}^{\text{true}}$

et  $p_T^{\text{area}}$  est tracée en fonction de  $p_T^{\text{area}}$ . Ensuite ces courbes sont ajustées en utilisant la fonction :

$$\langle p_T^{\text{true}} - p_T^{\text{area}} \rangle = a + b \times p_T^{\text{area}} + c \times \log(p_T^{\text{area}}) \quad (5.4)$$

Ce type de fonction est utilisé de façon empirique, car il permet de bien ajuster les courbes en fonction de  $p_T^{\text{area}}$ .

Enfin, ces fonctions, une pour chaque intervalle en  $(N_{\text{PV}}, \mu, \eta)$ , sont utilisées pour corriger l'impulsion transverse des jets. Cette nouvelle correction résiduelle permet de réduire la dépendance en empilement de l'impulsion transverse des jets par rapport à la correction résiduelle utilisée jusqu'à présent. Elle est dérivée entre  $10 < p_T^{\text{true}} < 120$  GeV. Auparavant, seuls les jets à partir de  $p_T^{\text{true}} = 20$  GeV étaient utilisés pour dériver l'étalonnage. Cependant, au moment de commencer à travailler sur ce sujet, ATLAS avait pour objectif d'étalonner des jets de plus basse impulsion transverse, jusqu'à  $p_T^{\text{true}} = 10$  GeV. La correction n'est pas dérivée à partir de jets de plus haute impulsion transverse que 120 GeV, car la contribution provenant de l'empilement est supposée être négligeable pour ceux-ci.

La correction de l'empilement sera notée « *PU* » par la suite.

### 5.2.2 La correction en énergie, masse et pseudo-rapacité

La deuxième étape de la chaîne d'étalonnage (appelée « Absolute MC-based calibration » sur la Figure 5.7) correspond à la correction absolue en énergie, en masse et en pseudo-rapacité. Elle a pour objectif de corriger le quadri-vecteur du jet reconstruit afin de correspondre au quadri-vecteur *vrai*. Pour ce faire, cette étape prend en compte le caractère non compensatoire des calorimètres, les pertes d'énergie dans les matériaux passifs, ou encore les biais dans la reconstruction en pseudo-rapacité du jet. Ces biais sont principalement causés par la transition entre différentes technologies de calorimètres et par des changements soudains dans la granularité des calorimètres. Cette correction est dérivée après que la correction d'empilement a été appliquée. Pour ce faire l'énergie et la masse d'un jet sont corrigées selon les formules suivantes :

$$E_{\text{PU}} = E_{\text{reco}} \times \frac{p_T^{\text{PU}}}{p_T^{\text{reco}}} \quad (5.5)$$

$$m_{\text{PU}} = m_{\text{reco}} \times \frac{p_T^{\text{PU}}}{p_T^{\text{reco}}} \quad (5.6)$$

où  $E_{\text{reco}}$  et  $m_{\text{reco}}$  représentent l'énergie et la masse du jet reconstitué à partir du détecteur (avant tout étalonnage) respectivement, et  $E_{\text{PU}}$  et  $m_{\text{PU}}$  représentent l'énergie et la masse du jet après correction d'empilement respectivement.

Afin de dériver cette correction, une quantité est introduite : la Réponse,  $\mathcal{R}$ , qui est définie comme la valeur moyenne d'un ajustement avec une Gaussienne des distributions  $\frac{X_{\text{PU}}}{X_{\text{true}}}$ , où  $X$  peut représenter l'énergie ou la masse d'un jet. La réponse est aussi appelée « *Échelle* », ou « *Scale* » en anglais. La réponse est schématisée Figure 5.8.

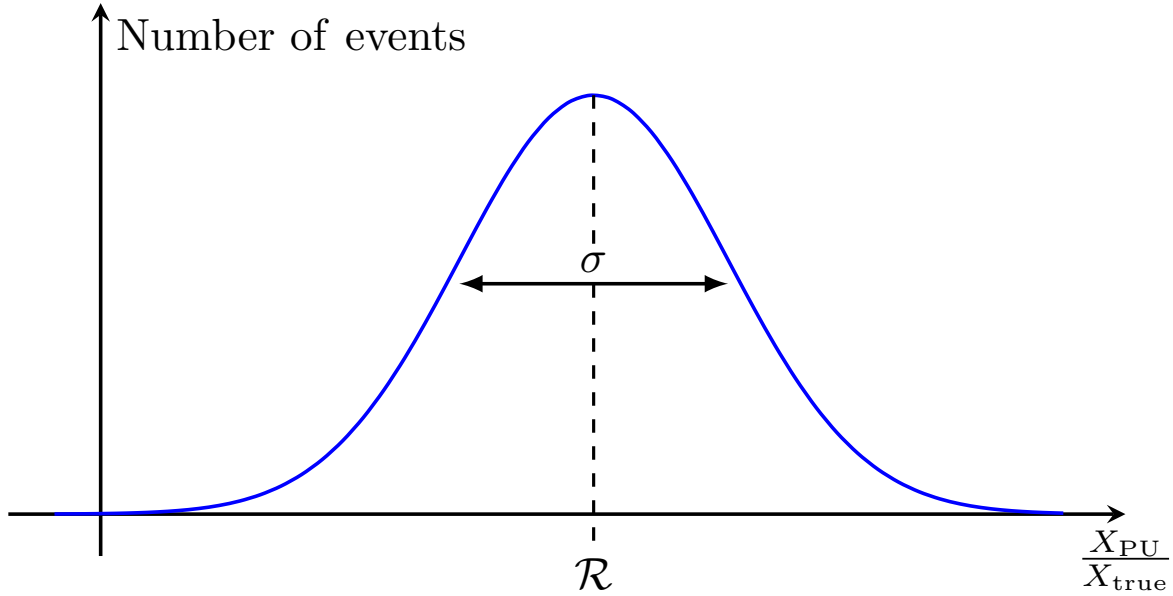


FIGURE 5.8 – Schéma de la réponse  $\mathcal{R}$  et de la résolution  $\sigma$  d'une quantité  $X$  d'un jet (par exemple l'énergie ou la masse).

La correction en énergie est dérivée en fonction de  $E_{\text{true}}$  et de  $\eta$ , c'est-à-dire que la réponse en énergie est calculée en intervalles de  $(E_{\text{true}}, \eta)$  séparément.

Un exemple de courbes de réponse en énergie en fonction de  $\eta$  et de  $E_{\text{reco}}$  est représenté Figure 5.9 [39].

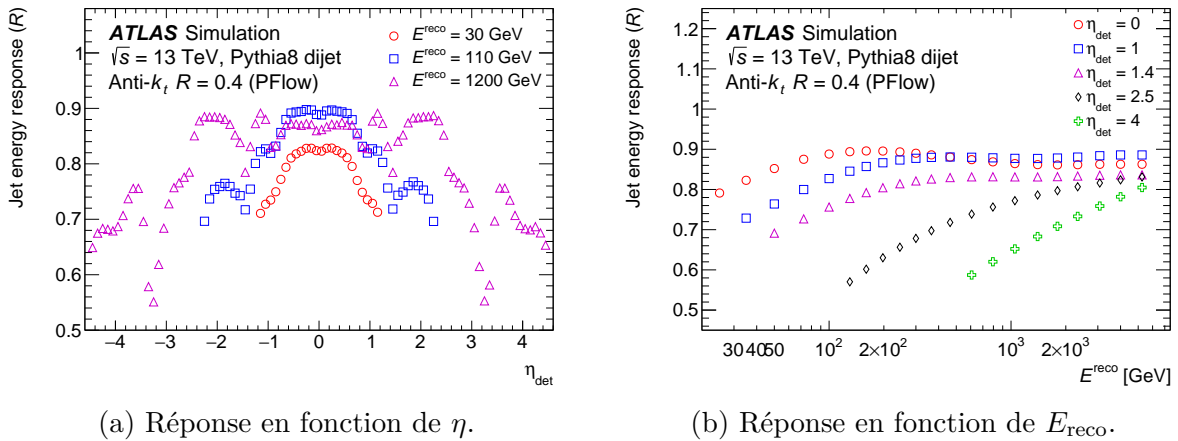


FIGURE 5.9 – Réponse en fonction de  $\eta$  et de  $E_{\text{reco}}$ , avant étalonnage [39].

La réponse est tout d'abord dérivée en fonction de  $E_{\text{true}}$ , et non en fonction de  $E_{\text{PU}}$ , car à un intervalle fixé en  $E_{\text{true}}$ , la réponse est distribuée de façon Gaussienne, ce qui n'est pas le cas pour  $E_{\text{PU}}$ . Cependant, la réponse doit être exprimée en fonction de  $E_{\text{PU}}$ , et non de  $E_{\text{true}}$ , afin de pouvoir être appliquée aux données réelles. C'est pourquoi une procédure d'inversion numérique est utilisée [67]. Tout d'abord, pour chaque intervalle en  $\eta$  séparément, la réponse en énergie est tracée en fonction de  $E_{\text{true}}$ , et est ajustée à l'aide

d'une fonction lisse, ce qui permet d'enlever les fluctuations statistiques. Jusqu'à présent la fonction utilisée était un polynôme logarithmique en énergie :

$$\mathcal{R}_1(E_{\text{true}}) = \sum_{i=0}^{N_{\text{max}}} a_i (\ln(E_{\text{true}}))^i \quad (5.7)$$

où le degré maximum du polynôme,  $N_{\text{max}}$ , était choisi selon l'ajustement ayant le plus petit  $\chi^2$  divisé par le nombre de degrés de liberté. Cependant, plus récemment une autre fonction a été utilisée par ATLAS : les *penalized splines* (ou plus brièvement *p-splines*) [68]. Les p-splines sont des fonctions non-paramétriques définies par morceaux par des polynômes et ajustées de façon à minimiser la quantité :

$$\sum_{i=0}^n (y_i - f(x_i))^2 + \lambda \int_a^b (f''(x))^2 dx \quad (5.8)$$

où  $\lambda$  représente un paramètre libre qui contrôle le degré de lissage. Le premier terme correspond aux moindres carrés usuels, et le second terme correspond à une pénalité sur la courbure de  $f$ . Si  $\lambda = 0$ , il n'y a pas de contrainte sur  $f$ , et cette équation revient à une interpolation normale. Si  $\lambda = \infty$ , cette équation revient à une régression linéaire.

Originellement, la valeur de  $\lambda$  était définie dans chaque intervalle en  $\eta$  par :

$$\lambda = \frac{\alpha}{n} \times \sum_{i=0}^n w_i \quad (5.9)$$

où  $i$  parcourt les  $n$  points à ajuster dans l'intervalle en  $\eta$  considéré, et les  $w_i$  sont définis par l'inverse de la racine carré de l'incertitude sur la valeur de la réponse déterminée avec un ajustement avec une gaussienne :  $w = \frac{1}{\sqrt{\sigma_{\mathcal{R}}}}$ . Le degré des p-splines et la valeur de  $\alpha$  étaient égaux à 3 et à 0,1 respectivement.

Cependant, lors de mes études, j'ai remarqué que cette procédure menait parfois à un *sur-ajustement* de la courbe de la réponse, en particulier dans la région avant du détecteur. Ceci est dû au fait que le paramètre de lissage, le même pour tous les intervalles en  $\eta$ , n'était pas forcément optimal pour tous ces intervalles. C'est pourquoi, j'ai légèrement modifié le code, afin que le paramètre de lissage puisse changer d'un intervalle en  $\eta$  à l'autre. Pour ce faire, j'ai utilisé la procédure de *validation croisée* afin de déterminer la valeur optimale de  $\lambda$  dans chaque intervalle en  $\eta$  indépendamment. Cette procédure consiste, pour chaque valeur de  $\lambda$ , et pour chaque point de la courbe, à enlever ce point de la courbe, et à ajuster la courbe avec ce point enlevé. Ensuite, la différence entre la valeur de la réponse en ce point et la valeur de l'ajustement (dérivé sans ce point) évalué en ce point est calculée. Cette valeur est ensuite élevée au carré, puis, la somme est effectuée sur tous les points. Cette procédure est répétée pour chaque valeur de  $\lambda$ . Plus formellement, cette procédure peut s'écrire selon l'équation suivante :

$$\text{CV}(\lambda) = \sum_{i=1}^{n-1} (y_i - f_{\lambda}^{(-i)}(x_i))^2 \quad (5.10)$$

où  $f_{\lambda}^{(-i)}(x_i)$ <sup>1</sup> représente la valeur de l'ajustement obtenu sans le point  $(x_i, y_i)$  et éva-

---

1. L'indice  $(-i)$  ne représente pas une puissance ici.

lué à  $x_i$ . La valeur de  $\lambda$  choisie est celle qui minimise  $CV(\lambda)$ . Pour minimiser  $CV(\lambda)$  en fonction de  $\lambda$  j'ai utilisé *SciPy* [69]. Cette procédure permet d'obtenir un bon ajustement, mais tout en ayant une fonction lisse. Le seul désavantage de cette méthode est que le premier et le dernier point ne peuvent pas être utilisés pour trouver la valeur de  $\lambda$ , car les p-splines ne sont pas définies en dehors de l'intervalle ajusté. Pour l'ajustement final (avec tous les points), une extrapolation linéaire à partir du premier ou du dernier point est utilisée.

Ensuite, pour chaque jet, l'énergie *vraie* est multipliée par la réponse  $\mathcal{R}_1(E_{\text{true}})$  évaluée à l'énergie vraie de ce jet. Le produit  $E_{\text{true}} \times \mathcal{R}_1(E_{\text{true}})$  est appelé l'énergie *inversée numériquement* (ou  $E_{\text{NI}}$ , pour *Numerically Inversed* en anglais).  $E_{\text{NI}}$  est une procuration pour l'énergie moyenne reconstruite (après correction d'empilement) correspondant à  $E_{\text{true}}$  (car  $E_{\text{true}} \times \mathcal{R}_1(E_{\text{true}}) \simeq E_{\text{true}} \times \frac{E_{\text{PU}}}{E_{\text{true}}} = E_{\text{PU}}$ ). La réponse ( $\frac{E_{\text{PU}}}{E_{\text{true}}}$ ) est tracée en fonction de  $E_{\text{NI}}$ , puis ajustée avec les mêmes fonctions que lors de la première étape. Ces fonctions, une pour chaque intervalle de  $\eta$ , définissent  $\mathcal{R}_2(E_{\text{NI}})$ .

Enfin, l'énergie est corrigée selon la formule suivante :

$$E_{\text{JES}} = \frac{E_{\text{PU}}}{\mathcal{R}_2(E_{\text{PU}})} \quad (5.11)$$

où  $E_{\text{JES}}$  représente l'énergie après correction, et  $\mathcal{R}_2(E_{\text{PU}})$  représente la courbe définie en fonction de  $E_{\text{NI}}$ , mais évaluée à  $E_{\text{PU}}$ . Ce même facteur de correction multiplicatif, donné par  $\frac{1}{\mathcal{R}_2(E_{\text{PU}})}$ , est aussi appliqué au reste du quadri-vecteur du jet (c'est-à-dire aussi à la masse :  $m_{\text{JES}} = \frac{m_{\text{PU}}}{\mathcal{R}_2(E_{\text{PU}})}$ , mais pas à  $\eta$  et à  $\phi$ ).

Si cette procédure fonctionne, la réponse après correction devrait être proche de 1, car  $\frac{E_{\text{JES}}}{E_{\text{true}}} \simeq \frac{E_{\text{PU}} \times \frac{E_{\text{true}}}{E_{\text{PU}}}}{E_{\text{true}}} = 1$ . Ce test est appelé test « de fermeture » de la méthode.

La correction en pseudo-rapacité est évaluée de la même façon que la correction en énergie, si ce n'est que ce n'est pas la réponse en pseudo-rapacité qui est utilisée, mais la différence entre la pseudo-rapacité reconstruite et la pseudo-rapacité *vraie* :  $\eta_{\text{reco}} - \eta_{\text{true}}$ . La correction en pseudo-rapacité prend donc la forme :  $\eta_{\text{JES}} = \eta_{\text{reco}} + \Delta\eta$ .

La correction en énergie et en pseudo-rapacité sera notée « *JES* » par la suite. Cette correction est dérivée à partir de jets avec  $10 < p_{\text{T}}^{\text{true}} < 4000$  GeV.

La masse est étalonnée de la même façon, si ce n'est que la réponse est évaluée en intervalles de  $(E, \log(\frac{m}{E}), \eta)$ , à la place de simplement  $(E, \eta)$  pour l'énergie. Le facteur de correction n'est appliqué qu'à la masse, et non au quadri-vecteur entier, en gardant l'énergie du jet fixée, et donc en permettant à l'impulsion transverse de varier. La correction en masse sera notée « *JMS* » par la suite, et le facteur de correction  $c_{\text{JMS}}$ . Cette correction est dérivée à partir de jets avec  $40 < m_{\text{true}} < 200$  GeV et  $200 < p_{\text{T}}^{\text{true}} < 4000$  GeV.

En résumé, la correction totale est donnée par :

$$E_{\text{JES}} = \frac{E_{\text{PU}}}{\mathcal{R}_2(E_{\text{PU}})} \quad (5.12)$$

$$\eta_{\text{JES}} = \eta_{\text{reco}} + \Delta\eta \quad (5.13)$$

$$m_{\text{JMS}} = \frac{c_{\text{JMS}} \times m_{\text{PU}}}{\mathcal{R}_2(E_{\text{PU}})} \quad (5.14)$$

$$p_{\text{T}}^{\text{JES+JMS}} = \frac{1}{\mathcal{R}_2(E_{\text{PU}})} \times \frac{\sqrt{E_{\text{PU}}^2 - c_{\text{JMS}}^2 \times m_{\text{PU}}^2}}{\cosh(\eta_{\text{reco}} + \Delta\eta)} \quad (5.15)$$

### 5.2.3 Global Sequential Calibration

La troisième étape (appelée « Global sequential calibration » (ou plus brièvement *GSC*) sur la Figure 5.7) a pour objectif d'améliorer la résolution en impulsion transverse des jets, définie par la largeur de l'ajustement avec une Gaussienne de la distribution en  $\frac{p_{\text{T}}^{\text{JES+JMS}}}{p_{\text{T}}^{\text{true}}}$ , comme représenté par la flèche horizontale sur la Figure 5.8. En pratique, la résolution est évaluée par l'écart interquantile à 16% et à 84% (noté « IQR », pour Inter-Quantile Range en anglais) des histogrammes dans ce document, puis divisée par deux, afin de correspondre à la largeur d'un ajustement avec une gaussienne.

En effet, après la correction absolue en énergie, en masse et en pseudo-rapacité, la réponse individuelle peut varier d'un jet à l'autre selon la saveur et la distribution en énergie des particules qui constituent le jet, de leur distribution transverse ou encore des fluctuations dans le développement du jet dans les calorimètres. De plus, le nombre moyen de particules dans le jet et la forme de la gerbe dépendent de la particule qui a initié le jet, et en particulier si c'est un quark ou un gluon : les jets initiés par des gluons tendent à avoir plus de constituants et à être plus larges que les jets initiés par des quarks. Ceci est dû au fait que les quarks ne transportent qu'une charge de couleur, tandis que les gluons en transportent deux. Plus précisément, les fonctions de fractionnement de Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (ou plus brièvement *DGLAP*) [70, 71, 72] contiennent un facteur  $C_A = 3$  pour les gluons émis à partir d'un autre gluon, et un facteur  $C_F = \frac{4}{3}$  pour les gluons émis à partir d'un quark. Ainsi, un jet initié par un quark contiendra souvent des hadrons avec une fraction plus élevée de l'impulsion transverse du jet qui pénètrent plus loin dans le calorimètre, tandis qu'un jet initié par un gluon contiendra typiquement plus de particules d'impulsion transverse plus faible, ce qui conduit à une réponse plus faible du calorimètre.

Réduire la différence en réponse entre les jets initiés par des quarks et les jets initiés par des gluons est important, car cette différence est une source d'incertitude sur l'échelle en énergie des jets. En effet, pour un même échantillon de données, la fraction de jets initiés par des quarks et de jets initiés par des gluons peut varier, et donc la réponse totale peut varier. Cette incertitude est appelée « incertitude sur la composition en saveurs » et est donnée par :

$$\sigma_{\text{composition}} = \sigma_{f_g} \times \frac{|\mathcal{R}_g - \mathcal{R}_q|}{f_g \times \mathcal{R}_g + (1 - f_g) \times \mathcal{R}_q} \quad (5.16)$$

où :

- $f_g$  représente la fraction de jets initiés par des gluons dans l'analyse. Cette fraction dépend de l'analyse considérée, par exemple, des processus de QCD pure contiennent plus de jets initiés par des gluons que des processus avec des bosons  $W$  qui se désintègrent en quarks,
- $\sigma_{f_g}$  représente une incertitude sur cette fraction de jets initiés par des gluons, car cette valeur n'est pas connue exactement, même pour un type de processus donné. Cette incertitude est généralement obtenue en prenant la différence entre la fraction de jets initiés par des gluons pour des jets générés avec Pythia et un autre générateur (généralement Herwig),
- $\mathcal{R}_g$  représente la réponse en  $p_T$  pour des jets initiés par des gluons avec le générateur Pythia,
- $\mathcal{R}_q$  représente la réponse en  $p_T$  pour des jets initiés par des quarks avec le générateur Pythia.

Le GSC correspond à une série de corrections multiplicatives qui servent à réduire les effets de ces fluctuations et à améliorer la résolution du jet sans modifier la réponse moyenne du jet. Pour ce faire, le GSC est basé sur des observables du jet telles que la structure longitudinale des dépôts d'énergie dans les calorimètres, des informations sur les traces associées au jet, et des informations liées à l'activité dans les chambres à muons derrière un jet. Six observables sont utilisées pour les Particle Flows :

- $f_{\text{charged}}$  : la fraction en impulsion transverse du jet mesurée à partir des traces avec  $p_T > 500$  MeV qui lui sont associées. Cette variable permet de caractériser la structure longitudinale de la gerbe d'un jet. Cette correction n'est appliquée qu'aux jets qui sont dans la région couverte par le détecteur interne, c'est-à-dire  $|\eta| < 2,5$ ,
- $f_{\text{Tile0}}$  : la fraction en énergie du jet mesurée dans la première couche du calorimètre hadronique en tuiles scintillatrices (TileCal). Cette variable permet de caractériser la structure longitudinale de la gerbe d'un jet. Cette correction n'est appliquée qu'aux jets qui sont dans la région couverte par ce calorimètre, c'est-à-dire  $|\eta| < 1,7$ ,
- $f_{\text{LAR3}}$  : la fraction en énergie du jet mesurée dans la troisième couche du calorimètre électromagnétique en argon liquide (EMCal). Cette correction n'est appliquée qu'aux jets qui sont dans la région couverte par ce calorimètre, c'est-à-dire  $|\eta| < 3,5$ ,
- $n_{\text{trk}}$  : le nombre de traces avec  $p_T > 1$  GeV qui sont associées au jet. Cette variable permet de discriminer les quarks des gluons, car elle est liée au nombre de constituants contenus dans le jet. Cette correction n'est appliquée qu'aux jets qui sont dans la région couverte par le détecteur interne, c'est-à-dire  $|\eta| < 2,5$ ,
- $w_{\text{trk}}$  : la distance moyenne pondérée en impulsion transverse dans le plan  $(\eta, \phi)$  entre l'axe du jet et toutes les traces avec  $p_T > 1$  GeV qui lui sont associées :



$w_{\text{trk}} = \frac{\sum_{\text{tracks}} p_{\text{T}}^{\text{track}} \times \Delta R(\text{track, jet})}{\sum_{\text{tracks}} p_{\text{T}}^{\text{track}}}$ . Cette variable permet aussi de discriminer les quarks

des gluons, car elle est reliée à la largeur du jet. Cette correction n'est appliquée qu'aux jets qui sont dans la région couverte par le détecteur interne, c'est-à-dire  $|\eta| < 2,5$ ,

- $n_{\text{segments}}$  : le nombre de segments de traces dans le spectromètre à muons qui sont associés au jet. Cette variable permet de corriger les jets de haute impulsion transverse qui ne sont pas entièrement contenus dans les calorimètres. Cette correction n'est appliquée qu'aux jets qui sont dans la région couverte par le spectromètre à muons, c'est-à-dire  $|\eta| < 2,7$ .

Pour les UFOs, ces observables sont les mêmes, si ce n'est que la façon dont est calculée  $f_{\text{charged}}$  diffère légèrement : seuls les UFOs *chargés* sont considérés dans le calcul de cette quantité, et non pas les UFOs *chargés et combinés*, c'est-à-dire aussi en considérant les Particle Flows *neutres* qui sont associés à une trace via l'algorithme des TCCs. Dû à l'algorithme des TCCs, un topo-cluster peut être associé à plusieurs traces, ce qui résulte en une  $f_{\text{charged}}$  beaucoup plus élevée que pour les Particle Flows. Cependant, tous les Particle Flows chargés ne correspondent pas à un UFO chargé, étant donné que les Particle Flows chargés qui sont dans un environnement dense sont donnés en entrée à l'algorithme des TCCs, comme décrit sur la Figure 5.1. C'est pourquoi  $f_{\text{charged}}$  est légèrement plus faible pour les UFOs que pour les Particle Flows.

Toutes ces corrections sont dérivées en fonction de l'impulsion transverse, sauf  $n_{\text{segments}}$ , qui est dérivée en fonction de l'énergie puisque cet effet est plus corrélé avec l'énergie qui s'échappe des calorimètres.

Pour chaque observable, la procédure d'inversion numérique est appliquée en intervalles de  $p_{\text{T}}$  (ou  $E$  pour  $n_{\text{segments}}$ ),  $\eta$ , et de l'observable considérée :  $f_{\text{charged}}$ ,  $f_{\text{Tile0}}$  etc. Ces corrections prennent la forme de facteurs multiplicatifs, et sont appliquées de façon indépendante et séquentielle au quadri-vecteur du jet, car les corrélations entre les observables sont supposées être négligeables.

Ainsi, après chaque correction, la dépendance de la réponse en fonction de la variable considérée est réduite, de façon à ce que la réponse soit approximativement 1 pour tous les jets. C'est pourquoi le GSC permet d'améliorer la résolution *totale*. Cette procédure est schématisée sur la Figure 5.10.

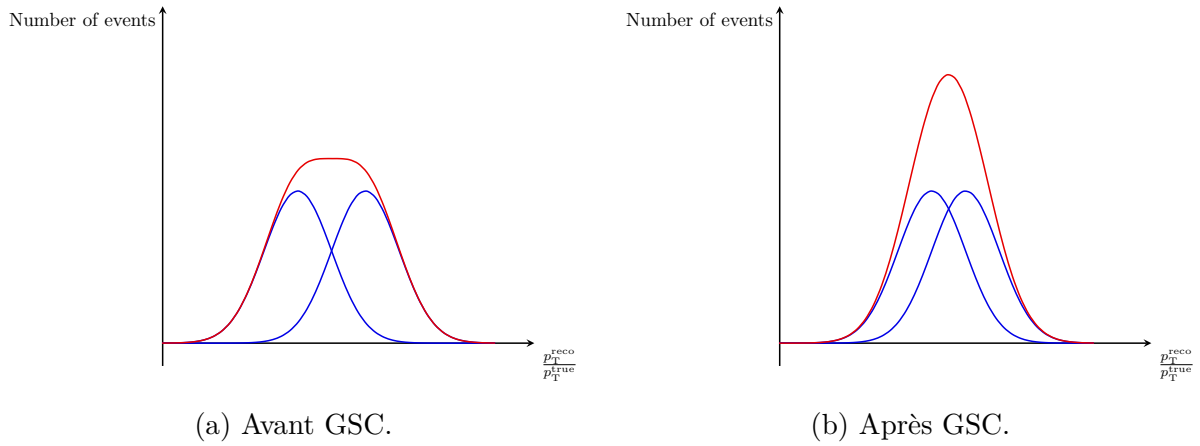


FIGURE 5.10 – Résolution en impulsion transverse avant GSC et après GSC.

La résolution en impulsion transverse après chaque étape du GSC est représentée Figure 5.11 [39]. Sur cette Figure, les variables dans la légende sont celles décrites précédemment :  $f_{\text{charged}}$ ,  $f_{\text{Tile0}}$ ,  $f_{\text{LAR3}}$ , etc . Après chaque étape, la résolution en impulsion transverse transverse s'améliore.

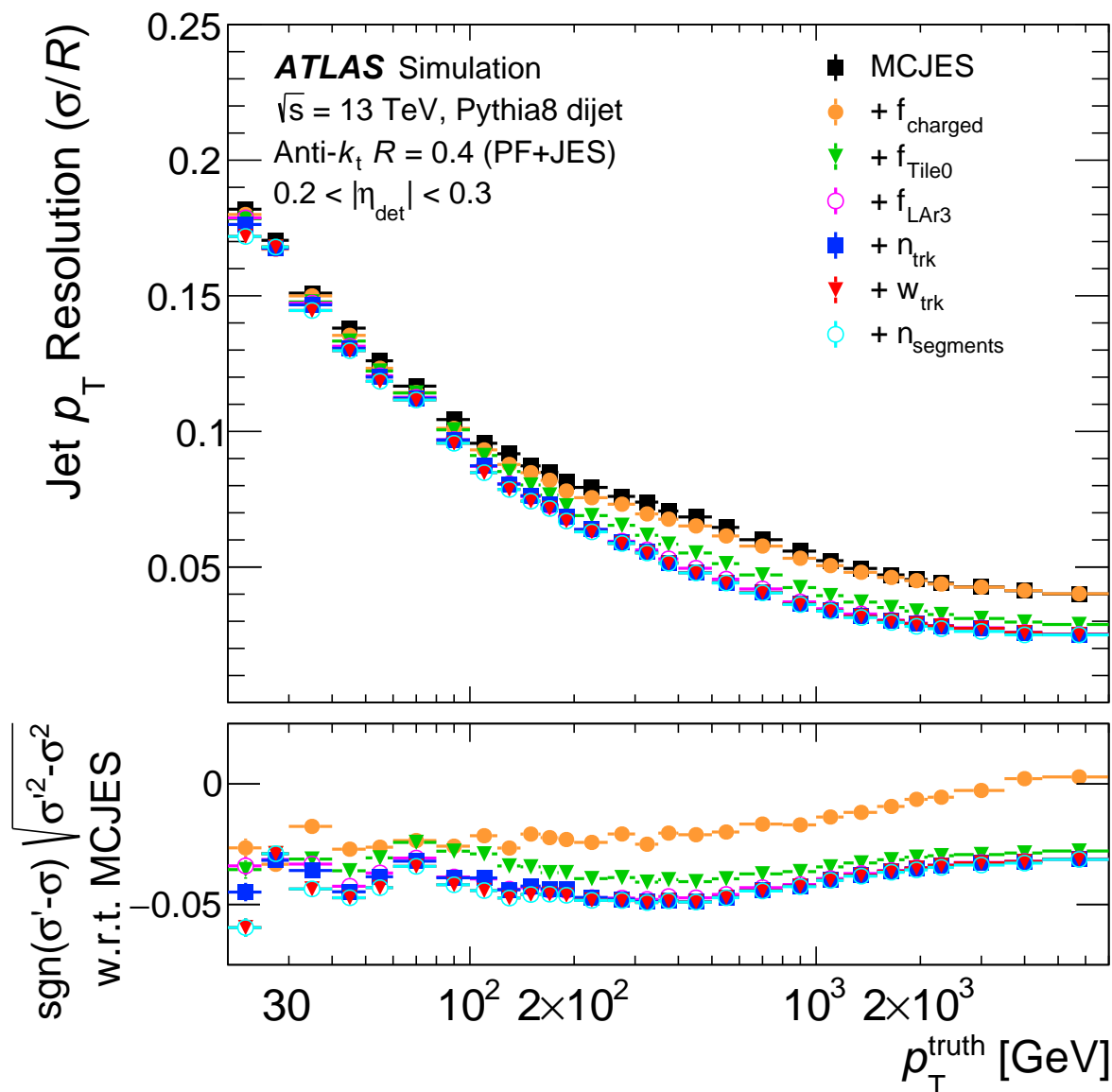


FIGURE 5.11 – Résolution en impulsion transverse après chaque étape du GSC [39]. Le panneau du bas représente la différence en quadrature entre la résolution avant GSC ( $\sigma$ ) et après l'application de l'étape correspondante du GSC ( $\sigma'$ ).

### 5.2.4 La correction « *in-situ* »

La dernière étape (appelée « Residual *in-situ* calibration » sur la Figure 5.7) a pour objectif de corriger les différences entre les données réelles et les simulations, toutes les corrections précédentes étant dérivées à partir des simulations seulement. Cette correction est dérivée en utilisant des objets de référence bien mesurés, tels que des bosons  $Z$ , des photons, ou plusieurs jets de plus basse impulsion transverse [39]. Cette étape n'est pas décrite plus en détail ici, étant donné que je n'y ai pas contribué.

## 5.3 Performances et étalonnage des UFOs de rayon $R=0,4$

Mon travail a consisté à dériver la chaîne d'étalonnage pour des jets de rayon  $R = 0,4$  reconstruits à partir des Unified Flow Objects, et à comparer leur performances avec la référence actuelle d'ATLAS, c'est-à-dire les Particle Flows avec l'algorithme de Charged Hadron Subtraction appliqué aux constituants chargés, et en utilisant des topo-clusters à l'échelle d'énergie électromagnétique.

### 5.3.1 Échantillons utilisés

Les échantillons utilisés dans ces études sont des échantillons purement de QCD avec deux partons générés dans l'état final. Les principaux échantillons utilisés, avec lesquels l'étalonnage est dérivé, ont été générés avec le générateur *PYTHIA 8.235* [41], la collection de *Parton distribution functions* (ou plus brièvement *PDF*) *NNPDF2.3LO* [73], la collection de paramètres *A14* [74] et une énergie dans le centre de masse  $\sqrt{s} = 13$  TeV. *PYTHIA* est un générateur dont les éléments de matrice sont au premier ordre en théorie des perturbations, dont le modèle d'hadronisation est le modèle *Lund string* [75] et dont le modèle de gerbe de partons est le modèle *dipole* [76] ordonné en  $p_T$ . Afin de générer des événements dans toute la gamme en impulsion transverse, les événements sont générés en 13 *tranches* en impulsion transverse vraie de jets reconstruits avec l'algorithme anti- $k_t$  [55] avec comme paramètre de rayon  $R = 0,6$ . Ces tranches sont appelées « *JZ* », avec un suffixe qui représente la gamme d'impulsion transverse correspondant à la tranche. Ce nombre varie entre 0 et 12. La gamme en impulsion transverse vraie, la section efficace, ainsi que l'efficacité du filtre du générateur (voir la Section 3.5) sont représentées dans la Table 5.1.

JZ	Gamme en $p_T^{\text{true}}$ [GeV]	$\sigma$ [fb]	Efficacité du filtre du générateur
0	0 – 20	$7,80 \times 10^{13}$	0,97533
1	20 – 60	$7,80 \times 10^{13}$	0,024425
2	60 – 160	$2,43 \times 10^{12}$	0,0098631
3	160 – 400	$2,64 \times 10^{10}$	0,011658
4	400 – 800	$2,55 \times 10^8$	0,013366
5	800 – 1300	$4,55 \times 10^6$	0,014526
6	1300 – 1800	$2,58 \times 10^5$	0,0094734
7	1800 – 2500	$1,62 \times 10^4$	0,011097
8	2500 – 3200	$6,25 \times 10^2$	0,010156
9	3200 – 3900	$1,96 \times 10^1$	0,012056
10	3900 – 4600	$1,20 \times 10^0$	0,0058933
11	4600 – 5300	$4,23 \times 10^{-2}$	0,002673
12	5300 – $\infty$	$1,04 \times 10^{-3}$	0,00042889

TABLE 5.1 – Gamme en impulsion transverse vraie, section efficace, et efficacité du filtre du générateur pour les 13 tranches en  $p_T^{\text{true}}$ .

Comme décrit dans la Section 3.5, l'empilement est incorporé dans les simulations en superposant des interactions inélastiques simulées, en plus de l'interaction d'intérêt. Ces interactions inélastiques sont simulées avec *PYTHIA 8.210*, la collection de PDF NNPDF2.3LO et la collection de paramètres *A3* [77]. Les événements simulés avec des conditions d'empilement particulières sont ensuite comparés avec les données de la période de données correspondante. Trois collections de simulations sont utilisées : *MC16a*, correspondant aux données des années 2015 et 2016, *MC16d*, correspondant aux données de l'année 2017, et *MC16e*, correspondant aux données de l'année 2018. Le poids des événements simulés de MC16a sont multipliés par la luminosité intégrée des années 2015 et 2016, c'est-à-dire  $36, 2 \text{ fb}^{-1}$ , ceux de MC16d par celle 2017 ( $44, 3 \text{ fb}^{-1}$ ), et ceux de MC16e par celle de 2018 ( $58, 5 \text{ fb}^{-1}$ ).

Quatre définitions de jets sont comparées :

- la référence actuelle d'ATLAS : les *CHS PFlow*,
- les PFlow auxquels les algorithmes de suppression d'empilement *CS+SK* (voir la Section 4.3.4) sont appliqués aux constituants neutres : les *CS+SK PFlow*,
- les Unified Flow Objects avec l'algorithme de *CHS* (voir la Section 4.3.1.2) appliqué aux constituants chargés : les *CHS UFO*,
- les Unified Flow Objects avec les algorithmes de *CS+SK* appliqués aux constituants neutres : les *CSSK UFO*

Pour les quatre collections de jets ci-dessus, les topo-clusters sont à l'échelle d'énergie électromagnétique.

Les jets sont reconstruits avec l'algorithme anti- $k_t$  avec comme paramètre de rayon  $R = 0, 4$ , à l'aide du logiciel *FASTJET* [56].

### 5.3.2 Sélections

L'évènement doit contenir au moins un vertex primaire *vrai* et un vertex primaire reconstruit à partir du détecteur, auquel au moins deux traces doivent être associées. Ces vertex doivent être proches :  $|z_{\text{true vertex}} - z_{\text{reco vertex}}| < 2 \text{ mm}$ . Les jets *vrais* avec une énergie négative ou avec  $|\eta| > 5$  ne sont pas utilisés. Il doit y avoir au moins un jet *vrai* et deux jets reconstruits à partir du détecteur associés à l'évènement. La moyenne en impulsion transverse des deux jets reconstruits à partir du détecteur avec la plus haute impulsion transverse (appelés le « *leading* » et le « *sub-leading* » jets respectivement en anglais),  $p_{\text{T}}^{\text{avg}} = \frac{p_{\text{T}}^{\text{leading reco}} + p_{\text{T}}^{\text{sub-leading reco}}}{2}$ , divisée par l'impulsion transverse du jet *vrai* de plus haute impulsion transverse :  $\frac{p_{\text{T}}^{\text{avg}}}{p_{\text{T}}^{\text{leading true}}}$  doit être inférieure à 1,4. Cette condition permet d'enlever les pics dans la distribution en  $p_{\text{T}}^{\text{reco}}$  dans les simulations qui proviennent d'évènements non physiques dont les jets sont générés dans des tranches de bas  $p_{\text{T}}^{\text{true}}$  et qui se retrouvent dans des tranches de plus haut  $p_{\text{T}}$  avec un poids très important comparé aux poids des évènements de cette tranche. La condition requiert que les deux jets reconstruits de plus haut  $p_{\text{T}}$  aient une impulsion transverse approximativement similaire (c'est-à-dire qu'il y ait un *équilibre* entre les jets), et que leur impulsion transverse moyenne soit proche (de moins de 40%) de l'impulsion transverse du jet *vrai* de plus haute impulsion transverse.

Si toutes ces conditions ne sont pas remplies pour un évènement, les jets associés à cet évènement ne sont pas utilisés.

De plus, les jets *vrais* et les jets reconstruits sont associés géométriquement si la distance angulaire qui les sépare,  $\Delta R$ , est inférieure à 0,3. Si aucun jet *vrai* n'est associé à un jet reconstruit, le jet reconstruit n'est pas utilisé. Si aucun jet reconstruit n'est associé à un jet *vrai*, le jet *vrai* n'est pas utilisé. Enfin, les jets *vrais* (reconstruits) ne doivent avoir aucun autre jet *vrai* (reconstruit) dans un rayon de  $\Delta R = 1,0$  ( $\Delta R = 0,6$ ). Cette dernière condition est appelée condition *d'isolation*, et permet d'enlever les jets qui sont proches d'un autre, ce qui viendrait contaminer la reconstruction en énergie et en direction du jet considéré. La condition d'isolation n'est utilisée que pour l'étalonnage des jets.

### 5.3.2.1 Distributions en impulsion transverse vraie

La Figure 5.12 représente la distribution en  $p_T^{\text{true}}$  de tous les jets *vrais* (et pas seulement le jet de plus haute impulsion transverse par exemple) de chacune des tranches, et ce pour les trois collections MC16a, d et e. Les jets *vrais* sont associés à au moins un jet reconstruit avec l'algorithme des CHS PFlow. Les distributions sont lisses, ce qui est l'objectif de la méthode d'assignement des poids. Les queues des distributions à basse  $p_T^{\text{true}}$  proviennent des jets autres que les deux jets de plus haute impulsion transverse. Bien que ces jets ne soient pas très bien modélisés par un générateur tel que PYTHIA (qui est au premier ordre), ces jets sont tout de même utilisés pour l'étalonnage, mais ils n'influencent pas les résultats de manière significative, car leur poids est très faible par rapport au poids des évènements de la tranche correspondant à leur impulsion transverse.

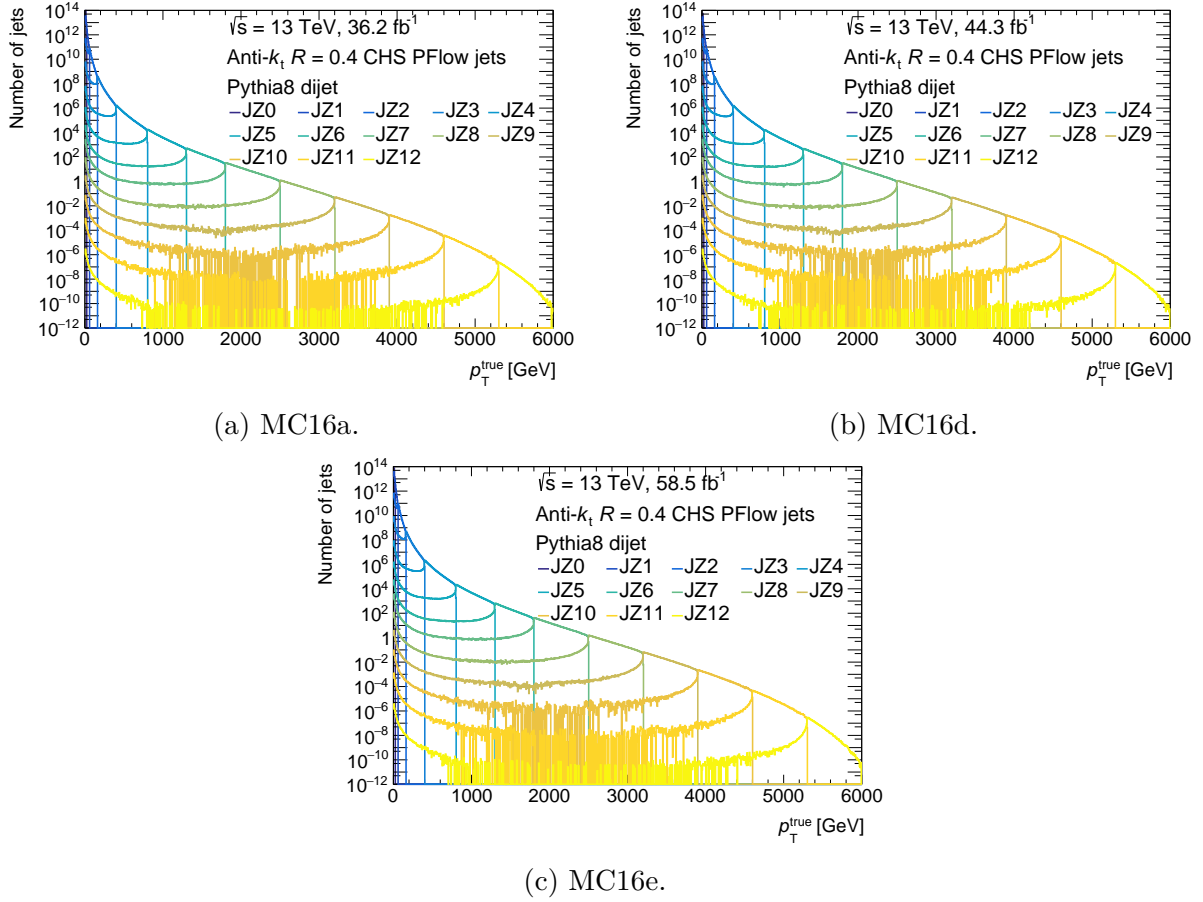


FIGURE 5.12 – Distributions en  $p_T^{\text{true}}$  de tous les jets de chacune des tranches, et ce pour les trois collections MC16a, d et e.

### 5.3.3 Résultats avant étalonnage

Cette Section décrit certaines propriétés des quatre collections de jets utilisées avant tout étalonnage (c'est-à-dire au niveau reconstruit), afin d'étudier les différences dues aux constituants directement, et non pas dues à des effets liés à l'étalonnage.

#### 5.3.3.1 Efficacité et pureté des jets

La Figure 5.13 représente une comparaison de l'efficacité de reconstruction en fonction de  $\eta$ ,  $\mu$  et  $N_{PV}$  pour les quatre différentes collections de jets étudiées. Les jets reconstruits sont associés à des jets *vrais* qui satisfont  $15 \text{ GeV} < p_T^{\text{true}} < 20 \text{ GeV}$  dans la première colonne et  $20 \text{ GeV} < p_T^{\text{true}} < 30 \text{ GeV}$  pour la seconde colonne. L'efficacité est définie par le nombre de jets reconstruits qui sont associés à un jet *vrai*, divisé par le nombre de jets *vrais* :  $\varepsilon = \frac{N_{\text{reco jets matched to truth jet}}}{N_{\text{truth jets}}}$ . Aucune condition sur l'isolation des jets n'est appliquée. Ainsi, plus l'efficacité est proche de 1, plus les performances sont bonnes. Afin d'éviter que plusieurs jets reconstruits puissent être associés à un même jet *vrai*, si un jet reconstruit est associé à un jet *vrai* auquel un jet reconstruit a déjà été associé, alors ce jet reconstruit n'entre pas dans le calcul du numérateur. Ceci peut arriver lorsque des jets provenant de l'empilement se trouvent proches d'un jet *vrai*. Seules des valeurs faibles

pour l'impulsion transverse sont montrées, car l'efficacité devient très proche de 1 pour tous les jets (et notamment dans toutes les régions du détecteur) à plus haute impulsion transverse (par exemple, c'est déjà le cas dans l'intervalle suivant, pour  $30 \text{ GeV} < p_T^{\text{true}} < 60 \text{ GeV}$ ). La première ligne représente l'efficacité en fonction de  $\eta$ , la seconde en fonction de  $\mu$ , et la troisième en fonction de  $N_{\text{PV}}$ . Sur la Figure 5.13a, on peut remarquer que l'efficacité est proche de 1 dans la région couverte par le détecteur interne, ce qui permet d'utiliser l'information des traces, qui permet d'améliorer la reconstruction des jets, et en particulier à basse impulsion transverse. L'efficacité se dégrade lorsque l'information des traces ne peut être utilisée, c'est-à-dire pour  $|\eta| > 2,5$ . On peut aussi remarquer que l'efficacité de reconstruction est plus faible pour les jets auxquels les algorithmes *CS+SK* sont appliqués que pour les jets auxquels ils ne sont pas appliqués, en particulier dans la région  $|\eta| > 2,5$ . Ceci peut s'expliquer par le fait que ces algorithmes permettent d'enlever certains constituants, en particuliers ceux qui proviennent probablement de l'empilement. Cependant, il peut arriver que ces algorithmes enlèvent de « vrais » constituants, ce qui rend la reconstruction des jets plus difficile. La différence due aux UFOs (par exemple entre les *CS+SK* PFlow et les *CS+SK* UFO) est négligeable sur cette Figure. Enfin, sur cette Figure, la transition avec le calorimètre avant à  $|\eta| \simeq 3,2$  est visible, ce qui se manifeste par une dégradation de l'efficacité dans cette région. Sur la Figure 5.13c, on peut remarquer que l'efficacité décroît en fonction de  $\mu$ , et donc se dégrade avec l'empilement, pour les jets auxquels les algorithmes *CS+SK* sont appliqués, ce qui n'est pas le cas pour les autres collections de jets. Au contraire, sur la 5.13e, l'efficacité est plus indépendante en fonction de  $N_{\text{PV}}$  pour les jets auxquels les algorithmes *CS+SK* sont appliqués. Ceci peut s'expliquer par le fait que  $N_{\text{PV}}$  est une variable permettant de décrire l'empilement dans le temps, alors que  $\mu$  permet de décrire l'empilement hors-temps. Or, l'empilement dans le temps se manifeste par des dépôts d'énergie positive dans les calorimètres, alors que l'empilement hors-temps se manifeste par des dépôts d'énergie négative, tel que décrit dans la Section 4.3.1.1. Les algorithmes *CS+SK* sont conçus de manière à enlever principalement les dépôts d'énergie positive dûs à l'empilement, et non pas les dépôts d'énergie négative. Sur ces deux Figures aussi, la différence due aux UFOs est négligeable. Pour des jets *vrais* qui satisfont  $20 \text{ GeV} < p_T^{\text{true}} < 30 \text{ GeV}$ , l'efficacité est plus proche de 1 et est moins dépendante en fonction des variables étudiées, ce qui indique que ces jets sont mieux reconstruits. Ceci s'explique par le fait que ces jets ont moins de chance d'être mal reconstruits et sont moins sensibles à l'empilement, car les dépôts d'énergie dans les calorimètres sont plus significatifs.



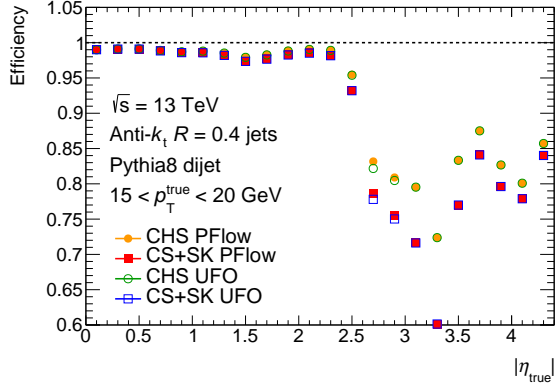
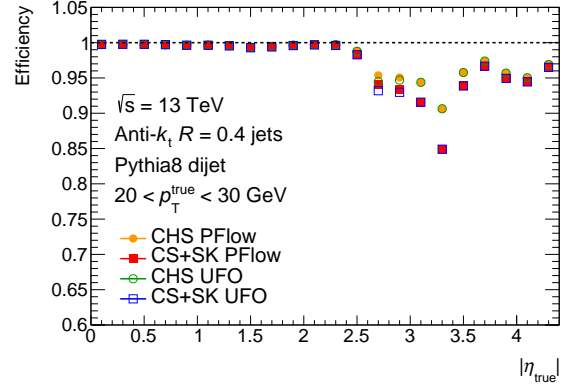
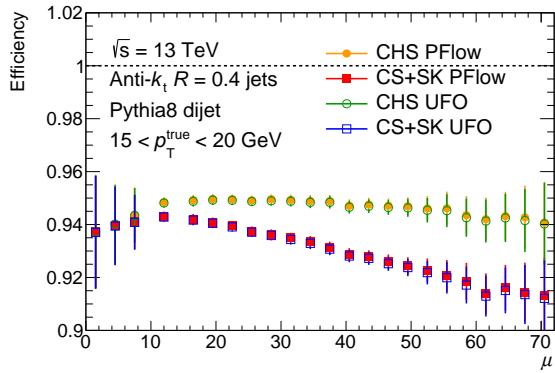
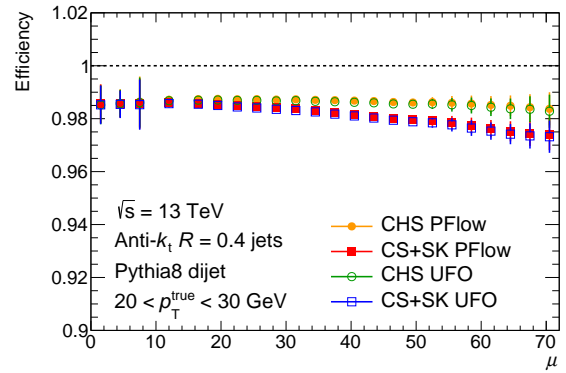
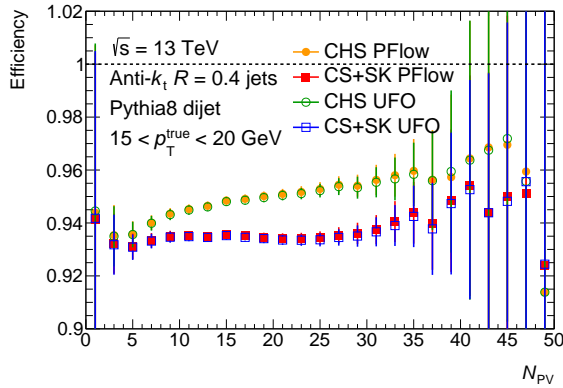
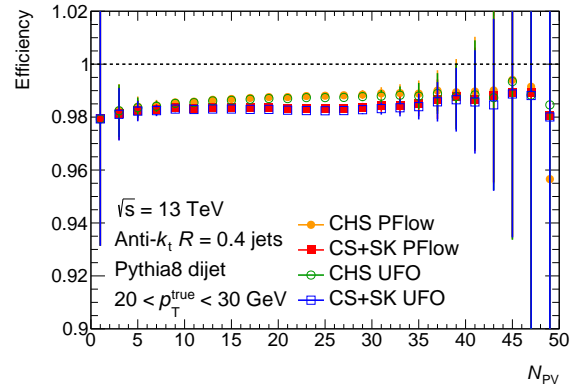

 (a) En fonction de  $\eta$ , pour  $15 \text{ GeV} < p_T^{\text{true}} < 20 \text{ GeV}$ .

 (b) En fonction de  $\eta$ , pour  $20 \text{ GeV} < p_T^{\text{true}} < 30 \text{ GeV}$ .

 (c) En fonction de  $\mu$ , pour  $15 \text{ GeV} < p_T^{\text{true}} < 20 \text{ GeV}$ .

 (d) En fonction de  $\mu$ , pour  $20 \text{ GeV} < p_T^{\text{true}} < 30 \text{ GeV}$ .

 (e) En fonction de  $N_{PV}$ , pour  $15 \text{ GeV} < p_T^{\text{true}} < 20 \text{ GeV}$ .

 (f) En fonction de  $N_{PV}$ , pour  $20 \text{ GeV} < p_T^{\text{true}} < 30 \text{ GeV}$ .

 FIGURE 5.13 – Efficacité en fonction de  $\eta$  (première ligne),  $\mu$  (seconde ligne) et  $N_{PV}$  (troisième ligne) pour des jets avec  $15 \text{ GeV} < p_T^{\text{true}} < 20 \text{ GeV}$  (première colonne) et  $20 \text{ GeV} < p_T^{\text{true}} < 30 \text{ GeV}$  (seconde colonne).

La Figure 5.14 représente une comparaison de la pureté de reconstruction en fonction de  $\eta$  (première ligne),  $\mu$  (seconde ligne) et  $N_{PV}$  (troisième ligne) pour les quatre différentes collections de jets étudiées. Les jets reconstruits satisfont  $15 \text{ GeV} < p_T^{\text{reco}} < 20 \text{ GeV}$

(première colonne) et  $20 \text{ GeV} < p_T^{\text{reco}} < 30 \text{ GeV}$  (seconde colonne), et sont associés à des jets vrais. La pureté est définie par le nombre de jets reconstruits qui sont associés à un jet *vrai* divisé par le nombre de jets reconstruits :  $\varepsilon = \frac{N_{\text{reco jets matched to truth jet}}}{N_{\text{reco jets}}}$ . Aucune condition sur l'isolation des jets n'est appliquée. Ainsi, plus la pureté est proche de 1, plus les performances sont bonnes. Seules des valeurs faibles pour l'impulsion transverse sont montrées, car la pureté devient proche de 1 pour tous les jets à plus haute impulsion transverse, par exemple pour  $30 \text{ GeV} < p_T^{\text{reco}} < 60 \text{ GeV}$  (bien que pour cet intervalle, la pureté ne soit pas exactement égale à 1, elle en est proche. Des valeurs plus faibles pour l'impulsion transverse sont montrées ici car les différences sont plus significatives). La pureté est un indicateur du nombre de jets *faux*, c'est-à-dire du nombre de jets reconstruits qui ne proviennent pas d'un jet *vrai*, mais de l'empilement. De la même façon que sur la Figure 5.13, si un jet reconstruit est associé à un jet *vrai* auquel un jet reconstruit a déjà été associé, alors ce jet reconstruit n'entre pas dans le calcul du numérateur. Sur la Figure 5.14a, la pureté est plus proche de 1 dans la région couverte par le détecteur interne ( $|\eta| < 2,5$ ), ce qui permet d'utiliser l'information des traces, et est meilleure pour les jets auxquels les algorithmes *CS+SK* sont appliqués, car moins de jets *faux* sont reconstruits, car ces algorithmes permettent d'enlever les constituants provenant de l'empilement. Les transitions entre les différentes parties du calorimètres à  $|\eta| \simeq 1,5$  et  $|\eta| \simeq 3,2$  sont aussi visibles sur cette Figure. Sur les Figures 5.14c et 5.14e, la pureté est meilleure pour les jets auxquels les algorithmes *CS+SK* sont appliqués, et est légèrement moins dépendante en fonction de l'empilement, ce qui indique que moins de jets *faux* sont reconstruits. Sur ces trois Figures, la différence apportée par les UFOs est négligeable. Pour des jets reconstruits qui satisfont  $20 \text{ GeV} < p_T^{\text{reco}} < 30 \text{ GeV}$ , la pureté est plus proche de 1, ce qui indique qu'il y a moins de jets *faux* qui sont reconstruits.

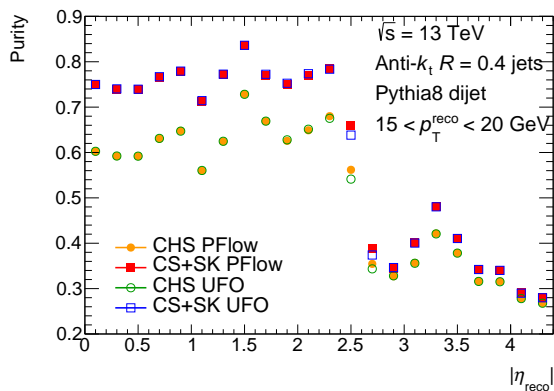
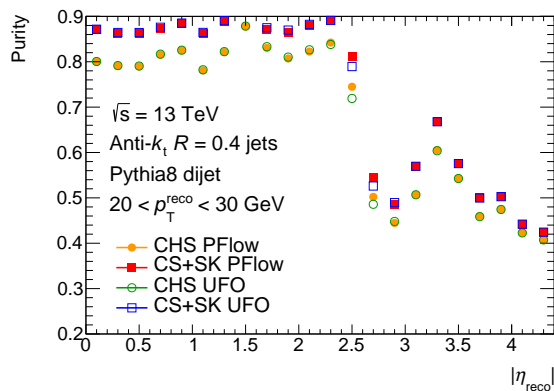
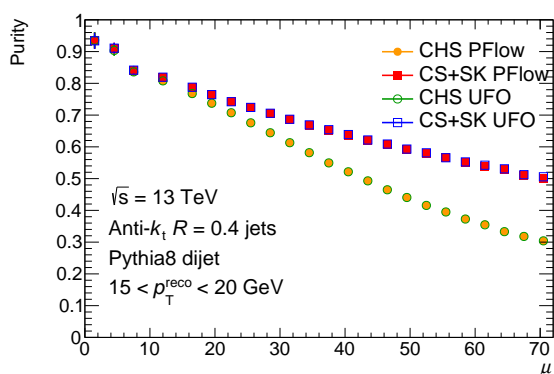
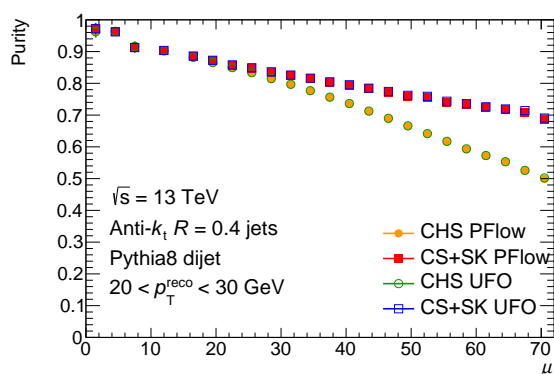
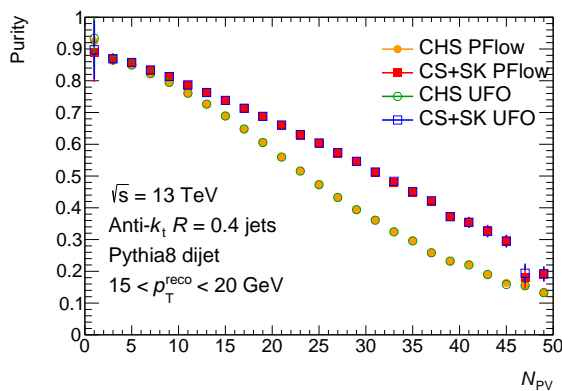
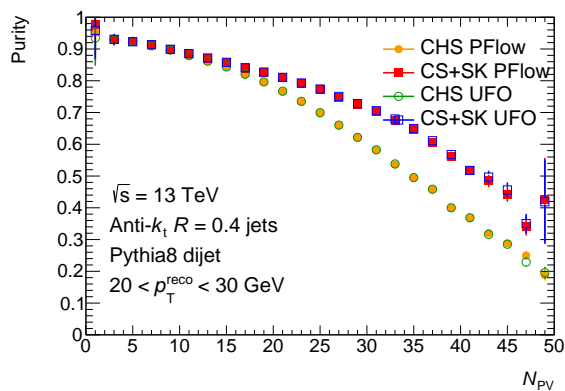

 (a) En fonction de  $\eta$ , pour  $15 \text{ GeV} < p_T^{\text{reco}} < 20 \text{ GeV}$ .

 (b) En fonction de  $\eta$ , pour  $20 \text{ GeV} < p_T^{\text{reco}} < 30 \text{ GeV}$ .

 (c) En fonction de  $\mu$ , pour  $15 \text{ GeV} < p_T^{\text{reco}} < 20 \text{ GeV}$ .

 (d) En fonction de  $\mu$ , pour  $20 \text{ GeV} < p_T^{\text{reco}} < 30 \text{ GeV}$ .

 (e) En fonction de  $N_{PV}$ , pour  $15 \text{ GeV} < p_T^{\text{reco}} < 20 \text{ GeV}$ .

 (f) En fonction de  $N_{PV}$ , pour  $20 \text{ GeV} < p_T^{\text{reco}} < 30 \text{ GeV}$ .

