



**HAL**  
open science

# Recherche d'un boson de Higgs additionnel léger se désintégrant en deux photons au sein de l'expérience CMS au LHC et interprétations

Antoine Lesauvage

► **To cite this version:**

Antoine Lesauvage. Recherche d'un boson de Higgs additionnel léger se désintégrant en deux photons au sein de l'expérience CMS au LHC et interprétations. Physique des accélérateurs [physics.acc-ph]. Université de Lyon, 2021. Français. NNT : 2021LYSE1156 . tel-03495493

**HAL Id: tel-03495493**

**<https://theses.hal.science/tel-03495493>**

Submitted on 20 Dec 2021

**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.



Université Claude Bernard  Lyon 1

N°d'ordre NNT : 2021LYSE1156

**THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON**

Opérée au sein de :

**l'Université Claude Bernard Lyon 1**

**École Doctorale N°52**

**École Doctorale de Physique et Astrophysique de Lyon**

**Spécialité de doctorat : Physique des particules**

**Discipline : Physique**

Soutenue publiquement le 03/09/2021, par :

**Antoine Lesauvage**

---

**Recherche d'un boson de Higgs additionnel léger se désintégrant en deux photons au sein de l'expérience CMS au LHC et interprétations**

---

Devant le jury composé de :

<b>A. EALET</b>	<b>Présidente du jury</b>
<b>R. NIKOLAIDOU</b>	<b>Rapportrice</b>
<b>P. VANLAER</b>	<b>Rapporteur</b>
<b>N. BERGER</b>	<b>Examineur</b>
<b>G. HAMEL de MONCHENAULT</b>	<b>Examineur</b>
<b>N. MAHMOUDI</b>	<b>Examinatrice</b>
<b>G. MOREAU</b>	<b>Examineur</b>
<b>S. GASCON-SHOTKIN</b>	<b>Directrice de thèse</b>
<b>A. DEANDREA</b>	<b>Co-Directeur de thèse, Invité</b>
<b>M. LETHUILLER</b>	<b>Co-Directeur de thèse, Invité</b>

---

## Résumé

Ce manuscrit présente la recherche d'une nouvelle particule au sein du détecteur CMS au LHC (CERN). On y détaille notamment l'analyse permettant la recherche d'un boson de Higgs additionnel de masse inférieure à 125 GeV qui se désintégrerait en deux photons ainsi que des interprétations phénoménologiques présentant de telles particules.

L'existence d'un nouveau boson de Higgs est, en effet, motivée par de nombreuses théories allant au-delà du Modèle Standard de la physique des particules, dont quelques unes sont présentées ici.

Le détecteur CMS est aussi décrit afin de comprendre en profondeur la physique présentée dans l'analyse. Le canal de désintégration étudié étant le canal diphoton, il est important d'étudier en détails la façon dont les photons sont reconstruits et identifiés au sein du détecteur CMS. Les corrections apportées à l'énergie des photons sont notamment cruciales pour les analyses travaillant sur des bosons de Higgs. On explicite ici, à cet effet, une méthode de validation de l'échelle d'énergie des photons sur des échantillons de photons. On est alors en mesure de présenter les méthodes d'analyse de données, utilisant notamment des méthodes d'apprentissage automatisé, pour la recherche d'un nouveau boson de Higgs de basse masse et les résultats obtenus pour les prises de données des années 2017 et 2018.

Dans cette continuité, différentes pistes d'amélioration à mettre en place, notamment en ce qui concerne la modélisation de la composante Drell-Yan du bruit de fond de l'analyse, seront aussi abordées.

Finalement, des interprétations phénoménologiques seront présentées dans l'objectif de poser des contraintes sur des modèles théoriques spécifiques et de statuer sur la possibilité de découvrir d'autres bosons de Higgs additionnels au LHC et dans les collisionneurs de particules à venir.

---

## Abstract

This thesis presents the search for a new particle within the CMS experiment in the LHC (CERN). The search analysis for an additional low mass (lower than 125 GeV) Higgs boson decaying into two photons is shown with some phenomenological interpretations.

The very existence of this new Higgs boson is, indeed, motivated by a lot of beyond the Standard Model theories, some of which contain low mass Higgs bosons and are presented here.

The CMS detector is described so that one can understand the physical aspects of the search analysis. It is also essential that the way photons are reconstructed and identified, within the CMS detector, is studied because of the diphoton decay channel that is considered in this thesis. Especially, the reconstructed photon energy needs to be corrected to account for energy leakage within the CMS electromagnetic calorimeter and a dedicated method, that enables the validation of such corrections on photons, is presented here.

With this in mind, the search analysis methods, widely using multivariate variable tools, which goal is to find a new low mass Higgs boson are shown together with the derived results using 2017 and 2018 data years. Following this step, some elements and discussions for analysis improvements, especially concerning the modeling of the Drell-Yan background, will be studied.

Finally, constraints on some theoretical models are derived and the possibility of discovering other Higgs bosons at the LHC and in future colliders is discussed thanks to phenomenological works.



## Remerciements

Mes premiers remerciements vont à mes directeurices de thèse, Suzanne Gascon-Shotkin, Morgan Lethuiller et Aldo Deandrea, pour leurs conseils, supports et implications dans ce projet de recherche. Je tiens à remercier plus particulièrement Suzanne pour son appui et ses recommandations, ainsi que Morgan pour les captivantes discussions échangées, professionnelles ou non, et l'aide apportée.

Je remercie sincèrement les chercheurs et chercheuses qui ont aimablement accepté de prendre le temps de faire partie de ce jury de thèse. Mes plus chaleureux remerciements à Nicolas Berger, Anne Ealet, Gautier Hammel de Monchenault, Nazila MahMoudi, Grégory Moreau, ainsi qu'à Rosy Nikolaidou et Pascal Vanlaer qui ont endossé le rôle de rapporteurices et ont ainsi contribué, par leurs retours et commentaires, à rendre plus agréable la lecture de ce manuscrit.

Je souhaite aussi remercier l'entièreté des (ex-)membres du groupe CMS de l'Institut de Physique de Deux Infinis de Lyon pour leur accueil, leur enthousiasme et tous les moments partagés ensemble. Un franc merci à Bernard pour tous ses précieux conseils et échanges, ainsi que sa bienveillante relecture de ce manuscrit.

Mes remerciements à tous les membres de l'Institut de Physique de Deux Infinis de Lyon pour m'avoir permis de mener à bien mes travaux dans de très bonnes conditions de travail, malgré les crises sanitaires.

Merci finalement à toute les personnes qui ont eu un rôle, peu importe l'ampleur, dans la réalisation et la rédaction de ce manuscrit. Une liste exhaustive serait encombrante mais sachez que vos actes comptent.

*Merci au curieux.se lisant ce manuscrit, puisse sa lecture être agréable.*

*A l'illusion que la science  
Insuffle à mon existence.*

*Une forme d'innocence du vide  
Inclue dans un modèle répétitif général.*

# Table des matières

<b>Table des matières</b>	<b>3</b>
<b>1 Le Modèle Standard et le boson de Higgs</b>	<b>9</b>
1.1 Le formalisme du Modèle Standard	10
1.1.1 Électrodynamique Quantique	10
1.1.2 Chromodynamique Quantique	11
1.1.3 Interaction faible et électrofaible	14
1.1.4 Mécanisme de Higgs	17
1.2 Le boson de Higgs au LHC	20
1.2.1 Le grand collisionneur à hadrons	20
1.2.2 Les propriétés du boson de Higgs	25
1.3 Les limites du Modèle Standard	31
1.3.1 Problèmes de naturalité	31
1.3.2 Phénomènes cosmologiques	32
1.3.3 Autres limites	33
<b>2 Les bosons de Higgs au-delà du Modèle Standard</b>	<b>35</b>
2.1 Modèles à deux doublets de Higgs	36
2.1.1 Motivations	36
2.1.2 Théorie générale	36
2.1.3 Bosons de Higgs	38
2.2 Modèles de Higgs Composites	40
2.2.1 Introduction aux théories Technicouleur et Higgs Composites	40
2.2.2 Dynamique Composite Fondamentale $SU(4) \rightarrow Sp(4)$	42
2.3 Modèles Supersymétriques	46
<b>3 Reconstruction des évènements dans le détecteur CMS</b>	<b>51</b>
3.1 Le détecteur CMS du LHC	52
3.1.1 Présentation du détecteur CMS	52
3.2 Reconstruction et simulation des évènements physiques	68
3.2.1 Algorithme de Particule Flow	69
3.2.2 Simulation du détecteur	72
<b>4 Reconstruction et identification des photons</b>	<b>75</b>

4.1	Reconstruction de l'énergie des photons par cristal . . . . .	76
4.1.1	Reconstruction des signaux $A$ dans le calorimètre électromagnétique . . . . .	76
4.1.2	Réponse et monitoring dans le calorimètre électromagnétique $LC$ . . . . .	77
4.1.3	Calibrations du calorimètre électromagnétique $IC$ . . . . .	78
4.2	Reconstruction des photons au sein de l'algorithme Particle Flow . . . . .	79
4.3	Corrections à l'énergie des photons . . . . .	82
4.3.1	Correction des pertes d'énergie $F_\gamma$ . . . . .	83
4.3.2	Corrections à l'échelle d'énergie des données et dégradation de la résolution en énergie dans les simulations $s_\gamma$ . . . . .	84
4.3.3	Validation et performance des corrections . . . . .	88
<b>5</b>	<b>Extraction de l'échelle d'énergie des photons dans CMS</b> . . . . .	<b>91</b>
5.1	Les évènements $Z \rightarrow \mu\mu\gamma$ . . . . .	92
5.1.1	La désintégration $Z \rightarrow \mu\mu\gamma$ . . . . .	92
5.1.2	Description des bruits de fond $\mu\mu\gamma$ . . . . .	93
5.1.3	Sélection des évènements . . . . .	93
5.2	L'échelle d'énergie des photons . . . . .	94
5.2.1	Définition de l'échelle d'énergie . . . . .	94
5.2.2	Expression littérale . . . . .	95
5.3	Extraction de l'échelle d'énergie . . . . .	97
5.3.1	Méthode d'extraction . . . . .	97
5.3.2	Résultats obtenus . . . . .	103
<b>6</b>	<b>Analyse pour la recherche de <math>h \rightarrow \gamma\gamma</math> à basses masses</b> . . . . .	<b>107</b>
6.1	Stratégie de l'analyse . . . . .	109
6.1.1	Stratégie générale . . . . .	109
6.1.2	Stratégie détaillée . . . . .	111
6.2	Échantillons utilisés . . . . .	113
6.2.1	Données . . . . .	113
6.2.2	Simulations . . . . .	113
6.3	Chemin de déclenchement et efficacités . . . . .	114
6.3.1	Description des chemins HLT . . . . .	114
6.3.2	Efficacités des chemins de déclenchement . . . . .	117
6.4	Sélection de l'analyse . . . . .	119
6.4.1	Critères de présélection . . . . .	119
6.4.2	Efficacités de présélection . . . . .	121
6.5	Identification des vertex diphoton . . . . .	123
6.5.1	BDT d'identification de vertex . . . . .	124
6.5.2	Probabilité d'estimer le bon vertex . . . . .	125
6.5.3	Validation et facteurs d'échelle . . . . .	125
6.6	Identification des photons . . . . .	126
6.6.1	Photon ID BDT . . . . .	127
6.6.2	Validation du photon ID BDT . . . . .	130

6.7	Classification des évènements . . . . .	134
6.7.1	Classificateur d'évènements : le diphoton BDT . . . . .	134
6.7.2	Catégorisations . . . . .	143
6.8	Modélisation du signal . . . . .	147
6.9	Modélisation du bruit de fond . . . . .	156
6.9.1	Modélisation de la composante Drell-Yan . . . . .	156
6.9.2	Modèle de bruit de fond total . . . . .	160
6.9.3	Incertitudes systématiques associées . . . . .	164
6.10	Incertitudes systématiques . . . . .	170
6.10.1	Listes des incertitudes systématiques pour le signal . . . . .	170
6.10.2	Impacts des incertitudes . . . . .	173
6.11	Résultats finaux . . . . .	178
6.11.1	Analyse statistique . . . . .	178
6.11.2	Résultats des tests statistiques . . . . .	183
6.12	Résultats avec catégorie VBF . . . . .	184
6.12.1	Catégorie VBF . . . . .	184
6.12.2	Résultats finaux . . . . .	187
6.12.3	Comparaison avec ATLAS . . . . .	191
<b>7</b>	<b>Études et perspectives d'amélioration</b> . . . . .	<b>195</b>
7.1	Relâchement des coupures sur la gamme de masse . . . . .	196
7.1.1	Extension de la gamme de masse . . . . .	196
7.1.2	Relâchement des coupures d'impulsion transverse normalisée . . . . .	198
7.2	Réoptimisation des catégories <i>diphoton BDT</i> pour 2018 . . . . .	203
7.3	Études de la modélisation de la composante Drell-Yan . . . . .	210
7.3.1	Ajout de la composante exponentielle . . . . .	210
7.3.2	Accords données/simulations pour les évènements Drell-Yan <i>double-fake</i> . . . . .	212
7.3.3	Régions d'inefficacité en $\phi$ . . . . .	230
<b>8</b>	<b>Études phénoménologiques</b> . . . . .	<b>235</b>
8.1	Recherche de désintégrations exotiques au LHC $h \rightarrow Za$ . . . . .	236
8.1.1	Introduction . . . . .	236
8.1.2	Désintégration $a \rightarrow \gamma\gamma$ . . . . .	237
8.1.3	Désintégration $a \rightarrow \mu\mu$ . . . . .	240
8.1.4	Interprétations en terme d'un modèle de particules axion-like . . . . .	241
8.2	Recherche de nouveaux scalaires au LHC par fusion de bosons vecteur . . . . .	245
8.2.1	Introduction . . . . .	245
8.2.2	Modèle effectif . . . . .	246
8.2.3	Distributions cinématiques simulées . . . . .	247
8.3	Recherche d'une nouvelle particule via un couplage $h\eta\gamma$ au HL-LHC . . . . .	250
8.3.1	Cadre du modèle . . . . .	250
8.3.2	Proto-analyse avec simulations . . . . .	251
8.3.3	Cas des collisionneurs $e^+e^-$ . . . . .	254

<b>A</b>	<b>Les arbres de décision boostés</b>	<b>263</b>
<b>B</b>	<b>Les variables du photon</b>	<b>267</b>
B.1	Les variables cinématiques . . . . .	267
B.2	Les variables de forme de gerbe . . . . .	267
B.3	Les variables d'isolation . . . . .	268
<b>C</b>	<b>Liste des échantillons utilisés</b>	<b>269</b>
C.1	Données . . . . .	269
C.2	Simulations - Signal . . . . .	270
C.3	Simulations - Bruit de fond . . . . .	278
<b>D</b>	<b>Figures de validation du diphoton BDT</b>	<b>279</b>
D.1	Validations zones de contrôle de données et simulations . . . . .	279
D.2	Validation $Z \rightarrow ee$ . . . . .	286
	<b>Bibliographie</b>	<b>293</b>

# Introduction

L'ambitieux objectif de la physique des particules est d'arriver à décrire le Monde et la matière au niveau le plus élémentaire afin de mieux les appréhender.

A cet effet, une théorie fut construite dans la deuxième partie du 20<sup>ème</sup> siècle réussissant le tour de force de dépendre - avec une précision inégalée - les interactions, considérées comme fondamentales, dites faible, forte et électromagnétique. Cette théorie, sobrement nommée Modèle Standard de la physique des particules, multiplia les succès théoriques et expérimentaux jusqu'à être couronnée par la découverte d'une particule compatible avec le boson de Higgs au LHC en 2012.

Malgré tous ses atouts, le Modèle Standard peine à décrire certaines phénomènes physiques ayant été observés, mettant alors en évidence des lacunes et les limites de ce modèle. De nombreuses théories étendant et dépassant le Modèle Standard ont alors été élaborées pour pallier ces problèmes. Ces nouveaux modèles prédisent souvent de nouvelles particules scalaires et pseudo-scalaires correspondant à des bosons de Higgs additionnels dont certains possèdent une masse inférieure à celle du boson de Higgs découvert en 2012. On présentera ici quelques unes de ces extensions à travers les modèles à deux doublets de Higgs, de Higgs composites et les modèles supersymétriques.

Afin de contraindre cet éventail de modèles, il s'avère opportun de chercher à observer de tels bosons de Higgs additionnels dans des détecteurs dédiés. On se concentrera ici sur des signatures comportant deux photons dans l'état final.

Le détecteur dans lequel on cherche à observer une telle particule est le détecteur CMS, l'un des deux détecteurs généralistes du LHC au CERN. Il apparaît donc essentiel de présenter ce détecteur et ses caractéristiques afin d'étudier son fonctionnement et comprendre comment les particules y sont reconstruites. On détaillera, à cet effet, les parties du détecteur associées à la détection des photons, c'est-à-dire les différentes parties du calorimètre électromagnétique.

Étant donné que l'on recherche ici une particule qui se désintégrerait en deux photons, il est important d'approfondir les méthodes utilisées pour reconstruire les photons au sein du détecteur, les méthodes permettant leur identification ainsi que celles employées pour corriger l'énergie des photons mesurés.

Une de ces méthodes corrige l'énergie en utilisant des électrons et doit donc être validée

sur des photons. Afin d'effectuer cette validation sur des échantillons photon de haute pureté, une technique basée sur l'observation de désintégrations  $Z \rightarrow \mu\mu\gamma$  a été mise au point et sera présentée ici.

Les photons du détecteur CMS étant reconstruits, on peut ensuite analyser les données correspondant à la recherche d'un boson de Higgs additionnel de masse inférieure à 125 GeV. L'analyse associée est complexe et repose, notamment, sur des méthodes d'apprentissage automatique comme les arbres de décision boostés. Les différentes étapes de la chaîne d'analyse ainsi que les résultats obtenus pour les données de 2017 et de 2018 du Run 2 du LHC seront alors présentés.

Dans la continuité, différentes études, concernant notamment la modélisation du bruit de fond de l'analyse, ont été réalisées évoquant alors des pistes d'amélioration et des éléments en développement ou à développer dans le futur.

Enfin, des interprétations phénoménologiques seront présentées dans l'objectif de poser des contraintes sur des modèles théoriques spécifiques et de statuer sur la possibilité de découvrir d'autres bosons de Higgs additionnels, à nouveau pour des modèles plus spécifiques, au LHC ou encore dans de futurs accélérateurs de particules comme le LHC à haute luminosité ou le Futur Collisionneur Circulaire à électron positron.



# Le Modèle Standard et le boson de Higgs

## Sommaire

---

1.1	Le formalisme du Modèle Standard . . . . .	10
1.1.1	Électrodynamique Quantique . . . . .	10
1.1.2	Chromodynamique Quantique . . . . .	11
1.1.3	Interaction faible et électrofaible . . . . .	14
1.1.4	Mécanisme de Higgs . . . . .	17
1.2	Le boson de Higgs au LHC . . . . .	20
1.2.1	Le grand collisionneur à hadrons . . . . .	20
	Présentation générale . . . . .	20
	Luminosité et empilement . . . . .	22
	Les détecteurs du LHC . . . . .	24
1.2.2	Les propriétés du boson de Higgs . . . . .	25
	Modes de production et de désintégration du boson de Higgs au LHC . . . . .	25
	Propriétés du boson de Higgs mesurées au LHC . . . . .	29
1.3	Les limites du Modèle Standard . . . . .	31
1.3.1	Problèmes de naturalité . . . . .	31
1.3.2	Phénomènes cosmologiques . . . . .	32
1.3.3	Autres limites . . . . .	33

---

Afin d'étudier d'éventuels nouvelles particules de type boson de Higgs, il semble important de définir rigoureusement ce qu'est la particule qu'on appelle "boson de Higgs" ainsi que son rôle au sein de la théorie actuelle de la physique des particules, le Modèle Standard (MS).

## 1.1 Le formalisme du Modèle Standard

Le Modèle Standard de la physique des particules est un formalisme théorique basé sur la théorie quantique des champs renormalisable, la théorie des perturbations et l'invariance sous transformation de jauge locale décrivant les interactions électromagnétique, forte et faible. On définit l'invariance sous transformation de jauge locale comme la propriété pour un champ de rester identique suite à une transformation<sup>1</sup> de ses attributs  $\psi(x) \rightarrow \psi'(x) = e^{i\theta_a(x)T^a} \psi$  où  $\psi$  est une fonction d'onde,  $x$  est un quadrivecteur temps-espace et  $\theta_a(x)T^a$  correspond à la décomposition sur les générateurs ( $T^a$ ) du groupe de symétrie associé à la transformation de jauge considérée. Dans le MS,  $\psi$  est un champ représentant des particules fermioniques (spin demi entier) et les générateurs  $T^a$  vont être liés aux bosons (spin entier) de spin 1 d'interaction.

Afin de comprendre en profondeur comment le MS fut construit nous allons brièvement détailler la description de chacune des trois interactions fondamentales ci-dessus.

### 1.1.1 Électrodynamique Quantique

L'électrodynamique quantique (QED) est la formulation de la théorie quantique des champs relativistes décrivant l'interaction électromagnétique, interaction entre les fermions et le photon. La QED est fondée sur la transformation de jauge locale du groupe de symétrie  $\mathcal{U}(1)_Q$  correspondant au groupe contenant les nombres complexes de module 1 et l'équation quantique relativiste de Dirac :

$$(i\gamma^\mu \partial_\mu - m)\psi(x) = 0 \quad (1.1)$$

où  $\gamma^\mu$  sont les matrices de Dirac,  $\partial_\mu$  est la dérivée covariante associée au groupe de Poincaré et  $m$  la masse du fermion considéré.

Considérons la transformation de jauge  $\psi(x) \rightarrow \psi'(x) \equiv U(x)\psi(x) = e^{ieQ\chi(x)}\psi(x)$  avec  $U^\dagger U = 1$ ,  $eQ=q$  la charge électrique du fermion et  $\chi$  champ scalaire générateur de  $\mathcal{U}(1)$ , l'équation de Dirac n'est à priori pas invariante sous cette transformation puisqu'on obtient après composition par  $U^\dagger$  :

$$i\gamma^\mu (\partial_\mu + ieQ\partial_\mu\chi)\psi - m\psi = 0 \quad (1.2)$$

Il est donc nécessaire de modifier l'équation de Dirac pour la rendre invariante sous cette transformation. On introduit, dans ce but, un nouveau champ scalaire ( $A_\mu$ ) qui se transforme comme  $A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu\chi$ . L'équation de Dirac modifiée est alors :

$$i\gamma^\mu (\partial_\mu + ieQA_\mu)\psi - m\psi = 0 \quad (1.3)$$

1. A priori, toutes les notions présentées ici dépendent de la représentation choisie pour le groupe de symétrie considéré. Dans un souci de simplification, nous omettrons de noter cette dépendance dans la suite.

manifestement invariante sous transformation de jauge locale. Ce terme supplémentaire  $-eQ\gamma^\mu A_\mu\psi$  peut être interprété comme un terme d'interaction entre le champ scalaire ( $A_\mu$ ) et le champ fermionique  $\psi$ . Ici, ( $A_\mu$ ) est en fait identifié comme le photon, vecteur de l'interaction électromagnétique de masse nulle.

Si on veut maintenant décrire la totalité de l'interaction électromagnétique, il ne faut pas oublier que le champ ( $A_\mu$ ) a aussi sa propre équation dynamique. On sait par ailleurs que la dynamique du photon en tant que particule libre est décrite par les équations de Maxwell dans le vide qui utilisent le tenseur de Faraday  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ .

En regroupant tous ces éléments sous une formulation lagrangienne des équations du mouvement, on obtient le lagrangien de QED avec  $\bar{\psi} = \psi^\dagger \gamma^0$  :

$$\mathcal{L}_{QED} = \underbrace{\bar{\psi}^{(f)}(i\gamma^\mu\partial_\mu - m_f)\psi^{(f)}}_{\text{Dirac : champ } \psi \text{ libre}} - \underbrace{eQ\bar{\psi}^{(f)}\gamma^\mu A_\mu\psi^{(f)}}_{\text{interaction entre } A_\mu \text{ et } \psi} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{Maxwell : champ } A_\mu \text{ libre}} \quad (1.4)$$

où on a ici sommé sur tous les fermions (f). Remarquons qu'un éventuel terme de masse pour le photon ( $\propto m_\gamma^2 A^\mu A_\mu$ ) ne respecterait pas l'invariance de jauge. Le boson de jauge doit donc être sans masse ici.

Ainsi nous avons vu dans cette section comment construire un lagrangien pour l'interaction électromagnétique en utilisant l'invariance sous transformation de jauge locale du groupe  $\mathcal{U}(1)$  pour faire apparaître l'interaction entre des fermions et un boson de jauge qui est, ici, le photon.

### 1.1.2 Chromodynamique Quantique

La chromodynamique quantique (QCD) est la formulation de la théorie quantique des champs relativistes décrivant l'interaction forte, interaction prédominante dans les noyaux d'atome, entre les quarks et les gluons. Pour connaître le groupe de symétrie utilisé, il faut s'intéresser à quelques observations expérimentales : observation des baryons suivant  $\Delta^{++}=(uuu)$ ,  $\Delta^-(ddd)$  et  $\Omega^-(sss)$ , soit à chaque fois 3 quarks dans le même état de spin, ce qui est a priori contraire au principe de Pauli à moins de postuler l'existence d'un nombre quantique supplémentaire. Ce nombre quantique fut appelé couleur et peut prendre trois valeurs (r,b,g). Par des considérations de symétrie des groupes et sous l'hypothèse que seules les particules de couleur neutre peuvent être observées<sup>2</sup>, on peut considérer une symétrie sous transformation de jauge locale du groupe  $SU(3)_C$  composé des matrices complexes unitaires de déterminant égal à 1.

Ce groupe est structurellement différent du groupe  $\mathcal{U}(1)$  utilisé en QED, notamment  $SU(3)$  possède  $n^2 - 1=8$  générateurs (contre 1 pour  $\mathcal{U}(1)$ ) qui ne commutent pas nécessairement entre eux. On parle de théorie de jauge non abélienne ou théorie de Yang-Mills.

2. Un gluon dans un état de singulet de couleur générerait une composante non confinée de l'interaction forte, ce qui serait en désaccord avec les observations expérimentales

Considérons donc la transformation de jauge :

$$\psi(x) \rightarrow \psi'(x) = e^{ig_s \alpha_a(x) T^a} \psi(x) \text{ avec } [T_a, T_b] = if_{abc} T_c \quad (1.5)$$

où  $T_a$  sont les générateurs<sup>3</sup> de  $SU(3)$ ,  $a=1, \dots, 8$ ,  $g_s$  est la constante de couplage de l'interaction forte,  $\alpha_a$  champs de décomposition sur les générateurs précédent et  $f_{abc}$  sont les constantes de structure du groupe  $SU(3)$ .

A nouveau, il faut modifier l'équation de Dirac pour conserver l'invariance sous transformation de jauge locale. En faisant explicitement apparaître un terme d'interaction entre les quarks et les huit nouveaux champs  $G_\mu^a$  correspondant aux gluons, on obtient alors une nouvelle équation de Dirac :

$$i\gamma^\mu (\partial_\mu + ig_s G_\mu^a T_a) \psi - m\psi = 0 \quad (1.6)$$

La transformation de ces nouveaux champs est néanmoins plus délicate que dans le cas QED à cause du caractère non-abélien du groupe de symétrie. Pour déterminer cette transformation, considérons une transformation infinitésimale pour le groupe  $SU(3)$  donnée par  $\delta\psi(x) = (\mathbb{1} + ig_s \alpha_a(x) T^a) \psi(x)$ . La condition d'invariance de l'équation 1.6 sous transformation de jauge infinitésimale donne :

$$i\gamma^\mu (\partial_\mu \delta\psi + ig_s \delta G_\mu^a T_a \psi + ig_s G_\mu^a T_a \delta\psi) - m\delta\psi = 0 \quad (1.7)$$

En injectant la formule de transformation infinitésimale dans l'équation précédente, on obtient :

$$i\gamma^\mu \partial_\mu \psi - g_s \gamma^\mu (\partial_\mu \alpha_a(x)) T^a \psi - g_s \gamma^\mu \alpha_a(x) T^a \partial_\mu \psi - g_s \gamma^\mu \delta G_\mu^a T_a \psi - g_s \gamma^\mu G_\mu^a T_a \psi - ig_s^2 \gamma^\mu G_\mu^a \alpha^b \underbrace{T_a T_b}_{T_b T_a + if_{abc} T^c} \psi - m\delta\psi = 0 \quad (1.8)$$

Cette équation se simplifie en reconnaissant deux fois l'équation modifiée 1.6 après avoir fait intervenir le commutateur des générateurs. On peut ensuite procéder par identification pour trouver la transformation que doivent satisfaire les champs  $G_\mu^a$  pour garantir l'invariance de jauge de l'équation 1.6 :

$$G_\mu^a \rightarrow (G_\mu^a)' = G_\mu^a - \partial_\mu \alpha^a - g_s f^{abc} \alpha^b G_\mu^c \quad (1.9)$$

Ceci est en fait une propriété plus générale des théories de jauge non-abéliennes que l'on peut aisément retrouver en faisant intervenir la dérivée covariante  $D_\mu \psi$  telle que  $D_\mu \psi$  se transforme comme  $\psi$  (valable pour tout multiplet de champs) :

$$D_\mu \psi \equiv (\partial_\mu + ig G_\mu) \psi \quad (1.10)$$

pour n'importe quel champ vectoriel relativiste  $G_\mu = G_\mu^a T_a$  associé à un groupe de symétrie de dimension finie (à valeurs dans l'algèbre de Lie associé). Pour une transformation de jauge

3. Notons que les notations  $T^a$  et  $T_a$  sont complètement équivalentes ici,  $a$  n'étant que de simples indices de sommation

locale quelconque  $\psi'(x) = U(x)\psi(x) = e^{igw^a(x)T_a}\psi(x)$ , on veut donc que la dérivée covariante satisfasse la relation :

$$(D_\mu\psi)' = UD_\mu\psi \quad (1.11)$$

$$\iff \partial_\mu\psi' + igG'_\mu\psi' = U(\partial_\mu\psi + igG_\mu\psi) \quad (1.12)$$

Comme  $U^{-1}U = \mathbb{1}$ , on peut écrire  $\partial_\mu\psi' = U\partial_\mu\psi - U(\partial_\mu U^{-1})U\psi$ , ce qui permet d'obtenir après composition par  $U^{-1}$  à droite et simplification de la relation de transformation générale pour un champ  $G_\mu$  quelconque :

$$G'_\mu = UG_\mu U^{-1} - \frac{i}{g}U(\partial_\mu U^{-1}) \quad (1.13)$$

En développant  $U$  au premier ordre en  $w$ ,  $U = \mathbb{1} + igw + o(w^2)$ , on peut retrouver la relation 1.9 avec les notations adéquates :

$$G'_\mu = G_\mu + \partial_\mu w + ig[w, G_\mu] + o(w^2) \quad (1.14)$$

Comme dans le cas de la QED, on veut rajouter un terme dans le lagrangien décrivant la dynamique du champ  $G_\mu$ . Pour cela on définit l'analogue du tenseur de Faraday :

$$F_{\mu\nu} \equiv -\frac{i}{g}[D_\mu, D_\nu] \quad (1.15)$$

En développant cette relation et en simplifiant grâce à la symétrie des termes en présence, on obtient une équation pour le tenseur de Faraday non-abélien (ou, en terme géométrique, courbure de la connexion du groupe de symétrie considéré) :

$$F_{\mu\nu} = \partial_\mu G_\nu - \partial_\nu G_\mu - g[G_\mu, G_\nu] \quad (1.16)$$

Dans le cas spécifique de QCD, on a simplement

$$G_{\mu\nu} = \partial_\mu G_\nu - \partial_\nu G_\mu + ig_s f_{abc} G_\mu^b G_\nu^c \quad (1.17)$$

Il est aisé de voir que  $F$  se transforme comme  $F'_{\mu\nu} = UF_{\mu\nu}U^{-1}$ , ainsi une contraction de type  $F_a^{\mu\nu}F_{\mu\nu}^a = Tr(F^{\mu\nu}F_{\mu\nu})$  est manifestement invariant de jauge par propriété cyclique de la trace matricielle.

Avec tous ces éléments, il devient naturel d'écrire le lagrangien QCD comme :

$$\mathcal{L}_{QCD} = \underbrace{\bar{\psi}^{(q)}(i\gamma^\mu\partial_\mu - m)\psi(x)^{(q)}}_{\text{Dirac : champ } \psi \text{ libre}} - \underbrace{g_s\bar{\psi}^{(q)}\gamma^\mu G_\mu^a T_a\psi(x)^{(q)}}_{\text{interactions}} - \underbrace{\frac{1}{4}G_{\mu\nu}G^{\mu\nu}}_{G_\mu \text{ libres}} \quad (1.18)$$

où on a ici sommé sur tous les quarks (q)=u,d,s,c,b,t. Remarquons à nouveau qu'un éventuel terme de masse pour les gluons ( $\propto m_g^2 G_\mu^a G_\mu^a$ ) ne respecterait pas l'invariance de jauge. Les gluons sont donc de masse nulle. Le terme en  $G_\mu^b G_\nu^c$  de l'Eq. 1.17 n'est pas présent dans le cas abélien de QED et va conduire aux interactions entre les gluons.

Ainsi nous avons vu dans cette section comment construire un lagrangien pour l'interaction forte en utilisant l'invariance sous transformation de jauge locale du groupe  $SU(3)$  non-abélien pour faire apparaître l'interaction entre des quarks et des bosons de jauge qui sont, ici, les gluons.

### 1.1.3 Interaction faible et électrofaible

L'interaction faible est la seule des quatre interactions à ne pas produire d'états liés (alors qu'il existe l'atome pour la force électromagnétique et le noyau pour l'interaction forte), ce qui motive son nom d'interaction "faible". Ceci s'explique par la masse de ses bosons médiateurs de l'interaction, les bosons  $W^\pm$  et  $Z^0$ , qui est relativement importante. Elle est aussi particulière puisqu'elle est la seule à violer la symétrie P (parité) et la symétrie CP (charge parité). Elle implique aussi tous les leptons (processus leptonique) et les quarks (semi-leptonique, hadronique) avec possibilité de changement de saveur.

Pour comprendre l'interaction faible, il est nécessaire de faire un rappel sur la notion de chiralité. La chiralité est une symétrie des champs de fermions  $\psi$  (avec  $\psi$  spineur de Dirac), plus spécifiquement on appelle états propres de chiralité les états propres de l'opérateur

$$\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & \mathbb{1}_2 \\ \mathbb{1}_2 & 0 \end{pmatrix} \text{ qui échange les composantes hautes et basses des spineurs.}$$

Ces états propres de chiralité peuvent être droit (R) ou gauches (L) selon leur valeur propre associée au cas particule ( $\psi^P$ ) ou antiparticule ( $\psi^A$ ) :

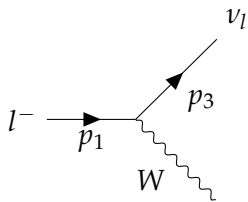
$$\gamma^5\psi_R^P = +\psi_R^P \quad \gamma^5\psi_L^P = -\psi_L^P \quad \gamma^5\psi_R^A = -\psi_R^A \quad \gamma^5\psi_L^A = +\psi_L^A \quad (1.19)$$

En définissant les projecteurs  $P_R = \frac{1}{2}(1 + \gamma^5)$  et  $P_L = \frac{1}{2}(1 - \gamma^5)$  projetant respectivement sur les états droits/gauches des particules/antiparticules et gauches/droites des particules/antiparticules tels que :

$$P_R\psi_R^P = \psi_R^P \quad P_R\psi_L^P = 0 \quad P_R\psi_R^A = 0 \quad P_R\psi_L^A = \psi_L^A \quad (1.20)$$

On peut alors décomposer tout spineur sur sa partie droite et gauche  $\psi = \psi_R + \psi_L = P_R\psi + P_L\psi$ .