 FIGURE 5.14 – Pureté en fonction de  $\eta$  (première ligne),  $\mu$  (seconde ligne) et  $N_{PV}$  (troisième ligne) pour des jets avec  $15 \text{ GeV} < p_T^{\text{reco}} < 20 \text{ GeV}$  (première colonne) et  $20 \text{ GeV} < p_T^{\text{reco}} < 30 \text{ GeV}$  (seconde colonne).

### 5.3.3.2 Dépendance de l'impulsion transverse en fonction de variables d'empilement

La Figure 5.15 représente une comparaison de la dépendance de  $p_T^{\text{reco}}$  en fonction de  $\mu$  entre les quatre différentes collections de jets étudiées, et ce dans plusieurs intervalles en  $p_T^{\text{true}}$ . Cette Figure est obtenue en produisant des histogrammes de  $\Delta p_T = p_T^{\text{reco}} - p_T^{\text{true}}$  en intervalles en  $(p_T^{\text{true}}, |\eta_{\text{true}}|, \mu$  et  $N_{\text{PV}}$ ). Ensuite, pour chaque intervalle en  $(p_T^{\text{true}}, |\eta_{\text{true}}|, N_{\text{PV}})$ , la dépendance de  $\Delta p_T$  est tracée en fonction de  $\mu$ , et ajustée avec une fonction affine ( $a + b \times \mu$ ). La pente de cette courbe ( $b$ ) est ensuite tracée en fonction de  $N_{\text{PV}}$  en intervalles en  $(p_T^{\text{true}}, |\eta_{\text{true}}|)$ . Puis, ces courbes sont ajustées avec une constante, afin d'en prendre la valeur moyenne. Enfin, la valeur de cette constante, qui représente la pente en fonction de  $\mu$ , moyennée en fonction de  $N_{\text{PV}}$ , est tracée en fonction de  $|\eta_{\text{true}}|$ , en intervalles de  $p_T^{\text{true}}$ . Ainsi, les meilleures performances sont obtenues pour les courbes qui sont les plus proches de 0, ce qui indique une plus faible dépendance en fonction de  $\mu$ . Sur ces Figures, on peut remarquer qu'à basse impulsion transverse, les meilleures performances sont obtenues par les jets auxquels les algorithmes *CS+SK* sont appliqués dans la région centrale du détecteur, car ces courbes sont plus proches de 0. Mais ce n'est pas le cas dans la région avant, où les performances sont légèrement moins bonnes. À haute  $p_T^{\text{true}}$ , les performances sont moins bonnes. La dépendance n'est pas nulle pour les jets auxquels les algorithmes *CS+SK* sont appliqués, ce qui indique qu'une correction d'empilement est tout de même nécessaire. La différence due aux UFOs est négligeable. La dépendance négative en fonction de  $\mu$  résulte de la forme de l'impulsion électrique dans le calorimètre en Argon liquide, qui est négative durant la période qui suit le signal, tel que décrit dans la Section 4.3.1.1. Or les dépôts d'énergie négative ainsi produits sont inclus dans les jets. La transition avec le calorimètre avant à  $|\eta| \simeq 3,1 - 3,2$  est visible sur ces figures. Enfin, on peut remarquer que la dépendance s'accroît en fonction de  $p_T^{\text{true}}$ .

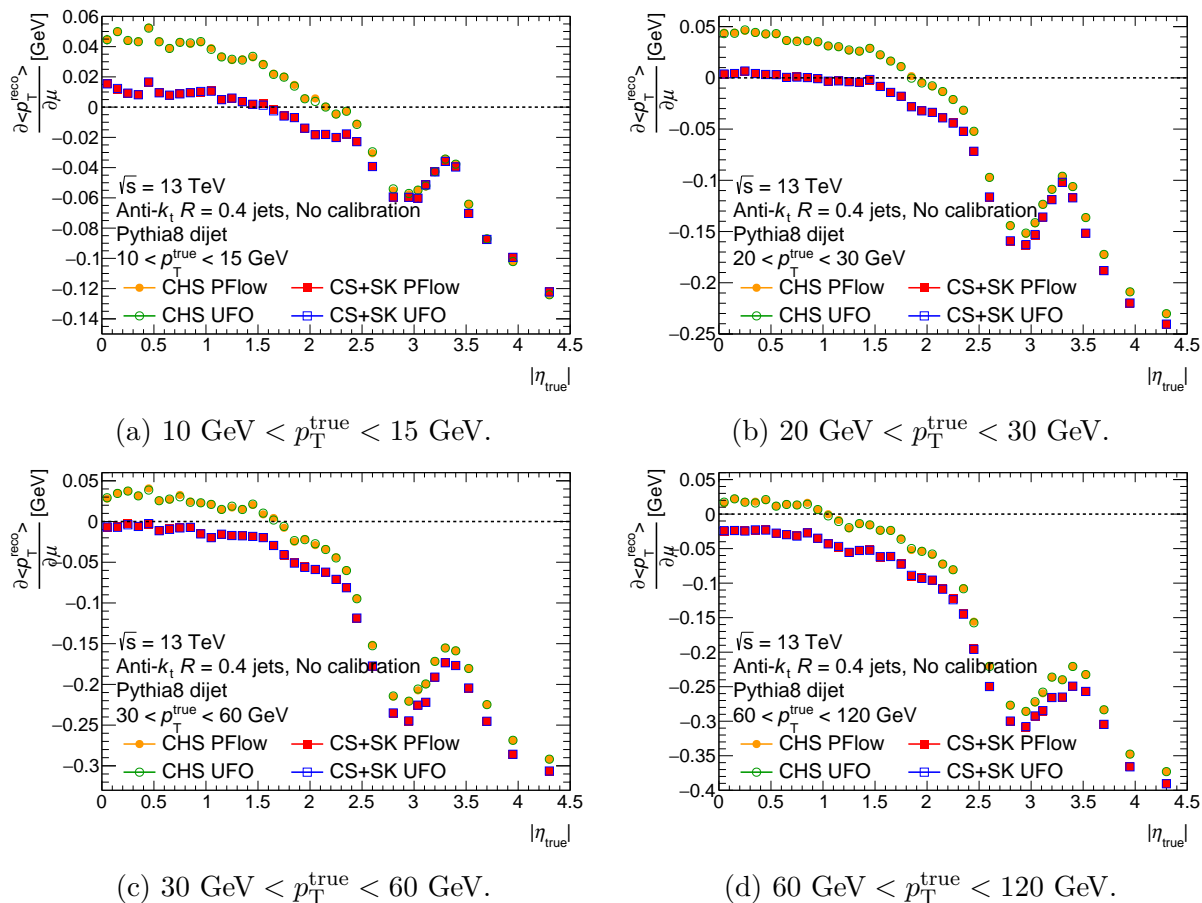


FIGURE 5.15 – Dépendance de  $p_T^{\text{reco}}$  en fonction de  $\mu$ , tracée en fonction de  $|\eta_{\text{true}}|$ , pour quatre différents intervalles en  $p_T^{\text{true}}$ .

La Figure 5.16 représente la même quantité que dans la Figure 5.15, mais cette fois-ci pour la pente en fonction de  $N_{\text{PV}}$ . La procédure pour obtenir ces courbes est similaire à la procédure précédente, mais en remplaçant  $N_{\text{PV}}$  par  $\mu$ , et  $\mu$  par  $N_{\text{PV}}$ . Sur ces Figures, on peut remarquer que les meilleures performances sont obtenues par les jets auxquels les algorithmes *CS+SK* sont appliqués dans toutes les régions du détecteur et dans toute la gamme cinématique. Cependant, la dépendance n'est pas nulle pour les jets auxquels les algorithmes *CS+SK* sont appliqués, ce qui indique qu'une correction d'empilement est tout de même nécessaire. La différence due aux UFOs est négligeable. La pente est cette fois-ci positive, car  $N_{\text{PV}}$  est une variable permettant de mesurer l'empilement *dans le temps*, c'est-à-dire des collisions qui se produisent en même temps que la collision d'intérêt, ce qui vient ajouter du bruit dans le détecteur. La transition avec le calorimètre avant à  $|\eta| \simeq 3,1 - 3,2$  est visible sur ces Figures. Enfin, on peut remarquer que la dépendance en fonction de  $N_{\text{PV}}$  augmente avec  $p_T^{\text{true}}$ .

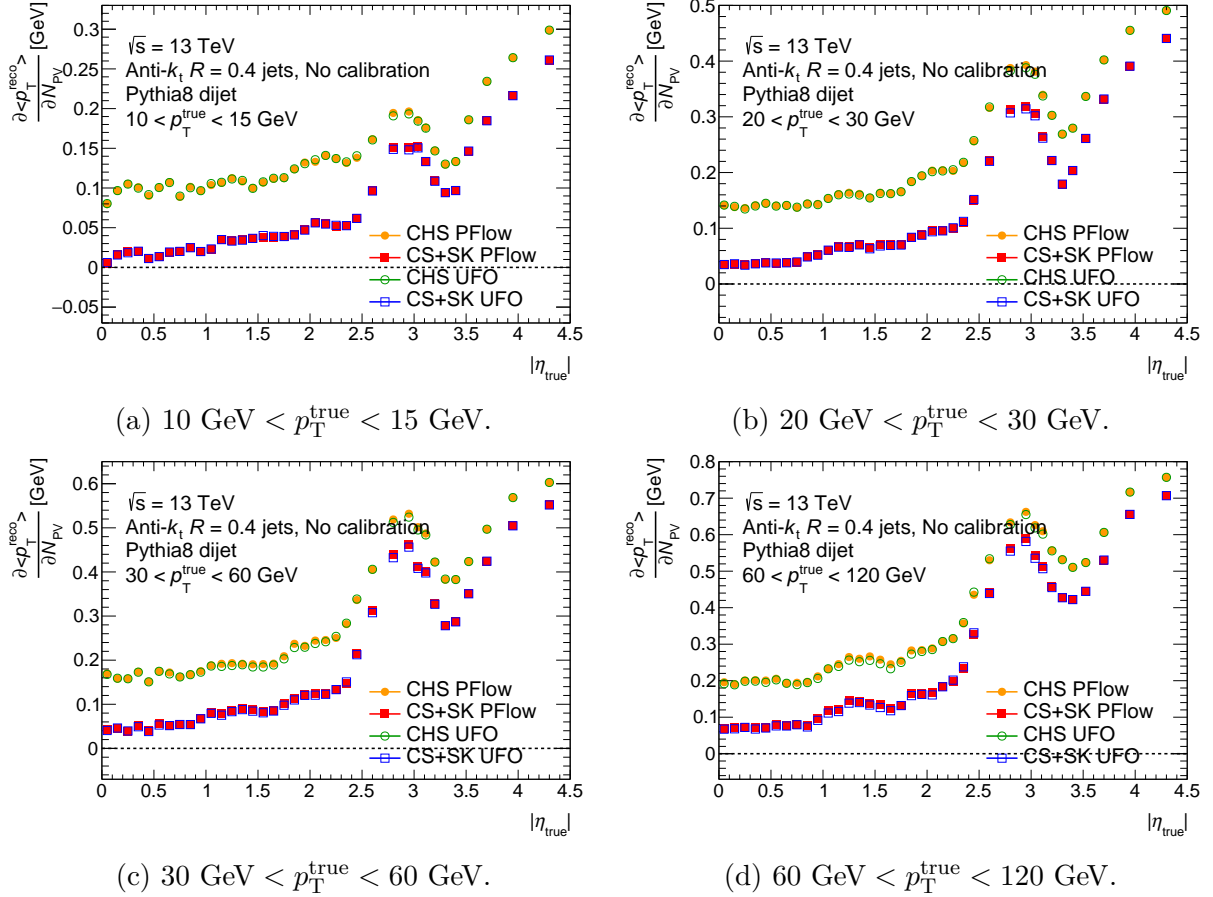


FIGURE 5.16 – Dépendance de  $p_T^{\text{reco}}$  en fonction de  $N_{\text{PV}}$ , tracée en fonction de  $|\eta_{\text{true}}|$ , pour quatre différents intervalles en  $p_T^{\text{true}}$ .

### 5.3.3.3 Échelle en énergie et en masse

La Figure 5.17 représente l'échelle en énergie (la moyenne d'ajustements avec des gaussiennes des distributions  $\frac{E_{\text{reco}}}{E_{\text{true}}}$ ) en fonction de  $E_{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$ . Sur les trois Figures, l'échelle en énergie est similaire entre les collections de jets auxquels les algorithmes *CS+SK* sont appliqués, et ceux auxquels ils ne sont pas appliqués. Cela peut s'expliquer par le fait que les Particle Flows et les UFOs ont la même énergie : l'algorithme des TTCs qui s'applique aux Particle Flows dans l'algorithme des UFOs ne fait que redistribuer l'énergie entre différents constituants, et crée plusieurs constituants, mais l'énergie totale de ces constituants reste la même. Ceci motive le fait de pouvoir utiliser le  $\rho_{\text{PUSB}}$  venant des Particle Flows pour les CHS UFO, comme discuté dans la Section 5.2.1. Dans la région avant du détecteur (Figure 5.17c), la différence entre les jets avec et sans les algorithmes *CS+SK* est plus faible que dans la région centrale, car le  $\rho_{\text{PUSB}}$  utilisé par les algorithmes de *Constituent Subtraction* et de *SoftKiller* est calculé dans la région centrale du détecteur ( $|\eta| < 2, 0$ ) et car la granularité du calorimètre y est plus faible.

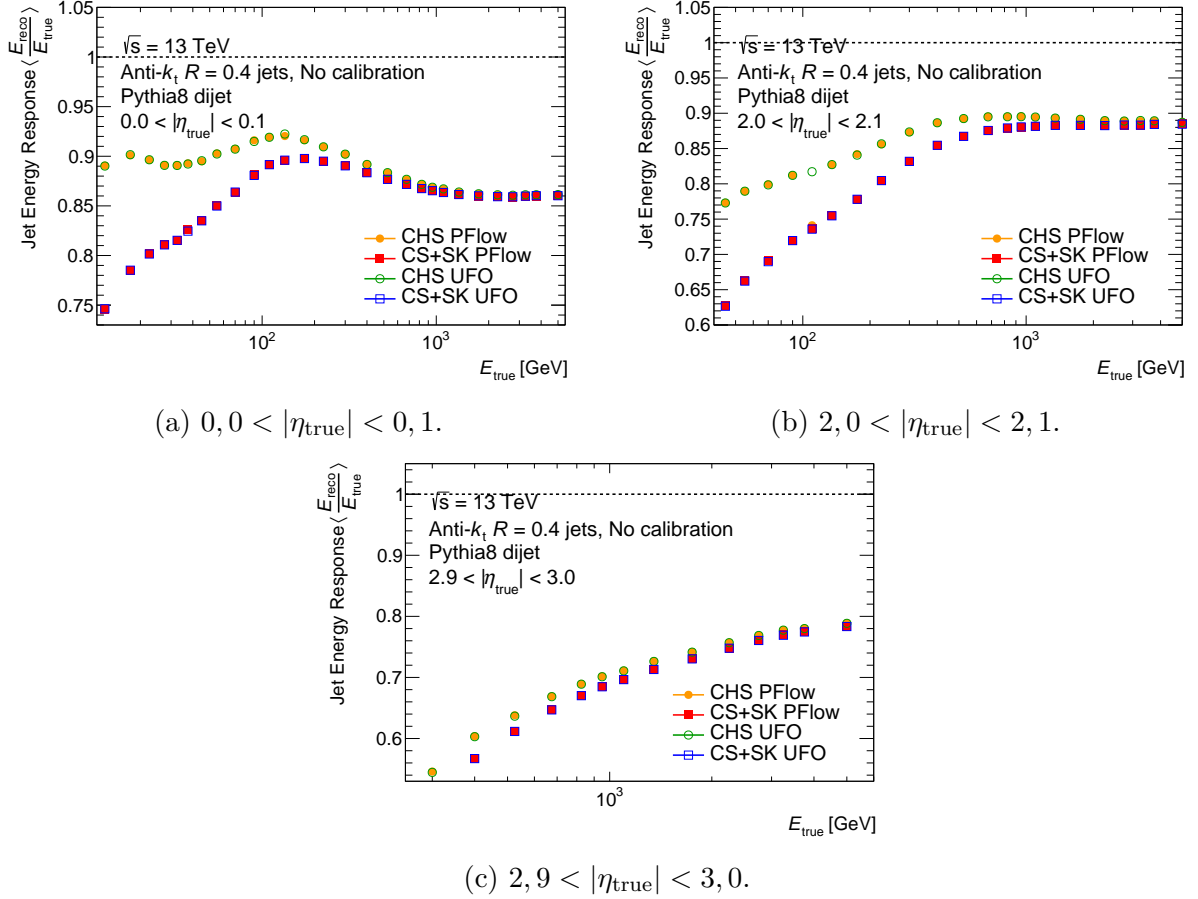


FIGURE 5.17 – Échelle en énergie en fonction de  $E_{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$ .

Dans la Figure 5.17a, la structure à  $E_{\text{true}} \simeq 30 - 40$  GeV présente pour les CHS PFlow et les CHS UFO, mais pas pour les jets auxquels les algorithmes *CS+SK* sont appliqués, provient de l'empilement, qui est en partie enlevé par ces derniers. Ceci peut être démontré dans la Figure 5.18, où sont comparées les échelles d'énergie avant (c'est-à-dire au niveau reconstruit) et après correction d'aire pour les CHS PFlow jets : la structure à  $E_{\text{true}} \simeq 30 - 40$  GeV disparaît en effet après correction d'aire. La structure n'est pas non plus présente à plus haut  $|\eta|$  sur les Figures 5.17b et 5.17c, car il n'y a pas suffisamment d'événements dans ces régions de l'espace des phases. L'échelle d'énergie est plus basse après correction d'aire, car celle-ci correspond à une soustraction (voir l'Équation 5.1), car l'énergie provenant de l'empilement est soustraite. Cet effet est plus significatif à basse énergie, car l'énergie provenant de l'empilement représente une plus grande proportion de l'énergie des jets qu'à haute énergie. Le fait que l'échelle en énergie diminue près la correction d'aire ne signifie pas que les performances se dégradent, car l'objectif de la correction absolue en énergie qui est appliquée ensuite est que l'échelle d'énergie soit approximativement 1. Enfin, sur la Figure 5.18, la courbe après correction d'aire peut être comparée aux résultats « officiels » d'ATLAS de la Figure 5.9b. Les deux courbes sont très similaires, ce qui permet de valider (partiellement) ma méthode pour obtenir les résultats.

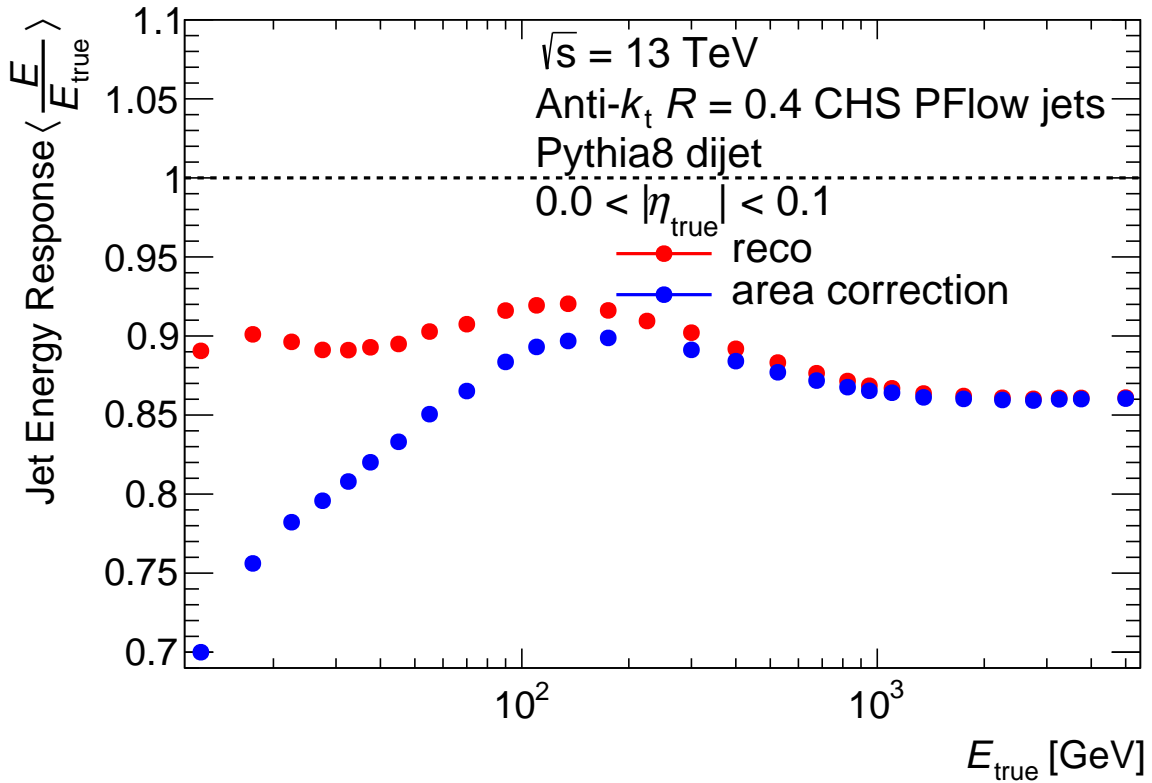


FIGURE 5.18 – Échelle en énergie en fonction de  $E_{\text{true}}$  avant et après correction d’aire pour les CHS PFlow jets.

La Figure 5.19 représente l’échelle en masse en fonction de  $p_{\text{T}}^{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$  et en  $m_{\text{true}}$ . Cette Figure met en évidence la différence majeure entre les Particle Flows et les UFOs : la masse. L’algorithme des TCCs utilisé lors de la construction des UFOs permet d’améliorer la résolution spatiale des constituants qui composent les jets. La masse est une variable extrêmement sensible à cette résolution, et c’est pourquoi elle diffère tant entre les UFOs et les Particle Flows. Plus il y a de constituants, plus la masse est élevée, et c’est pourquoi ce sont les CHS UFO qui ont l’échelle en masse la plus élevée. Ensuite les algorithmes de  $CS+SK$  permettent de retirer certains constituants, et c’est pourquoi la masse des  $CS+SK$  UFO est plus faible que celle des CHS UFO. À haute impulsion transverse, la plupart de l’énergie des jets provient des calorimètres, et non des traces. Or, la résolution angulaire des calorimètres n’est pas très bonne. Ainsi, la séparation angulaire des constituants est sur-estimée, et c’est pourquoi l’échelle en masse devient plus supérieure à 1 à haute impulsion transverse. La masse des jets de rayon  $R = 0,4$  n’est pas fréquemment utilisée, mais peut être utile dans le cadre de topologies extrêmement boostées, comme pour des bosons  $W$  ou des quarks  $t$  avec une impulsion transverse suffisamment élevée pour être contenus dans des jets de rayon  $R = 0,4$  (pour rappel, le rayon d’un jet est donné par  $R \sim \frac{2m}{p_{\text{T}}}$ ), ou encore dans le contexte de jets de rayon  $R = 1,0$  re-regroupés à partir de jets de rayon  $R = 0,4$  [78] à la place des constituants directement.



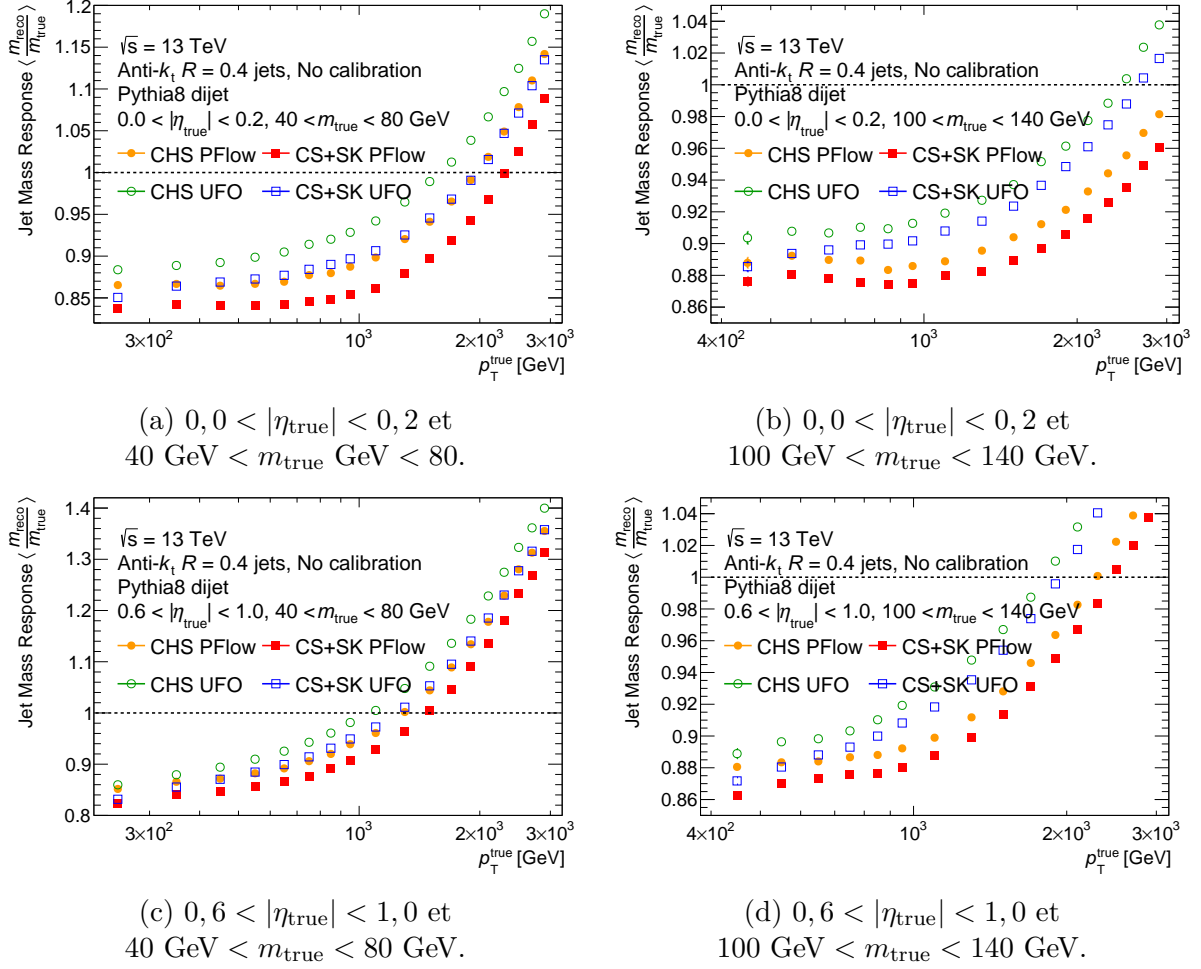
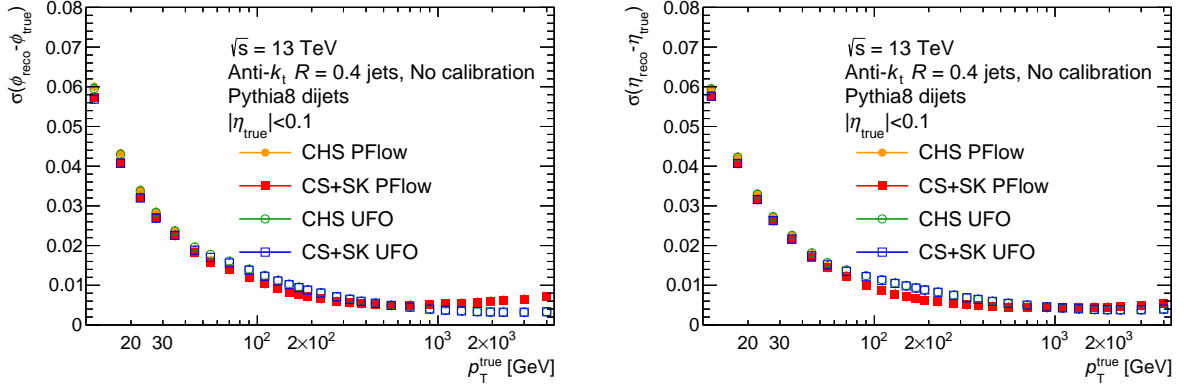


FIGURE 5.19 – Échelle en masse en fonction de  $p_{\text{T}}^{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$  et en  $m_{\text{true}}$ .

### 5.3.3.4 Résolution angulaire

La Figure 5.20 représente la résolution angulaire (c'est-à-dire en  $\phi$  et en  $\eta$ ) pour les quatre collections de jets étudiées en fonction de  $p_{\text{T}}^{\text{true}}$ . Ces Figures sont obtenues en remplissant des histogrammes de  $\phi_{\text{reco}} - \phi_{\text{true}}$  et de  $\eta_{\text{reco}} - \eta_{\text{true}}$  en intervalles de  $p_{\text{T}}^{\text{true}}$ . Ensuite, pour chaque intervalle en  $p_{\text{T}}^{\text{true}}$ , les histogrammes sont ajustés avec des gaussiennes. La résolution correspond alors à la largeur des gaussiennes ainsi obtenues. On peut remarquer qu'à basse impulsion transverse, les algorithmes *CS+SK* permettent d'améliorer la résolution angulaire. Pour  $50 \text{ GeV} < p_{\text{T}}^{\text{true}} < 300 \text{ GeV}$ , ce qui correspond à la région où la soustraction des traces est désactivée dans l'algorithme des Particle Flow, la résolution angulaire est dégradée avec les UFOs. Cette caractéristique n'est pas comprise pour le moment. Enfin, à haute  $p_{\text{T}}^{\text{true}}$ , la résolution angulaire est améliorée pour les UFOs, ce qui peut s'expliquer par le fait que l'algorithme des TCCs y est utilisé pour améliorer la résolution angulaire des constituants. Ces Figures sont obtenues avant tout étalonnage, mais les résultats sont similaires après étalonnage. En particulier, il n'y a pas d'étalonnage pour  $\phi$ .

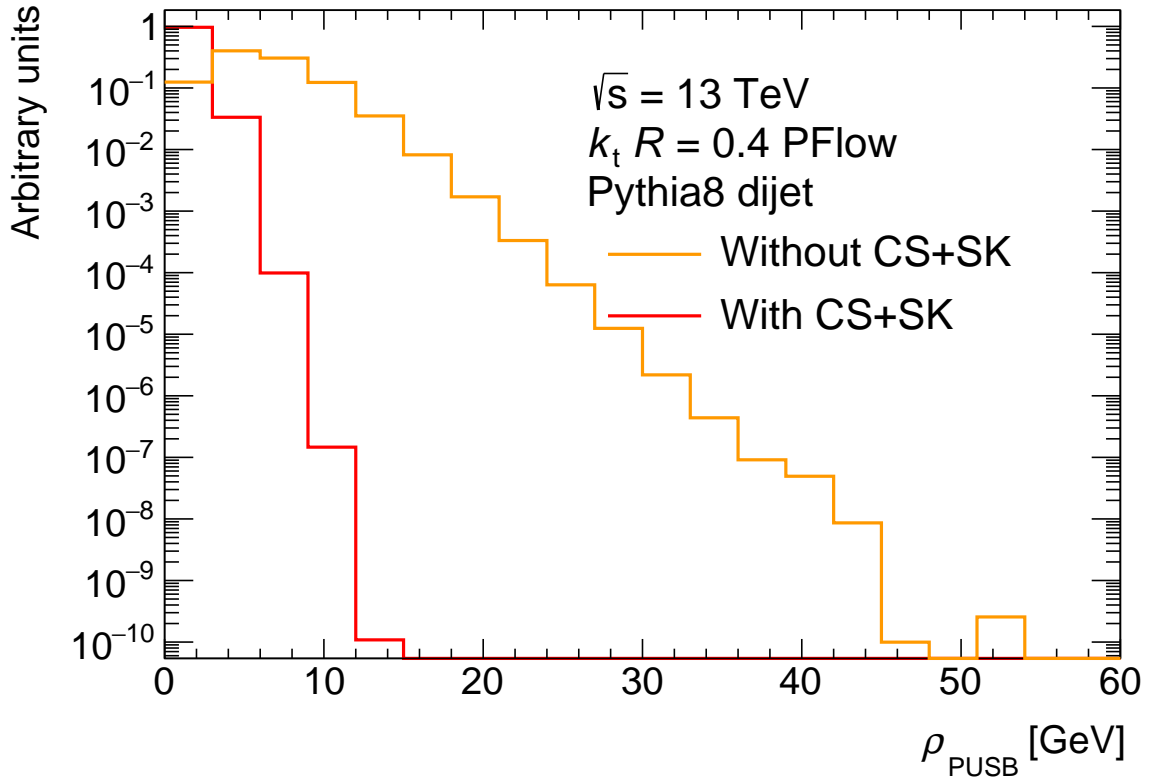

 (a) Résolution en  $\phi$ .

 (b) Résolution en  $\eta$ .

 FIGURE 5.20 – Résolution angulaire en fonction de  $p_T^{\text{true}}$ .

### 5.3.4 Résultats après correction de l'empilement

La Figure 5.21 représente la valeur de  $\rho_{\text{PUSB}}$  obtenue avec et sans les algorithmes *CS+SK* appliqués aux constituants. Comme décrit dans la Section 4.3.4, lorsque les algorithmes *CS+SK* sont appliqués, la valeur de  $\rho_{\text{PUSB}}$  est plus proche de 0, et c'est pourquoi nous avons décidé de n'appliquer que la correction résiduelle aux jets auxquels les algorithmes *CS+SK* sont appliqués.


 FIGURE 5.21 –  $\rho_{\text{PUSB}}$  avec et sans *CS+SK*.

Les Figures 5.22 et 5.23 représentent les dépendances en fonction de  $\mu$  et de  $N_{\text{PV}}$

(comme dans les Figures 5.15 et 5.16), mais ici seulement pour les CHS PFlow jets après les quatre étapes de correction d'empilement : sans aucune correction (seulement au niveau reconstruit), après correction d'aire seulement, après correction d'aire et l'ancienne correction d'empilement (appelée « *1D correction* ») et après correction d'aire et la nouvelle correction d'empilement (appelée « *3D correction* »). Sur les Figures 5.22a et 5.23a ( $10 \text{ GeV} < p_T^{\text{true}} < 15 \text{ GeV}$ ), on peut remarquer que la correction d'aire réduit la dépendance de  $p_T$  en  $\mu$  et en  $N_{PV}$  dans la région centrale du détecteur, mais pas dans la région *avant*. Ceci est dû au fait que  $\rho_{\text{PUSB}}$  est défini par des jets qui satisfont  $|\eta| < 2,0$ , et donc la correction n'est pas parfaite au-delà. Cependant, même après la correction d'aire il persiste une dépendance aux variables d'empilement, ce qui motive l'utilisation de corrections résiduelles. Sur ces Figures, l'ancienne correction résiduelle (« *1D correction* ») sur-corrige l'empilement, car cette correction est dérivée à partir de jets pour lesquels  $20 \text{ GeV} < p_T^{\text{true}} < 30 \text{ GeV}$  où la dépendance en empilement est plus forte, puis est appliquée à un intervalle plus faible en  $p_T^{\text{true}}$ . Enfin, la nouvelle correction d'empilement (« *3D correction* ») permet de corriger presque totalement la dépendance en  $\mu$  et  $N_{PV}$ . Sur les Figures 5.22b et 5.23b ( $20 \text{ GeV} < p_T^{\text{true}} < 30 \text{ GeV}$ ), l'ancienne correction d'empilement présente de meilleures performances que la nouvelle, car elle est dérivée dans cet intervalle en  $p_T^{\text{true}}$ . Enfin, sur les Figures 5.22c, 5.22d, 5.23c et 5.23d, la nouvelle correction d'aire présente de meilleures performances que l'ancienne, qui sous-corrige l'empilement, car dérivée à partir de jets avec une plus basse  $p_T^{\text{true}}$ . En général, la nouvelle correction d'empilement permet de réduire considérablement la dépendance en  $\mu$  et  $N_{PV}$ , et présente de meilleures performances que l'ancienne correction lorsque toute la gamme cinématique est considérée. C'est pourquoi c'est cette correction qui sera appliquée par la suite.

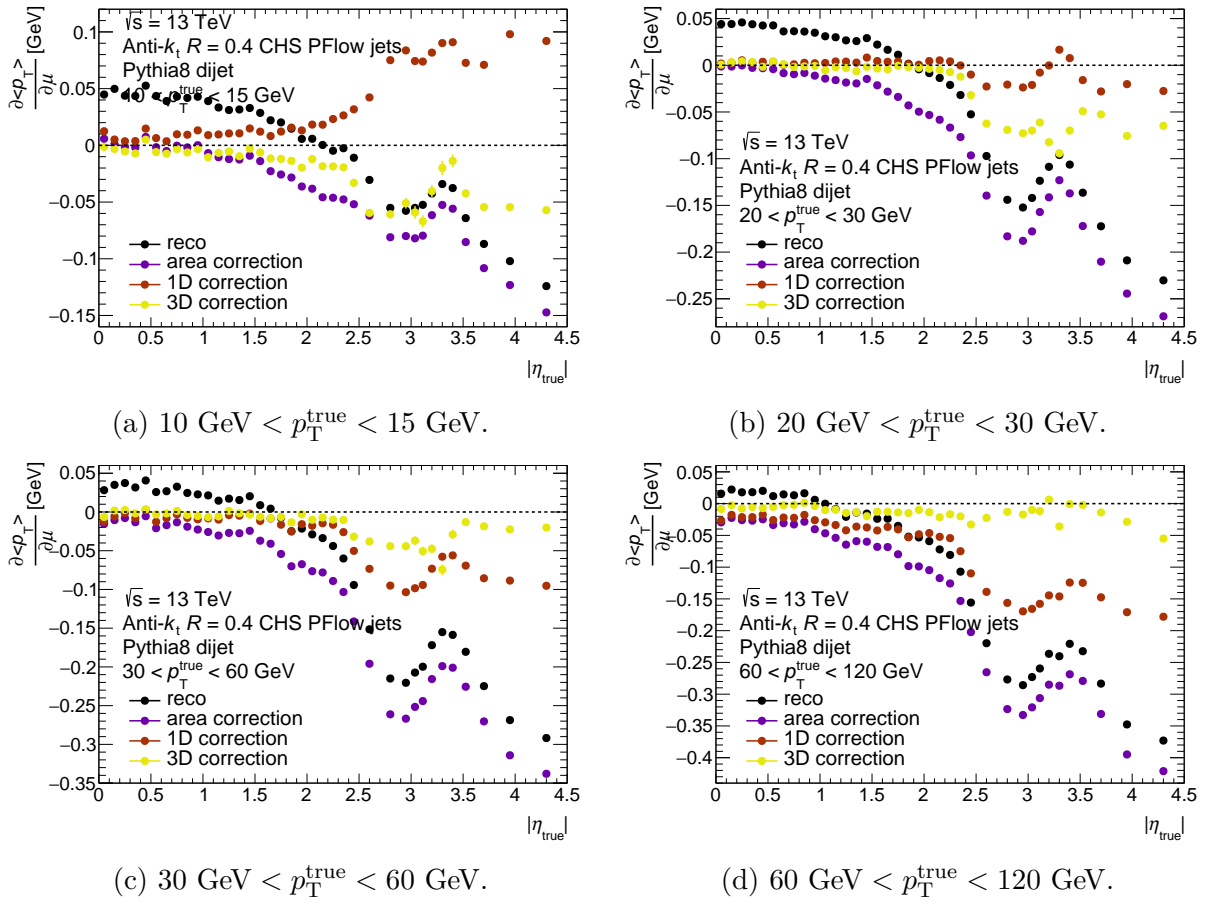


FIGURE 5.22 – Dépendance de  $p_T$  en fonction de  $\mu$ , tracée en fonction de  $|\eta_{\text{true}}|$ , après chaque étape de la correction d'empilement et pour quatre différents intervalles en  $p_T^{\text{true}}$ .

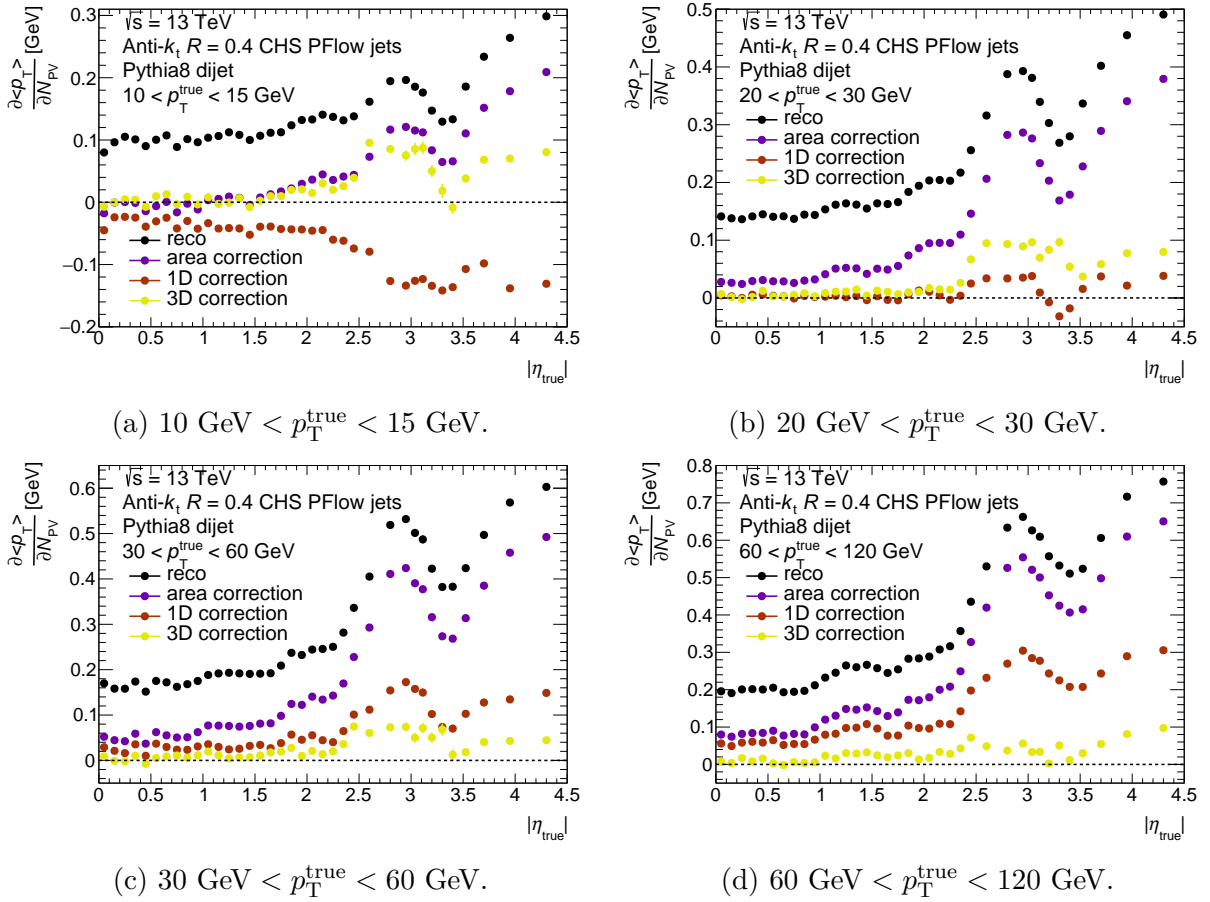


FIGURE 5.23 – Dépendance de  $p_T$  en fonction de  $N_{PV}$ , tracée en fonction de  $|\eta_{\text{true}}|$ , après chaque étape de la correction en empilement et pour quatre différents intervalles en  $p_T^{\text{true}}$ .

Les Figures 5.24 et 5.25 représentent les mêmes variables, mais cette fois-ci après que la nouvelle correction d'empilement (celle notée « *3D correction* » sur les Figures précédentes, et prenant en compte la dépendance de la correction résiduelle en fonction de  $p_T$ ) a été appliquée, pour les quatre types de jets étudiés. Après cette correction, les dépendances en  $\mu$  et  $N_{PV}$  sont très faibles. Sur les Figures 5.24a, 5.24b, 5.25a et 5.25b, les jets auxquels les algorithmes *CS+SK* sont appliqués ont une plus faible dépendance en  $\mu$  et  $N_{PV}$  par rapport aux jets sur lesquels ils ne sont pas appliqués, ce qui indique que l'étalonnage est plus facile à dériver pour ces jets. La différence due aux UFOs est négligeable. Sur les Figures 5.24c, 5.24d, 5.25c et 5.25d, les collections de jets ne peuvent pas être départagées.

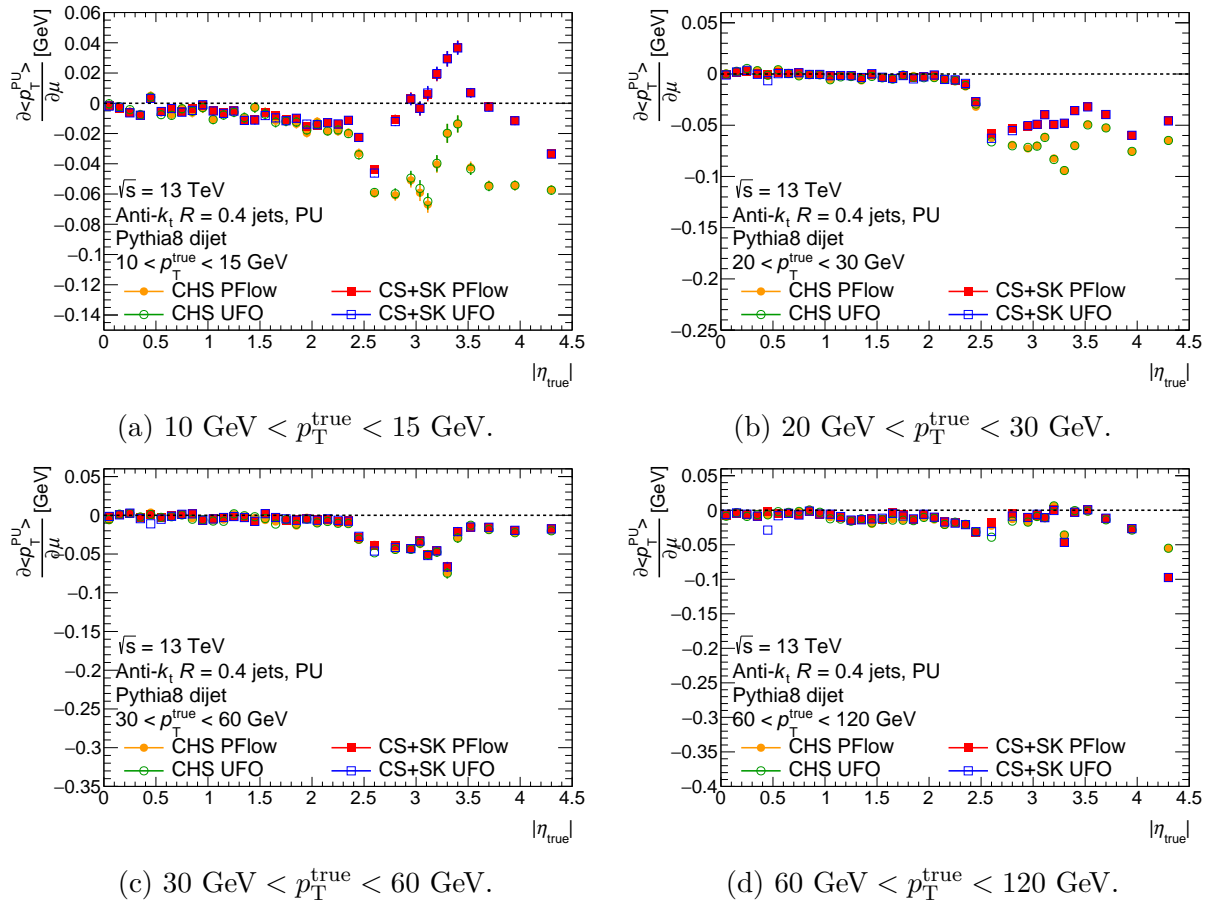


FIGURE 5.24 – Dépendance de  $p_T^{PU}$  en fonction de  $\mu$ , tracée en fonction de  $|\eta_{\text{true}}|$ , après correction d'empilement et pour quatre différents intervalles en  $p_T^{\text{true}}$ .

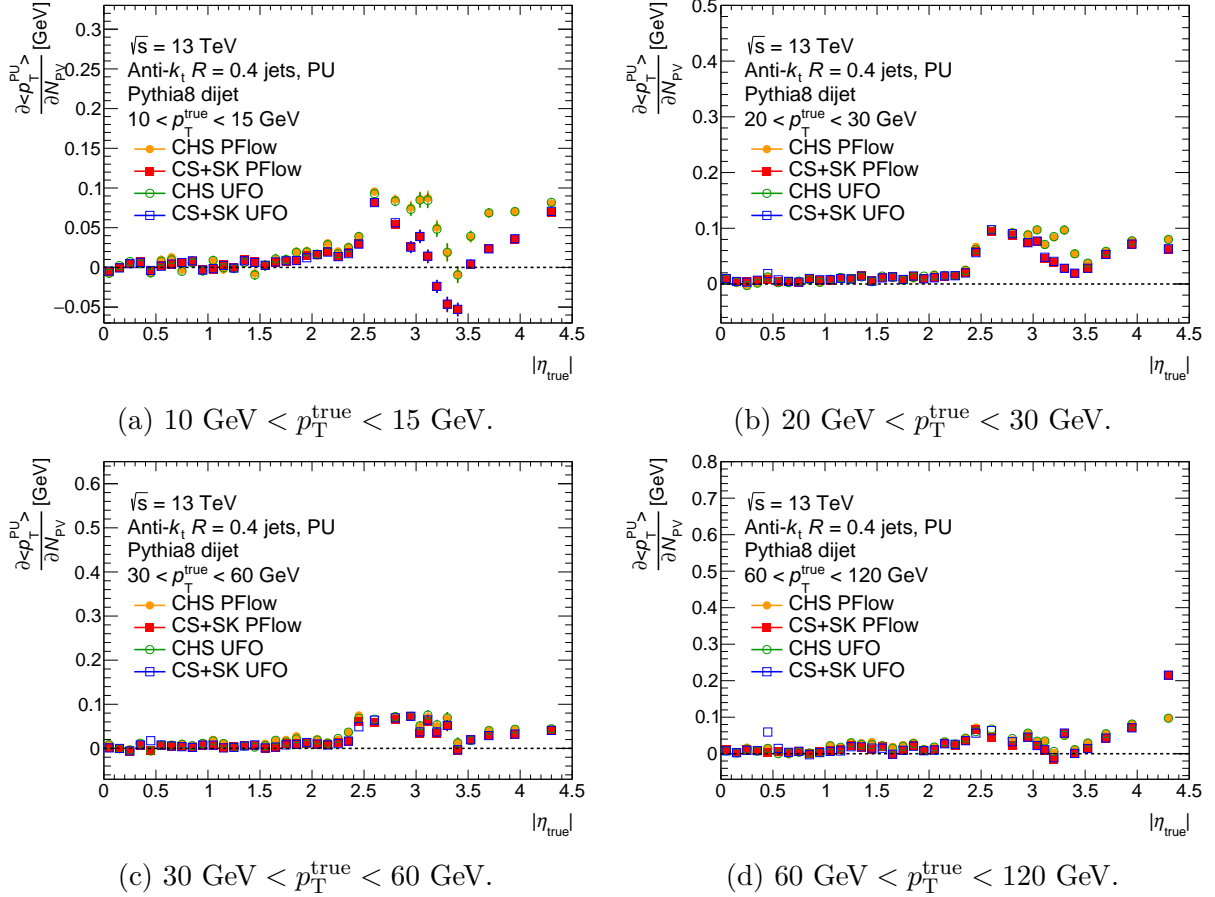
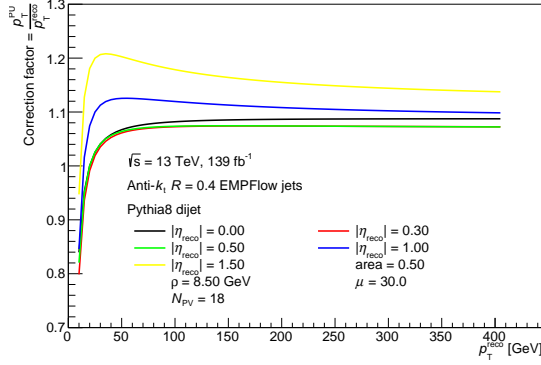
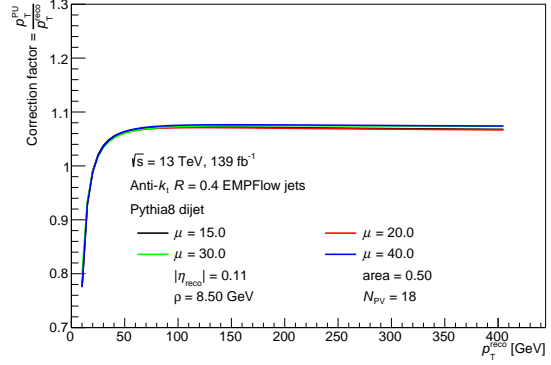
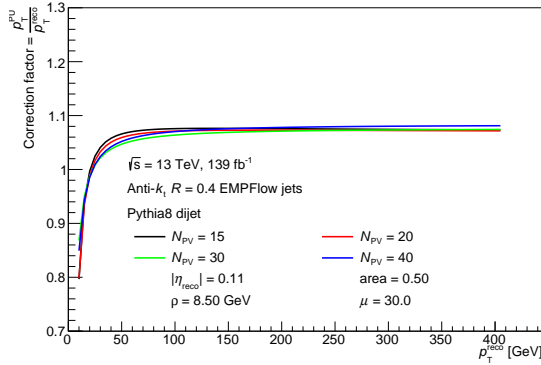
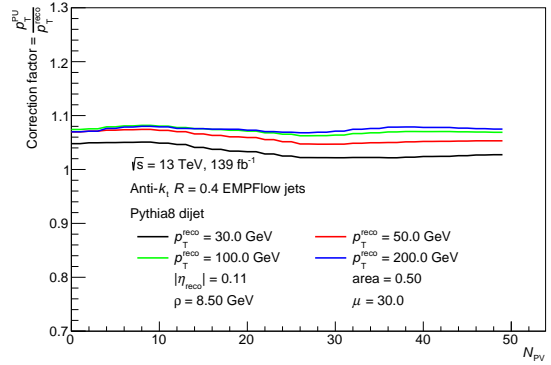
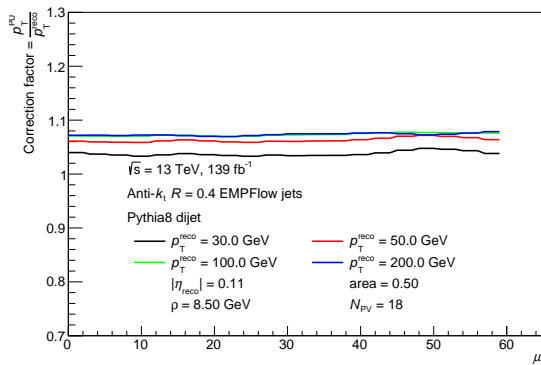
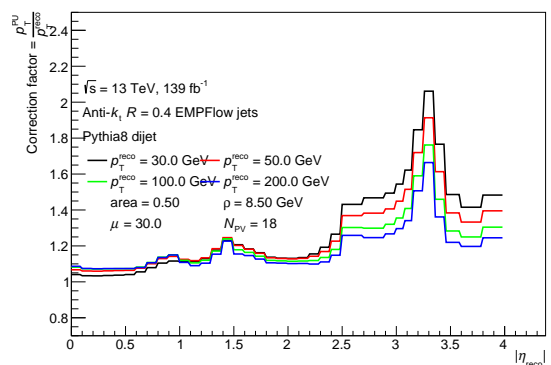
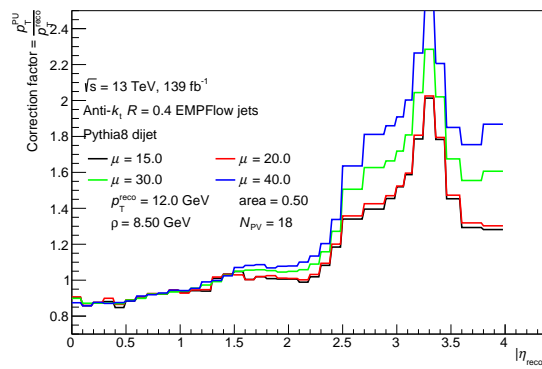


FIGURE 5.25 – Dépendance de  $p_T^{\text{PU}}$  en fonction de  $N_{\text{PV}}$ , tracée en fonction de  $|\eta_{\text{true}}|$ , après correction d'empilement et pour quatre différents intervalles en  $p_T^{\text{true}}$ .

Les Figures 5.26 et 5.27 représentent le facteur de correction ( $\frac{p_T^{\text{PU}}}{p_T^{\text{reco}}}$ ) qui doit être appliqué à l'énergie et la masse de chaque jet pour les CHS PFlow et les CS+SK PFlow respectivement. Sur la Figure 5.26a, le facteur de correction augmente avec  $p_T$ . Sur les Figures 5.26d et 5.26e, le facteur de correction diminue avec  $N_{\text{PV}}$  et augmente avec  $\mu$  respectivement, car  $N_{\text{PV}}$  permet de mesurer l'empilement dans le temps, qui se manifeste par des dépôts d'énergie positive, alors que  $\mu$  permet de mesurer l'empilement hors-temps, qui se manifeste par des dépôts d'énergie négative. Enfin, sur la Figure 5.26f les transitions du calorimètre à  $|\eta| \simeq 1,5$  et à  $|\eta| \simeq 3,2$  sont visibles. Le facteur de correction est très élevé dans la région à  $|\eta| \simeq 3,2$  car la dépendance de l'impulsion transverse en fonction de l'empilement y est très importante. Pour les CS+SK PFlow, les caractéristiques de ces courbes sont les mêmes, si ce n'est que le facteur de correction décroît en fonction de  $p_T$  à la place d'augmenter, ce qui indiquerait que les algorithmes *CS+SK* auraient tendance à sur-corriger l'empilement à basse impulsion transverse.


 (a) Facteur de correction en fonction de  $p_T$  pour différentes valeurs de  $|\eta|$ .

 (b) Facteur de correction en fonction de  $p_T$  pour différentes valeurs de  $\mu$ .

 (c) Facteur de correction en fonction de  $p_T$  pour différentes valeurs de  $N_{PV}$ .

 (d) Facteur de correction en fonction de  $N_{PV}$  pour différentes valeurs de  $p_T$ .

 (e) Facteur de correction en fonction de  $\mu$  pour différentes valeurs de  $p_T$ .

 (f) Facteur de correction en fonction de  $|\eta|$  pour différentes valeurs de  $p_T$ .

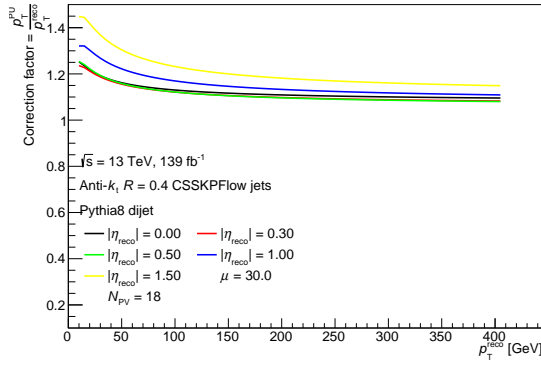




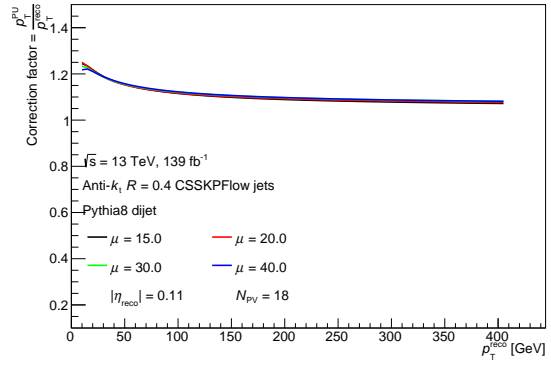
(g) Facteur de correction en fonction de  $|\eta|$  pour différentes valeurs de  $\mu$ .

FIGURE 5.26 – Facteur de correction en fonction de différentes variables pour les CHS PFlow jets.

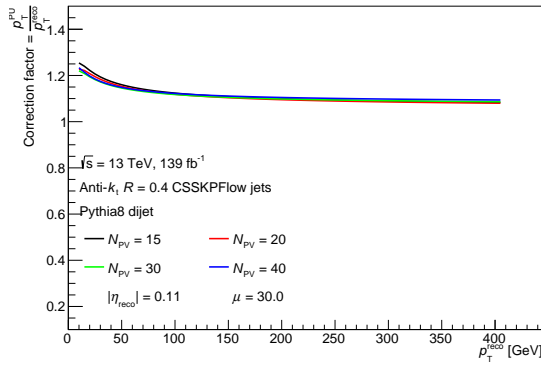
### 5.3. Performances et étalonnage des UFOs de rayon $R=0,4$



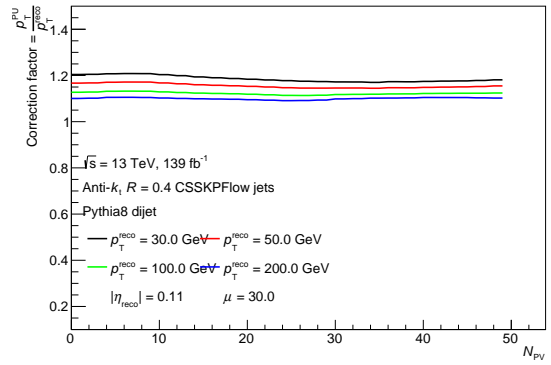
(a) Facteur de correction en fonction de  $p_T$  pour différentes valeurs de  $|\eta|$ .



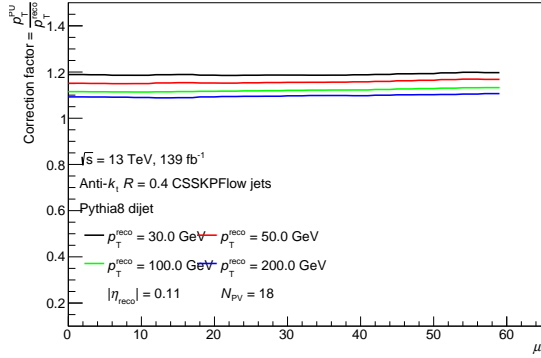
(b) Facteur de correction en fonction de  $p_T$  pour différentes valeurs de  $\mu$ .



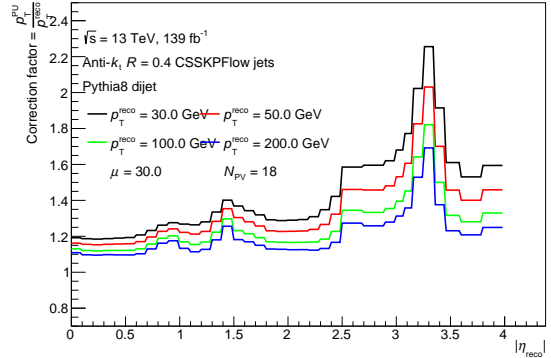
(c) Facteur de correction en fonction de  $p_T$  pour différentes valeurs de  $N_{PV}$ .