Il est possible de montrer [1] que la violation maximale de la parité par l'interaction faible implique une structure Vectorielle-Axiale pour les interactions de courant chargé (avec échange de  $W^\pm$ ) impliquant une valeur du vertex du diagramme de Feynman associée proportionnelle à  $\gamma^\mu \frac{1}{2}(1 - \gamma^5)$ . Cela a des implications sur la chiralité des interactions physiquement réalisables. Considérons par exemple les interactions leptoniques par courant chargé :



Au vertex d'interaction on a

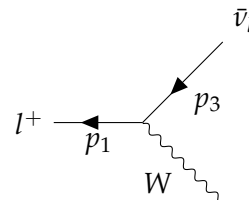
$$\begin{aligned} & \bar{\psi}^P(p_3)\gamma^\mu \frac{1}{2}(1 - \gamma^5)\psi^P(p_1) \\ &= \bar{\psi}^P(p_3)\gamma^\mu \psi_L^P(p_1) \\ &= [\bar{\psi}_R^P(p_3) + \bar{\psi}_L^P(p_3)]\gamma^\mu \psi_L^P(p_1) \end{aligned}$$

$$= \bar{\psi}_L^P(p_3)\gamma^\mu \psi_L^P(p_1)$$

car  $\bar{\psi}_R^P\gamma^\mu\psi_L^P = \bar{\psi}^P P_L\gamma^\mu\psi_L^P = \bar{\psi}^P\gamma^\mu P_R\psi_L^P = 0$  d'après les propriétés d'anticommutation des matrices de Dirac.

Pour le cas antiparticule on a de même

$$\begin{aligned} & \bar{\psi}^A(p_1)\gamma^\mu \frac{1}{2}(1 - \gamma^5)\psi^A(p_3) \\ &= \bar{\psi}^A(p_1)\gamma^\mu \psi_R^A(p_3) \end{aligned}$$



$$= [\bar{\psi}_R^A(p_1) + \bar{\psi}_L^A(p_1)]\gamma^\mu\psi_R^A(p_3)$$

$$= \bar{\psi}_L^A(p_1)\gamma^\mu\psi_R^A(p_3)$$
 car  $\bar{\psi}_R^A\gamma^\mu\psi_R^A = \bar{\psi}^A P_R \gamma^\mu \psi_R^A = \bar{\psi}^A \gamma^\mu P_L \psi_R^A = 0$  d'après les propriétés d'anticommutation des matrices de Dirac.

La conséquence de ces calculs qualitatifs est que seuls les spineurs avec une chiralité gauche, pour les particules, et les spineurs avec une chiralité droite, pour les antiparticules, participent aux interactions par courant chargé.

Pour résumer, on a vu que l'interaction faible possède 3 bosons médiateurs de l'interaction, dont les  $W$  qui ne se couplent qu'aux particules/antiparticules d'un spineur de chiralité gauche/droite. Ainsi on peut choisir le groupe de symétrie  $SU(2)_L$  comme jauge locale où  $L$  souligne le caractère chirale de l'interaction. Il devient alors commode de placer les particules dans des doublets (2 composantes) de spineurs pour celles participant à l'interaction faible par courants chargés. Les autres (fermions droits et anti-fermions gauches) sont placés dans des singulets de  $SU(2)$ , invariants sous toute transformation de  $SU(2)$ , puisque n'intervenant pas par interaction faible par courant chargé. On écrira par exemple :

$$\varphi_L^{e^-}(x) = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \equiv \begin{pmatrix} \psi_{\nu_e}(x) \\ \psi_{e^-}(x) \end{pmatrix}_L \quad \text{et} \quad \varphi_R^{e^-}(x) \equiv e_R^-(x) \quad (1.21)$$

De manière plus exhaustive, les fermions sont classés selon leur isospin faible (charge associée à la symétrie  $SU(2)$ ) [1, 2, 3]. Il faut néanmoins prendre en compte que les états propres de l'interaction faible, intervenants dans les vertex de l'interaction faible, (notés  $'$ ) ne sont pas les mêmes que les états propres de masse correspondant aux spineurs usuellement utilisés pour décrire les quarks. Ainsi on doit faire intervenir une matrice décrivant ce changement de base, la matrice CKM [4, 5], permettant le changement de saveur des quarks ; cette matrice satisfait la relation :

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (1.22)$$

Ainsi les fermions de  $SU(2)_L$  s'écrivent comme :

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \quad \begin{pmatrix} u \\ d' \end{pmatrix}_L \quad \begin{pmatrix} c \\ s' \end{pmatrix}_L \quad \begin{pmatrix} t \\ b' \end{pmatrix}_L \quad (1.23)$$

$$e_R^-, \quad \mu_R^-, \quad \tau_R^-, \quad u_R, \quad d_R, \quad c_R, \quad s_R, \quad t_R, \quad b_R \quad (1.24)$$

Et les anti-fermions (avec un isospin faible opposé) :

$$\begin{pmatrix} e^+ \\ \bar{\nu}_e \end{pmatrix}_R \quad \begin{pmatrix} \mu^+ \\ \bar{\nu}_\mu \end{pmatrix}_R \quad \begin{pmatrix} \tau^+ \\ \bar{\nu}_\tau \end{pmatrix}_R \quad \begin{pmatrix} \bar{d}' \\ \bar{u} \end{pmatrix}_R \quad \begin{pmatrix} \bar{s}' \\ \bar{c} \end{pmatrix}_R \quad \begin{pmatrix} \bar{b}' \\ \bar{t} \end{pmatrix}_R \quad (1.25)$$

$$e_L^+, \quad \mu_L^+, \quad \tau_L^+, \quad \bar{u}_L, \quad \bar{d}_L, \quad \bar{c}_L, \quad \bar{s}_L, \quad \bar{t}_L, \quad \bar{b}_L \quad (1.26)$$

On peut maintenant jager la théorie en considérant une transformation de jauge locale de  $SU(2)_L$ ,  $\varphi(x) \rightarrow \varphi'(x) = e^{ig_w \alpha^a(x) T_a} \varphi(x)$  avec  $a = 1, 2, 3$ ,  $g_w$  la constante de couplage faible et  $T_a$  des générateurs de  $SU(2)$  tels que  $[T_a, T_b] = i\epsilon_{abc} T_c$  où  $\epsilon_{abc}$  est le tenseur antisymétrique de

Levi-Civita.

A nouveau, imposer l'invariance de jauge locale fait apparaître un terme d'interaction entre les fermions et les bosons de l'interaction faible dans l'équation de Dirac :

$$ig_w T_a W_\mu^a \gamma^\mu \varphi_L \quad (1.27)$$

avec  $W_\mu^a$  qui se transforme comme le champ de gluons (groupe non-abélien) :

$$W_a^{\mu'} = W_a^\mu - \partial^\mu \alpha_a - g_w \epsilon_{bca} \alpha_b W_c^\mu \quad (1.28)$$

On peut alors identifier [3] les courants chargés ce qui permet de mettre en évidence les bosons physiques  $W_\pm^\mu = \frac{1}{\sqrt{2}} (W_1^\mu \mp iW_2^\mu)$ . On pourrait, ensuite, assimiler le boson  $Z^0$  au champ  $W_3$ , cela contraindrait cependant les courants associés, courants neutres, à ne se coupler qu'aux particules gauches et antiparticules droites. Or, expérimentalement, on observe aussi des couplages du  $Z^0$  aux particules droites. Le modèle de l'interaction faible paraît alors incomplet.

Pour le compléter il peut être intéressant de coupler le boson  $Z^0$  à un autre boson neutre. Le photon apparaît comme un bon candidat : c'est le modèle d'unification électrofaible de Glashow, Salam et Weinberg [1, 2, 3]. Dans ce modèle on unifie l'interaction faible et l'interaction électromagnétique en une seule interaction électrofaible dont le groupe de symétrie associé est  $SU(2)_L \times U(1)_Y$ . On voit que le groupe de symétrie  $U(1)_Q$  de l'interaction électromagnétique a été remplacé par le groupe  $U(1)_Y$  avec  $Y$  appelée hypercharge faible. La transformation associée est  $\psi(x) \rightarrow \psi'(x) = e^{ig' \frac{Y}{2} \zeta(x)} \psi(x)$  qui donne lieu à un terme d'interaction dans l'équation de Dirac qui est :

$$g' \frac{Y}{2} \gamma^\mu B_\mu \psi. \quad (1.29)$$

Dans ce modèle le photon et le boson  $Z^0$  sont construits comme combinaison linéaire des champs  $B_\mu$  et  $W_\mu^3$  fonction de l'angle de mélange faible  $\theta_w$  :

$$A_\mu = B_\mu \cos \theta_w + W_\mu^3 \sin \theta_w \quad (1.30)$$

$$Z_\mu = -B_\mu \sin \theta_w + W_\mu^3 \cos \theta_w \quad (1.31)$$

En calculant et identifiant les courants associés aux bosons physiques et ceux associés à l'hypercharge et l'isospin faibles, on obtient [1]  $e = g_w \sin \theta_w = g' \cos \theta_w$  et le lien entre l'hypercharge faible, l'isospin faible  $I_w^3$  et la charge électromagnétique  $Q$ ,  $Y = 2(Q - I_w^3)$ . On peut alors écrire le lagrangien décrivant l'interaction électrofaible de manière analogue à QED et QCD :

$$\mathcal{L}_{EW} = \bar{\varphi}^f (i\gamma^\mu D_\mu - m) \varphi^f - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^k W_k^{\mu\nu} \quad (1.32)$$

où  $k=1,2,3$ , la dérivée covariante est

$$D_\mu = \partial_\mu + ig' \frac{1}{2} Y B_\mu \mathbb{1}_2 + ig_w T_a W_\mu^a \quad (1.33)$$

et les tenseurs de champs sont

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad W_{\mu\nu}^k = \partial_\mu W_\nu^k - \partial_\nu W_\mu^k + g_w \epsilon_{ijk} W_\mu^i W_\nu^j \quad (1.34)$$



Ce modèle permet donc de conserver les propriétés de la QED et celle de l'interaction électrofaible en garantissant que le boson  $Z^0$  interagisse avec les particules droites (grâce à sa composante en  $B_\mu$ ).

### 1.1.4 Mécanisme de Higgs

L'invariance de jauge nous a permis de construire des lagrangiens décrivant les interactions forte et électrofaible (faible et électromagnétique). Cette invariance de jauge est nécessaire pour assurer le caractère renormalisable de la théorie [6] permettant de compenser les divergences apparaissant dans les diagrammes de Feynman. Cependant les lagrangiens, tels qu'ils sont présentés dans les Eq. 1.18 et 1.32, comportent des termes non invariants, dans le cadre de la formulation  $SU(2)_L$  de la théorie, sous transformation de jauge locale : les termes de masse.

En ce qui concerne les bosons de jauge, on voit d'ailleurs qu'aucun terme de masse n'est présent alors que les bosons  $W^\pm$  et  $Z^0$  possèdent une masse non nulle. Pour le boson  $B$ , il est facile de voir qu'une transformation de jauge  $B_\mu \rightarrow B_\mu - \partial_\mu \chi$  donnerait un terme de masse se transformant comme :

$$\frac{1}{2}m_B^2 B_\mu B^\mu \rightarrow \frac{1}{2}m_B^2 (B_\mu - \partial_\mu \chi)(B^\mu - \partial^\mu \chi) \neq \frac{1}{2}m_B^2 B_\mu B^\mu \quad (1.35)$$

donc un terme non invariant sous transformation de jauge locale. Le même raisonnement peut être fait pour les bosons de  $SU(2)$  (et  $SU(3)$ ). Ainsi il n'est pas possible d'avoir des termes explicites de masse pour les bosons de jauge dans la formulation lagrangienne actuelle.

Pour les fermions, le problème est le même. En effet les termes de masse pour les fermions de Dirac s'écrivent, avec des doublets de  $SU(2)$ , tels (le cas antiparticule est similaire) :

$$m_f \bar{\psi} \psi = m_f (\bar{\psi}_R + \bar{\psi}_L)(\psi_R + \psi_L) = m_f (\bar{\psi}_R P_L + \bar{\psi}_L P_R)(P_R \psi_R + P_L \psi_L) = m_f (\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R) \quad (1.36)$$

en utilisant les relations 1.20. Si on considère une transformation de jauge  $SU(2)_L$  telle que  $\varphi_R \rightarrow \varphi_R$  et  $\varphi_L \rightarrow e^{ig_w \alpha^a(x) T_a} \varphi_L$ , on observe que les termes de masse précédents ne respectent pas l'invariance de jauge.

Il fallait donc trouver un moyen d'obtenir des termes de masse pour les particules sans briser explicitement les symétries de jauge. Un tel mécanisme fut introduit de manière indépendante par Brout, Englert, Higgs, Hagen, Guralnik et Kibble en 1964 [7, 8, 9] et porte aujourd'hui leur nom (bien qu'il soit souvent abrégé en "mécanisme de Higgs"). Ce mécanisme se base sur l'ajout d'un champ scalaire  $\phi$  et sur la **brisure de symétrie spontanée** : les équations du lagrangien respectent bien les symétries de jauge mais la valeur moyenne, asymétrique, dans le vide du champ  $\phi$  permet l'apparition de termes de masse.

Pour expliciter ce mécanisme, on peut écrire le lagrangien électrofaible l'incorporant :

$$L_{EW+H} = \underbrace{-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}}_{\text{bosons de jauge}} + \underbrace{i\bar{\varphi}^f \not{D}\varphi^f}_{\text{termes de Dirac}} + \underbrace{(D_\mu\phi)^2 - V(\phi) + y_f \left( \bar{\varphi}_L^f \phi \varphi_R^f + \left( \bar{\varphi}_L^f \phi \varphi_R^f \right)^\dagger \right)}_{\text{termes de Higgs}} \quad (1.37)$$

où on a posé  $F^{\mu\nu} = B_{\mu\nu}B^{\mu\nu} + W_{\mu\nu}^k W_k^{\mu\nu}$  et  $\not{D} = \gamma^\mu D_\mu$ .  $y$  est une matrice  $3 \times 3$ , et  $\phi$  est un doublet de champs scalaires, appelé "champs de Higgs", dont le potentiel associé est  $V$ .

Dans le Modèle Standard (MS), ce potentiel est fixé [10], avec deux constantes  $\lambda > 0$  et  $\sigma$  par :

$$V(\phi) = \frac{1}{2}\sigma^2\phi^2 + \frac{1}{4}\lambda\phi^4 \quad (1.38)$$

Ce potentiel est le point clef permettant la brisure de symétrie spontanée et donc la possibilité d'écrire des masses pour les particules dans le lagrangien. Pour illustrer cette brisure de symétrie, considérons le cas simplifié où le champ de Higgs  $\phi$  est un singulet en place d'un doublet et peut donc s'écrire  $\phi(x) = \frac{1}{\sqrt{2}}(\phi_1(x) + i\phi_2(x))$  où  $\phi_1$  et  $\phi_2$  sont des champs scalaires réels. On a alors deux possibilités pour le potentiel  $V$  selon le signe de  $\sigma^2$  comme on peut le voir sur la figure 1.1.

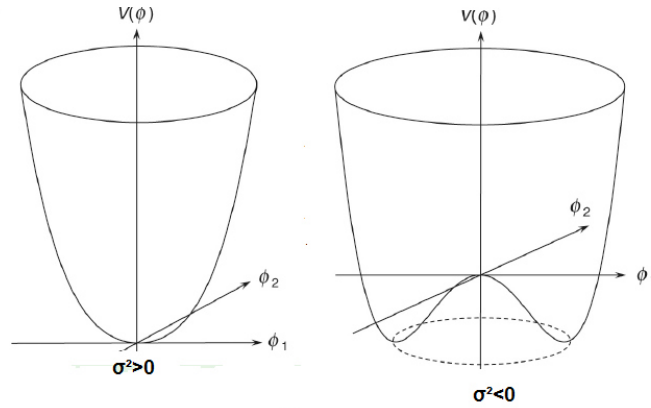


FIGURE 1.1 – Potentiel de Higgs en fonction de  $\phi_1$  et  $\phi_2$ . Si  $\sigma^2 > 0$  (gauche), le potentiel est symétrique en son minimum global mais pas dans le cas  $\sigma^2 < 0$  (droite) [10].

Si  $\sigma^2 > 0$ , le minimum du potentiel est zéro qui est une valeur symétrique. Mais si  $\sigma^2 < 0$  le minimum est  $|\phi|^2 = \frac{-\sigma^2}{\lambda} \equiv v^2$  et le potentiel n'est plus symétrique vers ce minimum  $|\phi| = v$ . C'est la brisure spontanée de symétrie : le choix du minimum, le vide, a brisé la symétrie. Considérons maintenant que  $\sigma^2 < 0$ ,  $\phi$  peut être développé<sup>4</sup> autour du minimum  $v$  en écrivant  $\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\zeta(x))$  (avec deux champs scalaires réels  $\eta$  et  $\zeta$ ), pour obtenir

$$(D_\mu\phi)^2 - V(\phi) = \underbrace{\frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta)}_{\text{champ scalaire massif}} - \lambda v^2\eta^2 + \underbrace{\frac{1}{2}(\partial_\mu\zeta)(\partial^\mu\zeta)}_{\text{boson de Goldstone}} + \underbrace{\frac{1}{2}v^2 F_\mu F^\mu}_{\text{masse bosons de jauge}} + \underbrace{V_{int}(\eta, \zeta, F)}_{\text{termes d'interaction}} \quad (1.39)$$

Un tel développement permet effectivement [10] d'obtenir des termes de masse pour les bosons de jauge. On observe aussi la présence de bosons de Goldstone de masse nulle, ils sont

4. Cela correspond à des fluctuations quantiques et thermiques.

caractéristiques d'une brisure de symétrie spontanée.

On peut faire disparaître ces termes du lagrangien par une transformation de jauge, dite de jauge unitaire,  $\phi(x) \rightarrow \phi'(x) = e^{i\zeta(x)/v} \phi(x)$ . On peut montrer que cette transformation de jauge se ramène (on développe les champs à l'ordre 1) à choisir le champ scalaire  $\phi$  comme réel; proche du minimum, on a alors  $\phi(x) = \frac{1}{\sqrt{2}}(v + h(x))$  avec  $h$  un champ scalaire réel.

Revenons maintenant au cas du MS, le champ de Higgs  $\phi$ , le doublet de Higgs, s'écrit

$$\phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad (1.40)$$

avec  $\phi^+$  et  $\phi^0$  des champs scalaires complexes. On considère  $\sigma^2 < 0$  et la valeur attendue dans le vide du champ est, avec  $v = 246$  GeV environ :

$$\langle 0|\phi(x)|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (1.41)$$

En suivant ce qui a été effectué précédemment c'est-à-dire en développant autour du minimum  $v$  avec des champs réels  $\phi_i$  tel que

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ v + \phi_3(x) + i\phi_4(x) \end{pmatrix} \quad (1.42)$$

et en injectant ce développement dans le lagrangien Eq. 1.37, on obtient les termes de masse pour toutes les particules massives du MS (bosons de jauge et fermions) [8]. Il y a cette fois trois bosons de Goldstone de masse nulle  $\phi_1$ ,  $\phi_2$  et  $\phi_4$  correspondant au degré de liberté des bosons  $W^\pm$  et  $Z^0$  que l'on peut, à nouveau, éliminer du lagrangien en se plaçant dans la jauge unitaire [10] ce qui nous permet d'écrire :

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (1.43)$$

avec  $h$  un champ scalaire réel massif correspondant à une particule nommée "boson de Higgs".

Ce développement induit par la brisure de symétrie spontanée permet d'obtenir des termes de masse pour les bosons de jauge :

$$L_{EW+H} \supset \frac{g^2 v^2}{4} W_\mu^+ W^{-\mu} + \frac{(g'^2 + g^2)v^2}{8} Z_\mu Z^\mu \quad (1.44)$$

permettant d'identifier  $m_W = \frac{gv}{2}$  et  $m_Z = \frac{m_W}{\cos\theta_w}$  tandis que le photon a une masse nulle.

Des termes d'interaction des bosons de jauge avec le boson de Higgs apparaissent de la même manière :

$$L_{EW+H} \supset \frac{2m_W^2}{v} W_\mu^+ W^{-\mu} h + \frac{m_W^2}{v^2} W_\mu^+ W^{-\mu} h^2 + \frac{m_Z^2}{v} Z_\mu Z^\mu h + \frac{m_Z^2}{2v^2} Z_\mu Z^\mu h^2 \quad (1.45)$$

ils décrivent des interactions à trois et quatre points ; le couplage de ces interactions est directement proportionnel à la masse du boson de jauge au carré.

Si les masses des bosons de jauge apparaissent naturellement, il est nécessaire d'ajouter un terme supplémentaire pour générer la masse des fermions. Ce terme, terme de Yukawa, est le terme de droite dans l'Eq. 1.37 et il permet d'obtenir des termes de masse et d'interaction avec le boson de Higgs pour les fermions :

$$L_{EW+H} \supset \frac{y_f v}{\sqrt{2}} \bar{\psi}^f \psi^f + \frac{y_f}{\sqrt{2}} \bar{\psi}^f \psi^f h \quad (1.46)$$

Les masses des fermions du MS s'écrivent alors en fonction des coefficients de la matrice  $y$ , les couplages de Yukawa, comme :

$$m_f = \frac{y_f v}{\sqrt{2}} \quad (1.47)$$

Le couplage de fermions au boson de Higgs est, ici, directement proportionnel à leur masse.

La masse  $m_H = \sqrt{-2\mu^2}$  et les auto-interactions du boson de Higgs apparaissent de la même manière à partir du potentiel  $V$  :

$$L_{EW+H} \supset -\frac{m_H^2}{2} h^2 - \frac{m_H^2}{2v} h^3 - \frac{m_H^2}{8v^2} h^4 \quad (1.48)$$

Finalement, le lagrangien du MS s'écrit en ajoutant les termes du lagrangien QCD (en prenant soin de retirer les termes de masse des fermions) :

$$L_{MS} \supset L_{EW+H} + L_{QCD} \quad (1.49)$$

## 1.2 Le boson de Higgs au LHC

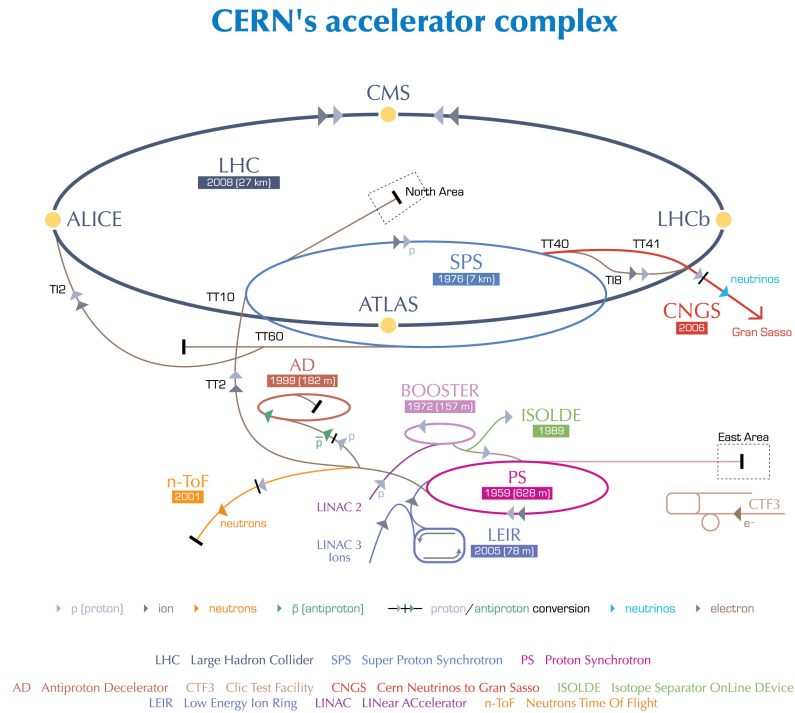
Le Modèle Standard de la physique des particules est à la fois un succès théorique et expérimental couronné par la découverte du boson de Higgs en 2012 par deux expériences du grand collisionneur à hadrons au CERN, ATLAS [11] et CMS [12]; toutes les particules prédites par le MS ont donc été découvertes.

On présente ici brièvement le grand collisionneur à hadrons ayant mené à la découverte du boson de Higgs et les caractéristiques de ce dernier.

### 1.2.1 Le grand collisionneur à hadrons

#### Présentation générale

Le grand collisionneur à hadrons (LHC) est actuellement le plus grand, avec une circonférence de 27 km, et plus puissant accélérateur de particules - protons ou ions  $Pb^+$  - avec une énergie dans le centre de masse devant atteindre 14 TeV [13] pour les collisions proton-proton, sur lesquelles on se concentre dans la suite. Situé à la frontière franco-suisse au CERN (Organisation européenne pour la recherche nucléaire), son objectif est de délivrer des collisions avec assez d'énergie et en nombre suffisant afin de pouvoir effectuer des mesures précises et



European Organization for Nuclear Research | Organisation européenne pour la recherche nucléaire

© CERN 2008

FIGURE 1.2 – Schéma de la chaîne d'accélération du CERN. Avant d'être injectés dans le LHC les protons sont accélérés par le LINAC 2, le Booster, le PS et le SPS [15].

détecter d'éventuelles nouvelles particules.

Situé entre 47 et 170 m sous la surface, dans le tunnel abritant précédemment le grand collisionneur électron-positron (LEP) [14], le LHC n'est que la dernière étape d'une chaîne d'accélérateurs complexe récapitulée sur la figure 1.2. Les protons proviennent d'une simple bouteille de dihydrogène  $H_2$  reliée à un cylindre métallique soumis à un champ électrique permettant de séparer électrons et protons. A la sortie de ce cylindre, les protons sont accélérés à une énergie de 750 keV, collimatés et séparés par paquets. Ils sont envoyés dans un accélérateur linéaire (LINAC 2) constitué de cavités radiofréquences permettant d'atteindre une énergie de 50 MeV. Ils sont ensuite accélérés par le Booster du Synchrotron à protons (PSB), un accélérateur circulaire constitué de 277 électroaimants, jusqu'à une énergie de 25 GeV dans des paquets de 81 protons séparés d'un intervalle de 25 ns. L'étape suivante est le Super Synchrotron à protons (SPS), constitué de 1317 électroaimants, permettant d'obtenir une énergie de 450 GeV. Finalement, les protons atteignent le LHC où ils sont accélérés jusqu'à une énergie de 6,5 TeV actuellement.

Les composantes clés du LHC sont ses 1232 aimants dipolaires, ses 392 aimants quadripolaires et ses 16 cavités radiofréquences répartis dans des sections droites et des sections en arc de cercle. Les aimants dipolaires sont refroidis à l'hélium superfluide à une température de 1,9 K et produisent ainsi un champ magnétique d'environ 8,3 T qui permet de courber la trajectoire des faisceaux de protons. Les aimants quadripolaires sont utilisés pour focaliser les

faisceaux et ainsi augmenter la probabilité d'interaction proton-proton aux points d'interaction. Les cavités radiofréquences sont responsables de l'accélération des protons et délivrent un champ de 5 MV/m à une fréquence de 400 MHz tout en permettant de garder la forme des paquets de protons dans les faisceaux. Les collisions ont ensuite lieu dans un des quatre points d'interaction correspondant aux quatre grands détecteurs du LHC toutes les 25 ns.

## Luminosité et empilement

### Luminosité

Le nombre d'évènements d'un processus physique pouvant être observé pendant la prise de données est directement proportionnel à la luminosité délivrée par le collisionneur au point d'interaction de l'expérience considérée. Ainsi plus la luminosité est élevée, plus on peut obtenir d'évènements physiques rares en plus de permettre une réduction des différentes incertitudes statistiques. Soient  $N$  le nombre d'évènement du processus considéré,  $\sigma$  sa section efficace et  $T$  le temps de prise de données, on a la relation :

$$N = \sigma \times L_{int} = \sigma \int_0^T L(t) dt \quad (1.50)$$

où  $L_{int}$  est la luminosité intégrée et  $L$  est la luminosité instantanée qui dépend du temps et des propriétés des faisceaux de protons.

La luminosité instantanée nominale du LHC est de  $10^{-34} \text{ cm}^{-2} \cdot \text{s}^{-1}$  et cette valeur a notamment été dépassée pendant la majorité de la prise des données de 2018 où elle est d'environ  $2 \times 10^{-34} \text{ cm}^{-2} \cdot \text{s}^{-1}$ . Expérimentalement, on utilise en général plutôt la luminosité intégrée. La figure 1.3 présente ainsi l'évolution de la luminosité délivrée à l'un des quatre détecteurs du LHC en fonction du temps et de l'année de prise de données avec des énergies dans le centre de masse de 7 TeV (en 2010-11) et de 8 TeV (en 2012) pour le Run 1, et de 13 TeV (2015-2018) pour le Run 2. L'augmentation de la luminosité intégrée en 2018 est de 23 % par rapport à 2017 et d'environ 39 % par rapport à 2016. L'énergie dans le centre de masse devrait monter à  $\sim 14$  TeV pour le Run 3 (prévu de 2022 à 2024) pour lequel on attend une luminosité intégrée totale d'environ  $300 \text{ fb}^{-1}$ .

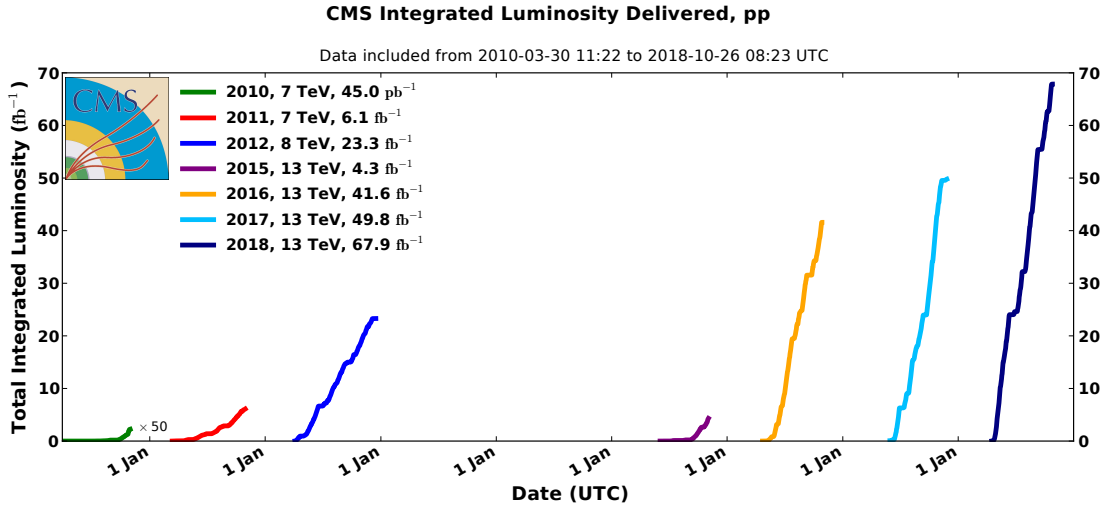


FIGURE 1.3 – Luminosité délivrée dans le détecteur CMS en fonction du temps pour 2010-2012 (Run 1) et 2015-2018 (Run 2) pour des collisions proton-proton [16].

### Empilement

Un grand nombre de collisions proton-proton a lieu lors du croisement des faisceaux de protons à l'un des quatre points d'interaction, il devient alors difficile de distinguer les événements issus d'une collision ou d'une autre ce qui peut gêner l'observation de processus d'intérêt. Ce phénomène, dont l'importance augmente avec la luminosité, est nommé empilement (PU).

On distingue en général deux types d'empilement : l'empilement "en temps" et l'empilement "hors temps". L'empilement en temps correspond au nombre d'interactions ayant lieu par croisement de faisceaux de protons et est donné par la formule :

$$PU_{in-time} = \sigma_{p-p} L(t) \tau_{coll} \quad (1.51)$$

où  $\sigma_{p-p}$  est la section efficace d'interaction inélastique proton-proton et  $\tau_{coll}$  le temps entre deux croisements de faisceaux qui est ici de 25 ns.

L'empilement hors-temps correspond au nombre d'interactions ayant lieu lors des croisements de faisceaux antérieurs ou postérieurs au croisement considéré et associées à ce dernier à cause du temps de mise en forme du système électronique qui est en général supérieur au temps entre deux croisements de faisceaux.

Un exemple de distribution du PU en temps est visible pour le détecteur CMS sur la figure 1.4. Les algorithmes de reconstruction utilisés, et présentés dans la suite, prennent, en règle générale, en compte l'effet du PU et corrigent les données en fonction.



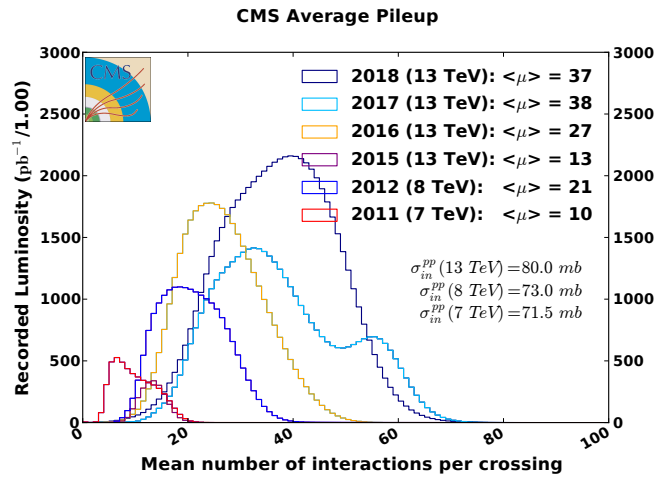


FIGURE 1.4 – Distribution du nombre moyen d’interaction par croisement de faisceaux pour des collisions proton-proton pour différentes années de prise de données dans le détecteur CMS du LHC [17].

### Les détecteurs du LHC

Aux quatre points d’intersection des faisceaux ont été construits quatre détecteurs étudiant, comme illustré sur la figure 1.5, les produits des collisions y ayant lieu. Ces quatre détecteurs principaux sont ALICE [18] (A Large Ion Collider Experiment), ATLAS [19] (A Toroidal LHC ApparatuS), CMS (Compact Muon Solenoid) détaillé ci-après et LHCb [20] (Large Hadron Collider beauty experiment). Ces détecteurs sont associés à des collaborations internationales du même nom.

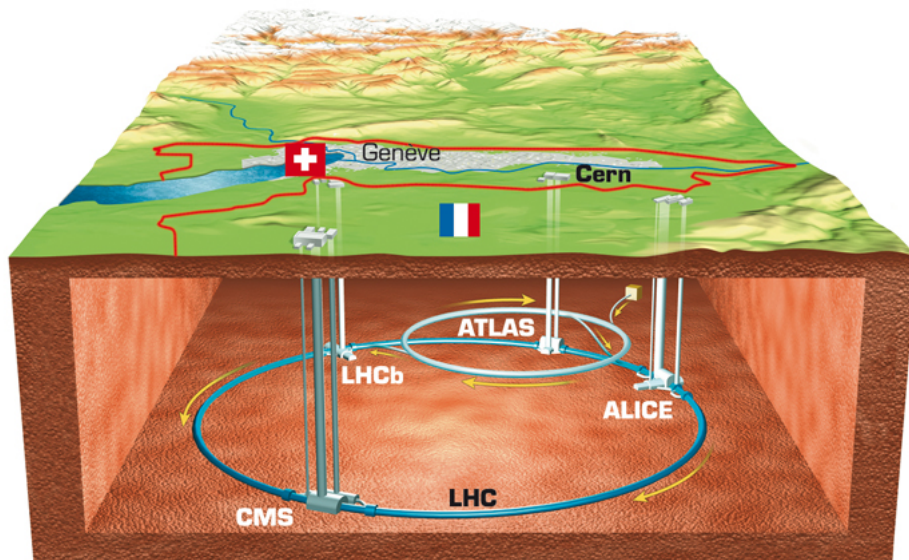


FIGURE 1.5 – Schéma du LHC au CERN avec ses quatre détecteurs principaux [21].



L'objectif de ces quatre détecteurs est différent. Les détecteurs ATLAS et CMS sont des expériences généralistes couvrant les mesures de précision du MS, les mesures des propriétés du boson de Higgs suite à sa découverte, la recherche de nouvelle particule ou encore l'étude du plasma quark-gluon. L'avantage de disposer de deux détecteurs étudiant des sujets similaires est qu'il est possible de recouper et comparer des résultats obtenus de manière indépendante avec des technologies de détections différentes. Une découverte de nouvelle particule est d'autant plus probante qu'elle est à la fois observée par ATLAS et CMS comme ce fut le cas pour le boson de Higgs.

Les détecteurs ALICE et LHCb sont plus spécialisés. L'expérience LHCb s'intéresse à l'étude des mésons et baryons à travers la recherche de désintégrations rares associées à un quark bottom ou charm, la mesure de violation CP, ainsi qu'à la recherche de nouvelles particules exotiques.

ALICE se concentre sur les collisions d'ions lourds afin d'étudier les interactions QCD et le plasma quark-gluon.

Il existe également plusieurs autres détecteurs de taille plus modeste comme TOTEM [22] (TOTal Elastic and diffractive cross section Measurement), LHCf [23] (Large Hadron Collider forward) étudiant les propriétés de particules émises avec des angles très faibles par rapport aux faisceaux, ou encore MoEDAL [24] (Monopole and Exotics Detector At LHC) qui cherche à mettre en évidence l'existence d'un monopole magnétique.

### 1.2.2 Les propriétés du boson de Higgs

Découvrir le boson de Higgs était un des objectifs du LHC qui a donc été atteint en 2012. Depuis, plusieurs modes de production et de désintégration du boson de Higgs ont été, et sont toujours, étudiés afin de mesurer les propriétés de ce boson le plus précisément possible. De telles mesures expérimentales permettent de tester directement le MS : si des différences sont observées par rapport aux quantités prédites par le MS, c'est le signe de présence d'une nouvelle physique au-delà du MS (BSM), le cas échéant des contraintes sur les formes possibles de cette nouvelle physique sont dérivées.

On présente ici les principaux modes de production et de désintégration du boson de Higgs au LHC ainsi qu'un bref état de l'art concernant les mesures effectuées.

### Modes de production et de désintégration du boson de Higgs au LHC

#### Principaux modes de production du boson de Higgs au LHC

Le boson de Higgs peut être produit de différentes façons au LHC selon l'énergie dans le centre de masse des collisions proton-proton et selon la masse considérée pour le boson de Higgs. On dénombre quatre modes de production dits principaux de part leur relative importance en terme de section efficace. Ces modes sont décrits ci-après et sont visibles sur la figure 1.6.

Pour une énergie de collision  $\sqrt{s}=13$  TeV et une masse  $m_H=125$  GeV, le mode de production

principal est la fusion de gluons (ggH). La prédominance de ce mode, près de 87 % [25] de la section efficace de production totale, s'explique par la prépondérance des gluons dans la fonction de densité partonique du proton et la valeur élevée du couplage entre le boson de Higgs et les quarks massifs comme le top ou, dans une moindre mesure, le bottom.

Le second mode de production en terme d'importance, de l'ordre de 7 %, est la production par fusion de bosons vecteurs (VBF). Dans ce cas, deux quarks sont aussi produits, en plus du boson de Higgs, ce qui se traduit expérimentalement par la présence de deux jets caractéristiques dans l'état final qui peuvent être utilisés comme identification d'un tel processus.

Les deux autres modes de production principaux sont la production d'un boson de Higgs associée à un boson vecteur (VH) ou associée avec un quark et antiquark top (ttH), ces deux modes de production ont été mis en évidence au début du Run 2 du LHC [26, 27] alors que les deux premiers modes évoqués ci-dessus ont été observés au Run 1. Il est usuel de séparer le mode de production VH selon la nature du boson vecteur associé soit WH ou ZH. Une désintégration leptonique de ces bosons vecteurs associés est aussi une signature expérimentale caractéristique de ce type de mode de production.

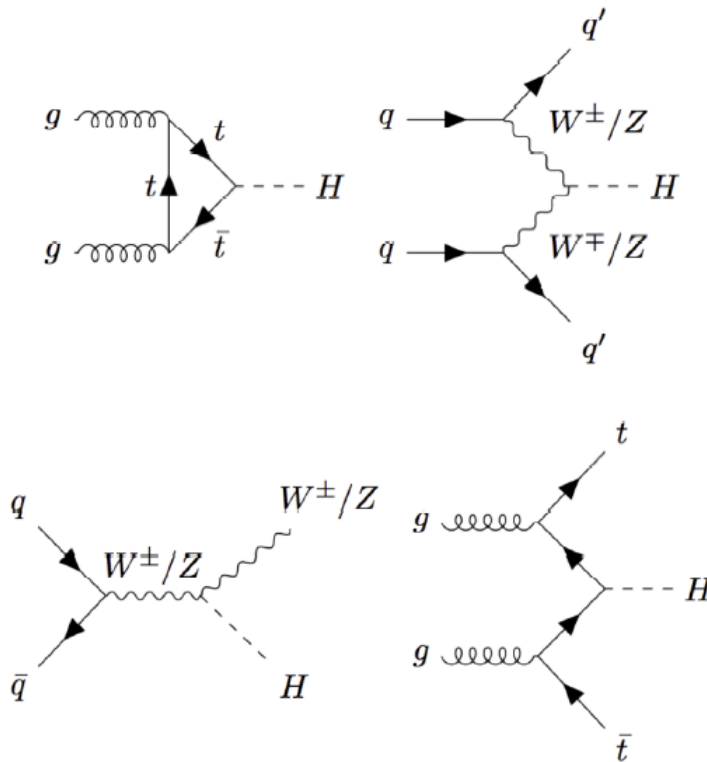


FIGURE 1.6 – Diagrammes de Feynman correspondant aux quatre modes de production principaux du boson de Higgs au LHC par ordre de valeur de section efficace décroissante : ggH (côté supérieur gauche), VBF (côté supérieur droit), VH (côté inférieur gauche) et ttH (côté inférieur droit).

La valeur de la section efficace pour différents modes de production du boson de Higgs est visible sur la figure 1.7. Si la masse du boson de Higgs découvert en 2012 est proche de 125 GeV, il est intéressant d'étudier le comportement de la section efficace pour des masses

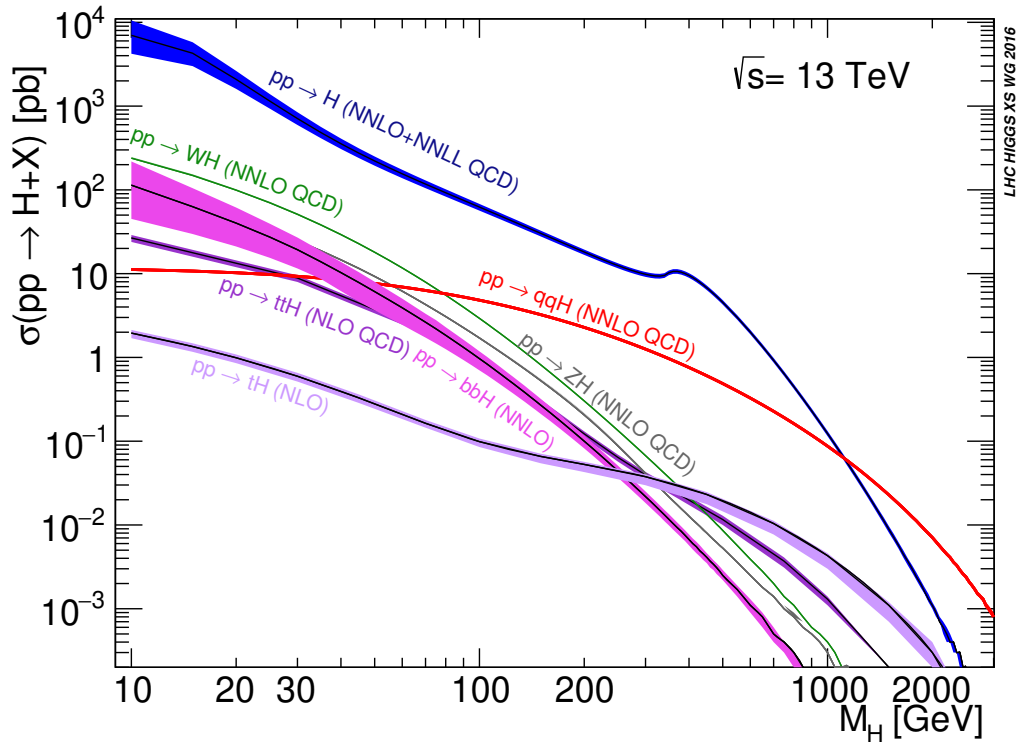


FIGURE 1.7 – Section efficace de la production d’un boson de Higgs selon différents modes de production au LHC pour  $\sqrt{s}=13$  TeV [25].

inférieures à cette valeur puisque l’on va s’intéresser, dans le chapitre 6, à étudier la présence hypothétique d’un boson de Higgs additionnel de masse inférieure à 125 GeV.

### Principaux modes de désintégration du boson de Higgs au LHC

Le temps de vie du boson de Higgs étant prédit [25] comme étant très court, le boson de Higgs n’est jamais observé directement mais sa présence est déduite par observation de ses produits de désintégration. En effet, la largeur de désintégration du boson de Higgs est prédite comme étant de l’ordre de 4 MeV (et encore plus petite pour des masses inférieures à 125 GeV) et est donc négligeable devant la résolution du détecteur (de l’ordre du GeV) lors de la reconstruction des produits de désintégration.

Comme étudié dans la section précédente, le boson de Higgs se couple avec toutes les particules massives (et d’autant plus que leur masse est importante) et il y a donc de nombreux modes de désintégration possibles. D’autant plus que les diagrammes à boucles permettent des désintégrations du boson de Higgs en des particules sans masse comme les photons et les gluons.

La figure 1.8 montre la valeur du rapport de branchement (BR), la largeur de désintégration relative, pour les principaux modes de désintégration du boson de Higgs. A nouveau, on considère un boson de Higgs du MS avec une masse proche de 125 GeV mais il est aussi intéressant pour la suite d’étudier les variations des rapports de branchement en fonction de la masse de ce dernier. On observe donc que la désintégration  $H \rightarrow b\bar{b}$  possède le BR le plus

important. Ce n'est cependant pas le canal de désintégration privilégié pour l'observation du boson de Higgs puisqu'il est très difficile à utiliser expérimentalement à cause de la présence importante de bruit de fond hadronique au LHC.

Les canaux de plus grandes sensibilités sont les canaux de désintégration  $H \rightarrow \gamma\gamma$  et  $H \rightarrow ZZ^* \rightarrow 4l$ . En effet malgré leur BR bien plus faible, ces canaux permettent une bonne signature des états finaux avec un bruit de fond relativement faible. Ces canaux ont notamment été d'une importance cruciale pour la découverte du boson de Higgs et sont aujourd'hui toujours des canaux de choix pour étudier ses propriétés.

Les principales contributions à la désintégration  $H \rightarrow \gamma\gamma$  sont visibles sur la figure 1.9, elles présentent des couplages entre le boson de Higgs et le quark top et le boson  $W^\pm$ .

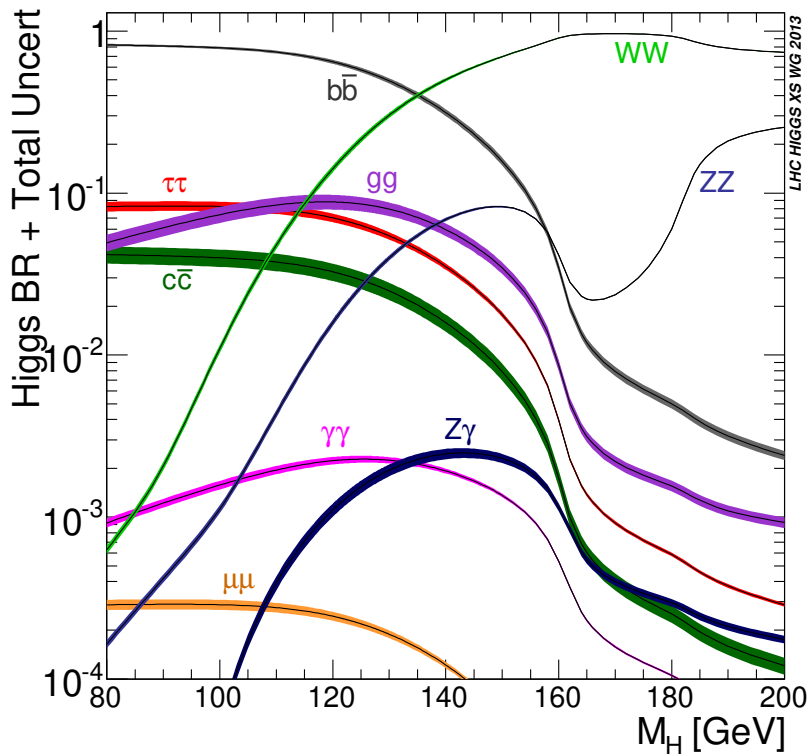


FIGURE 1.8 – Rapports de branchement pour différents canaux de désintégration du boson de Higgs en fonction de sa masse au LHC pour  $\sqrt{s}=13$  TeV [25].

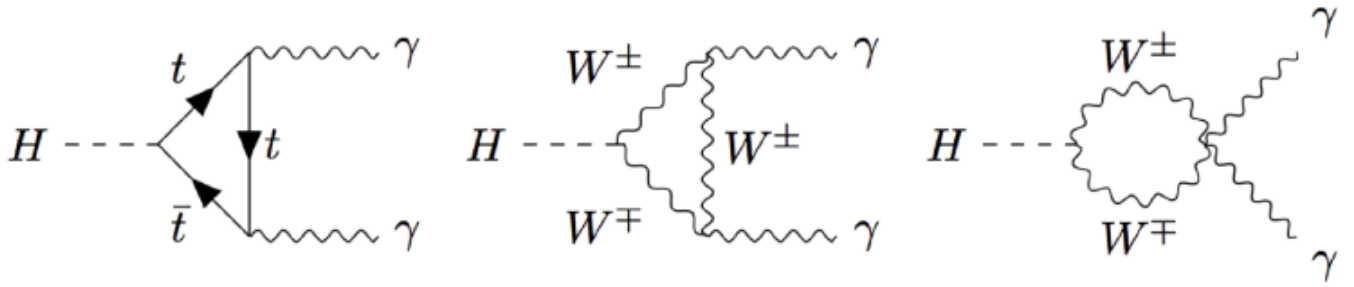


FIGURE 1.9 – Exemple de diagrammes de Feynman contribuant à la désintégration  $H \rightarrow \gamma\gamma$ .

### Propriétés du boson de Higgs mesurées au LHC

Depuis la découverte du boson de Higgs en 2012, de nombreuses mesures ont été réalisées pour étudier ses propriétés avec une précision remarquable et tester ainsi la structure du MS.

#### Masse

On a précédemment supposé l’hypothèse d’une masse du boson de Higgs autour de 125 GeV, sa valeur a cependant été précisément mesurée. Les combinaisons des résultats ATLAS et CMS [28] pour le Run 1 du LHC ont permis d’obtenir  $m_H = 125,09 \pm 0,24$  GeV. A l’heure actuelle, la mesure la plus précise de cette masse est celle effectuée dans CMS [29] pour le Run 2 du LHC qui donne :

$$m_H = 125,38 \pm 0,14 \text{ GeV} \quad (1.52)$$

#### Couplages

Les couplages du boson de Higgs avec les autres particules du MS sont mesurés expérimentalement en utilisant notamment les modificateurs de la force du signal  $\mu_i$  pour la production d’un boson de Higgs, où  $i$  est le mode de production, et  $\mu^f$  pour la désintégration d’un boson de Higgs, où  $f$  est le mode désintégration considéré. Ils sont définis en fonction des valeurs de section efficace et rapport de branchement obtenus divisés par leur valeur attendue dans le MS tel :

$$\mu_i = \frac{\sigma_i}{(\sigma_i)_{MS}} \quad \mu^f = \frac{BR^f}{(BR^f)_{MS}} \quad (1.53)$$

Une combinaison préliminaire [30] des résultats obtenus pendant le Run 2 du LHC pour les analyses CMS  $H \rightarrow \gamma\gamma$ ,  $H \rightarrow ZZ$ ,  $H \rightarrow WW$ ,  $H \rightarrow bb$ ,  $H \rightarrow \tau\tau$  et  $H \rightarrow \mu\mu$  sous l’hypothèse d’une masse fixée  $m_H=125,09$  GeV<sup>5</sup> permet d’obtenir un modificateur de la force du signal global  $\mu$  tel que

$$\mu = 1,02^{+0,07}_{-0,06} = 1,02 \pm 0,04 \text{ (th)} \pm 0,04 \text{ (exp)} \pm 0,04 \text{ (stat)} \quad (1.54)$$

où les incertitudes systématiques ont été séparées selon source théorique et expérimentale. Les valeurs des modificateurs de la force du signal observés expérimentalement sous ces hypothèses

5. La valeur plus précise de la masse mentionnée précédemment n’était pas encore publique au moment de la combinaison.

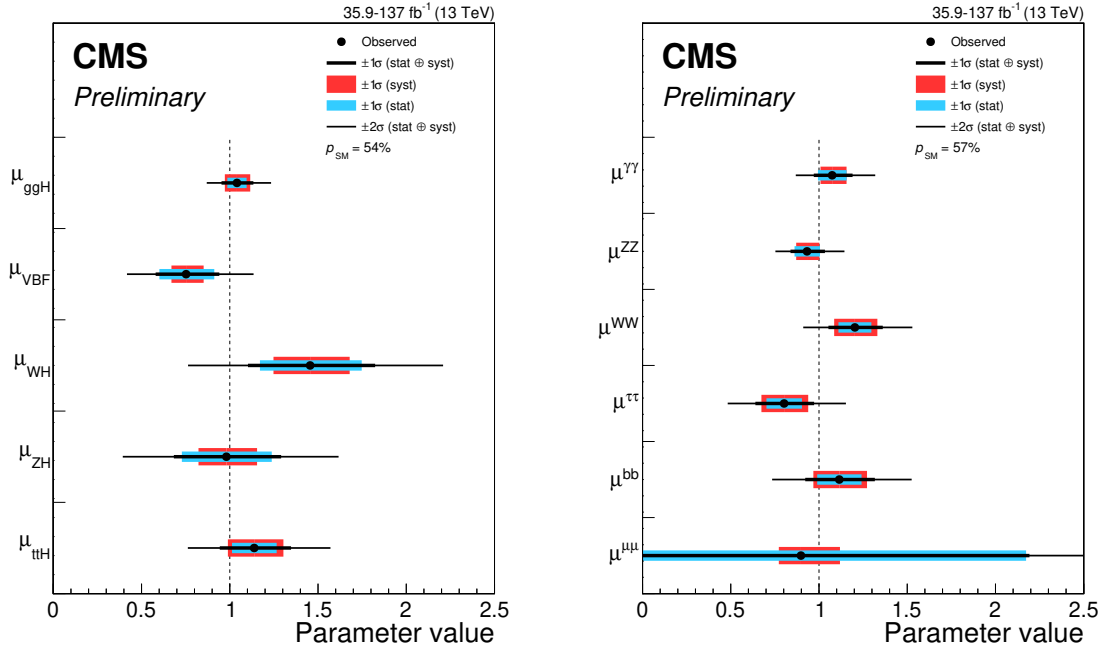


FIGURE 1.10 – Modificateurs de la force du signal pour les modes de production (gauche) et de désintégration (droite) dans CMS pour une masse  $m_H=125,09$  GeV pour une combinaison des analyses  $H \rightarrow \gamma\gamma$ ,  $H \rightarrow ZZ$ ,  $H \rightarrow WW$ ,  $H \rightarrow bb$ ,  $H \rightarrow \tau\tau$  et  $H \rightarrow \mu\mu$  et pour le Run 2 du LHC avec une luminosité intégrée allant de  $36 \text{ fb}^{-1}$  à  $137 \text{ fb}^{-1}$  selon l'analyse considérée. La ligne noire épaisse (fine) représente l'intervalle de confiance à  $1\sigma$  ( $2\sigma$ ) et les lignes bleu et rouge représentent respectivement les composantes des incertitudes statistiques et systématiques de l'intervalle de confiance  $1\sigma$  [30].

sont visibles sur la figure 1.10. Une meilleure compréhension des quantités présentées ici peut être obtenue suite à la lecture du chapitre 6.

Si les valeurs de ces modificateurs diffèrent de 1, on peut être en présence de signes de nouvelle physique puisqu'on aurait alors des déviations observées par rapport au MS. Cependant on observe ici que tous les modificateurs de force de signal présentés ici sont compatibles avec la valeur 1 et sont donc compatibles avec les valeurs attendues dans le cadre du MS.

### Spin et parité

Étant donné que le boson de Higgs est actuellement la seule particule scalaire dans le MS, l'observation expérimentale de son spin est une mesure importante. Du fait de l'observation de la désintégration du boson de Higgs en paire de photons de spin 1, la particule ne peut pas avoir un spin 1 [31]. De plus, l'hypothèse d'un spin 2 ainsi que l'hypothèse d'un spin 0 avec une parité négative (donc un pseudo-scalaire) ont été exclues, pour des modèles particuliers, par ATLAS [32] et CMS [33]. L'hypothèse d'un scalaire de spin 0, compatible avec les prédictions du MS, est donc privilégiée bien que le cas d'un mélange d'une composante scalaire et pseudo-scalaire n'a pas été complètement exclu et fait toujours l'objet de recherches.

La particule découverte en 2012 a donc pour l'instant toutes les caractéristiques du boson de

Higgs du MS même si de nombreuses études sont encore en cours pour vérifier cette hypothèse et mesurer les propriétés de cette particule avec une précision permettant de la confirmer.

### 1.3 Les limites du Modèle Standard

Bien que le MS soit robuste expérimentalement, comme étudié précédemment, et théoriquement, des désaccords expérimentaux avec ses prédictions ont commencé à émerger. Le Fermi National Accelerator Laboratory "Muon g-2 Experiment" mesurant le moment magnétique anormal du muon a en effet récemment observé une différence dans la valeur de g-2 par rapport à la valeur prédite par le MS de  $4,2\sigma$  [34] en terme de déviation standard (sachant qu'une découverte de nouvelle physique peut être proclamée à partir d'une différence de  $5\sigma$ ). Une déviation, plus faible, de  $3,1\sigma$ , a aussi été observée dans l'expérience LHCb sur la violation de l'universalité des saveurs leptoniques [35], si cet écart avec le MS n'est, pour l'instant, pas complètement significatif, ceci est un résultat encourageant, parmi d'autres de moindre signifiante, qui pourrait amener à la découverte de nouvelle physique.

En effet, il y a des phénomènes physiques expérimentaux et théoriques que le MS peine à décrire. Il pourrait alors n'être qu'une théorie effective d'un modèle de plus haute énergie et il est donc nécessaire d'avoir recours à de la nouvelle physique.

#### 1.3.1 Problèmes de naturalité

Les principales limites directes du MS et plus spécifiquement du secteur du Higgs sont les problèmes de naturalité. C'est-à-dire que certaines valeurs des paramètres du MS sont "surprenantes" et apparaissent comme non naturelles. Si cela n'est pas nécessairement incorrect, cette non naturalité n'est pas élégante et suggère la présence d'une autre théorie.

Parmi ces problèmes de naturalité, on trouve celui de la hiérarchie de la masse du boson de Higgs. En effet, la masse physique, mesurée, des particules élémentaires s'obtient en prenant en compte des corrections quantiques à boucles. Des exemples de termes, sous forme de diagrammes de Feynman, à considérer pour le boson de Higgs sont représentés sur la figure 1.11. A l'ordre dominant la masse physique s'obtient selon la formule :

$$m_{phys}^2 = m^2 + b\Lambda^2 + \dots \quad (1.55)$$

où  $m$  est un paramètre libre,  $b$  est de l'ordre de l'unité et  $\Lambda$  est l'échelle de validité de la théorie considérée. Pour les particules du MS autres que le boson de Higgs, la masse n'apparaît que lors de la brisure de symétrie spontanée électrofaible (à une énergie de l'ordre de  $\Lambda \sim 100$  GeV), et les divergences quadratiques ne sont présentes grâce aux symétries sous-jacentes. Aucune symétrie ne protège cependant la masse du boson de Higgs et, si l'on considère que le MS est valable jusque l'échelle de Planck, la valeur de  $\Lambda$  est de l'ordre de  $\sim 10^{19}$  GeV. Pour obtenir la masse mesurée du boson de Higgs  $m_{phy} \sim 125$  GeV, il faut donc que le paramètre libre  $m$  soit précis à 17 chiffres significatifs près (pour compenser la très haute valeur de  $\Lambda$ ) ce qui suggère

un ajustement extrêmement fin non naturel.

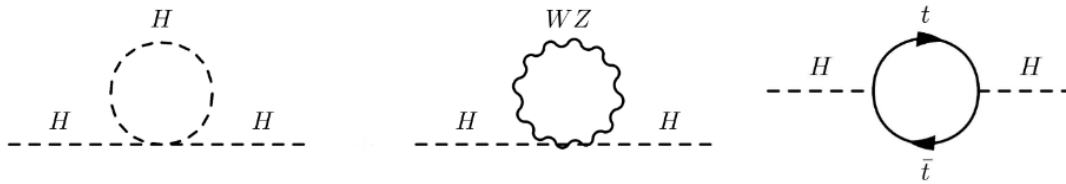


FIGURE 1.11 – Exemples de diagrammes de Feynman illustrant les contributions à boucles à la masse physique du boson de Higgs.

Le MS ne fournit, de plus, aucune contrainte sur les 22 paramètres qu’il contient (masse des particules, constantes de couplages et paramètres de mélange) et n’explique pas les différents ordres de grandeur qui existent, par exemple, entre les couplages de Yukawa.

On observe aussi qu’il y a exactement trois familles de fermions dans le MS mais aucune explication ne justifie cela.

Finalement, on s’attend à avoir une violation de la symétrie CP dans le secteur de l’interaction forte comme pour l’interaction faible. Une telle violation n’a cependant jamais été observée expérimentalement ce qui suggère une valeur de la phase quantifiant la violation de CP dans l’interaction forte très proche de zéro sans aucune forme de justification.

### 1.3.2 Phénomènes cosmologiques

Une autre limite, d’importance capitale pour avoir une théorie complète, du MS est qu’il ne prend pas en compte l’interaction gravitationnelle. La fusion de la théorie quantique des champs et de la relativité générale est en effet mathématiquement difficile et une simple quantification de la gravitation en terme d’un hypothétique graviton se heurte à des problèmes de non-renormalisation.

De nombreuses observations cosmologiques et astrophysiques ont mis en évidence l’existence de matière noire froide, composant 23 % de l’univers observable d’après le modèle standard de la cosmologie, expliquant notamment les comportements des courbes de vitesse de rotation des galaxies spirales ou encore les fluctuations de fond diffus cosmologique. L’univers serait aussi composé à 72 % d’énergie noire ayant pour effet l’accélération de l’expansion de l’univers. Le MS ne fournit aucune explication sur ces deux phénomènes.

L’asymétrie entre matière et antimatière dans l’univers  $\eta_B = \frac{n_B - n_{\bar{B}}}{s}$ , où  $n_B$  et  $n_{\bar{B}}$  sont la densité de baryon et anti-baryon et  $s$  est la densité d’entropie, a été mesurée [36] comme valant actuellement  $\eta_B \sim 6,5 \times 10^{-10}$  : autrement dit cette valeur élevée suggère que la quantité d’antimatière dans l’univers actuellement observable est bien plus faible devant la quantité de matière. Les modèles du Big-Bang prédisent une production symétrique entre particules



de matière et antimatière, il existerait alors des phénomènes physiques qui interagiraient différemment entre matière et antimatière [37]. Si des violations de symétrie CP existent bien dans le MS, elles sont bien trop faibles pour expliquer une telle asymétrie entre la densité de baryon et d'antibaryon donc le MS ne permet pas de rendre compte de l'asymétrie entre matière et antimatière.

#### 1.3.3 Autres limites

Des oscillations de la saveur des neutrinos ont été observées expérimentalement [38] et impliquent nécessairement que les neutrinos aient une masse très faible mais non nulle. Ce phénomène n'est pas prédit dans le MS. S'il est possible de construire une extension relativement simple du MS expliquant ces oscillations, la nature des neutrinos reste une question ouverte<sup>6</sup>. Les faibles couplages de Yukawa (masse du neutrino électronique inférieure à l'eV) qui y seraient associés posent à nouveau un problème de hiérarchie.

Les procédés de renormalisation permettent de réécrire les couplages des trois interactions fondamentales contenues dans le MS en fonction de l'échelle d'énergie considérée. Ce faisant, on observe (voir notamment la figure 2.1) que les couplages ne se recoupent jamais tous les trois quelle que soit l'échelle d'énergie considérée. Or on s'attend à ce qu'une théorie fondamentale soit capable d'unir, au moins, les interactions électromagnétique, faible et forte.

## Conclusion

Nous avons vu dans ce chapitre que le Modèle Standard permet de décrire les interactions électromagnétique, faible et forte d'un point de vue quantique. Ce modèle est basé sur la théorie quantique des champs et la théorie des perturbations combinées aux principes d'invariance sous transformation de jauge locale. Le Modèle Standard est un succès théorique et expérimental couronné par la découverte du boson Higgs au LHC plus de cinquante ans après sa conceptualisation.

Le boson de Higgs découvert en 2012 possède des propriétés compatibles avec celles prédites par le Modèle Standard. De nombreuses analyses continuent cependant de mesurer ses propriétés de manière extrêmement précise à la recherche de déviations par rapport au Modèle Standard pouvant conduire à la découverte de nouvelle physique.

En effet, des modèles de nouvelle physique sont nécessaires puisqu'il existe des phénomènes physiques que le Modèle Standard n'explique pas comme la force gravitationnelle, la matière noire ou encore la masse des neutrinos. Le Modèle Standard est sujet à des problèmes de naturalité et d'ajustements fins. Une solution pour résoudre ces problèmes de naturalité est de considérer le Modèle Standard comme une théorie effective valable jusqu'à une certaine

---

6. Les neutrinos pourraient par exemple être des particules de Majorana.

échelle d'énergie et englobée dans une symétrie plus globale.

Il existe de nombreux modèles de nouvelle physique impliquant souvent de nouvelles particules permettant d'étendre le Modèle Standard et ainsi de résoudre les limites et problèmes de ce dernier mais aucun d'entre eux n'a pour l'instant été mis en évidence expérimentalement.

# Les bosons de Higgs au-delà du Modèle Standard

## Sommaire

---

2.1	Modèles à deux doublets de Higgs . . . . .	36
2.1.1	Motivations . . . . .	36
2.1.2	Théorie générale . . . . .	36
2.1.3	Bosons de Higgs . . . . .	38
2.2	Modèles de Higgs Composites . . . . .	40
2.2.1	Introduction aux théories Technicouleur et Higgs Composites . . . . .	40
	La Technicouleur . . . . .	40
	Les Higgs Composites . . . . .	41
	Dynamique Composite Fondamentale . . . . .	42
2.2.2	Dynamique Composite Fondamentale $SU(4) \rightarrow Sp(4)$ . . . . .	42
	Symétries et vides . . . . .	43
	Lagrangien effectif FCD . . . . .	43
	Génération de la masse des pNGB . . . . .	45
2.3	Modèles Supersymétriques . . . . .	46

---

On présente dans ce chapitre quelques modèles de nouvelle physique permettant de résoudre une partie des limites du Modèle Standard abordées dans le chapitre précédent. Étant donné que l'objectif de ce manuscrit est l'étude d'une analyse recherchant expérimentalement un boson de Higgs additionnel de basse masse, on ne s'intéresse ici qu'à des modèles autorisant la présence d'un boson de Higgs additionnel de masse inférieure à 125 GeV.

On commence par présenter des modèles généraux possédant deux doublets de Higgs avant d'explicitier ce que sont les modèles de Higgs composites. Finalement, quelques brefs éléments de modèles supersymétriques seront abordés. Tous ces modèles ont la similarité de présenter des interprétations (voir par exemple [39]) au-delà du Modèle Standard pour un éventuel boson de Higgs additionnel de basse masse.

## 2.1 Modèles à deux doublets de Higgs

### 2.1.1 Motivations

Les modèles à deux doublets de Higgs (2HDM) sont des modèles assez généraux<sup>1</sup> consistant à étendre le secteur scalaire du MS par un doublet de Higgs supplémentaire.

Ce doublet supplémentaire permet d'obtenir 8 champs réels ; 3 d'entre eux sont des bosons de Goldstone nécessaires pour générer la masse des bosons  $W^\pm$  et  $Z^0$ , il reste donc 5 champs physiques de type bosons de Higgs : deux scalaires (CP-pair) neutres  $h$  et  $H$ , un pseudo-scalaire (CP-impair) neutre  $a$  et deux boson de Higgs chargés  $H^\pm$ .

Le 2HDM reste cependant une théorie effective qui ne résout notamment pas les problèmes de hiérarchie présentés précédemment. Ce modèle reste intéressant puisque de nombreuses théories BSM, comme celles abordées dans la suite, possèdent un secteur de Higgs contenant deux doublets de Higgs ; le 2HDM permet ainsi une étude simple du secteur de Higgs de ces modèles.

Les nouveaux scalaires présents dans les 2HDM peuvent être interprétés en terme de candidats à la matière noire ou encore en tant qu'axion [40] permettant alors d'élucider le problème CP-fort. Contrairement au MS, ils permettent en plus de créer une asymétrie baryonique [41] grâce à de nouveaux termes de violation CP [42].

### 2.1.2 Théorie générale

Les 2HDM contiennent donc deux champs complexe  $\phi_1$  et  $\phi_2$  de doublets  $SU(2)$  possédant les mêmes propriétés que le doublet de Higgs du MS soit une hypercharge +1 et se comportant

---

1. On peut considérer un cas plus général avec un nombre  $N$  de doublets de Higgs, mais le cas à 2 doublets reste la plus simple addition.

comme un singulet sous  $SU(3)_C$ .