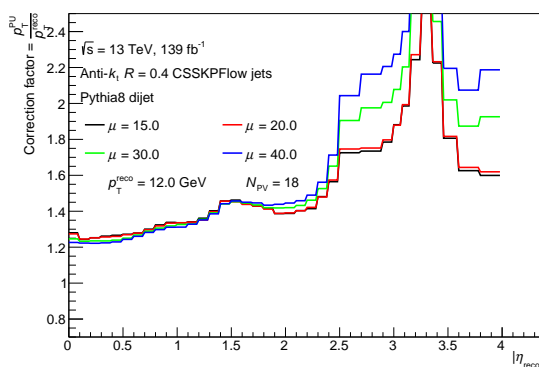
(d) Facteur de correction en fonction de  $N_{PV}$  pour différentes valeurs de  $p_T$ .



(e) Facteur de correction en fonction de  $\mu$  pour différentes valeurs de  $p_T$ .



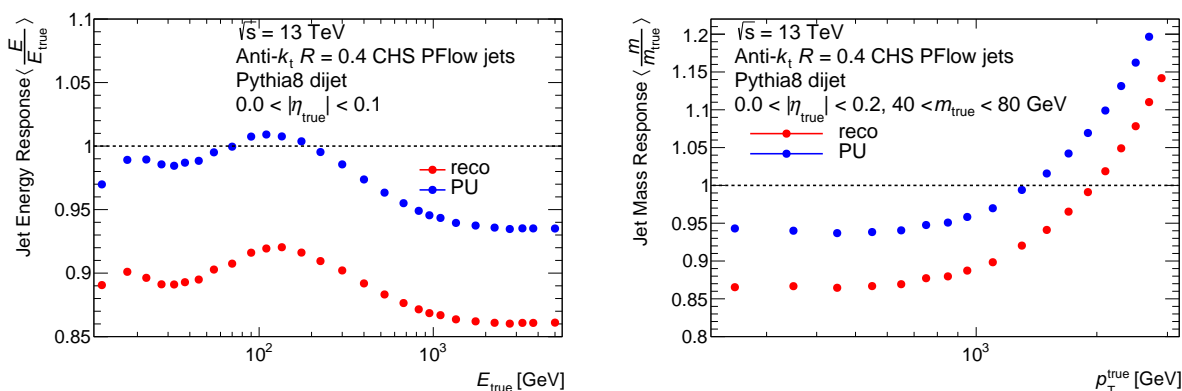
(f) Facteur de correction en fonction de  $|\eta|$  pour différentes valeurs de  $p_T$ .



(g) Facteur de correction en fonction de  $|\eta|$  pour différentes valeurs de  $\mu$ .

FIGURE 5.27 – Facteur de correction en fonction de différentes variables pour les CS+SK PFlow jets.

La Figure 5.28 représente l'échelle en énergie en fonction de  $E_{\text{true}}$  et de masse en fonction de  $p_{\text{T}}^{\text{true}}$  avant et après correction d'empilement pour les CHS PFlow. Sur la Figure 5.28a, l'échelle en énergie devient proche de 1 car la correction d'empilement résiduelle est dérivée par rapport  $p_{\text{T}}^{\text{true}}$  (voir l'Équation 5.3), et donc corrige déjà pour l'échelle en énergie (qui est l'objectif de l'étape de la correction absolue en énergie, masse et pseudo-rapidité). Cependant, cette correction n'est pas parfaite, et notamment à haute  $E_{\text{true}}$ , où  $\frac{E_{\text{PU}}}{E_{\text{true}}} < 1$ , car la correction résiduelle n'est dérivée qu'en dessous de  $p_{\text{T}}^{\text{true}} = 120$  GeV (bien qu'elle soit appliquée en fonction de  $p_{\text{T}}^{\text{area}}$ ). C'est pourquoi la correction absolue en énergie, masse et pseudo-rapidité est tout de même nécessaire.



(a) Échelle en énergie en fonction de  $E_{\text{true}}$ .

(b) Échelle de masse en fonction de  $p_{\text{T}}^{\text{true}}$ .

FIGURE 5.28 – Échelle en énergie et de masse avant et après correction d'empilement pour les CHS PFlow.

La Figure 5.29 représente la résolution en énergie en fonction de  $E_{\text{true}}$  et en masse en fonction de  $p_{\text{T}}^{\text{true}}$  avant et après correction d'empilement pour les CHS PFlow. Sur les deux Figures on peut remarquer que la résolution en énergie et en masse est dégradée par la correction d'empilement.

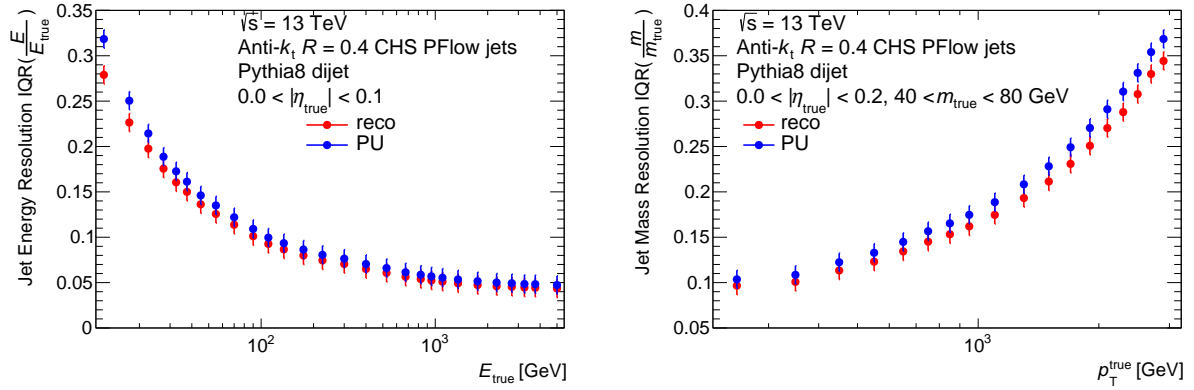

 (a) Résolution en énergie en fonction de  $E_{\text{true}}$ . (b) Résolution en masse en fonction de  $p_{\text{T}}^{\text{true}}$ .

FIGURE 5.29 – Résolution en énergie et en masse avant et après correction d’empilement pour les CHS PFlow.

La Figure 5.30 représente l’échelle en énergie après correction d’empilement pour les CHS PFlow, mais cette fois-ci en fonction de  $\eta_{\text{reco}}$  et ce pour plusieurs valeurs de l’énergie. Sur cette Figure aussi, on peut voir que la correction d’empilement corrige pour l’échelle en énergie, car l’échelle en énergie est proche de 1, bien que de façon imparfaite. Sur cette Figure, les transitions du calorimètre à  $|\eta| \simeq 1,5$  et à  $|\eta| \simeq 3,2$  sont visibles.

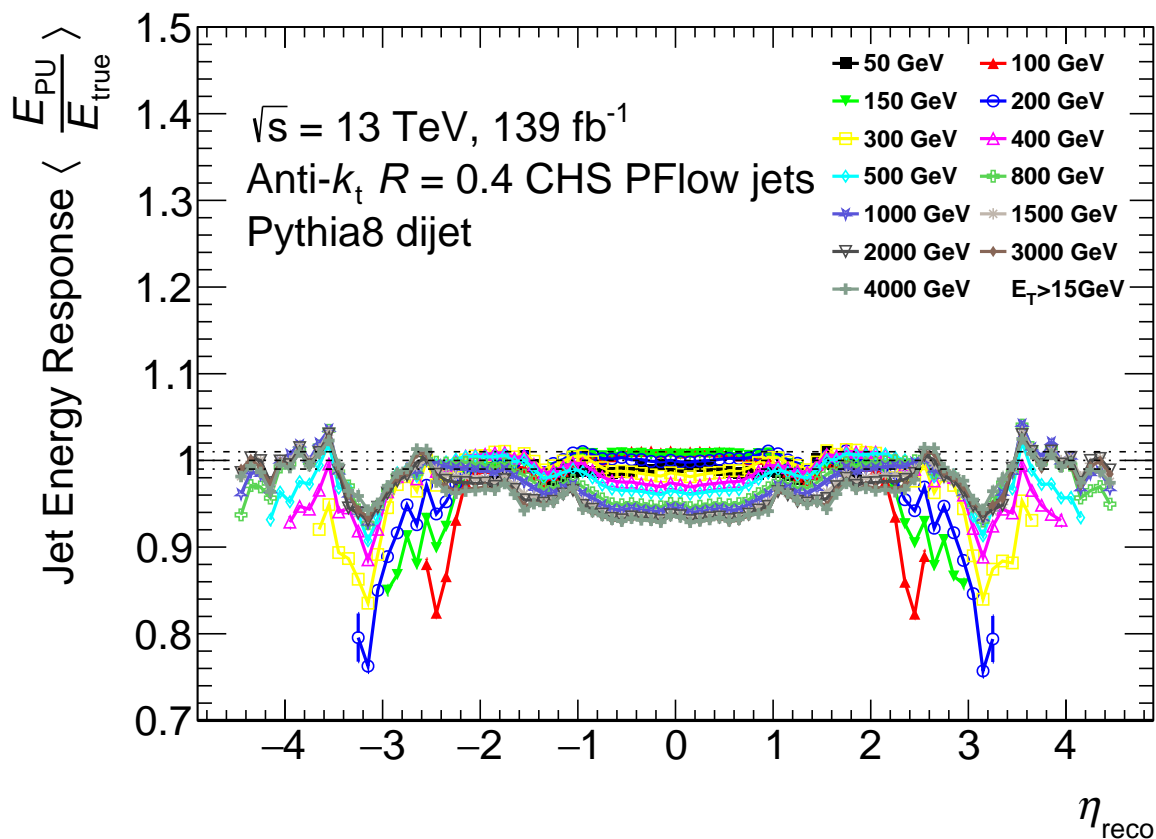


FIGURE 5.30 – Échelle en énergie en fonction de  $\eta_{reco}$  après correction d'empilement pour les CHS PFlow et pour plusieurs valeurs de  $E$ .

### 5.3.5 Résultats après correction absolue en énergie, masse et pseudo-rapidité

La Figure 5.31 représente l'échelle en énergie en fonction de  $\eta_{reco}$  après correction absolue en énergie, et ce pour plusieurs valeurs de l'énergie. L'échelle en énergie est approximativement 1 pour toutes les collections de jets, ce qui est l'objectif de la correction absolue en énergie. Cependant des déviations par rapport à 1 persistent, et ce en particulier à haute  $|\eta_{reco}|$  et basse  $E$ , ce qui indique que la méthode n'est pas parfaite dans ces régions de l'espace des phases, et en particulier que les ajustements de la réponse en énergie en fonction de l'énergie  $y$  est difficile à paramétrer. Cependant, il ne s'agit que de jets de basse énergie dans les régions avant du calorimètre (et donc de faible impulsion transverse), qui ne sont pas utilisés par les analyses. On peut remarquer que la méthode fonctionne légèrement moins bien dans la région avant pour les jets auxquels les algorithmes  $CS+SK$  sont appliqués, car les courbes de la réponse en fonction de l'énergie sont plus difficiles à ajuster.

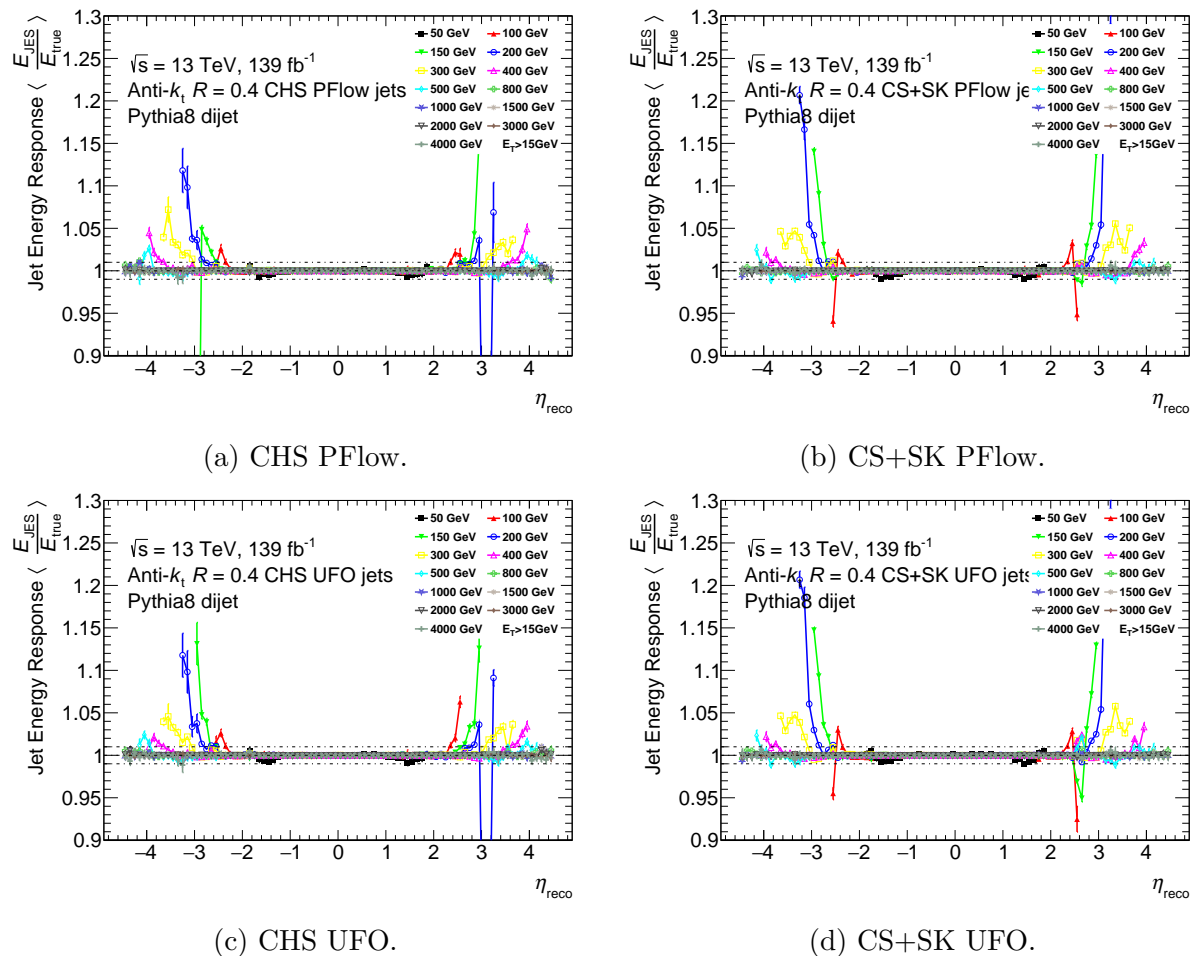
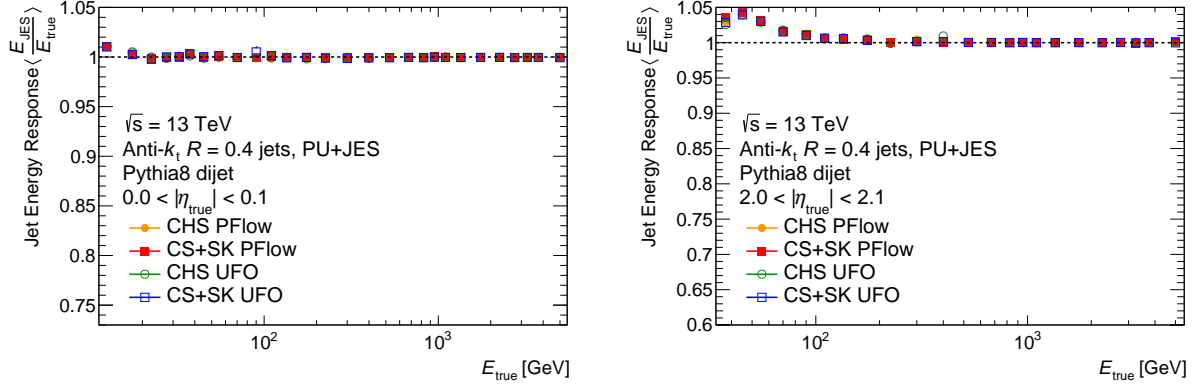


FIGURE 5.31 – Échelle en énergie en fonction de  $\eta_{\text{reco}}$  après correction absolue en énergie pour plusieurs valeurs de  $E$ .

La Figure 5.32 représente l'échelle en énergie en fonction de  $E_{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$  après correction absolue en énergie. L'échelle en énergie est environ la même pour les quatre collections de jets, ce qui indique que la méthode présente des performances similaires. La déviation par rapport à 1 à basse  $E_{\text{true}}$  est due au fait que la méthode utilisée n'est pas parfaite.


 (a)  $0, 0 < |\eta_{\text{true}}| < 0, 1$ .

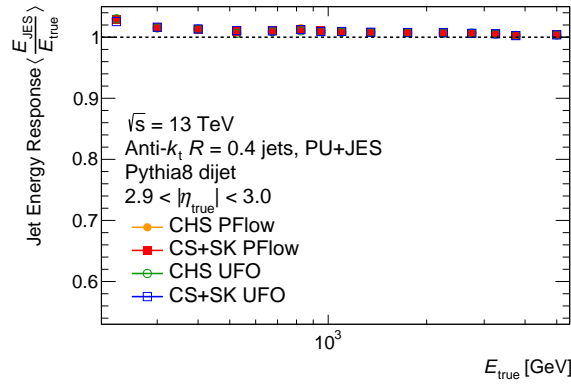
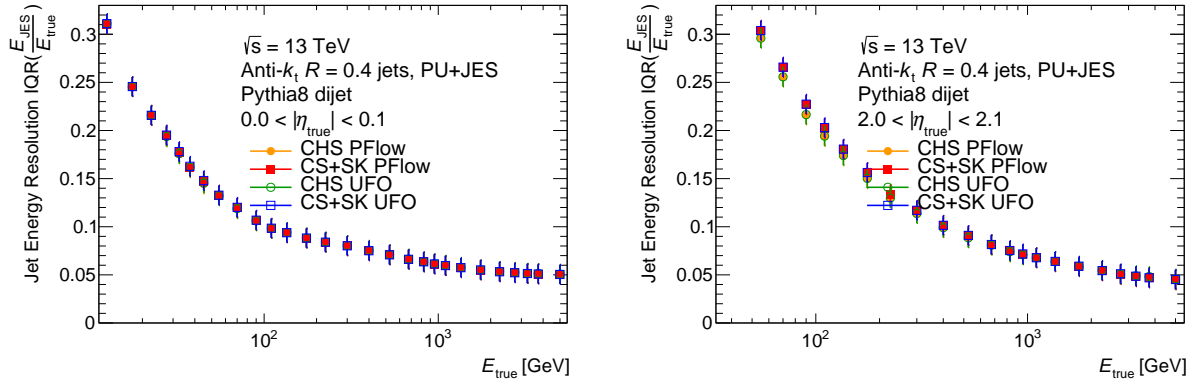
 (b)  $2, 0 < |\eta_{\text{true}}| < 2, 1$ .

 (c)  $2, 9 < |\eta_{\text{true}}| < 3, 0$ .

 FIGURE 5.32 – Échelle en énergie en fonction de  $E_{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$  après correction absolue en énergie.

La Figure 5.33 représente la résolution en énergie en fonction de  $E_{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$  après correction absolue en énergie. Après étalonnage, la résolution en énergie est désormais similaire pour les quatre collections de jets.


 (a)  $0,0 < |\eta_{\text{true}}| < 0,1$ .

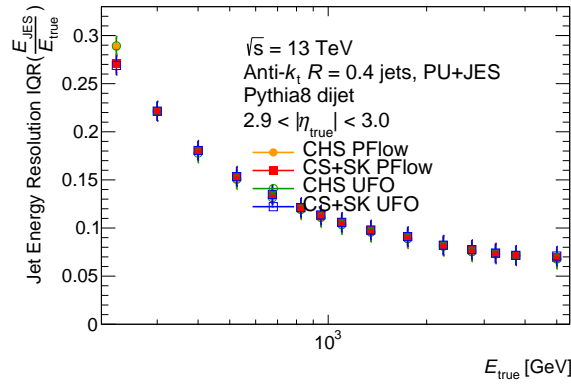
 (b)  $2,0 < |\eta_{\text{true}}| < 2,1$ .

 (c)  $2,9 < |\eta_{\text{true}}| < 3,0$ .

 FIGURE 5.33 – Résolution en énergie en fonction de  $E_{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$  après correction absolue en énergie.

La Figure 5.34 représente l'échelle en masse en fonction de  $p_{\text{T}}^{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$  et en  $m_{\text{true}}$  après correction absolue en énergie, masse et pseudo-rapidité. La méthode n'est pas parfaite, et en particulier à basse masse, où les meilleures performances sont obtenues pour les CS+SK UFO (car la courbe est la plus proche de 1).



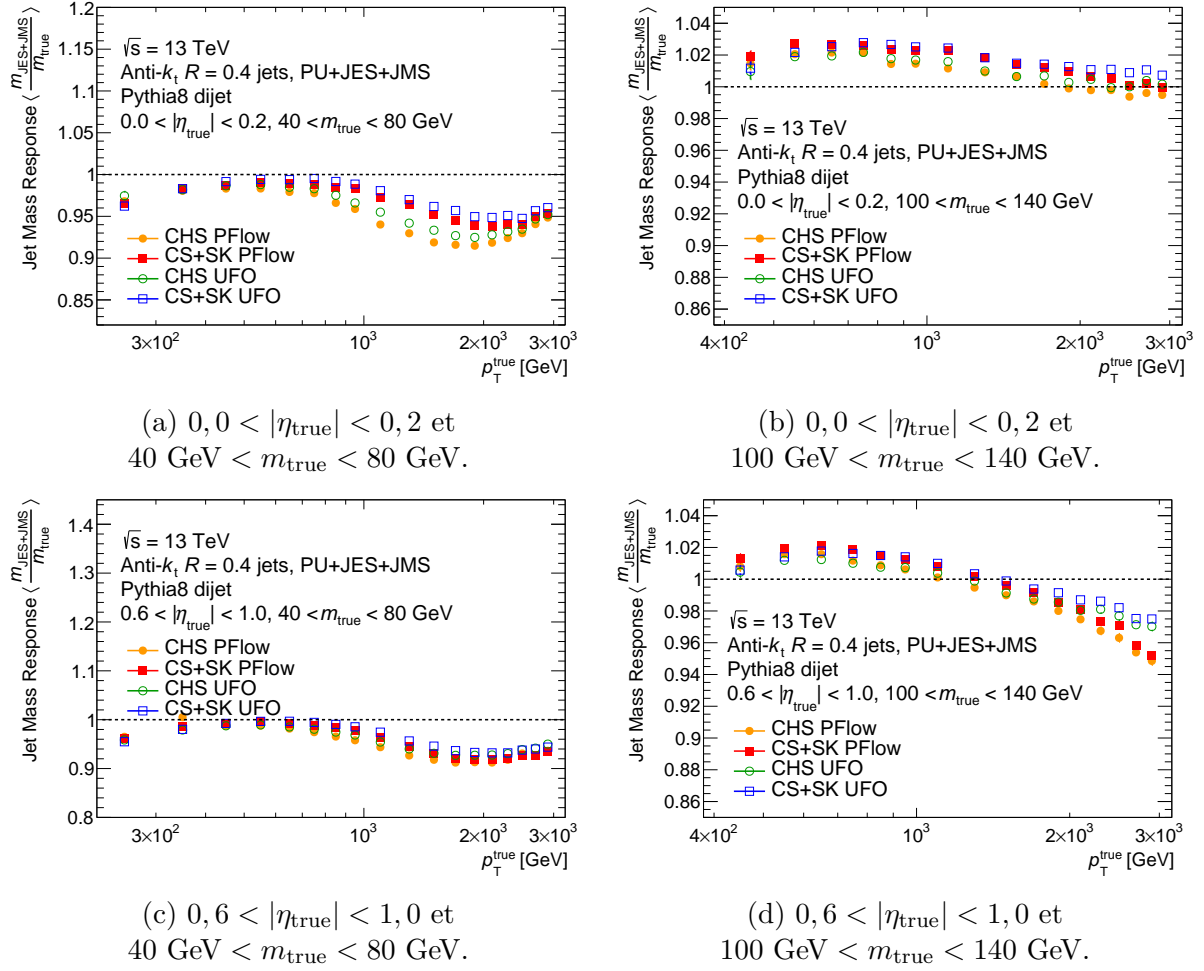
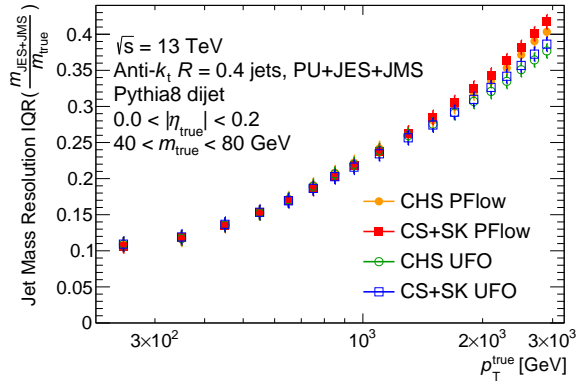
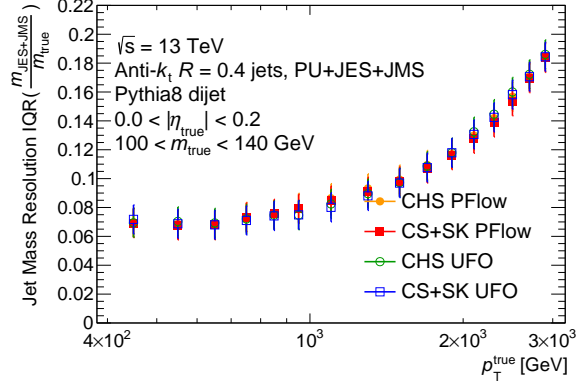


FIGURE 5.34 – Échelle en masse en fonction de  $p_T^{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$  et en  $m_{\text{true}}$  après correction absolue en énergie, masse et pseudo-rapidité.

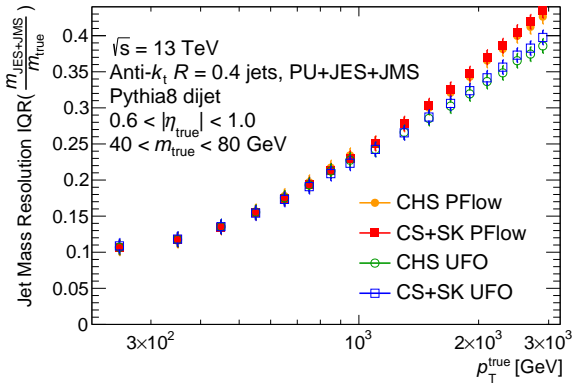
La Figure 5.35 représente la résolution en masse en fonction de  $p_T^{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$  et en  $m_{\text{true}}$  après correction absolue en énergie, masse et pseudo-rapidité. Les meilleures performances sont obtenues pour les UFOs à haute impulsion transverse.



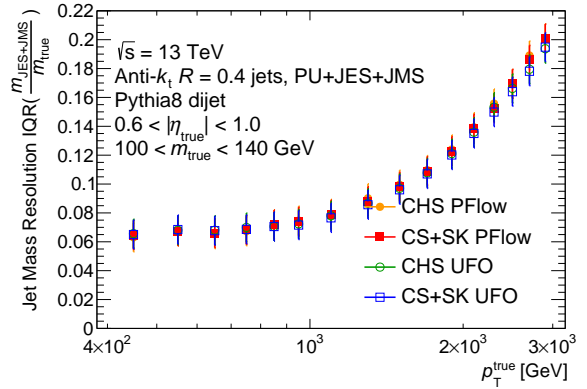
(a)  $0, 0 < |\eta_{\text{true}}| < 0, 2$  et  
 $40 \text{ GeV} < m_{\text{true}} < 80 \text{ GeV}$ .



(b)  $0, 0 < |\eta_{\text{true}}| < 0, 2$  et  
 $100 \text{ GeV} < m_{\text{true}} < 140 \text{ GeV}$ .



(c)  $0, 6 < |\eta_{\text{true}}| < 1, 0$  et  
 $40 \text{ GeV} < m_{\text{true}} < 80 \text{ GeV}$ .

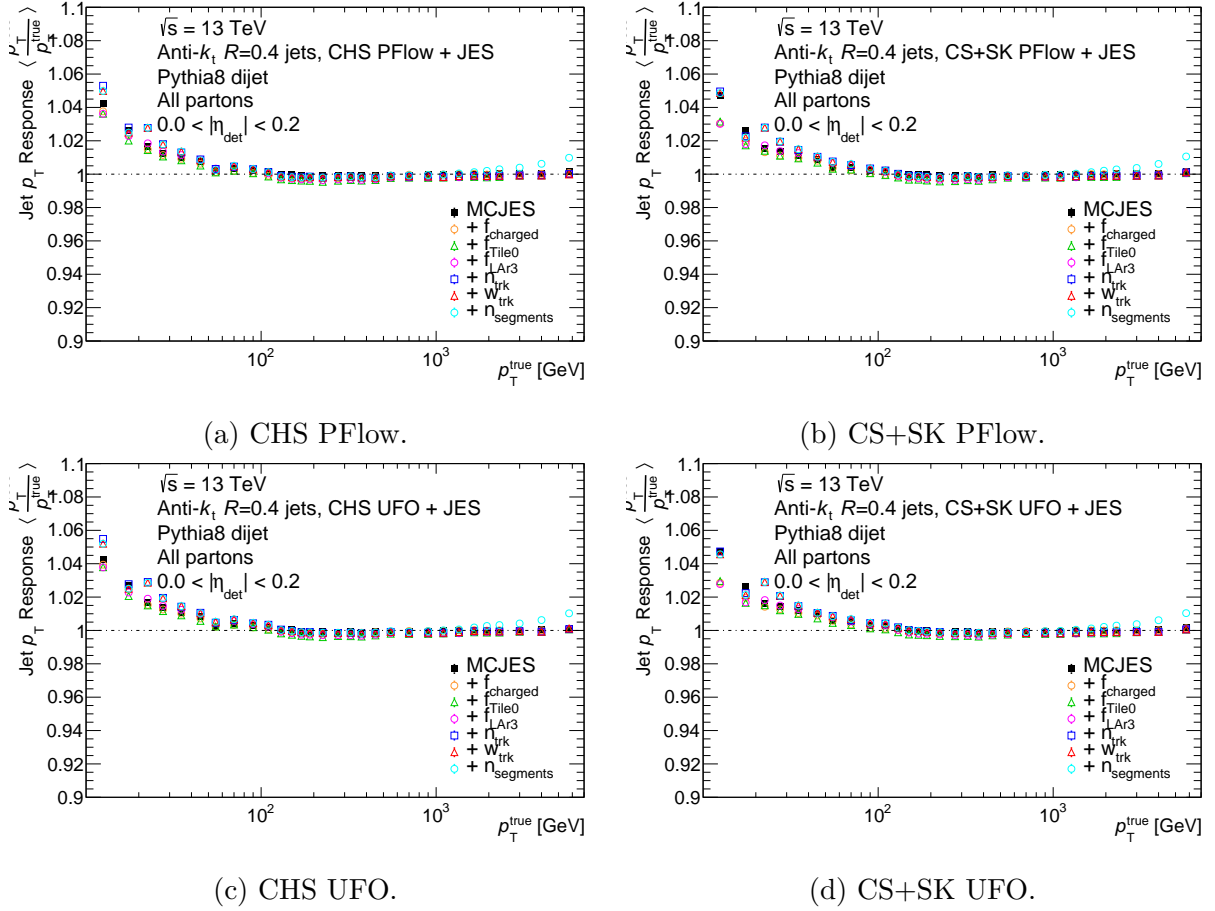


(d)  $0, 6 < |\eta_{\text{true}}| < 1, 0$  et  
 $100 \text{ GeV} < m_{\text{true}} < 140 \text{ GeV}$ .

FIGURE 5.35 – Résolution en masse en fonction de  $p_T^{\text{true}}$  dans différents intervalles en  $|\eta_{\text{true}}|$  et en  $m_{\text{true}}$  après correction absolue en énergie, masse et pseudo-rapidité.

### 5.3.6 Résultats après GSC

La Figure 5.36 représente l'échelle en impulsion transverse en fonction de  $p_T^{\text{true}}$  après chaque étape de GSC. On peut remarquer que l'échelle en impulsion transverse ne change pas de manière significative après chaque étape, ce qui est un des objectifs (secondaire) de la méthode. Les changements de l'échelle en impulsion transverse sont dus à des imperfections de la méthode utilisée, mais ne sont pas problématiques, puisque l'objectif premier de GSC est d'améliorer la résolution, et que l'échelle en énergie ne change pas de manière significative.


 FIGURE 5.36 – Échelle en  $p_T$  en fonction de  $p_T^{\text{true}}$  après chaque étape du GSC.

La Figure 5.37 représente la résolution en impulsion transverse en fonction de  $p_T^{\text{true}}$  après chaque étape de GSC. On peut remarquer que la résolution en impulsion transverse diminue après chaque étape (sauf à très basse impulsion transverse), et de manière similaire pour toutes les collections de jets.

### 5.3. Performances et étalonnage des UFOs de rayon $R=0,4$

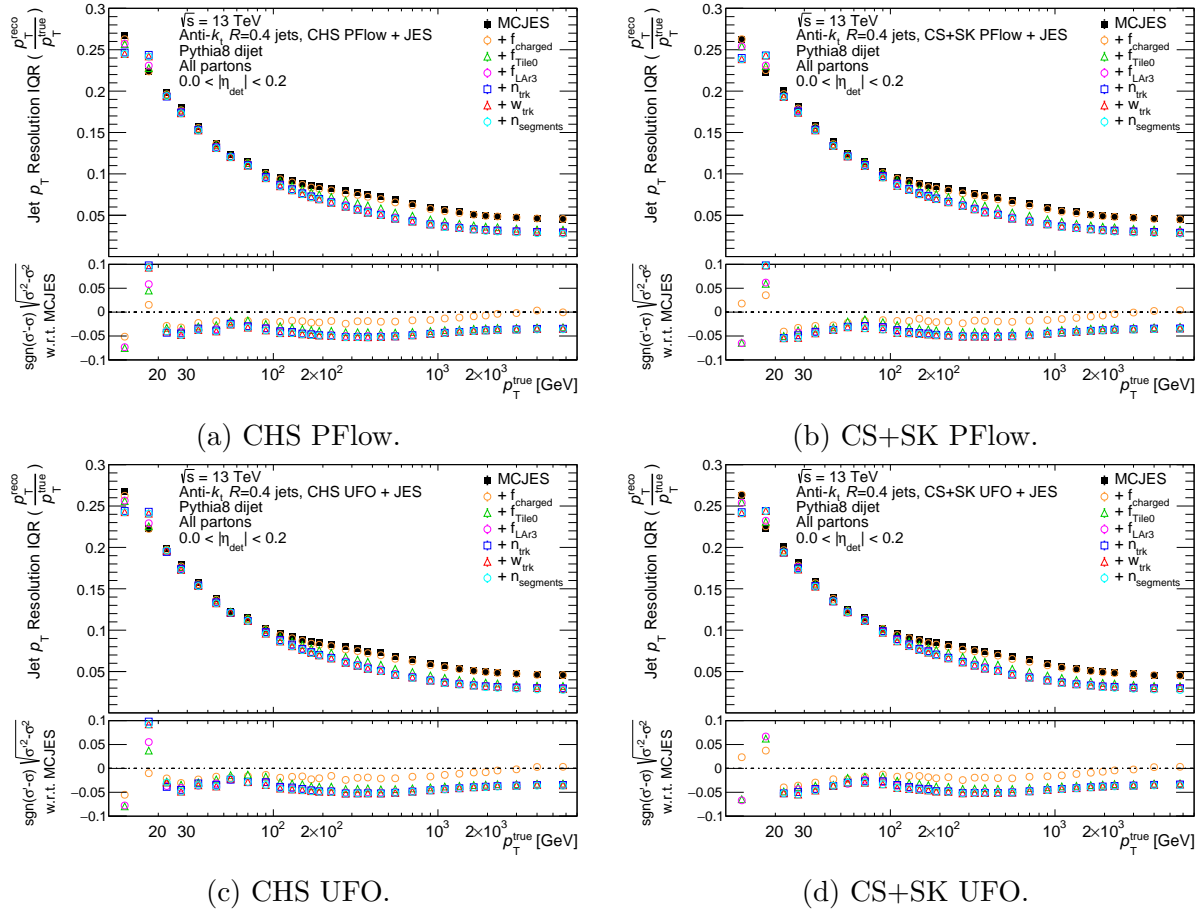


FIGURE 5.37 – Résolution en  $p_T$  en fonction de  $p_T^{\text{true}}$  après chaque étape de GSC.

La Figure 5.38 représente une comparaison pour toutes les collections de jets de la résolution en impulsion transverse en fonction de  $p_T^{\text{true}}$  après toutes les étapes de GSC. On peut remarquer que la résolution en impulsion transverse est similaire entre toutes les collections de jets.

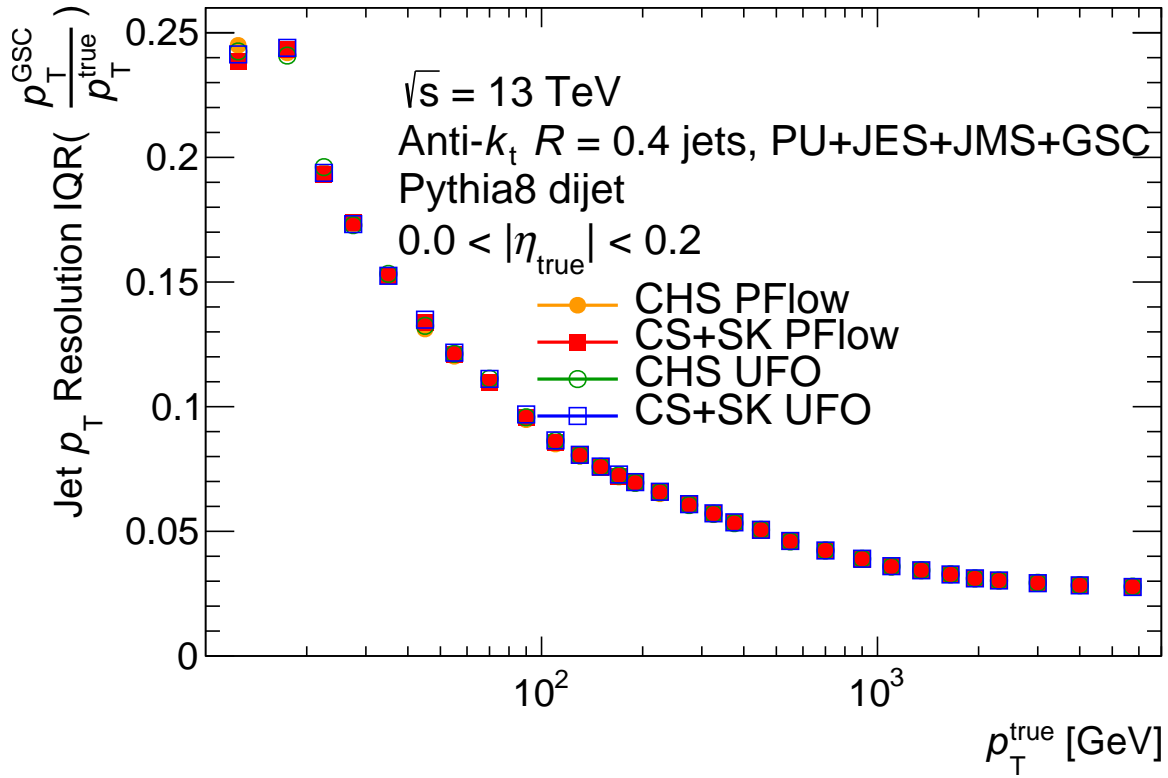


FIGURE 5.38 – Résolution en  $p_T$  en fonction de  $p_T^{\text{true}}$  après toutes les étapes de GSC.

La Figure 5.39 représente la différence de réponse en impulsion transverse entre les jets initiés par des quarks et les jets initiés par des gluons en fonction de  $p_T^{\text{true}}$  avant et après GSC. En comparant la Figure 5.39a avec la Figure 5.39b, on peut remarquer que la différence de réponse en impulsion transverse entre les quarks et les gluons diminue après GSC à haute impulsion transverse, ce qui est un des objectifs de la méthode. On peut aussi remarquer qu'à basse impulsion transverse, la différence en impulsion transverse entre les jets initiés par des quarks et des gluons est plus élevée pour les jets auxquels les algorithmes CS+SK sont appliqués, ce qui peut être expliqué par le fait que ces algorithmes enlèvent les constituants de basse impulsion, qui sont plus présents dans les jets initiés par des gluons, ce qui implique que ces algorithmes sont plus susceptibles à la différence entre quarks et gluons. C'est le deuxième point négatif apporté par l'utilisation des UFOs avec les algorithmes CS+SK appliqués aux constituants neutres, après la dégradation en résolution angulaire dans la gamme intermédiaire en impulsion transverse.

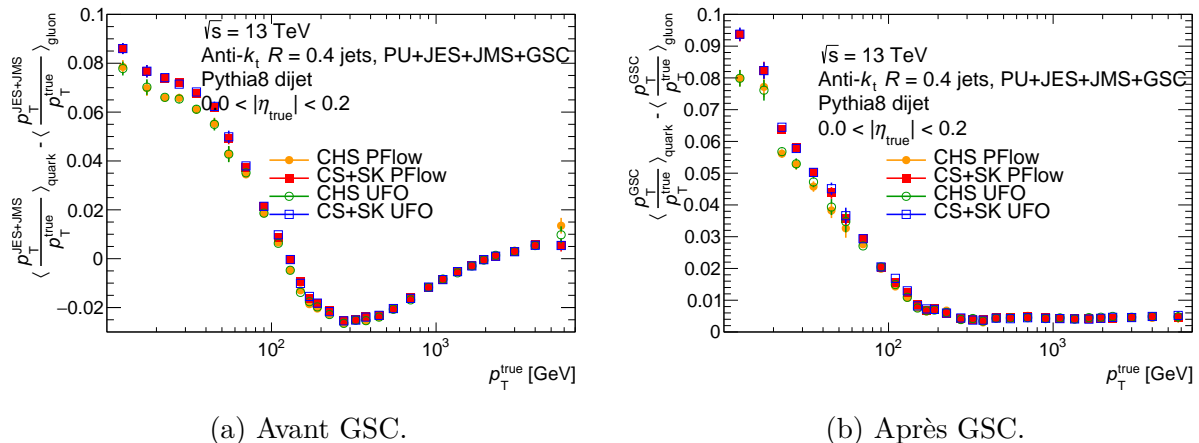


FIGURE 5.39 – Différence de réponse en impulsion transverse entre des jets initiés par des quarks et des jets initiés par des gluons en fonction de  $p_T^{\text{true}}$ .

## 5.4 Conclusion

En conclusion, j'ai dérivé toute la chaîne d'étalonnage impliquant seulement les simulations pour quatre collections de jets : les CHS PFlow, les CS+SK PFlow, les CHS UFO, et les CS+SK UFO. À chaque étape de la chaîne d'étalonnage, les performances respectives de ces quatre collections de jets ont été comparées. De manière générale, les différences entre les Particle Flow et les UFO, ainsi que celles entre les jets auxquels les algorithmes CS+SK sont appliqués aux constituants neutres ou non sont faibles. Cependant, quelques différences persistent tout de même :

- à basse impulsion transverse, les jets auxquels les algorithmes CS+SK sont appliqués présentent une plus faible dépendance en impulsion transverse par rapport aux variables d'empilement,
- la masse des UFO est plus facile à étalonner que celle des Particle Flow,
- la masse des jets auxquels les algorithmes CS+SK sont appliqués est plus facile à étalonner que celle des jets auxquels ces algorithmes ne sont pas appliqués,
- à haute impulsion transverse, la résolution en masse est meilleure pour les UFO que pour les Particle Flow,
- dans la gamme intermédiaire en impulsion transverse ( $p_T \simeq 100$  GeV), la résolution angulaire est dégradée par l'utilisation des UFO,
- à basse impulsion transverse, la différence en réponse en impulsion transverse entre des jets initiés par des quarks et des jets initiés par des gluons est plus élevée pour les jets auxquels les algorithmes CS+SK sont appliqués.

Pour les jets de rayon  $R = 1,0$  la définition des jets de référence désormais utilisée par ATLAS sont les UFO avec les algorithmes CS+SK appliqués aux constituants neutres avant reconstruction et avec l'algorithme Soft-drop appliqués aux constituants après que les jets ont été reconstruits. D'après mes études, pour les jets de rayon  $R = 0,4$ , utiliser les UFOs avec les algorithmes CS+SK permettrait d'en améliorer les performances, en particulier celles qui concernent la masse, ce qui en ferait potentiellement la définition de

jets de référence utilisée dans le futur.

Dans le cadre de mon travail sur les jets, j'ai aussi participé à des « recommandations de précision » pour les jets de rayon  $R = 0,4$ , en dérivant l'étape de calibration GSC uniquement. Un article résumant ces recommandations de précision est en cours d'approbation dans ATLAS. Ce travail n'est pas décrit plus en détail ici, car il est similaire à celui de la Section 5.3.6. Lors de ce projet, j'ai aussi travaillé sur le fait d'essayer d'inclure une nouvelle étape dans le GSC, prenant en compte la fraction d'énergie déposée dans le trou à  $|\eta| \simeq 1,4$  dans les calorimètres (la variable associée est nommée *TileGap3*, voir la Figure 3.9b). Cependant, l'étalonnage pour cette variable était difficile à dériver.

# Chapitre 6

## La recherche d'un secteur sombre

### Sommaire du Chapitre

---

<b>6.1</b>	<b>Description des modèles étudiés</b>	<b>124</b>
<b>6.2</b>	<b>Échantillons utilisés</b>	<b>129</b>
6.2.1	Données réelles	129
6.2.2	Simulations du bruit de fond	130
6.2.3	Simulations des signaux	130
<b>6.3</b>	<b>Reconstruction des jets</b>	<b>132</b>
<b>6.4</b>	<b>Pré-sélections</b>	<b>135</b>
6.4.1	Nettoyage des évènements	135
6.4.2	Sélection du vertex primaire	136
6.4.3	Système de déclenchement	136
6.4.4	Nettoyage des jets	137
6.4.5	Au niveau de l'analyse	137
<b>6.5</b>	<b>Stratégie de l'analyse</b>	<b>138</b>
<b>6.6</b>	<b>Optimisation des coupures de sélection</b>	<b>141</b>
<b>6.7</b>	<b>Comparaison des données et des simulations</b>	<b>156</b>
<b>6.8</b>	<b>Effet de <math>n_{\text{trk}}^{\varepsilon}</math> dans les données</b>	<b>159</b>
<b>6.9</b>	<b>Incertitudes systématiques</b>	<b>161</b>
6.9.1	Incertitudes expérimentales	162
6.9.2	Incertitudes théoriques	173
6.9.3	Incertitudes sur les signaux	178
6.9.4	Incertitudes sur le bruit de fond	178
<b>6.10</b>	<b>Interpolation des signaux</b>	<b>181</b>
6.10.1	Ajustement de la section efficace	182
6.10.2	Ajustement de l'acceptance $\times$ l'efficacité	182
6.10.3	Nombre d'évènements dans la Région de Signal	183
6.10.4	Interpolation de la forme des signaux	184
<b>6.11</b>	<b>Interprétation statistique</b>	<b>186</b>
6.11.1	Statistiques	186
6.11.2	Tests de faux signal	193
6.11.3	Tests d'injection de signal	198



6.11.4 Classement des paramètres de nuisance . . . . .	201
6.11.5 Limites . . . . .	204
<b>6.12 Conclusion . . . . .</b>	<b>204</b>

Ce Chapitre représente mon travail sur l'analyse pour la recherche d'un secteur sombre. J'ai contribué à toutes les étapes de cette analyse, sauf la génération des signaux : productions des fichiers, évaluation des incertitudes systématiques, optimisation des coupures de sélection, estimation du bruit de fond, interprétation statistique. Lors de ma dernière année de thèse, j'ai été nommé contact pour cette analyse, ce qui signifie devoir gérer l'analyse : organiser les réunions, mener les discussions, écrire une documentation interne à ATLAS, vérifier que les échantillons utilisés par l'analyse ne soient pas supprimés etc.

Au moment où j'écris cette thèse, l'analyse est en cours d'approbation dans ATLAS. Par conséquent, ce manuscrit ne pourra pas inclure les données dans la région de signal.

## 6.1 Description des modèles étudiés

Les modèles étudiés correspondent à ceux décrits dans la Section 2.5, et en particulier à ceux dans lesquels la plupart des mésons sombres se désintègrent rapidement en particules du Modèle Standard, produisant des jets qui ressemblent à ceux du Modèle Standard.

La désintégration des hadrons sombres dépend de leur spin, de leur masse, et du médiateur vers le secteur visible, ici considéré être un  $Z'$ . Afin de couvrir la diversité des modèles de QCD sombre, plusieurs modèles présentant différents spectres, échelles de confinement et modes de désintégration sont étudiés. Ces modèles, nommés Modèles  $A$ ,  $B$ ,  $C$  et  $D$ , sont décrits dans la Table 6.1 [20].

Modèle	$N_d$	$N_f$	$\Lambda_d$ (GeV)	$\tilde{m}_{q'}$ (GeV)	$m_{\pi_d}$ (GeV)	$m_{\rho_d}$ (GeV)	$\rho_d$ : Modes de Désintégration	$\pi_d$ : Modes de Désintégration
$A$	3	2	15	20	10	50	$\rho_d \rightarrow \pi_d \pi_d$	$\pi_d \rightarrow c\bar{c}$
$B$	3	6	2	2	2	4,67	$\rho_d \rightarrow \pi_d \pi_d$	$\pi_d \rightarrow s\bar{s}$
$C$	3	2	15	20	10	50	$\rho_d \rightarrow \pi_d \pi_d$	$\pi_d \rightarrow \gamma' \gamma'$ avec $m_{\gamma'} = 4,0$ GeV
$D$	3	6	2	2	2	4,67	$\rho_d \rightarrow \pi_d \pi_d$	$\pi_d \rightarrow \gamma' \gamma'$ avec $m_{\gamma'} = 0,7$ GeV

TABLE 6.1 – Modèles de référence considérés. Tous les mésons sombres sont supposés se désintégrer rapidement. Dans ce contexte,  $N_d$  est la dimension du nouveau groupe de symétrie  $SU(N_d)$ ,  $N_f$  est le nombre de saveurs de quark sombre,  $\Lambda_d$  est l'échelle de confinement sombre,  $\tilde{m}_{q'}$  est la masse constitutive du quark sombre, et les deux masses des mésons sombres sont données par  $m_{\pi_d}$  et  $m_{\rho_d}$  [20].

Les rapports d'embranchement du photon sombre  $\gamma'$  pour les Modèles C et D sont

présentés dans la Table 6.2 [79].

Modèle C	Modèle D
$\gamma' \rightarrow u\bar{u} : 22\%$	
$\gamma' \rightarrow c\bar{c} : 22\%$	
$\gamma' \rightarrow e^-e^+ : 17\%$	$\gamma' \rightarrow \pi^-\pi^+ : 70\%$
$\gamma' \rightarrow \mu^-\mu^+ : 17\%$	$\gamma' \rightarrow e^-e^+ : 15\%$
$\gamma' \rightarrow \tau^-\tau^+ : 10\%$	$\gamma' \rightarrow \mu^-\mu^+ : 15\%$
$\gamma' \rightarrow d\bar{d} : 6\%$	
$\gamma' \rightarrow s\bar{s} : 6\%$	

TABLE 6.2 – Rapports d'embranchement du photon sombre  $\gamma'$  pour les Modèles C et D [79].

Les paramètres sous-jacents du secteur sombre affectent la phénoménologie d'un jet sombre au sein des collisionneurs. La constante de couplage du secteur sombre,  $\alpha_d(\mu)$ , est contrôlée par l'équation suivante (équivalente à l'Équation 1.5) :

$$\frac{d}{d \ln \mu^2} \alpha_d^{-1}(\mu) = \frac{1}{12\pi} (11N_d - 2N_f) \quad (6.1)$$

Une comparaison entre la constante de couplage de la QCD du Modèle Standard et celle des divers modèles de QCD sombre est représentée dans la Figure 6.1.

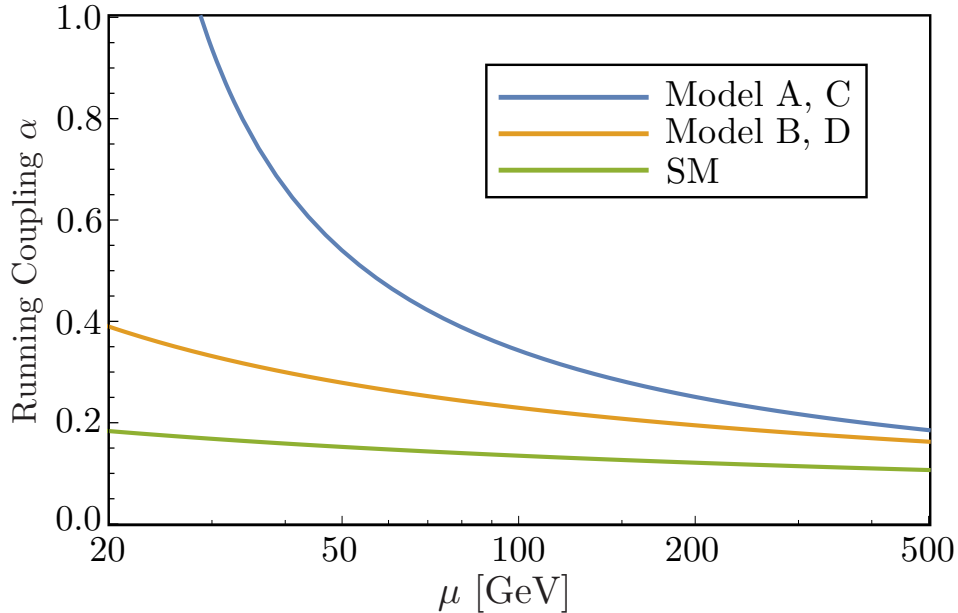


FIGURE 6.1 – Constante de couplage dans le secteur de QCD sombre et dans celui du Modèle Standard. Les définitions des modèles A, B, C et D peuvent être trouvées dans la Table 6.1 [20].

L'objectif de cette analyse est de rechercher une résonance dans la masse invariante des deux jets de plus haute impulsion (dénotée  $m_{JJ}$  par la suite) par rapport au bruit

de fond du Modèle Standard, c'est-à-dire un excès d'évènements par rapport au nombre d'évènements attendus par le Modèle Standard. Cette résonance se situerait à la masse du  $Z'$ , dénotée  $M_{Z'}$ . Cette procédure est schématisée dans la Figure 6.2.

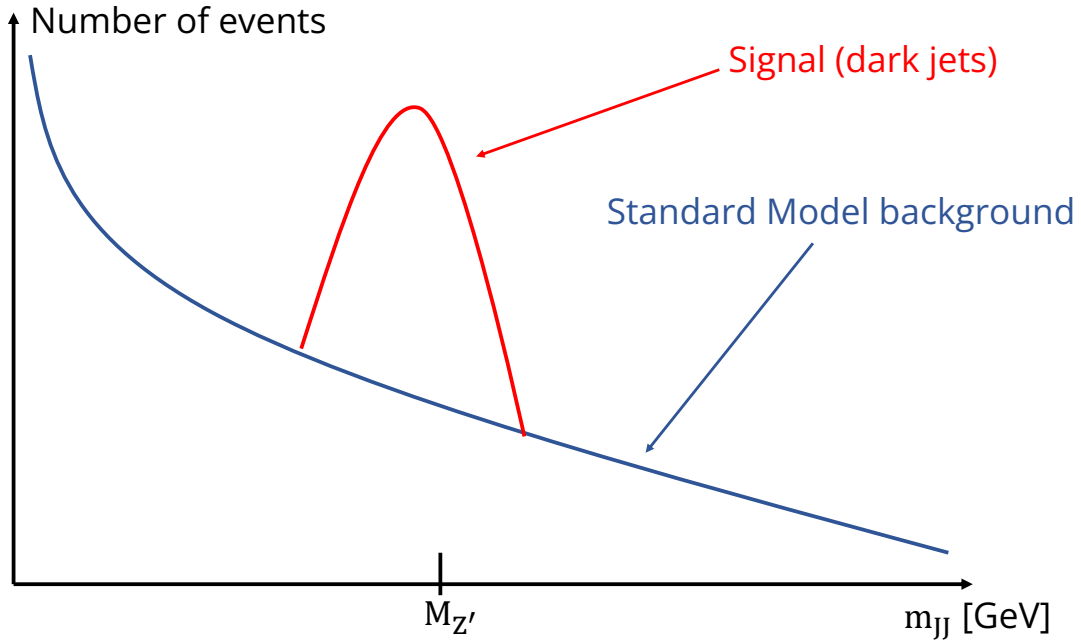


FIGURE 6.2 – Schéma de la résonance en masse invariante des deux jets,  $m_{JJ}$ , qui correspondrait au  $Z'$ , par rapport au bruit de fond du Modèle Standard.

Pour ce faire, l'analyse a pour objectif de trouver des variables qui permettent de discriminer les évènements de signal (c'est-à-dire de QCD sombre) de ceux du bruit de fond. Ensuite, des coupures sur ces variables sont appliquées, ce qui définit une région de Signal (*Signal Region* en anglais, ou plus brièvement « *SR* »), où le nombre d'évènements de signal est significatif par rapport au nombre d'évènements de bruit de fond. Cette procédure est schématisée dans la Figure 6.3, et les variables pouvant être utilisées pour définir la Région de Signal, ainsi que leur motivation, sont introduites ci-dessous.

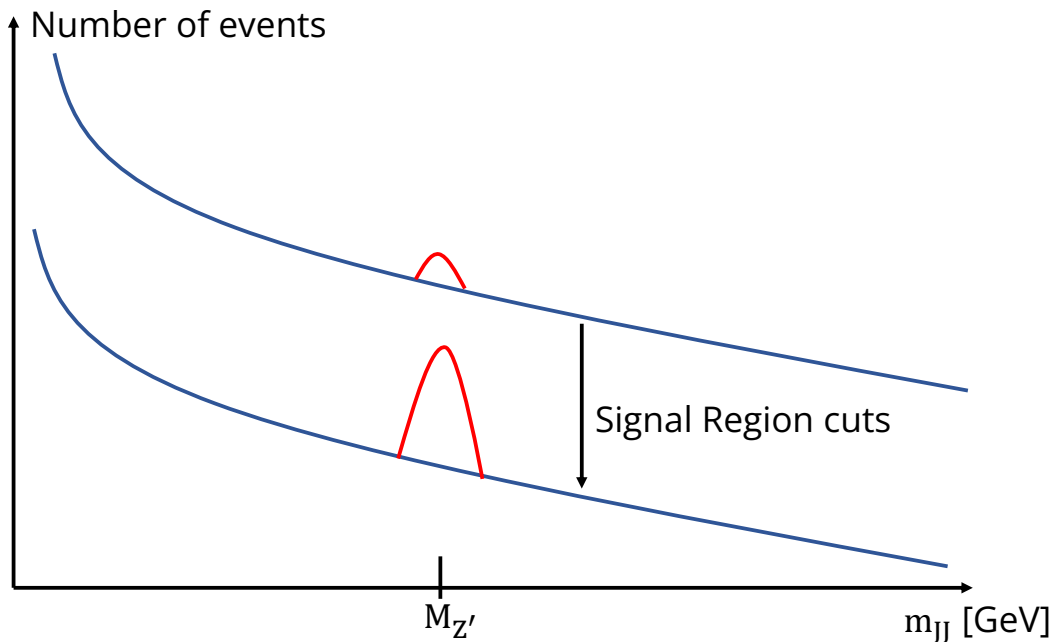


FIGURE 6.3 – Schéma de la Région de Signal où le nombre d'évènements de signal par rapport au nombre d'évènements de bruit de fond est plus important que dans la région sans coupure (« *inclusive* »).

La constante de couplage détermine l'intensité de la gerbe de partons. Ainsi, de la même façon que les jets initiés par des gluons sont plus larges et contiennent plus de particules que les jets initiés par des quarks (la constante de couplage de l'interaction forte est la même entre les jets initiés par des quarks ou des gluons, cette différence est due aux facteurs  $C_A$  et  $C_F$ , comme décrit dans la Section 5.2.3), les jets de QCD sombre sont plus larges (ce qui est aussi dû à la « double hadronisation » : une dans le secteur sombre, puis une dans le Modèle Standard) et contiennent plus de particules que les jets de QCD du Modèle Standard. Ainsi, des variables telles que le nombre de traces associées à un jet ou la corpulence d'un jet (*girth* en anglais), dont la définition est donnée par :

$$\text{girth} = \sum_{i \in J} \frac{p_{T,i}}{p_T^{\text{jet}}} \Delta R_{i,\text{jet}} \quad (6.2)$$

où la somme  $i$  parcourt les constituants du jet  $J$  et  $\Delta R_{i,\text{jet}}$  représente la distance angulaire entre le constituant  $i$  et l'axe du jet  $J$ , s'avèrent être des variables qui permettent de discriminer les jets de QCD sombre de ceux du Modèle Standard.

D'autres variables permettant de discriminer les jets de QCD sombre de ceux du Modèle Standard sont introduites dans [20], parmi lesquelles la fraction d'énergie du jet déposée dans le calorimètre électromagnétique :

$$E - \text{ratio} = \frac{\text{Energy deposited in the electromagnetic calorimeter}}{E_{\text{jet}}} \quad (6.3)$$

où  $E_{\text{jet}}$  représente l'énergie totale du jet, et la fonction de corrélation d'énergie à deux points [53] :

$$C_1^\beta = \sum_{i < j \in J} z_i z_j (\Delta R_{i,j})^\beta \quad (6.4)$$

où la somme  $i$  parcourt les constituants du jet  $J$ ,  $z_i = \frac{p_{T,i}}{p_T^{\text{jet}}}$  et  $\Delta R_{i,j}$  représente la distance angulaire entre les composants  $i$  et  $j$ . La valeur  $\beta = 0, 2$  est utilisée dans cette section, car cette valeur permet de sonder plus efficacement les radiations collinéaires à petite échelle, qui s'avèrent utiles pour discriminer les jets initiés par des quarks des jets initiés par des gluons.

La Figure 6.4 représente une comparaison pour les quatre variables définies précédemment entre des jets issus des différents modèles étudiés et ceux du Modèle Standard [20]. Sur la Figure 6.4a, on peut remarquer que le nombre de traces est une variable particulièrement discriminante pour les Modèles A et D, alors que sur la Figure 6.4b on peut remarquer que la corpuence est une variable particulièrement discriminante pour les Modèles A et C. Sur la Figure 6.4c, on peut remarquer que la fraction d'énergie d'un jet déposée dans le calorimètre électromagnétique est une variable particulièrement discriminante pour le Modèle B. Ceci est dû au fait que dans le Modèle B, les mésons sombres se désintègrent principalement en une paire de quarks étranges. La plus grande partie de l'énergie du jet est donc transportée par des kaons à longue durée de vie qui vont se désintégrer principalement dans le calorimètre hadronique. Ainsi, pour le modèle B, la fraction d'énergie déposée dans le calorimètre électromagnétique est plus faible que par rapport aux jets de QCD du Modèle Standard. Enfin, sur la Figure 6.4d, on peut remarquer que  $C_1^\beta$  est une variable particulièrement discriminante pour les Modèles A et C.

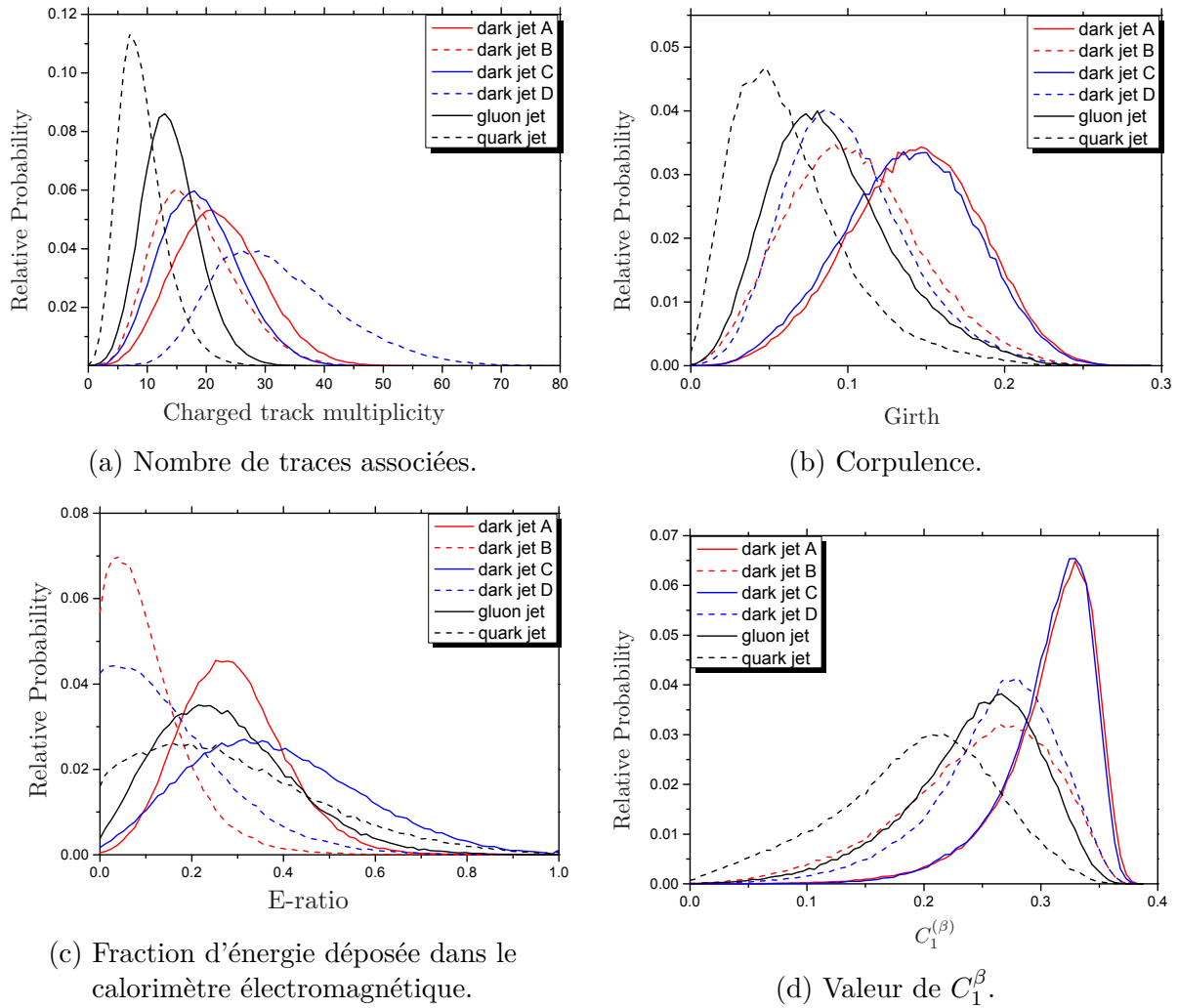


FIGURE 6.4 – Comparaison de différentes variables discriminantes entre des jets issus des différents modèles étudiés et ceux du Modèle Standard [20].

Ces études phénoménologiques, tirées de [20], seront revisitées dans le cadre de l'analyse dans la Section 6.6.

## 6.2 Échantillons utilisés

De manière générale, les analyses utilisent trois types de données : les données réelles acquises grâce au détecteur ATLAS lors des collisions, des simulations du bruit de fond et des simulations des signaux étudiés. Ces trois types de données sont décrites ci-dessous.

### 6.2.1 Données réelles

Cette analyse utilise les données collectées en 2015, 2016, 2017 et 2018 produites par des collisions protons-protons à  $\sqrt{s} = 13$  TeV au LHC. Les événements pris dans de bonnes conditions du détecteur sont sélectionnés grâce aux Good Run Lists (voir la Section 3.4).

Cela correspond à  $3,2 \text{ fb}^{-1}$  en 2015,  $33,0 \text{ fb}^{-1}$  en 2016,  $44,3 \text{ fb}^{-1}$  en 2017, et  $58,5 \text{ fb}^{-1}$  en 2018, pour un total de  $139 \text{ fb}^{-1}$ .

## 6.2.2 Simulations du bruit de fond

Les simulations de bruit de fond utilisées sont des simulations purement de QCD avec deux partons générés dans l'état final, telles que décrites dans la Section 5.3.1. Ainsi, le générateur utilisé est PYTHIA [41], qui est un générateur au premier ordre en théorie des perturbations, et qui est le générateur utilisé par défaut pour les simulations de QCD pure dans ATLAS. Les valeurs de la section pour chacune des 13 tranches en impulsion transverse peuvent être trouvées dans la Table 5.1.

D'autres processus, tels que des productions de deux bosons via le Modèle Standard, des bosons  $W/Z$  avec des jets, des quarks  $t\bar{t}$ , ne sont pas utilisés, étant donné que leur contribution est négligeable en comparaison avec les processus de QCD avec deux partons. En effet, par exemple, la section efficace du processus  $t\bar{t}$  est égale à  $7,30 \times 10^5 \text{ fb}$ , ce qui est négligeable comparé aux valeurs de la Table 5.1.

## 6.2.3 Simulations des signaux

Les signaux sont générés grâce au module *Hidden Valley* [80] de PYTHIA [41] via le processus  $q\bar{q} \rightarrow Z' \rightarrow q_d\bar{q}_d$ . Puisque le  $Z'$  doit pouvoir être produit par des quarks du Modèle Standard, il doit aussi pouvoir se désintégrer en quarks du Modèle Standard. Ces deux processus sont représentés dans la Figure 6.5, qui a été produite avec TikZ-Feynman [81].

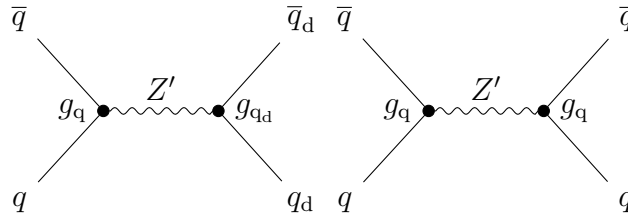


FIGURE 6.5 – Modes de désintégration du  $Z'$ . Cette Figure a été produite avec TikZ-Feynman [81].

Dans ces Figures,  $g_q$  représente la constante de couplage du  $Z'$  aux quarks du Modèle Standard et  $g_{q_d}$  représente la constante de couplage du  $Z'$  aux quarks sombre.

La section efficace du premier processus est proportionnelle à [4] :

$$\sigma_{q\bar{q} \rightarrow Z' \rightarrow q_d\bar{q}_d} \sim B_{Z' \rightarrow q\bar{q}} B_{Z' \rightarrow q_d\bar{q}_d} \frac{(\frac{\Gamma}{2})^2}{(\sqrt{s} - M_{Z'})^2 + (\frac{\Gamma}{2})^2} \quad (6.5)$$

où :

- $B_{Z' \rightarrow q\bar{q}}$  représente le rapport d'embranchement du  $Z'$  en quarks du Modèle Standard,
- $B_{Z' \rightarrow q\bar{a}q\bar{a}}$  représente le rapport d'embranchement du  $Z'$  en quarks sombres,
- $\Gamma$  représente la largeur de désintégration totale du  $Z'$ ,
- $\sqrt{s}$  représente l'énergie dans le centre de masse,
- $M_{Z'}$  représente la masse du  $Z'$ .

La section efficace du second processus est proportionnelle à :

$$\sigma_{q\bar{q} \rightarrow Z' \rightarrow q\bar{q}} \sim B_{Z' \rightarrow q\bar{q}}^2 \frac{(\frac{\Gamma}{2})^2}{(\sqrt{s} - M_{Z'})^2 + (\frac{\Gamma}{2})^2} \quad (6.6)$$

La largeur de désintégration pour le  $Z'$  est artificiellement petite ( $\Gamma = 0,1$  GeV), afin de le produire sur la couche de masse seulement. Le rapport d'embranchement du  $Z'$  en quarks du Modèle Standard est choisi artificiellement petit afin que le  $Z'$  se désintègre essentiellement en quarks sombres, mais aussi afin que la section efficace pour le processus  $q\bar{q} \rightarrow Z' \rightarrow q\bar{q}$  reste faible :  $B_{Z' \rightarrow q\bar{q}} = 0,0001$ . Par conséquent, le rapport d'embranchement du  $Z'$  en quarks sombres est égal à 0,9999.

La collection de *PDFs NNPDF2.3LO* [73] et la collection de paramètres *A14* [74] sont utilisées. Les signaux sont générés pour les quatre modèles étudiés. Pour chacun des modèles étudiés, sept signaux, correspondant à sept différentes masses du  $Z'$ ,  $M_{Z'}$ , sont générés : 1,5 TeV, 1,75 TeV, 2 TeV, 2,25 TeV, 2,5 TeV, 3 TeV et 3,5 TeV.  $M_{Z'} = 3,5$  TeV est la plus haute masse que nous avons pu générer, car le logiciel de reconstruction d'ATLAS ne pouvait supporter le grand nombre de traces au-delà.

Pour chacun des signaux, trois collections, MC16a, d et e, correspondant respectivement aux années 2015+2016, 2017 et 2018, sont utilisées. 20 000 évènements sont générés pour MC16a, 30 000 évènements pour MC16d et 40 000 pour MC16e.

La section efficace, ainsi que l'efficacité du filtre du générateur (voir la Section 3.5) pour chacun des signaux sont représentées dans la Table 6.3. On peut remarquer que la section efficace des signaux est très faible, mais la théorie contient un paramètre libre pour le couplage du  $Z'$  aux quarks ( $g_q$ ), comme discuté auparavant. Sa normalisation est donc arbitraire. De plus, la section efficace est extrêmement similaires entre les signaux des différents modèles ayant la même masse. Ceci est dû au fait que les différents modèles étudiés ne diffèrent que par les modes de désintégration des particules sombres.



Modèle	$M_{Z'}$ [GeV]	Section efficace [fb]	Efficacité du filtre du générateur
A	1500	$2,84 \times 10^{-4}$	0,771
	1750	$1,15 \times 10^{-4}$	0,835
	2000	$5,04 \times 10^{-5}$	0,876
	2250	$2,35 \times 10^{-5}$	0,905
	2500	$1,15 \times 10^{-5}$	0,923
	3000	$3,04 \times 10^{-6}$	0,940
	3500	$8,85 \times 10^{-7}$	0,942
B	1500	$2,86 \times 10^{-4}$	0,860
	1750	$1,15 \times 10^{-4}$	0,898
	2000	$5,01 \times 10^{-5}$	0,925
	2250	$2,35 \times 10^{-5}$	0,938
	2500	$1,15 \times 10^{-5}$	0,950
	3000	$3,05 \times 10^{-6}$	0,959
	3500	$8,87 \times 10^{-7}$	0,954
C	1500	$2,83 \times 10^{-4}$	0,651
	1750	$1,15 \times 10^{-4}$	0,750
	2000	$5,04 \times 10^{-5}$	0,810
	2250	$2,35 \times 10^{-5}$	0,821
	2500	$1,14 \times 10^{-5}$	0,879
	3000	$3,03 \times 10^{-6}$	0,911
	3500	$8,88 \times 10^{-7}$	0,921
D	1500	$2,84 \times 10^{-4}$	0,801
	1750	$1,15 \times 10^{-4}$	0,856
	2000	$5,01 \times 10^{-5}$	0,890
	2250	$2,34 \times 10^{-5}$	0,914
	2500	$1,14 \times 10^{-5}$	0,931
	3000	$3,03 \times 10^{-6}$	0,945
	3500	$8,86 \times 10^{-7}$	0,945

TABLE 6.3 – Section efficace et efficacité du filtre du générateur pour chacun des signaux.

### 6.3 Reconstruction des jets

Dans cette Section, toutes les études sont effectuées à partir des particules *vraies*, c'est-à-dire à partir des particules simulées au niveau du générateur.

Comme décrit dans la Section 6.1, les jets de QCD sombre sont plus larges que les jets de QCD du Modèle Standard. ATLAS utilise des jets de rayon  $R = 0,4$  pour les jets de QCD du Modèle Standard, car ce rayon est suffisant pour contenir la plupart des radiations. La Figure 6.6 représente la distribution de la masse invariante entre les deux jets *vrais* de plus haute impulsion transverse (*leading jets* en anglais) pour les différents modèles étudiés et avec une masse pour le  $Z'$ ,  $M_{Z'} = 3,5$  TeV. Plusieurs techniques de

reconstruction de jets sont étudiées : des jets reconstruits avec un rayon  $R = 0,4$ ,  $R = 1,0$  (l'autre rayon utilisé par ATLAS),  $R = 1,0$  avec l'algorithme de *trimming* (voir la Section 4.3.5.1) appliqué aux constituants après que le jet est formé,  $R = 1,0$  avec l'algorithme de *Soft-drop* (voir la Section 4.3.5.2), et enfin des jets de rayon  $R = 1,0$  dont l'algorithme de regroupement est appliqué sur des jets de rayon  $R = 0,4$  formés à partir des particules vraies, à la place des particules vraies directement (ces jets sont appelés *re-clustered jets* (ou plus brièvement *RC jets*) en anglais [78], et sont appelés « R=1.0 RC from R=0.4 » sur la Figure 6.6). Plusieurs algorithmes de *grooming* sont étudiés, car, de par leur taille, les jets de rayon  $R = 1,0$  contiennent plus d'empilement que les jets de rayon  $R = 0,4$ , et par conséquent des algorithmes pour enlever cet empilement sont nécessaires pour ces jets, en particulier pour étudier les variables de sous-structure de ces jets, qui y sont très sensibles. Les re-clustered jets sont intéressants à étudier, car les algorithmes de grooming ne sont pas nécessairement appliqués. Le choix de  $M_{Z'} = 3,5$  TeV permet de mieux visualiser l'impact des différentes techniques de reconstruction de jets qu'à basse masse pour le  $Z'$ , car les différences y sont plus importantes.

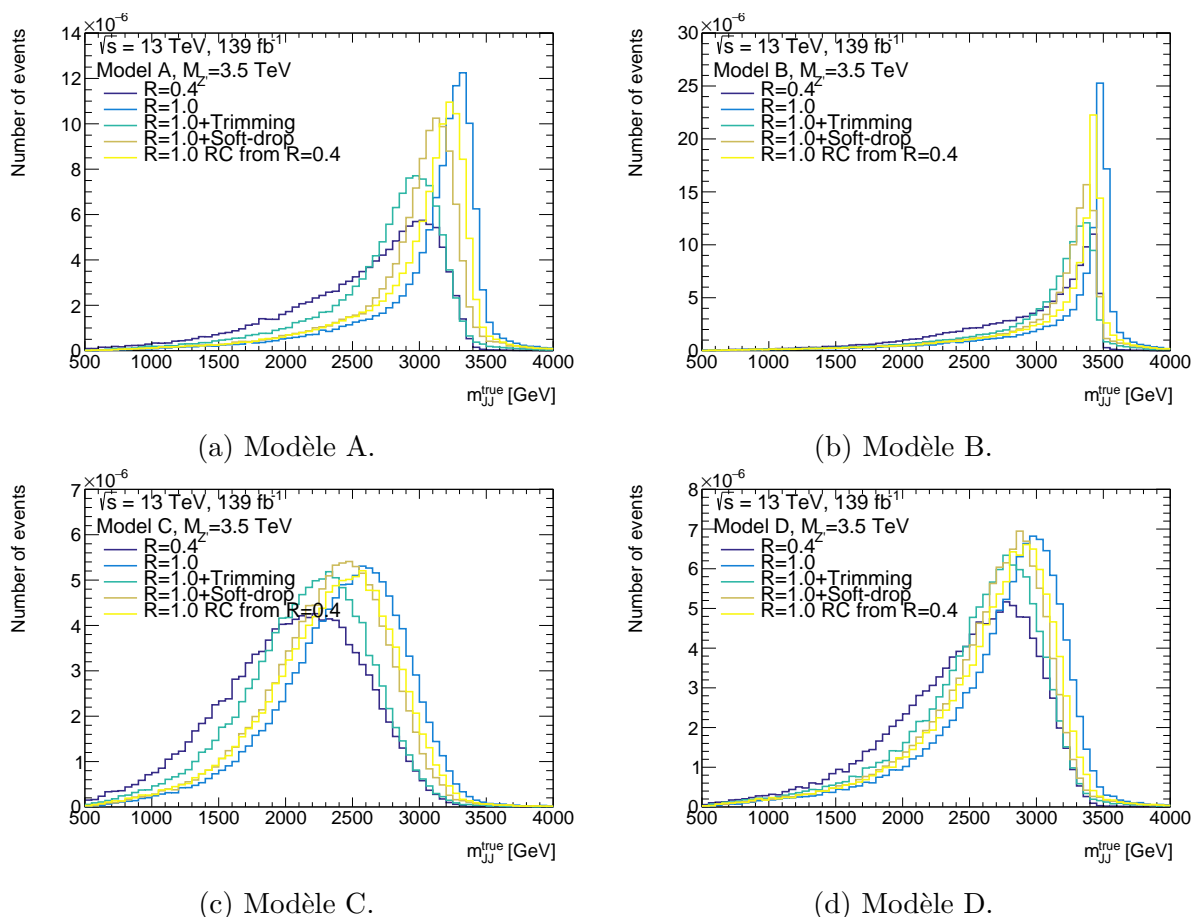


FIGURE 6.6 – Comparaison de différentes techniques de reconstruction de jets appliquées sur les deux jets vrais de plus haute impulsion des quatre différents modèles étudiés, avec une masse du  $Z'$ ,  $M_{Z'} = 3,5$  TeV.

Sur les Figures 6.6a et 6.6b, on peut remarquer que la distribution pour des jets de rayon  $R = 0,4$  ne pique pas à 3500 GeV (la masse du  $Z'$ ). Ceci est dû au fait que les jets de rayon  $R = 0,4$  ne sont pas suffisant pour contenir toutes les radiations provenant

des jets de QCD sombre. Au contraire pour des jets de rayon  $R = 1,0$  sans grooming, la distribution pique à la masse du  $Z'$ . Cependant, les jets de rayon  $R = 1,0$  sans grooming ne sont pas recommandés par ATLAS, c'est-à-dire, qu'il n'y a pas d'étalonnage disponible pour les jets de rayon  $R = 1,0$  sans grooming, et donc qu'ils ne peuvent pas être utilisés au niveau des analyses. En effet, les jets de rayon  $R = 1,0$  sont utilisés pour reconstruire des objets boostés, qui peuvent être identifiés grâce à leur sous-structure. Or cette sous-structure est très sensible à l'empilement. L'objectif des algorithmes de grooming est d'enlever cet empilement afin d'améliorer la résolution de ces variables. C'est pourquoi les jets de rayon  $R = 1,0$  sans grooming ne sont pas recommandés dans ATLAS. Sur les Figures 6.6a et 6.6b, les jets de rayon  $R = 1,0$  avec trimming présentent de meilleures performances que les jets de rayon  $R = 0,4$ , mais la dégradation, notamment en résolution, par rapport aux jets de rayon  $R = 1,0$  sans trimming est non négligeable. Ceci est dû au fait que les jets de QCD sombre contiennent un large nombre de constituants de faible impulsion, qui sont enlevés du jet par l'algorithme. Les jets de rayon  $R = 1,0$  avec Soft-drop présentent de meilleures performances que les jets de rayon  $R = 1,0$  avec trimming. Enfin, les re-clustered jets semblent montrer des performances prometteuses, puisque ce sont les jets qui se rapprochent le plus des jets de rayon  $R = 1,0$  sans grooming.

Sur les Figures 6.6c et 6.6d, une autre remarque peut être émise : aucune des distributions ne pique à la masse du  $Z'$  (même pas les jets de rayon  $R = 1,0$  sans grooming). Ceci est dû au fait que les photons sombres  $\gamma'$  peuvent se désintégrer en muons de façon non négligeable (voir la Table 6.2). Or les muons ne sont pas inclus dans les jets *vrais* (voir la Section 4.3.1.4), car leur dépôt d'énergie dans les calorimètres est négligeable.

Lorsque cette analyse a débuté, la seule collection de jets de rayon  $R = 1,0$  recommandée par ATLAS était les jets de rayon  $R = 1,0$  reconstruits à partir des constituants étalonnés avec la technique *LCW* (voir la Section 4.3.1.1) et avec le trimming appliqué aux constituants. Étant donné que les jets de rayon  $R = 1,0$  avec trimming présentaient de meilleures performances que les jets de rayon  $R = 0,4$  (pour lesquels il n'y a pas de grooming), nous avons décidé d'utiliser cette collection de jets.

Cependant, plus récemment, comme décrit dans la Section 5.1, ATLAS recommande désormais un autre type de jets de rayon  $R = 1,0$  : les jets de rayon  $R = 1,0$  reconstruits à partir des Unified Flow Objects avec les algorithmes *CS+SK* et Soft-drop appliqués. L'analyse était alors trop avancée pour changer de collection de jets. De plus, l'étalonnage *in-situ* (voir la Section 5.2.4) n'est toujours pas disponible au moment d'écrire cette thèse. Qui plus est, les performances avec l'algorithme *CS+SK* sont encore à démontrer, car cet algorithme peut aussi enlever les constituants de basse impulsion. Enfin, la résolution en impulsion transverse est plus élevée pour les jets reconstruits à partir des UFOs, que pour les jets reconstruits à partir des *LCTopos* (voir la Figure 5.5).

Enfin, les re-clustered jets semblent montrer des performances prometteuses lorsqu'ils sont appliqués aux particules vraies. La Figure 6.7 représente une comparaison de la masse invariante reconstruite des deux jets de plus haute impulsion transverse pour des *LCTopo* jets de rayon  $R = 1,0$  avec trimming et des re-clustered formés de jets de rayon  $R = 0,4$  formés à partir des *CHS PFlow*, et ce pour les Modèles A et D avec une masse pour le  $Z'$ ,

$M_{Z'} = 3,5$  TeV. Les jets reconstruits sont étalonnés avec une méthode similaire à celle de la Section 5.2.2, et les coupures de sélection décrites dans la Section 6.4 sont appliquées. Sur cette Figure, les re-clustered montrent de meilleures performances que les LCTopo jets de rayon  $R = 1,0$  avec trimming. En particulier, pour le Modèle D, la distribution des re-clustered pique à la masse du  $Z'$ , ce qui n'était pas le cas au niveau *vrai* sur la Figure 6.6d. Ceci est dû au fait que ces jets sont formés à partir des Particle Flows, qui utilisent les traces, et donc celles des muons. La distribution n'aurait pas piqué à la masse du  $Z'$  si les jets étaient formés à partir des EMTopos.

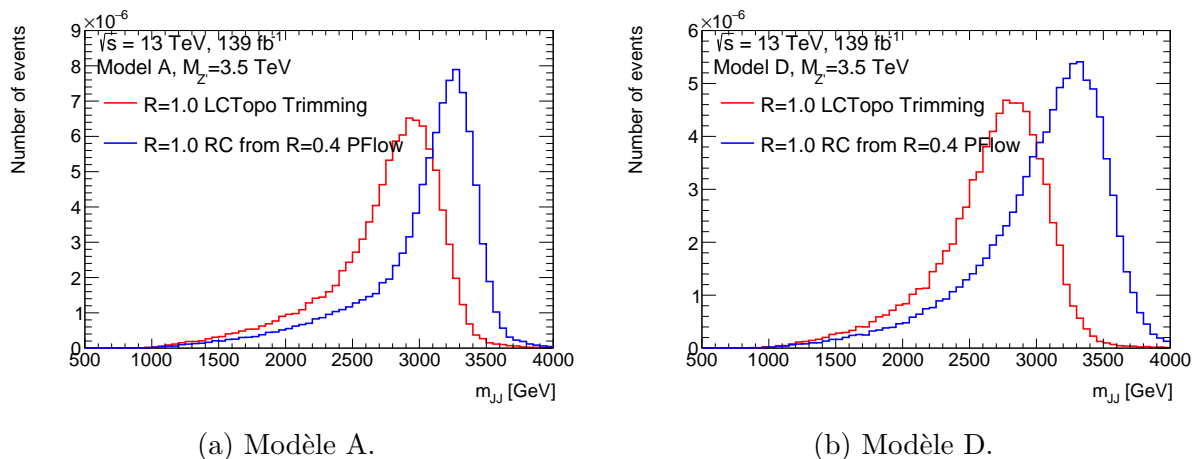


FIGURE 6.7 – Comparaison de la masse invariante reconstruite des deux jets de plus haute impulsion transverse pour des LCTopo jets de rayon  $R = 1,0$  avec trimming et des re-clustered jets formés de jets de rayon  $R = 0,4$  formés à partir des CHS PFlow.

Cependant, à cause de la coupure en masse au niveau des pré-sélections (voir Section 6.4), l'amélioration en rapport signal sur bruit de fond apportée par les re-clustered jets devient moins importante. C'est pourquoi nous avons décidé d'utiliser tout de même les LCTopo jets de rayon  $R = 1,0$  avec trimming.

## 6.4 Pré-sélections

Plusieurs niveaux de sélections sont appliqués : au niveau du nettoyage des évènements, de la sélection du vertex primaire, du système de déclenchement, du nettoyage des jets, et enfin de l'analyse. Tout ces niveaux sont décrits plus en détail ci-dessous.

### 6.4.1 Nettoyage des évènements

Les évènements qualifiés de « *mauvais* », c'est-à-dire avec un problème dans le calorimètre à argon liquide, ou dans le calorimètre en tuiles scintillatrices, ou dans le SCT ne sont pas utilisés. Ces problèmes peuvent par exemple venir d'un problème dans le système à haute tension ou encore de bruits soudains dans le détecteur. Les évènements incomplets, c'est-à-dire avec de l'information manquante dans un détecteur, ne sont pas non plus utilisés. Cette procédure s'appelle « *event cleaning* » en anglais. L'inefficacité de cette coupure est négligeable.

## 6.4.2 Sélection du vertex primaire

Les évènements doivent avoir un vertex primaire avec lequel sont associées au moins deux traces. Le vertex primaire est sélectionné comme étant celui avec la plus haute somme des impulsions transverses au carré des traces qui lui sont associées :  $\sum_{\text{tracks}} (p_T^{\text{track}})^2$ .

## 6.4.3 Système de déclenchement

Pour les années 2017 et 2018, les systèmes de déclenchement non pré-échelonnés (voir la Section 3.3) avec le seuil de sélection le plus bas sont utilisés. Pour toutes les années, des systèmes de déclenchement avec des conditions sur un seul jet sont utilisés [82] (ces systèmes de déclenchement sont appelés des *single jet triggers* en anglais). Pour les années 2015 et 2016, le système de déclenchement utilisé peut s'écrire : HLT\_j460\_a10\_sub\_L1J100. Ceci signifie que le jet au niveau du système de déclenchement est reconstruit avec l'algorithme anti- $k_t$  avec  $R = 1,0$  sans trimming, après une correction d'empilement telle que décrite dans la Section 5.2.1, et que ce jet doit satisfaire  $E_T > 460$  GeV, où  $E_T$  représente l'énergie transverse. Pour 2017 et 2018, le système de déclenchement utilisé peut s'écrire : HLT\_j460\_a10t\_lcw\_jes\_L1J100. Cela signifie que le jet au niveau du système de déclenchement est reconstruit avec l'algorithme anti- $k_t$  avec  $R = 1,0$  avec trimming à partir des constituants étalonnés avec la technique LCW, que l'étalonnage absolu (voir la Section 5.2.2) est appliqué, et que ce jet doit satisfaire  $E_T > 460$  GeV.

L'efficacité de ces systèmes de déclenchement est égale à 100% après les sélections d'analyse décrites dans la Section 6.4.5. Ceci peut être démontré sur la Figure 6.8 [83], qui représente l'efficacité de sélection du système de déclenchement en fonction de l'impulsion transverse du jet reconstruit au niveau des analyses pour des données de 2017. L'efficacité est obtenue en divisant le nombre de jets qui passent le critère de sélection sur l'impulsion transverse considéré par le nombre de jets avant cette sélection. Les jets avant sélection sont obtenus en utilisant un système de déclenchement avec un seuil en impulsion transverse plus petit que le critère de sélection du seuil considéré (460 GeV ici). Sur cette Figure, le système de déclenchement dont la condition est  $E_T > 460$  GeV est efficace à 100% à approximativement  $p_T = 480$  GeV. C'est ce système de déclenchement en particulier qui est utilisé dans l'analyse. L'autre système de déclenchement représenté sur la Figure 6.8 (avec la condition sur la masse) n'est pas utilisé.

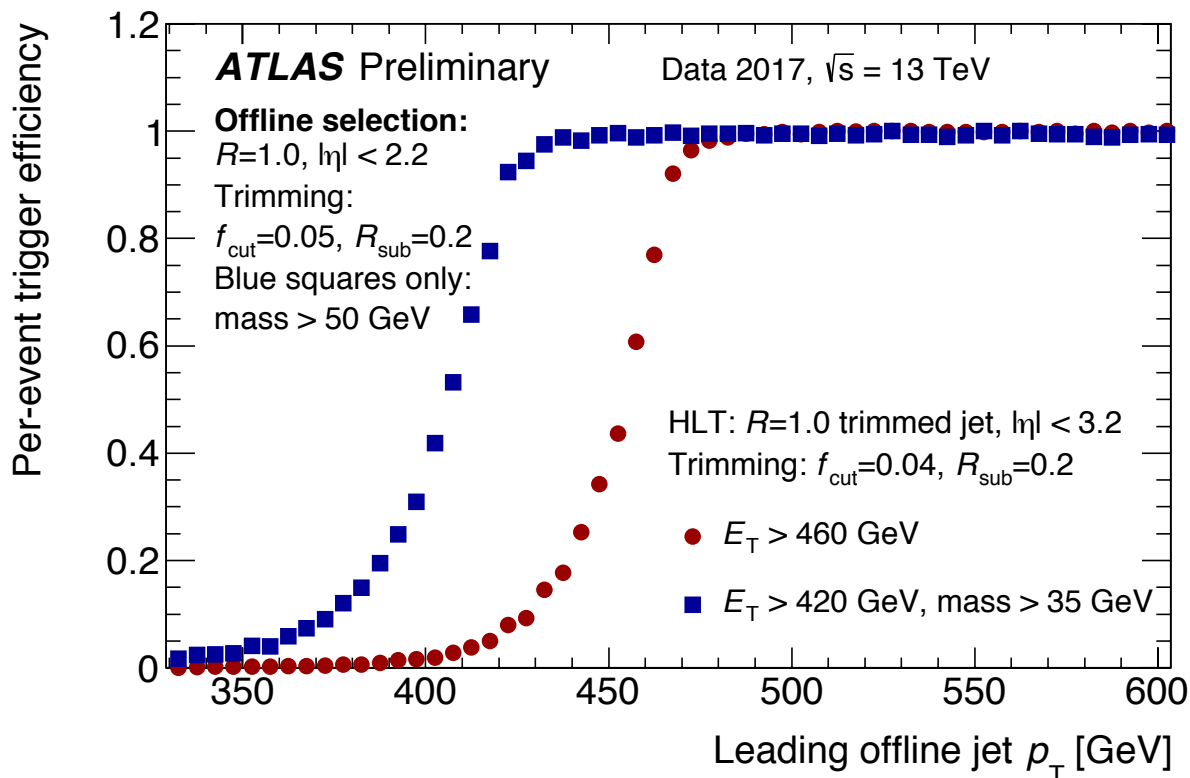


FIGURE 6.8 – Efficacité de sélection du système de déclenchement en fonction de l’impulsion transverse du jet reconstruit au niveau des analyses [83]. Les barres d’erreurs ne sont pas affichées.

#### 6.4.4 Nettoyage des jets

Afin de retirer les jets qui proviendraient de bruit du calorimètre ou encore de rayons cosmiques, tous les évènements qui contiennent au moins un jet de rayon  $R = 0,4$  qui ne satisfait pas la condition « *LooseBad* » telle que définie dans [84] ne sont pas utilisés. Cette condition utilise des informations telles que la forme du jet ou le temps auquel il parvient au détecteur. Tous les évènements contenant un tel jet sont enlevés afin de ne pas contaminer les quantités au niveau des évènements telles que la masse invariante. Cette procédure est appelée « *jet cleaning* » en anglais. L’efficacité de sélection de cette procédure est de 99,5% (99,9%) pour des jets avec  $p_T > 20$  (100) GeV [84], donc elle est égale à 100% pour des jets avec des sélections telles que décrites dans la Section suivante.

#### 6.4.5 Au niveau de l’analyse

Les jets sont tout d’abord étalonnés avec l’étalonnage absolu en énergie, masse et direction, tel que décrit dans la Section 5.2.2.

Plusieurs sélections sont alors appliquées au niveau de l’analyse :

- au moins 2 jets,

- $p_{\text{T}}^{\text{leading jet}} > 500 \text{ GeV}$ ,  $p_{\text{T}}^{\text{subleading jet}} > 400 \text{ GeV}$  et  $p_{\text{T}}^{\text{jet}} < 3000 \text{ GeV}$  pour les deux jets,
- $|\eta_{\text{jet}}| < 2.0$  pour les deux jets,
- $50 \text{ GeV} < m_{\text{jet}} < 600 \text{ GeV}$  pour les deux jets,
- pour les simulations du bruit de fond seulement :

$$0,7 < \frac{p_{\text{T}}^{\text{leading jet}} + p_{\text{T}}^{\text{subleading jet}}}{2 \times p_{\text{T}}^{\text{leading truth jet}}} < 1,3 \quad (6.7)$$

Ces conditions sont appliquées pour deux raisons. La première est qu'elles correspondent aux intervalles dans lesquels les incertitudes systématiques sur l'échelle et la résolution en énergie et en masse sont disponibles. En particulier, les conditions supérieures sur l'impulsion transverse ( $p_{\text{T}}^{\text{jet}} < 3000 \text{ GeV}$ ) et sur la masse ( $m_{\text{jet}} < 600 \text{ GeV}$ ) correspondent aux seuils recommandés pour les incertitudes sur l'échelle et la résolution en énergie et en masse fournies par ATLAS. À plus haute impulsion transverse, ou à plus haute masse, ces incertitudes devraient être encore valides mais ne sont pas recommandées. La seconde raison est pour améliorer la correspondance entre les données réelles et les simulations dans certaines régions de l'espace des phases telles qu'à basse impulsion ou à basse masse. La dernière condition permet d'enlever les pics artificiels dans la distribution en  $p_{\text{T}}^{\text{reco}}$ , comme décrit dans la Section 5.3.2. Cette condition n'introduit pas de biais dans le spectre en impulsion transverse par exemple. Elle permet seulement d'enlever les pics provenant d'évènements avec un poids très élevé.

Les systèmes de déclenchement utilisés sont efficaces à 100% pour une masse invariante des deux jets de plus haute impulsion transverse  $m_{\text{JJ}} \geq 1300 \text{ GeV}$ , ce qui correspond donc à la borne inférieure de l'intervalle sur lequel l'analyse est effectuée.

Par la suite, toutes ces coupures seront appliquées aux évènements des données réelles, des simulations de bruit de fond et des simulations de signal.

## 6.5 Stratégie de l'analyse

La Figure 6.9 représente la distribution en masse invariante des deux jets de plus haute impulsion pour toutes les masses du  $Z'$  de chaque modèle étudié après les coupures de pré-sélection. On peut remarquer que le nombre d'évènements diminue avec la masse du  $Z'$ , ce qui est dû à la section efficace (voir Table 6.3). On peut aussi remarquer que les signaux deviennent de plus en plus larges au fur et à mesure que la masse du  $Z'$  augmente.

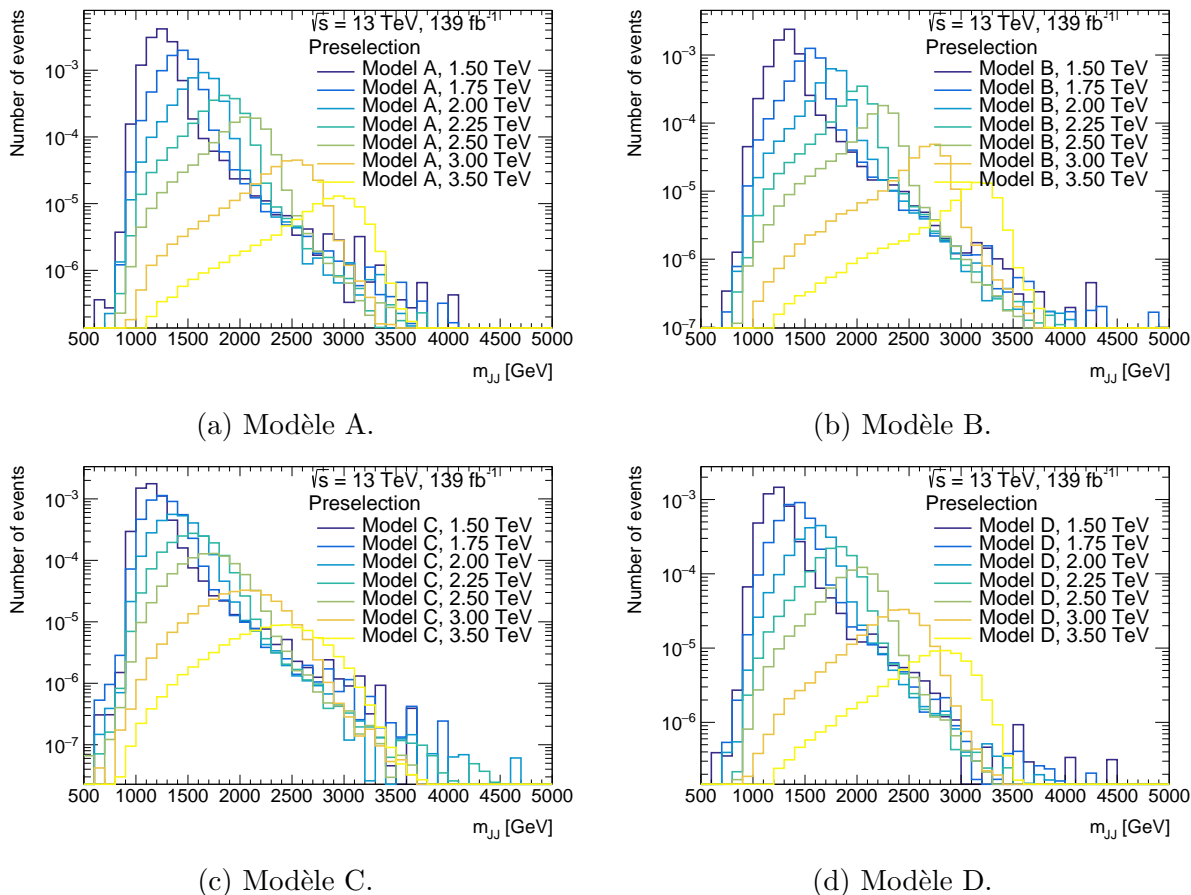


FIGURE 6.9 – Distribution en masse invariante des deux jets de plus haute impulsion pour toutes les masses du  $Z'$  de chaque modèle étudié après les coupures de pré-sélection.

La largeur des signaux est quantifiée par la Table 6.4, qui représente l'écart-type de la distribution en masse invariante reconstruite, divisé par le maximum de cette distribution. On peut remarquer que les signaux sont très larges ( $> 10\%$ ), que ce soit à basse et à haute masse pour le  $Z'$ , ce qui est dû à une reconstruction imparfaite et en particulier au trimming. En particulier, le Modèle C est le modèle avec les distributions en masse invariante les plus larges.

Modèle	1,5 TeV	1,75 TeV	2 TeV	2,25 TeV	2,5 TeV	3 TeV	3,5 TeV
A	12,9	12,5	12,6	13,2	13,3	14,1	14,8
B	14,4	13,3	12,8	12,5	12,7	13,0	13,5
C	16,2	15,4	15,6	16,6	16,6	17,0	17,3
D	16,0	14,8	14,0	13,8	13,8	14,8	14,5

TABLE 6.4 – Largeur de la distribution en masse invariante reconstruite divisée par le maximum de la distribution ( $\frac{\text{Ecart type}}{\text{Maximum}}$ ) exprimé en pourcentages pour tous les signaux après les coupures de pré-sélection.

Originellement, la stratégie de l'analyse était d'estimer le bruit de fond avec un ajustement



tement des données dans la Région de Signal à l'aide d'une « *fonction de bruit de fond seulement* », c'est-à-dire une fonction décroissant de façon monotone. Cependant, dû à la largeur des signaux que nous étudions, nous avons remarqué que, lorsque nous injectons du signal artificiellement dans les données (pour plus de détails, voir la Section 6.11.3), cet ajustement avec une fonction de bruit de fond seulement absorbe en partie le signal que nous avons injecté, et ce même avec des fonctions contenant un petit nombre de paramètres libres. Par conséquent nous ne récupérerions pas le nombre d'évènements de signal que nous avons injecté. La Figure 6.10 représente une comparaison de l'ajustement avec une fonction de bruit de fond seulement des simulations de bruit de fond (de QCD pure) sans et avec du signal injecté artificiellement à 5 fois la racine carré du nombre d'évènements de bruit de fond attendus dans une fenêtre centrée sur le signal. Les simulations de bruit de fond sans signal injecté, représentées par les simulations du bruit de fond dans la Région de Signal (définie dans la Section 6.6), sont en noir. Les simulations de bruit de fond avec signal injecté sont en rouge. L'ajustement de bruit de fond seulement sur les simulations de bruit de fond sans signal injecté est en bleu. L'ajustement de bruit de fond seulement sur les simulations de bruit de fond avec signal injecté est en vert. On peut remarquer que les deux fonctions de bruit de fond seulement diffèrent, car celle qui est ajustée aux simulations de bruit de fond avec du signal injecté absorbe en partie ce signal.

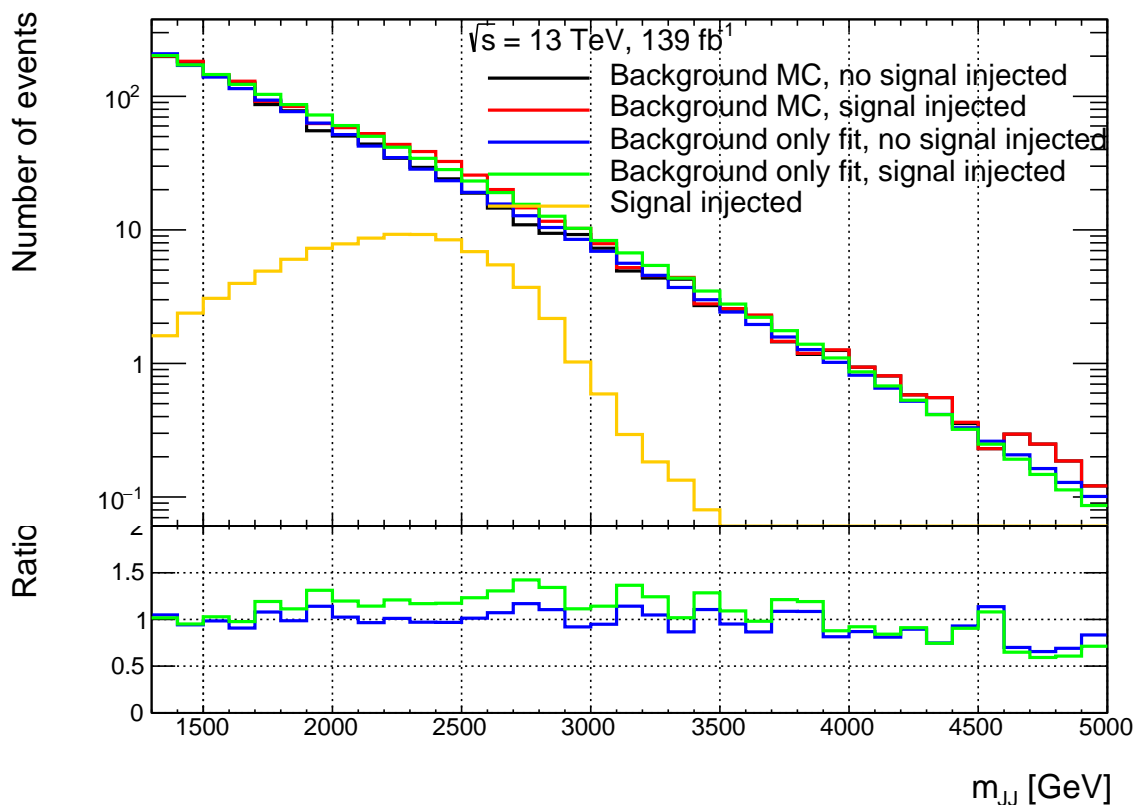


FIGURE 6.10 – Comparaison d'un ajustement avec une fonction de bruit de fond seulement sans signal injecté, et avec signal injecté. Le panneau du bas représente le rapport entre les simulations de bruit de fond, et les ajustements avec des fonctions de bruit de fond seulement sur les simulations de bruit de fond avec et sans signal injecté.

C'est pourquoi nous avons décidé de changer de stratégie d'analyse et d'estimer le bruit de fond à l'aide d'une Région de Contrôle (*Control Region* en anglais, ou plus brièvement « *CR* »), dans laquelle la même forme que la distribution en masse invariante du bruit de fond dans la Région de Signal est attendue. Cette Région de Contrôle, définie dans les données, est une région avec un rapport signal sur bruit de fond extrêmement faible afin de limiter l'éventuelle contamination provenant de jets de QCD sombre.

## 6.6 Optimisation des coupures de sélection

Afin d'optimiser le rapport signal sur bruit de fond, plusieurs variables permettant de discriminer les jets de QCD sombre de ceux du Modèle Standard, inspirées par les études phénoménologiques présentées dans la Section 6.1, sont étudiées : le nombre de traces associées au jet,  $C_1^\beta$  et la fraction d'énergie du jet déposée dans le calorimètre électromagnétique (*E-ratio*). Les traces qui sont associées aux jets doivent remplir les critères de sélection correspondant au point de travail *Loose*, tel que décrit dans la Section 4.1. Les traces sont alors associées à un jet à l'aide de la technique d'« *association fantôme* » [85] : le processus de regroupement avec l'algorithme Anti- $k_t$  des jets est répété avec l'ajout de versions « fantômes » des traces, qui ont la même direction mais une impulsion infiniment petite, de sorte qu'elles ne changent pas les propriétés du jet. Une trace est associée à un jet si sa version fantôme est contenue dans le jet après le regroupement avec l'algorithme Anti- $k_t$ . Deux possibilités sont disponibles pour le nombre de traces pour des jets auxquels un algorithme de grooming est appliqué : les traces peuvent être associées via la technique d'association fantôme au jet avant (dans ce cas ces traces sont appelées *ungroomed tracks*) ou après (auquel cas il n'y a pas de dénomination particulière) que le grooming soit appliqué. Ces deux cas sont étudiés. Enfin le paramètre  $\beta$  par défaut utilisé dans ATLAS dans le calcul de  $C_1^\beta$  est  $\beta = 1$  ( $\beta = 0, 2$  était utilisé dans la Section 6.1). C'est donc  $\beta = 1$  qui sera utilisé dans cette Section. Cette valeur n'est pas optimale pour sonder les radiations collinéaires à petite échelle, mais cela n'impacte pas de manière significative les résultats.

La figure de mérite pour déterminer quelle variable permet de discriminer au mieux les jets de QCD sombre de ceux du Modèle Standard est la signification statistique  $S(c)$ , qui est donnée par la formule suivante [86] :

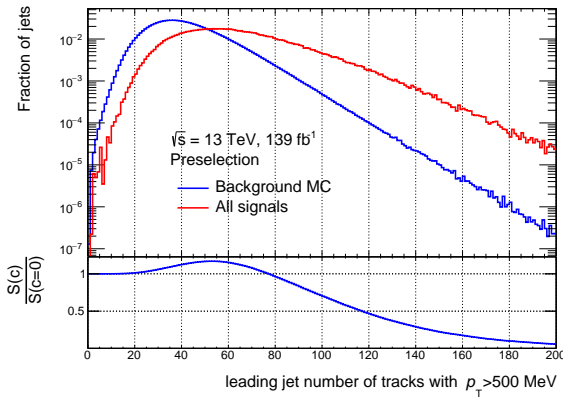
$$S(c) = \sqrt{2 \times ((s(c) + b(c)) \ln(1 + \frac{s(c)}{b(c)}) - s(c))} \quad (6.8)$$

où :

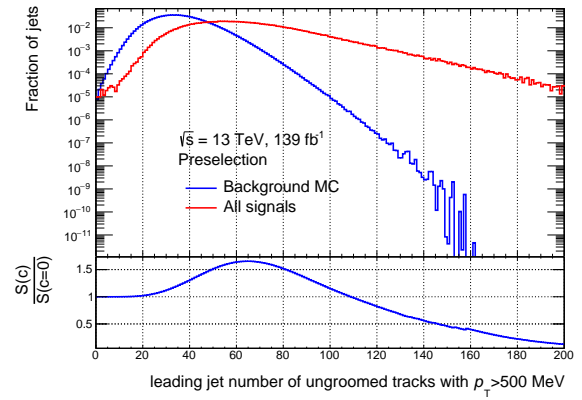
- $c$  représente une valeur de coupure (qui peut être inférieure ou supérieure),
- $s(c)$  représente le nombre d'évènements de signal qui passent la coupure de sélection  $c$ ,
- $b(c)$  représente le nombre d'évènements de bruit de fond qui passent la coupure de sélection  $c$ .

La variable la plus discriminante est alors celle qui permet de maximiser le gain en signification statistique  $\frac{S(c)}{S(c=0)}$ .

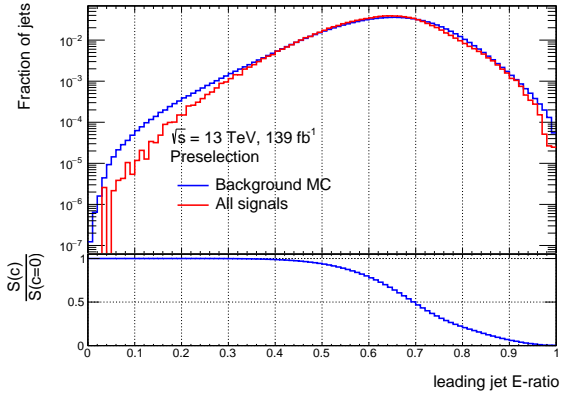
La Figure 6.11 représente une comparaison de plusieurs variables pour le jet de plus haute impulsion transverse entre le bruit de fond et tous les signaux des tous les modèles réunis. Sur la Figure 6.11a on peut remarquer que les jets de QCD sombre possèdent effectivement un plus grand nombre de traces qui leur sont associées. Sur la Figure 6.11b, on peut aussi remarquer que les jets de QCD sombre possèdent aussi un plus grand nombre de traces sans grooming. De plus cette variable permet de mieux discriminer les jets de QCD sombre que le nombre de traces avec grooming. Enfin, sur les Figures 6.11c et 6.11d, on peut remarquer que la fraction d'énergie déposée dans le calorimètre électromagnétique et  $C_1^{\beta=1}$  ne sont pas des variables très discriminantes, car il n'y a pas de séparation claire entre les signaux et le bruit de fond. Les conclusions sont similaires pour le second jet de plus haute impulsion transverse.



(a) Nombre de traces.



(b) Nombre de traces sans grooming.



(c) Fraction d'énergie déposée dans le calorimètre électromagnétique.

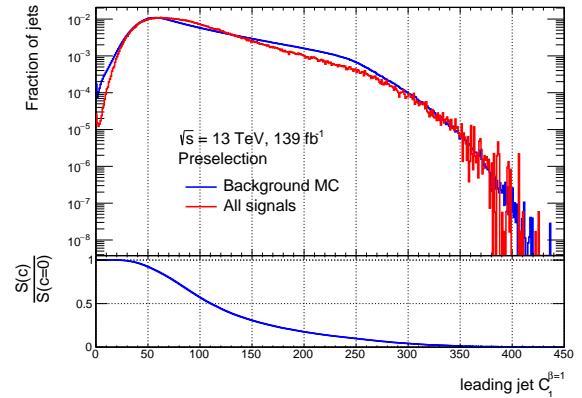

 (d)  $C_1^{\beta=1}$ .

FIGURE 6.11 – Comparaison de plusieurs variables pour le jet de plus haute impulsion transverse entre le bruit de fond et tous les signaux de tous les modèles réunis.

Ainsi, le nombre de traces sans grooming associées aux jets est la variable qui permet de mieux discriminer le signal du bruit de fond. Les autres variables étudiées ne permettent de discriminer que certains modèles spécifique, alors que le nombre de traces permet de discriminer tous les modèles. Par conséquent, étant donné que cette analyse représentait une première itération pour ce type de recherche, nous avons décidé de n'utiliser qu'une seule variable pour tous les modèles : le nombre de traces sans grooming associées aux

jets. Cette variable sera notée  $n_{\text{trk}}$  par la suite. Cependant, comme énoncé précédemment, certains modèles possèdent plus de traces que d'autres. Ceci peut être vu dans les Figures 6.12 et 6.13, qui représentent une comparaison du nombre de traces sans grooming associées aux jets pour un signal avec un grand nombre de traces (Modèle D,  $M_{Z'} = 3,5$  TeV) et signal avec un faible nombre de traces (Modèle B,  $M_{Z'} = 1,5$  TeV) respectivement. On peut remarquer que le gain en sensibilité est beaucoup plus faible pour le signal de Modèle B, mais que ce signal possède tout de même plus de traces que le bruit de fond.

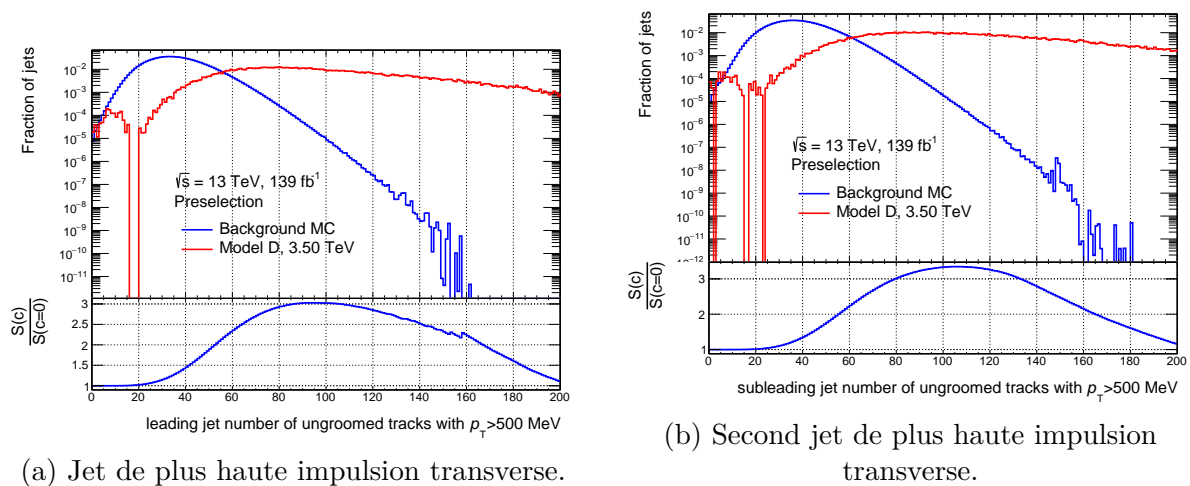


FIGURE 6.12 – Comparaison du nombre de traces sans grooming associées au jet de plus haute impulsion transverse (gauche) et du second jet de plus haute impulsion transverse (droite) entre le bruit de fond et Modèle D avec  $M_{Z'} = 3,5$  TeV.

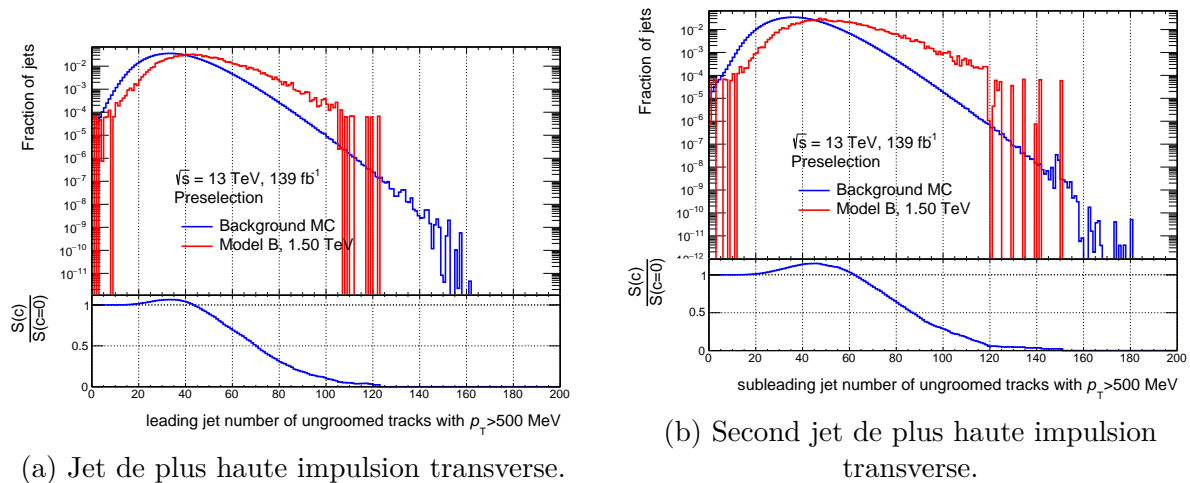


FIGURE 6.13 – Comparaison du nombre de traces sans grooming associées au jet de plus haute impulsion transverse (gauche) et du second jet de plus haute impulsion transverse (droite) entre le bruit de fond et Modèle B avec  $M_{Z'} = 1,5$  TeV.

En effet, les Figures 6.14 et 6.15 représentent les distributions du nombre de traces sans grooming pour le jet de plus haute impulsion transverse et du second jet de plus haute impulsion transverse respectivement, et ce pour toutes les masses de tous les modèles. On

peut remarquer que deux Modèles possèdent tout particulièrement un grand nombre de traces : les Modèles A et D. On peut aussi remarquer que le nombre de traces augmente avec la masse du  $Z'$ .

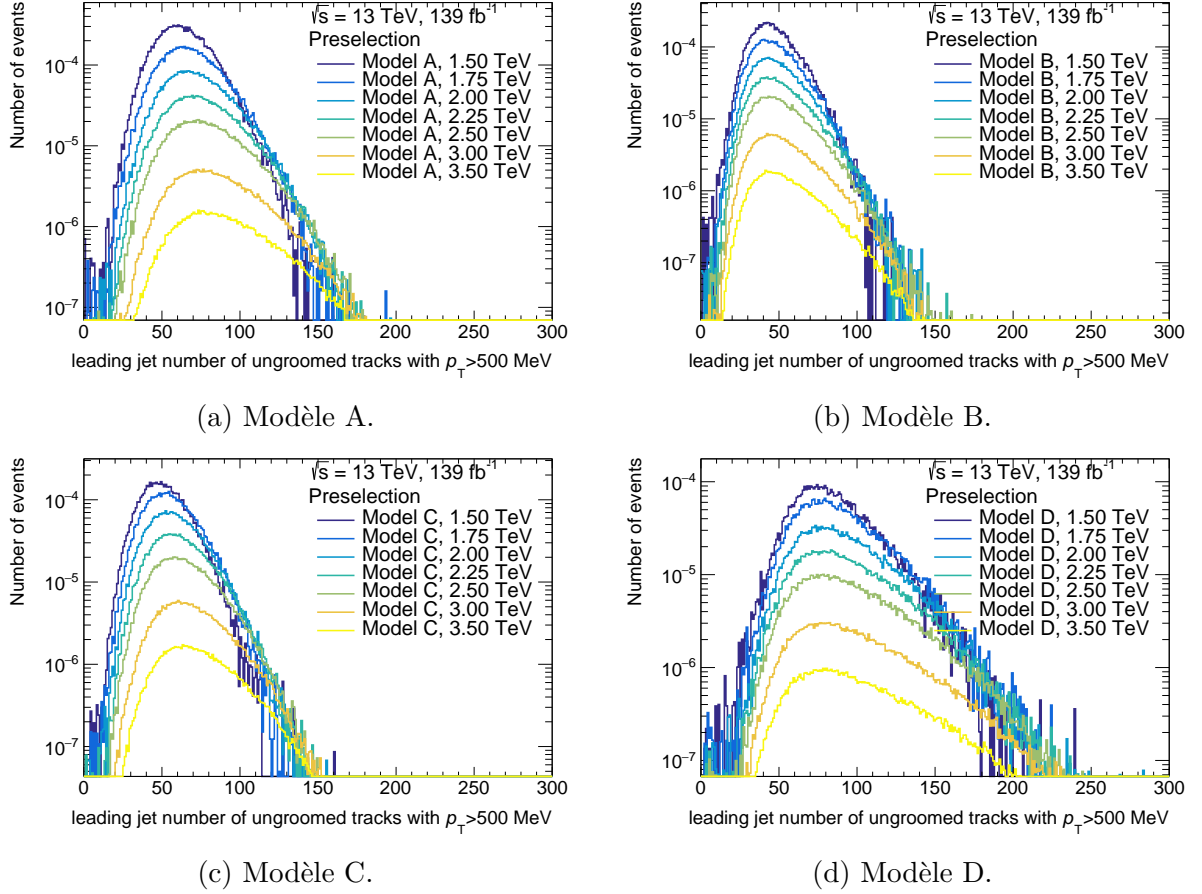
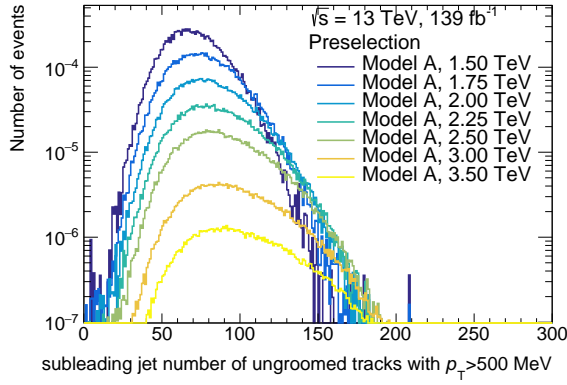
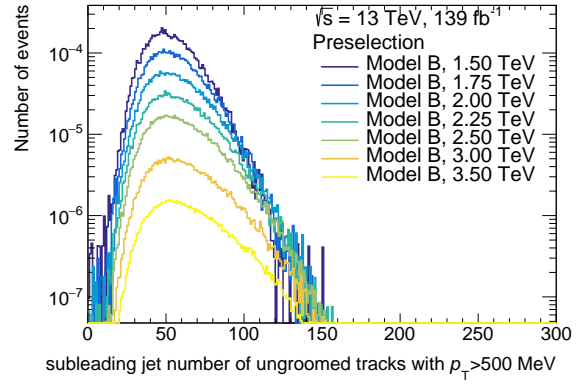


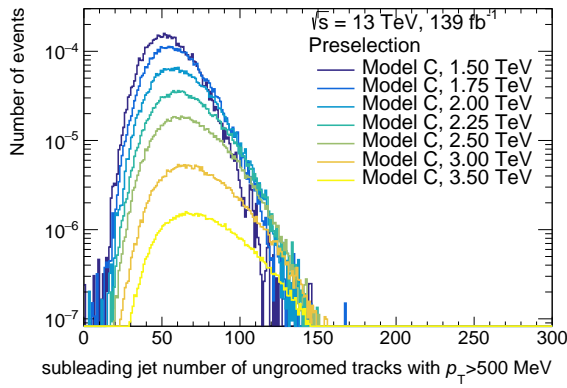
FIGURE 6.14 – Distribution en nombre de traces sans grooming du jet de plus haute impulsion pour toutes les masses du  $Z'$  de chaque modèle étudié après les coupures de pré-sélection.