Les valeurs attendues dans le vide de ces doublets sont :

$$\langle \phi_1 \rangle = \begin{pmatrix} 0 \\ \frac{v_1}{\sqrt{2}} \end{pmatrix} \quad \text{et} \quad \langle \phi_2 \rangle = \begin{pmatrix} 0 \\ \frac{v_2}{\sqrt{2}} \end{pmatrix} \quad (2.1)$$

où  $v_1$  et  $v_2$  sont deux réels tel que  $\sqrt{v_1^2 + v_2^2} = v$  la valeur attendue dans le vide du MS. On peut alors écrire ces doublets comme :

$$\phi_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1^+ \\ v_1 + \phi_1^0 + i\eta_1 \end{pmatrix} \quad \text{et} \quad \phi_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_2^+ \\ v_2 + \phi_2^0 + i\eta_2 \end{pmatrix} \quad (2.2)$$

avec  $\phi_i^0$  et  $\eta^i$  des champs réels et  $\phi_i^+$  des champs complexes. Des combinaisons de ces différents champs permettent d'obtenir les champs correspondant aux particules physiques  $h$ ,  $H$ ,  $a$  et  $H^\pm$  ainsi que les 3 bosons de Goldstone associés.

De la même manière que dans le MS, un potentiel pour le secteur de Higgs permettant la brisure spontanée de symétrie est nécessaire. Le potentiel le plus général s'écrit en prenant toutes les contributions, en matière des deux doublets, pour des termes quadratiques et quartiques. Il est cependant d'usage d'introduire une symétrie supplémentaire  $\mathbb{Z}_2$  pour éviter les changements de saveurs par courant neutre [43] non observés expérimentalement. Cette symétrie transforme les doublets  $\phi_i$  et les champs fermioniques  $\varphi^f$  tels que  $(\phi_1, \phi_2) \mapsto (\phi_1, -\phi_2)$  et  $(\varphi_L^f, \varphi_R^f) \mapsto (\varphi_L^f, \pm \varphi_R^f)$  où les deux signes peuvent être considérés. Afin de respecter cette symétrie additionnelle, les fermions du MS ne peuvent se coupler qu'avec un des deux doublets  $\phi_i$  selon le signe de la transformation  $\varphi_R^f$ . Ces différents couplages donnent lieu à 4 types de modèles différents lisibles dans le tableau 2.1. On observe notamment que les quarks de type up ne se couplent, par convention, qu'à  $\phi_2$ . Le modèle de type II peut notamment être utilisé en supersymétrie comme nous le verrons dans la suite.

Modèles	Type I	Type II	Lepton-specific	Flipped
Quarks de type up	$\phi_2$	$\phi_2$	$\phi_2$	$\phi_2$
Quarks de type down	$\phi_2$	$\phi_1$	$\phi_2$	$\phi_1$
Leptons	$\phi_2$	$\phi_1$	$\phi_1$	$\phi_2$

TABLE 2.1 – Les quatre façons possibles de coupler les fermions du MS aux deux doublets de Higgs des 2HDM et leur nom associé.

Le potentiel usuel est alors donné avec 8 paramètres libres réels  $m_{11}, m_{22}, m_{12}, \lambda_1, \lambda_2, \lambda_3, \lambda_4$  et  $\lambda_5$  :

$$V = m_{11}^2 \phi_1^\dagger \phi_1 + m_{22}^2 \phi_2^\dagger \phi_2 - m_{12}^2 (\phi_1^\dagger \phi_2 + \phi_2^\dagger \phi_1) + \frac{\lambda_1}{2} (\phi_1^\dagger \phi_1)^2 + \frac{\lambda_2}{2} (\phi_2^\dagger \phi_2)^2 + \lambda_3 (\phi_1^\dagger \phi_1) (\phi_2^\dagger \phi_2) + \lambda_4 (\phi_1^\dagger \phi_2) (\phi_2^\dagger \phi_1) + \frac{\lambda_5}{2} \left[ (\phi_1^\dagger \phi_2)^2 + (\phi_2^\dagger \phi_1)^2 \right] \quad (2.3)$$

On a donc 8 paramètres libres auxquels on peut ajouter  $v_1$  et  $v_2$ . La relation entre  $v_1$  et  $v_2$  ainsi que les conditions de minimisation du potentiel pour  $\phi_1$  et  $\phi_2$  permet cependant de réduire à 7 le nombre de degrés de liberté du secteur de Higgs.

### 2.1.3 Bosons de Higgs

Il est commode pour faire apparaître les bosons de Higgs physiques d'effectuer une rotation d'un angle  $\beta$  pour définir deux doublets de Higgs  $H_1$  et  $H_2$  de façon à ce que seul  $H_1$  acquière une valeur attendue dans le vide non nulle :

$$\begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \begin{pmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \quad \langle H_1 \rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \quad \langle H_2 \rangle = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (2.4)$$

Cette définition permet de lier les différentes valeurs attendues dans le vide en fonction de l'angle  $\beta$  :

$$v_1 = v \cos \beta, \quad v_2 = v \sin \beta, \quad \tan \beta = \frac{v_2}{v_1} \quad (2.5)$$

Cette nouvelle base permet alors de faire apparaître les bosons de Higgs chargés  $H^\pm$  et pseudo-scalaire  $a$  en réécrivant les doublets comme :

$$H_1 = \begin{pmatrix} G^\pm \\ \frac{1}{\sqrt{2}}(v + h_1^0 + iG^0) \end{pmatrix} \quad \text{et} \quad H_2 = \begin{pmatrix} H^\pm \\ \frac{1}{\sqrt{2}}(h_2^0 + ia) \end{pmatrix} \quad (2.6)$$

où  $G^\pm$  et  $G^0$  sont les bosons de Goldstone absorbés par les bosons de jauge  $W^\pm$  et  $Z^0$  pour obtenir une masse, et  $h_1^0$  et  $h_2^0$  sont deux champs scalaires neutres qui ne sont en général pas états propres de masse.

Pour obtenir des champs physiques pour les scalaires neutres, il est nécessaire d'effectuer une autre rotation par un angle  $\alpha$  de façon à définir les bosons de Higgs neutres  $h$  et  $H$  (avec par convention  $m_h < m_H$ ) tel que :

$$\begin{pmatrix} H \\ h \end{pmatrix} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \phi_1^0 \\ \phi_2^0 \end{pmatrix} = \begin{pmatrix} \cos(\beta - \alpha) & -\sin(\beta - \alpha) \\ \sin(\beta - \alpha) & \cos(\beta - \alpha) \end{pmatrix} \begin{pmatrix} h_1^0 \\ h_2^0 \end{pmatrix} \quad (2.7)$$

Cette nouvelle base permet donc de travailler avec les cinq champs physique  $h$ ,  $H$ ,  $a$  et  $H^\pm$ .

En développant le potentiel du secteur de Higgs autour des valeurs attendues dans le vide et avec ces nouvelles bases, on peut réécrire le potentiel en fonction des masses des particules physiques qui sont alors considérées comme des paramètres libres de la théorie. A la place des 7 paramètres libres du potentiel  $m_{22}, m_{12}, \lambda_1, \lambda_2, \lambda_3, \lambda_4$  et  $\lambda_5$ , on obtient 7 nouveaux paramètres libres caractéristiques de la masse de particules physiques  $m_{12}, m_h, m_H, m_a, m_{H^\pm}, \tan \beta$  et  $\sin(\beta - \alpha)$ .

Pour obtenir les couplages de ces nouveaux bosons de Higgs avec les bosons de jauge et fermions du MS, il faut considérer un terme cinétique légèrement différent de celui du MS pour les doublets de Higgs :

$$L_{H,cine} = (D_\mu \phi_1)^2 + (D_\mu \phi_2)^2 \quad (2.8)$$

De la même manière, le terme de Yukawa doit être réécrit pour prendre en compte notamment la symétrie  $Z_2$  supprimant des couplages entre un des doublets et les fermions selon le modèle 2HDM considéré :

$$L_{Yukawa} = \sum_{i=1}^2 y_{f^{(i)}} \left( \bar{\varphi}_L^{f^{(i)}} \phi \varphi_R^{f^{(i)}} + \left( \bar{\varphi}_L^{f^{(i)}} \phi \varphi_R^{f^{(i)}} \right)^\dagger \right) \quad (2.9)$$

En développement ces différents termes après brisure de symétrie spontanée comme dans le cas du MS, on obtient les différents couplages entre les bosons de Higgs et les fermions et bosons de jauge pour les différents 2HDM. Ces couplages pour les bosons de Higgs neutres, auxquels on s'intéresse plus particulièrement ici, sont lisibles sur le tableau 2.2.

Modèles		Type I	Type II	Lepton-specific	Flipped
Quarks de type up	h	$\frac{\cos \alpha}{\sin \beta}$			
	H	$\frac{\sin \alpha}{\sin \beta}$			
	a	$\cot \beta$			
Quarks de type down	h	$\frac{\cos \alpha}{\sin \beta}$	$-\frac{\sin \alpha}{\cos \beta}$	$-\frac{\sin \alpha}{\cos \beta}$	$\frac{\cos \alpha}{\sin \beta}$
	H	$\frac{\sin \alpha}{\sin \beta}$	$\frac{\cos \alpha}{\cos \beta}$	$\frac{\cos \alpha}{\cos \beta}$	$\frac{\sin \alpha}{\sin \beta}$
	a	$\cot \beta$	$\tan \beta$	$\tan \beta$	$\cot \beta$
Leptons	h	$\frac{\cos \alpha}{\sin \beta}$	$-\frac{\sin \alpha}{\cos \beta}$	$\frac{\cos \alpha}{\sin \beta}$	$-\frac{\sin \alpha}{\sin \beta}$
	H	$\frac{\sin \alpha}{\sin \beta}$	$\frac{\cos \alpha}{\cos \beta}$	$\frac{\sin \alpha}{\sin \beta}$	$\frac{\cos \alpha}{\cos \beta}$
	a	$\cot \beta$	$\tan \beta$	$\cot \beta$	$\tan \beta$
WW et ZZ	h	$\sin(\beta - \alpha)$			
	H	$\cos(\beta - \alpha)$			
	a	0			

TABLE 2.2 – Couplages normalisés à leur valeur dans le MS entre les bosons de Higgs neutres et les bosons de jauge et fermions pour différents types de 2HDM.

Si l'on considère que  $H$  correspond à la particule découverte à une masse d'environ 125 GeV alors  $h$  correspond à un boson de Higgs additionnel de masse inférieure à 125 GeV qui peut être détecté expérimentalement. Si l'on identifie  $h$  comme la particule découverte à une masse de 125 GeV alors  $a$  peut correspondre à un boson de Higgs pseudo-scalaire de masse inférieure à 125 GeV pouvant de même être détecté expérimentalement. Dans tous les cas, la recherche d'un boson de Higgs additionnel de masse inférieure à 125 GeV permettrait de poser des contraintes fortes sur les différents couplages et donc sur les paramètres libres des modèles 2HDM [44]. Ces différents éléments motivent donc la recherche d'une telle particule.

## 2.2 Modèles de Higgs Composites

On étudie ici une combinaison de deux théories : la Technicouleur (TC) et modèles de Higgs composites (CH). Ces deux théories reposent en effet sur des principes semblables rendant leur combinaison aisée, notamment toute théorie HC admet une limite TC (l'inverse n'est cependant pas vrai). Ces deux théories se basent sur le postulat qu'il existe une nouvelle interaction à couplage fort à des énergies plus élevées que celle du MS. Elles permettent ainsi de régler le problème de hiérarchie de la masse du boson de Higgs, présentent des candidats à la matière noire (plus de détails dans la section 8.3) et proposent une explication dynamique de la brisure de symétrie spontanée pouvant expliquer l'asymétrie entre matière et antimatière [45]. On présente ici brièvement ces deux théories avant d'étudier, avec un peu plus de détails, un exemple de combinaison possible, en se basant sur [44].

### 2.2.1 Introduction aux théories Technicouleur et Higgs Composites

#### La Technicouleur

L'idée des théories TC est de considérer une nouvelle théorie de jauge à couplage fort associé à un groupe de symétrie  $G_{TC} = SU(N_T)_{TC}$  contenant alors  $N_T^2 - 1$  "techniglons". Des fermions sensibles à cette nouvelle interaction, appelés alors "technifermions", peuvent alors être considérés sous l'hypothèse qu'il se transforment sous une symétrie de saveur  $G_f$  qui doit au moins<sup>2</sup> contenir la symétrie électrofaible  $SU(2)_L \times U(1)_Y$ . Les fermions classiques sont considérés comme des singulets selon  $G_{TC}$  mais non nécessairement selon  $G_f$ .

Comme dans QCD, un condensat de technifermions peut apparaître à des échelles d'énergie plus faibles, brisant alors la symétrie initiale  $G_f$  en un sous-groupe  $F \subset G_f$ . D'après le théorème de Goldstone, cette brisure de symétrie génère  $\dim(G_f) - \dim(F) = n$  bosons de Goldstone dont 3 d'entre eux doivent être de masse nulle afin d'être absorbés par les bosons  $W^\pm$  et  $Z^0$ . Les autres peuvent être aussi de masse nulle ou bien acquérir une masse si la symétrie est brisée de manière explicite<sup>3</sup>; dans ce cas ils deviennent des pseudo-bosons de Nambu-Goldstone (pNGB).

La brisure de la symétrie de saveur donne naissance à de nombreuses particules composites composées de (anti)techniquarks que l'on appelle "technimésons" et "technibaryons" selon s'ils ont un spin entier ou demi-entier. Toutes ces particules acquièrent une masse  $M$  de l'ordre de l'échelle de la brisure de symétrie  $\Lambda_{TC}$  telle que  $M \sim \frac{\Lambda_{TC}}{4\pi}$ . La brisure de symétrie fait aussi apparaître des particules scalaires qui sont des singulets de  $F$  pouvant être identifiés comme des bosons de Higgs.

Les fermions classiques sont cependant toujours sans masse après cette brisure de symétrie. Une possibilité pour remplacer les termes de Yukawa du MS est d'ajouter une brisure de

2. Elle doit aussi contenir la symétrie custodiale  $SU(2)_L \times SU(2)_R$  afin de garantir un rapport  $\frac{m_W^2}{\cos^2 \theta_w m_Z^2}$  proche de la valeur expérimentale.

3. Les termes brisant explicitement la symétrie sont d'ordre faibles.



symétrie supplémentaire permettant de générer un couplage entre fermions classiques et technifermions, étendant alors la théorie TC [46].

Si les théories TC ont été très attractives par le passé, elles rendent difficilement compte des observations expérimentales. Notamment la masse la plus faible des condensats pouvant être interprétés comme des bosons de Higgs a été prédite par des calculs sur réseaux comme étant de l'ordre du TeV [47] soit bien plus importante que la masse de la particule découverte au LHC. Les théories TC ont alors besoin d'ajustements fins pour expliquer une masse mesurée expérimentalement aussi faible, et on retrouve un problème de naturalité. Plus généralement, il est nécessaire d'apporter beaucoup de modifications à la théorie initiale pour l'ajuster aux résultats expérimentaux. Une des modifications possibles est d'unifier les théories TC avec des modèles de HC.

### Les Higgs Composites

Des modèles similaires à ceux des théories TC sont les modèles de Higgs composites. Dans ces théories, une symétrie de jauge à interaction forte, associée à une symétrie de saveur  $G_f$  agissant sur un nouveau secteur fermionique, est aussi présente à hautes énergies. La symétrie de saveur  $G_f$  est toujours brisée à une échelle  $\Lambda$  en un sous-groupe  $F$  qui contient cette fois-ci la symétrie électrofaible  $SU(2)_L \times U(1)_Y$ .

Cette brisure de symétrie donne naissance à  $\dim(G_f) - \dim(F) = n$  bosons de Goldstone ou pNGB. Les théories HC requièrent qu'il y ait au moins un doublet de  $SU(2)_L$  parmi eux, ce doublet est alors identifié comme le doublet de Higgs du MS. Un mécanisme de brisure de la symétrie électrofaible similaire à celui du MS permet alors générer la masse des bosons de jauge associés.

Dans les modèles HC, des termes supplémentaires brisent explicitement, au niveau de l'arbre ou à boucles, la symétrie  $G_f$  permettant aux pNGB d'acquérir une masse. Ces pNGB peuvent alors être identifiés comme des bosons de Higgs. L'avantage de considérer les bosons de Higgs, et notamment la particule découverte au LHC à une masse de 125 GeV, comme des pNGB est que les pNGB ont une masse naturellement faible comparée à celle des condensats de technimésos et technibaryons qui apparaissent aussi dans les théories HC. Il est donc possible d'obtenir un candidat Higgs de masse 125 GeV sans ajustement fin. De plus, les candidats pNGB ne sont définis que jusqu'à l'échelle  $\Lambda$  et non plus jusqu'à l'échelle de Planck ce qui permet de résoudre le problème de l'hierarchie de la masse du boson de Higgs.

La masse des fermions est souvent obtenue grâce à de la composition partielle [48] : les fermions classiques du MS se mélangent avec les fermions lourds du nouveau secteur, que l'on appelle techniquarks par analogie avec les théories TC, et leur masse apparaît lorsque ces techniquarks se condensent en dessous de l'échelle d'énergie  $\Lambda$ . La hiérarchie des masses des fermions du MS est obtenue en variant le degré de couplage avec les techniquarks : les fermions de masse faible sont considérés comme quasiment élémentaires alors que les fermions

lourds de troisième génération sont fortement couplés aux techniquarks lourds. A noter que le mécanisme de composition partielle peut être utilisé pour les théories TC, et inversement le mécanisme d'extension des théories TC générant la masse des fermions peut être utilisé pour les modèles CH.

## Dynamique Composite Fondamentale

On a vu que les théories TC and HC sont basées sur une nouvelle interaction à couplage fort impliquant de nouvelles particules, les techniquarks, se transformant sous une symétrie de saveur globale  $G_f$ . Cette symétrie est brisée à une échelle  $\Lambda$  et produit alors des technimésons et technihadrons par condensation des techniquarks. Cette brisure de symétrie produit aussi des bosons de Goldstone qui peuvent être massifs, bien que la masse générée soit faible par rapport à celle des autres condensats.

Il y a cependant deux différences fondamentales entre ces deux théories. Premièrement, la symétrie électrofaible est entièrement brisée lors de la brisure de la symétrie  $G_f$  dans les théories TC alors qu'elle est préservée dans les théories HC, et un autre mécanisme doit être pris en compte dans ce cas pour briser la symétrie  $SU(2)_L \times U(1)_Y$ . La seconde est que la particule découverte en 2012 au LHC est considérée comme étant un scalaire neutre issu d'un condensat dans les théories TC alors qu'elle est considérée comme un pNGB faisant parti d'un doublet  $SU(2)_L$  dans les théories HC.

Finalement, on peut combiner ces deux théories au niveau effectif en considérant un état fondamental, un vide,  $\Sigma_0$  qui est un mélange des vides des deux théories  $\Sigma_{TC}$  pour les théories TC et  $\Sigma_{CH}$  pour les modèles HC. Le vide  $\Sigma_0$  le plus général brise donc partiellement la symétrie électrofaible. On doit aussi imposer que les bosons de Goldstone résultant de la brisure de symétrie  $G_f$  contiennent au moins un doublet  $SU(2)_L$ .

La particule de masse 125 GeV sera alors considérée comme étant un mélange entre un pNGB, venant de la partie HC de la théorie, et d'un scalaire neutre issu d'un condensat de masse  $M$ , appelé  $\sigma$ , provenant de la partie TC. Une telle combinaison est appelée Dynamique Composite Fondamentale (FCD).

### 2.2.2 Dynamique Composite Fondamentale $SU(4) \rightarrow Sp(4)$

On présente ici brièvement un modèle spécifique effectif de FCD basé sur une brisure de symétrie  $G_f$  telle que  $SU(4) \rightarrow Sp(4)$ . On assume qu'il existe un condensat brisant la symétrie initiale  $SU(4)$  [49] et, par souci de simplification et clarté, on se concentre ici sur une approche effective sans entrer dans les détails concernant l'interaction à couplage fort initiale ou ses extensions permettant la génération de la masse des fermions. Une description phénoménologique de ce modèle est présentée dans la section 8.3.

### Symétries et vides

La symétrie du groupe spécial unitaire de dimension 4 est brisée en une symétrie de groupe symplectique de dimension 4 ce qui génère alors  $15-10=5$  bosons de Goldstone. Ces bosons sont placés dans un singulet et un bi-doublet de  $SU(2)_L \times SU(2)_R \subset Sp(4)$  [50]. Le bi-doublet peut être assimilé au doublet du MS contenant un candidat Higgs  $h$  ainsi que les 3 bosons de Goldstone absorbés par les bosons  $W^\pm$  et  $Z^0$ . Le singulet est alors un pseudo-scalaire  $\eta$  pouvant correspondre à un candidat à la matière noire (voir notamment la section 8.3).

On peut définir deux vides distincts qui brisent la symétrie  $SU(4)$  mais conservent  $Sp(4)$  [44] :

$$\Sigma_{TC} = \begin{pmatrix} 0 & \mathbb{1}_2 \\ -\mathbb{1}_2 & 0 \end{pmatrix} \quad \text{et} \quad \Sigma_{HC} = \begin{pmatrix} i\sigma_2 & 0 \\ 0 & -i\sigma_2 \end{pmatrix} \quad (2.10)$$

où  $\sigma_2$  est une matrice de Pauli. Le vide  $\Sigma_{TC}$  brise la symétrie électrofaible et est donc considéré comme le vide des théories TC, le vide  $\Sigma_{HC}$  ne brise pas cette symétrie et est donc associé aux théories HC. Ces deux vides préservent donc des sous-groupes différents de  $Sp(4)$  et sont des vides viables pour la brisure de la symétrie de saveur  $SU(4)$ .

Afin d'unir les théories TC et HC, on définit un vide général  $\Sigma_0$  en tant que combinaison linéaire des deux vides précédents :

$$\Sigma_0 = \cos \theta \Sigma_{CH} + \sin \theta \Sigma_{TH} \quad (2.11)$$

où  $\theta \in [0, \frac{\pi}{2}]$  quantifie le mélange entre les approches HC et TC et est un paramètre libre du modèle FCD considéré ici. Ainsi dans la limite  $\theta \rightarrow 0$  on retrouve une théorie purement HC et dans celle  $\theta \rightarrow \pi/2$  on retrouve une théorie purement TC.

### Lagrangien effectif FCD

La composante dynamique des bosons de Goldstone dans le lagrangien peut être déterminée en suivant le formalisme de Callan-Coleman-Wess-Zumino (CCWZ) [51, 52]. On définit pour cela le champ  $\Sigma$  contenant les cinq champs de Goldstone :

$$\Sigma = e^{i \sum_{j=1}^5 \phi_j Y^j} \cdot \Sigma_0 \quad (2.12)$$

où  $\phi_j, j \in 1, \dots, 5$  sont les champs des bosons de Goldstone et  $Y^i$  les générateurs de la brisure de symétrie associés (ils sont explicitement visibles dans [44]),  $f$  est la constante de désintégration de ces bosons par analogie avec  $f_\pi$  la constante de désintégration des pions en QCD.

La partie dynamique du lagrangien contenant les bosons de Goldstone est alors donnée à l'ordre le plus faible par [44] :

$$L_{dyn} = f^2 \text{Tr} \left[ (D_\mu \Sigma)^\dagger (D^\mu \Sigma) \right] \quad (2.13)$$

avec  $D_\mu$  la dérivée covariante définie comme :

$$D_\mu = \partial_\mu \Sigma - \left[ igW_\mu^a (S_a \Sigma + \Sigma S_a^T) + ig'B_\mu (S_6 \Sigma + \Sigma S_6^T) \right] \quad (2.14)$$

avec  $S_a$ ,  $a=1,2,3$  et  $S_6$  des générateurs de  $Sp(4)$  associés respectivement à  $SU(2)_L$  et  $U(1)_Y$ . On peut effectuer une rotation appropriée, en jauge unitaire, afin de supprimer les trois champs  $\phi_i$   $i=1,2,3$  des bosons de Goldstone absorbés par les bosons  $W^\pm$  et  $Z^0$  pour obtenir une masse.

Dans le contexte d'une théorie effective, le lagrangien de Yukawa équivalent donnant la masse des fermions du MS est donné par [53] :

$$L_{Yukawa} = f \left( y'_u{}^{ij} (Q_{L,i} \tilde{u}_{R,j})^\dagger_\alpha + y'_d{}^{ij} (Q_{L,i} \tilde{d}_{R,j})_\alpha + y'_l{}^{ij} (L_{L,i} \tilde{l}_{R,j})^\dagger_\alpha \right) \text{Tr}[P^\alpha \Sigma] + h.c. \quad (2.15)$$

avec  $\alpha = 1,2$  indices  $SU(2)_L$ ,  $(i,j) = 1,2,3$ ,  $y'$  sont les nouveaux couplages de yukawa,  $\tilde{f}_R$  ( $f = u, d, l$ ) est le singulet du conjugué de charge pour les fermions (quarks de type up, down et les leptons) du MS,  $Q_L$  et  $L_L$  étant les doublets des quarks et leptons du MS, et  $P^\alpha$  est une matrice de projection projetant hors de  $\Sigma$  [44].

Pour obtenir le lagrangien FCD il faut cependant prendre en compte le scalaire composite neutre  $\sigma$ . Comme  $\sigma$  est un singulet sous  $Sp(4)$ , on ajoute simplement [44] des termes cinématiques correspondant et un facteur  $\zeta(\theta)$  dépendant du champ  $\sigma$  devant chacun des termes du lagrangien évoqué précédemment. Le facteur  $\zeta(\theta)$  est a priori différent selon le terme  $X$  considéré (Goldstone, condensat ou fermions) et peut être écrit comme [53] :

$$\zeta_X(\sigma) = 1 + \frac{\zeta_X^{(1)}}{4\pi f} \sigma + \frac{1}{2} \frac{\zeta_X^{(2)}}{(4\pi f)^2} \sigma^2 + \dots \quad (2.16)$$

où les  $\zeta_X^{(1)}$  et  $\zeta_X^{(2)}$  ( $X = G, M, f$ ) sont des paramètres inconnus de l'ordre de l'unité.

Le lagrangien total s'écrit alors comme :

$$L_{FCD} = \zeta_G(\sigma) f^2 \text{Tr} \left[ (D_\mu \Sigma)^\dagger (D^\mu \Sigma) \right] + \frac{1}{2} \partial_\mu \sigma \partial^\mu \sigma - \frac{1}{2} M^2 \zeta_M(\sigma) \sigma^2 + V_{corr} \\ + f \zeta_f(\sigma) \left( y'_u{}^{ij} (Q_{L,i} \tilde{u}_{R,j})^\dagger_\alpha + y'_d{}^{ij} (Q_{L,i} \tilde{d}_{R,j})_\alpha + y'_l{}^{ij} (L_{L,i} \tilde{l}_{R,j})^\dagger_\alpha \right) \text{Tr}[P^\alpha \Sigma] + h.c. \quad (2.17)$$

où  $V_{corr}$  est un potentiel prenant en compte les corrections venant des termes brisant explicitement la symétrie  $SU(4)$  [44].

En développant le premier terme de ce lagrangien, on obtient les termes de masse des bosons  $W^\pm$  et  $Z^0$ , de couplage entre les bosons  $h$ ,  $\eta$  et  $\sigma$  et les paires de bosons de jauge  $W^+W^-$  et  $Z^0Z^0$ , ainsi que les couplages à trois champs pour les bosons  $h$ ,  $\eta$  et  $\sigma$ . Le terme de Yukawa permet lui d'obtenir les termes de masse des fermions du MS ainsi que les couplages entre  $h$ ,  $\eta$  et  $\sigma$  et les paires de fermions.

Ces différentes masses sont alors données par [53] avec  $\sin \theta = s_\theta$  et  $\cos \theta = c_\theta$  :

$$m_W = \sqrt{2} g f s_\theta, \quad m_Z = \frac{m_W}{\cos \theta_W}, \quad m_f = y'_f s_\theta f \quad (2.18)$$

En identifiant ces masses avec les masses du MS, on obtient des relations entre la valeur attendue dans le vide  $v$  du MS et la constante de désintégration  $f$ , ainsi qu'entre les couplages de Yukawa du MS  $y_f$  et  $y'_f$  :

$$v \equiv \frac{2m_W}{g} = 2\sqrt{2}fs_\theta \quad \text{et} \quad y_f = \frac{y'_f}{2} \quad (2.19)$$

Les valeurs des couplages précédemment mentionnées sont lisibles [53], normalisées à la valeur du couplage obtenue avec le boson de Higgs dans le MS, dans le tableau 2.3. On observe que  $\eta$  ne se couple pas, au niveau de l'arbre, directement<sup>4</sup> avec une paire de bosons de jauge ou de fermions faisant de  $\eta$  une particule stable qui est alors un bon candidat à la matière noire, étant donné que  $\eta$  est le nouveau boson le plus léger ici.

Champs	$W^+W^+, Z^0Z^0$	$f\bar{f}$
$h$	$c_\theta$	$c_\theta$
$\sigma$	$\frac{\xi_G^{(1)}}{2\sqrt{2}}s_\theta$	$\frac{\xi_f^{(1)}}{\sqrt{2}}s_\theta$
$\eta$	-	-

TABLE 2.3 – Couplages à trois particules pour des champs  $h$ ,  $\sigma$  et  $\eta$  et des paires de bosons  $W$  et  $Z^0$  ainsi que des fermions normalisés à leur valeur obtenue dans le cas du boson de Higgs du MS.

### Génération de la masse des pNGB

Nous avons dérivé la masse des bosons de jauge électrofaible et des fermions du MS mais nous n'avons pas encore étudié comment les bosons  $h$  et  $\eta$  acquièrent leur masse. On rappelle que leur masse, de faible valeur, émerge de corrections provenant de termes brisant explicitement la symétrie  $SU(4)$  dans le lagrangien, formant alors deux pNGB et modifiant la masse  $M$  du condensat  $\sigma$ .

Ces corrections sont prises en compte dans le terme de potentiel  $V_{corr}$  dans le lagrangien de l'Eq. 2.17. Les termes de corrections ne seront pas explicités ici mais plus de détails peuvent être trouvés dans [53]. Les masses obtenues suite au développement des termes du potentiel sont alors [53] :

$$m_\eta^2 = \frac{f^2}{4}C_t y_t'^2 \quad (2.20)$$

$$m_h^2 = m_\eta^2 s_\theta^2 \quad (2.21)$$

$$m_\sigma^2 = M^2 - m_\eta^2 \left( (\xi_t^2 + \xi_t^{(2)})s_\theta^2 + \xi_m^{(2)}c_\theta^2 \right) \quad (2.22)$$

où  $C_t$  et  $\xi_m^{(2)}$  sont deux nouveaux paramètres de l'ordre de l'unité apparaissant dans le potentiel.

4. Des couplages  $\eta\eta VV$  sont néanmoins possibles.

A cause des facteurs  $\zeta(\sigma)$  présents dans le potentiel, le scalaire  $h$  n'est pas un état propre de masse et se mélange avec  $\sigma$ . Les états propres de masse  $h_1$  et  $h_2$  avec  $m_{h_1} < m_{h_2}$  sont, à nouveau, obtenus par rotation d'un angle  $\alpha$  des champs  $h$  et  $\sigma$  tel que :

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} h \\ \sigma \end{pmatrix} \quad (2.23)$$

On fait l'hypothèse que l'on est dans la limite  $\theta$  faible qui est justifié puisque  $M \sim m_\eta \sim f \gg v \sim 2\sqrt{2}f\theta$ . Sous cette hypothèse, les masses de  $h_1$  et  $h_2$  sont données par [53] :

$$m_{h_2}^2 \sim M^2 - m_\eta^2 \zeta_m^{(2)} \quad (2.24)$$

$$m_{h_1}^2 \sim m_\eta^2 \theta^2 \left( 1 - 4 \frac{m_\eta^2}{m_{h_2}^2} (\zeta_t - \zeta_m)^2 \right) \quad (2.25)$$

L'angle  $\alpha$  de mélange entre les deux états satisfait d'ailleurs, toujours dans l'approximation d'un petit angle  $\theta$ , la relation :

$$\tan \alpha \sim \alpha \sim 2 \frac{m_\eta^2}{m_{h_2}^2} (\zeta_t - \zeta_m) \theta \quad (2.26)$$

On peut finalement identifier  $h_2$  comme étant la particule découverte en 2012 à une masse de près de 125 GeV.  $h_1$  est alors un boson de Higgs de masse plus légère et  $\eta$  est un pseudo-scalaire neutre pouvant également posséder une masse inférieure à 125 GeV.

## 2.3 Modèles Supersymétriques

D'autres classes de modèles impliquent la présence de nouvelles particules dont de nouveaux bosons de Higgs de plus faible masse : les **modèles de Supersymétrie** (SUSY). La SUSY est, là encore, une extension au MS : on rajoute à ce dernier une symétrie supplémentaire qui associe à chaque particule du MS un partenaire qui est appelé superparticule. Ainsi cette symétrie associe aux fermions du MS des bosons, et aux bosons de MS des fermions. Le formalisme théorique est sensiblement différent de celui décrit plus haut, on utilise par exemple des objets comme les superchamps et les superpotentiels. Pour ces raisons, on n'effectuera ici qu'une brève introduction à la SUSY.

La SUSY a connu un engouement notable il y a quelques dizaines d'années. En effet, la SUSY est un modèle élégant qui permet de résoudre le problème de la hiérarchie de la masse du boson de Higgs : les superpartenaires des particules du MS ajoutent une contribution aux corrections à boucles à la masse du boson de Higgs identifié qui compensent la contribution aux termes quadratiques apportée par les particules du MS. Ainsi le facteur  $b$  de l'équation 1.55 serait égal à zéro faisant alors disparaître la divergence quadratique<sup>5</sup>.

5. Il reste, sous cette hypothèse, un terme logarithmique n'induisant pas de problème de naturalité.

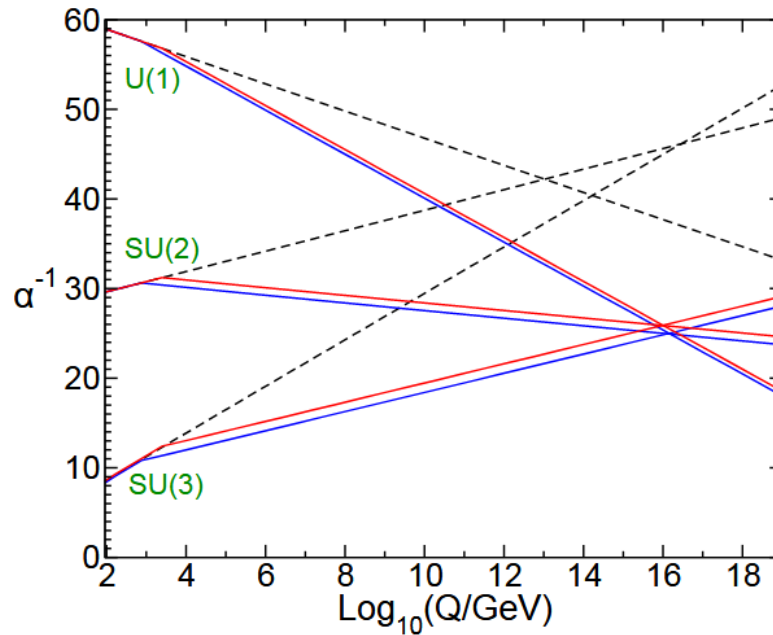


FIGURE 2.1 – Variation des constantes de couplages, obtenues à partir des équations du groupe de renormalisation, pour les symétries  $U(1)$ ,  $SU(2)$  et  $SU(3)$  en fonction de l'énergie. Les traits pointillés correspondent au cas MS et les traits pleins au cas SUSY (MSSM) avec des variations de la masse des superpartenaires entre 750 GeV (rouge) et 2,5 TeV (bleu) [54].