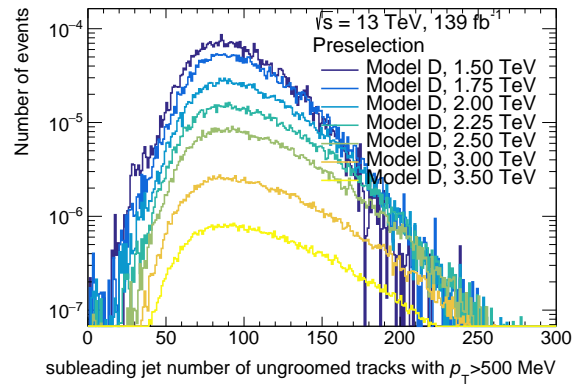
(a) Modèle A.



(b) Modèle B.



(c) Modèle C.



(d) Modèle D.

FIGURE 6.15 – Distribution en nombre de traces sans grooming du second jet de plus haute impulsion pour toutes les masses du  $Z'$  de chaque modèle étudié après les coupures de pré-sélection.

Cependant, pour le bruit de fond, le nombre de traces est corrélé avec la masse invariante des deux jets. Par conséquent, une coupure sur le nombre de traces sans grooming sculpte le spectre en masse invariante du bruit de fond, comme représenté sur la Figure 6.16.

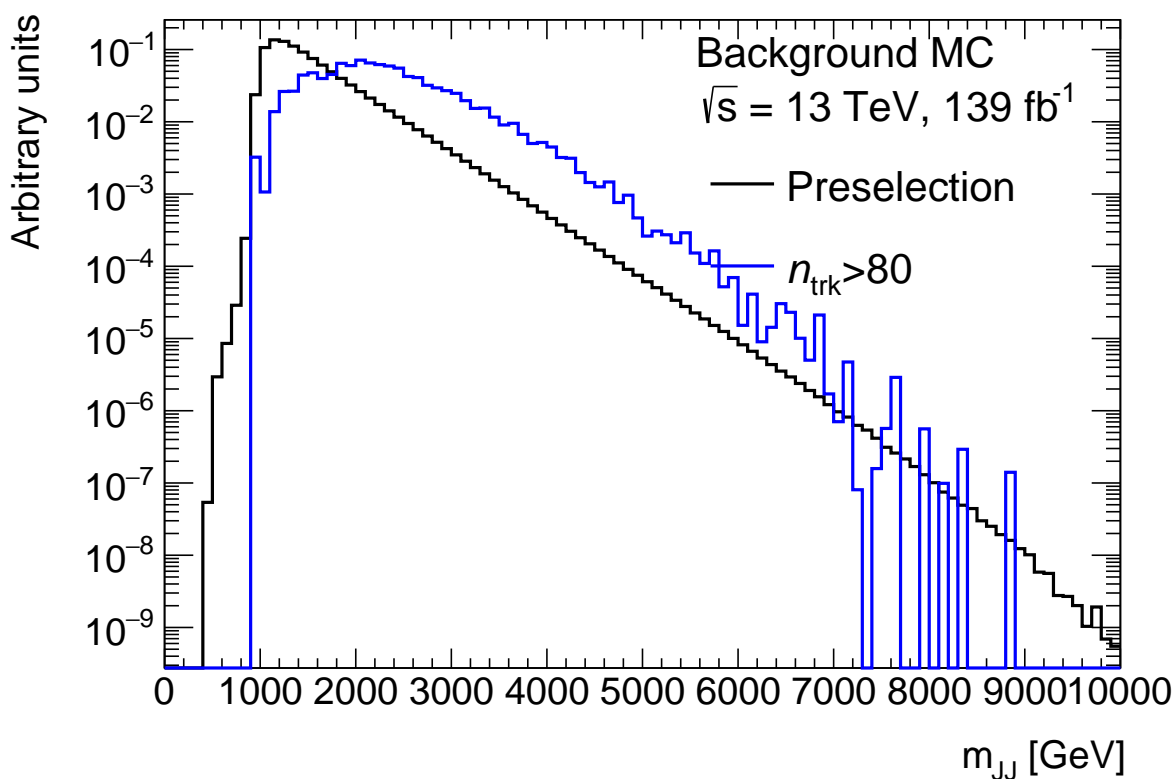


FIGURE 6.16 – Comparaison du spectre en masse invariante du bruit de fond simulé dans différentes régions : la région « inclusive » et après une coupure directe sur le nombre de traces.

Ainsi, si par exemple la Région de Contrôle était définie avec un bas nombre de traces et la Région de Signal avec un haut nombre de traces, ces deux régions n'auraient pas le même spectre en masse invariante, et la Région de Contrôle ne pourrait pas servir à estimer le bruit de fond dans la Région de Signal.

Afin d'assurer qu'une coupure sur le nombre de traces ne sculpte pas le spectre en masse invariante, une méthode de dé-corrélation entre la masse invariante et le nombre de traces a été mise en place. Cette méthode est similaire à celle qui peut être trouvée dans [87] et est appelée « *régression à efficacité fixée* ». Pour ce faire, le spectre en masse invariante est divisé en intervalles de 100 GeV, ce qui correspond à la résolution expérimentale sur la masse invariante, comme représenté sur la Figure 6.17.

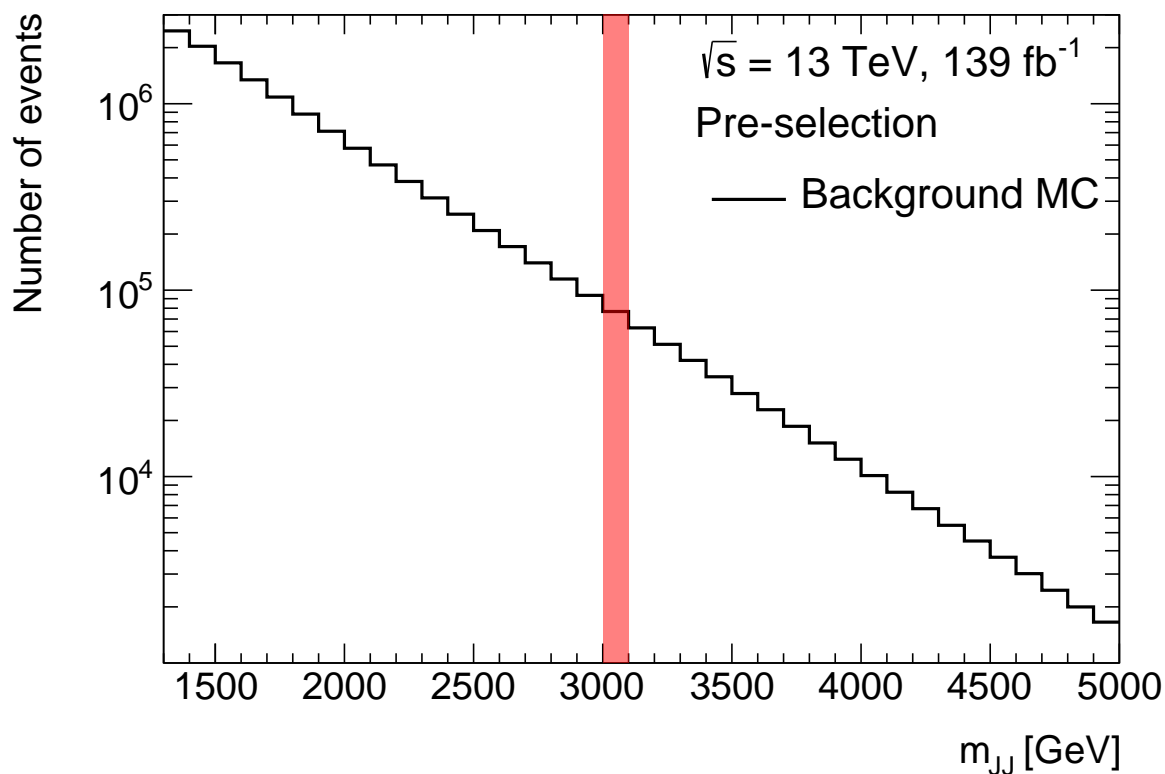


FIGURE 6.17 – Distribution de la masse invariante des simulations de bruit de fond après les coupures de pré-sélection. Un exemple d’intervalle de 100 GeV en masse invariante des deux jets est représenté.

Dans chaque intervalle en masse invariante indépendamment, la valeur de la coupure sur le nombre de traces qui résulte en une certaine efficacité de coupure de bruit fond est calculée. La valeur cible de cette efficacité, appelée  $\varepsilon$ , est fixée à l’avance. La valeur de la coupure sur le nombre de traces qui mène à cette efficacité est calculée à l’aide de la fonction `GetQuantiles` de ROOT [88]. Cette valeur est notée  $P^\varepsilon$ . Cette procédure est représentée sur la Figure 6.18.



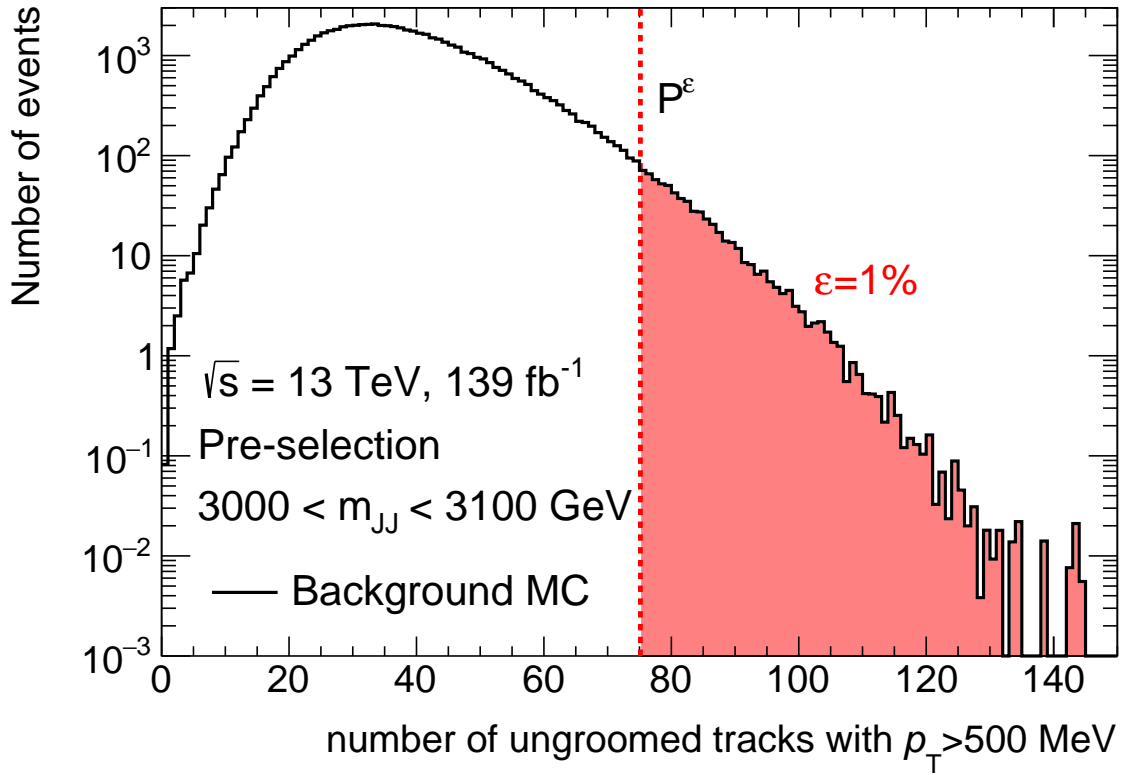


FIGURE 6.18 – Distribution du nombre de traces sans grooming qui satisfont  $p_T > 500$  MeV dans un certain intervalle donné en  $m_{JJ}$  pour les simulations de bruit de fond. La ligne en pointillés rouge verticale correspond à la valeur de la coupure sur le nombre de traces qui correspond à une certaine efficacité de sélection  $\varepsilon$  (ici égale à 1%). La valeur de cette coupure est dénotée  $P^\varepsilon$ .

Ensuite cette procédure est répétée pour tous les intervalles en masse invariante, et la valeur de la coupure sur le nombre de traces qui résulte en cette certaine efficacité de sélection  $\varepsilon$  est tracée en fonction de la masse invariante, comme représenté sur la Figure 6.19.

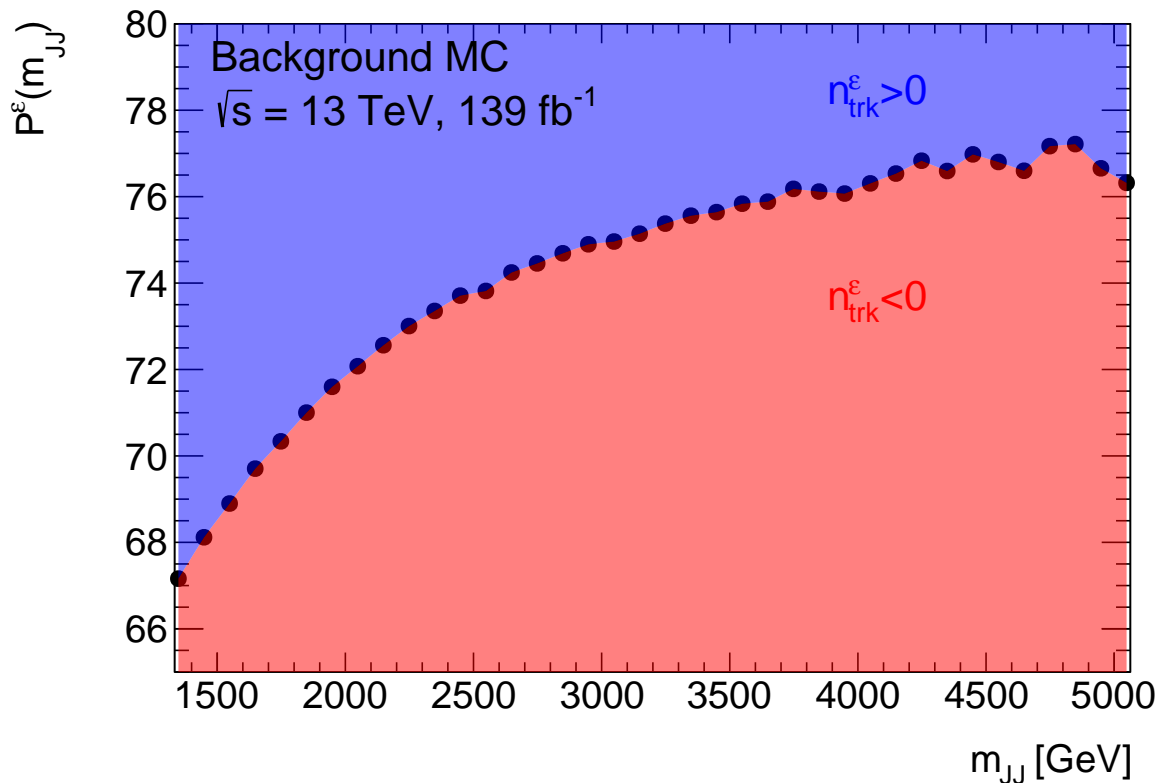


FIGURE 6.19 – Courbe de la coupure sur le nombre de traces sans grooming satisfaisant  $p_T > 500$  MeV qui résulte une certaine efficacité de sélection  $\varepsilon = 1\%$  en fonction de la masse invariante pour les simulations de bruit de fond.

Afin de ne pas tenir compte des fluctuations statistiques de bruit de fond simulé, cette courbe de la coupure sur le nombre de traces résultant en une certaine efficacité de sélection en fonction de la masse invariante est ensuite ajustée en utilisant plusieurs polynômes. Celui qui est retenu est celui avec le plus petit  $\frac{\chi^2}{\text{number of degrees of freedom}}$ . La fonction `GetQuantiles` de ROOT permet d'obtenir une courbe lisse, facilement ajustable, en fonction de la masse invariante, car une interpolation linéaire est effectuée par cette fonction. Ceci résulte en une valeur de la coupure qui peut être non-entière alors que le nombre de traces est une valeur entière, et donc en une valeur de la coupure un peu plus élevée que ce qu'elle devrait être. Par conséquent l'efficacité de sélection est plus petite que l'efficacité cible. Le profil ajusté en fonction de la masse invariante est appelé *percentiles*, et est noté  $P^\varepsilon(m_{JJ})$ . Enfin, une nouvelle observable notée  $n_{\text{trk}}^\varepsilon$  est définie pour chaque jet par :

$$n_{\text{trk}}^\varepsilon(m_{JJ}) = n_{\text{trk}} - P^\varepsilon(m_{JJ}) \quad (6.9)$$

Ainsi, une coupure telle que  $n_{\text{trk}}^\varepsilon(m_{JJ}) > 0$  ( $\Leftrightarrow n_{\text{trk}} > P^\varepsilon(m_{JJ})$ ) permet d'obtenir la même efficacité de sélection dans chaque intervalle en masse invariante, et par conséquent le spectre en masse invariante n'est pas modifié par cette coupure. Par la suite,  $n_{\text{trk}}^\varepsilon(m_{JJ})$  sera noté  $n_{\text{trk}}^\varepsilon$ .

Cette procédure est dérivée pour le jet de plus haute impulsion en premier, puis pour

le second jet de plus haute impulsion, avec la coupure  $n_{\text{trk}}^\varepsilon > 0$  du jet de plus haute impulsion déjà appliquée. Ceci permet de minimiser les effets de corrélation entre les jets.

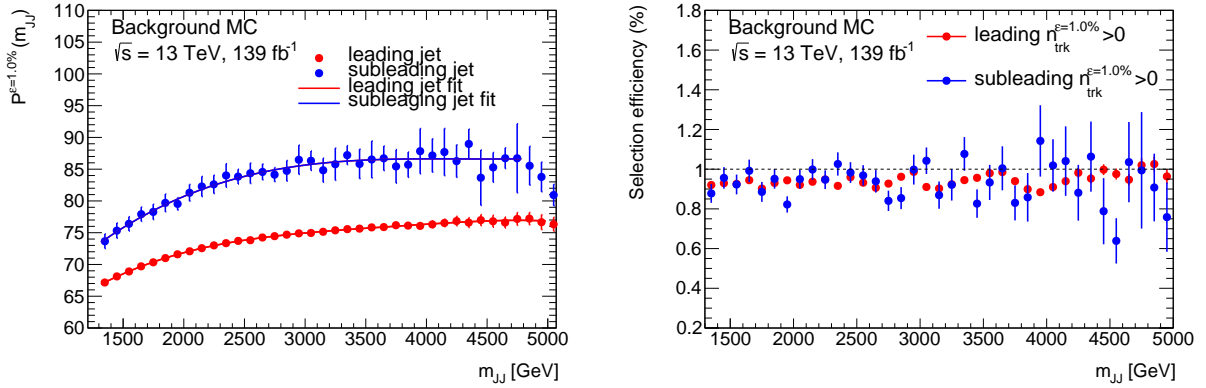
Pour cette analyse, nous avons choisi d'utiliser une valeur de l'efficacité de sélection du bruit de fond cible  $\varepsilon = 1\%$  pour chaque jet (donc pour les deux jets l'efficacité est  $1\% \times 1\%$ ), afin qu'il y ait un équilibre entre la réduction du nombre d'événements de bruit de fond et de ne pas être trop sensible aux incertitudes systématiques.

Les courbes de  $P^\varepsilon(m_{JJ})$  sont définies à partir des simulations du bruit de fond car s'ils étaient définis dans les données et qu'un signal était présent, nous pourrions éventuellement voir une bosse dans les courbes de la valeur de la coupure sur le nombre de traces en fonction de la masse invariante qui correspondrait à ce signal, car il faudrait couper plus fort sur le nombre de traces afin de garder la même efficacité. De plus, la bosse correspondant à ce signal disparaîtrait après coupure, par définition de la méthode. Les percentiles ne peuvent pas non plus être définis dans les données dans une Région de Contrôle (par exemple en appliquant la sélection  $n_{\text{trk}} < 20$ ), car la sélection sculpte le spectre en masse invariante. Ainsi, cette Région de Contrôle n'aurait pas le même spectre en masse invariante que la région originelle (« inclusive »), et les percentiles dérivés à partir de la Région de Contrôle ne permettraient pas de garder une efficacité constante sur toute la gamme en masse invariante s'ils étaient appliqués à la région originelle.

C'est pourquoi les percentiles sont dérivés à partir des simulations du bruit de fond, puis appliqués aux données, c'est-à-dire que pour construire les variables  $n_{\text{trk}}^\varepsilon$  pour les données, ce sont les courbes  $P^\varepsilon(m_{JJ})$  qui sont définies à partir des simulations de bruit de fond qui sont utilisées. Un bon accord entre les données et le bruit de fond est donc nécessaire. Des incertitudes systématiques liées à ce choix seront appliquées (voir la Section 6.9).

La Figure 6.20a représente les percentiles, ainsi que leur ajustement, définis dans les simulations du bruit et avec une efficacité cible  $\varepsilon = 1\%$ . On peut remarquer que les points ne sont pas des entiers, comme discuté auparavant, mais aussi que la courbe pour le second jet de plus haute impulsion transverse est plus haute (en bleu) que la courbe pour le jet de plus haute impulsion transverse (en rouge). Ceci signifie que de manière générale, le second jet de plus haute impulsion transverse contient plus de traces que le jet de plus haute impulsion transverse. Ceci est dû au fait que le second jet est plus souvent initié par un gluon, et que la réponse du calorimètre à des jets initiés par des gluons est plus basse que pour des jets initiés par des quarks (voir la Section 5.2.3), mais aussi que les jets de plus haute impulsion transverse sont plus collimés. Enfin on peut remarquer que les barres d'erreur sur les points sont plus grandes pour le second jet de plus haute impulsion transverse que pour le jet de plus haute impulsion transverse. Ceci est dû au fait que les percentiles du second jet de plus haute impulsion transverse sont dérivés avec la coupure sur le jet de plus haute impulsion transverse déjà appliquée. La Figure 6.20b représente l'efficacité de sélection après la coupure  $n_{\text{trk}}^\varepsilon > 0$  pour les deux jets de plus haute impulsion transverse. On peut remarquer que l'efficacité de sélection est légèrement plus faible que l'efficacité cible (1%), comme discuté auparavant. Ces deux courbes sont plates, ce qui indique que la distribution en masse invariante est semblable, ce qui est l'objectif de la méthode.

## 6.6. Optimisation des coupures de sélection



(a) Percentiles en fonction de la masse invariante pour les deux jets de plus haute impulsion transverse.

(b) Efficacité de sélection après les coupures  $n_{\text{trk}}^{\epsilon} > 0$  pour les deux jets de plus haute impulsion transverse.

FIGURE 6.20 – Gauche : percentiles en fonction de la masse invariante pour les deux jets de plus haute impulsion transverse. Droite : efficacité de sélection après les coupures  $n_{\text{trk}}^{\epsilon} > 0$  pour les deux jets de plus haute impulsion transverse. Les deux Figures n'impliquent que les simulations de bruit de fond.

La Région de Signal de l'analyse est définie en appliquant la sélection  $n_{\text{trk}}^{\epsilon} > 0$  pour les deux jets. La Figure 6.21 représente une comparaison pour la simulation de bruit de fond du spectre en masse invariante dans différentes régions : la région « inclusive » (nommée *Preselection*), après une coupure directe sur le nombre de traces, et après les coupures correspondant à la Région de Signal. On peut remarquer que la coupure sur le nombre de traces sculpte effectivement le spectre en masse invariante par rapport à la région inclusive, ce qui n'est pas le cas des coupures de la Région de Signal, ce qui confirme que la méthode fonctionne.

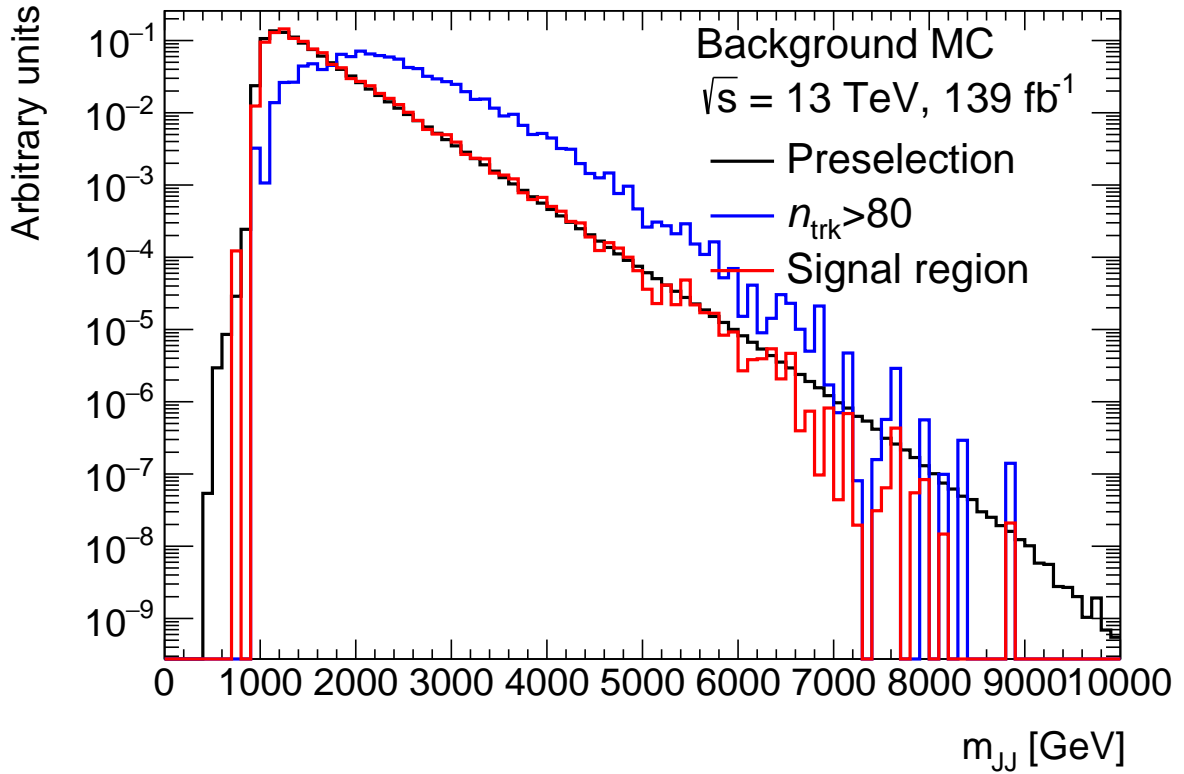


FIGURE 6.21 – Comparaison du spectre en masse invariante du bruit de fond simulé dans différentes régions : la région « inclusive », après une coupure directe sur le nombre de traces, et après les coupures correspondant à la Région de Signal.

La méthode est validée dans les simulations du bruit de fond en construisant plusieurs régions, dont le spectre en masse invariante doit avoir la même forme. Ces régions sont les suivantes : la Région de Signal, définie précédemment, la Région de Contrôle, et la Région de Validation. La Région de Contrôle est définie en exigeant que  $n_{\text{trk}}^{\varepsilon} < 0$  pour les deux jets. Cette Région possède une contamination de signal extrêmement faible car elle correspond à une coupure à bas nombre de traces sur les deux jets, et que 99% des événements du bruit de fond persistent avec une coupure sur un jet. Enfin, la Région de Validation représente une Région intermédiaire entre la Région de Contrôle et la Région de Signal. La Région de Validation est définie en exigeant que  $n_{\text{trk}}^{\varepsilon} > 0$  pour le jet de plus haute impulsion transverse et  $n_{\text{trk}}^{\varepsilon} < 0$  pour le second jet de plus haute impulsion transverse. Cette Région sera utilisée par la suite afin de tester que la méthode fonctionne dans les données réelles. La région avec  $n_{\text{trk}}^{\varepsilon} < 0$  pour le jet de plus haute impulsion transverse et  $n_{\text{trk}}^{\varepsilon} > 0$  pour le second jet de plus haute impulsion transverse n'est pas incluse dans la Région de Validation, car les percentiles pour le second jet de plus haute impulsion transverse étant définis avec la coupure sur le jet de plus haute impulsion transverse appliquée, la méthode de dé-corrélation ne fonctionnerait pas. Les différentes régions sont schématisées sur la Figure 6.22.

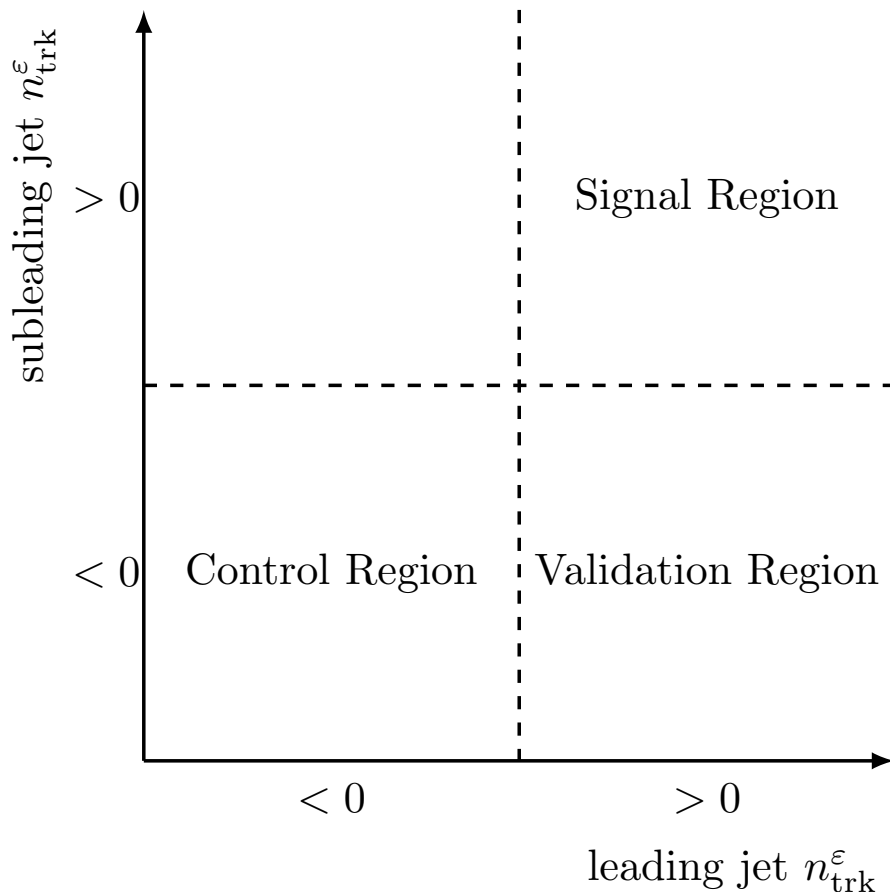


FIGURE 6.22 – Schéma des différentes Régions : de Signal, de Contrôle, de Validation, selon la coupure en  $n_{\text{trk}}^{\varepsilon}$  sur le jet de plus haute impulsion ou sur le second jet de plus haute impulsion.

Les Tables 6.5 et 6.6 représentent les efficacités de sélections pour le bruit de fond simulé et tous les signaux avec  $M_{Z'} = 2$  TeV et  $M_{Z'} = 3, 5$  TeV respectivement dans chacune des régions. On peut remarquer que l'efficacité de sélections pour tous les signaux dans la Région de Contrôle est plutôt élevée (42% et 31% respectivement). Ceci est dû au fait que les Modèles B et C ne contiennent pas beaucoup plus de traces que les simulations de bruit de fond. Cependant, l'efficacité de sélection des simulations de bruit de fond dans la Région de Contrôle est extrêmement élevée (99%). Ainsi, puisque le nombre d'évènements de bruit de fond est beaucoup plus important que le nombre d'évènements de signal, la contamination en signal reste extrêmement faible dans la Région de Contrôle.

Sample	Signal Region	Validation Region	Control Region
Background MC	$8,6 \times 10^{-5}$	$9,2 \times 10^{-3}$	$9,9 \times 10^{-1}$
All signals	$2,3 \times 10^{-1}$	$1,6 \times 10^{-1}$	$4,2 \times 10^{-1}$
Model A	$2,9 \times 10^{-1}$	$2,1 \times 10^{-1}$	$2,6 \times 10^{-1}$
Model B	$1,9 \times 10^{-2}$	$8,5 \times 10^{-2}$	$7,8 \times 10^{-1}$
Model C	$5,4 \times 10^{-2}$	$1,7 \times 10^{-1}$	$6,3 \times 10^{-1}$
Model D	$5,8 \times 10^{-1}$	$1,7 \times 10^{-1}$	$6,8 \times 10^{-2}$

TABLE 6.5 – Efficacités de sélections pour le bruit de fond simulé et tous les signaux avec  $M_{Z'} = 2$  TeV dans chacune des régions.

Sample	Signal Region	Validation Region	Control Region
Background MC	$8,6 \times 10^{-5}$	$9,2 \times 10^{-3}$	$9,9 \times 10^{-1}$
All signals	$3,1 \times 10^{-1}$	$2,0 \times 10^{-1}$	$3,1 \times 10^{-1}$
Model A	$4,6 \times 10^{-1}$	$2,0 \times 10^{-1}$	$1,3 \times 10^{-1}$
Model B	$3,8 \times 10^{-2}$	$1,3 \times 10^{-1}$	$6,7 \times 10^{-1}$
Model C	$1,9 \times 10^{-2}$	$2,6 \times 10^{-1}$	$3,6 \times 10^{-1}$
Model D	$5,3 \times 10^{-1}$	$2,0 \times 10^{-1}$	$8,4 \times 10^{-2}$

TABLE 6.6 – Efficacités de sélections pour le bruit de fond simulé et tous les signaux avec  $M_{Z'} = 3,5$  TeV dans chacune des régions.

La Figure 6.23 représente une comparaison de la distribution en masse invariante normalisée pour le bruit de fond simulé de toutes ces régions. On peut remarquer que les distributions en masse invariante ont toutes la même forme et que par conséquent la méthode fonctionne comme attendu.

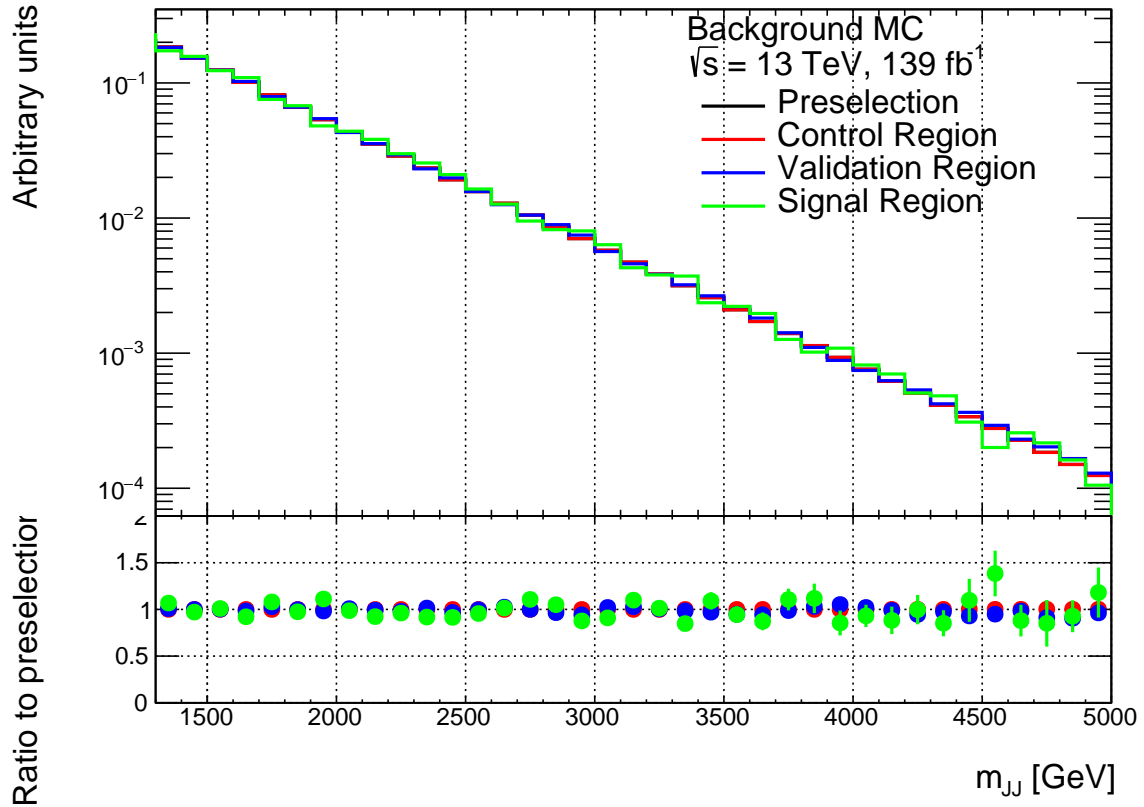


FIGURE 6.23 – Comparaison du spectre en masse invariante des simulations de bruit de fond dans différentes régions : la région « inclusive », la Région de Contrôle, la Région de Validation et la Région de Signal pour le bruit de fond simulé.

La Figure 6.24 représente la distribution en masse invariante de tous les signaux utilisés après que les coupures dans la Région de Signal ont été appliquées. On peut remarquer que le signal à 1,5 TeV pour le Modèle C est presque entièrement coupé, et que le pic de la distribution manque.



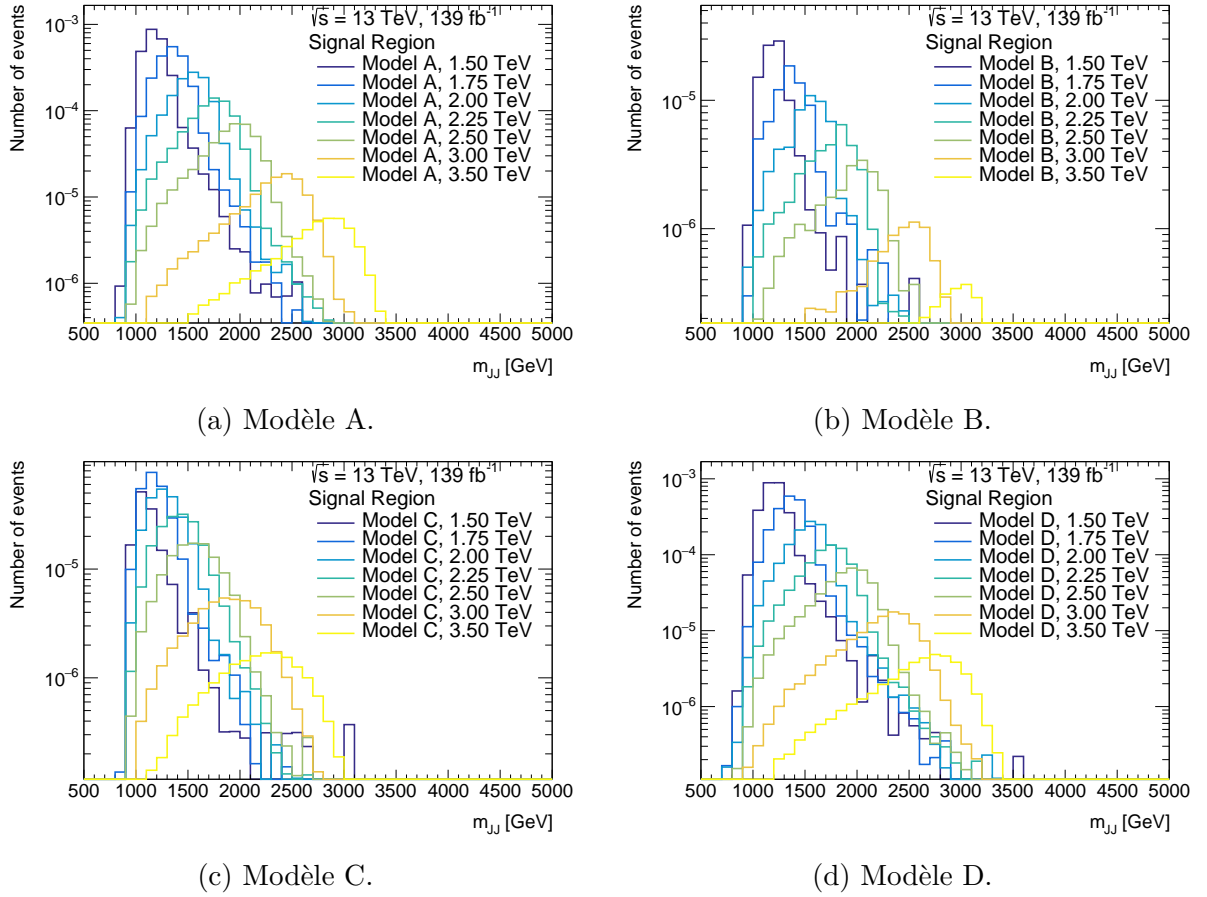


FIGURE 6.24 – Distribution en masse invariante des deux jets de plus haute impulsion pour toutes les masses du  $Z'$  de chaque modèle étudié dans la Région de Signal.

## 6.7 Comparaison des données et des simulations

La Figure 6.25 représente une comparaison entre les données et les simulations du bruit de fond de la distribution en masse invariante des deux jets de plus haute impulsion transverse après les coupures de pré-sélection. Les simulations du bruit de fond modélisent bien les données, puisque que le rapport entre les deux courbes est approximativement plat (la pente est égale à  $-2,82 \times 10^{-5} \pm 5,87 \times 10^{-7} \text{ GeV}^{-1}$ ). La normalisation des données n'est pas la même que dans les simulations de bruit de fond, mais ce n'est pas important pour l'analyse, qui n'utilise pas cette normalisation, comme il sera expliqué plus loin.

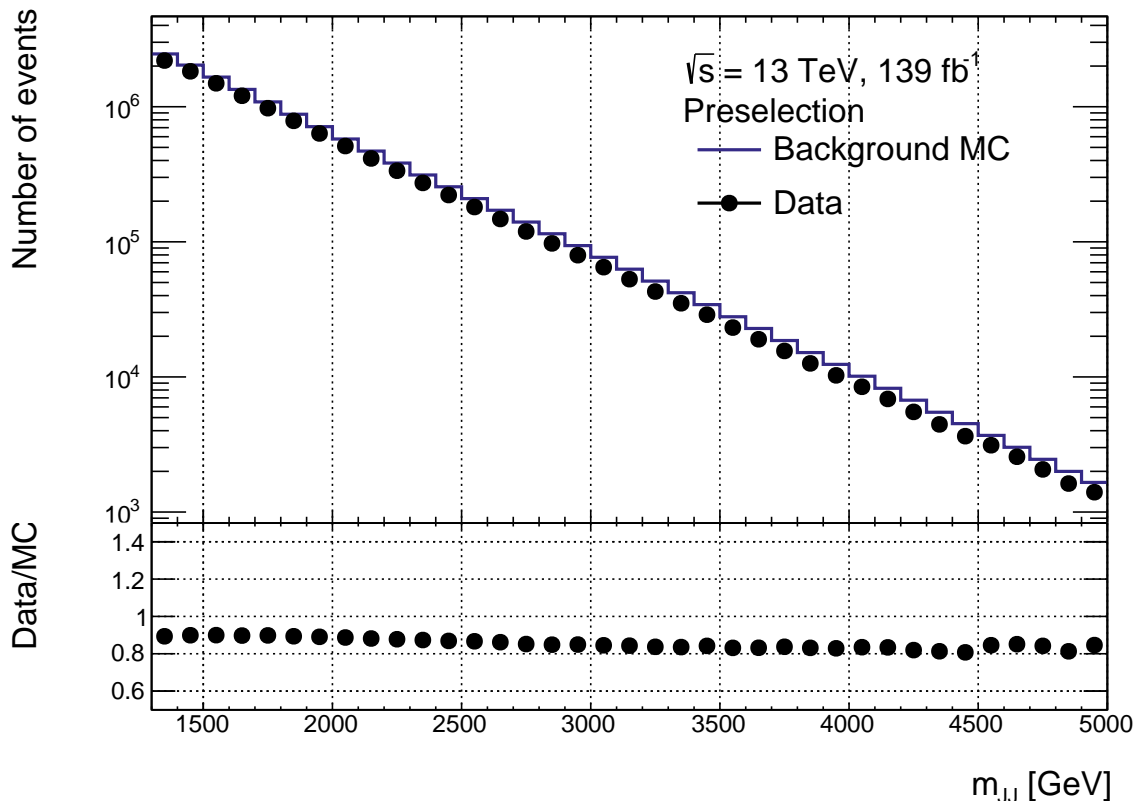
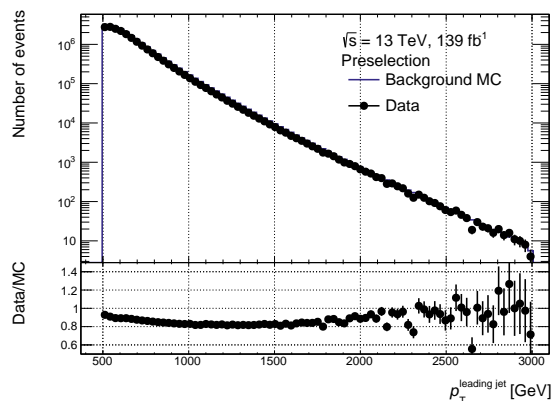
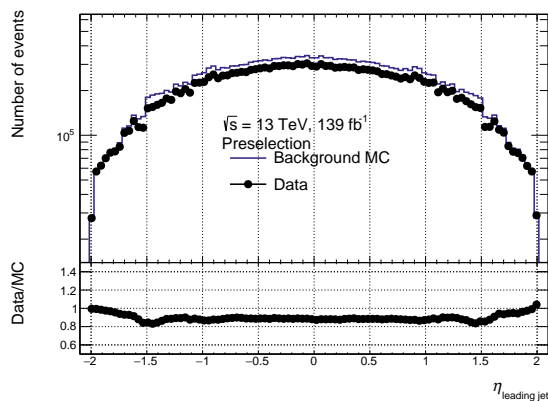
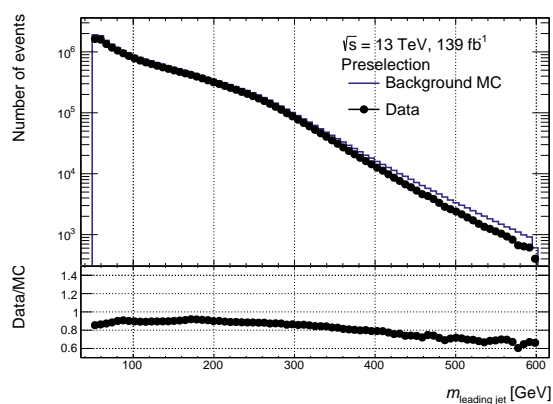


FIGURE 6.25 – Comparaison entre les données et les simulations du bruit de fond de la distribution en masse invariante des deux jets de plus haute impulsion transverse après les coupures de pré-sélection.

La Figure 6.26 représente une comparaison entre les données et les simulations du bruit de fond de la distribution en impulsion transverse,  $\eta$ , masse et du nombre de traces sans grooming qui satisfont  $p_T > 500$  MeV pour le jet de plus haute impulsion transverse après les coupures de pré-sélection. À noter que l'axe des ordonnées du panneau du bas est plus élevé pour le nombre de traces que pour les autres variables. L'accord entre les données et les simulations est plutôt bon pour  $p_T$ ,  $\eta$  et  $m$ , mais pas pour le nombre de traces sans grooming qui satisfont  $p_T > 500$  MeV, qui est la principale variable discriminante de l'analyse. Les conclusions sont similaires pour le second jet de plus haute impulsion transverse.



(a) Impulsion transverse.


 (b)  $\eta$ .


(c) Masse.

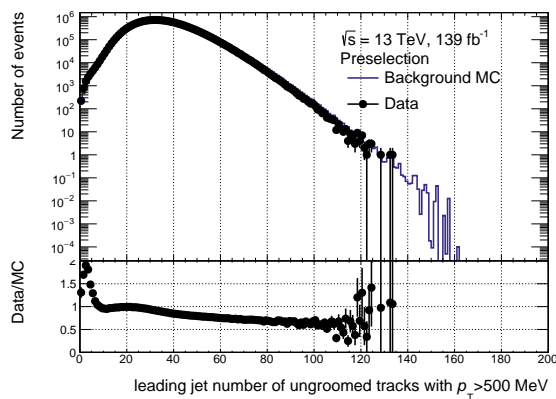

 (d) Nombre de traces sans grooming qui satisfont  $p_T > 500$  MeV du jet avec la plus haute impulsion transverse.

 FIGURE 6.26 – Comparaison entre les données et les simulations du bruit de fond de la distribution en  $p_T$ ,  $\eta$ ,  $m$  et du nombre de traces sans grooming qui satisfont  $p_T > 500$  MeV pour le jet de plus haute impulsion transverse.

La Figure 6.27 représente une comparaison entre les données et les simulations du bruit de fond de la distribution en nombre de traces sans grooming qui satisfont  $p_T > 500$  MeV pour le jet de plus haute impulsion transverse après les coupures de pré-sélection. Une enveloppe des différentes incertitudes systématiques associées aux variations internes des poids de Pythia (voir la Section 6.9) est représentée sur la Figure, et permet de couvrir la différence entre les données et les simulations, sauf à très bas nombre de traces. Cela ne devrait pas influencer l'analyse, étant donné que cela ne concerne qu'un petit nombre d'évènements.

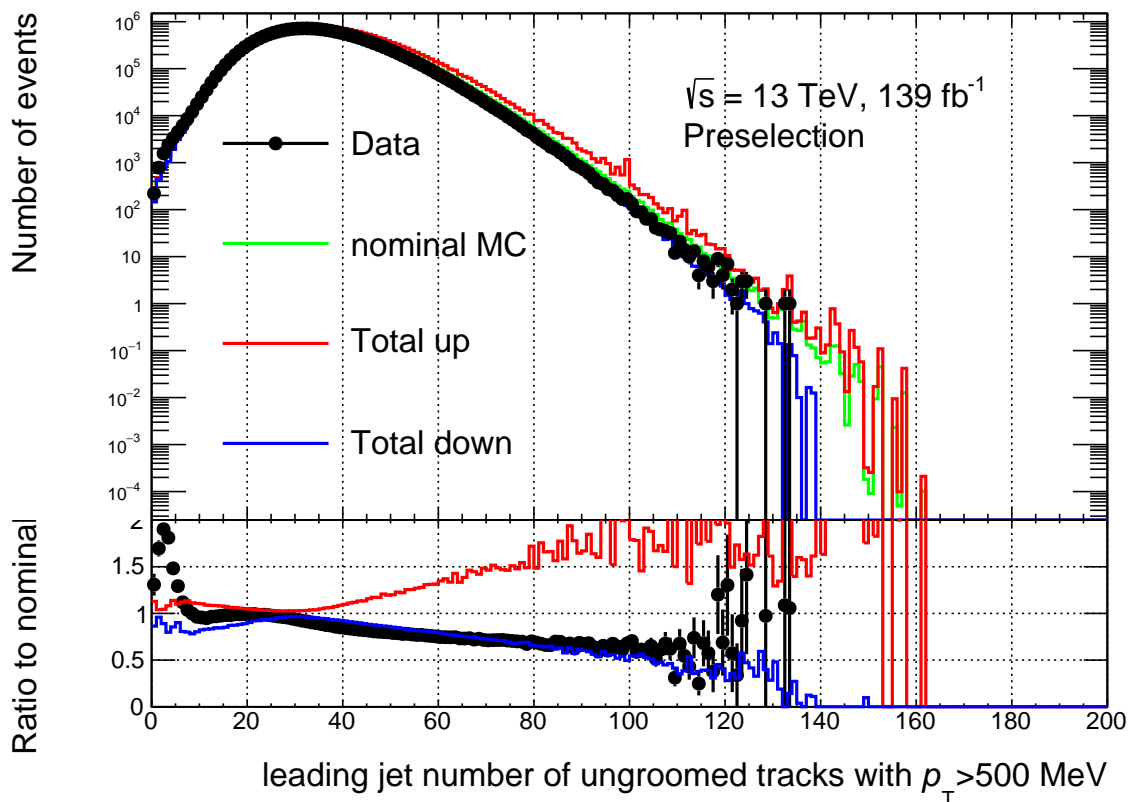


FIGURE 6.27 – Comparaison entre les données et les simulations du bruit de fond de la distribution en nombre de traces sans grooming qui satisfont  $p_T > 500$  MeV pour le jet de plus haute impulsion transverse.

## 6.8 Effet de $n_{\text{trk}}^\varepsilon$ dans les données

Le bruit de fond du Modèle Standard est estimé empiriquement à partir de la distribution en masse invariante dans les données réelles avec les coupures correspondant à la Région de Contrôle appliquées. Comme décrit dans la Section 6.6, cette Région de Contrôle est définie en exigeant que  $n_{\text{trk}}^\varepsilon < 0$  pour les deux jets, alors que la Région de Signal est définie par  $n_{\text{trk}}^\varepsilon > 0$  pour les deux jets. La Région de Contrôle est donc une région avec une faible contamination en signal, puisqu'elle correspond à une coupure à bas nombre de traces et que la plupart des événements de bruit de fond persistent. Par définition de  $n_{\text{trk}}^\varepsilon$ , la Région de Contrôle a, en théorie, la même distribution en masse invariante que le bruit de fond provenant du Modèle Standard dans la Région de Signal. C'est pourquoi cette analyse repose sur un bon accord entre la distribution en masse invariante dans la Région de Contrôle et celle du bruit de fond dans la Région de Signal.

Cependant, comme décrit dans la Section 6.6, les percentiles sont définis à partir des simulations du bruit fond, puis appliqués aux données réelles. Or le nombre de traces est une variable qui n'est pas très bien modélisée par les simulations, comme décrit dans la Section 6.7 (bien que le désaccord soit couvert par les incertitudes systématiques sur les simulations du bruit de fond). Ainsi, la dé-corrélation entre le nombre de traces et la distribution en masse invariante peut être imparfaite lorsque les coupures en  $n_{\text{trk}}^\varepsilon$  sont

appliquées aux données réelles, ce qui peut mener à une distribution en masse invariante qui est différente entre la Région de Contrôle et la Région de Signal.

Afin de tester cette idée, les coupures correspondant à la Région de Validation (voir Section 6.6) sont appliquées aux données réelles. Cette Région de Validation représente une région intermédiaire entre la Région de Contrôle et la Région de Signal, et permet de tester la méthode, sans regarder les données dans la Région de Signal (stratégie de « *blinding* »). La Figure 6.28 représente la distribution en masse invariante du rapport dans la Région de Validation et la Région de Contrôle, et ce pour les données réelles et les simulations du bruit de fond. Par construction de la méthode, ces deux courbes sont censées être plates, ce qui voudrait dire que la distribution en masse invariante a la même forme dans la Région de Contrôle et dans la Région de Validation. On peut remarquer que la courbe est plus plate pour les simulations du bruit de fond que pour les données réelles, ce qui est dû à une modélisation imparfaite des données par les simulations, puisque les percentiles sont définis à partir des simulations. Cependant, la courbe est aussi à une bonne approximation plate dans les données, ce qui indique que la méthode fonctionne aussi dans les données, bien que de façon imparfaite. Ces deux courbes sont ajustées avec une fonction constante, afin d'obtenir la valeur de l'efficacité après la coupure de la Région de Validation. Une fonction constante est choisie car ces deux courbes sont censées être plates. Les valeurs de l'ajustement avec la constante sont indiquées sur la Figure. On peut remarquer que l'efficacité pour la courbe des simulations n'est pas à 1%, mais à une valeur plus basse, dont la raison a déjà été expliquée dans la Section 6.6. Enfin, on peut remarquer que l'efficacité de sélection est plus basse dans les données, ce qui est aussi dû à une modélisation imparfaite des données par les simulations. Cette différence en efficacité de coupure est utilisée par la suite pour réduire le nombre d'évènements des simulations lorsque des coupures sur le nombre de traces sont appliquées. Puisque la courbe est moins plate pour les données que pour les simulations, une fonction affine a aussi été utilisée pour l'ajuster, à la place d'utiliser simplement une constante. La valeur de la pente de la fonction affine obtenue avec cet ajustement est égale à  $3,82 \times 10^{-5} \pm 1,20 \times 10^{-4} \text{ GeV}^{-1}$ . Ainsi, cette pente n'est pas significative, puisque la valeur nulle est couverte par son erreur. Par conséquent, aucune incertitude systématique n'est ajoutée.

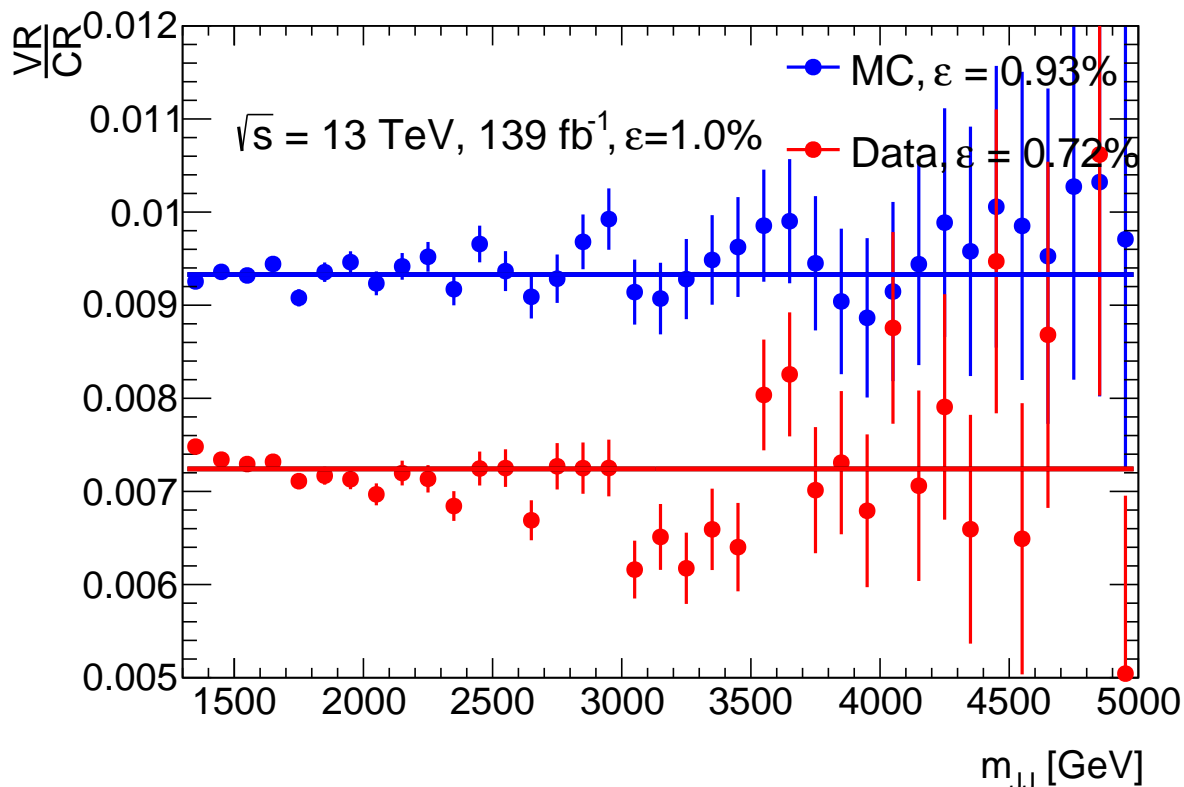


FIGURE 6.28 – Distribution en masse invariante du rapport entre la Région de Validation et la Région de Contrôle pour les données réelles et les simulations du bruit de fond.

## 6.9 Incertitudes systématiques

Les incertitudes systématiques sont toutes les incertitudes qui ne sont pas directement dues à la statistique des données, c'est-à-dire une variation inconnue possible d'une mesure, qui ne varie pas de façon aléatoire d'un ensemble de données à l'autre. Les incertitudes systématiques reflètent donc une connaissance imparfaite des données. Elles sont évaluées et exprimées en termes de paramètres de nuisance dans l'analyse statistique, comme indiqué dans la Section 6.11.1. Des incertitudes sur les signaux et sur le bruit de fond sont considérées.

Deux types d'incertitudes sont considérées dans cette analyse :

- les incertitudes expérimentales : sur l'échelle en énergie et en masse des jets, sur la résolution en énergie des jets, sur l'efficacité et le taux de fausses traces pour les traces à l'intérieur des jets, et sur la valeur de la luminosité intégrée,
- les incertitudes théoriques : sur la valeur de  $\alpha_s$ , sur la collection de *Parton Distribution Functions (PDFs)*, sur les variations internes des poids de Pythia.

Ces deux types d'incertitudes sont décrits plus en détail ci-dessous.

Comme les simulations de bruit de fond ne sont utilisées que pour dé-corréler le nombre de traces de la distribution en masse invariante, afin de s'assurer que la forme du bruit de fond dérivée à partir des données dans la Région de Contrôle puisse être utilisée dans la Région de Signal, les incertitudes systématiques sur ces simulations n'auront un impact sur l'analyse que si elles changent l'accord de cette forme entre la Région de Contrôle et la Région de Signal.

Pour toutes les incertitudes systématiques (en particulier celles des simulations de bruit de fond) décrites ci-dessous, lorsque les coupures correspondant à la Région de Signal sont appliquées, les courbes des percentiles utilisés pour calculer la valeur de  $n_{\text{trk}}^\varepsilon$  sont ceux dérivés à partir des simulations de bruit de fond sans aucune incertitude systématique, c'est-à-dire, ceux de la Figure 6.20a. En effet, pour chaque incertitude systématique, les percentiles pourraient être dérivés, puis utilisés afin de calculer la valeur de  $n_{\text{trk}}^\varepsilon$ . Cependant, par construction de la méthode, la dé-corrélation entre le nombre de traces et la masse invariante fonctionnerait. Cela ne permettrait donc pas d'utiliser les incertitudes systématiques (en particulier celles des simulations de bruit de fond) afin de calculer l'impact du fait d'utiliser les percentiles dérivés à partir des simulations de bruit de fond pour les données. C'est pourquoi, pour chaque incertitude systématique, les percentiles utilisés pour calculer la valeur de  $n_{\text{trk}}^\varepsilon$  sont ceux dérivés à partir des simulations de bruit de fond sans aucune incertitude systématique.