La SUSY permet aussi d'aboutir à une convergence des couplages des interactions électromagnétique, faible et forte à une échelle d'énergie élevée (de l'ordre de  $10^{16}$  GeV pour le modèle de Minimal Supersymmetric Model [54]) comme visible sur la figure 2.1.

De plus, la SUSY propose un candidat à la matière noire, la particule SUSY la plus légère qui est souvent, selon les modèles, un neutralino qui est une particule constituée d'un mélange de Higgsino, photino et Zino (voir tableau 2.4).

Finalement, certains modèles de SUSY essaient de décrire la gravitation en intégrant un graviton ainsi que son superpartenaire. La SUSY est aussi utilisée dans la plupart des théories des cordes.

Il existe de nombreux modèles de SUSY. Le modèle le plus simple est appelé Minimal Supersymmetric Model (MSSM). Ce modèle, que l'on peut inclure dans les modèles 2HDM de type II, présente 5 bosons de Higgs ainsi que leur superpartenaires. Cependant la masse théorique de ces diverses particules est difficilement compatible avec les masses et les contraintes mesurées expérimentalement [55]. Dans ce modèle, la masse des superpartenaires est égale à celle des particules du MS. Or aucun de ces superpartenaires n'a été découvert, leur masse est donc potentiellement plus élevée ce qui signifie que la symétrie SUSY est brisée aux échelles d'énergie du LHC. A cause de cela, les termes de corrections en masse ne se compensent plus et le problème de la hiérarchie de la masse du boson de Higgs réapparaît.

On lui préférera donc le Next-to Minimal Supersymmetric Model (NMSSM) [56] qui est, comme son nom l'indique, une simple extension du MSSM auquel on ajoute un superchamp scalaire complexe, ce qui a pour principal avantage de moins contraindre les masses théoriques et donc de les rendre possiblement compatibles avec les observations expérimentales. Le NM-



Spin	Particules du Modèle Standard avec secteur de Higgs étendu et graviton	Superpartenaires
0	Higgs scalaires $h, H$ Higgs pseudo-scalaire $a$ Higgs chargés $H^\pm$	squarks $\tilde{q}$ sleptons $\tilde{l}, \tilde{\nu}$
1/2	quarks $q$ leptons $l, \nu$	neutralinos $\tilde{\chi}^0$ (higgsinos $\tilde{h}, \tilde{a}$ , zino $\tilde{Z}$ et photino $\tilde{\gamma}$ ) charginos $\tilde{\chi}^\pm$ (higgsinos $\tilde{H}^\pm$ et winos $\tilde{W}^\pm$ ) gluinos $\tilde{g}$
1	photon $\gamma$ bosons $Z^0$ et $W^\pm$ gluons $g$	-
3/2		gravitino $\tilde{G}$
2	graviton $G$	

TABLE 2.4 – Particules et superpartenaires pour le MSSM selon leur spin. Les neutralinos et charginos sont composés des champs marqués entre parenthèses.

SSM prévoit de nombreuses nouvelles particules dont notamment 7 particules de type bosons de Higgs : 2 bosons chargés  $H^\pm$ , 2 pseudo-scalaires neutres  $a_1$  et  $a_2$  et 3 scalaires neutres  $H_1, H_2$  et  $H_3$  avec  $m_{H_1} < m_{H_2} < m_{H_3}$ . Si l'on identifie  $H_2$  ou  $H_3$  (même si plus difficile) avec le boson de Higgs détecté au LHC, ce modèle pourrait conjecturer un, ou deux, boson(s) scalaire(s) de plus faible masse et potentiellement deux pseudo-scalaires.

La théorie SUSY prédit donc de nombreuses nouvelles particules dont plusieurs d'entre elles peuvent correspondre à un boson de Higgs additionnel de masse inférieure à 125 GeV. Le principal problème de la SUSY est qu'aucune de ces nouvelles particules n'a pour l'instant été découverte. Des analyses recherchent cependant toujours de telles particules à travers des signatures multiples comme une importante énergie transverse manquante, des particules à longue durée de vie ou encore en recherchant de nouveaux bosons de Higgs.

## Conclusion

On a donc ici présenté quelques théories BSM permettant de pallier certaines des limites du MS. Les modèles à deux doublets de Higgs, les modèles de Higgs composites (incluant ou non la Technicouleur) et les modèles supersymétriques proposent des solutions aux problèmes de naturalité, de la matière noire ou encore de l'asymétrie entre matière et antimatière selon le modèle considéré. Il existe cependant de nombreux autres théories et modèles BSM comme les modèles à dimensions supplémentaires type Randall-Sundrum [57].

Ces théories au-delà du MS, en essayant de corriger les insuffisances de ce dernier, prédisent la présence de nouvelles particules dont parfois de nouveaux bosons de Higgs. Si l'on identifie la particule découverte en 2012 à un de ces bosons de Higgs, il resterait, dans ces modèles, d'autres particules scalaires de même type, dont certaines pourraient être plus légères que la particule détectée à 125 GeV, qui pourraient être découvertes.



Ces prédictions sont autant de motivations nous encourageant à rechercher de nouvelles particules afin d'effectuer une sélection dans tous ces modèles théoriques. Même en absence de découverte, de telles recherches sont intéressantes puisqu'elles permettent de poser des contraintes sur les différents paramètres de ces modèles.

Notre objectif est donc de rechercher la présence d'un tel boson additionnel, qui se désintégrerait en deux photons, au sein d'un des quatre détecteurs du LHC, le détecteur CMS.



# Reconstruction des événements dans le détecteur CMS

## Sommaire

---

3.1	Le détecteur CMS du LHC . . . . .	52
3.1.1	Présentation du détecteur CMS . . . . .	52
	L'aimant solénoïde . . . . .	54
	Le trajectographe . . . . .	56
	Le calorimètre électromagnétique . . . . .	58
	Le calorimètre hadronique . . . . .	63
	Les chambres à muons . . . . .	65
	Le système de déclenchement . . . . .	66
3.2	Reconstruction et simulation des événements physiques . . . . .	68
3.2.1	Algorithme de Particule Flow . . . . .	69
	Les éléments Particle Flow . . . . .	69
	Algorithme de lien . . . . .	70
	Identification des particules . . . . .	70
3.2.2	Simulation du détecteur . . . . .	72

---

Afin de permettre la recherche de n'importe quelle signature physique impliquant des photons, il faut commencer par étudier la manière dont sont reconstruits les événements impliquant des photons au sein d'un détecteur ; à savoir comment, à partir d'interactions de particules à l'échelle des interactions fondamentales, obtenir des données informatiques contenant des informations sur la trajectoire et l'énergie des particules en jeu.

## 3.1 Le détecteur CMS du LHC

### 3.1.1 Présentation du détecteur CMS

Le détecteur CMS (pour "Compact Muon Solenoid") est un gigantesque détecteur de forme cylindrique de 21 m de long et 15 m de diamètre pour un poids de 14000 tonnes, construit autour d'une de ses caractéristiques principales : un aimant solénoïde supraconducteur refroidi à une température de  $-268.5\text{ °C}$  produisant un champ magnétique de presque 4 T [58].

CMS joue le rôle de détecteur dans le sens où il est constitué de différentes couches spécifiques permettant de stopper, suivre ou mesurer les particules émergeant des collisions proton-proton en son sein, ce qui fournit aussi des informations sur la trajectoire, l'énergie et finalement l'identité potentielle des particules en présence.

Il dispose ainsi d'une structure en oignon dont les constituants principaux sont, à partir du point d'interaction et comme on peut le voir sur la figure 3.1, le trajectographe en silicium permettant d'obtenir la trajectoire des particules chargées, le calorimètre électromagnétique ayant pour rôle de stopper et mesurer les électrons et photons issus d'interactions électromagnétiques, le calorimètre hadronique<sup>1</sup> ayant pour rôle de stopper et mesurer les produits d'interactions fortes, de l'aimant solénoïde mentionné ci avant, et de chambres à muons permettant la détection de muons avec une très bonne résolution en impulsion transverse.

Le poids, conséquent, du détecteur au vu de sa taille lui confère le caractère compact. Le fait de posséder un aimant solénoïde puissant et pouvoir caractériser les muons avec précision grâce aux chambres consacrées explique son appellation de Solénoïde Compact à Muons, soit CMS en anglais.

Outre ces couches spécifiques, le détecteur est séparé en deux parties selon un système de coordonnées propres aux expériences du LHC.

Le repère de coordonnées cartésiennes orthonormé direct  $(O, x, y, z)$  prend pour origine  $O$  le point de collision des faisceaux, donc le centre du détecteur ; les axes  $x$  et  $y$  sont perpendiculaires à la direction des faisceaux, l'axe  $x$  pointant vers le centre du LHC tandis que l'axe  $y$  pointe

---

1. Le calorimètre hadronique situé à l'intérieur de l'aimant est une spécificité notable de CMS.

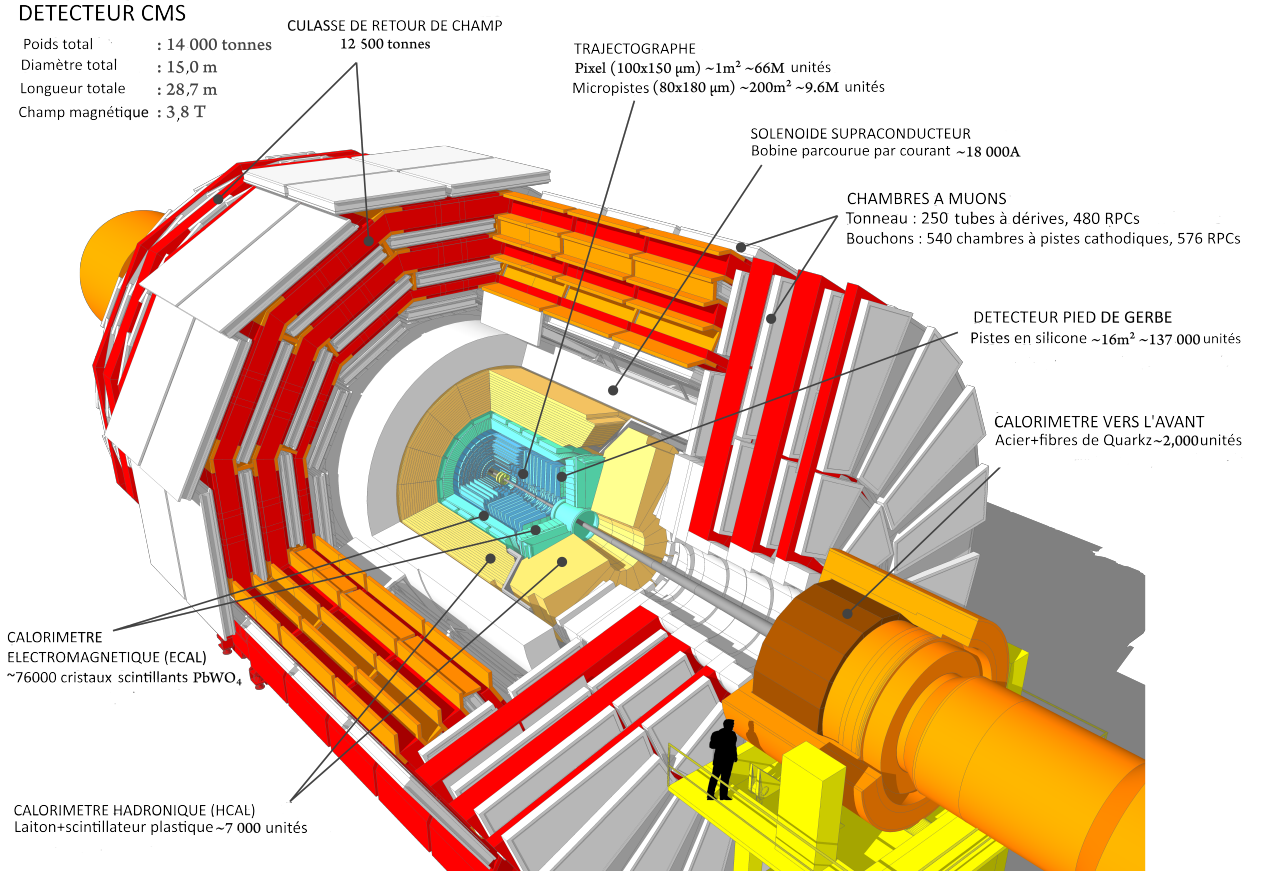


FIGURE 3.1 – Coupe du détecteur CMS présentant les différentes couches et éléments. Traduit de [58]

vers le ciel. L'axe  $z$  est, lui, tangent à la direction des faisceaux, c'est ainsi l'axe privilégié pour la symétrie cylindrique. On peut ainsi définir l'angle polaire  $\theta$  entre le point considéré et l'axe  $z$ , et l'angle azimutal  $\phi$  entre la projection du point considéré sur le plan  $(Oxy)$  et l'axe  $x$ . Ces deux angles permettent de définir des variables invariantes de Lorentz pour un boost selon l'axe  $z$ . Travailler avec des variables invariantes de Lorentz est important ici puisque, contrairement aux collisionneurs leptoniques, seule une fraction de l'impulsion du proton est emporté par le quark/gluon à l'origine de la collision ce qui résulte alors en un boost selon la direction  $z$  différent selon les référentiels considérés. On définit donc la pseudo-rapacité  $\eta$ , telle que  $\Delta\eta$  soit invariant de Lorentz dans les domaines ultra-relativistes, comme :

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \quad (3.1)$$

ainsi que l'angle de séparation  $\Delta R$ , alors <sup>2</sup> aussi invariant de Lorentz, entre deux particules possédant des angles respectifs  $(\eta_1, \phi_1)$  et  $(\eta_2, \phi_2)$  tel que :

$$\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2} \quad (3.2)$$

Comme annoncé, on sépare la partie "tonneau" du détecteur définie grossièrement par une couverture en pseudo-rapacité correspondant à  $|\eta| < 1,5$  couvrant la majorité du détecteur.

2.  $\Delta\phi$  est invariant selon un boost selon l'axe  $z$  puisque définit dans un plan orthogonal à la ligne de faisceaux.

L'autre partie, appelée "bouchon", concerne les régions proches de la ligne de faisceau comme on peut le voir sur la figure 3.2.

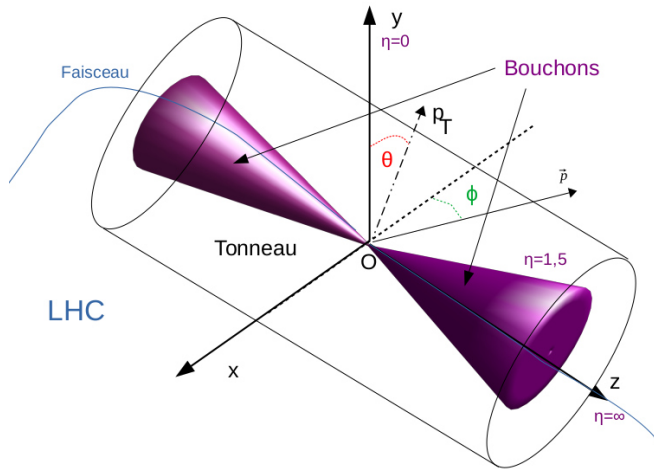


FIGURE 3.2 – Schéma du détecteur CMS faisant apparaître le système de coordonnées ainsi que la séparation entre tonneau et bouchon.

Les protons étant accélérés selon l'axe z, le moment total transverse est nul. Il est alors intéressant de définir des variables transverse (projections sur le plan (Oxy) transverse à l'axe z) tels l'énergie et le moment transverse à partir de l'énergie totale  $E$  et du moment total  $\vec{p}$  :

$$\begin{aligned}
 p_T &= \sqrt{p_x^2 + p_y^2} = \|\vec{p}\| \sin \theta = \frac{\|\vec{p}\|}{\cosh \eta} \\
 E_T &= E \sin \theta = \frac{E}{\cosh \eta}
 \end{aligned}
 \tag{3.3}$$

Pour comprendre plus profondément le fonctionnement du détecteur CMS, il est ensuite nécessaire de décrire avec plus de détails les différents éléments présentés ci-dessus.

### L'aimant solénoïde

L'aimant solénoïde n'est pas l'élément le plus proche de la ligne de faisceau mais il a une incidence sur les composants des couches plus proches de cette ligne (trajectographe et calorimètre électromagnétique et hadronique). En effet, avec sa taille importante (près de 13 m de long pour un rayon externe de 3 m), il est permis de délivrer un champ magnétique de 3,8 T avec un courant d'environ 18 kA généré dans des câbles en Niobium-Titane. Pour produire un champ magnétique d'une telle intensité il est nécessaire de refroidir l'aimant à une température de près de 4,6 K pour passer dans un régime supraconducteur (sans résistance donc). Ce champ magnétique dépend, notamment, du rayon polaire où l'on se place et de la distance z sur l'axe du faisceau comme on peut l'observer sur la figure 3.3.

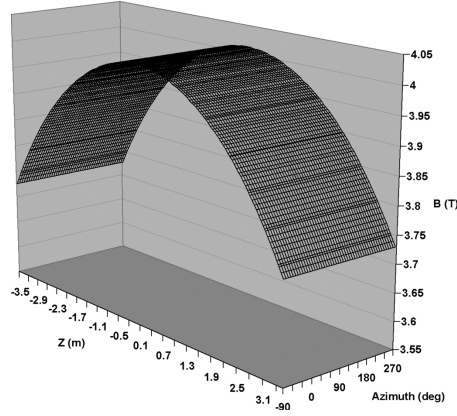


FIGURE 3.3 – Flux de densité magnétique mesurée à un rayon  $r=92$  mm en fonction de la distance sur l’axe  $z$  et l’angle azimutal  $\phi$  correspondant aux coordonnées détaillées précédemment [59].

Or des particules chargées évoluant en présence d’un champ magnétique sont soumises à une force  $F_{mag}$  qui a pour effet de courber leur trajectoire telle que :

$$\begin{aligned}\vec{F}_{mag} &= \frac{q}{m} \vec{p} \wedge \vec{B} \\ \rho &= \frac{\|\vec{p}\|}{q\|\vec{B}\|}\end{aligned}\tag{3.4}$$

où  $q$  est la charge de la particule considérée,  $m$  sa masse,  $\vec{p}$  son impulsion et  $\vec{B}$  le champ magnétique en présence. Ce dernier étant orienté suivant l’axe  $z$ , la courbure s’effectue donc dans le plan transverse. Le rayon de courbure  $\rho$  dépend alors de la charge, de la masse et de l’impulsion de la particule considérée.

Ainsi une charge positive ou négative donne des trajectoires à courbure opposée et un champ  $B$  important induit un rayon de courbure petit donc plus facile à détecter. L’observation de ces trajectoires courbes permet donc d’obtenir<sup>3</sup> le signe de la charge. Sachant que le champ  $B$  peut être déterminé avec une bonne précision [59], on peut ainsi aussi obtenir des informations sur les caractéristiques de la particule, notamment son impulsion qui joue un rôle crucial dans l’identification des évènements physiques ayant lieu suite aux collisions de particules.

A noter que le retour des lignes du champ magnétique vers l’intérieur du détecteur CMS s’effectue, comme on peut l’observer sur la figure 3.4, grâce à la présence de culasses en acier d’un poids total de 12 500 tonnes assurant le retour du champ.

Ainsi l’intense champ magnétique délivré par cet aimant permet de courber les particules, ce qui permet, couplé avec les autres mesures de précision fournies par les autres éléments du détecteur, d’obtenir des mesures de précisions sur les moments de particules possédant même une grande énergie.

3. A hautes impulsions transverses, la trajectoire n’est que faiblement courbée et l’identification de la charge est difficile.

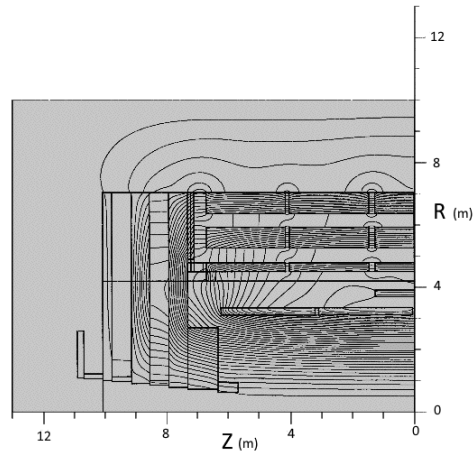


FIGURE 3.4 – Distribution du flux de densité magnétique réalisée avec des simulations en présence de culasses de retour de champ dans le détecteur CMS en coupe transverse [60].

### Le trajectographe

Comme on l'a vu, il est important de pouvoir identifier avec extrême précision les trajectoires des particules dans le détecteur : c'est en partie le rôle du trajectographe. Il fonctionne selon le principe suivant : les particules passant en son sein peuvent interagir avec, en enregistrant les signaux de ces interactions en une multitude de points spatiaux on peut alors remonter à la trajectoire effective. Ce trajectographe a été construit en silicium de manière à ne pas interagir trop lourdement avec les particules et ne pas les freiner<sup>4</sup>. Ainsi, les particules arrivant dans le trajectographe vont être confrontées à une multitude de couches de silicium de fine granularité ; elles vont ensuite interagir avec ce silicium ce qui va perturber l'état fondamental de ce dernier typiquement à travers une ionisation des atomes de silicium. Le silicium étant un semi-conducteur, ici dopé, le mouvement des électrons créé par ces ionisations produit un courant électrique que l'on peut ensuite mesurer.

Le trajectographe de CMS mesure environ 5,4 m de long pour un rayon de 2,1 m. Il couvre une surface de près de 200 m<sup>2</sup> et une zone angulaire de  $|\eta| < 2,5$ . Il est composé de deux sous-détecteurs possédant chacun leur propre système électriques de fonctionnement :

- Le détecteur à pixel qui se situe au plus proche de l'origine O, couvre une surface de plus de 1 m<sup>2</sup> et est composé, comme l'indique son nom, de plus de 66 millions de pixels de taille  $100 \times 150 \mu\text{m}$  et d'épaisseur  $285 \mu\text{m}$  répartis en trois couches dans le tonneau et deux disques dans chacun des bouchons sur une distance radiale  $4 \text{ cm} < r < 15 \text{ cm}$  et sur longueur de 98 cm sur l'axe z. Les pixels sont regroupés en matrice de  $52 \times 80$  et les signaux créés par le passage des particules sont lus par 15840 cartes de lecture dédiées. De part sa proximité avec l'origine de la collision, il joue un rôle essentiel dans la reconstruction des vertex secondaires de désintégrations de mésons B et de tau, ou encore la reconstruction de traces complexes.

4. Ceci permet aussi de diminuer les dommages causés au trajectographe et donc d'augmenter sa durée de vie



- Le détecteur à (micro)pistes qui entoure le détecteur à pixels, est situé dans des zones  $20 < r < 115$  cm, et est composé de plus de 9,6 millions de pistes réparties en 10 couches dans le tonneau et 12 disques dans chaque bouchon avec une épaisseur d'environ  $320 \mu\text{m}$  ou  $500 \mu\text{m}$ . Il est séparé en 4 sous-parties, les plus proches du point d'interaction étant le TIB (tracker inner barrel) dans le tonneau et les TID (tracker inner disks) dans les bouchons; viennent ensuite le TOB (tracker outer barrel) dans le tonneau et les TEC (tracker endcap) dans les bouchons comme cela est schématisé sur le figure 3.5

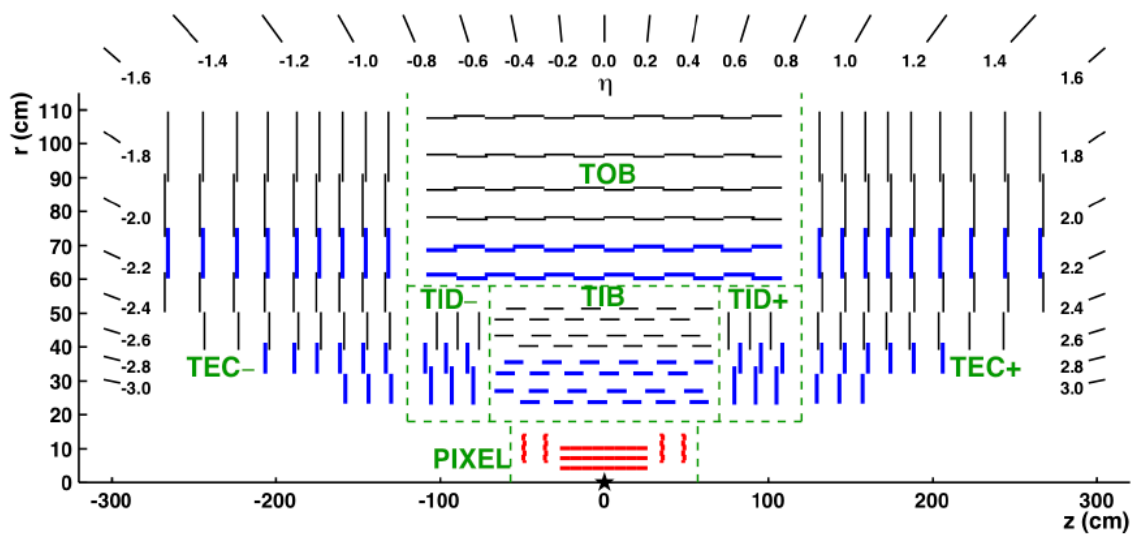


FIGURE 3.5 – Schéma en coupe du trajectographe de CMS. En rouge le détecteur à pixels. Les parties du détecteur à pistes sont visibles en bleu (mono-couche) et noir (double couche) [61].

A noter que le détecteur à pixels a été à la base construit pour résister à une luminosité instantanée, correspondant à celle considérée lors du développement du LHC, de  $10^{34} \text{cm}^{-2} \text{s}^{-1}$ . Cette luminosité a été dépassée lors de la prise des données 2016 et est encore plus importante pour les prises de données de 2017 et 2018 conduisant alors à des inefficacités donnant lieu à des pertes de données. On observe notamment sur la figure 3.6 que la couche la plus proche du point de collision enregistre une perte de 5 % des données pour 2016 ce qui est problématique. Afin de maintenir une bonne efficacité et les performances du détecteur à pixels à plus haute luminosité, un nouveau détecteur à pixels fut construit et installé après la prise des données 2016 [62].

Ce nouveau détecteur inclut une quatrième couche de pixels dans le tonneau et un troisième disque par côté pour les bouchons. De plus, le nombre de pixels a doublé (de 66 millions à 124 millions), la résistance aux radiations a augmenté et le budget matériel, voir ci-après, a diminué suite à cette installation.

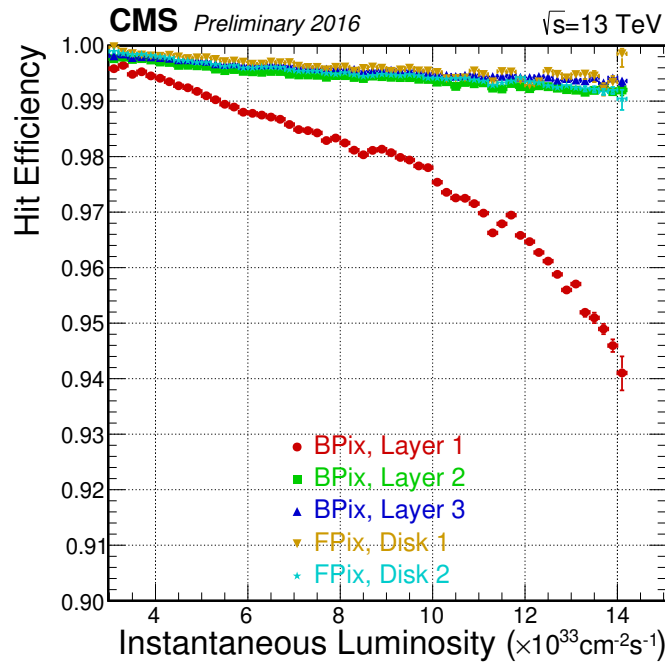


FIGURE 3.6 – Efficacité mesurée dans les différentes couches du détecteur à pixels en fonction de la luminosité instantanée pour les données 2016 [62].

Il faut en plus de cela prendre en compte tout l'équipement permettant le fonctionnement, la lecture, et le refroidissement (entre  $-10$  et  $-30$  °C) qui peut aussi interagir avec les particules en présence, on parle alors de budget matériel.

Le trajectographe joue finalement un rôle prépondérant dans le système de déclenchement des évènements, présenté dans la suite, utilisé pour réduire de manière significative le flux de données enregistrées.

### Le calorimètre électromagnétique

Le rôle du calorimètre électromagnétique (ECAL) est d'arrêter les électrons et les photons afin d'obtenir des informations très précises sur ces types de particule. Il joue donc un rôle clef dans la reconstruction des photons. Pour ce faire, il a donc fallu trouver à la fois un matériau lourd capable d'absorber et à la fois scintillateur comme décrit dans la suite. Ainsi le ECAL comporte 75848 cristaux scintillants en tungstate de plomb ( $\text{PbWO}_4$ ) répartis entre le tonneau (EB pour "ECAL barrel") défini entre  $0 < |\eta| < 1,479$  avec 61200 cristaux et les bouchons (EE pour "ECAL endcap") définies entre  $1,566 < |\eta| < 3,0$  avec 7324 cristaux.

#### Cristaux scintillants

Les cristaux de tungstate de plomb ont une densité très importante de  $8,28 \text{ g.cm}^{-3}$ , sa petite longueur de radiation de  $0,89 \text{ cm}$ , caractérisant la distance moyenne d'interaction électromagnétique, ainsi que son rayon de Molière, caractéristique de l'étalement transverse d'une gerbe électromagnétique, permettent d'assurer le caractère compact du détecteur CMS. Ils sont aussi scintillants. Au passage de particules interagissant de manière électromagnétiques

(majoritairement du rayonnement de photons par effet bremsstrahlung pour les électrons et production de paires électron-positron pour les photons), les électrons de la couche de valence des cristaux vont recevoir de l'énergie et quitter leur état fondamental avant de le retrouver en libérant des photons caractéristiques de l'énergie initialement reçue : c'est le phénomène de scintillation. Les photons ainsi émis sont directement enregistrés par des photo-détecteurs situés derrière les cristaux. Ainsi l'enregistrement de l'intensité du flux de photons émis par les cristaux selon leur position dans le ECAL permet de remonter à l'énergie initialement déposée dans ces mêmes cristaux ce qui permet, in fine, de reconstruire les particules ayant interagi avec les cristaux.

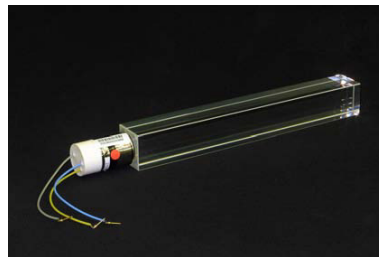


FIGURE 3.7 – Photographie d'un cristal  $\text{PbWO}_4$  provenant d'un bouchon accompagné de son photo-détecteur (phototriode à vide) [63].

Ces cristaux, visibles sur la figure 3.7 ont une forme trapézoïdale de 22 ou 23 cm de longueur selon qu'ils se situent dans le tonneau ou les bouchons. Leur face avant présente une section de  $2,2 \times 2,2 \text{ cm}^2 / 2,9 \times 2,9 \text{ cm}^2$  selon s'ils sont situés dans les tonneau/bouchons et leur face arrière, où sont situés les photo-détecteurs décrits ci-après, mesure  $2,6 \times 2,6 \text{ cm}^2 / 3,0 \times 3,0 \text{ cm}^2$ . Ils émettent préférentiellement dans le domaine visible bleu-vert avec un pic d'émission aux alentours de 420-430 nm, longueurs d'onde où les cristaux sont transparents.

Le caractère transparent des cristaux, notamment aux longueurs d'onde précédentes, est fondamental puisqu'il permet de mesurer l'intensité de la lumière induite. Comme la majorité des sous-détecteurs, les cristaux de tungstate de plomb du ECAL sont soumis à de l'usure par rayonnement. Le passage incessant de particules dans le ECAL créent, en effet, des défauts de structure, parfois irrémédiables (notamment dans des interactions avec des hadrons) dans les cristaux appelés centres de couleur qui absorbent une partie de la lumière émise et provoquent ainsi une perte de transparence. S'il paraît inconcevable de remplacer les cristaux de façon périodique, on peut par contre mesurer la réponse optique des cristaux en temps réel pour pouvoir corriger, comme on le verra dans le chapitre suivant, les informations mesurées et notamment l'énergie des photons reconstruits.

Ceci est effectué en utilisant un monitoring par impulsion laser d'intensité connue et émettant à une longueur d'onde d'environ 440 nm, donc proche du pic d'émission des cristaux, pour suivre l'évolution de la transparence. Des impulsions d'une longueur d'onde plus élevée, 796 nm environ, donc peu affectée par les changements de transparence, sont aussi utilisées pour vérifier la stabilité des autres éléments du sous-détecteur. Sur la figure 3.8, on peut voir que la réponse moyenne des cristaux ECAL au test de transparence par laser diminue effectivement

avec le temps et la luminosité du LHC ; on observe cependant des phases de rémission notamment lors de pauses où le LHC n'est pas en activité ce qui met bien en évidence la présence de défauts non irrémédiables (notamment dus à des interactions électromagnétiques) et le pouvoir de récupération du tungstate de plomb. Notons que cette réponse moyenne peut atteindre des valeurs de 13 % de la transparence nominale dans les hautes valeurs de  $\eta$  mais qu'elle reste quand même au dessus de 90 % dans la majorité du tonneau. De telles mesures sont effectuées toutes les quarante minutes, le but étant à nouveau de pouvoir apporter des corrections aux données mesurées.

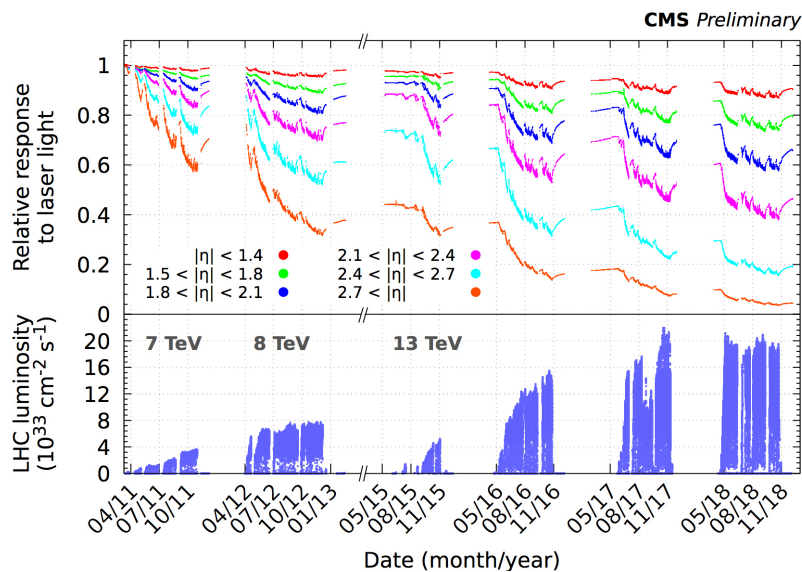


FIGURE 3.8 – Réponse moyenne des cristaux ECAL à de la lumière laser (de longueur d'onde 440 nm pour 2011 et 447 nm pour les années suivantes), mesurée par le système de monitoring laser du ECAL, et luminosité du LHC en fonction de la date d'acquisition des données jusqu'à la fin du Run 2 [64].