## 6.9.1 Incertitudes expérimentales

### 6.9.1.1 Incertitudes liées à l'étalonnage des jets

La correction « *in-situ* » a pour objectif de corriger les différences en échelle ou résolution en énergie ou en masse après étalonnage entre les données et les simulations.

Pour évaluer la différence d'échelle en énergie, l'équilibre entre les jets de rayon  $R = 1,0$  par rapport à un objet de référence bien défini est comparé entre les données et les simulations, et est utilisé pour corriger les éventuelles différences entre les deux. Il existe des incertitudes sur ces corrections provenant de la précision avec laquelle les facteurs de correction sont dérivés. Il s'agit notamment de sources liées aux effets du détecteur, aux sélections appliquées, à la statistique et à la modélisation des simulations. Ces variations constituent la grande majorité des composantes de l'incertitude sur l'échelle en énergie. La Figure 6.29 [89] représente le rapport entre les données et les simulations du rapport de l'impulsion transverse du jet et l'impulsion transverse de l'objet de référence en fonction de l'impulsion transverse du jet. L'erreur systématique totale sur l'impulsion transverse du jet est affichée en vert.

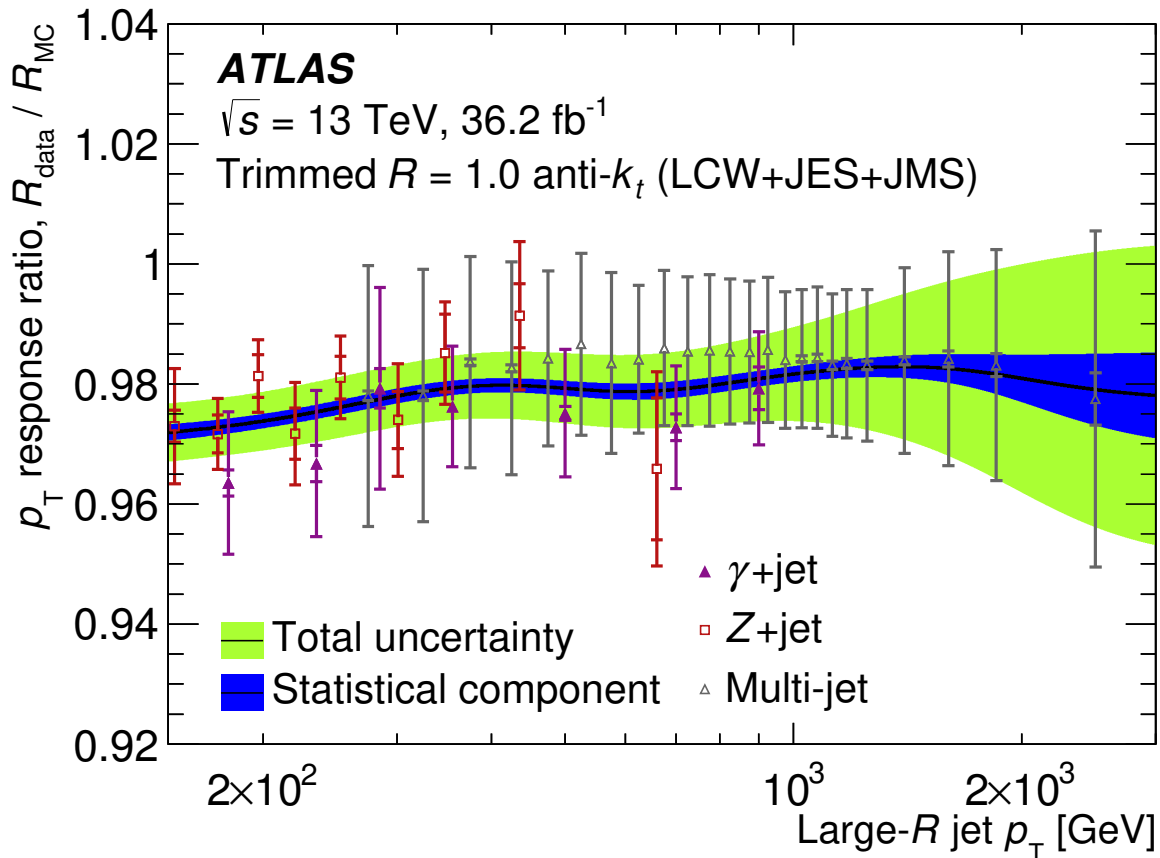


FIGURE 6.29 – Rapport entre les données et les simulations du rapport de l’impulsion transverse du jet et l’impulsion transverse de l’objet de référence en fonction de l’impulsion transverse du jet [89]. L’erreur systématique totale sur l’échelle en impulsion transverse du jet est affichée en vert.

D’autres incertitudes systématiques, qui ne peuvent pas être prises en compte par la procédure *in-situ*, sont appliquées sur l’échelle en énergie des jets, parmi lesquelles :

- Incertitude sur la réponse en fonction de la saveur : cette incertitude tient compte de la différence en réponse d’un jet initié par un quark par rapport à un jet initié par un gluon, et en particulier de la façon dont elle est modélisée dans différents générateurs. En général, la réponse des jets initiés par des quarks est très bien modélisée par les générateurs, tandis que la réponse des jets initiés par des gluons peut varier de manière significative entre générateurs. Cette incertitude est donnée par :

$$\sigma_{\text{response}} = f_g (\mathcal{R}_g^{\text{Pythia}} - \mathcal{R}_g^{\text{Alternative}}) \quad (6.10)$$

où :

- $f_g$  représente la fraction de jets initiés par des gluons dans l’analyse,
- $\mathcal{R}_g^{\text{Pythia}}$  représente la réponse en  $p_T$  pour des jets initiés par des gluons avec le générateur Pythia,
- $\mathcal{R}_g^{\text{Alternative}}$  représente la réponse en  $p_T$  pour des jets initiés par des gluons avec un générateur alternatif à Pythia (généralement Herwig).
- Incertitude sur la composition en saveurs : étant donné que la modélisation de la



réponse des jets initiés par des quarks et par des gluons diffère, il est important de prendre également en compte que la fraction de jets initiés par des quarks et des gluons peut varier. Cette incertitude est donnée par :

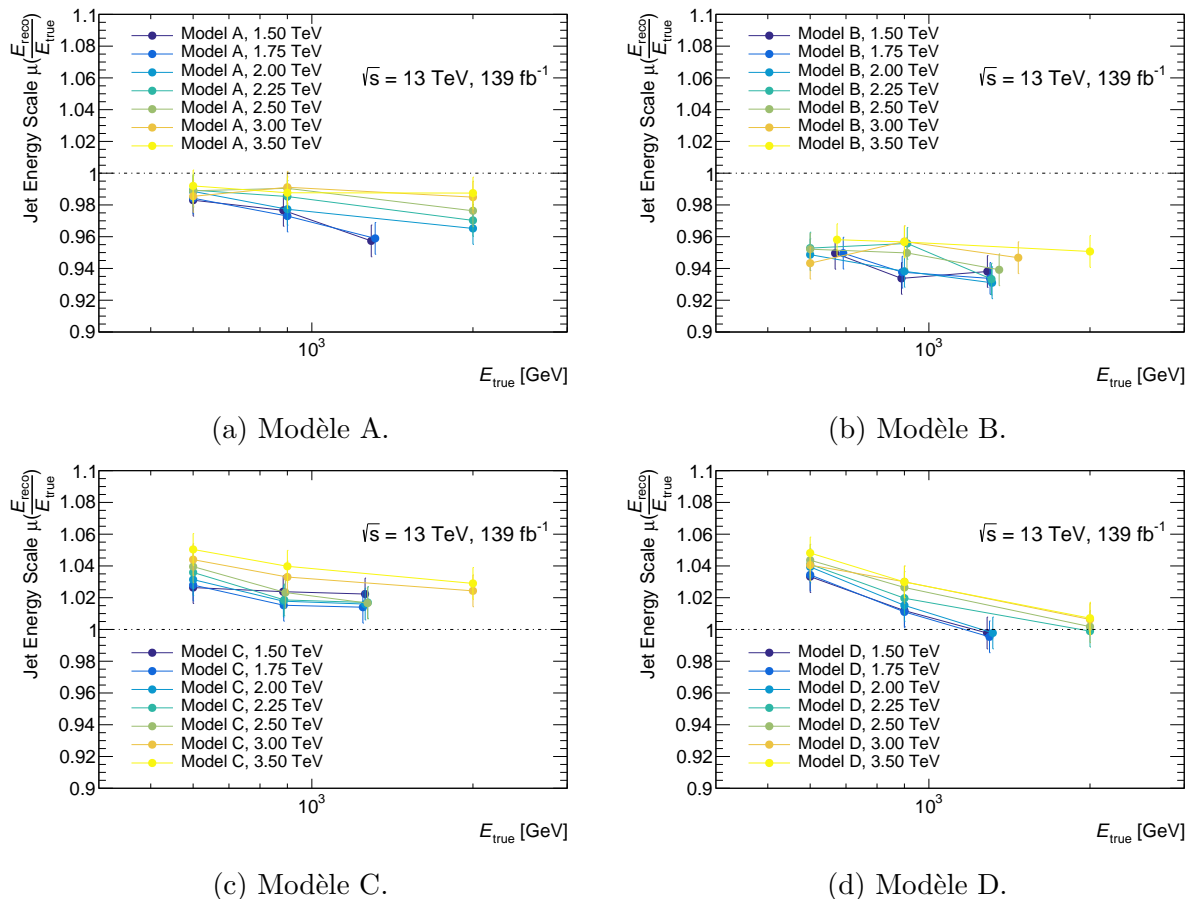
$$\sigma_{\text{composition}} = \sigma_{f_g} \times \frac{|\mathcal{R}_g - \mathcal{R}_q|}{f_g \times \mathcal{R}_g + (1 - f_g) \times \mathcal{R}_q} \quad (6.11)$$

où :

- $\sigma_{f_g}$  représente une incertitude sur la fraction de jets initiés par des gluons dans l'analyse, généralement obtenue en comparant la fraction de jets initiés par des gluons avec Pythia et Herwig,
- $\mathcal{R}_g$  représente la réponse en  $p_T$  pour des jets initiés par des gluons avec le générateur Pythia,
- $\mathcal{R}_q$  représente la réponse en  $p_T$  pour des jets initiés par des quarks avec le générateur Pythia.

Ces incertitudes s'appliquent sur la valeur nominale de l'échelle en énergie, c'est-à-dire l'échelle en énergie moyenne, et donc décalent le spectre en énergie des jets.

Les jets de QCD sombre sont étalonnés avec les mêmes facteurs de correction que les jets provenant du Modèle Standard. Or les jets de QCD sombre sont différents des jets du Modèle Standard (voir par exemple la Table 6.1), et par conséquent il n'est pas évident que ces facteurs permettent de corriger efficacement l'échelle en énergie des jets de QCD sombre. Afin de vérifier que les facteurs de correction dérivés à partir des jets du Modèle Standard permettent de restaurer l'énergie des jets de QCD sombre à l'énergie des jets *vrais* qui leur sont associés, le rapport entre l'énergie des jets de QCD sombre après étalonnage (dérivé à partir des jets du Modèle Standard) et l'énergie du jet *vrai* associé est ajusté avec une gaussienne en intervalles de l'énergie du jet *vrai*. Ensuite la moyenne de l'ajustement avec la gaussienne est tracée en fonction de l'énergie du jet *vrai*. C'est ce qui est représenté sur la Figure 6.30, pour chacun des Modèles étudiés et pour toutes les masses du  $Z'$ . Sur cette Figure, les coupures correspondant à la Région de Signal sont appliquées aux signaux. Le trimming est aussi appliqué aux jets *vrais* (car c'est le cas dans la procédure d'étalonnage standard d'ATLAS), et c'est pourquoi l'échelle en énergie est parfois supérieure à 1, alors que la distribution en masse invariante reconstruite pique toujours à une masse inférieure à celle du  $Z'$ . On peut remarquer que l'échelle en énergie dépend très fortement du modèle considéré. Pour tous les modèles, les courbes ne sont pas à 1, ce qui indiquerait que l'étalonnage, dérivé à partir des jets de QCD du Modèle Standard, ne permet pas de corriger efficacement l'énergie des jets de QCD sombre. C'est pourquoi, une incertitude systématique, donnée par la différence entre ces courbes et 1, sera ajoutée. Cependant, cette incertitude systématique n'est pour le moment pas mise en place pour les études de la Section 6.11.

FIGURE 6.30 – Échelle en énergie après étalonnage en fonction de l'énergie du jet *vrai*.

Pour les jets de QCD sombre, l'incertitude sur la réponse en fonction de la saveur ne peut être évaluée, car le module *Hidden Valley* est spécifique à Pythia. Par conséquent, pour les jets de QCD sombre, la valeur utilisée est directement celle dérivée avec les jets de QCD du Modèle Standard.

Enfin, pour les jets de QCD sombre, l'incertitude sur la composition en saveurs n'est pas appliquée, étant donné que cela ne concerne que des jets initiés par des quarks ou des gluons. Or les jets de QCD sombre ne sont générés que par des quarks sombres (voir la Figure 2.11). Cependant, une incertitude sur la composition pourrait être appliquée pour les Modèles C et D, car la composition de ces jets peut changer d'un jet à l'autre (voir la Table 6.2), et donc la réponse peut changer d'un jet à l'autre. Or les valeurs de la Table 6.2 sont des valeurs présumées fixes : il n'y a pas d'incertitude associées à ces valeurs (comme  $\sigma_{f_g}$  dans l'équation 6.11). C'est pourquoi l'incertitude sur la composition en saveurs n'est pas appliquée sur les jets de QCD sombre.

Afin d'évaluer la différence de résolution en énergie entre les données et les simulations, l'asymétrie de l'équilibre entre les deux jets de plus haute impulsion transverse est utilisée. La différence en résolution en énergie entre les données et les simulations est considérée comme une incertitude, en plus de celles provenant des effets du détecteur, des sélections appliquées, de la statistique et de la modélisation. Enfin, les incertitudes sur l'échelle en

énergie sont propagées. La Figure 6.31 représente les différentes sources d'incertitudes sur la résolution en impulsion transverse [89]. Ces incertitudes sont appliquées en élargissant artificiellement les distributions en impulsion transverse, et par conséquent, elles modifient la largeur des signaux.

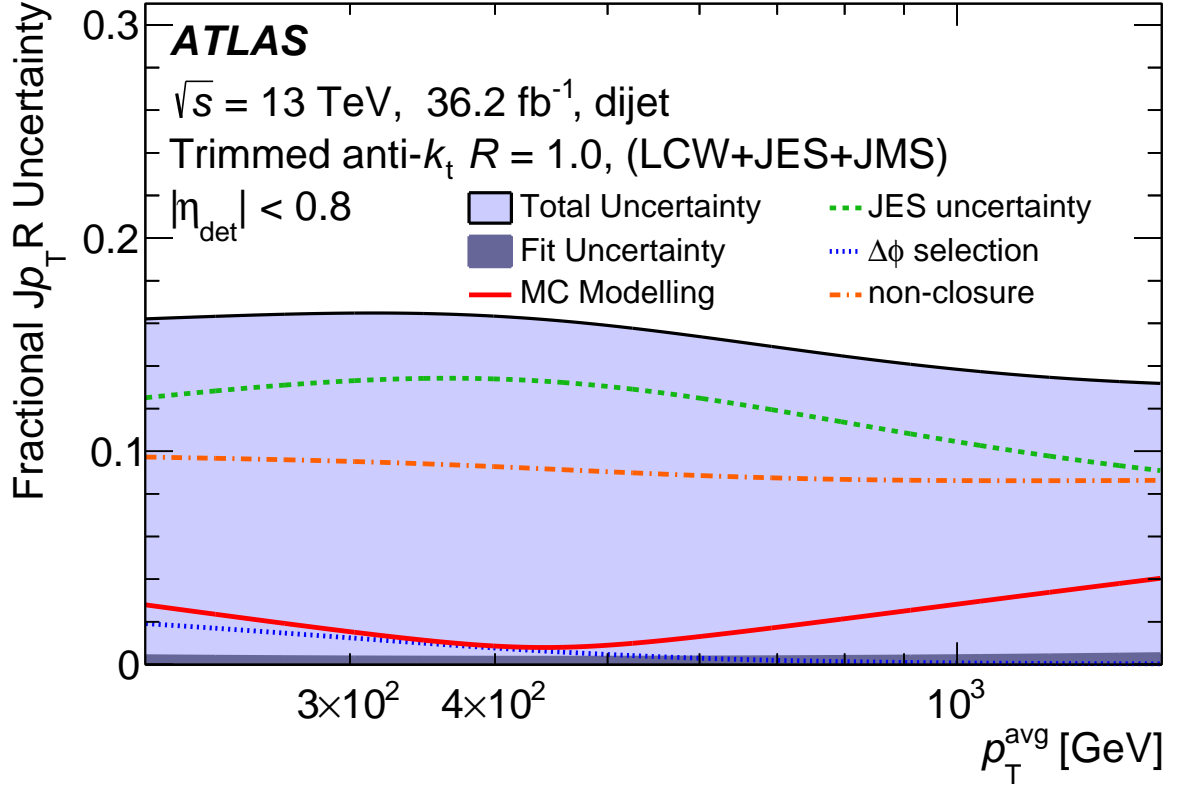


FIGURE 6.31 – Les différentes sources d'incertitudes sur la résolution en impulsion transverse [89].

Pour les jets de QCD sombre, la résolution en énergie est similaire à celle des jets de QCD du Modèle Standard. Par conséquent, aucune incertitude systématique supplémentaire n'est ajoutée pour les jets de QCD sombre.

Les dernières sources d'incertitudes liées à l'étalonnage des jets concernent l'échelle en masse. Ces incertitudes sont dérivées en utilisant deux méthodes. La première utilise un ajustement de la distribution en masse pour des événements contenant des bosons  $W$  ou des quarks  $t$ . La seconde exploite le double rapport entre la masse construite à partir des calorimètres et la masse construite à partir du détecteur interne entre les données et les simulations. Les incertitudes sur l'échelle en masse proviennent principalement d'effets du détecteur, de la sélection, de la statistique et de la modélisation. La Figure 6.32 représente le rapport entre les données et les simulations du rapport de la masse du jet et de la masse de référence en fonction de l'impulsion transverse du jet [89]. L'erreur systématique totale sur l'échelle en masse du jet est affichée en vert.

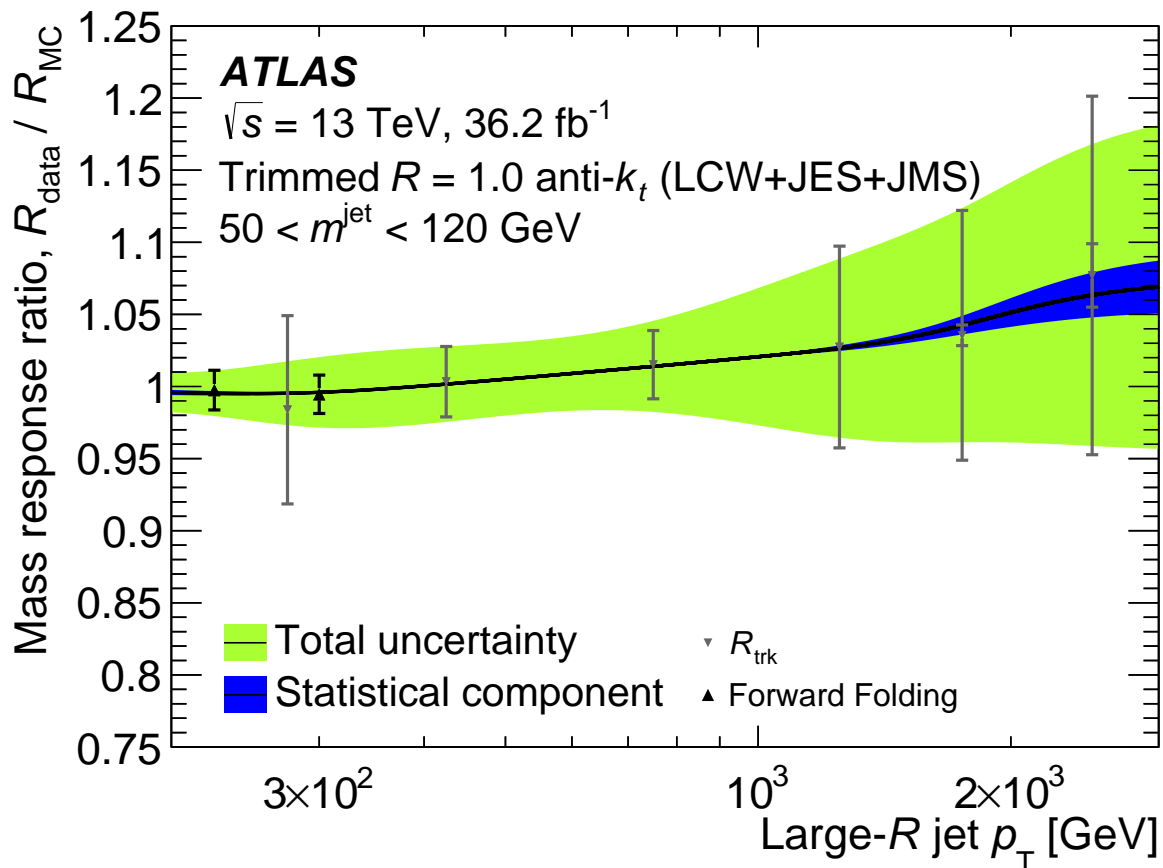
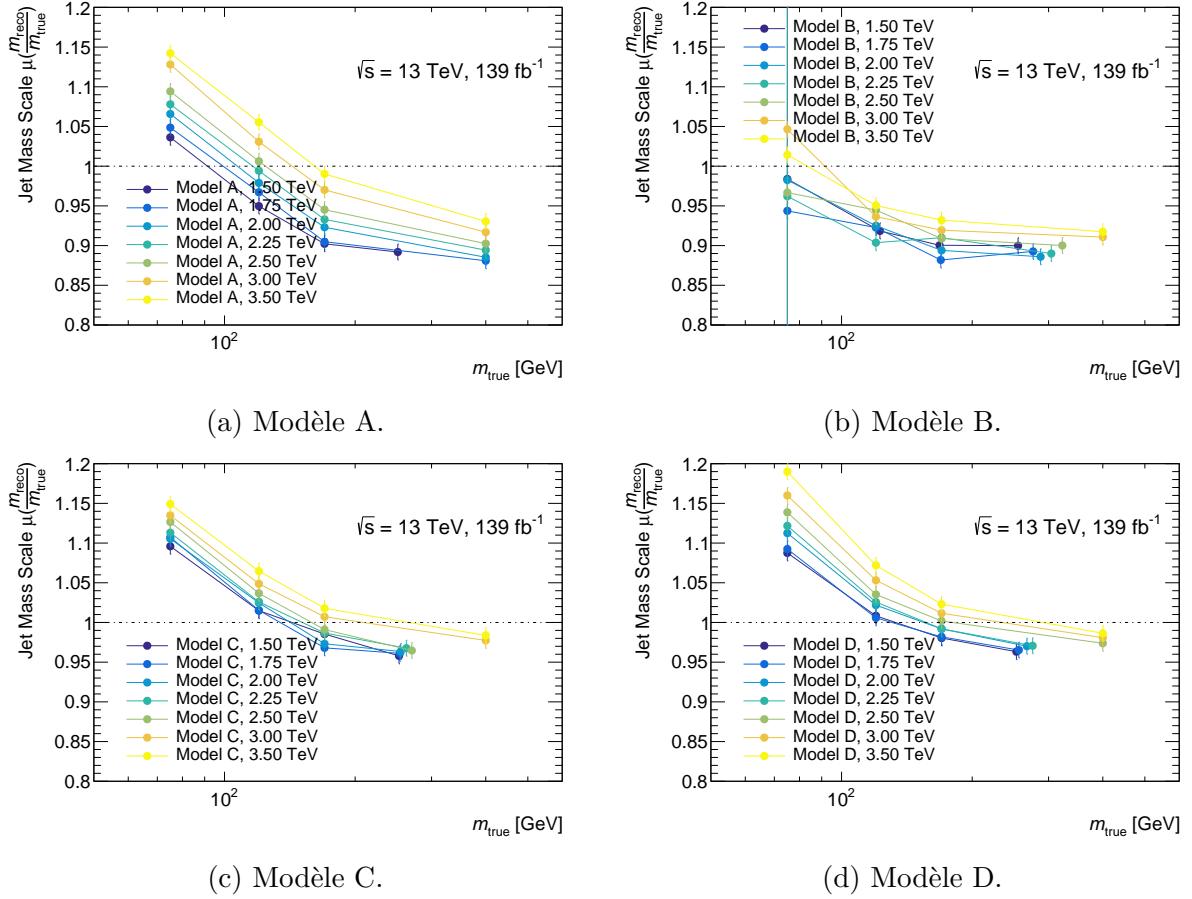


FIGURE 6.32 – Rapport entre les données et les simulations du rapport de la masse du jet et de la masse de référence en fonction de l’impulsion transverse du jet [89]. L’erreur systématique totale sur l’échelle en masse transverse du jet est affichée en vert.

Enfin, la Figure 6.33 est équivalente à la Figure 6.30, mais cette fois-ci, pour la masse. On peut remarquer que l’échelle en masse peut être très différente de 1, par exemple, pour le Modèle D. Une incertitude systématique qui permet de couvrir cette différence sera ajoutée. Cependant, elle n’est pas non plus pour le moment mise en place pour les études de la Section 6.11.


 FIGURE 6.33 – Échelle en masse après étalonnage en fonction de la masse du jet *vrai*.

### 6.9.1.2 Incertitudes liées à la reconstruction des traces au sein des jets

Le cœur des jets est un environnement dense où la séparation des particules chargées peut devenir de l'ordre de la granularité des capteurs du détecteur interne. Par conséquent, les amas de pixels dans le détecteur interne commencent à fusionner, ce qui résulte en une dégradation de la reconstruction des traces à l'intérieur des jets.

La Figure 6.34 [34] représente le nombre moyen de traces par unité de surface angulaire en fonction de la distance angulaire par rapport à l'axe du jet pour des données réelles et des simulations. Plusieurs intervalles en impulsion transverse du jet sont représentés. On peut remarquer que les traces sont très denses dans le cœur du jet, et que le nombre moyen de traces par unité angulaire augmente avec l'impulsion transverse du jet.

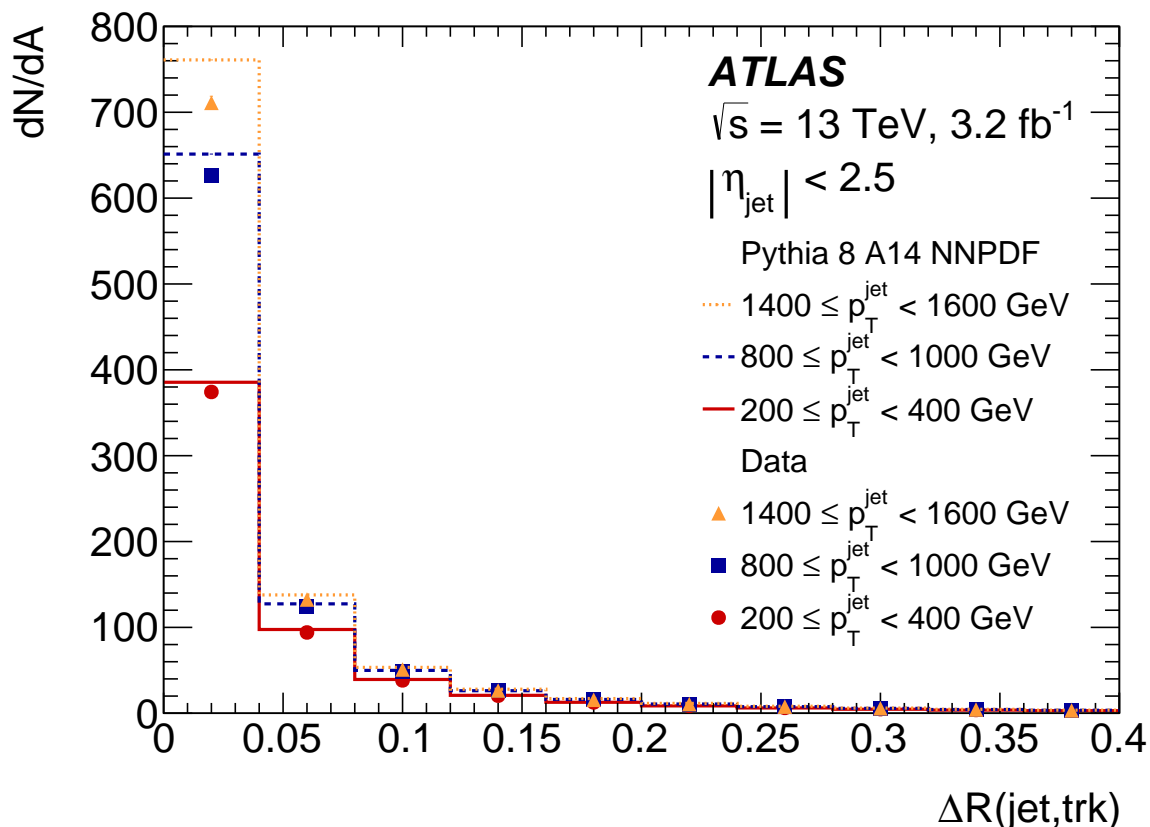


FIGURE 6.34 – Nombre moyen de traces par unité de surface angulaire en fonction de la distance angulaire par rapport à l’axe du jet pour des données réelles et des simulations [34]. Plusieurs intervalles en impulsion transverse du jet sont représentés.

L’efficacité de reconstruction des traces, c’est-à-dire le nombre de traces *vraies* divisé par le nombre de traces reconstruites, diminue. Cet effet est évalué en étudiant la perte d’énergie par ionisation ( $\frac{dE}{dx}$ ) dans les amas de pixels du détecteur interne. Les amas avec des dépôts d’énergie compatibles avec de multiples particules qui ne sont associées qu’à une seule trace sont utilisés pour déterminer la fraction de traces perdues qui peut être comparé dans les données réelles et dans les simulations. Ces études mènent à une incertitude systématique sur l’efficacité de reconstruction des traces estimée à 0,4%. Cette incertitude systématique ne dépend pas de l’impulsion transverse du jet considéré. Cette valeur est dérivée en utilisant des jets de rayon  $R = 0,4$ , alors que l’analyse utilise des jets de rayon  $R = 1,0$ . Mais cette incertitude systématique n’est appliquée qu’aux traces dans le coeur des jets ( $\Delta R(\text{jet}, \text{track}) < 0,1$ ). Vers l’extérieur des jet ( $\Delta R(\text{jet}, \text{track}) > 0,1$ ), cette incertitude n’est pas appliquée. De plus, le bruit de fond simulé correspond à des jets initiés par des quarks et de gluons, qui sont entièrement contenus par des jets de rayon  $R = 0,4$ . Ainsi, le nombre de traces qui satisfont  $0,4 < \Delta R(\text{jet}, \text{track}) < 1,0$  est très faible pour les jets de rayon  $R = 1,0$  pour les simulations de bruit de fond. Enfin c’est le nombre de traces sans grooming qui est utilisé par l’analyse comme principale variable discriminante, et non pas le nombre de traces avec grooming. Par conséquent, aucune trace n’est enlevée par l’algorithme de grooming, et donc la densité de traces qui satisfont  $\Delta R(\text{jet}, \text{track}) < 0,4$  est la même pour des jets de rayon  $R = 0,4$  et des jets

de rayon  $R = 1, 0$ . C'est pourquoi l'analyse peut utiliser directement les valeurs obtenues avec des jets de rayon  $R = 0, 4$ , même si des jets de rayon  $R = 1, 0$  sont utilisés.

Avec l'augmentation de la densité des jets, un autre effet apparaît : le nombre de *fausses* traces, c'est-à-dire les traces qui ne proviennent pas d'une particule chargée, mais d'une mauvaise combinaison d'amas de pixels dans le détecteur interne, augmente. Cet effet est estimé en construisant une Région de Contrôle enrichie en fausses traces, et en effectuant un ajustement pour extraire une estimation de la fraction de vraies et de fausses traces, qui peut être comparée entre les données réelles et les simulations. Le nombre de fausses traces est environ égal à 0,5%, et l'incertitude systématique sur ce nombre est environ égal à 25%.

Ces deux incertitudes systématiques s'appliquent en enlevant certaines traces à l'intérieur des jets aléatoirement. Par conséquent, seules les variations « *basses* » sont disponibles. Les variations « *hautes* » sont obtenues en symétrisant les variations basses par rapport aux distributions obtenues sans incertitudes systématiques appliquées.

La Figure 6.35 représente une comparaison de la distribution en masse invariante des simulations du bruit de fond dans la Région Signal avec les systématiques liées à la reconstruction des traces à l'intérieur des jets. On peut remarquer que l'ordre de grandeur de ces systématiques est plutôt faible (<5%).

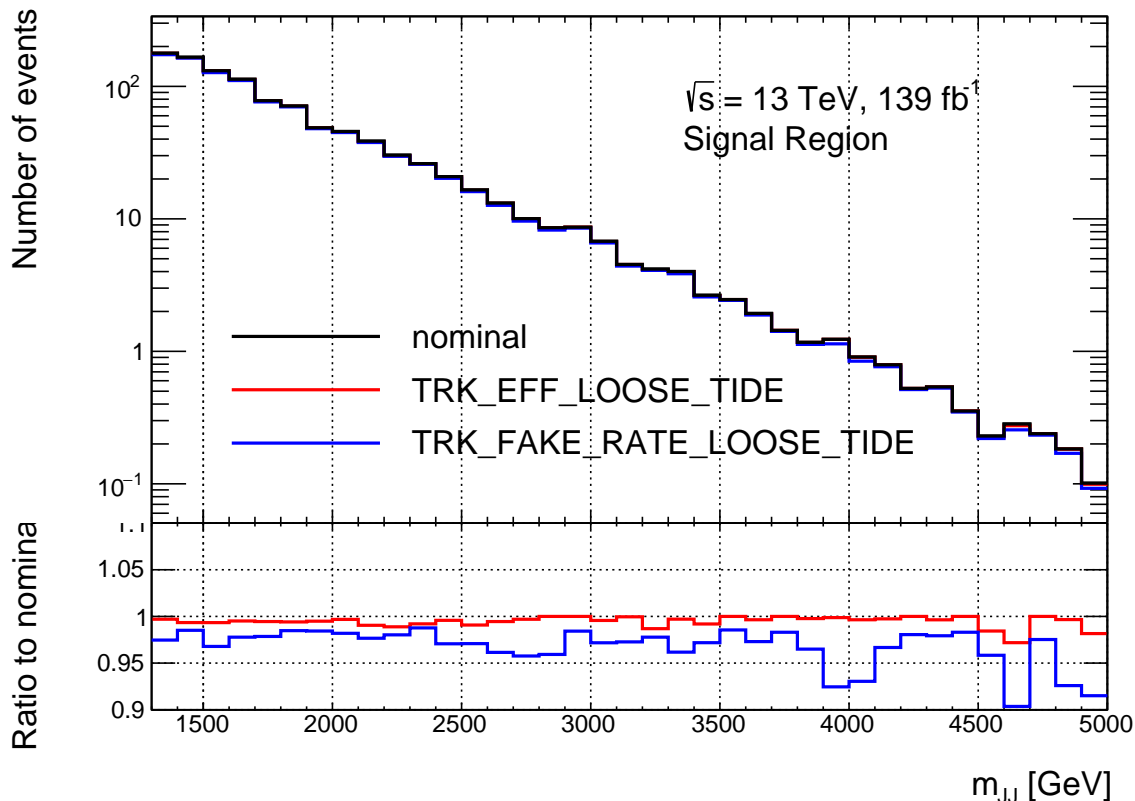


FIGURE 6.35 – Comparaison de la distribution en masse invariante des simulations du bruit de fond dans la Région Signal avec les systématiques liées à la reconstruction des traces à l’intérieur des jets.

Cependant, les jets de QCD sombre sont plus denses que les jets de QCD du Modèle Standard. Il n’est donc pas évident que les chiffres décrits ci-dessus, dérivés à partir de jets de QCD du Modèle Standard, puissent aussi s’appliquer aux jets de QCD sombre. Afin de tester cette idée, la Figure 6.34 a été reproduite pour deux signaux en particulier : Modèle B avec  $M_{Z'}$  = 1,5 TeV, qui est le Modèle avec le plus bas nombre de traces et Modèle D avec  $M_{Z'}$  = 3,5 TeV, qui est le Modèle avec le plus haut nombre de traces. C’est ce qui est représenté sur la Figure 6.36. À noter que l’axe des abscisses utilise la rapidité à la place de la pseudo-rapidité, car les jets de rayon  $R = 1,0$  sont supposés être massifs. En comparant avec la Figure 6.34, et en particulier à la gamme de données de l’axe des ordonnées, on peut remarquer que les jets de QCD sombre sont beaucoup plus denses que les jets de QCD du Modèle Standard. Par conséquent, les valeurs dérivées à partir des jets de QCD du Modèle Standard pourraient être sous-estimées pour les jets de QCD sombre. Cependant, l’impact de cette incertitude systématique est plutôt petit comparé à d’autres incertitudes systématiques. Donc, même si cette systématique était doublée, l’impact sur l’analyse serait négligeable. De plus, même pour les jets de QCD du Modèle Standard, les valeurs sont indépendantes de l’impulsion transverse du jet, qui est une quantité corrélée avec le nombre de traces. Par conséquent, nous n’avons pas dérivé de valeurs spéciales pour les jets de QCD sombre.



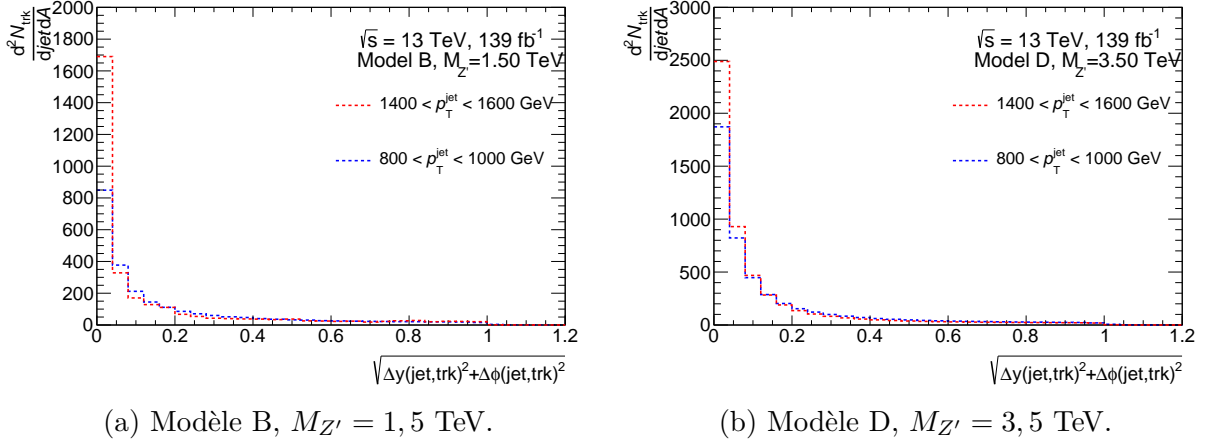


FIGURE 6.36 – Nombre moyen de traces par unité de surface angulaire en fonction de la distance angulaire par rapport à l'axe du jet pour deux signaux. Plusieurs intervalles en impulsion transverse du jet sont représentés.

La Figure 6.37 représente une comparaison de la distribution en masse invariante de ces deux signaux dans la Région Signal avec les systématiques liées à la reconstruction des traces à l'intérieur des jets. On peut remarquer que l'impact de ces incertitudes systématiques est plutôt faible ( $<2\%$ ).

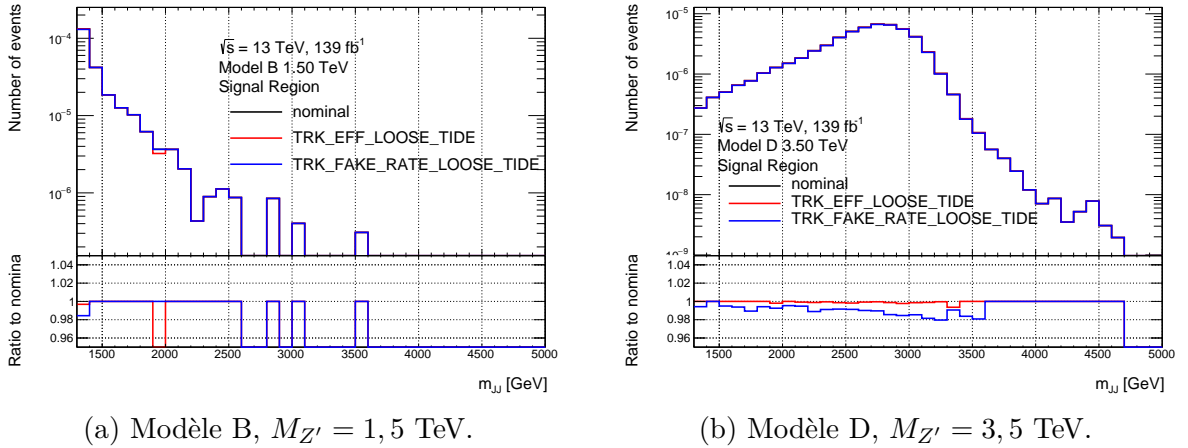


FIGURE 6.37 – Comparaison de la distribution en masse invariante de deux signaux dans la Région Signal avec les systématiques liées à la reconstruction des traces à l'intérieur des jets.

### 6.9.1.3 Incertitude sur la luminosité intégrée

L'incertitude sur la luminosité intégrée combinée des années 2015 à 2018 est de 1,7% [90]. Cette valeur est obtenue en utilisant le détecteur LUCID-2 [91]. Cette incertitude n'est appliquée qu'aux signaux, étant donné que la normalisation du bruit de fond est définie dans les données réelles.

## 6.9.2 Incertitudes théoriques

Toutes les incertitudes décrites dans cette section sont implémentées en modifiant les poids des évènements simulés  $w_{MC}$  (voir la Section 3.5).

En théorie des perturbations, la section efficace à l'ordre  $n$  est donnée de manière schématique par :

$$\sigma^{(n)} = \text{PDF}(x_1, \mu_F) \times \text{PDF}(x_2, \mu_F) \times \hat{\sigma}^{(n)}(x_1, x_2, \mu_R) \quad (6.12)$$

où :

- $\sigma^{(n)}$  représente la section efficace totale à l'ordre  $n$ ,
- PDF représente la *Parton Distribution Function*,
- $x_1$  et  $x_2$  représentent le Bjorken- $x$  des partons entrant,
- $\mu_F$  représente le facteur de factorisation,
- $\hat{\sigma}^{(n)}$  représente la section efficace :

$$\hat{\sigma}^{(n)} = \alpha_s \times \hat{\sigma}^{(0)} + \alpha_s^2 \times \hat{\sigma}^{(1)} + \dots + \alpha_s^n \times \hat{\sigma}^{(n-1)} + \mathcal{O}(\alpha_s^{n+1}), \quad (6.13)$$

- $\mu_R$  représente le facteur de renormalisation.

Les PDF sont estimées en supposant une fonction donnée pour le contenu en partons du proton à une échelle  $Q$  donnée. Les équations de DGLAP sont ensuite utilisées pour estimer les PDF à différentes échelles :

$$\frac{d\text{PDF}(x, Q^2)}{d \log(Q^2)} = P(\alpha_s, z) \times \text{PDF}\left(\frac{x}{z}, Q^2\right) \quad (6.14)$$

où :

$$P(\alpha_s, z) = \alpha_s \times P^{\text{LO}} + \alpha_s^2 \times P^{\text{NLO}} + \dots \quad (6.15)$$

La constante de couplage de l'interaction forte est également déterminée expérimentalement et est généralement donnée à l'échelle  $Q = m_Z$ . L'équation du groupe de renormalisation est ensuite utilisée dans les générateurs Monte-Carlo pour prédire la valeur de  $\alpha_s$  à différentes échelles d'énergie :

$$\frac{d\alpha_s(Q)}{d \log(Q^2)} = -b_0 \times \alpha_s^2(Q) - b_1 \times \alpha_s^3(Q) + \dots \quad (6.16)$$

Ainsi, il y a majoritairement 4 sources d'incertitudes dans les calculs en théorie des perturbations :

- Termes d'ordres supérieurs manquants dans l'expansion perturbative de la section efficace : ces incertitudes fournissent une estimation du terme  $\mathcal{O}(\alpha_s^{n+1})$  de l'Équation 6.13. La manière habituelle de les estimer est d'effectuer des variations d'échelle, c'est-à-dire de faire varier les facteurs de renormalisation et de factorisation d'un facteur deux. Ces variations sont implémentées en faisant varier les facteurs de renormalisation et de factorisation indépendamment :

$$\{\mu_R, \mu_F\} = \{1, 1\}, \{1, 0, 5\}, \{0, 5, 1\}, \{2, 1\}, \{1, 2\} \quad (6.17)$$

En général, les variations d'échelle entraînent une variation de la section efficace et une variation de la forme différentielle des observables. L'incertitude systématique totale est donnée par l'enveloppe de toutes les incertitudes systématiques individuelles. Ces incertitudes ne sont pour l'instant disponibles que pour les simulations du bruit de fond.

- Incertitudes sur les PDFs, c'est-à-dire les incertitudes expérimentales entrant dans les données utilisées et sur la fonction utilisée dans les ajustements des PDFs. Ces incertitudes sont incorporées dans des *collections de PDFs*. Le générateur utilisé pour les simulations de bruit de fond est Pythia, qui utilise la collections de PDFs NNPDF2.3LO. Cette collection est dénotée *NNPDF23\_lo\_as\_0130\_qed* [92], où « 0130 » signifie que la valeur de  $\alpha_s$  utilisée est 0,130. Cette collection de PDFs contient  $N = 100$  différentes valeurs avec lesquelles les poids des événements simulés doivent être modifiés, en plus de la valeur centrale. L'incertitude totale correspond alors à l'écart type de ces 100 variations :

$$\Delta X = \sqrt{\frac{1}{N} \sum_i (X_i - X_0)^2} \quad (6.18)$$

où  $X$  représente l'observable considérée,  $X_0$  représente la valeur centrale de cette observable, et les  $X_i$  représentent la valeur de l'observable pour les 100 variations.

- Incertitudes sur la valeur de  $\alpha_s$ . Ces incertitudes proviennent des erreurs expérimentales sur la détermination de la valeur de  $\alpha_s$ . Ces incertitudes sont estimées en utilisant la même collection de PDFs, mais avec une différente valeur de  $\alpha_s$ . Pour ce faire, la collection de PDFs utilisée, dénotée *NNPDF23\_lo\_as\_0119\_qed*, utilise une valeur de  $\alpha_s = 0,119$ . Ensuite l'incertitude liée à la valeur de  $\alpha_s$  est calculée en prenant la différence entre l'écart type des deux collections de PDFs :

$$\Delta\alpha_s = |\Delta X_{\alpha_s=0,130} - \Delta X_{\alpha_s=0,119}| \quad (6.19)$$

L'incertitude évaluée de cette manière ne correspond alors qu'à une variation basse. La variation haute est obtenue en symétrisant la variation basse par rapport à la valeur centrale. Enfin les incertitudes sur les PDFs et sur la valeur de  $\alpha_s$  peuvent être ajoutées en quadrature :

$$\Delta(\text{PDF} + \alpha_s) = \sqrt{(\Delta X_{\alpha_s=0,130})^2 + (\Delta\alpha_s)^2} \quad (6.20)$$

- Incertitudes algorithmiques ou paramétriques sur la gerbe de partons, sur l'hadronisation et sur la collections de paramètres A14. Deux incertitudes en particulier sont considérées : « *VAR3c* », qui correspond à la variation de la constante de couplage de l'interaction forte dans la gerbe de l'état initial, et « *cns* » qui correspond à des termes additifs non singuliers dans les fonctions de fractionnement. Ces incertitudes ne sont pour l'instant disponibles que pour les simulations du bruit de fond.

La Figure 6.38 représente une comparaison entre les données et les simulations du bruit de fond de la distribution en nombre de traces sans grooming qui satisfont  $p_T > 500$  MeV pour le jet de plus haute impulsion transverse après les coupures de pré-sélection. Les différentes systématiques liées aux termes d'ordres supérieurs manquants dans l'expansion perturbative de la section efficace ( $\mu_R$ ), ainsi que les incertitudes algorithmiques ou

paramétriques sur la gerbe de partons, sur l'hadronisation et sur la collections de paramètres A14 (VAR3c et cns) sont affichées. On peut remarquer que les variations les plus importantes sont données par «  $f_{sr} : \mu_R$  » et «  $f_{sr} : cns$  ».

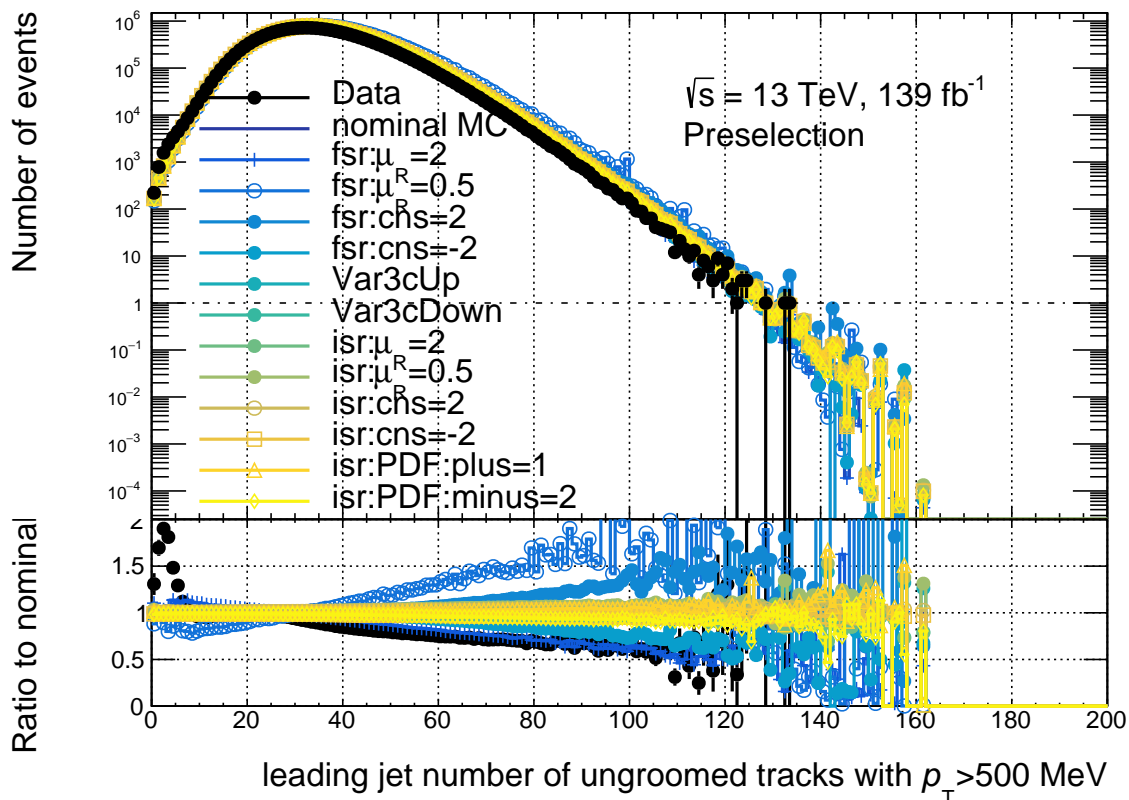


FIGURE 6.38 – Comparaison entre les données et les simulations du bruit de fond de la distribution en nombre de traces sans grooming qui satisfont  $p_T > 500$  MeV pour le jet de plus haute impulsion transverse. Différentes incertitudes systématiques sont affichées.

Certaines permettent de couvrir la différence entre les données et les simulations du bruit de fond.

La Figure 6.39 représente la distribution en masse invariante des simulations du bruit de fond dans la Région de Signal pour les différentes systématiques liées aux termes d'ordres supérieurs manquants dans l'expansion perturbative de la section efficace, ainsi que les incertitudes algorithmiques ou paramétriques sur la gerbe de partons, sur l'hadronisation et sur la collections de paramètres A14 sont affichées. On peut remarquer que les variations les plus importantes sont encore données par «  $f_{sr} : \mu_R$  » et «  $f_{sr} : cns$  ».

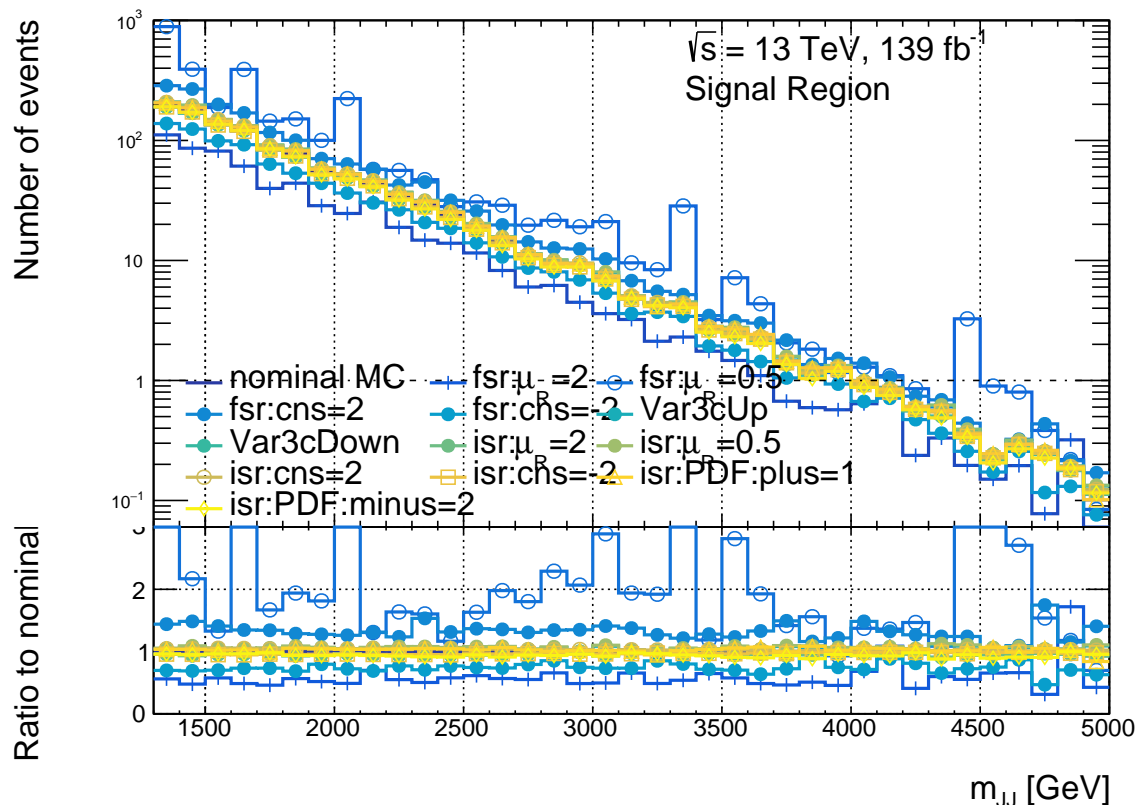
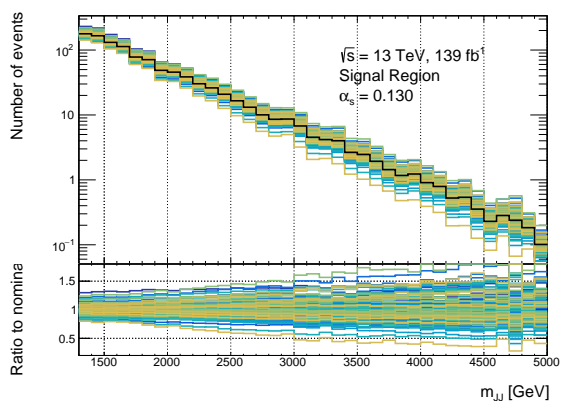
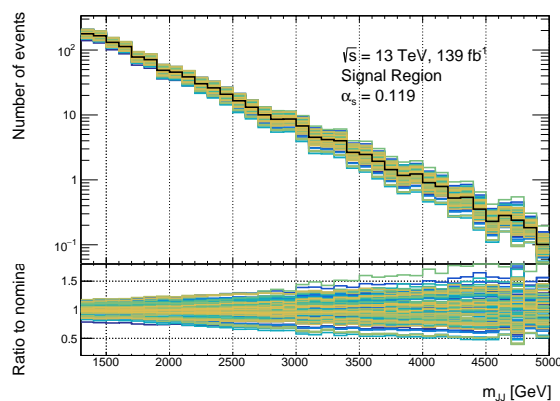


FIGURE 6.39 – Distribution en masse invariante des simulations de bruit de fond dans la Région de Signal pour certaines incertitudes systématiques.

La Figure 6.40 représente la distribution en masse invariante des simulations du bruit de fond dans la Région de Signal pour les 100 variations contenues dans les deux collections de PDFs avec deux différentes valeurs de  $\alpha_s$  : 0,130 et 0,119. L'incertitude systématique totale est donnée par l'écart type de ces deux distributions indépendamment, comme décrit dans l'Équation 6.18.



(a)  $\alpha_s = 0,130$ .



(b)  $\alpha_s = 0,119$ .

FIGURE 6.40 – Comparaison de la distribution en masse invariante des simulations de bruit de fond dans la Région Signal. Les 100 variations contenues dans les collections de PDFs, avec deux différentes valeurs de  $\alpha_s$  sont affichées.

La Figure 6.40 représente la distribution en masse invariante des simulations du bruit de fond dans la Région de Signal pour l'incertitude totale sur les deux collections de PDFs avec deux valeurs différentes de  $\alpha_s$ . L'incertitude finale sur  $\alpha_s$  est donnée par la différence entre les distributions sur la Figure 6.41a et sur la Figure 6.41b.

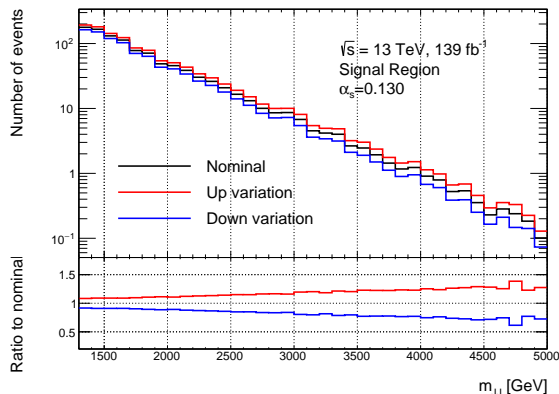
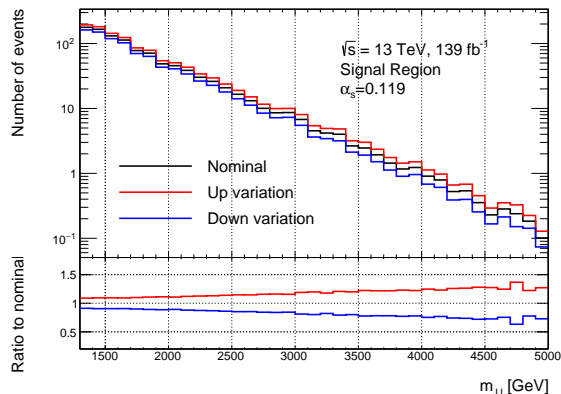
(a)  $\alpha_s = 0, 130$ .(b)  $\alpha_s = 0, 119$ .

FIGURE 6.41 – Comparaison de la distribution en masse invariante des simulations de bruit de fond dans la Région Signal. Les 100 variations contenues dans les collections de PDFs sont combinées. Deux différentes valeurs de  $\alpha_s$  sont affichées.

Les Figures 6.42 et 6.43 représentent les mêmes quantités, mais cette fois-ci pour deux signaux : le Modèle B avec  $M_{Z'}$  = 1,5 TeV, qui est un signal avec un bas nombre de traces, et le Modèle D avec  $M_{Z'}$  = 3,5 TeV, qui est un signal avec un haut nombre de traces.

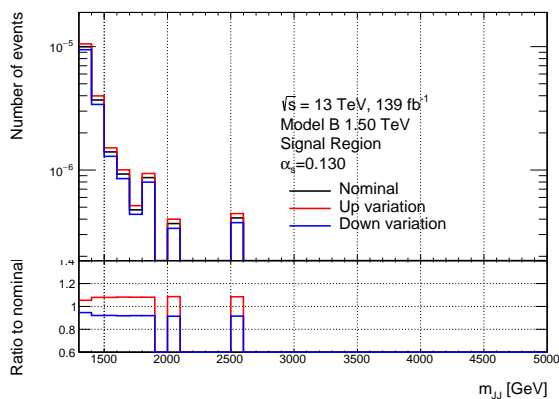
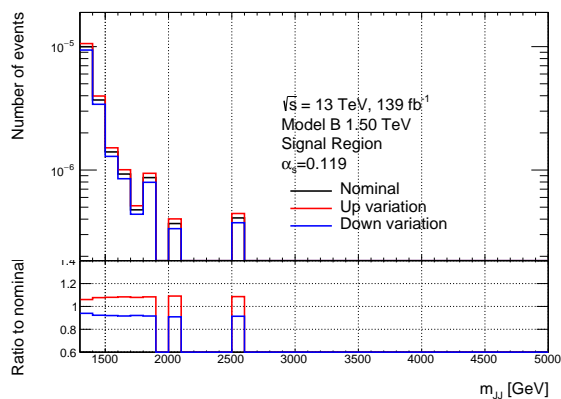
(a)  $\alpha_s = 0, 130$ .(b)  $\alpha_s = 0, 119$ .

FIGURE 6.42 – Comparaison de la distribution en masse invariante pour le Modèle B à 1,5 TeV dans la Région Signal. Les 100 variations contenues dans les collections de PDFs sont combinées. Deux différentes valeurs de  $\alpha_s$  sont affichées.

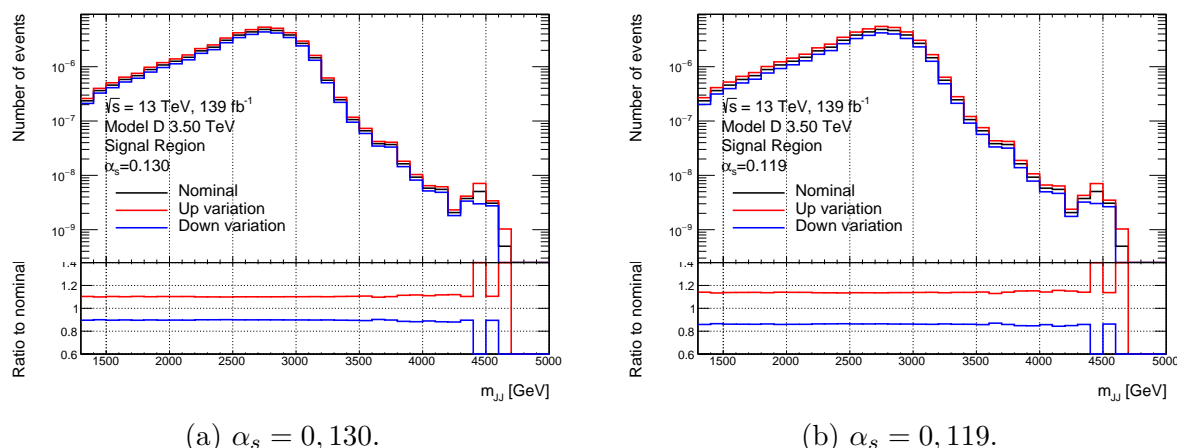


FIGURE 6.43 – Comparaison de la distribution en masse invariante pour le Modèle D à 3,5 TeV dans la Région Signal. Les 100 variations contenues dans les collections de PDFs sont combinées. Deux différentes valeurs de  $\alpha_s$  sont affichées.

### 6.9.3 Incertitudes sur les signaux

En résumé, les incertitudes qui sont considérées pour les signaux sont les suivantes :

- incertitudes sur l'échelle en énergie, sans l'incertitude sur la composition en saveur, mais avec l'incertitude permettent de corriger pour l'étalonnage imparfait,
- incertitudes sur la résolution en énergie,
- incertitudes sur l'échelle en masse, avec l'incertitude permettent de corriger pour l'étalonnage imparfait,
- incertitudes sur la reconstruction des traces à l'intérieur des jets,
- incertitude sur la luminosité intégrée,
- incertitude sur les PDFs,
- incertitude sur la valeur de  $\alpha_s$ .

Pour les signaux, ces incertitudes sont implémentées en faisant varier directement les histogrammes (comme sur la Figure 6.37 par exemple).

Pour les signaux, il existe également une autre incertitude, liée au fait que l'ajustement peut trouver de « faux évènements de signal ». Cette incertitude sera décrite plus en détail dans la Section 6.11.2.

### 6.9.4 Incertitudes sur le bruit de fond

Les incertitudes qui sont considérées pour le bruit de fond sont les suivantes :

- incertitudes sur l'échelle en énergie,
- incertitudes sur la résolution en énergie,
- incertitudes sur l'échelle en masse,
- incertitudes sur la reconstruction des traces à l'intérieur des jets,

- incertitudes sur les termes d’ordres supérieurs manquants dans l’expansion perturbative de la section efficace,
- incertitude sur les PDFs,
- incertitude sur la valeur de  $\alpha_S$ ,
- incertitudes algorithmiques ou paramétriques sur la gerbe de partons, sur l’hadronisation et sur la collections de paramètres A14.

Puisque les percentiles sont dérivés à partir des simulations du bruit de fond, puis appliqués aux données, la dé-corrélation entre la masse invariante et le nombre de traces peut être imparfaite dans les données, ce qui implique que les distributions en masse invariante des données dans les différentes régions (de Contrôle, de Validation, de Signal) n’ont peut-être pas exactement la même forme. Il est à noter que seule la variation de la forme de la distribution en  $m_{JJ}$  peut affecter l’estimation du bruit de fond dans la Région de Signal : tout changement global de normalisation dû à une systématique étant absorbé par le paramètre libre de normalisation du bruit de fond dans la Région de Signal.

Afin d’évaluer cet effet, les percentiles, dérivés à partir des simulations de bruit de fond sans incertitudes systématiques, sont appliqués à toutes les variations des simulations du bruit de fond décrites ci-dessus. Ensuite, l’accord entre la distribution en masse invariante dans la Région de Signal et dans la Région de Contrôle est vérifiée pour chacune de ces incertitudes. Pour ce faire, le rapport entre la distribution en masse invariante dans la Région Signal et dans la Région de Contrôle pour les distributions sans systématiques est divisé par le même rapport, mais cette fois-ci avec les incertitudes systématiques appliquées. La Figure 6.44 représente la distribution en masse invariante de ce double rapport pour les incertitudes systématiques sur  $fsr : \mu_R$  (des distributions similaires sont obtenues pour toutes les autres systématiques). Cette distribution est lissée pour éviter les fluctuations dues à la faible statistique des simulations dans la Région de Signal.



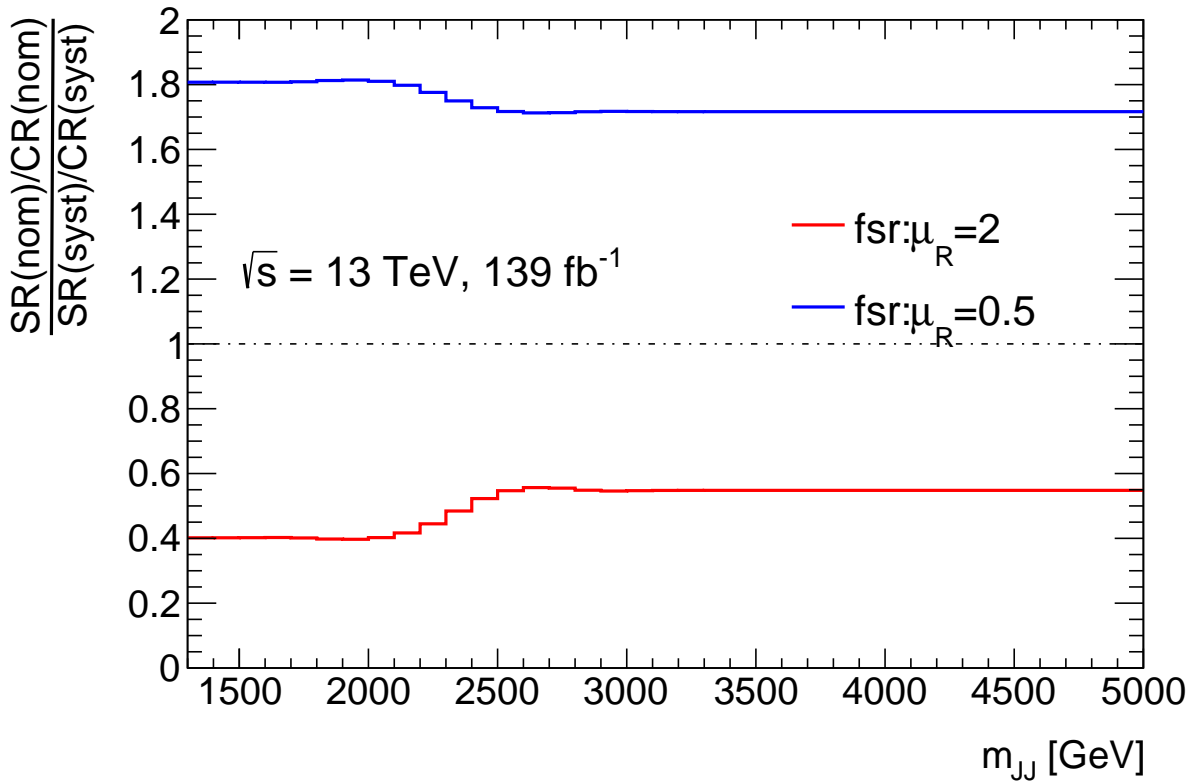


FIGURE 6.44 – Distribution en masse invariante après lissage du rapport des simulations de bruit de fond sans systématiques dans la Région de Signal et dans la Région de Signal sans systématiques divisé par ce même rapport lorsque les incertitudes systématiques sur  $f_{sr} : \mu_R$  sont appliquées.

Toutes ces courbes de double-rapport après lissage entre les simulations de bruit de fond dans la Région de Signal et dans la Région de Contrôle avec et sans incertitudes systématiques sont utilisées afin d'assigner des incertitudes systématiques sur l'accord entre les différentes régions dans les données réelles. Pour ce faire, la distribution en masse invariante des données dans la Région de Contrôle est divisée par ces courbes, une pour chaque incertitude systématique.

Sur la Figure 6.44, on peut remarquer que le double rapport change la forme du bruit de fond, mais aussi sa normalisation (car les courbes ne sont pas à 1). Or ce double rapport n'est utilisé que pour estimer les incertitudes sur les différences potentielles des distributions en masse invariante entre les données dans Région de Signal et celles dans la Région de Contrôle, pas leur normalisation relative. La normalisation du bruit de fond est laissée libre lors de l'ajustement, ce qui absorbe ces différences de normalisation. Par conséquent, les queues de distribution de ces double-rapports sont tout d'abord fixées à 1, avant d'être utilisés afin de diviser la distribution en masse invariante des données dans la Région de Contrôle. Le résultat pour  $f_{sr} : \mu_R$  est représenté dans la Figure 6.45.

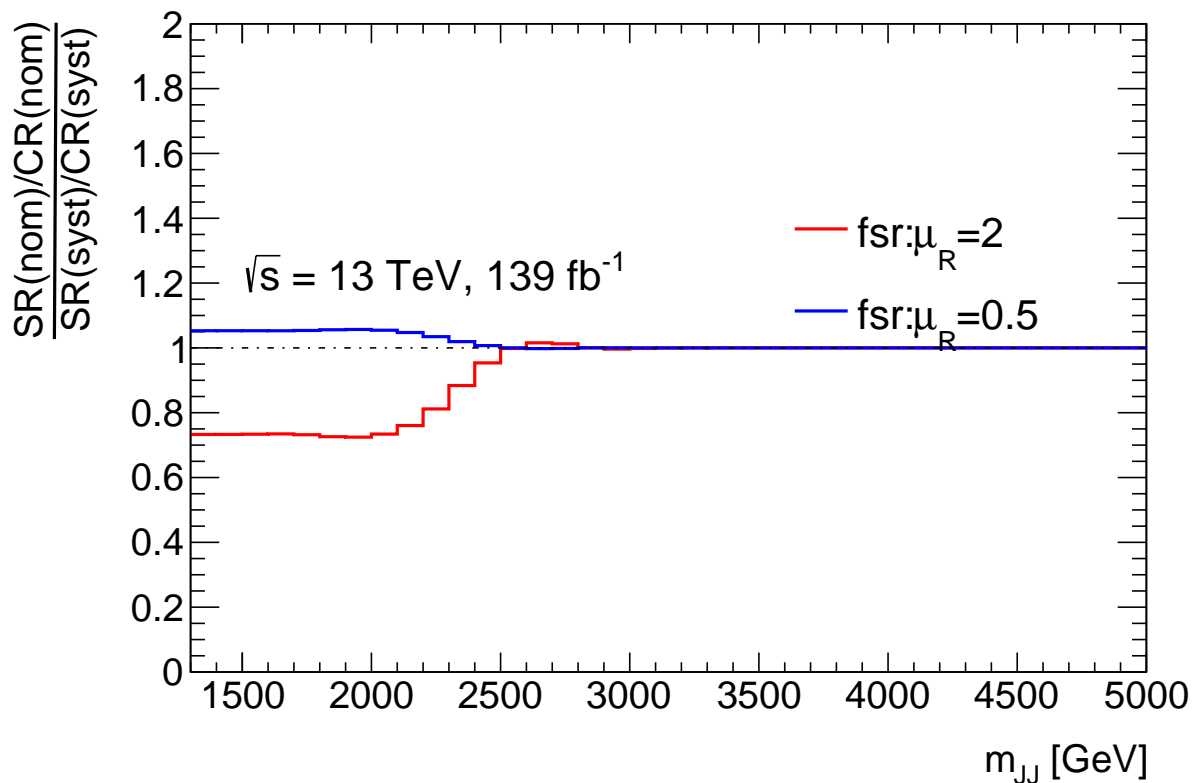


FIGURE 6.45 – Distribution en masse invariante après lissage du rapport des simulations de bruit de fond sans systématiques dans la Région de Signal et dans la Région de Contrôle sans systématiques divisé par ce même rapport lorsque les incertitudes systématiques sur  $f_{sr} : \mu_R$  sont appliquées. Les queues de distribution fixées à 1, contrairement à la Figure 6.44.

## 6.10 Interpolation des signaux

Cette analyse recherche un excès d'évènements par rapport à la distribution en masse invariante du bruit de fond. Pour ce faire, la forme des signaux est utilisée lors de l'analyse statistique. Cependant, les signaux générés (1,5 TeV, 1,75 TeV, 2 TeV, 2,25 TeV, 2,5 TeV, 3 TeV et 3,5 TeV) ne sont pas suffisant pour couvrir tout l'intervalle en  $m_{JJ}$ , et en particulier les intervalles entre les signaux. C'est pourquoi une procédure d'interpolation a été mise en place. Cette procédure comprend quatre étapes :

- La section efficace pour chacun des Modèles est ajustée en fonction de la masse du  $Z'$ .
- Le produit de l'acceptance  $\times$  l'efficacité (définie ci-dessous) pour chacun des Modèles est ajustée en fonction de la masse du  $Z'$ .
- Combinées, les deux étapes précédentes permettent de paramétrer le nombre d'évènements dans la Région de Signal, ce qui sera ensuite utilisé pour obtenir le nombre d'évènements correct pour les signaux interpolés.
- La forme des signaux est interpolée.

Toutes ces étapes sont décrites plus en détail ci-dessous.

### 6.10.1 Ajustement de la section efficace

La section efficace est ajustée en fonction de la masse du  $Z'$  avec la fonction paramétrique suivante :

$$\sigma(m_{Z'}) = p_0 m_{Z'}^{p_1-4} \left(1 - \frac{m_{Z'}}{13 \text{ TeV}}\right)^{p_2} \quad (6.21)$$

L'ajustement de la section efficace en fonction de la masse du  $Z'$  pour tous les modèles étudiés indépendamment est représentée sur la Figure 6.46.

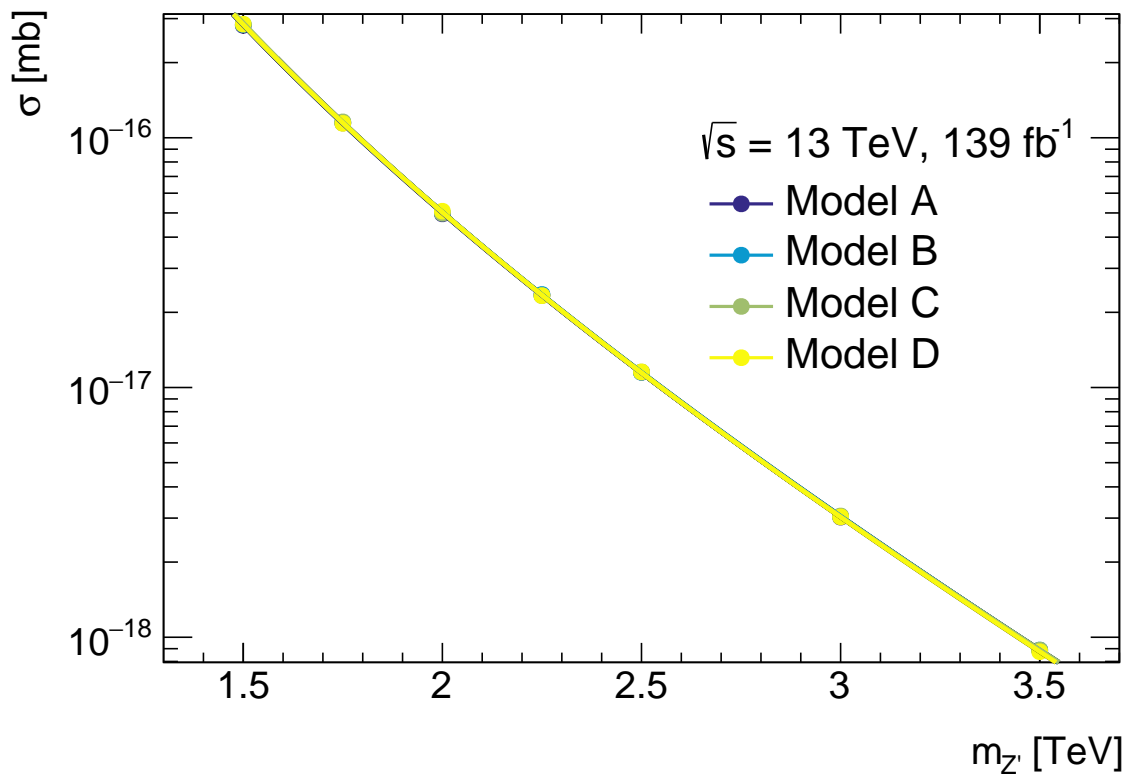


FIGURE 6.46 – Ajustement de la section efficace en fonction de la masse du  $Z'$  pour tous les Modèles.

### 6.10.2 Ajustement de l'acceptance $\times$ l'efficacité

L'acceptance  $\times$  l'efficacité (ou plus brièvement  $A \times \varepsilon$ ) représente le nombre d'évènements dans la Région de Signal divisé par la section efficace  $\times$  la luminosité, c'est-à-dire le nombre d'évènements observés si le détecteur était parfait et que les sélections appliquées n'enlevaient aucun évènement de signal.

L'acceptance  $\times$  l'efficacité est ajustée en fonction de la masse du  $Z'$  avec la fonction paramétrique suivante :

$$f_{A \times \varepsilon}(m_{Z'}) = \frac{p_0}{(1 + e^{-(m_{Z'} - p_1)p_2}) \times (1 + e^{-(m_{Z'} - p_1)p_2})} \quad (6.22)$$

L'ajustement de l'acceptance  $\times$  l'efficacité en fonction de la masse du  $Z'$  pour tous les modèles étudiés indépendamment est représentée sur la Figure 6.47. Le point à 1,5 TeV pour le Modèle C n'est pas utilisé pour l'ajustement, car la plupart des événements de ce signal sont enlevés par les coupures de pré-sélection, comme discuté dans la Section 6.6. On peut remarquer que les Modèles A et D sont les Modèles avec l'acceptance  $\times$  l'efficacité la plus élevée, car ce sont les Modèles avec le plus grand nombre de traces.

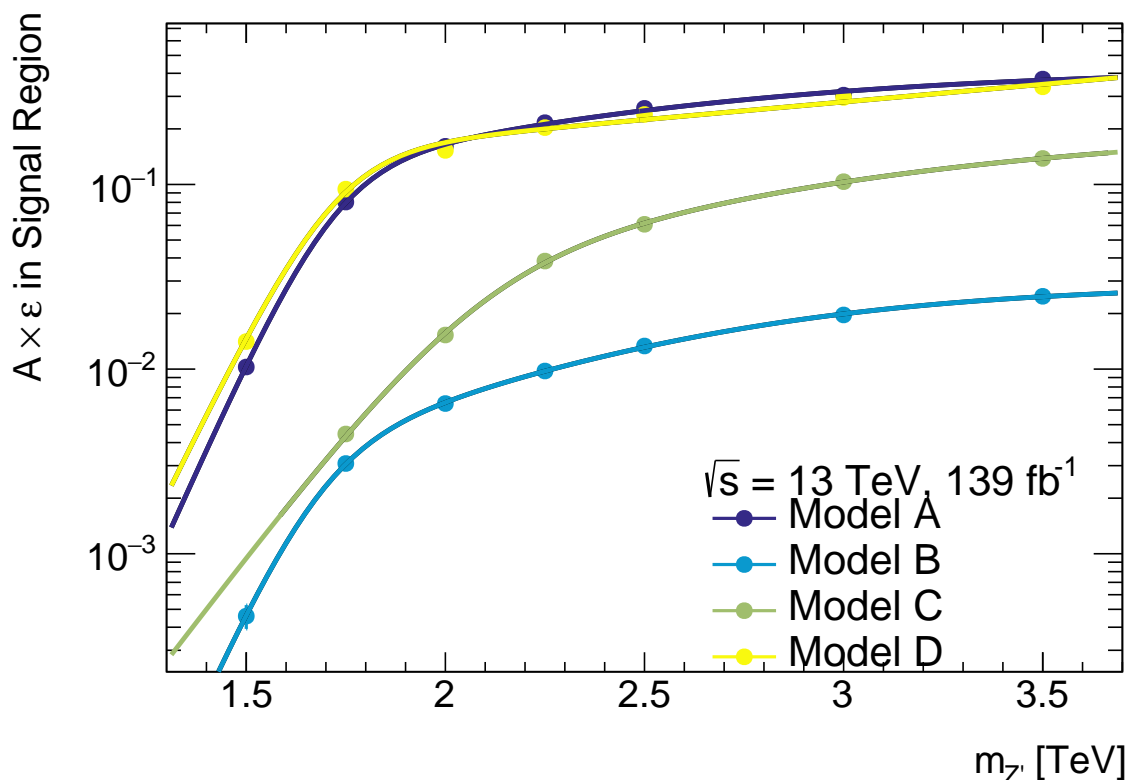


FIGURE 6.47 – Ajustement de l'acceptance  $\times$  l'efficacité en fonction de la masse du  $Z'$  pour tous les Modèles.

### 6.10.3 Nombre d'évènements dans la Région de Signal

Ainsi, à l'aide des deux étapes précédentes, le nombre d'évènements dans la Région de Signal peut être paramétré à l'aide de la formule suivante :

$$\text{Number of events in Signal Region} = \mathcal{L} \times (\sigma \text{ fit}) \times (A \times \varepsilon \text{ fit}) \quad (6.23)$$

La Figure 6.48 représente le nombre d'évènements dans la Région de Signal, ainsi que la paramétrisation donnée par l'équation 6.23. On peut remarquer que la paramétrisation est plutôt bonne.

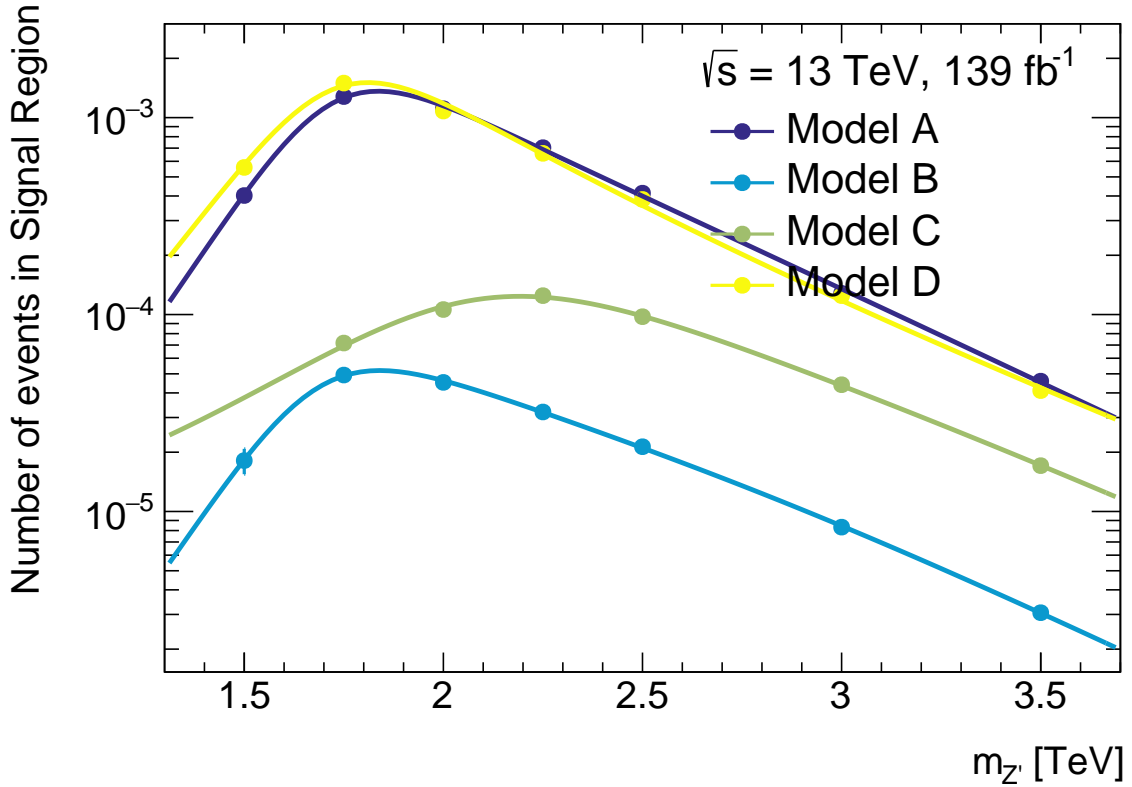


FIGURE 6.48 – Ajustement du nombre d'évènements dans la Région de Signal en fonction de la masse du  $Z'$  pour tous les Modèles. La normalisation est arbitraire.

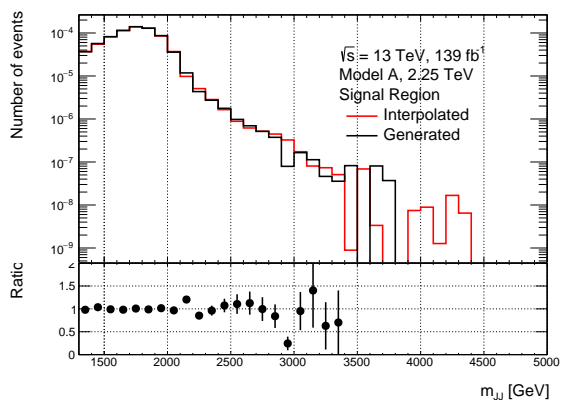
Cette procédure est répétée pour toutes les incertitudes systématiques indépendamment, qui sont listées dans la Section 6.9.

#### 6.10.4 Interpolation de la forme des signaux

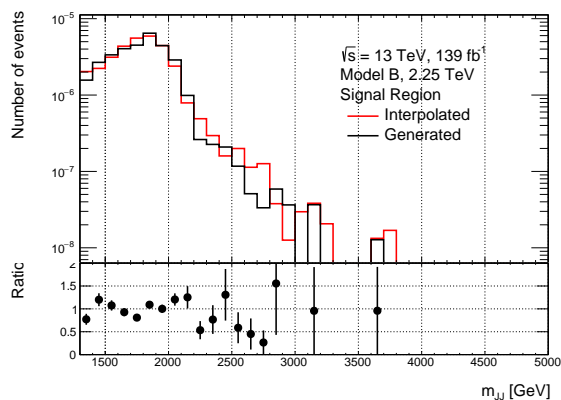
La distribution en masse invariante dans la Région de Signal des signaux intermédiaires est dérivée à l'aide du paquet logiciel *RoMomentMorph* [93]. Cette procédure utilise une interpolation linéaire entre la distribution en masse invariante dans la Région de Signal des signaux générés, afin de prédire la forme des signaux intermédiaires. Pour chaque signal interpolé et pour chaque modèle indépendamment, tous les signaux générés sont utilisés, sauf pour le Modèle C, pour lequel le signal à 1,5 TeV n'est pas utilisé, car sa forme est biaisée par les coupures de pré-sélection (voir la Figure 6.24c). Le premier signal utilisé pour l'interpolation pour le Modèle C est donc 1,75 TeV. Par conséquent, il n'y a pas de signal interpolé entre 1,5 et 1,75 TeV pour Modèle C. Des signaux tous les 100 GeV (ce qui correspond à la résolution expérimentale) sont ainsi interpolés, et ensuite leur normalisation est estimée en évaluant les courbes de la Figure 6.48 à la valeur de la masse du  $Z'$  interpolée. Cette procédure est répétée pour chaque incertitude systématique indépendamment.

Afin de tester si la forme des signaux intermédiaire est prédite correctement, une procédure a été mise en place. Pour ce faire, un signal généré est enlevé de la liste des

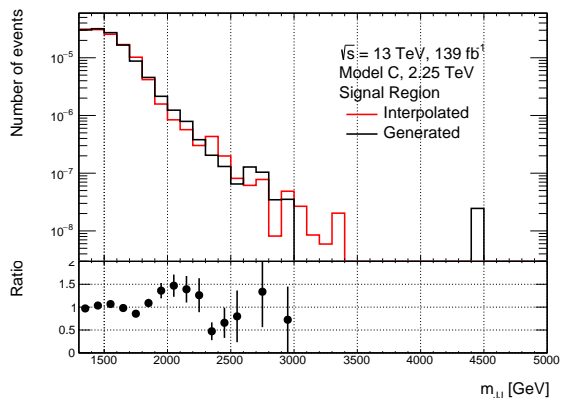
distributions en masse invariante utilisées par le paquet logiciel RooMomentMorph. Ensuite, l'interpolation est effectuée pour ce signal, et les distributions en masse invariante du signal généré et du signal interpolé sont comparées. La Figure 6.49 représente cette comparaison, pour les signaux à 2,25 TeV. Le signal généré et sa prédiction sont en général en bon accord dans le pic de la distribution, sauf pour le Modèle B qui a peu d'évènements dans la Région de Signal, et dont les différences sont couvertes par les incertitudes statistiques.



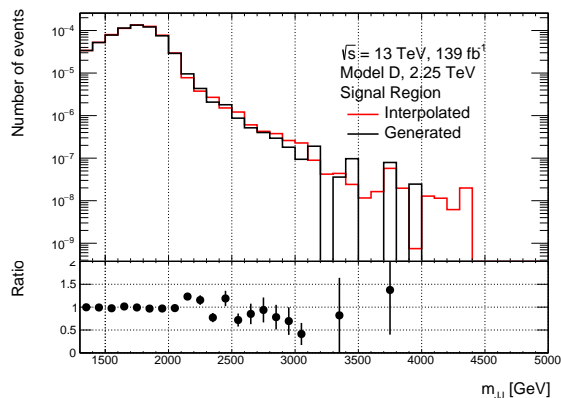
(a) Modèle A.



(b) Modèle B.



(c) Modèle C.



(d) Modèle D.

FIGURE 6.49 – Comparaison de la distribution en masse invariante entre le signal généré à 2,25 TeV et l'interpolation de ce même signal lors que le signal généré n'est pas utilisé lors de la procédure, et ce pour tous les modèles.

Enfin, la Figure 6.50 représente tous les signaux interpolés et générés.

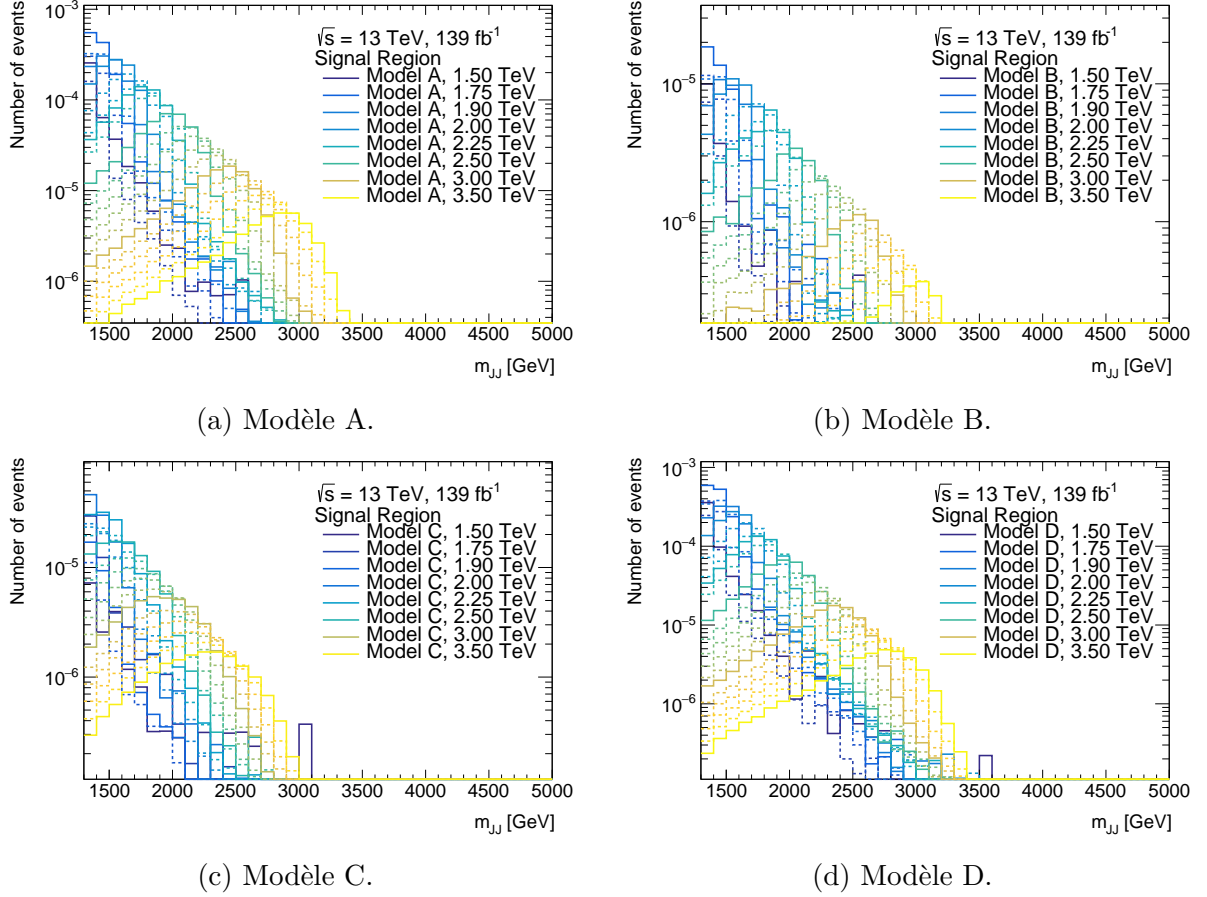


FIGURE 6.50 – Tous les signaux générés (en traits pleins) et interpolés (en pointillés) pour tous les Modèles dans la Région de Signal.

## 6.11 Interprétation statistique

### 6.11.1 Statistiques

En raison de la nature intrinsèquement non déterministe de la physique des particules, les résultats d'une expérience peuvent changer d'un jeu de données à l'autre. Par conséquent, ces résultats doivent être interprétés à l'aide d'une analyse statistique, afin de déterminer s'ils sont significatifs ou non. Cette Section résume quelques résultats de statistique appliquée à la physique des particules nécessaires à la compréhension des prochaines sections.

La probabilité d'observer  $n$  évènements, sachant que  $\nu$  sont attendus en moyenne est donnée par la distribution de Poisson :

$$P(n; \nu) = \frac{\nu^n e^{-\nu}}{n!} \quad (6.24)$$

Cette probabilité, lorsqu'elle est vue comme une fonction des paramètres (ici  $\nu$ ) avec les résultats des données (ici  $n$ ) fixés, est appelée la fonction de vraisemblance :

$$L(\nu) = P(n; \nu) \quad (6.25)$$

La recherche d'un secteur sombre est effectuée sur plusieurs intervalles en masse invariante de 100 GeV. Le nombre d'évènements dans ces intervalles sont supposés être indépendants les uns des autres, et par conséquent, la fonction de vraisemblance totale est donnée par le produit des fonctions de vraisemblance à l'intérieur de ces intervalles :

$$L(\nu) = \prod_{i=1}^{n_{\text{bins}}} L_i(\nu_i) \quad (6.26)$$

où  $\nu$  représente un vecteur des paramètres  $\nu_i$  dans chaque intervalle.

Le maximum de cette fonction de vraisemblance peut être utilisé afin de déterminer les valeurs des paramètres  $\nu_i$  qui permettent d'ajuster au mieux les données :

$$\frac{\partial L}{\partial \nu_i} = 0, \text{ pour } i = 1, \dots, n_{\text{bins}} \quad (6.27)$$

La valeur  $\nu$  qui permet de maximiser la fonction de vraisemblance est notée  $\hat{\nu}$  :

$$\hat{\nu} = \operatorname{argmax}_{\nu} (L(\nu)) \quad (6.28)$$

De manière générale, et par convention, les valeurs des paramètres qui permettent de maximiser la fonction de vraisemblance seront notées avec un « ^ » par la suite.

En physique des particules, le nombre d'évènements attendus  $\nu$  est décomposé en fonction de l'intensité du signal  $\mu$ , du nombre d'évènements de signal attendu  $S$ , et du nombre d'évènements de bruit de fond attendu  $B$  :

$$\nu_i = \mu \times S_i + B_i \quad (6.29)$$

Ainsi, si  $\mu = 0$ , il n'y a pas de signal, et les données ne contiennent que des évènements de bruit de fond. Ce paramètre est souvent appelé le *paramètre d'intérêt*, car c'est celui que les analyses cherchent à déterminer : si la valeur de  $\mu$  qui maximise la fonction de vraisemblance est plus grande que 0, cela pourrait indiquer la présence d'un signal, et par conséquent l'existence potentielle de nouvelle physique. L'intensité du signal peut aussi être formulée selon le rapport des sections efficaces observée et théorique :  $\mu = \frac{\sigma_{\text{observed}}}{\sigma_{\text{theory}}}$ . Par la suite, la fonction de vraisemblance sera notée en fonction de  $\mu$  :  $L(\mu)$ .

Les paramètres  $S_i$  et  $B_i$  peuvent être des fonctions de paramètres inconnus, notés  $\theta$ , et appelés *paramètres de nuisances*, car ils ne sont pas utilisés dans l'interprétation finale de l'analyse :

$$S_i = S_i(\theta) \text{ et } B_i = B_i(\theta) \quad (6.30)$$

Ces paramètres de nuisance permettent de tenir compte du fait que les modèles utilisés pour décrire les données sont imparfaits. Les incertitudes systématiques sont des exemples de paramètres de nuisance. Les paramètres de nuisance détériorent la sensibilité de l'analyse en donnant plus de flexibilité dans l'estimation de la valeur du paramètre d'intérêt. Ils sont inclus dans la fonction de vraisemblance en étant contraints :



$$L(\mu, \theta) = \left( \prod_{i=1}^{n_{\text{bins}}} L_i(\mu_i, \theta) \right) \times \left( \prod_{j=1}^{n_{\text{sys}}} \mathcal{C}(\theta_j) \right) \quad (6.31)$$

où  $\mathcal{C}(\theta)$  représente des connaissances supplémentaires sur les paramètres de nuisance. Cette connaissance supplémentaire provient de données auxiliaires, en plus des données étudiées. La fonction la plus souvent utilisée pour  $\mathcal{C}(\theta)$  est la fonction gaussienne :

$$\mathcal{C}(\theta) = \mathcal{G}(\theta_{\text{obs}}; \theta, \sigma_{\text{syst}}) \quad (6.32)$$

où  $\theta_{\text{obs}}$  et  $\sigma_{\text{syst}}$  proviennent de données auxiliaires.

Les paramètres  $\theta$  sont ajustés, au même nom que le paramètre d'intérêt  $\mu$ . Dans certains cas, il peut arriver que les paramètres  $\theta$  soient contraints de manière plus efficace par l'ajustement que par les données auxiliaires. Afin d'estimer cet effet, les *pulls* sont calculés pour chacun des paramètres de nuisance. Ces pulls sont donnés par :

$$\frac{\hat{\theta} - \theta_0}{\sigma_\theta} \quad (6.33)$$

où  $\hat{\theta}$  représente la valeur de  $\theta$  qui permet de maximiser la fonction de vraisemblance,  $\theta_0$  représente la valeur initiale de  $\theta$  obtenue à l'aide des données auxiliaires et  $\sigma_\theta$  représente l'incertitude sur la valeur de  $\theta$  obtenue lors de l'ajustement. Normalement, les pulls sont distribués de façon gaussienne, avec une moyenne égale à 0 et un écart type égal à 1. Si la valeur centrale d'un pull est différente de 0, cela indique que certaines caractéristiques des données diffèrent du modèle utilisé. Si l'incertitude d'un pull est inférieure à 1, cela indique que le paramètre de nuisance est contraint par les données, ce qui peut être le cas si les données sont particulièrement riches en information sur ce paramètre.

L'impact d'un paramètre de nuisance représente la différence entre la valeur de  $\mu$  qui maximise la fonction de vraisemblance,  $\hat{\mu}$ , et  $\hat{\mu}_{\theta_0 \pm \sigma_\theta}$ , la valeur du paramètre d'intérêt lorsque que le paramètre de nuisance est fixé à sa valeur attendue (observée)  $\theta_0$  plus ou moins un écart type avant (après) ajustement  $\sigma_\theta$  :

$$\Delta \hat{\mu}^\pm = \hat{\mu}_{\theta_0 \pm \sigma_\theta} - \hat{\mu} \quad (6.34)$$

L'impact permet de classifier les paramètres de nuisance par ordre d'importance.

En statistique fréquentiste, la compatibilité d'un modèle avec les données est effectuée en testant une hypothèse nulle, notée  $H_0$ , en comparaison avec une hypothèse alternative  $H_1$ .

Le niveau de concordance des données observées avec une hypothèse  $H$  donnée est quantifié en calculant la *valeur-p*, dénotée  $p$ , c'est-à-dire la probabilité, sous l'hypothèse que  $H$  est vraie, de trouver des données d'égale ou de plus grande incompatibilité avec les prédictions de  $H$ . Si la valeur-p est en dessous d'une certaine valeur  $\alpha$ , alors l'hypothèse  $H$  est rejetée. Habituellement,  $\alpha = 0,05$  est utilisée.  $\alpha$  correspond à la probabilité de rejeter l'hypothèse  $H$  alors que celle-ci est vraie. La valeur-p peut être convertie en une variable

appelée la signification,  $Z$ , qui correspond au nombre d'écarts types de la distribution gaussienne, via la formule suivante :

$$Z = \Phi^{-1}(1 - p) \quad (6.35)$$

où  $\Phi$  représente la distribution cumulative de la distribution gaussienne.

Afin de pouvoir discriminer deux hypothèses  $H_0$  et  $H_1$ , une quantité est introduite : la *statistique du test*, notée  $t$ . Cette quantité est une fonction qui prend en entrée les données  $x$  et qui retourne un simple nombre. Selon le lemme de Neyman et de Pearson [94], la statistique du test qui permet de discriminer au mieux deux hypothèses est donnée par le rapport entre les fonctions de vraisemblance sous les deux hypothèses :

$$t(x) = \frac{P(x, H_0)}{P(x, H_1)} = \frac{L(H_0)}{L(H_1)} \quad (6.36)$$

Cependant, ce lemme n'est valide que dans le cadre d'hypothèses simples, c'est-à-dire sans paramètres libres.

Dans le cadre des expériences au LHC, qui contiennent des paramètres libres, c'est une quantité qui s'en rapproche, appelée *profile likelihood ratio* en anglais, qui est utilisée :

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} \quad (6.37)$$

où :

- $\hat{\theta}(\mu)$  représentent les valeurs de  $\theta$  qui maximisent la fonction de vraisemblance pour une valeur de  $\mu$  donnée,
- $\hat{\mu}$  et  $\hat{\theta}$  représentent les valeurs des paramètres qui maximisent la fonction de vraisemblance, sans condition sur la valeur de  $\hat{\mu}$ .

La présence des paramètres de nuisance élargit la fonction de vraisemblance en fonction de  $\mu$  par rapport à ce que l'on aurait si leurs valeurs étaient fixes. Cela reflète la perte d'information sur  $\mu$  due aux incertitudes systématiques.

La statistique du test est alors donnée par :

$$t_\mu = -2 \times \ln(\lambda(\mu)) \quad (6.38)$$

Des valeurs plus élevées de  $t(\mu)$  correspondent donc à une incompatibilité croissante entre les données et  $\mu$ . Par conséquent, la valeur-p est donnée par :

$$p_\mu = \int_{t_{\mu, \text{obs}}}^{+\infty} f(t_\mu | \mu) dt_\mu \quad (6.39)$$

où  $t_{\mu, \text{obs}}$  est la valeur de la statistique du test  $t_\mu$  observée à partir des données et  $f(t_\mu | \mu)$  désigne la distribution de probabilité de  $t_\mu$  sous l'hypothèse de l'intensité du signal  $\mu$ . Dans cette notation, l'indice de  $t$  se réfère à l'hypothèse testée, et le deuxième

argument de  $f(t_\mu|\mu)$  donne la valeur de  $\mu$  supposée dans la distribution des données.

Dans le contexte de la recherche de signal, l'hypothèse nulle correspond à l'hypothèse de bruit de fond seulement ( $\mu = 0$ ), et l'hypothèse alternative correspond à l'hypothèse de signal et de bruit de fond ( $\mu > 0$ ). Le rejet de l'hypothèse  $\mu = 0$  conduit donc à la découverte d'un nouveau signal. Cependant, seuls les modèles avec  $\mu \geq 0$  sont considérés ici, car le signal représenterait un excès d'évènements seulement. Par conséquent, une sous-fluctuation du bruit de fond ne doit pas servir de preuve contre le bruit de fond, et la statistique de test devient [86] :

$$q_0 = \begin{cases} -2 \times \ln(\lambda(0)) & \text{si } \hat{\mu} \geq 0 \\ 0 & \text{si } \hat{\mu} < 0 \end{cases} \quad (6.40)$$

Et la p-valeur est donnée par [86] :

$$p_0 = \int_{q_{0,\text{obs}}}^{+\infty} f(q_0|0)dq_0 \quad (6.41)$$

où  $f(q_0|0)$  représente la distribution de probabilité de  $q_0$  sous l'hypothèse de bruit de fond seulement ( $\mu = 0$ ).

Dans le contexte d'absence de signal observé, des limites supérieures sont émises sur la valeur de l'intensité de signal, c'est-à-dire que  $\mu < \mu_{up}$  à un certain niveau de confiance  $1 - \alpha$  (95% habituellement). Pour ce faire, les hypothèses nulles et alternatives sont inversées, c'est-à-dire que l'hypothèse nulle correspond au bruit de fond avec du signal avec une intensité  $\mu = \mu_0$  alors que l'hypothèse alternative correspond à  $\mu < \mu_0$ . De la même façon que pour la recherche de signal, les limites sont émises pour des valeurs de  $\mu \geq 0$ . De plus, une sur-fluctuation du bruit de fond ne doit pas servir de preuve contre l'hypothèse de signal et de bruit de fond, et par conséquent la statistique de test prend la forme [86] :

$$\tilde{q}_\mu = \begin{cases} -2 \times \ln\left(\frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))}\right) & \text{si } \hat{\mu} < 0 \\ -2 \times \ln\left(\frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})}\right) & \text{si } 0 \leq \hat{\mu} \leq \mu \\ 0 & \text{si } \hat{\mu} > \mu \end{cases} \quad (6.42)$$

Dans ce cas, la valeur-p est donnée par [86] :

$$p_\mu = p_{s+b} = \int_{\tilde{q}_{\mu,\text{obs}}}^{+\infty} f(\tilde{q}_\mu|\mu)d\tilde{q}_\mu \quad (6.43)$$

Afin de trouver la limite supérieure sur l'intensité du signal, toutes les valeurs de  $\mu$  sont parcourues. La limite supérieure est calculée en cherchant la valeur de  $\mu$  pour laquelle  $p_\mu = 5\%$ .

Lorsque l'analyse a très peu de sensibilité au signal, c'est-à-dire que les hypothèses de bruit de fond seulement, et de signal et de bruit de fond sont indistinguables, il peut arriver qu'une fluctuation basse du bruit de fond mène à rejeter des très faibles valeurs de  $\mu$ . Afin d'éviter que cela n'arrive, la probabilité  $p_\mu$  remplacée par [95] :

$$CL_s = \frac{p_\mu}{1 - p_b} = \frac{p_{s+b}}{1 - p_b} \quad (6.44)$$

où  $p_b = 1 - \int_{\tilde{q}_{\mu, \text{obs}}}^{+\infty} f(\tilde{q}_{\mu} | 0, \hat{\theta}(0, \text{obs})) d\tilde{q}_{\mu}$ .

Ainsi,  $1 - p_b$  est petit lorsque l'analyse a peu de sensibilité, et mène donc à des larges valeurs de  $\text{CL}_s$ , évitant l'exclusion du signal auquel l'analyse ne devrait pas être sensible. Cette méthode est appelée la *méthode*  $\text{CL}_s$ . Les définitions de  $p_{\mu} = p_{s+b}$  et de  $p_b$  sont représentées sur la Figure 6.51 [96].

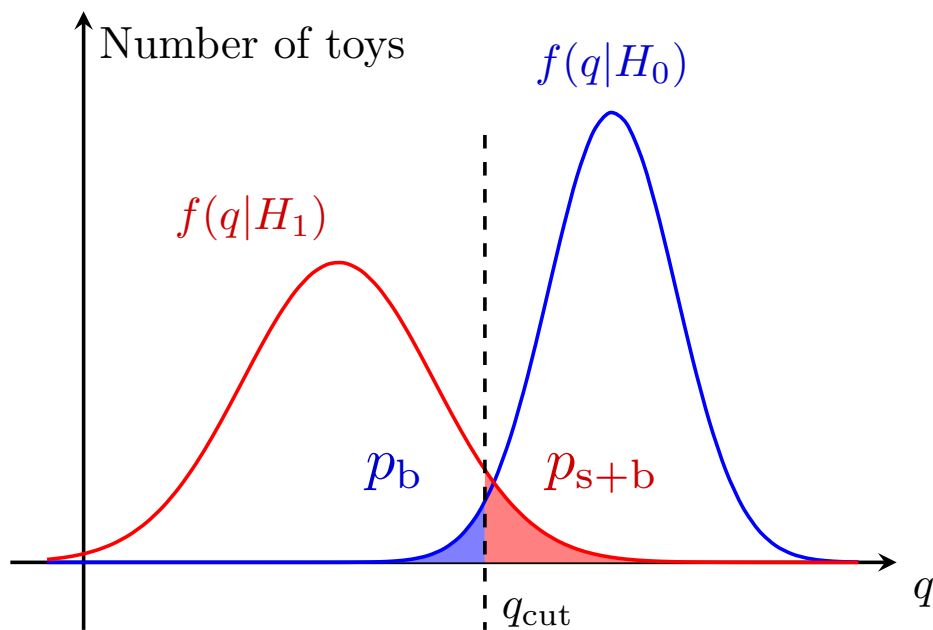


FIGURE 6.51 – Définition de  $p_{\mu} = p_{s+b}$  et de  $p_b$  utilisées dans la méthode  $\text{CL}_s$  [96].

La limite supérieure attendue correspond à la limite supérieure obtenue dans le cas où l'hypothèse de bruit de fond seulement est vraie. Elle est obtenue en prenant la médiane de la distribution de  $\mu_{\text{up,exp}}$  calculée avec la méthode  $\text{CL}_s$  sous l'hypothèse de bruit de fond seulement. De manière générale, les valeurs à  $\pm 1\sigma$  et à  $\pm 2\sigma$  de la distribution de  $\mu_{\text{up}}$  sont aussi reportées en vert et en jaune respectivement, comme représenté sur la Figure 6.52 [96].

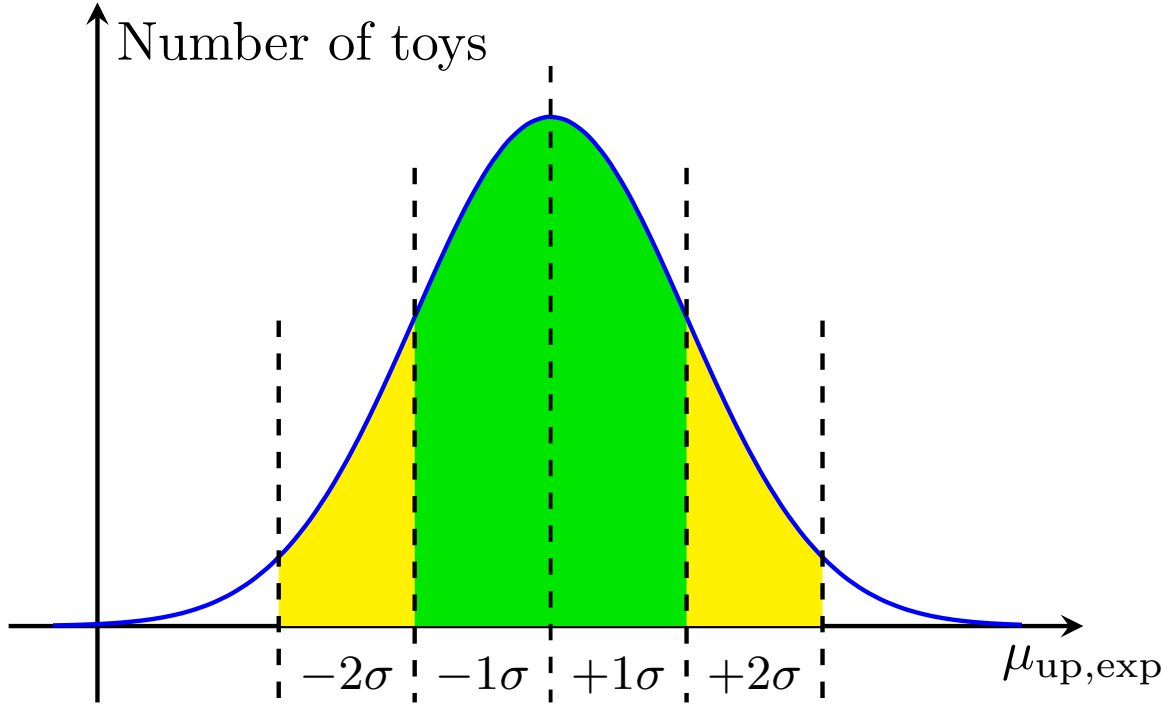


FIGURE 6.52 – Définition de la limite supérieur attendue, ainsi que des valeurs à  $\pm 1\sigma$  et à  $\pm 2\sigma$  [96].

Afin de calculer les probabilités présentées ci-dessus, les distributions de probabilité  $f$  des statistiques de test doivent être connues. Plusieurs méthodes existent pour les estimer. La première consiste à générer des pseudo-données sous les hypothèses considérées et à remplir des histogrammes des statistiques de test. Cependant, cette méthode est très intensive en calcul, car plusieurs millions de pseudo-données, si ce n'est plus, doivent être générées afin de remplir les queues des distributions. Dans la limite asymptotique, c'est-à-dire d'un grand nombre d'évènements (en pratique environ 10 évènements suffisent), une autre méthode peut être utilisée, qui est moins intensive en calcul. Dans cette limite, le théorème de Wilks [97] prévoit que la statistique de test  $t_\mu$  suit la distribution d'un  $\chi^2$ . Lorsque l'hypothèse en train d'être testée est différente de l'hypothèse supposée dans les données, les distributions  $f(q_\mu|\mu')$  (c'est-à-dire la distribution de probabilité de  $q_\mu$  sous l'hypothèse de l'intensité du signal  $\mu'$ ) peuvent être déterminées en utilisant le théorème de Wald [98]. Selon ce théorème :

$$-2 \times \ln(\lambda(\mu)) \simeq \frac{(\mu - \hat{\mu})^2}{\sigma^2} \quad (6.45)$$

où  $\sigma$  représente l'écart type de  $\hat{\mu}$ . Cela permet de montrer que la statistique de test  $t_\mu$  suit une distribution de  $\chi^2$  non centrale :

$$f(t_\mu; \Lambda) \simeq \frac{1}{2\sqrt{t_\mu}} \frac{1}{\sqrt{2\pi}} \left[ e^{\left(-\frac{1}{2}(\sqrt{t_\mu} + \sqrt{\Lambda})^2\right)} + e^{\left(-\frac{1}{2}(\sqrt{t_\mu} - \sqrt{\Lambda})^2\right)} \right] \quad (6.46)$$

où le paramètre de non centralité  $\Lambda = \frac{(\mu - \mu')^2}{\sigma}$ .

Cette distribution dépend de l'écart type de  $\hat{\mu}$ ,  $\sigma$ . Afin d'estimer cette quantité, un ensemble de données artificielles, nommé « *ensemble de données Asimov* » [86], est utilisé. L'ensemble de données Asimov est défini de sorte que lorsqu'on l'utilise pour évaluer les estimateurs de tous les paramètres, on obtient les vraies valeurs de ces paramètres. L'ensemble de données Asimov peut être utilisé pour calculer la fonction de vraisemblance  $L_A$ , et le profile likelihood ratio correspondant  $\lambda_A$  :

$$\lambda_A(\mu) = \frac{L_A(\mu, \hat{\theta}(\mu))}{L_A(\hat{\mu}, \hat{\theta})} \quad (6.47)$$

Pour l'ensemble de données Asimov, l'Équation 6.45 devient :

$$-2 \times \ln(\lambda_A(\mu)) \simeq \frac{(\mu - \mu')^2}{\sigma^2} = \Lambda \quad (6.48)$$

Cette Équation peut être utilisée pour calculer le paramètre de non centralité  $\Lambda$ .

Par la suite, ce sera la limite asymptotique qui sera utilisée dans l'analyse.

### 6.11.2 Tests de faux signal

S'il y a un biais intrinsèque entre le bruit de fond et les données dans la Région de Signal, dû par exemple à une différence entre la distribution en masse invariante des données dans la Région de Contrôle et celle des données dans la Région de Signal, ce biais pourrait éventuellement être interprété de façon incorrecte comme provenant de la présence d'un signal. Cette déviation provenant du biais entre l'estimation du bruit de fond et les données est appelé « faux signal » (« *spurious signal* » en anglais).

C'est pourquoi des « tests de faux signal » sont mis en place. Ces tests sont effectués avec les données dans la Région de Validation. Pour ce faire, les coupures de la Région de Signal sont tout d'abord appliquées aux simulations du bruit de fond. Ensuite, le nombre d'évènements des simulations de bruit de fond dans la Région de Signal est diminué d'un facteur prenant en compte la différence en nombre d'évènements entre les données et les simulations du bruit de fond après les coupures de pré-sélection, comme décrit dans la Section 6.7. Enfin, la différence en efficacité de sélection après les coupures en  $n_{\text{trk}}^\epsilon$  entre les données réelles et les simulations du bruit de fond est utilisée pour diminuer le nombre d'évènements des simulations de bruit dans la Région de Signal, afin d'estimer le nombre d'évènements qu'auront les données réelles dans la Région de Signal. Cette différence d'efficacité de sélection est élevée au carré, afin de prendre en compte que la Région de Signal correspond à une coupure sur les deux jets de plus haute impulsion transverse, alors que la Région de Validation correspond à une coupure sur le jet de plus haute impulsion transverse seulement. En résumé, l'estimation du nombre d'évènements attendus des données réelles dans la Région de Signal est donné par :

$$N_{\text{expected data SR}} = \frac{N_{\text{MC SR}} \times \left(\frac{R_{\text{data}}}{R_{\text{MC}}}\right)_{\text{leading jet}} \times \left(\frac{R_{\text{data}}}{R_{\text{MC}}}\right)_{\text{subleading jet}}}{\frac{N_{\text{MC preselection}}}{N_{\text{data preselection}}}} \quad (6.49)$$

avec

$$R = \frac{N_{\text{VR}}}{N_{\text{CR}}} \quad (6.50)$$

La Figure 6.53 représente une comparaison de la distribution en masse invariante des données dans la Région de Contrôle et dans la Région de Validation, une fois que leur nombre d'évènements a été normalisé aux nombres d'évènements attendus des données réelles dans la Région de Signal.

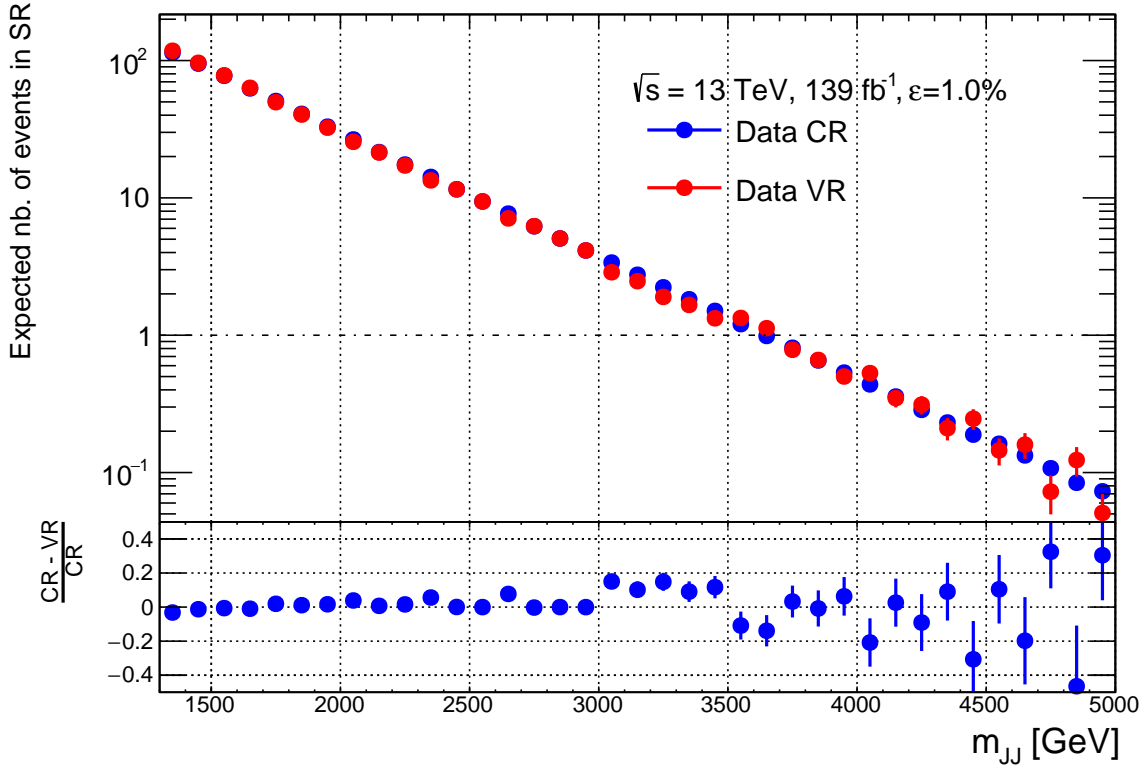


FIGURE 6.53 – Comparaison de la distribution en masse invariante des données dans la Région de Contrôle et dans la Région de Validation. Les données sont normalisées au nombre d'évènements attendus des données dans la Région de Signal.

Lors des tests de faux signal, les données dans la Région de Validation normalisées au nombre d'évènements attendus pour les données dans la Région de Signal sont ajustées avec du signal et du bruit de fond (ce qui est noté S+B). La forme des signaux provient des simulations décrites précédemment. Les signaux interpolés sont aussi utilisés. La forme du bruit de fond provient de la distribution en masse invariante des données dans la Région de Contrôle normalisées. Seules les normalisations du bruit de fond et du signal sont laissées libres lors de l'ajustement. Les tests de faux signal sont alors des tests de faux positifs : il n'y a pas de signal attendu dans les données dans la Région de Validation, mais on cherche à savoir si l'ajustement peut quand même en trouver un.

Le nombre d'évènements de faux signal, noté  $S_{\text{spur}}$ , correspond alors à la différence entre le nombre d'évènements de signal ajusté et le nombre d'évènements de signal attendu (c'est-à-dire 0 ici) :

$$S_{\text{spur}} = S_{\text{fit}} - S_{\text{expected}} = S_{\text{fit}} \quad (6.51)$$

Dans le cadre de cette analyse, le nombre d'évènements de faux signal est évalué pour tous les modèles en fonction de la masse du  $Z'$ . Le résultat est représenté sur la Figure 6.54, lorsque les incertitudes systématiques sur le signal et le bruit de fond ne sont pas appliquées. Un nombre d'évènements de faux signal négatif représente un déficit en évènements. Sur cette Figure, on peut remarquer que les Modèles A, B et D présentent un nombre d'évènements de faux signal similaires. Le nombre d'évènements de faux signal est plus grand pour le Modèle C, ce qui peut s'expliquer par le fait que c'est le Modèle avec les signaux les plus larges, et donc par conséquent qu'ils accommodent plus les différences entre le bruit de fond et les données dans la Région de Validation. La courbe pour ce Modèle est aussi déplacée par rapport aux autres Modèles, car le Modèle C est un Modèle avec la plus grande différence entre la masse reconstruite et la masse du  $Z'$ .