### Photo-détecteurs et électronique de lecture

Deux types de photo-détecteurs, placés à l'arrière des cristaux, sont utilisés dans le ECAL permettant de convertir le signal lumineux issu des photons en signal électrique amplifié : les photodiodes à avalanche (APD) et les phototriodes à vide (VPT).

Les APD sont utilisées, par paire, dans les cristaux du tonneau. De surface active de  $5 \times 5$  mm<sup>2</sup> et composée de silicium, elles fonctionnent selon le principe de l'avalanche électronique : les photons de scintillation vont interagir avec les atomes de silicium, soumis à un important champ électrique, arrachant alors des électrons qui se retrouvent ensuite accélérés et vont donc ioniser d'autres atomes de silicium, et ainsi de suite. Il en résulte un gain important, d'environ 50, pour les APD ce qui permet de mesurer de faibles phénomènes de scintillation.

L'inconvénient principal de ces photodiodes est que leur gain est fortement dépendant de la température, avec une perte enregistrée d'environ 2,2 %/°C à 18 °C, et de la tension appliquée, avec une variation d'environ 3,3 %/V nécessitant un contrôle précis de ces deux paramètres.

Notamment, dans le tonneau la température est contrôlée à  $\pm 0,05$  °C [65] autour de 18 °C par des systèmes de refroidissement utilisant notamment de l'eau à 18 °C et de la mousse isolante.

Les radiations sont trop importantes pour utiliser des APD dans les bouchons, on utilise donc des VPT qui sont bien plus résistantes aux radiations. Leur surface active est de 280 mm<sup>2</sup> et leur gain est inférieur à 10. Les VPT fonctionnent aussi selon un principe d'avalanches successives entre trois électrodes.

Le signal obtenu est ensuite amplifié et mis en forme par des composants électroniques de lecture. Une partie de ces composants est embarquée à l'arrière des cristaux et l'autre partie est située dans la salle de contrôle de CMS.

La partie embarquée est principalement composée de deux types de cartes. Une première, fonctionnant sur la sortie de 4 cristaux, qui permet d'amplifier le signal selon un gain adapté et de le numériser. Elle contient aussi des modules de mesure de la température et de détection des courants dits d'obscurité correspondant au courant résiduel observé dans les photodiodes en absence de luminosité.

Le second type de carte, fonctionnant sur la sortie de 5 des cartes précédentes, calcule notamment les primitives de déclenchement, voir la section 3.1.1, dans le ECAL et stocke les données enregistrées dans les cristaux dans une mémoire tampon. Ces informations sont ensuite envoyées dans des cartes dédiées dans la salle de contrôle [63].

#### Description du ECAL

Un schéma présentant les différents éléments constituant le calorimètre électromagnétique est visible sur la figure 3.9.

Les 61200 cristaux du tonneau forment un volume de près de 8,14 m<sup>3</sup> et sont contenus dans des structures en nid d'abeille appelées sous-modules. Chaque sous-module contient dix cristaux. Les sous-modules sont assemblés au sein de modules contenant de 400 à 500 cristaux selon leur distance au centre. Un ensemble de quatre modules est ensuite assemblé au sein d'un super-module contenant alors 1700 cristaux. Ces cristaux sont positionnés à un angle de près de 3° en  $\phi$  et  $\eta$  par rapport au point d'interaction afin de minimiser la probabilité qu'un électron ou un photon tombe dans un des interstices, de quelques millimètres, présents notamment entre les différents cristaux, sous-modules, modules et super-modules et échappe alors à la détection ainsi que de minimiser la probabilité pour ces mêmes particules d'interagir avec les autres sous-composantes du ECAL (c'est encore la notion de budget matériel). A noter qu'il existe tout de même des pertes d'énergie à cause de ces interstices (où l'énergie n'est alors pas mesurable). Nous verrons dans le chapitre suivant comment tenir compte de cet effet.

Les 7324 cristaux présents dans les bouchons sont, à nouveau, regroupés dans des structures alvéolaires appelés super-cristaux. Notamment, il est commun de séparer chaque bouchon en deux parties selon la verticale (l'axe  $y$ ) appelées "Dee". Chaque Dee est lui-même composé de 138 super-cristaux ainsi que de 18 super-cristaux spéciaux situés au centre et aux extrémités des

bouchons. A nouveau, les cristaux ne sont pas orientés vers le point d'interaction mais pointent vers un point situé à 1 300 mm de ce dernier, ce qui donne des angles d'inclinaison entre 2 et 8 °.

Un détecteur à pied de gerbe (ES ou Preshower) est aussi présent dans chaque bouchon. Son rôle est de permettre d'identifier les pions neutres  $\pi^0$ , d'aider à l'identification des électrons et d'améliorer la détermination des caractéristiques des électrons et des photons grâce à sa très fine granularité avec des cellules de 2 mm de largeur et près de 137000 pistes de lecture au total. Avec une circonférence de 2,5 m, ils couvrent les régions  $1,653 < |\eta| < 2,6$  et sont constitués de couches de plomb suivies (vers l'extérieur du détecteur) de couches de silicium pour une épaisseur totale de 20 cm. Les couches de plomb permettent d'initier des gerbes d'interactions électromagnétiques alors que les couches de silicium sont en fait des détecteurs semblables à ceux du trajectographe. Ces sous-détecteurs jouent un rôle dans la reconstruction des photons. En effet, les pions neutres sont une source importante de bruit de fond pour les photons puisqu'ils se désintègrent facilement (temps de vie de l'ordre de  $10^{-16}$  s) en deux photons émis de façon presque collinéaires ce qui ressemble fortement, dans les détecteurs, à l'émission d'un seul photon de haute énergie donc provenant d'une interaction d'intérêt. La granularité du ES n'est cependant pas assez fine pour complètement séparer les deux photons et rejeter efficacement les pions neutres.

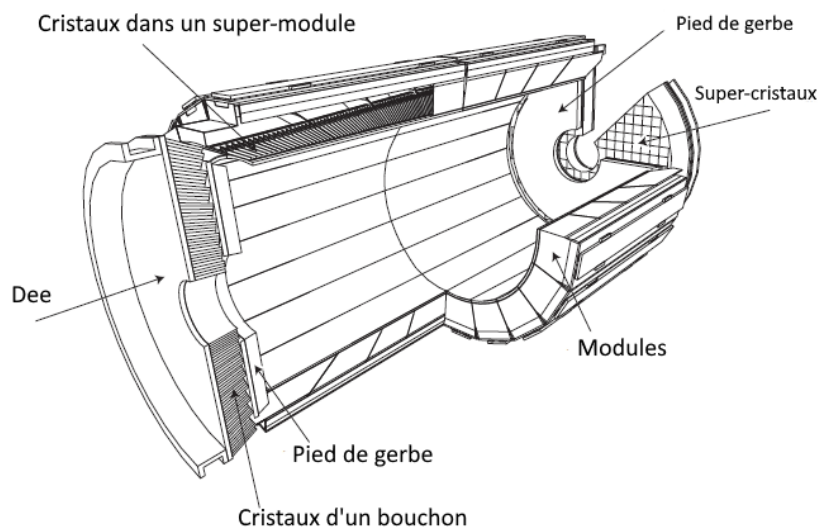


FIGURE 3.9 – Schéma en coupe du calorimètre électromagnétique de CMS présentant les principaux éléments le constituant [63].

A nouveau, il est nécessaire de prendre en compte les interactions avec les composants des services assurant le fonctionnement, le refroidissement et toutes les connectiques du ECAL qui n'ont pas été décrites ici. Ce budget matériel est de l'ordre de 0,4 fois la longueur de radiation des matériaux considérés près de  $|\eta| = 0$  et va jusqu'à des valeurs de deux fois cette longueur d'onde dans des régions proches de  $|\eta| = 1,4$  [66].

### Résolution en énergie du ECAL

La résolution en énergie du ECAL est une grandeur caractéristique de sa performance puisqu'elle est directement liée à la précision qu'il est possible d'atteindre sur les énergies des photons et électrons reconstruits, ainsi que sur les masses invariantes faisant intervenir ces photons et électrons. Elle est en générale de l'ordre de 1 % dans le domaine d'énergie qui nous intéresse et peut être modélisée selon la formule suivante pour des énergies inférieures à 500 GeV [67] :

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E(\text{GeV})}}\right)^2 + \left(\frac{B}{E(\text{GeV})}\right)^2 + C^2 \quad (3.5)$$

avec  $\sigma_E$  la résolution en énergie,  $E$  l'énergie considérée,  $S$  un terme stochastique,  $B$  un terme de bruit et  $C$  un terme constant. Les principales contributions au terme stochastique sont dues à des fluctuations, en terme de nombre d'évènements, sur l'étalement latéral des gerbes électromagnétiques, des fluctuations en termes de nombre de photons émis par les photo-diodes et des fluctuations dans l'énergie déposée au sein des couches absorbantes des détecteurs à pied de gerbe par rapport à celle mesurée dans la couche de détection. Le terme de bruit est composé principalement de termes provenant du bruit de l'électronique et du système de numérisation ainsi que la prise en compte de l'empilement. Enfin le terme constant modélise la non-uniformité de la réponse lumineuse des cristaux et les pertes d'énergie à l'arrière des cristaux.

Cette résolution en énergie a été mesurée en 2004 [67] dans le ECAL seul en utilisant des électrons d'énergie variant entre 20 et 250 GeV dans un faisceau test reconstruits dans une matrice de 3×3 cristaux. Comme on peut l'observer sur la figure 3.10, cette résolution varie entre 0,3 et 1,5 %. Ces valeurs sont cependant obtenues dans des conditions idéales, loin de celles correspondant aux cristaux ECAL en fonctionnement dans CMS.

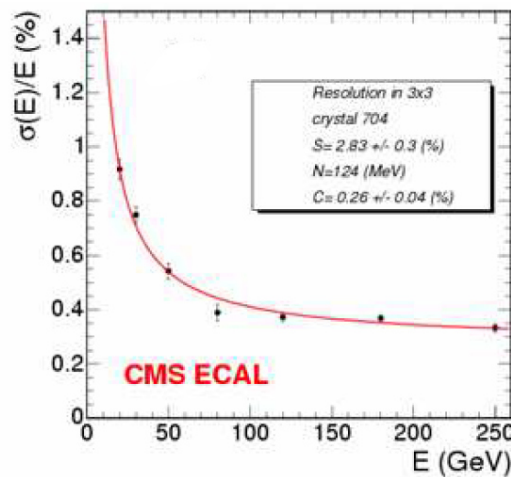


FIGURE 3.10 – Résolution en énergie en fonction de l'énergie pour des électrons [67].

### Le calorimètre hadronique

Le calorimètre hadronique (HCAL) a pour but de stopper et mesurer les particules hadroniques qui ont traversé le ECAL, c'est-à-dire des produits d'interactions fortes mettant en jeu des quarks et des gluons tels les protons, les neutrons ou des jets qui sont des gerbes dues aux



propriétés de confinement des quarks et des gluons ne pouvant être observés de manière isolés aux échelles d'énergie actuelles du LHC.

Le HCAL est un détecteur dit "à échantillonnage". Il permet en effet de mesurer le moment et l'énergie des particules grâce à une série alternée de plaques d'absorbant ayant pour rôle de ralentir les particules et de créer un début de gerbes, et de scintillateurs qui émettent des photons au passage et interaction de particules. A nouveau, la lumière émise par ces scintillateurs est détectée par des photo-diodes et son intensité est directement liée à l'énergie des particules interagissant ; on a donc accès à l'énergie des particules en présence en sommant ces échantillons de mesure.

Un des rôles du HCAL est aussi de mesurer indirectement l'énergie des particules n'interagissant que très peu avec le détecteur comme les neutrinos. Pour ce faire il faut que le HCAL puisse mesurer l'énergie de toutes les particules autres que les électrons et photons mesurés dans le ECAL, et les muons (mesurés dans les chambres à muons), relativement à leur force d'interaction, venant des collisions proton-proton originelles, afin de pouvoir en déduire l'énergie manquante transverse (MET). Le HCAL doit alors être au maximum hermétique, avec une bonne couverture spatiale et angulaire tout en minimisant les interstices entre sous-détecteurs.

Les gerbes hadroniques peuvent être très larges et s'étaler sur un mètre ce qui rend difficile de satisfaire la contrainte de compacité du détecteur CMS. Ainsi le HCAL est séparé en quatre parties : le tonneau interne (HB) accompagné de ses bouchons (HE), un tonneau externe (HO) et un calorimètre vers l'avant (HF) tous deux situés hors du solénoïde de CMS comme on peut le voir sur la figure 3.11.

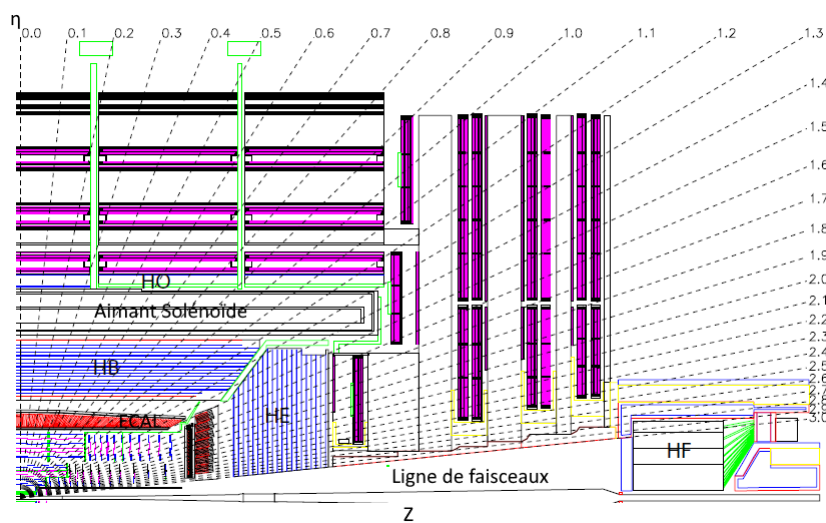


FIGURE 3.11 – Schéma en coupe représentant un quart du calorimètre hadronique de CMS. La partie interne (par rapport à l'aimant) du HCAL est représentée en bleu (HB et HE). Au dessus de l'aimant se trouve le tonneau externe HO. Le calorimètre vers l'avant (HF) est lui situé en aval de CMS [63].



Le tonneau interne est contenu dans la partie d'environ  $1,77 \text{ m} < r < 2,95 \text{ m}$ , soit entre le ECAL et l'aimant solénoïde avec une couverture angulaire de  $0 < |\eta| < 1,3$  environ. Les plaques d'absorbant sont constituées de quinze plaques de laiton (mélange de cuivre et de zinc) et deux plaques en acier inoxydable (pour la première et la dernière plaque) allant de 4 à 7,5 cm d'épaisseur. Entre ces plaques sont insérées des plaques scintillatrices en plastique complexe [63] de 3,7 mm (9 mm pour celles aux extrémités du tonneau).

Les bouchons internes couvrent les zones d'environ  $1,3 < |\eta| < 3$ , ils ont été construits de façon à minimiser les interstices entre sous-détecteurs ce qui impose une géométrie différente de celle du tonneau. Les différentes couches sont semblables à celles du tonneau interne mais elles sont plus épaisses (7,9 cm d'épaisseur pour les couches de laiton) afin de couvrir l'intégralité des gerbes hadroniques.

Pour compléter la couverture angulaire, des calorimètres vers l'avant sont situés à chaque extrémité du détecteur CMS comme on peut l'observer sur la figure 3.1. Ils ont un rayon interne de 1,3 m et sont situés à  $z = \pm 11,2 \text{ m}$ . Leur rôle est de détecter les particules ayant un  $\eta$  élevé, typiquement  $3 < |\eta| < 5$ . Ces zones étant soumises à de nombreuses interactions, les HF se doivent d'être très résistants et sont donc composés de matériaux différents des autres parties du HCAL. Les plaques d'absorbant sont ainsi en acier alors que les plaques actives sont constituées de fibres de quartz fonctionnant par effet Tcherenkov (les particules se déplacent à une vitesse supérieure à celle de la lumière dans le matériau considéré créant une interférence constructive de la lumière émise). Le HF est aussi l'un des luminomètres de CMS qui permet de mesurer avec précision la luminosité intégrée de CMS, voir [16] pour plus de détails.

Finalement pour garantir le caractère hermétique du HCAL, le tonneau externe situé après l'aimant solénoïde vient compléter le tonneau interne pour permettre de mesurer les gerbes hadroniques très énergétiques qui auraient traversé le HB et l'aimant. Il utilise comme absorbant l'acier présent pour garantir le retour du champ magnétique au sein de l'aimant.

#### Les chambres à muons

La dernière couche du détecteur CMS est composée de juxtapositions de chambres à muons insérées dans la culasse en acier assurant le retour du champ magnétique. Les chambres à muons couvrent une surface de détection de  $25\,000 \text{ m}^2$  et, comme l'indique leur nom, permettent de détecter les muons faisant partie de la catégorie des particules qui ne sont pas stoppées par le ECAL et le HCAL. L'identification et la mesure du moment des muons ont un rôle essentiel dans CMS, jusqu'à être présent dans le nom du détecteur, puisqu'ils peuvent être la signature visible de phénomènes de nouvelle physique.

Cette détection s'effectue grâce à l'utilisation de sous-détecteurs gazeux répartis concentriquement afin d'identifier les muons. En effet, au passage des particules chargées que sont les muons, un phénomène d'ionisation va se produire dans les gaz qui sera ensuite transcrit en signal électrique grâce à la présence d'une différence de potentiel au sein du sous-détecteur

considéré. A nouveau, selon la position spatio-temporelle du dépôt d'énergie et grâce au champ magnétique courbant les trajectoires, on peut remonter, en utilisant en plus les informations du trajectographe, à l'énergie et l'impulsion des muons.

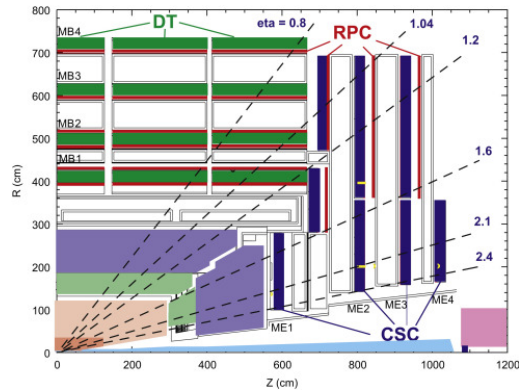


FIGURE 3.12 – Schéma en coupe représentant un quart du détecteur CMS et ses chambres à muons. Les tubes à dérives sont visibles en vert foncé, les chambres à pistes cathodiques en bleu et les chambres à plaques résistives en rouge. Sont aussi visibles : le HCAL en violet, le ECAL en vert clair, le trajectographe en beige et le calorimètre vers l'avant en rose [68].

Il existe environ 1850 chambres à muons réparties en trois types de sous-détecteur gazeux [69] : environ 250 tubes à dérives (DT), 1056 chambres à pistes cathodiques (CSC) et 540 chambres à plaques résistives (RPC). Les DT sont en aluminium et remplis de gaz inertes. Ils couvrent la partie tonneau du détecteur CMS tel que  $|\eta| < 1,2$  comme on peut l'observer sur la figure 3.12, et permettent de mesurer le passage des muons avec une excellente résolution temporelle ( $\sim 5$  ns). Les CSC sont, elles, présentes dans les bouchons et plus globalement dans les régions  $0,9 < |\eta| < 2,3$  où le champ magnétique est non-uniforme et où le flux de particules est élevé. Elles sont composées de réseaux d'anodes entremêlés de pistes cathodiques en cuivre dans un volume de gaz inerte. Elles ont ainsi été construites pour offrir une bonne résistance aux radiations, une bonne résolution temporelle et spatiale permettant de différencier efficacement les muons d'autres particules (comme les neutrons).

Les RPC viennent épauler les DT et CSC dans leur rôle dans le système de déclenchement. Elles sont composées de deux plaques, anode et cathode, parallèles en plastique complexe très résistant séparées par du gaz inerte. Elles offrent la meilleure résolution temporelle ( $\sim 1$  ns) et permettent un déclenchement indépendant et parallèle, vis-à-vis des autres types de chambres, permettant de rejeter efficacement les évènements non désirés.

### Le système de déclenchement

Effectivement, sachant que LHC envoie des paquets de protons toutes les 25 ns, correspondant à une fréquence de 40 MHz, dans le détecteur, que l'empilement des évènements<sup>5</sup> est d'environ<sup>6</sup> 20, et que l'enregistrement d'un de ces évènements bruts fait environ 1 MB, on

5. Un évènement correspond à un vertex d'interaction

6. C'est-à-dire qu'il y a environ 20 collisions proton-proton en même temps. Pour le Run 2, l'empilement est plutôt de l'ordre de 50.

aurait, en l'absence de sélection des événements, un taux d'enregistrement aux alentours  $800 \text{ TB}\cdot\text{s}^{-1}$  dans le détecteur CMS ce qui est, avec les technologies actuelles, impossible à réaliser et à stocker. Par conséquent, il est nécessaire pour CMS de posséder un système de déclenchement (trigger) rapide, avec une faible latence, permettant de sélectionner les événements physiques selon s'ils sont considérés comme intéressants ou non.

Cette sélection est réalisée par le système de déclenchement composé de deux étapes successives : le déclenchement de niveau 1 ou Level-1 Trigger (L1) fonctionnant surtout avec de l'électronique au sein du détecteur, suivi par le système de déclenchement de haut niveau ou High-Level Trigger (HLT) basé plutôt sur des reconstructions utilisant des logiciels informatiques. Le L1 permet de réduire le taux d'évènement total à un ordre de  $100 \text{ kHz}$ , et le HLT permet d'obtenir ensuite un taux de quelques centaines de Hz, maintenant enregistrable et stockable.

Le L1 est composé de deux chaînes de déclenchement logiques comme on peut le voir sur la figure 3.13 basées sur les muons et sur les informations des calorimètres obtenues à un niveau local puis régional et enfin global. Elles sont donc chacune construites grâce à des informations locales, les générateurs de primitives de déclenchement (TPG), récupérées directement dans les sous-détecteurs comme les énergies déposées dans différentes parties des calorimètres, pour la chaîne calorimètre, ou des coups dans les chambres à muons, pour la chaîne muonique.

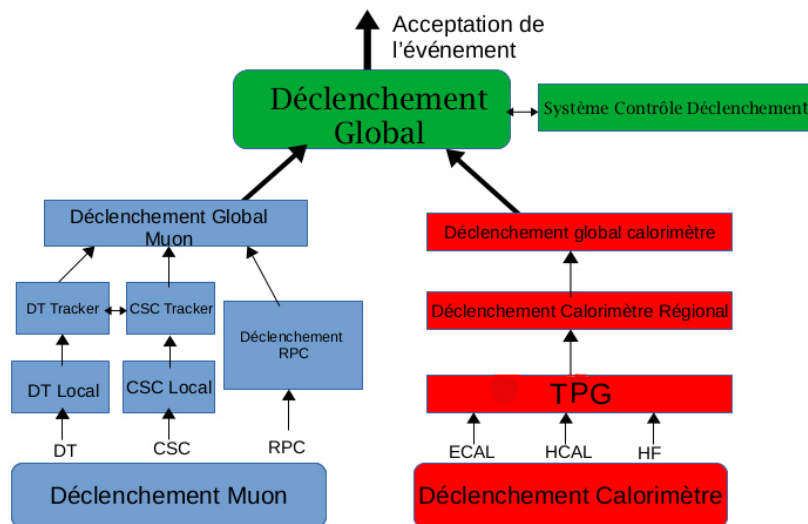


FIGURE 3.13 – Schéma de fonctionnement du déclenchement de niveau 1. En bleu la partie basée sur le déclenchement des muons, en rouge celle basée sur les informations des différents calorimètres. En vert le système de déclenchement global qui prend la décision finale.

Des informations régionales permettent ensuite, par combinaison des informations obtenues, de reconstruire simplement des électrons, photons, muons selon des critères de position, d'isolation et en fonction de l'énergie, du moment ou de la qualité des informations obtenues, ce qui permet d'établir un classement des objets reconstruits.

Des reconstructions globales des objets (jets, énergie manquante, photons isolés, muons isolés, ...) sont finalement obtenues notamment grâce à des algorithmes embarqués avant de les transmettre au système de déclenchement général ; pour la partie muonique les différences entre les informations obtenues par les DT et CSC, et les RPC permettent de supprimer une partie du bruit de fond et d'augmenter l'efficacité du système de déclenchement [63].

C'est ce système de déclenchement général, ou global, qui prend la décision finale d'accepter ou de rejeter l'évènement considéré en utilisant des algorithmes et informations des sous-détecteurs spécifiques. Les algorithmes sont nommés "candidat L1" et utilisent des critères sur un ou plusieurs objets reconstruits selon leur énergie ou encore leur moment transverse et sont regroupés en "menu" ; ainsi si un évènement passe la sélection d'au moins un candidat L1 d'un menu, il est accepté. Le temps de latence du système de déclenchement, entre une collision d'un paquet proton-proton considéré et la prise de décision, est d'environ  $3 \mu\text{s}$ , les informations étant stockées en mémoire tampon pendant ce temps.

Les évènements acceptés par le système de déclenchement général du L1 sont ensuite envoyés en entrée du HLT qui, contrairement au L1, n'utilise que des infrastructures informatiques. Le HLT correspond en effet à un ferme de processeurs situés à côté de la salle de contrôle de CMS. En utilisant les informations des calorimètres, des chambres à muons, mais aussi du trajectographe, les algorithmes du HLT reconstruisent des objets physiques complexes qui doivent ensuite satisfaire à différents critères pour que les évènements correspondants puissent être finalement acceptés, ou non, et être stockés dans des serveurs dédiés. Les reconstructions au niveau du HLT sont plus sophistiquées que celle utilisées dans le L1 mais, puisque le système de déclenchement doit agir rapidement (environ 150 ms par évènement pour le HLT), elles sont moins détaillées que les reconstructions finales utilisées dans les analyses et précisées ci-après.

Les conditions utilisées pour sélectionner les évènements reconstruits (coupure en moment transverse, coupure en énergie, ou encore coupure en isolation) définissent un "chemin" de déclenchement. Les évènements acceptés par le HLT sont donc catégorisés en fonction de leur chemin de déclenchement. Un évènement peut donc être associé à plusieurs chemins cependant il est essentiel de choisir ces chemins avec réflexion puisque un évènement qui ne satisfait aucun des chemins désignés ne serait tout simplement pas enregistré et on pourrait alors louper des évènements appartenant à de la nouvelle physique.

### 3.2 Reconstruction et simulation des évènements physiques

Nous avons vu que les différentes particules issues des collisions proton-proton interagissent avec l'ensemble des sous-détecteurs de CMS, ces interactions étant ensuite transformées en signaux électroniques. Il est alors nécessaire, en vue d'études physiques, de pouvoir reconstruire avec confiance ces signaux et identifier les particules auxquelles ils correspondent. Ceci est effectué par des algorithmes avancés dont la base est l'algorithme de Particle Flow (PF).

### 3.2.1 Algorithme de Particule Flow

L'algorithme de PF permet ainsi une reconstruction des hadrons neutres et chargés, des électrons, des photons, des muons en corrélant les différents signaux mesurés dans les différents sous-détecteurs, en associant par exemple une trace dans le trajectographe à un dépôt d'énergie dans le ECAL. Il a été construit de manière à obtenir un taux d'efficacité de reconstruction très élevé ainsi qu'un taux de mauvaises identifications faible [70]. L'algorithme commence par chercher et construire des éléments PF au niveau des sous-détecteurs, comme des traces dans le trajectographe ou les chambres à muons, ou des agrégats de dépôts d'énergie dans les calorimètres ; il tisse ensuite des liens entre ces éléments PF pour isoler les particules et les identifier avant d'appliquer des corrections dues notamment à l'empilement dans le détecteur.

#### Les éléments Particle Flow

Pour reconstruire les traces, éléments de la trajectoire des particules chargées, on utilise un algorithme d'identification combinatoire basé sur un filtre de Kalman (KF), voir [71] pour plus de détails, fonctionnant en trois étapes : on commence par la génération d'une "graine" c'est-à-dire plusieurs coups dans un sous-détecteur compatibles avec la trajectoire d'une particule chargée définissant les paramètres initiaux de la trajectoire. On construit ensuite la trajectoire en rassemblant tous les coups dans les couches du trajectographe proches de celle de la graine. Un ajustement est finalement réalisé pour déterminer les propriétés (origine, direction et impulsion transverse) de la particule chargée considérée. Pour que la trajectoire reconstruite soit acceptée dans les analyses suivantes, il faut que la graine de la trajectoire soit composée de deux coups dans des couches consécutives du détecteur à pixels, que la reconstruction totale possède au moins 8 coups avec au plus un coup manquant sur la trajectoire totale, et que leur origine soit située dans un cylindre de quelques mm de rayon centré sur l'axe z avec un moment transverse supérieur à 0,9 GeV afin de garantir un taux faible de mauvaises reconstructions [72].

Cet algorithme est appliqué de manière itérative (une dizaine de fois) en masquant les coups associés à la trajectoire finale acceptée afin de permettre aux coups restants d'être utilisés pour former de nouvelles graines et traces ; ce faisant l'efficacité de la reconstruction des trajectoires est augmenté et le taux de mauvaises reconstructions reste fixe.

Les trajectoires ainsi reconstruites dans le trajectographe peuvent ensuite être agrégées en vertex [73] représentant les différents points d'interaction. On distingue les vertex primaires durs, dus aux interactions proton-proton d'intérêt, des vertex d'empilement. Ces vertex sont reconstruits en sélectionnant les traces ayant la même origine et satisfaisant différents critères de qualité d'ajustement [73].

Pour la reconstruction des éléments PF dans les calorimètres, un algorithme d'agrégation a été développé dans le but de détecter et mesurer l'énergie et la direction des particules neutres, de séparer les dépôts correspondant à ces particules neutres de ceux correspondant à des particules chargées, de reconstruire et identifier les électrons et les photons qu'ils rayonnent par effet bremsstrahlung et aussi de mesurer l'énergie des hadrons chargés, dont les informations sur la trajectoire seraient insuffisantes, tout en permettant une forte efficacité de détection et de

TABLE 3.1 – Paramètres d’agrégation pour les différents calorimètres. Ces valeurs sont le résultat de simulation utilisant des photons, des pions neutres, des kaons neutres et des jets [70].

	ECAL		HCAL		Preshower
	Tonneau	Bouchons	Tonneau	Bouchons	
Coupure $E$ agrégat topologique (MeV)	80	300	800	800	0,06
Nombre de voisins pour la graine	8	8	4	4	8
Coupure $E$ graine (MeV)	230	600	800	1100	0,12
Coupure $E_T$ graine (MeV)	0	150	0	0	0
Largeur de la gaussienne (cm)	1,5	1,5	10,0	10,0	0,2

séparer des dépôts d’énergies proches. Pour cela, les dépôts sont agrégés dans chaque détecteur séparément en commençant à nouveau par déterminer les graines des agrégats c’est-à-dire les cellules, les cristaux, des calorimètres ayant une énergie locale plus élevée que les cellules avoisinantes et supérieure à une certaine valeur fixée dépendant du calorimètre, comme visible dans le tableau 3.1. Ensuite, les cellules sont agrégées autour de la graine si elles sont au moins en contact avec une cellule déjà présente dans l’agrégat et si l’énergie déposée y est supérieure à deux fois l’énergie du bruit ambiant afin de former des agrégats topologiques. L’agrégat est finalement reconstruit à partir des agrégats topologiques grâce à un modèle de mixture de gaussiennes permettant de décrire les dépôts d’énergie selon le nombre de graines dans l’agrégat topologique considéré [74].

### Algorithme de lien

Une particule donnée n’interagit à priori pas avec un unique sous-détecteur de CMS mais avec plusieurs d’entre eux, ainsi une particule est en générale associée à plusieurs éléments PF. Pour reconstruire la particule en question il faut donc être capable de relier tous ses éléments PF entre eux : c’est le rôle de l’algorithme de lien. Un lien entre les éléments PF est établi selon des critères propres à chaque élément PF, si deux éléments PF sont liés l’algorithme définit une distance entre eux qui quantifie la qualité du lien. Les éléments PF ainsi associés par des liens sont regroupés au sein de blocs PF. Ainsi un lien est créé entre une trace dans le trajectographe et un agrégat calorimétrique si, en extrapolant cette trace à partir du dernier coup mesuré jusqu’à deux couches/la longueur typique d’une gerbe électromagnétique/une longueur d’interaction pour le preshower/ECAL/HCAL, son extrapolation correspond à la position d’un agrégat calorimétrique. Des liens sont aussi créés entre une trace dans le trajectographe et des informations dans les chambres à muons, entre différents agrégats calorimétriques, entre les traces d’une particule chargée et un vertex secondaire, et ils sont essentiels pour récupérer l’énergie perdue par un électron par émission de photon par effet bremsstrahlung.

### Identification des particules

On peut alors identifier les particules dans chaque bloc PF en fonction des liens établis entre les différents éléments PF et de leur qualité. L’identification se fait dans un ordre précis : on commence par identifier les muons puis les électrons et les photons isolés avant d’identifier les hadrons et les photons non isolés et de finir par les particules interagissant par interaction nu-



cléaire dans le trajectographe. En général les éléments PF utilisés pour identifier une particule sont masqués dans la suite de l'algorithme.

On distingue trois différents types de **muons** reconstruits : les muons *standalone* qui possèdent uniquement des traces dans les détecteurs des chambres à muons, les muons *tracker* qui possèdent des traces dans le trajectographe de CMS extrapolées jusqu'aux chambres à muons à la recherche de segment muonique, et les muons globaux composés de muons *standalone* dont les trajectoires extrapolées sont compatibles avec des traces dans le trajectographe de CMS. On identifie ensuite les muons isolés grâce à des critères sur l'isolation, l'énergie et le moment transverse. Les muons ne satisfaisant pas les critères d'isolation doivent satisfaire des critères de qualité de trajectoire ou une sélection sur les muons resserrée [75]. L'entièreté des muons n'est pas forcément reconstruite à cette étape puisqu'on peut identifier des muons parmi les candidats hadrons chargés.

Les **électrons** peuvent perdre, surtout dans le trajectographe, une grande quantité d'énergie en émettant des photons par effet bremsstrahlung et les **photons** se convertissent souvent en une paire électron positron qui, à leur tour, peuvent émettre du rayonnement bremsstrahlung. Il est ainsi nécessaire, pour démêler les traces et agrégats associés à ces particules, de reconstruire ensemble électrons et photons isolés. Pour couvrir l'intégralité de l'énergie des électrons et donc du rayonnement bremsstrahlung, ainsi que la courbure de la trajectoire plus importante due à la perte d'énergie, il est nécessaire de regrouper les agrégats dans le ECAL dans la direction du candidat électron considéré en  $\eta$  et  $\phi$  en un super-agrégat. Pour prendre en compte ce genre de perte d'énergie le long de la trajectoire, le filtre de Kalman utilisé pour la reconstruction des traces et remplacé par un filtre de somme de gaussienne (GSF) [76]. Pour être identifiés en tant que candidats électrons et photons du ECAL, il faut que l'énergie contenue dans les cellules du HCAL à une distance angulaire  $\Delta R < 0,15$  du super-agrégat considéré soit plus petite que 10 % de l'énergie de ce super-agrégat. Pour les photons, le super-agrégat associé ne doit pas avoir de lien avec une trace GSF et doit avoir une énergie transverse de plus de 10 GeV. Pour les électrons, l'agrégat associé dans le ECAL ne doit pas être lié à plus de trois traces et ils doivent aussi satisfaire un critère de sélection basé sur la sortie d'arbres de décisions boostés<sup>7</sup>. Les critères de reconstruction permettant de distinguer électron et photon variant en fonction de l'analyse, l'historique de ces reconstructions est sauvegardé ce qui permet un ajustement facile.