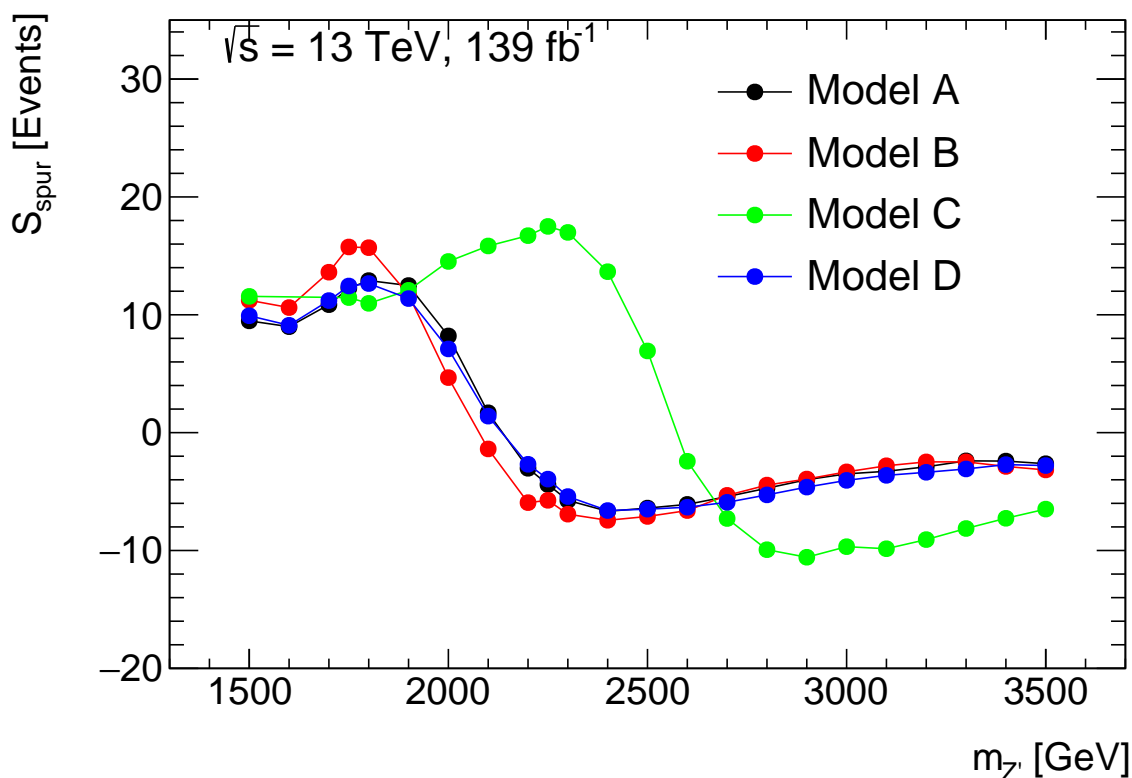


FIGURE 6.54 – Nombre d'évènements de faux signal sans incertitudes systématiques pour tous les modèles étudiés en fonction de la masse du  $Z'$ .

Idéalement, le nombre d'évènements absolu de faux signal,  $|S_{\text{spur}}|$ , devrait être petit. Le nombre d'évènements de faux signal  $|S_{\text{spur}}|$  est implémenté comme une incertitude systématique additionnelle sur le nombre d'évènements de signal (voir ci-dessous) associée à un paramètre de nuisance avec une contrainte gaussienne. Cela conduit à une incertitude totale, notée  $\sigma_{\text{tot}}$ , sur le nombre d'évènements de signal donnée par l'Équation [99] :

$$\sigma_{\text{tot}} = \sqrt{(\sigma_{\text{fit}})^2 + (S_{\text{spur}})^2} \quad (6.52)$$



où  $\sigma_{\text{fit}}$  représente l'incertitude sur  $S_{\text{fit}} = S_{\text{spur}}$  obtenue lors de l'ajustement. Ainsi, pour éviter que le faux signal implique une forte augmentation de l'incertitude totale, il est nécessaire que [99] :

$$|S_{\text{spur}}| \leq 0,5 \times \sigma_{\text{fit}} \Leftrightarrow \frac{|S_{\text{spur}}|}{\sigma_{\text{fit}}} \leq 0,5 \quad (6.53)$$

La Figure 6.55 représente le rapport entre  $S_{\text{spur}}$  et  $\sigma_{\text{fit}}$  lorsqu'aucune incertitude systématique n'est appliquée pour tous les modèles étudiés en fonction de la masse du  $Z'$ . On peut remarquer que ce rapport est très proche du critère de 50% décrit ci-dessus.

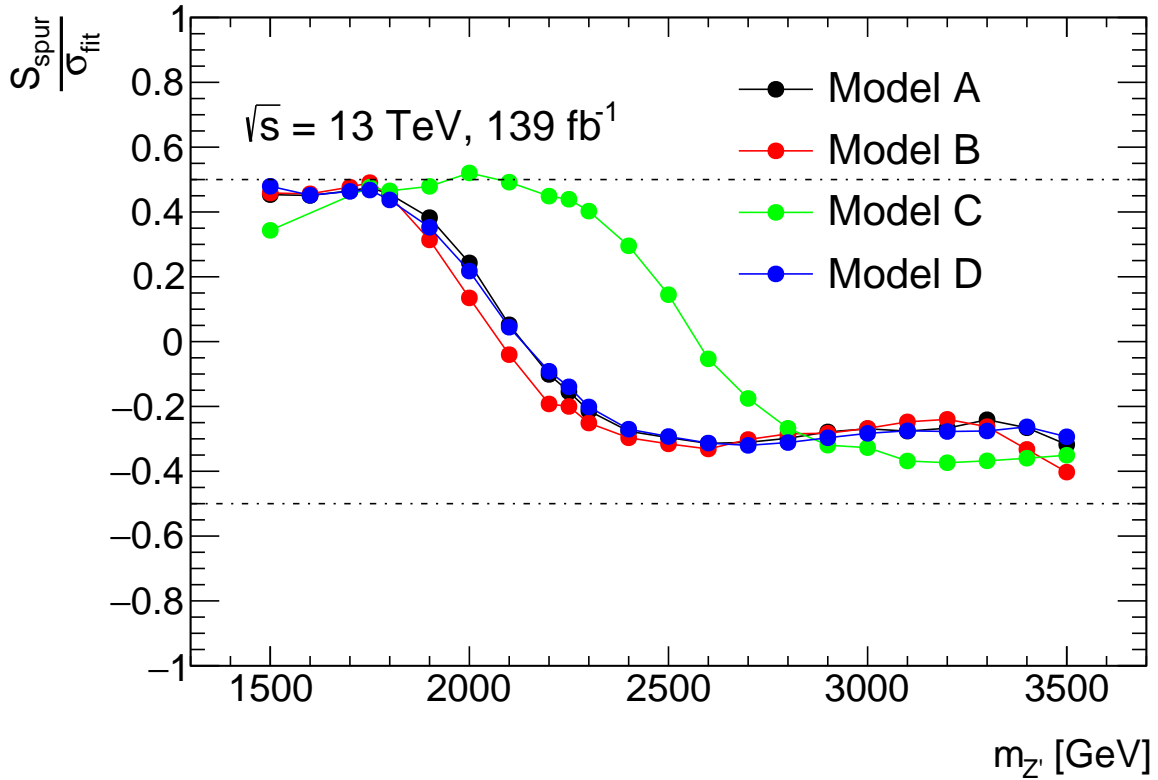


FIGURE 6.55 – Faux signal sans incertitudes systématiques pour tous les modèles étudiés en fonction de la masse du  $Z'$ .

La Figure 6.56 représente la même quantité que dans la Figure 6.55, mais cette fois-ci, une fois que toutes les incertitudes systématiques aient été incluses dans l'ajustement. On peut remarquer que le rapport  $\frac{|S_{\text{spur}}|}{\sigma_{\text{fit}}}$  est maintenant inférieur à 0,5, la diminution de l'importance du faux signal étant attendue lorsque les incertitudes reliées à l'ajustement augmentent.

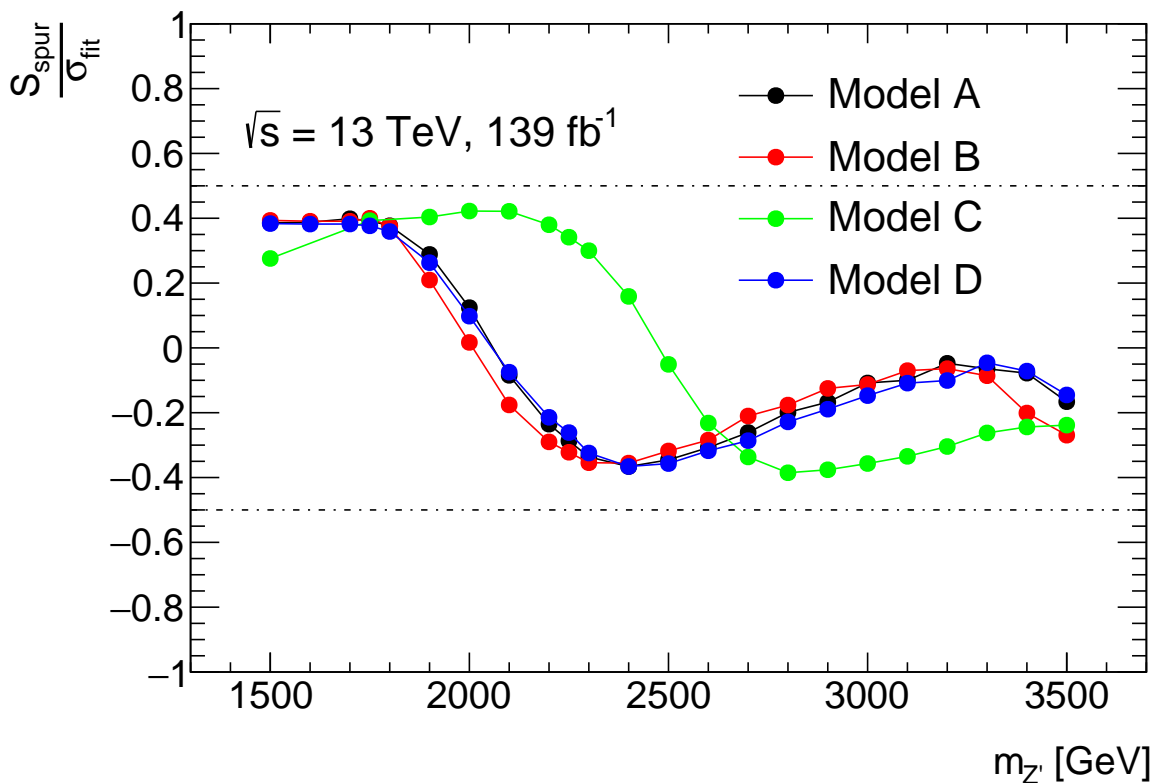


FIGURE 6.56 – Faux signal avec toutes les incertitudes systématiques pour tous les modèles étudiés en fonction de la masse du  $Z'$ .

La valeur de  $S_{\text{spur}}$  est considérée comme une incertitude systématique sur le nombre d'évènements de signal, afin de couvrir de potentiels biais. Cette incertitude est implémentée comme une contribution supplémentaire au signal. Par exemple, si le nombre d'évènements de signal est paramétré avec la section efficace  $\sigma_{\text{signal}}$ , alors le nombre total d'évènements de signal peut être écrit comme suit [99] :

$$N_{\text{signal}}(m_{Z'}) = \sigma_{\text{signal}} \times \mathcal{L} \times (A \times \varepsilon) + S_{\text{modelling}}(m_{Z'}) \times \theta_{\text{modelling}} \quad (6.54)$$

où le premier terme correspond au vrai nombre d'évènements de signal, et le second à l'incertitude systématique provenant du faux signal.  $\theta_{\text{modelling}}$  représente un paramètre de nuisance à cette incertitude à laquelle une contrainte gaussienne est appliquée avec une moyenne égale à 0 et un écart type égal à 1. Enfin  $S_{\text{modelling}}$  représente la valeur absolue du faux signal.

La valeur du faux signal change avec les incertitudes systématiques, et il y a donc une incertitude sur la valeur de faux signal attendue. Pour prendre en compte cet effet, la valeur finale de l'incertitude due au faux signal est ici obtenue en ajoutant en quadrature la différence entre la valeur du faux signal sans systématiques (notée  $S_0$ , c'est ce qui est représenté sur la Figure 6.54) et la valeur du faux signal avec toutes les incertitudes systématiques incluses (notée  $S_{\text{systs}}$ ) à la valeur du faux signal sans systématiques :

$$S_{\text{modelling}} = \sqrt{(S_0)^2 + (S_0 - S_{\text{systs}})^2} \quad (6.55)$$

### 6.11.3 Tests d'injection de signal

Les tests d'injection de signal correspondent à des tests de faux négatif : il y a du signal, mais l'ajustement n'en trouve pas. Pour ce faire, un ajustement de signal et de bruit de fond est effectué sur les données dans la Région de Validation (normalisées au nombre d'évènements attendu pour les données dans la Région de Signal) dans lesquelles du signal est artificiellement injecté. Tout d'abord, une fenêtre est calculée autour de chaque signal. Cette fenêtre est définie par le maximum de la distribution en masse invariante plus ou moins l'écart type de cette distribution. Ensuite, le nombre d'évènements, c'est-à-dire l'intégrale, du bruit de fond est calculé dans cette fenêtre. Enfin, la quantité de signal injectée correspond à  $\sigma = 3$  ou  $5$  fois la racine carrée du nombre d'évènements de bruit de fond dans la fenêtre définie précédemment. Cette procédure est répétée pour chacun des signaux indépendamment. Idéalement, le nombre d'évènements de signal observé, c'est-à-dire le nombre d'évènements de signal obtenu par l'ajustement, devrait être égal au nombre d'évènements de signal injecté et donc le rapport entre ces deux quantités devrait être proche de 1.

La Figure 6.57 représente ce rapport en fonction de la masse du  $Z'$  pour différentes quantités de signal injectées (ce qui est indiqué par  $\sigma$  dans la légende) et pour tous les modèles étudiés. On peut remarquer que le nombre d'évènements de signal ajusté et le nombre d'évènements de signal attendus est proche de 1, bien que les barres d'erreur soient très grandes, ce qui est encore en cours d'investigation.

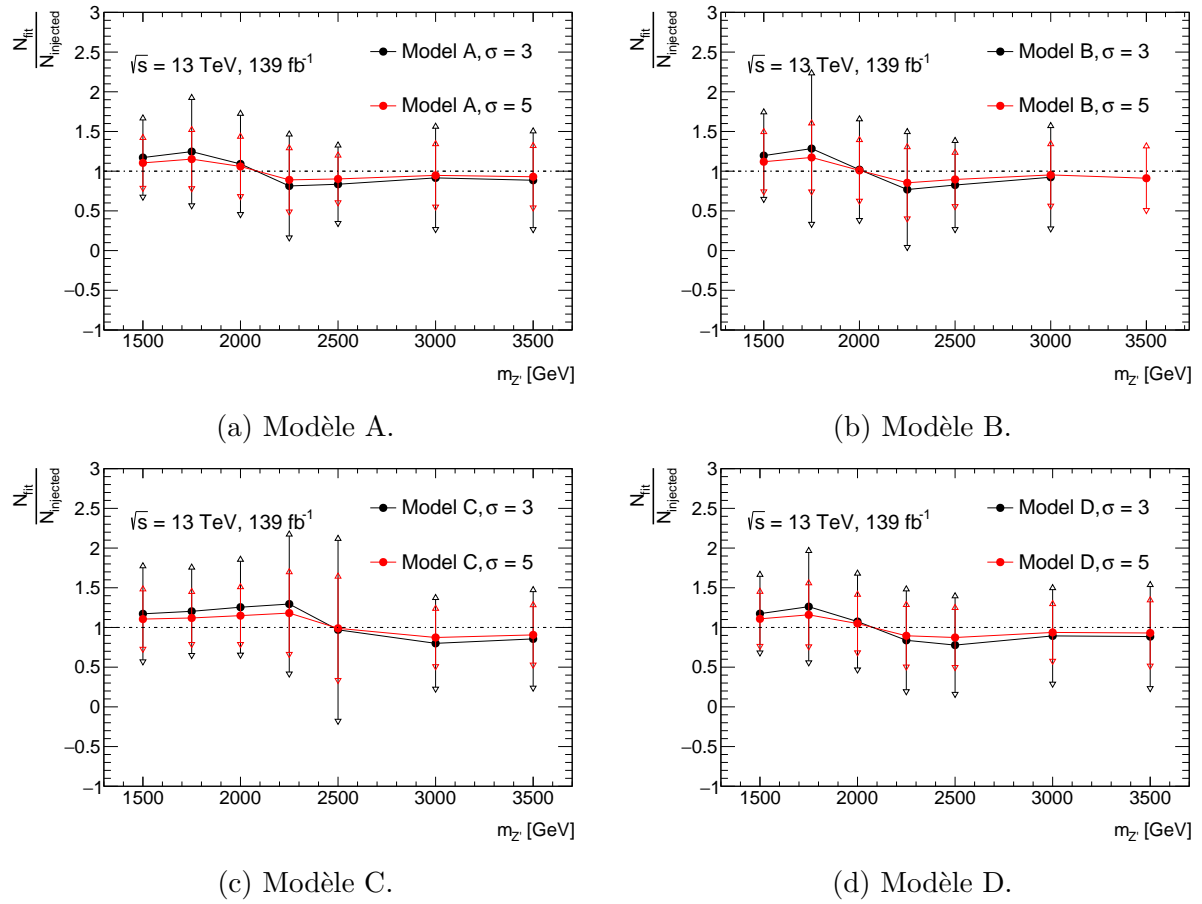


FIGURE 6.57 – Rapport entre le nombre d'évènements de signal ajusté et le nombre d'évènements de signal attendus en fonction de la masse du  $Z'$  pour différentes quantités de signal injectées et pour tous les modèles étudiés.

Un exemple de distribution des données dans la Région de Validation avec du signal injecté pour le Modèle D 3,5 TeV à  $5\sigma$  est représenté sur la Figure 6.58. Sur cette Figure, les distributions après ajustement sur les données avec le signal injecté du bruit de fond et du signal sont représentées. On peut remarquer que l'ajustement permet de bien modéliser l'excès qui correspond au signal injecté.

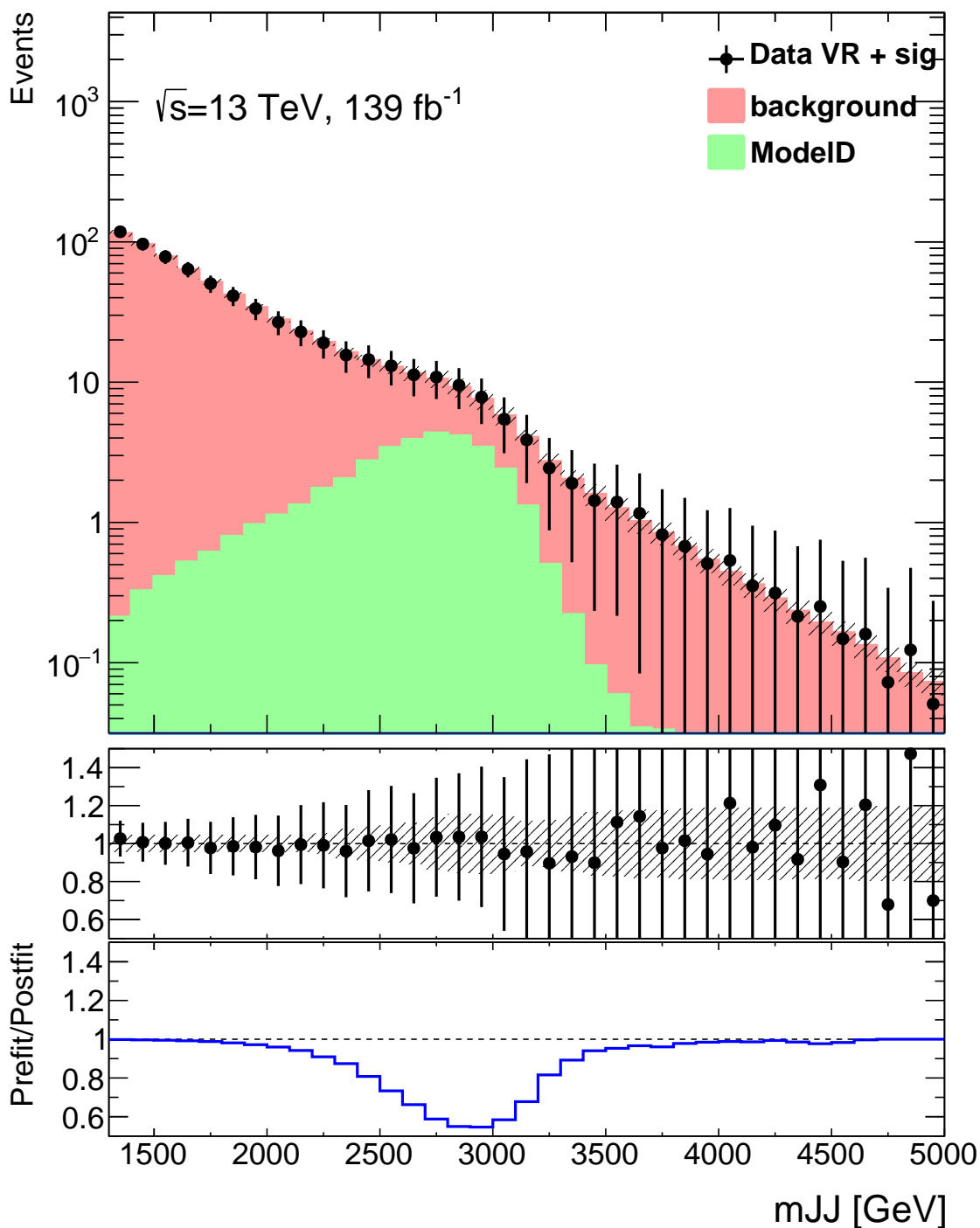


FIGURE 6.58 – Distribution en masse invariante des données dans la région de Validation avec du signal injecté pour le Modèle D 3,5 TeV à  $5 \sigma$  ainsi les distributions après ajustement du bruit de fond et du signal.

#### 6.11.4 Classement des paramètres de nuisance

Les Figures 6.59 et 6.60 représentent un classement des paramètres de nuisance pour les Modèles A et D respectivement avec  $M_{Z'} = 3,5$  TeV, tel qu'obtenu en performant un ajustement de type signal + bruit de fond sur les données dans la Région de Validation (sans signal injecté) normalisées au nombre d'évènements attendu des données dans la Région de Signal. Sur ces Figures, les points noirs et leur barre d'erreur représentent la valeur du paramètre d'intérêt après ajustement et son incertitude, ce qui correspond à l'axe des abscisses inférieur (voir la Section 6.11.1). Les bandes remplies correspondent à l'impact relatif de ces paramètres de nuisance, tel que décrit dans la Section 6.11.1, ce qui correspond à l'axe des abscisses supérieur. Plus l'impact est important, plus les paramètres de nuisance sont en haut. La couleur verte signifie un impact anti-corrélé, c'est-à-dire que l'impact à  $+1\sigma$  est négatif et l'impact  $-1\sigma$  est positif. Un impact de couleur bleue signifie un impact corrélé, c'est-à-dire que l'impact  $+1\sigma$  est positif et l'impact  $-1\sigma$  est négatif. Sur ces deux Figures, « norm\_qqqq\_bkg » représente le paramètre de normalisation du bruit de fond, qui est laissé libre lors de l'ajustement. On peut voir que parmi les paramètres de nuisance les plus importants figurent  $f_{SR} : \mu_R$ , c'est-à-dire des variations du facteur de renormalisation (voir la Section 6.9). Toutes les valeurs moyennes des pulls (voir la Section 6.11.1) se situent entre -1 et 1, et toutes leurs incertitudes sont d'écart type inférieure ou égale à 1. Enfin, ce sont les mêmes paramètres de nuisance qui ont le plus d'impact pour les deux Modèles.

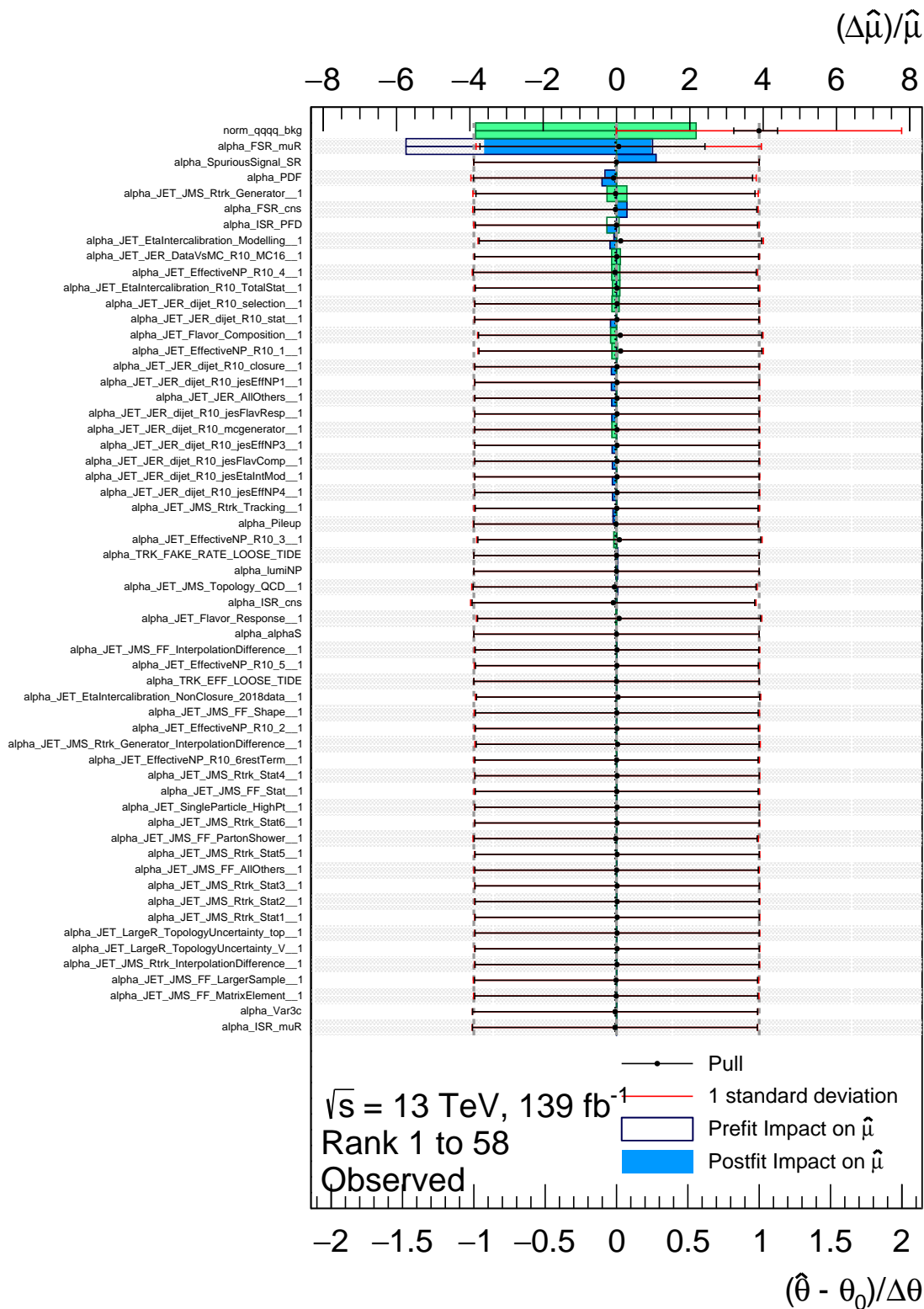


FIGURE 6.59 – Classement des paramètres de nuisance pour le Modèle A avec  $M_{Z'} = 3,5 \text{ TeV}$ .

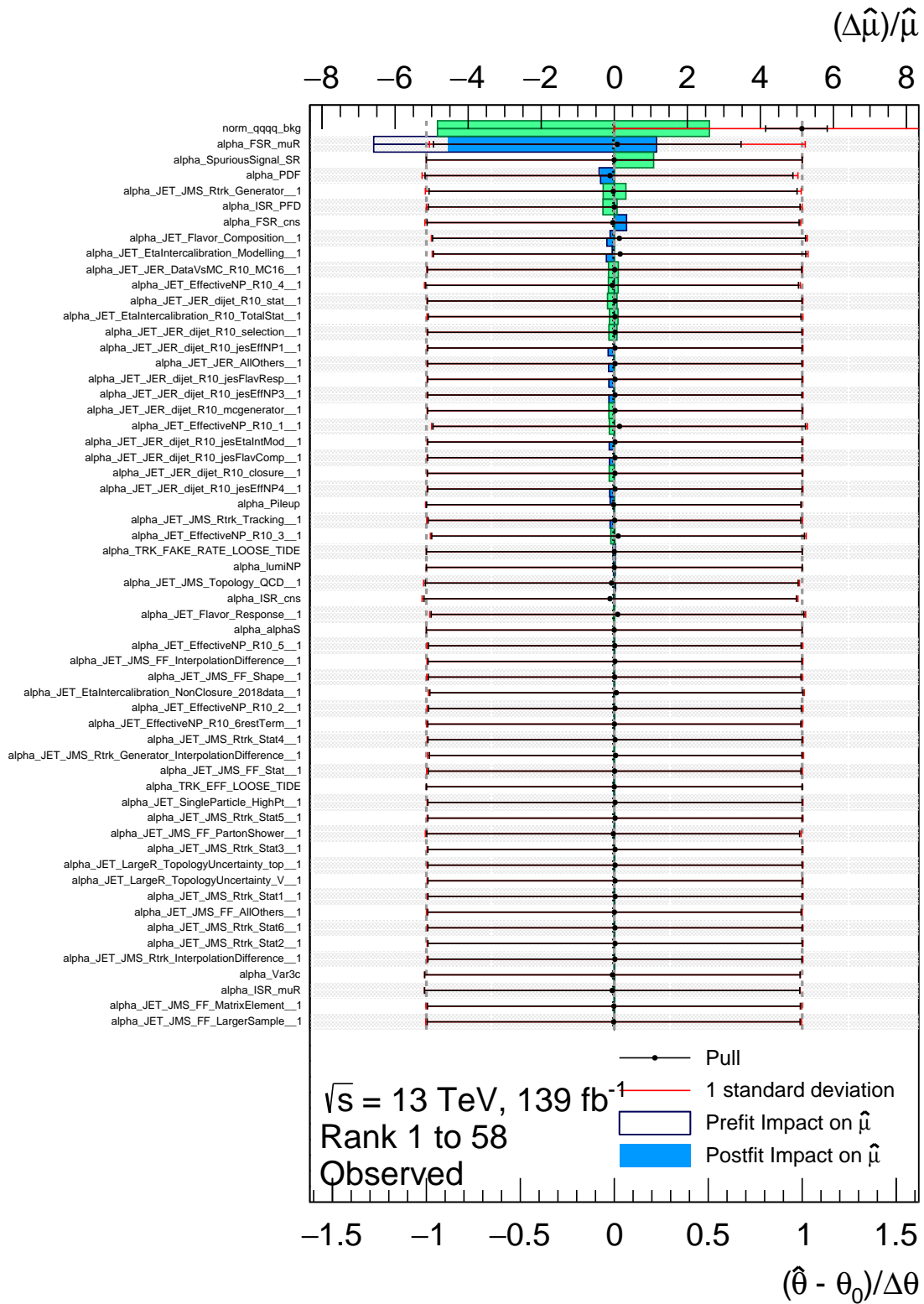
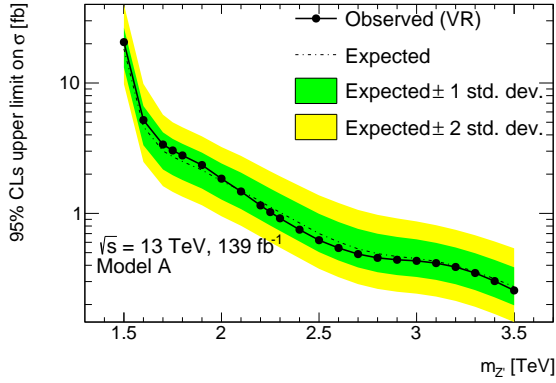


FIGURE 6.60 – Classement des paramètres de nuisance pour le Modèle D avec  $M_{Z'} = 3,5 \text{ TeV}$ .

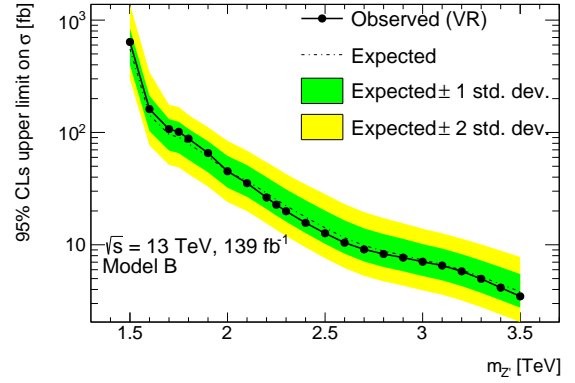


### 6.11.5 Limites

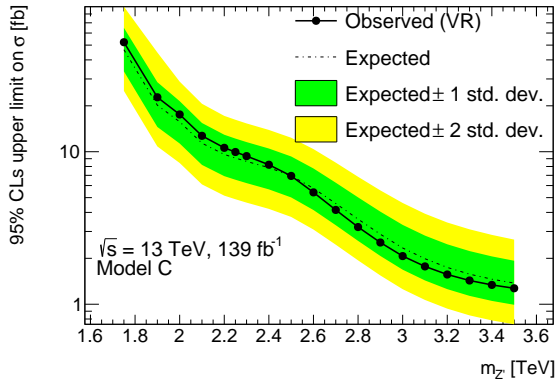
La Figure 6.61 représente la limite supérieure sur la section efficace, dont la procédure est décrite dans la Section 6.11.1, et ce pour tous les Modèles étudiés en fonction de la masse du  $Z'$  avec les incertitudes systématiques incluses. Sur ces Figures, « observed » représente les données dans la Région de Validation normalisées au nombre d'évènements attendus dans la Région de Signal, afin de tester la chaîne d'interprétation statistique. On peut remarquer que la gamme de valeur de l'axe des ordonnées est différent entre tous les Modèles, ce qui peut s'expliquer par une différence d'efficacité de sélection (voir la Figure 6.47).



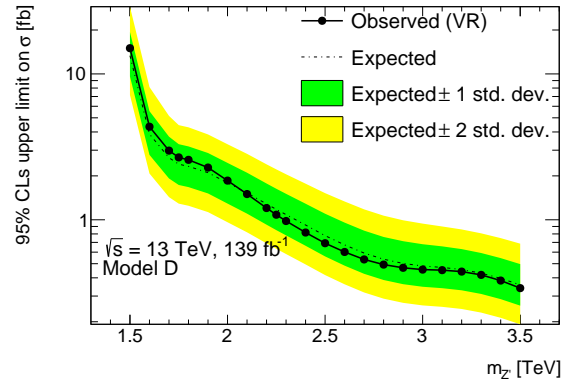
(a) Modèle A.



(b) Modèle B.



(c) Modèle C.



(d) Modèle D.

FIGURE 6.61 – Limites supérieures avec incertitudes systématiques pour tous les modèles étudiés en fonction de la masse du  $Z'$ . La valeur observée est obtenue avec les données dans la Région de Validation normalisées au nombre d'évènements attendus dans la Région de Signal.

## 6.12 Conclusion

En conclusion, cette analyse a pour objectif de rechercher un secteur sombre de type QCD qui serait lié au Modèle Standard via un médiateur  $Z'$ , une signature jamais explorée auparavant dans ATLAS. Les jets issus de ce secteur pourraient être alors identifiés

comme ayant un plus grand nombre de traces que ceux du Modèle Standard. Tous les jets sont reconstruits avec un rayon  $R = 1,0$ . L'existence du secteur sombre se manifesterait en une résonance dans la masse invariante de deux jets. Cependant, dû au *trimming* et aux effets de résolution, la résonance en masse invariante des jets sombres est très large, ce qui ne permet pas d'estimer le bruit de fond en ajustant directement les données dans la Région de Signal. C'est pourquoi une variable, appelée  $n_{\text{trk}}^\varepsilon$  est définie de façon à ce qu'une coupure sur le nombre de traces ne modifie pas la distribution en masse invariante du bruit de fond. Ceci permet d'estimer le bruit de fond à partir d'une région de contrôle. Cependant,  $n_{\text{trk}}^\varepsilon$ , est défini dans les simulations de bruit de fond et puis est utilisé pour les données. Des incertitudes systématiques sont associées à ce choix.

L'interprétation statistique a été dérivée, les résultats sont satisfaisants. En particulier, utiliser une région de contrôle afin d'estimer le bruit de fond permet d'obtenir un nombre d'évènements de signal injecté approximativement égal au nombre d'évènements de signal attendu. Les limites attendues sur la section efficace de production du  $Z'$  décroissent en fonction de sa masse : pour un médiateur de 3,5 TeV, les limites vont de 0,2 fb à 3 fb selon les modèles. Au moment d'écrire cette thèse, l'analyse débute son cycle d'approbation dans ATLAS, ce qui devrait mener à une publication de ces résultats.

La recherche de ce type de modèles est encore très récente puisqu'il n'y a eu jusqu'à présent qu'une seule analyse visant à la recherche d'un secteur de QCD sombre qui a été rendue publique par ATLAS [100]. Cependant, cette analyse est différente de celle présentée ici puisque la signature qui y est étudiée correspond à celle des jets « semi-visibles », tels que décrits dans la Section 2.5. C'est la première fois que la recherche d'un secteur de QCD sombre se manifestant par des particules ayant un faible temps de vie, et qui se désintègrent pour la plupart, est étudiée par ATLAS (aucune analyse n'a été rendue publique par CMS pour le moment). Par conséquent, cette analyse pourrait éventuellement bénéficier de futures itérations, par exemple avec les données du Run 3. Plusieurs points pourraient être étudiés afin de faciliter ces futures itérations. Le plus important d'entre eux consiste probablement à améliorer la reconstruction des jets, par exemple en utilisant les *Unified Flow Objects* avec *Soft-drop* à la place des LCTopos avec *trimming*. En effet, la recherche d'une résonance étroite permettrait d'ajuster directement les données dans la Région de Signal avec une fonction de bruit de fond seulement, à la place d'estimer le bruit de fond grâce à une Région de Contrôle. Un second point dont pourrait bénéficier l'analyse consiste à améliorer la modélisation des données par les simulations de certaines variables, et en particulier le nombre de traces qui sont associées aux jets.

# Conclusion générale

La nature de la matière noire demeure à ce jour un mystère. Une hypothèse permettant de résoudre ce problème implique l'existence d'un secteur sombre qui serait lié à celui du Modèle Standard, et dans lequel il y a des quarks sombres qui interagissent entre eux via une QCD sombre. Ce secteur sombre se manifesterait alors au LHC par la production de jets avec une signature particulière. Ce type de recherches est relativement récent et très peu de publications des expériences ATLAS et CMS n'existent au sujet de la QCD sombre, d'où l'importance d'exploiter ce type de signatures.

Durant ma thèse, j'ai contribué à la comparaison et à l'étalonnage d'une nouvelle technique de reconstruction des jets : les Unified Flow Objects. Pour ce faire, j'ai dérivé toute la chaîne d'étalonnage des jets : la correction d'empilement, l'étalonnage absolue en énergie, en masse et en direction, et le « GSC ». A chacune de ces étapes, les performances des jets ont été comparées. Ce travail permet de montrer que les UFOs permettent d'améliorer les performances des jets, et en particulier en termes de masse, par rapport aux définitions de jets jusqu'à présent utilisées dans ATLAS. Ce travail a donné lieu à l'élaboration d'une note publique [65]. Je suis allé présenter ce travail à la conférence BOOST 2022 au nom de la collaboration ATLAS. J'ai aussi contribué à la dérivation du GSC dans le cadre de « recommandations de précision » (c'est-à-dire une procédure d'étalonnage consolidée et très précise) dont un article est aussi en cours d'approbation dans ATLAS.

J'ai aussi contribué à une analyse recherchant un modèle particulier de secteur sombre, qui se manifesterait par un grand nombre de traces associées aux jets. J'ai été le principal contributeur à cette analyse. J'ai contribué à la plupart des étapes de l'analyse : de la production de fichiers à l'interprétation statistique. Cette analyse a pour stratégie de rechercher une résonance dans la distribution en masse invariante de deux jets. Cependant, dû à des signaux qui sont larges, l'estimation du bruit de fond ne peut être obtenue en ajustant directement les données dans la région de signal. Par conséquent, une nouvelle stratégie a été mise en place, qui décorrèle la sélection sur le nombre de traces de la masse invariante des deux jets. Cette stratégie permet d'estimer le bruit de fond à partir d'une région de contrôle. J'ai aussi contribué à l'évaluation des incertitudes systématiques. J'ai ensuite dérivé l'interprétation statistique de l'analyse : tests de faux signal, test d'injection de signal, production des limites etc. Ce travail de définition d'une toute nouvelle analyse a pris du temps, ce qui explique que l'analyse initie la phase d'approbation dans ATLAS au moment d'écrire cette thèse, ce qui devrait mener à une publication.

# Références

- [1] C. BURGARD. *Example : Standard model of physics*. 2016. URL : <https://texample.net/tikz/examples/model-physics/> (visité le 04/06/2022).
- [2] Peter W. HIGGS. « Broken Symmetries and the Masses of Gauge Bosons ». In : *Phys. Rev. Lett.* 13 (1964). Sous la dir. de J. C. TAYLOR, p. 508-509. DOI : 10.1103/PhysRevLett.13.508.
- [3] WIKIPEDIA. *Diagramme de Feynman*. 2022. URL : [https://fr.wikipedia.org/wiki/Diagramme\\_de\\_Feynman](https://fr.wikipedia.org/wiki/Diagramme_de_Feynman) (visité le 04/06/2022).
- [4] Particle Data GROUP. « Review of Particle Physics ». In : *PTEP* 2020.8 (2020). and 2021 update, p. 083C01. DOI : 10.1093/ptep/ptaa104.
- [5] Stefan HÖCHE. « Introduction to parton-shower event generators ». In : (2015), p. 235-295. DOI : 10.1142/9789814678766\_0005. arXiv : 1411.4085 [hep-ph].
- [6] F. ZWICKY. « Die Rotverschiebung von extragalaktischen Nebeln ». In : *Helv. Phys. Acta* 6 (1933), p. 110-127. DOI : 10.1007/s10714-008-0707-4.
- [7] V. C. RUBIN, N. THONNARD et W. K. FORD Jr. « Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/ ». In : *Astrophys. J.* 238 (1980), p. 471. DOI : 10.1086/158003.
- [8] K. G. BEGEMAN, A. H. BROEILS et R. H. SANDERS. « Extended rotation curves of spiral galaxies : Dark haloes and modified dynamics ». In : *Mon. Not. Roy. Astron. Soc.* 249 (1991), p. 523.
- [9] Agence Spatiale EUROPÉENNE. *The Bullet Cluster*. 2007. URL : [https://www.esa.int/ESA\\_Multimedia/Images/2007/07/The\\_Bullet\\_Cluster2](https://www.esa.int/ESA_Multimedia/Images/2007/07/The_Bullet_Cluster2) (visité le 03/01/2022).
- [10] M. MILGROM. « A Modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis ». In : *Astrophys. J.* 270 (1983), p. 365-370. DOI : 10.1086/161130.
- [11] Maxim MARKEVITCH et al. « Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56 ». In : *Astrophys. J.* 606 (2004), p. 819-824. DOI : 10.1086/383178. arXiv : astro-ph/0309303.
- [12] Arno A. PENZIAS et Robert Woodrow WILSON. « A Measurement of excess antenna temperature at 4080-Mc/s ». In : *Astrophys. J.* 142 (1965), p. 419-421. DOI : 10.1086/148307.

- 
- [13] N. AGHANIM et al. « Planck 2018 results. I. Overview and the cosmological legacy of Planck ». In : *Astron. Astrophys.* 641 (2020), A1. DOI : 10.1051/0004-6361/201833880. arXiv : 1807.06205 [astro-ph.CO].
- [14] N. AGHANIM et al. « Planck 2018 results. VI. Cosmological parameters ». In : *Astron. Astrophys.* 641 (2020). [Erratum : *Astron. Astrophys.* 652, C4 (2021)], A6. DOI : 10.1051/0004-6361/201833910. arXiv : 1807.06209 [astro-ph.CO].
- [15] Jacob D. BEKENSTEIN. « Relativistic gravitation theory for the MOND paradigm ». In : *Phys. Rev. D* 70 (2004). [Erratum : *Phys. Rev. D* 71, 069901 (2005)], p. 083509. DOI : 10.1103/PhysRevD.70.083509. arXiv : astro-ph/0403694.
- [16] Daniel BAUMANN. « Primordial Cosmology ». In : *PoS TASI2017* (2018), p. 009. DOI : 10.22323/1.305.0009. arXiv : 1807.03098 [hep-th].
- [17] Stephen P. MARTIN. « A Supersymmetry primer ». In : *Adv. Ser. Direct. High Energy Phys.* 18 (1998). Sous la dir. de Gordon L. KANE, p. 1-98. DOI : 10.1142/9789812839657\_0001. arXiv : hep-ph/9709356.
- [18] Matthew R. BUCKLEY. *Paper Explainer : Hiding Thermal Dark Matter with Leptons*. 2017. URL : <http://www.physicsmatt.com/blog/2017/5/13/paper-explainer-hiding-thermal-dark-matter-with-leptons> (visité le 03/01/2022).
- [19] ATLAS COLLABORATION. *Searching for Dark Matter with the ATLAS detector*. 2019. URL : <https://atlas.cern/updates/feature/dark-matter> (visité le 03/01/2022).
- [20] Myeonghun PARK et Mengchao ZHANG. « Tagging a jet from a dark sector with Jet-substructures at colliders ». In : *Phys. Rev. D* 100.11 (2019), p. 115009. DOI : 10.1103/PhysRevD.100.115009. arXiv : 1712.09279 [hep-ph].
- [21] Pedro SCHWALLER, Daniel STOLARSKI et Andreas WEILER. « Emerging Jets ». In : *JHEP* 05 (2015), p. 059. DOI : 10.1007/JHEP05(2015)059. arXiv : 1502.05409 [hep-ph].
- [22] Timothy COHEN et al. « LHC Searches for Dark Sector Showers ». In : *JHEP* 11 (2017), p. 196. DOI : 10.1007/JHEP11(2017)196. arXiv : 1707.05326 [hep-ph].
- [23] Elias BERNREUTHER et al. « Strongly interacting dark sectors in the early Universe and at the LHC through a simplified portal ». In : *JHEP* 01 (2020), p. 162. DOI : 10.1007/JHEP01(2020)162. arXiv : 1907.04346 [hep-ph].
- [24] Fabienne MARCASTEL. « CERN's Accelerator Complex. La chaîne des accélérateurs du CERN ». In : (sept. 2013). General Photo. URL : <https://cds.cern.ch/record/1621583>.
- [25] ALICE COLLABORATION. « The ALICE experiment at the CERN LHC ». In : *JINST* 3 (2008), S08002. DOI : 10.1088/1748-0221/3/08/S08002.
- [26] ATLAS COLLABORATION. « The ATLAS Experiment at the CERN Large Hadron Collider ». In : *JINST* 3 (2008), S08003. DOI : 10.1088/1748-0221/3/08/S08003.
- [27] CMS COLLABORATION. « The CMS Experiment at the CERN LHC ». In : *JINST* 3 (2008), S08004. DOI : 10.1088/1748-0221/3/08/S08004.
- [28] LHCb COLLABORATION. « The LHCb Detector at the LHC ». In : *JINST* 3 (2008), S08005. DOI : 10.1088/1748-0221/3/08/S08005.

- [29] ATLAS COLLABORATION. « ATLAS data quality operations and performance for 20152018 data-taking ». In : *JINST* 15.04 (2020), P04003. DOI : 10.1088/1748-0221/15/04/P04003. arXiv : 1911.04632 [physics.ins-det].
- [30] I. NEUTELINGS. *CMS coordinate system*. 2021. URL : [https://tikz.net/axis3d\\_cms/](https://tikz.net/axis3d_cms/) (visité le 03/01/2022).
- [31] Joao PEQUENAO. « Event Cross Section in a computer generated image of the ATLAS detector. » Mars 2008. URL : <https://cds.cern.ch/record/1096081>.
- [32] J.J. GOODSON. « Search for Supersymmetry in States with Large Missing Transverse Momentum and Three Leptons including a Z-Boson ». Presented 17 Apr 2012. Thèse de doct. Stony Brook University, mai 2012.
- [33] Quantum DIARIES. *ATLAS magnet system*. 2011. URL : <https://www.quantumdiaries.org/wp-content/uploads/2011/05/exp-magnets.png> (visité le 05/09/2022).
- [34] ATLAS COLLABORATION. « Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2 ». In : *Eur. Phys. J. C* 77.10 (2017), p. 673. DOI : 10.1140/epjc/s10052-017-5225-7. arXiv : 1704.07983 [hep-ex].
- [35] ATLAS COLLABORATION. « Performance of the reconstruction of large impact parameter tracks in the ATLAS inner detector ». In : (juill. 2017).
- [36] ATLAS COLLABORATION. *ATLAS Insertable B-Layer Technical Design Report*. Rapp. tech. Sept. 2010. URL : <https://cds.cern.ch/record/1291633>.
- [37] ATLAS COLLABORATION. « Jet reconstruction and performance using particle flow with the ATLAS Detector ». In : *Eur. Phys. J. C* 77.7 (2017), p. 466. DOI : 10.1140/epjc/s10052-017-5031-2. arXiv : 1703.10485 [hep-ex].
- [38] ATLAS COLLABORATION. « Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1 ». In : *Eur. Phys. J. C* 77 (2017), p. 490. DOI : 10.1140/epjc/s10052-017-5004-5. arXiv : 1603.02934 [hep-ex].
- [39] ATLAS COLLABORATION. « Jet energy scale and resolution measured in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector ». In : *Eur. Phys. J. C* 81.8 (2021), p. 689. DOI : 10.1140/epjc/s10052-021-09402-3. arXiv : 2007.02645 [hep-ex].
- [40] Arthur LINSS. « Search for Dark Matter in Invisible Higgs Boson Decays with the ATLAS Detector at the LHC ». Presented 14 Dec 2021. Oct. 2021. URL : <https://cds.cern.ch/record/2801844>.
- [41] Torbjorn SJOSTRAND, Stephen MRENNA et Peter Z. SKANDS. « PYTHIA 6.4 Physics and Manual ». In : *JHEP* 05 (2006), p. 026. DOI : 10.1088/1126-6708/2006/05/026. arXiv : hep-ph/0603175.
- [42] T. GLEISBERG et al. « Event generation with SHERPA 1.1 ». In : *JHEP* 02 (2009), p. 007. DOI : 10.1088/1126-6708/2009/02/007. arXiv : 0811.4622 [hep-ph].
- [43] G. CORCELLA et al. « HERWIG 6 : An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes) ». In : *JHEP* 01 (2001), p. 010. DOI : 10.1088/1126-6708/2001/01/010. arXiv : hep-ph/0011363.

- [44] S. AGOSTINELLI et al. « GEANT4—a simulation toolkit ». In : *Nucl. Instrum. Meth. A* 506 (2003), p. 250-303. DOI : 10.1016/S0168-9002(03)01368-8.
- [45] R. FRUHWIRTH. « Application of Kalman filtering to track and vertex fitting ». In : *Nucl. Instrum. Meth. A* 262 (1987), p. 444-450. DOI : 10.1016/0168-9002(87)90887-4.
- [46] D. WICKE. « A New Algorithm for Solving Tracking Ambiguities ». In : (oct. 1998).
- [47] T. CORNELISSEN et al. « The new ATLAS track reconstruction (NEWT) ». In : *J. Phys. Conf. Ser.* 119 (2008). Sous la dir. de Randall SOBIE, Reda TAFIROUT et Jana THOMSON, p. 032014. DOI : 10.1088/1742-6596/119/3/032014.
- [48] ATLAS COLLABORATION. *ATLAS Track Reconstruction – General Overview*. 2021. URL : <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/overview/> (visité le 04/06/2022).
- [49] ATLAS COLLABORATION. « Early Inner Detector Tracking Performance in the 2015 data at  $\sqrt{s} = 13$  TeV ». In : (2015).
- [50] CMS COLLABORATION. « Particle-flow reconstruction and global event description with the CMS detector ». In : *JINST* 12.10 (2017), P10003. DOI : 10.1088/1748-0221/12/10/P10003. arXiv : 1706.04965 [physics.ins-det].
- [51] CMS COLLABORATION. « Pileup mitigation at CMS in 13 TeV data ». In : *JINST* 15.09 (2020), P09018. DOI : 10.1088/1748-0221/15/09/P09018. arXiv : 2003.00503 [hep-ex].
- [52] ATLAS COLLABORATION. « Improving jet substructure performance in ATLAS using Track-CaloClusters ». In : (juill. 2017). URL : <https://cds.cern.ch/record/2275636>.
- [53] Andrew J. LARKOSKI, Gavin P. SALAM et Jesse THALER. « Energy Correlation Functions for Jet Substructure ». In : *JHEP* 06 (2013), p. 108. DOI : 10.1007/JHEP06(2013)108. arXiv : 1305.0007 [hep-ph].
- [54] Andrew J. LARKOSKI, Ian MOULT et Duff NEILL. « Power Counting to Better Jet Observables ». In : *JHEP* 12 (2014), p. 009. DOI : 10.1007/JHEP12(2014)009. arXiv : 1409.6298 [hep-ph].
- [55] Matteo CACCIARI, Gavin P. SALAM et Gregory SOYEZ. « The anti- $k_t$  jet clustering algorithm ». In : *JHEP* 04 (2008), p. 063. DOI : 10.1088/1126-6708/2008/04/063. arXiv : 0802.1189 [hep-ph].
- [56] Matteo CACCIARI, Gavin P. SALAM et Gregory SOYEZ. « FastJet User Manual ». In : *Eur. Phys. J. C* 72 (2012), p. 1896. DOI : 10.1140/epjc/s10052-012-1896-2. arXiv : 1111.6097 [hep-ph].
- [57] Peter BERTA et al. « Particle-level pileup subtraction for jets and jet shapes ». In : *JHEP* 06 (2014), p. 092. DOI : 10.1007/JHEP06(2014)092. arXiv : 1403.3108 [hep-ex].
- [58] ATLAS COLLABORATION. « Determination of jet calibration and energy resolution in proton-proton collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector ». In : *Eur. Phys. J. C* 80.12 (2020), p. 1104. DOI : 10.1140/epjc/s10052-020-08477-8. arXiv : 1910.04482 [hep-ex].

- [59] ATLAS COLLABORATION. « Constituent-level pile-up mitigation techniques in ATLAS ». In : (août 2017).
- [60] Matteo CACCIARI, Gavin P. SALAM et Gregory SOYEZ. « SoftKiller, a particle-level pileup removal method ». In : *Eur. Phys. J. C* 75.2 (2015), p. 59. DOI : 10.1140/epjc/s10052-015-3267-2. arXiv : 1407.0408 [hep-ph].
- [61] David KROHN, Jesse THALER et Lian-Tao WANG. « Jet Trimming ». In : *JHEP* 02 (2010), p. 084. DOI : 10.1007/JHEP02(2010)084. arXiv : 0912.1342 [hep-ph].
- [62] ATLAS COLLABORATION. « Performance of jet substructure techniques for large- $R$  jets in proton-proton collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector ». In : *JHEP* 09 (2013), p. 076. DOI : 10.1007/JHEP09(2013)076. arXiv : 1306.4945 [hep-ex].
- [63] Andrew J. LARKOSKI et al. « Soft Drop ». In : *JHEP* 05 (2014), p. 146. DOI : 10.1007/JHEP05(2014)146. arXiv : 1402.2657 [hep-ph].
- [64] ATLAS COLLABORATION. « Measurement of the Soft-Drop Jet Mass in pp Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector ». In : *Phys. Rev. Lett.* 121.9 (2018), p. 092001. DOI : 10.1103/PhysRevLett.121.092001. arXiv : 1711.08341 [hep-ex].
- [65] ATLAS COLLABORATION. « R=0.4 jets input comparison and Monte Carlo calibration with the ATLAS Detector ». In : (2022). All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-038>. URL : <https://cds.cern.ch/record/2824558>.
- [66] ATLAS COLLABORATION. « Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV protonproton collisions ». In : *Eur. Phys. J. C* 81.4 (2021), p. 334. DOI : 10.1140/epjc/s10052-021-09054-3. arXiv : 2009.04986 [hep-ex].
- [67] Aviv CUKIERMAN et Benjamin NACHMAN. « Mathematical Properties of Numerical Inversion for Jet Calibrations ». In : *Nucl. Instrum. Meth. A* 858 (2017), p. 1-11. DOI : 10.1016/j.nima.2017.03.038. arXiv : 1609.05195 [physics.data-an].
- [68] Bjarne GRIMSTAD et al. *SPLINTER : a library for multivariate function approximation with splines*. <http://github.com/bgrimstad/splinter>. Accessed : 2015-05-16. 2015.
- [69] Pauli VIRTANEN et al. « SciPy 1.0 : Fundamental Algorithms for Scientific Computing in Python ». In : *Nature Methods* 17 (2020), p. 261-272. DOI : 10.1038/s41592-019-0686-2.
- [70] Yuri L. DOKSHITZER. « Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. » In : *Sov. Phys. JETP* 46 (1977), p. 641-653.
- [71] Guido ALTARELLI et G. PARISI. « Asymptotic Freedom in Parton Language ». In : *Nucl. Phys. B* 126 (1977), p. 298-318. DOI : 10.1016/0550-3213(77)90384-4.
- [72] V. N. GRIBOV et L. N. LIPATOV. « Deep inelastic e p scattering in perturbation theory ». In : *Sov. J. Nucl. Phys.* 15 (1972), p. 438-450.



- [73] Richard D. BALL et al. « Parton distributions with LHC data ». In : *Nucl. Phys. B* 867 (2013), p. 244-289. DOI : 10.1016/j.nuclphysb.2012.10.003. arXiv : 1207.1303 [hep-ph].
- [74] ATLAS COLLABORATION. « ATLAS Pythia 8 tunes to 7 TeV data ». In : (nov. 2014).
- [75] Bo ANDERSSON et al. « Parton Fragmentation and String Dynamics ». In : *Phys. Rept.* 97 (1983), p. 31-145. DOI : 10.1016/0370-1573(83)90080-7.
- [76] Gosta GUSTAFSON et Ulf PETTERSSON. « Dipole Formulation of QCD Cascades ». In : *Nucl. Phys. B* 306 (1988), p. 746-758. DOI : 10.1016/0550-3213(88)90441-5.
- [77] ATLAS COLLABORATION. « The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model ». In : (août 2016).
- [78] ATLAS COLLABORATION. « Jet reclustering and close-by effects in ATLAS run II ». In : (juill. 2017).
- [79] Malte BUSCHMANN et al. « Lepton Jets from Radiating Dark Matter ». In : *JHEP* 07 (2015), p. 045. DOI : 10.1007/JHEP07(2015)045. arXiv : 1505.07459 [hep-ph].
- [80] Lisa CARLONI et Torbjorn SJOSTRAND. « Visible Effects of Invisible Hidden Valley Radiation ». In : *JHEP* 09 (2010), p. 105. DOI : 10.1007/JHEP09(2010)105. arXiv : 1006.2911 [hep-ph].
- [81] Joshua ELLIS. « TikZ-Feynman : Feynman diagrams with TikZ ». In : *Comput. Phys. Commun.* 210 (2017), p. 103-123. DOI : 10.1016/j.cpc.2016.08.019. arXiv : 1601.05437 [hep-ph].
- [82] ATLAS COLLABORATION. « The performance of the jet trigger for the ATLAS detector during 2011 data taking ». In : *Eur. Phys. J. C* 76.10 (2016), p. 526. DOI : 10.1140/epjc/s10052-016-4325-0. arXiv : 1606.07759 [hep-ex].
- [83] ATLAS COLLABORATION. *Public Jet Trigger Plots for Collision Data*. 2022. URL : <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetTriggerPublicResult> (visité le 04/06/2022).
- [84] ATLAS COLLABORATION. « Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector ». In : (juill. 2015).
- [85] Matteo CACCIARI, Gavin P. SALAM et Gregory SOYEZ. « The Catchment Area of Jets ». In : *JHEP* 04 (2008), p. 005. DOI : 10.1088/1126-6708/2008/04/005. arXiv : 0802.1188 [hep-ph].
- [86] Glen COWAN et al. « Asymptotic formulae for likelihood-based tests of new physics ». In : *Eur. Phys. J. C* 71 (2011). [Erratum : *Eur.Phys.J.C* 73, 2501 (2013)], p. 1554. DOI : 10.1140/epjc/s10052-011-1554-0. arXiv : 1007.1727 [physics.data-an].
- [87] ATLAS COLLABORATION. « Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS ». In : (2018).
- [88] ROOT. *GetQuantiles*. 2022. URL : <https://root.cern.ch/doc/master/classTH1.html#a2be4a5889f7421047d68b0a398675b8d> (visité le 04/06/2022).

- [89] ATLAS COLLABORATION. « In situ calibration of large-radius jet energy and mass in 13 TeV protonproton collisions with the ATLAS detector ». In : *Eur. Phys. J. C* 79.2 (2019), p. 135. DOI : 10.1140/epjc/s10052-019-6632-8. arXiv : 1807.09477 [hep-ex].
- [90] ATLAS COLLABORATION. « Luminosity determination in  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC ». In : (juin 2019).
- [91] G. AVONI et al. « The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS ». In : *JINST* 13.07 (2018), P07017. DOI : 10.1088/1748-0221/13/07/P07017.
- [92] Jon BUTTERWORTH et al. « PDF4LHC recommendations for LHC Run II ». In : *J. Phys. G* 43 (2016), p. 023001. DOI : 10.1088/0954-3899/43/2/023001. arXiv : 1510.03865 [hep-ph].
- [93] ROOT. *RoMomentMorph Class Reference*. 2022. URL : <https://root.cern.ch/doc/master/classRoMomentMorph.html> (visité le 04/06/2022).
- [94] Jerzy NEYMAN et Egon Sharpe PEARSON. « On the Problem of the Most Efficient Tests of Statistical Hypotheses ». In : *Phil. Trans. Roy. Soc. Lond. A* 231.694-706 (1933), p. 289-337. DOI : 10.1098/rsta.1933.0009.
- [95] Alexander L. READ. « Presentation of search results : The CL(s) technique ». In : *J. Phys. G* 28 (2002). Sous la dir. de M. R. WHALLEY et L. LYONS, p. 2693-2704. DOI : 10.1088/0954-3899/28/10/313.
- [96] I. NEUTELINGS. *Gaussian distributions and statistical tests*. 2021. URL : <https://tikz.net/gaussians/> (visité le 04/06/2022).
- [97] S. S. WILKS. « The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses ». In : *Annals Math. Statist.* 9.1 (1938), p. 60-62. DOI : 10.1214/aoms/1177732360.
- [98] Abraham WALD. « Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large ». In : *Transactions of the American Mathematical Society* 54.3 (1943), p. 426-482. ISSN : 00029947. URL : <http://www.jstor.org/stable/1990256> (visité le 30/07/2022).
- [99] ATLAS COLLABORATION. « Recommendations for the Modeling of Smooth Backgrounds ». In : (nov. 2020). URL : <https://cds.cern.ch/record/2743717>.
- [100] ATLAS COLLABORATION. « Search for non-resonant production of semi-visible jets using Run 2 data in ATLAS ». In : (2022).