Les éléments PF encore non masqués dans un bloc PF à cette étape sont utilisés pour identifier des **hadrons chargés** (typiquement  $\pi^\pm$ ,  $K^\pm$  ou des protons), **neutres** ( $K_L^0$  ou des neutrons), des **photons non isolés** (par exemple provenant de désintégration de pions neutres) et potentiellement des **muons** provenant de désintégrations d'hadrons chargés. Les agrégats calorimétriques qui n'ont aucun lien avec des traces tout en étant angulairement couverts par le trajectographe (soit  $|\eta| < 2,5$ ) sont identifiés comme des photons s'ils sont dans le ECAL et comme des hadrons neutres s'ils sont dans le HCAL. Les agrégats ECAL liés à des agrégats HCAL sont associés à un même hadron.

---

7. Dans la suite, nous verrons plus en détail en quoi consiste une telle technique d'analyse multivariée

Les énergies des particules identifiées sont ensuite estimées et calibrées, en utilisant notamment des simulations, afin de déterminer l'énergie calorimétrique totale. Les agrégats calorimétriques restant sont ensuite associés à des traces dont on calcule la somme des moments, sous l'hypothèse que ces traces sont dues à des hadrons chargés, afin de la comparer à l'énergie calorimétrique totale dans le but d'identifier la présence d'éventuels hadrons neutres ou photons (si l'énergie calorimétrique totale est plus importante), voire des muons (dans le cas où cette même énergie est plus faible).

Il reste ensuite à réduire le taux de mauvaises identifications en étudiant les particules identifiées ayant un moment transverse manquant important (souvent des muons mal reconstruits). Le moment transverse manquant, défini comme l'opposé de la somme vectorielle des moments transverses de toutes les particules reconstruites, permet de mettre en évidence la présence de particules n'interagissant quasiment pas avec le détecteur tels les neutrinos. La reconstruction des leptons  $\tau$ , dont l'algorithme non abordé ici est expliqué dans la référence [77], et des jets, par un algorithme anti- $k_T$  [78], viennent finalement conclure ces étapes de reconstruction et d'identification. A noter qu'il est nécessaire de prendre en compte la multiplicité des collisions proton-proton, l'empilement, notamment pour ces derniers éléments.

### 3.2.2 Simulation du détecteur

Il est essentiel de pouvoir simuler précisément les processus physiques ayant lieu au sein du détecteur CMS afin de pouvoir comparer les connaissances théoriques aux données expérimentales. De telles simulations, basées sur des modèles théoriques éprouvés et des reconstructions détaillées du détecteur CMS, ont un rôle crucial dans les analyses.

Pour obtenir ces évènements simulés, on commence par des calculs de quadri-vecteurs associés aux interactions physiques que l'on souhaite simuler avec des logiciels dits générateurs comme MadGraph [79], qui utilise les règles de Feynman pour des phénomènes partoniques, ou PYTHIA [80] qui permet justement de construire des gerbes hadroniques. Ces quadri-vecteurs sont ensuite envoyés, avec une description détaillée des sous-détecteurs et leur géométrie, comme entrées dans GEANT4 [81], un logiciel basé sur une méthode Monte-Carlo qui va simuler l'interaction des particules produites par les générateurs avec les sous-détecteurs de CMS.

On requiert ainsi pour le trajectographe que ses composantes actives et passives, ainsi que le rayonnement bremsstrahlung des électrons soient décrits avec grande précision. Pour le ECAL, on insiste sur la précision des descriptions de la géométrie et du budget matériel ainsi que la bonne implémentation des processus électromagnétiques, ce qui permet une bonne compréhension des gerbes électromagnétiques prenant en compte les spécificités du détecteur (interstices entre les différents amas de cristaux, ou encore les différentes calibrations). Pour le HCAL il est important d'avoir une description précise du bruit de fond ambiant, des temps de réponse et de la réponse en énergie du calorimètre. Pour les chambres à muons, une description physique des phénomènes utilisant des muons ( $\mu$  bremsstrahlung ou encore diffusion multiple  $\mu$ ) est nécessaire. Tout cela implique donc au préalable de connaître précisément la géomé-



trie et le fonctionnement de tous les sous-détecteurs ainsi que leurs composantes et autres équipements associés. Toutes ces simulations sont vérifiées, améliorées et validées en utilisant d'autres simulations ou des données réelles non directement utilisées dans les analyses. Pour un meilleur accord aux données, l'empilement est aussi simulé avec PYTHIA en utilisant un profil de luminosité intégrée donné et la section efficace inélastique totale [82].

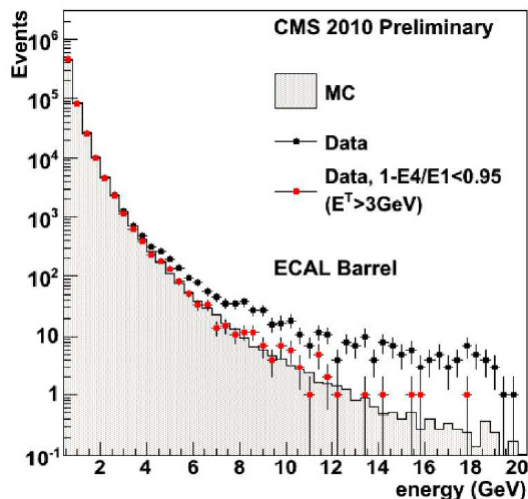


FIGURE 3.14 – Distribution de l'énergie déposée dans les cristaux du tonneau du calorimètre électromagnétique. En gris la distribution simulée par méthode Monte-Carlo que l'on compare aux données (points noirs). L'accord entre les simulations et les données devient bon lorsque l'on applique un filtre spécifique aux données (rouge) [82].

Les bénéfices de l'utilisation de simulations sont nombreux. Notamment les simulations peuvent permettre de mieux comprendre ce qu'il se passe au niveau des détecteurs. On peut, par exemple, observer sur la figure 3.14 un désaccord dans le ECAL entre les données et les simulations, la volonté de comprendre ce désaccord a mené à la découverte de coups anormaux dans les cristaux causés en fait par des dépôts d'énergie dans les couches actives des photo-diodes du tonneau ECAL. Les données ont été corrigées et l'accord données/simulation fut amélioré.

## Conclusion

Le détecteur CMS est un détecteur technologiquement très avancé possédant un intense champ magnétique axial délivré par un solénoïde, un trajectographe en silicium, un calorimètre électromagnétique composé de cristaux scintillants, un calorimètre hadronique à échantillonnage ainsi que de chambres à muons. Ses caractéristiques lui permettent de satisfaire les propriétés suivantes [63] :

- Bonne résolution sur le moment et efficacité de reconstruction pour les particules chargées dans le trajectographe. Système de déclenchement efficace.

- Excellente résolution en énergie électromagnétique ainsi qu'en masse invariante diphoton et diélectron. Bonne couverture géométrique du calorimètre électromagnétique. Rejet des pions neutres. Bonne isolation des photons et leptons à hautes luminosités.
- Bonne résolution en énergie transverse manquante et en masse dijet. Bonne herméticité et couverture angulaire, et fine granularité du calorimètre hadronique.
- Excellente identification des muons et résolution en moment des muons pour une large gamme de moments et d'angles. Bonne résolution sur la masse invariante dimuon. Détermination efficace de la charge des muons.

On a pu voir comment les particules interagissaient au sein du détecteur et comment on pouvait, à partir d'informations sur ces différentes interactions, et grâce aux propriétés des sous-détecteurs de CMS, identifier et reconstruire ces particules. L'objectif étant la recherche d'une potentielle nouvelle particule se désintégrant en deux photons, il est nécessaire de s'intéresser plus en détail à savoir comment les photons sont reconstruits dans les analyses de données de la collaboration CMS.

# Reconstruction et identification des photons

## Sommaire

---

4.1	Reconstruction de l'énergie des photons par cristal . . . . .	76
4.1.1	Reconstruction des signaux $A$ dans le calorimètre électromagnétique	76
4.1.2	Réponse et monitoring dans le calorimètre électromagnétique $LC$ . .	77
4.1.3	Calibrations du calorimètre électromagnétique $IC$ . . . . .	78
4.2	Reconstruction des photons au sein de l'algorithme Particle Flow . . . . .	79
4.3	Corrections à l'énergie des photons . . . . .	82
4.3.1	Correction des pertes d'énergie $E_\gamma$ . . . . .	83
4.3.2	Corrections à l'échelle d'énergie des données et dégradation de la résolution en énergie dans les simulations $s_\gamma$ . . . . .	84
	Calcul des échelles d'énergie . . . . .	84
	Rôles de ces corrections pour les analyses $H \rightarrow \gamma\gamma$ . . . . .	87
4.3.3	Validation et performance des corrections . . . . .	88

---

## 4.1 Reconstruction de l'énergie des photons par cristal

L'énergie reconstruite  $E_\gamma$  d'un photon est l'énergie contenue dans un super-agrégat identifié comme candidat photon. Elle s'exprime comme la somme des énergies déposées dans les cristaux du super-agrégat. Ainsi l'énergie déposée dans un cristal  $i$  du ECAL est reconstruite comme [83] :

$$E_{cristal,i} = LC_i \times \beta_i \times IC_i \times G \times A_i \quad (4.1)$$

où  $A$  est l'amplitude d'énergie brut mesurée dans un cristal par le système électronique,  $G$  étant le facteur de conversion de la sortie du système électronique des photodétecteurs permettant d'obtenir l'énergie en GeV qui est multiplié par un terme de correction linéaire  $\beta$  prenant en compte les effets de variation de gain des photodétecteurs, avec  $LC$  un facteur corrigeant les pertes de transparence dans les cristaux et de réponse dans les photo-diodes, et  $IC$  un terme d'intercalibration prenant en compte les différences intrinsèques entre les différents cristaux et photodétecteurs. Une description de quelques uns de ces facteurs sont détaillés dans les sous-sections suivantes.

Des coupures de qualités [83] sont appliquées pour réduire le bruit ambiant et obtenir un signal propre. L'énergie par cristal obtenue est ensuite celle utilisée par l'algorithme de reconstruction PF décrit dans le chapitre précédent.

### 4.1.1 Reconstruction des signaux $A$ dans le calorimètre électromagnétique

Le temps intrinsèque de fonctionnement du système d'acquisition des signaux d'énergie déposée dans le ECAL étant supérieure (il est de l'ordre d'une centaine de nanosecondes) à la fréquence de croisements des faisceaux, il s'étale sur plusieurs croisement de faisceaux et contient alors une partie importante d'évènements d'empilement hors du temps (OOT-PU) provenant des croisements de faisceaux suivants. Cet effet s'est accentué au Run 2 à cause de l'augmentation de la fréquences des collisions proton-proton au LHC ayant lieu toutes les 25 ns à la place de 50 ns.

Il est alors nécessaire d'utiliser un nouvel algorithme par rapport au Run 1 pour reconstruire les signaux ECAL afin de réduire ces effets de chevauchement provoquant une augmentation du bruit de fond et une dégradation de la résolution en énergie. Ainsi la "méthode des poids" [84] utilisée pendant le Run 1 a été remplacée par une méthode d'ajustement de modèle nommée "multifit" [85].

On commence alors par enregistrer dix échantillons de signaux toutes les 25 ns dès que, et si, le L1 accepte l'évènement considéré. On souhaite ensuite ajuster la distribution de ces échantillons par des modèles prédéfinis. Le modèle utilisé pour l'ajustement de la forme du signal est décrit par des distributions binnées de 15 bins de largeur de 25 ns censés refléter la distribution temporelle attendue d'un signal provenant d'énergie déposée dans les cristaux ECAL. La présence de 5 bins supplémentaires (par rapport aux 10 échantillons mesurées) a pour but la description précise des contributions dues aux OOT-PU. Ces modèles sont déterminés séparément sur des jeux de données adaptés, en réalisant un décalage temporel de 25 ns pour les bins supplémentaires [85]. Ils diffèrent selon les cristaux considérés (différence

de forme des signaux mesurés) et aussi du temps global à cause, notamment, de la perte de transparence des cristaux ; ils doivent ainsi être mis à jour fréquemment.

Les amplitudes  $A_i$  de signal d'énergie déposée dans les cristaux sont alors obtenues par un ajustement par les modèles décrits précédemment des dix échantillons mesurés, par minimisation d'une loi  $\chi^2$  prenant en compte le bruit de fond ambiant et ses corrélations avec les échantillons mesurés [85]. Des exemples d'amplitudes obtenues avec des simulations de photons de moment transverse  $p_T=10$  GeV peuvent être observés sur la figure 4.1.

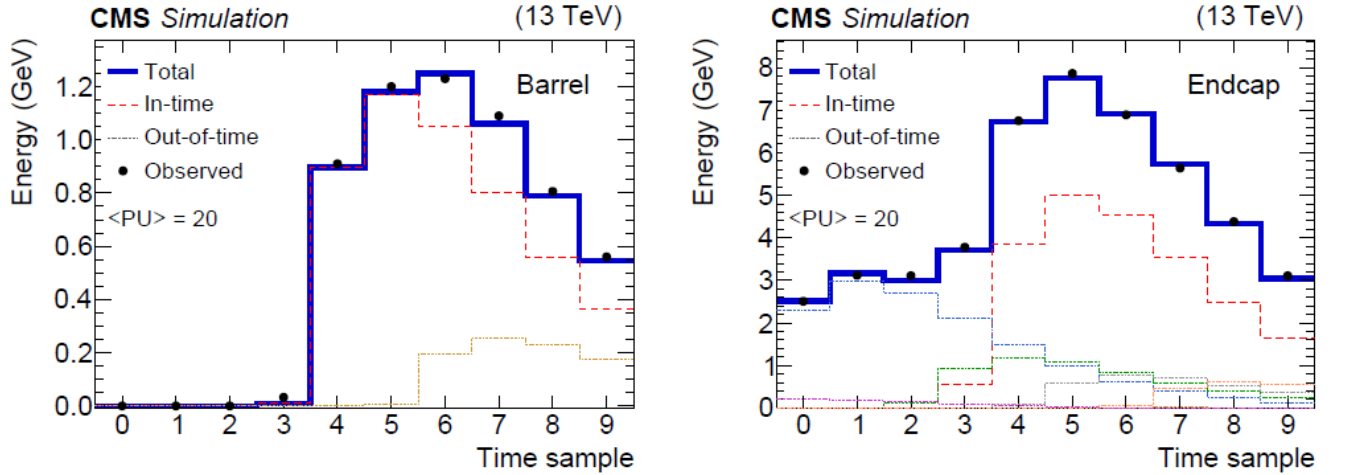


FIGURE 4.1 – Exemples d'amplitudes de signal obtenues par la méthode *multifit* pour des simulations d'une gerbe d'un photon de  $p_T=10$  GeV dans le tonneau (gauche) et dans un bouchon (droite) pour un empilement moyen de 20. Les points noirs représentent les 10 échantillons mesurés, les traits en pointillé rouge/points fins multicolores correspondent aux distributions en temps/hors temps de signal après ajustement. En bleu plein, la somme totale des contributions. Pour les points fins, la couleur, différente de celle de la légende, permet de distinguer les contributions OOT-PU appartenant à différents croisements de faisceaux.

#### 4.1.2 Réponse et monitoring dans le calorimètre électromagnétique LC

Comme entraperçu dans la section 3.1.1, la réponse en énergie du ECAL varie en fonction du temps à cause, notamment, des dégâts provoqués par les diverses radiations issues des collisions du LHC. Pour permettre une reconstruction stable de l'énergie déposée, il est ainsi nécessaire de corriger ces effets.

Pour ce faire, un monitoring laser est réalisé comme expliqué dans la section 3.1.1 afin de prendre en compte les pertes de transparence en fonction du temps dans les cristaux. L'évolution temporelle de cet effet est visible sur la figure 3.8. Le facteur de correction associé  $LC_i$  à appliquer dans l'équation 4.1 dépend évidemment du temps et du cristal  $i$  considéré ; il est décrit par la réponse à la lumière laser mesurée  $R(t)$  selon l'expression :

$$LC_i(t) = \left[ \frac{R_i(0)}{\beta(t)R_i(t)} \right]^\alpha \quad (4.2)$$

où  $R_i(0)$  correspond à la réponse à la lumière laser prise à un temps de référence choisi comme étant le début de la période de prise de données pour chaque année,  $\beta$  est le même facteur

que dans l'Eq. 4.1 et  $\alpha$  est un paramètre mesurable dépendant de la nature intrinsèque de chaque cristal. Cette formule n'est cependant valide que pour de faibles pertes de transparence. Pour des grandes pertes de transparence (typiquement dans les bouchons) une formule plus sophistiquée doit être utilisée mais elle est en pratique approximée en changeant la valeur de  $\alpha$  [83].

On utilise aussi des méthodes de monitoring basées sur des mesures de collisions in-situ pour déterminer plus globalement la réponse en énergie du ECAL :

- La méthode  $\pi^0$  qui utilise un jeu de données optimisé pour la reconstruction de  $\pi^0$  se désintégrant en deux photons. La stabilité de la réponse en énergie est obtenue en étudiant la masse invariante diphoton.
- La méthode  $E/p$  est basée sur le rapport entre l'énergie et l'impulsion pour des électrons de hautes énergies issus de désintégrations de  $Z$  ou de  $W$ . Des ajustements des distributions en énergie et impulsion sur des fonctions modèles prédéfinies selon différentes régions en  $\eta$  sont utilisées comme indicateur de la stabilité de la réponse en énergie.
- La méthode  $Z \rightarrow ee$  où la masse invariante de deux électrons venant de désintégrations de bosons  $Z$  est utilisé pour mesurer la stabilité de la réponse énergétique du ECAL.

Chacune de ces méthodes indépendantes ont leurs spécificités, et elles sont décrites plus en profondeur dans [83]. La réponse en énergie obtenue par ces méthodes permet de valider les corrections apportés par le monitoring laser ; on peut par exemple observer sur la figure 4.2 que, avec la méthode  $\pi^0$ , on réussit à obtenir une bonne estimation de la masse des  $\pi^0$  lorsque l'on inclut les corrections  $LC$ .

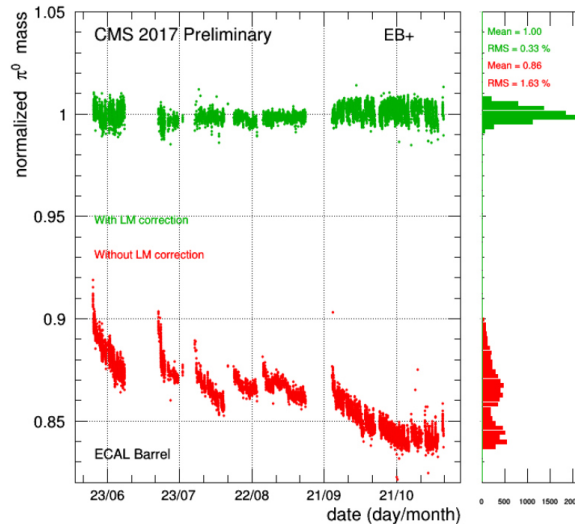


FIGURE 4.2 – Stabilité de l'échelle d'énergie associée à la masse invariante  $\pi^0 \rightarrow \gamma\gamma$  dans EB en fonction du temps. En rouge, les données obtenues sans correction  $LC$  et en vert avec corrections  $LC$  [86].

### 4.1.3 Calibrations du calorimètre électromagnétique $IC$

La calibration du ECAL consiste à stabiliser l'échelle d'énergie pour chaque région/cristal du ECAL vis-à-vis du temps et uniformiser la réponse en énergie pour différents cristaux

pour les mêmes régions angulaires en  $\eta$ . Pour effectuer cette calibration, on peut utiliser une combinaison des trois méthodes vues précédemment ainsi<sup>1</sup> qu'avec une méthode exploitant l'uniformité selon l'angle azimutal  $\phi$  des distributions en énergies déposées dans le ECAL nommée *méthode  $\phi$ -symétrie*<sup>2</sup>.

En effet, la distribution intégrée d'énergie déposée dans le ECAL après de nombreuses collisions p-p est supposément uniforme selon  $\phi$ , ainsi à  $\eta$  fixé, l'énergie déposée dans des cristaux de  $\phi$  différents est en moyenne la même et, par conséquent, si on mesure une différence dans l'énergie déposée celle-ci peut-être attribuée à la variation de la réponse énergétique du cristal considéré. La méthode  $\phi$ -symétrie exploite donc cette caractéristique en mesurant la calibration dans des anneaux, de la largeur d'un cristal, selon  $\phi$  définis à  $\eta$  fixé. Le facteur de calibration  $IC_i$  pour un cristal  $i$  est alors donné par le rapport entre l'énergie déposée dans un anneau autour du cristal  $i$  moyennée et l'énergie déposée dans le cristal  $i$  :

$$IC_i = \frac{\sum_{j=1}^N E_{T,j}^{(i)} / N^{(i)}}{E_{T,i}} \times \kappa \quad (4.3)$$

où le dénominateur  $E_{T,i}$  est l'énergie transverse déposée dans le cristal  $i$ , et le numérateur est la somme des énergies transverses déposées  $E_{T,j}^{(i)}$  dans l'anneau en  $\eta$  où se situe le cristal  $i$  divisée par le nombre de cristaux contenus dans cet anneau.  $\kappa$  est un facteur de correction, calculé séparément, de valeur typique autour de 2. Les constantes  $IC$  sont calculées pour chaque cristal pour différents intervalles de temps.

Les valeurs  $IC$  obtenues avec cette méthode doivent être cependant corrigées pour prendre en compte le budget matériel. La géométrie du ECAL et l'orientation des cristaux dans les bouchons font que ces corrections sont difficiles à calculer pour EE et par conséquent les valeurs de calibration  $IC$  y sont difficilement exploitables.

## 4.2 Reconstruction des photons au sein de l'algorithme Particle Flow

L'énergie obtenue dans l'équation 4.1 permet ensuite de reconstruire les photons au sein de la description générale par évènement de l'algorithme de PF tel que présenté dans la section 3.2.1. Ceci représente un changement important par rapport au Run 1 où les photons et les électrons étaient reconstruits par des algorithmes différents du PF.

Comme vu dans la section 3.2.1, les photons et les électrons sont gérés conjointement par l'algorithme PF étant données les similitudes entre les deux particules. Ainsi la reconstruction PF des photons, et des électrons, peut être résumée, voir [74] pour plus de détails, par différentes étapes pour chaque évènement physique :

1. On regroupe les cristaux possédant une énergie supérieure à un certain seuil fixé en agrégats et on identifie les agrégats graines comme les agrégats possédant une énergie

1. Pour le Run 2 l'incertitude totale de la méthode suivante étant bien plus importantes que celle des autres méthodes, elle n'est utilisée que pour des vérifications.

2. Cette méthode peut aussi permettre un monitoring efficace du ECAL.

déposée supérieure à une certaine valeur (voir tableau 3.1) qui pourraient correspondre à l'interaction d'une seule particule incidente dans le sous-détecteur. L'énergie déposée pouvant être partagée par plusieurs agrégats se superposant, on utilise un modèle de gerbes gaussien pour déterminer la fraction d'énergie déposée pour chaque agrégat.

2. Les agrégats dans le ECAL sont regroupés, selon une zone géométrique dépendant de  $E_T$  et  $\eta$ , autour des agrégats graines dans des super-agrégats afin de prendre en compte, et contenir, les phénomènes de gerbes électromagnétiques, les conversions de photons en paires électron-positron et l'effet bremsstrahlung. Cette étape de super-agrégation s'effectue grâce à deux algorithmes. Le premier est l'algorithme "moustache" qui permet l'ajout d'agrégats à l'agrégat graine selon des régions  $(\eta, \phi)$  légèrement courbées à cause du champ magnétique de CMS (qui rappellent alors la forme de moustaches). Un exemple de distribution simulée d'un super-agrégat reconstruit par l'algorithme "moustache" est visible sur la figure 4.3.

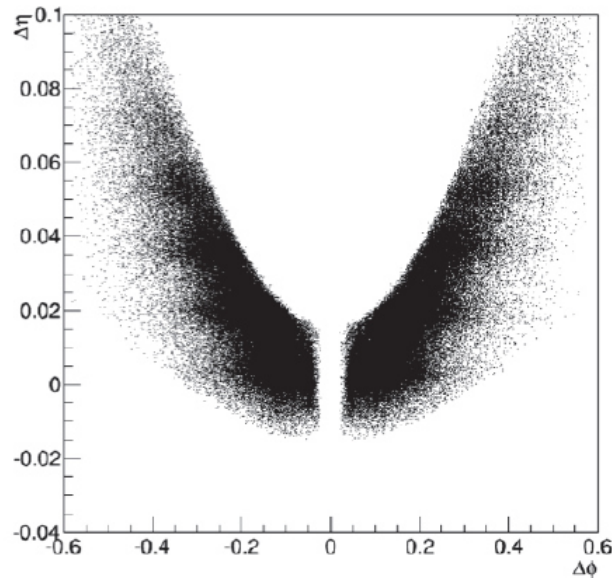


FIGURE 4.3 – Distribution de la position des agrégats par rapport à l'agrégat graine dans les directions  $\eta$  et  $\phi$  (pour des agrégats d'énergie comprise entre 5 et 10 GeV) [87].

Le deuxième algorithme utilisé est l'algorithme "raffiné" qui utilise des informations supplémentaires provenant du trajectographe pour extrapoler les traces venant de conversion et d'effet bremsstrahlung [66] afin de décider si l'agrégat considéré doit appartenir, ou non, à un super-agrégat grâce, notamment, à des arbres de décision boostés ou BDT décrits en détail dans l'annexe A. L'algorithme utilise les super-agrégats moustaches comme point de départ mais il peut aussi créer ses propres super-agrégats.

3. Les coups dans le détecteur à pixels compatibles avec la trajectoire d'un électron et la position d'un super-agrégat sont envoyés en entrée d'un algorithme utilisant le GSF pour estimer les paramètres de la trajectoire reconstruite. Pour un super-agrégat donné, la trajectoire de l'électron correspondant est supposée hélicoïdale et elle est calculée en



utilisant la valeur du champ magnétique, la position du super-agrégat et son énergie transverse.

4. En parallèle, toutes les trajectoires reconstruites à partir de coups dans le détecteur à pixels avec l'algorithme basé sur le KF satisfaisant  $p_T > 2 \text{ GeV}$ , les trajectoires "générales", sont testées vis-à-vis de leur compatibilité avec une trajectoire appartenant à un électron. En cas de compatibilité, ces trajectoires sont aussi envoyées en entrée de l'algorithme à GSF.
5. Les trajectoires générales provenant de conversion de photons sont identifiées par un algorithme dédié [66] permettant de créer des liens entre deux trajectoires compatibles avec une conversion de photon. Cet algorithme est validé par reconstruction de la masse invariante  $\mu\mu\gamma$  provenant de la désintégration d'un boson Z où une paire de trajectoires issues de conversion est assigné au photon. Pour déterminer les traces associées aux conversions de photon où une seule jambe a été reconstruite, un autre arbre de décision boosté est utilisé [74].
6. Les agrégats ECAL, les super-agrégats, les trajectoires GSF, les trajectoires générales et les trajectoires de conversion ainsi que leurs agrégats associés sont tous importés dans l'algorithme de lien PF afin de former des blocs PF.
7. Ces blocs sont résolus en objets électron et photon à partir d'une trajectoire GSF ou d'un super-agrégat c'est-à-dire qu'à cette étape il n'y a pas de différenciation entre photon et électron. Les agrégats ayant été liés pour former de tels objets sont promus en tant que super-agrégats raffinés.
8. Les objets électron ou photon sont reconstruits à partir des super-agrégats raffinés et en satisfaisant différents critères de coupures lâches, les électrons devant aussi posséder une trajectoire GSF associée. La sélection pour les photons est basée sur une coupure  $H/E_{SC} < 0,5$  où H est la somme des énergies déposées dans un cône de  $\Delta R < 0,15$  centré sur la position du super-agrégat dans le HCAL et  $E_{SC}$  est l'énergie du super-agrégat considéré; pour cette sélection l'efficacité de reconstruction des super-agrégats est estimée à 100 %. La sélection pour les électrons étant plus complexe, l'efficacité de reconstruction pour les électrons est légèrement plus faible avec des valeurs avoisinant les 95 % pour la plupart des régions angulaires et pour des électrons de  $p_T > 20 \text{ GeV}$  comme on peut l'observer sur la figure 4.4. Cette collection d'objets est la base utilisée dans le plupart des analyses utilisant les photons et les électrons.

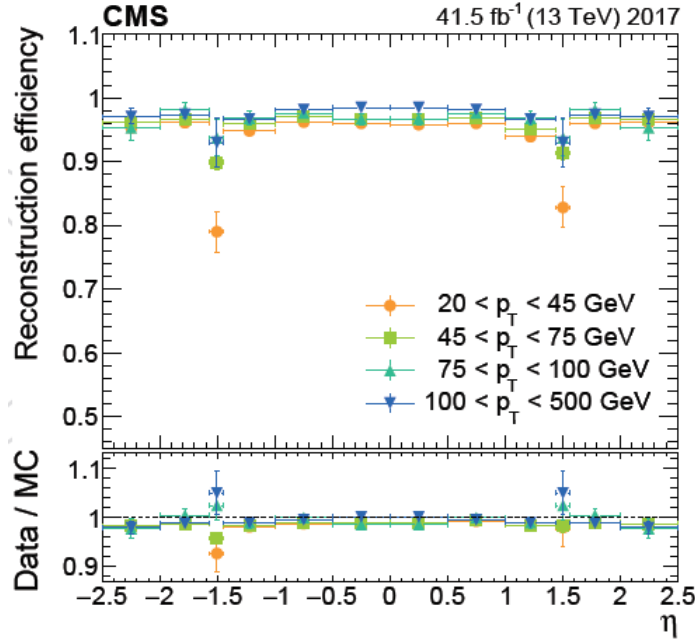


FIGURE 4.4 – Efficacité de reconstruction pour les électrons, définie comme le rapport du nombre de super-agrégats correspondant à des électrons reconstruits sur le nombre de tous les super-agrégats reconstruits pour une sélection donnée, en fonction de la pseudo-rapidité pour différentes valeurs de  $p_T$  dans les données (haut) et pour des rapports données sur simulations (bas) pour la prise de données de 2017. Cette efficacité est calculée avec une méthode "tag-and-probe", décrite plus en détail dans la section 6.3.2, utilisant des événements  $Z \rightarrow ee$  [74].

### 4.3 Corrections à l'énergie des photons

La connaissance précise de l'énergie et de la résolution des photons ainsi reconstruits est essentielle pour les analyses basées sur ces particules. Cette énergie est obtenue en sommant les énergies contenues dans les cristaux des super-agrégats identifiés comme photons. Cependant une partie de cette énergie est perdue dans la reconstruction à cause de fuites dues à l'étalement longitudinal de la gerbe électromagnétique, de fuites latérales dues aux défauts d'alignement des cristaux, de déficits dus aux espaces entre les modules de cristaux, d'électroniques de cristaux morts ou inefficaces, de déficits dus à l'énergie perdue dans le trajectographe, ou encore à cause des conversions. Ces effets résulteraient en des variations de l'énergie mesurée, dégradant alors la résolution en énergie des photons reconstruits, ainsi que des désaccords entre les données observées et les simulations MC qui ne seraient pas alors représentatives des performances mesurées du détecteur.

Pour éviter ces désaccords et préserver la résolution en énergie, élément crucial dans les méthodes utilisées par les analyses utilisant le canal  $H \rightarrow \gamma\gamma$  comme nous le verrons dans la suite, l'énergie d'un photon reconstruit est donc corrigée pour prendre en compte ces pertes :

$$E_\gamma = s_\gamma \times E_\gamma \sum_{i,cristal}^N E_{cristal,i} \quad (4.4)$$

où  $F_\gamma$  est un facteur de correction des effets de pertes présentés précédemment calculé par une méthode d'analyse multivariée basée sur des simulations et  $s_\gamma$  un terme de correction basé sur la comparaison entre données et simulation. Ces facteurs sont explicités dans les sous-sections suivantes.

#### 4.3.1 Correction des pertes d'énergie $F_\gamma$

La correction  $F_\gamma$  est calculée par une méthode de régression multivariée dont le but est d'estimer la fonction de correspondance  $F(\vec{x})$  entre une variable cible  $y$  et des variables d'entrée discriminantes  $\vec{x}$ . Cette méthode repose sur des arbres de régression boostés semi-paramétriques qui fonctionnent de façon similaire aux BDT présentés en annexe A : avec des coupures discriminantes sur les variables d'entrée, on construit des feuilles présentant des similarités en terme de la variable cible  $y$  dans lesquelles on ajuste - c'est le côté semi-paramétrique de la régression - la distribution obtenue en minimisant une fonction de vraisemblance donnée. Pour la correction d'énergie des photons et des électrons, on utilise des simulations MC contenant deux photons/électrons pour entraîner et valider les régressions.

La **variable cible**  $y$  est définie comme le facteur à appliquer à l'énergie mesurée pour obtenir l'énergie totale des objets reconstruits (en prenant en compte les pertes donc), soit le rapport entre l'énergie réelle totale, ou énergie vraie<sup>3</sup>, et l'énergie observée expérimentalement.

Les **variables d'entrée** utilisées lors des régressions sont [88] les coordonnées globales et locales en  $(\eta, \phi)$  des super-agrégats, les variables de forme de gerbe, la variable  $R_9$  définie comme le rapport entre l'énergie contenue dans une matrice  $3 \times 3$  de cristaux sur l'énergie brute du super-agrégat, le nombre d'agréats, le rapport  $H/E$  entre l'énergie dans le calorimètre hadronique sur l'énergie déposée dans le calorimètre électromagnétique dans une région  $\Delta R$  définie autour du super-agrégat ou encore l'énergie déposée dans le détecteur à pied de gerbe. Une définition plus détaillée de ces variables est lisible dans l'annexe B.

La **fonction de vraisemblance** employée utilise, en guise de fonction de densité de probabilité, une fonction double-sided Crystal Ball [88] (dCB), détaillée dans la section 6.9, décrite par 6 paramètres composés des paramètres gaussiens  $\mu$  et  $\sigma$  pour la valeur moyenne et la largeur du pic gaussien, les paramètres gauche et droite de coupure et de forme de la queue  $\alpha_{L/R}$  et  $n_{L/R}$ . La fonction de vraisemblance s'écrit alors comme une somme sur tous les objets  $\gamma$ /électrons présent dans les simulations :

$$\mathcal{L} = - \sum_{MC\gamma/e} \ln(p(y|\vec{x})) = - \sum_{MC\gamma/e} \ln [dCB(y; \mu(\vec{x}), \sigma(\vec{x}), \alpha_L(\vec{x}), n_L(\vec{x}), \alpha_R(\vec{x}), n_R(\vec{x}))] \quad (4.5)$$

Ainsi  $\mu$  peut être interprété comme l'estimation de la correction en énergie, la plus probable, à apporter à un  $\gamma$ /électron reconstruit avec  $\sigma$  la résolution en énergie correspondante. Ces deux paramètres dépendent donc de  $\vec{x}$  et sont donnés par la régression.

Cette régression globale est constituée de plusieurs étapes de régression (2 pour les photons

3. L'énergie vraie est connue dans les simulations mais pas dans les données où le but est toujours de l'estimer.

et 3 pour les électrons) : la première régression a pour but principal de tenir compte des pertes d'énergie détaillées précédemment, et la deuxième régression utilise les mêmes variables d'entrée que la première mais les énergies des super-agrégats sont corrigées par le facteur obtenu dans l'étape précédente et sa variable cible est le rapport entre l'énergie entièrement corrigée et l'énergie corrigée obtenue par la première régression, permettant alors d'estimer la résolution en énergie. Pour les électrons, comme la mesure d'énergie est effectuée de manière indépendante dans le trajectographe et dans le ECAL, une troisième étape de régression est effectuée en combinant l'énergie ECAL avec les moments mesurés dans le trajectographe [74].

L'effet des corrections calculées par ces régressions est visible pour des électrons sur la Fig. 4.5 avec, notamment, la troisième régression pour les électrons basée sur la méthode E-p abordée précédemment apportant des corrections de second ordre. Le gain en terme d'estimation de l'énergie vraie des électrons ainsi qu'en terme de résolution  $y$  est clairement visible.

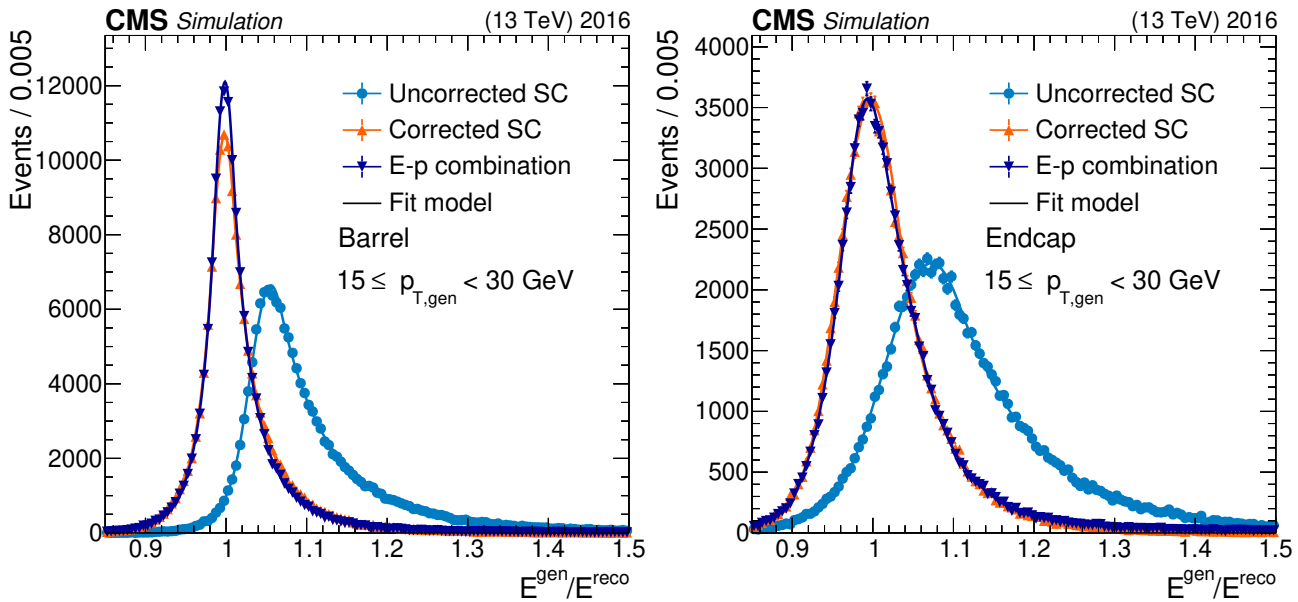


FIGURE 4.5 – Rapport entre l'énergie vraie corrigée et l'énergie non corrigée du super-agrégat pour des simulations d'électrons reconstruits avec un  $p_T$  entre 15-30 GeV avec et sans correction de régressions dans le tonneau (gauche) et les bouchons (droite); un ajustement par une dCB ainsi que la troisième étape de régression E-p pour les électrons sont aussi affichés. Les barres verticales représentent l'incertitude statistique des simulations MC [74].

### 4.3.2 Corrections à l'échelle d'énergie des données et dégradation de la résolution en énergie dans les simulations $s_\gamma$

#### Calcul des échelles d'énergie

Après les corrections en énergie effectuées précédemment, il reste de légères différences entre les données et les simulations pour l'échelle d'énergie des photons et leur résolution en énergie. Ces différences viennent principalement [89] de la description imparfaite du ECAL et du trajectographe dans les simulations, ainsi que la description de l'évolution de transparence

des cristaux et du bruit dans le ECAL. Il est alors nécessaire de corriger l'échelle d'énergie au niveau des données pour qu'elle corresponde à celle obtenue dans les simulations et de dégrader la résolution en énergie des photons dans les simulations pour qu'elle corresponde à celle observée dans les données.

Pour ce faire, des comparaisons entre simulations MC, prenant en compte des corrections fines sur les différents effets des sous-détecteurs, les défauts de reconstruction et les propriétés cinématiques du boson Z, et données sont effectuées sur des électrons provenant d'événements  $Z \rightarrow ee$  [83] :

- Pour corriger les données, on effectue un ajustement de la distribution de la masse invariante  $Z \rightarrow ee$  par une convolution entre une fonction de Breit-Weigner, caractéristique de la désintégration du boson Z dont la largeur est fixée à celle de la désintégration du Z, et d'une fonction "one-sided Crystal Ball", décrivant les effets de résolution en énergie du ECAL et les pertes par effet bremsstrahlung, séparément sur les données et les simulations MC. L'échelle corrective à appliquer aux données est ensuite extraite en comparant les résultats de ces ajustements entre données et simulations.
- Pour dégrader les simulations, on réalise un ajustement en maximisant une fonction de vraisemblance sur les données en utilisant la distribution de la masse invariante des simulations  $Z \rightarrow ee$  comme fonction de densité de probabilité. Les différences obtenues sont décrites par une fonction gaussienne qui est ensuite appliquée aux simulations pour dégrader leur résolution.

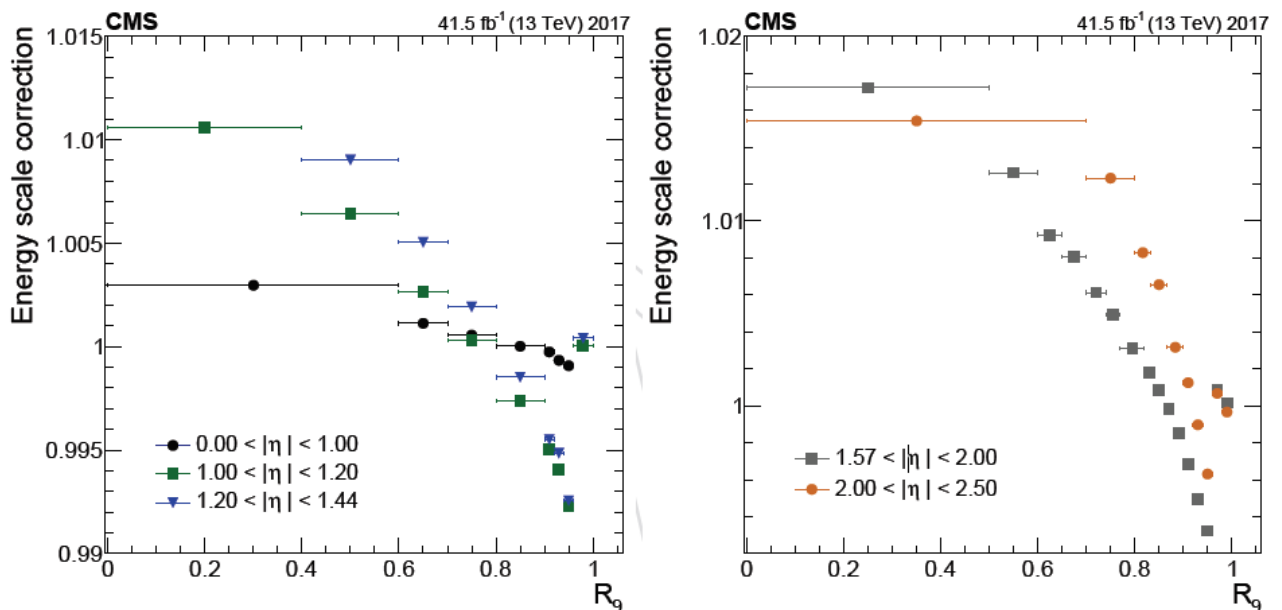


FIGURE 4.6 – Corrections à l'échelle d'énergie pour les données de 2017 en fonction de  $|\eta|$  et  $R_9$ . Les barres d'erreur représentent les incertitudes statistiques pour le tonneau (gauche) et les bouchons (droite) [74].

Pour déterminer les corrections et dégradations finales, ces méthodes sont appliquées de manière procédurale [74] afin de permettre de corriger l'échelle d'énergie en fonction du temps

et de définir des facteurs correctifs/dégradants dans différentes catégories en  $\eta$  et en  $R_9$  afin d'améliorer la précision des corrections et dégradations effectuées. On rappelle que  $R_9$  est une variable définissant le rapport entre l'énergie contenue dans une matrice  $3 \times 3$  cristaux ECAL autour du cristal graine et l'énergie du super-agrégat ; sa valeur est donc comprise entre 0 et 1, des valeurs proches de 1 indiquant la présence de photons non-convertis. Des exemples de corrections à l'échelle d'énergie sont visibles sur la figure 4.6. A noter qu'il faut aussi prendre en compte un facteur d'échelle supplémentaire dépendant du gain utilisé dans les systèmes électroniques de le ECAL [74]. Les corrections finales maximales atteignent 1,5 % de l'échelle d'énergie avec une incertitude totale estimée à moins de 0,1 % pour le tonneau et 0,3 % pour les bouchons.

Les incertitudes totales sont obtenues en additionnant en quadrature les incertitudes statistiques et les incertitudes systématiques dues à l'échelle d'énergie des électron et les corrections en résolution elles-mêmes calculées en faisant varier les distributions de  $R_9$ , les critères sur les électrons utilisés et les coupures en énergie transverse des électrons venant d'évènements  $Z \rightarrow ee$ .

Après avoir appliqué les corrections à l'échelle d'énergie et les dégradations des simulations, on peut observer sur la figure 4.7 que l'accord entre les données et les simulations est bon.

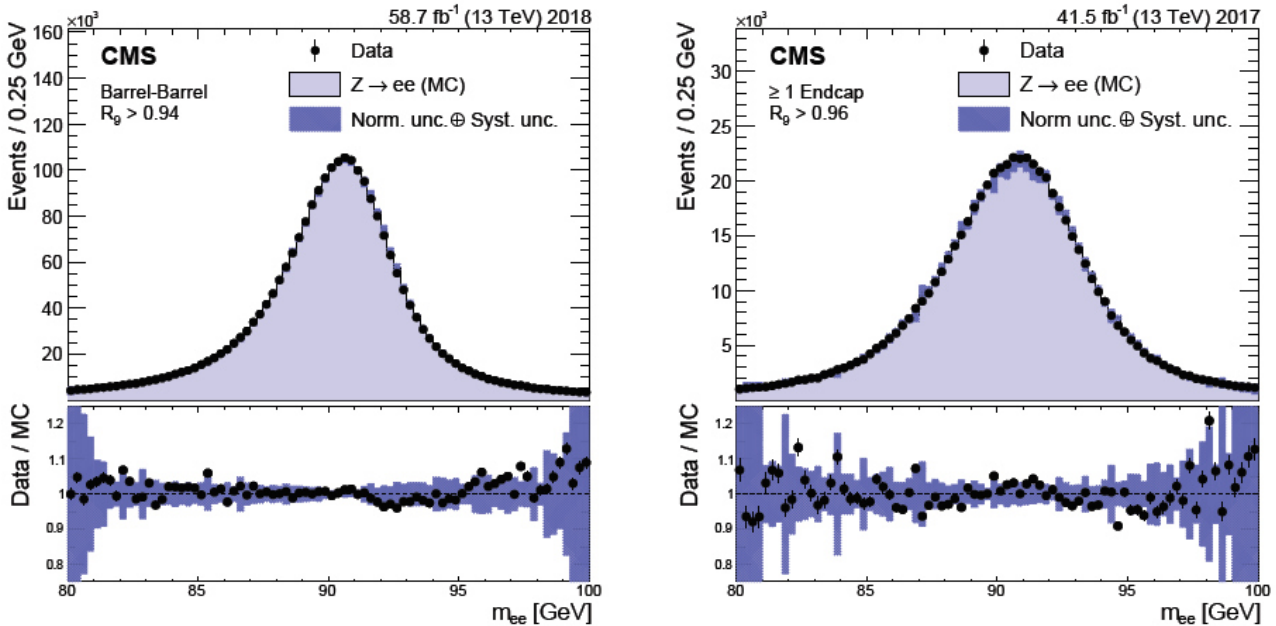


FIGURE 4.7 – Comparaison de la distribution en masse invariante  $Z \rightarrow ee$ , après application des corrections à l'échelle d'énergie des photons sur les données et après dégradation des simulations pour le tonneau (gauche) et les bouchons (droite), pour les données (points noirs) et les simulations (bleu plein) assortie des incertitudes totales (bleu foncé). La partie inférieure présente le rapport entre les données et les simulations [74].

On peut finalement observer l'effet des corrections  $F_\gamma$  (régression) et  $s_\gamma$  (échelle données/simulations) sur les données en visualisant sur la Fig. 4.8 les différences entre la distribution de masse invariante diélectron avant et après application des corrections. La méthode de régression améliore nettement la résolution en masse des électrons, provenant ici de désintégration de boson  $Z$ , et la correction à l'échelle d'énergie décale le pic de la distribution de

masse invariante qui est ainsi plus proche de la valeur tabulée de la masse du boson  $Z$ , elle permet aussi, par sa granularité en  $\eta$  et  $R_9$ , d'améliorer la résolution.

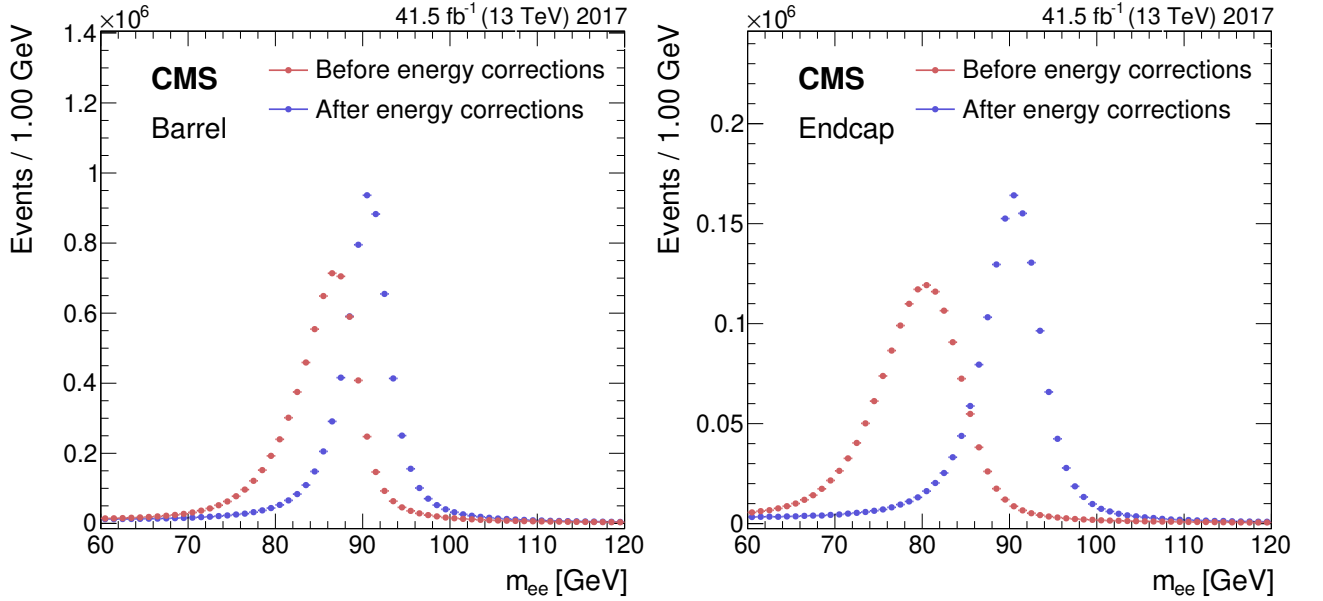


FIGURE 4.8 – Comparaison de la distribution en masse invariante diélectron pour des données  $Z \rightarrow ee$  après application des corrections de la régression et à l'échelle d'énergie des photons pour le tonneau (gauche) et les bouchons (droite). Les barres d'erreur représentent les incertitudes statistiques dans les données mais sont trop petites pour être directement vues ici. De [74].

### Rôles de ces corrections pour les analyses $H \rightarrow \gamma\gamma$

Dans les analyses impliquant une particule de type boson de Higgs dans le canal diphoton, comme les analyses mesurant les propriétés du boson de Higgs découvert à une masse proche de 125 GeV [29, 90, 91] ou encore dans les analyses de recherche de nouvelles particules telle l'analyse présentée dans cette thèse, la résolution en énergie des photons est une variable de première importance puisqu'elle est directement reliée à la masse invariante du système diphoton par la formule  $m_{\gamma\gamma} = \sqrt{E_{\gamma,1}E_{\gamma,2}(1 - \cos\Delta\alpha)}$  où  $\Delta\alpha$  est la différence d'angle entre les 2 photons. Il est donc nécessaire de minimiser la résolution en énergie pour minimiser la résolution en masse invariante, ce qui permet d'obtenir une meilleure précision en masse sur la particule recherchée, de minimiser ainsi la largeur du pic de signal recherché (ce qui est important dans la méthode de modélisation du bruit de fond comme nous l'étudierons dans la section correspondante) et de maximiser le ratio signal sur bruit de fond. Il est donc nécessaire, pour ce type d'analyse, d'avoir une description précise de la résolution en énergie des photons, notamment dans les simulations MC qui sont utilisées pour construire les signaux correspondant aux particules recherchées.



### 4.3.3 Validation et performance des corrections

Les corrections à l'échelle d'énergie et dégradations sont déterminées en utilisant, ainsi que décrit dans la section précédente, des électrons provenant de désintégrations du boson  $Z$  avec un  $p_T$  moyen de 45 GeV. Cependant elles sont aussi appliquées à des photons et des électrons dont le  $p_T$  peut atteindre plusieurs centaines de GeV. Il est donc important de valider les performances des corrections précédentes sur des électrons de hautes énergies et sur des photons.

Pour la validation sur les photons, on utilise des photons provenant d'évènements  $Z \rightarrow \mu\mu\gamma$ , selon une méthode décrite dans le chapitre suivant, permettant d'obtenir une pureté de près 99 %. On peut observer sur la figure 4.9 que la distribution en masse invariante  $m_{\mu\mu\gamma}$  obtenue après l'application des corrections à l'échelle d'énergie aux photons possède un excellent accord données/simulations ; la correction à l'échelle d'énergie est donc globalement valide pour les photons.

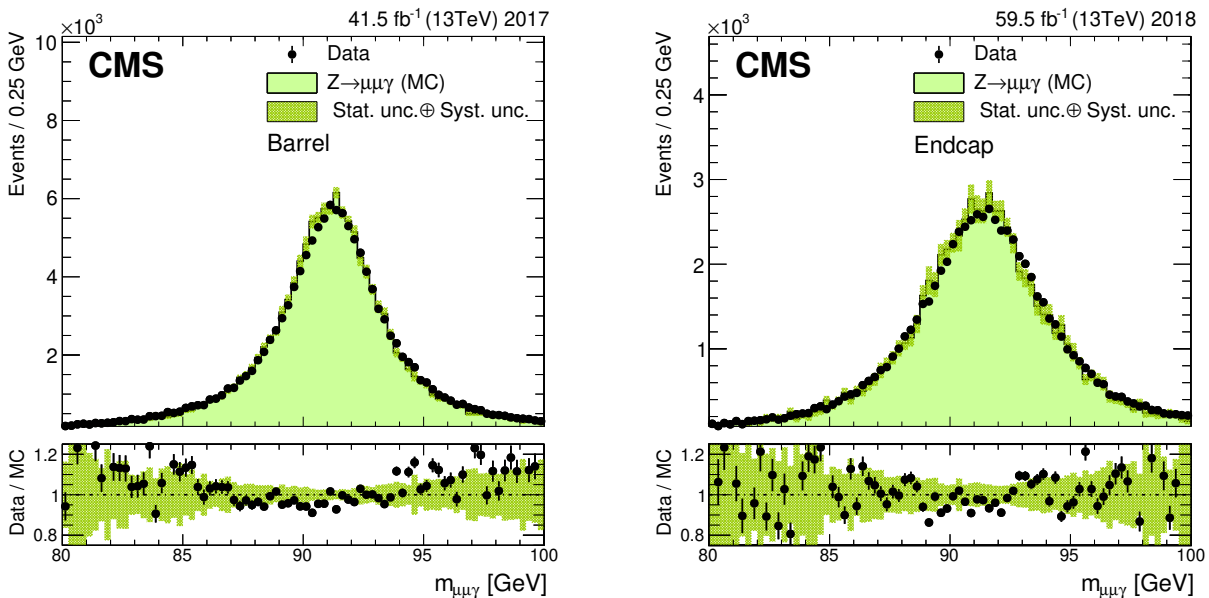


FIGURE 4.9 – Comparaison de la distribution en masse invariante  $Z \rightarrow \mu\mu\gamma$ , après application des corrections à l'échelle d'énergie des photons sur les données et après dégradation des simulations pour 2017 dans le tonneau (gauche) et pour 2018 dans les bouchons (droite), pour les données (points noirs) et les simulations (vert plein) assorties des incertitudes totales (vert foncé hachuré). La partie inférieure présente le rapport entre les données et les simulations [74].

La méthode utilisée pour effectuer cette validation pour les photons sera détaillée dans le chapitre suivant.

## Conclusion

La méthode de reconstruction globale des photons au sein de CMS a donc été présentée. Elle se base, notamment, sur l'énergie déposée au sein du ECAL permettant de créer des super-agrégats grâce à l'algorithme PF. L'énergie des super-agrégats est ensuite corrigée pour



prendre en compte différents effets de pertes d'énergie et les différences entre les distributions pour les données et les simulations MC. Après ces corrections, la résolution en énergie est meilleure et les accords entre données et simulations sont excellents.

Les photons ainsi reconstruits sont proches des électrons et il est nécessaire d'utiliser d'autres outils pour les discriminer du bruit de fond comme des jets ou autres mésons légers. Ces outils doivent être robustes dans le cas d'analyses étudiant un état final constitué de photons et deviennent alors dépendants de la nature de celles-ci. Ainsi nous présenterons dans le chapitre 6, les coupures et méthodes d'identification utilisées pour ce faire. L'outil principal de différenciation entre photons et électrons est le veto à électron, aussi abordé dans ce même chapitre.



# Extraction de l'échelle d'énergie des photons dans CMS

## Sommaire

---

5.1	Les évènements $Z \rightarrow \mu\mu\gamma$ . . . . .	92
5.1.1	La désintégration $Z \rightarrow \mu\mu\gamma$ . . . . .	92
5.1.2	Description des bruits de fond $\mu\mu\gamma$ . . . . .	93
	Bruit de fond irréductible . . . . .	93
	Bruit de fond réductible . . . . .	93
5.1.3	Sélection des évènements . . . . .	93
5.2	L'échelle d'énergie des photons . . . . .	94
5.2.1	Définition de l'échelle d'énergie . . . . .	94
5.2.2	Expression littérale . . . . .	95
5.3	Extraction de l'échelle d'énergie . . . . .	97
5.3.1	Méthode d'extraction . . . . .	97
	Ajustement . . . . .	97
	Choix du meilleur ajustement . . . . .	97
	Estimations des incertitudes systématiques . . . . .	100
5.3.2	Résultats obtenus . . . . .	103
	Valeur de l'échelle d'énergie extraite . . . . .	103
	Test de fermeture . . . . .	105

---

Nous avons étudié, dans le chapitre précédent, l'importance, notamment pour les analyses  $H \rightarrow \gamma\gamma$ , de corriger l'échelle d'énergie des photons dans les données et les simulations. La méthode utilisée pour ce faire étant basée sur des événements  $Z \rightarrow ee$ , soit des bosons  $Z$  se désintégrant en électrons, et étant donné que les photons et les électrons n'ont pas exactement le même comportement, il est nécessaire de savoir si les corrections obtenues sont valables ou non pour les photons. Cette validation est effectuée avec des événements  $Z \rightarrow \mu\mu\gamma$  présentés, ainsi que la méthode utilisée pour extraire l'échelle d'énergie, dans ce chapitre. La dérivation des échelles d'énergies pour les photons fait partie des tâches de service que j'ai effectuées au sein de la collaboration CMS où j'ai utilisé les données mesurées en 2017 avec les dernières corrections possibles (nommées *2017 Ultra Legacy*). Les données utilisées correspondent aux échantillons **DoubleMuon** et l'échantillon de simulation utilisé est **DYJetsToLL** qui correspond à des désintégrations leptoniques de bosons  $Z$ . L'énergie dans le centre de masse proton-proton est de 13 TeV et la luminosité intégrée est d'environ  $41,5 \text{ fb}^{-1}$ .

## 5.1 Les événements $Z \rightarrow \mu\mu\gamma$

### 5.1.1 La désintégration $Z \rightarrow \mu\mu\gamma$

Les événements  $Z \rightarrow \mu\mu\gamma$  proviennent de la désintégration d'un boson  $Z$  en deux muons dont l'un d'eux émet un photon comme illustré sur la figure 5.1, on parle alors d'émission dans l'état final (FSR).

Le principe est le même que pour les événements  $Z \rightarrow ee$  : comme les propriétés du boson  $Z$  sont connues avec une grande précision expérimentale [92], et que l'expérience CMS est capable de mesurer les muons avec une grande précision [93], on peut remonter aux caractéristiques du photon issu de la désintégration à trois corps à partir de la cinématique des autres objets.

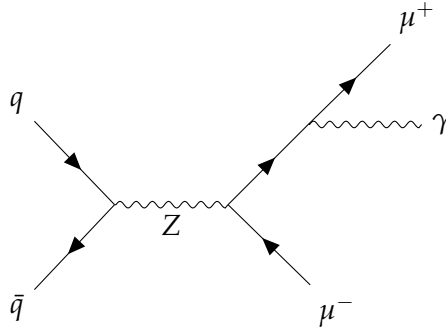


FIGURE 5.1 – Exemple de diagramme de Feynman pour des désintégration  $Z \rightarrow \mu\mu\gamma$

Ce type de désintégration offre une signature très claire puisqu'elle est purement électrofaible et le photon est le seul objet présent dans le ECAL. Les désintégrations  $Z \rightarrow \mu\mu\gamma$  peuvent donc fournir un échantillon de photon non biaisé et de haute pureté qui permet des validations dédiées de la simulation des photons, et peut même être utilisé comme méthode de calibration et de mesure d'efficacité ; comme nous le verrons par exemple dans le chapitre suivant.

### 5.1.2 Description des bruits de fond $\mu\mu\gamma$

Comme nous venons de le voir, les évènements que l'on souhaite utiliser pour extraire l'échelle d'énergie des photons sont des évènements  $Z \rightarrow \mu\mu\gamma$  avec  $\gamma$  FSR. Cependant les données ne contiennent pas uniquement ces évènements, il est alors nécessaire d'identifier les évènements pouvant être, à tort, identifiés comme des évènements  $Z \rightarrow \mu\mu\gamma$  FSR; ces évènements sont appelés bruits de fond, par rapport aux évènements d'intérêt nommés évènements signaux.

Il est usuel de séparer le bruit de fond en deux catégories, le bruit de fond "irréductible" dont l'état final est le même que celui des évènements d'intérêt, et le bruit de fond "réductible" où l'état final est composé d'évènements qui peuvent être mal identifiés comme évènements d'intérêt. Les principales sources de bruit de fond ici sont donc :

#### Bruit de fond irréductible

- Désintégrations  $Z \rightarrow \mu\mu\gamma$  mais le photon est émis dans l'état initial (ISR) comme représenté sur la Fig. 5.2
- Désintégrations  $Z \rightarrow \mu\mu$  et le photon est émis depuis un jet de recul ou un autre vertex d'interaction à cause de l'empilement

#### Bruit de fond réductible

- De manière générale des interactions QCD sont susceptibles de former des jets contenant des muons et des photons
- Deux muons provenant de désintégrations de boson  $W^\pm$  provenant d'un évènement  $t\bar{t}$  avec photon émis par radiation
- Muons produits par des désintégrations de  $W^\pm$  provenant d'un jet et de jets de recul

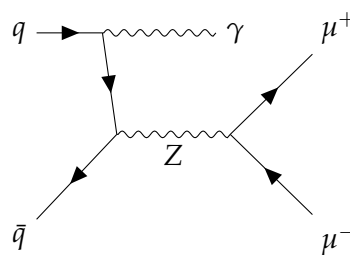


FIGURE 5.2 – Exemple de diagramme de Feynman pour des désintégration  $Z \rightarrow \mu\mu\gamma$  ISR

Il est important d'identifier les principales sources de bruit de fond afin d'optimiser des critères et coupures de sélection permettant de rejeter un maximum d'évènements de bruit de fond tout en gardant un maximum d'évènements d'intérêts.

### 5.1.3 Sélection des évènements

Pour rejeter les évènements de bruit de fond tout en sélectionnant les évènements de signal, on effectue une sélection des évènements en plusieurs étapes dont les valeurs de coupure sont

résumées dans le tableau 5.3.

Pour qu'un évènement soit accepté, il faut donc qu'il satisfasse aux conditions ci-dessous :

1. Chemin de déclenchement muon et critère de qualité des données certifiées. Le HLT utilisé est HLT\_Mu17\_TrkIsoVVL\_Mu8\_TrkIsoVVL\_DZ\_Mass8\_v\* sélectionnant 2 muons d'impulsion transverse de plus de 17 et 8 GeV isolés dans le trajectographe avec une masse invariante  $m_{\mu\mu} > 8$  GeV.
2. Au moins deux muons satisfaisant des critères plus contraignants d'identification. Le mode d'identification des muons *tight* [93] ou resserré, qui requiert un muon à la fois *standalone* et *tracker* muon bien reconstruit, auquel on ajoute notamment une condition sur l'isolation en hadron chargés  $I_{charged}^{\mu}$ <sup>1</sup>, afin de rejeter une partie importante des évènements QCD incluant des jets, sont requis.
3. La paire dimuon reconstruite doit être composée de deux muons de charge opposée et de masse invariante suffisamment élevée  $m_{\mu\mu} > 35$  GeV.
4. Un seul photon avec des coupures d'acceptation lâches  $E_T > 20$  GeV et  $\eta < 2,5$  (mais pas dans la zone  $1,4442 < |\eta| < 1,566$  entre EB et EE).
5. Des coupures cinématiques et d'isolation pour garantir une radiation FSR :
  - a) L'impulsion transverse du muon le plus éloigné du photon reconstruit doit être supérieur à 30 GeV et la distance angulaire entre le muon le plus proche du photon et ce dernier doit être relativement faible, soit  $\text{Min}(\Delta R(\mu^{\pm}, \gamma)) < 0,8$ , afin de rejeter les évènements non FSR.
  - b) La masse invariante du système à 3 corps doit être compatible avec la masse du boson Z afin de rejeter les évènements qui ne correspondent pas à une désintégration du Z,  $m_{\mu\mu\gamma} \in [60,120]$  GeV et  $m_{\mu\mu} + m_{\mu\mu} < 180$  GeV.

Pour finir dans les rares cas où il y a plus d'un candidat passant cette sélection pour une interaction proton-proton donnée, on choisit l'évènement dont la masse invariante  $m_{\mu\mu\gamma}$  est la plus proche de celle du boson Z.

Ces sélections permettent d'obtenir un lot de photons atteignant une pureté de plus de 98 % inclusive (95 % dans EE) [94] pouvant être utilisés dans l'extraction de l'échelle d'énergie.

## 5.2 L'échelle d'énergie des photons

### 5.2.1 Définition de l'échelle d'énergie

L'échelle d'énergie  $s$  est définie en fonction du rapport entre l'énergie des photons issus de désintégration  $Z \rightarrow \mu\mu\gamma$  mesurée expérimentalement et leur énergie réelle totale, appelée énergie vraie, telle :

$$s = \frac{E_{mesure}^{\gamma}}{E_{vraie}^{\gamma}} - 1 \quad (5.1)$$

Le problème de cette définition est que l'énergie vraie est accessible uniquement pour les simulations<sup>2</sup> et non pas pour les données. Or ce que l'on souhaite mesurer est un accord entre

1. Correspond à la somme des moments transverses des hadrons chargés dans un cône  $\Delta R < 0.4$  autour du muon considéré.

2. Elle correspond à l'énergie au niveau générateur.

## 5.2. L'ÉCHELLE D'ÉNERGIE DES PHOTONS

FIGURE 5.3 – Tableau résumant la liste des coupures permettant de sélectionner les événements  $Z \rightarrow \mu\mu\gamma$

Catégorie	Coupure
Déclenchement Muons	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v
Sélection des muons	muons <i>tight</i> [93] $I_{charged}/p_T < 0,2$ $p_T > 10 \text{ GeV}$ $ \eta  < 2,4$
Sélection des systèmes dimuon	paire de muons de charge opposée $m_{\mu\mu} > 35 \text{ GeV}$
Sélection photon	$E_T > 20 \text{ GeV}$ $ \eta  < 1,4442 \cup 1,566 <  \eta  < 2,5$
Sélection FSR	$\text{Min}(\Delta R(\mu^\pm, \gamma)) < 0,8$ $p_T^\mu > 30 \text{ GeV}$ pour le $\mu$ le plus éloigné de $\gamma$ $m_{\mu\mu\gamma} \in [60,120] \text{ GeV}$ $m_{\mu\mu\gamma} + m_{\mu\mu} < 180 \text{ GeV}$

données et simulations, il faut donc essayer de trouver une expression littérale de  $s$  qui pourrait être estimée à partir des données uniquement. La validité de cette approche a été établie dans des travaux antérieurs [95, 96] à ceux présentés dans cette thèse.

### 5.2.2 Expression littérale

La cinématique à trois corps des événements  $Z \rightarrow \mu\mu\gamma$  permet d'écrire dans le cas général :

$$\begin{aligned}
 m_{\mu\mu\gamma}^2 &= p_{\mu\mu\gamma}^\nu p_{\mu\mu\gamma,\nu} \\
 &= (p_{\mu\mu}^\nu + p_\gamma^\nu)(p_{\mu\mu,\nu} + p_{\gamma,\nu}) \\
 &= m_{\mu\mu}^2 + 2p_{\mu\mu}^\nu p_{\gamma,\nu} + 0 \\
 &= m_{\mu\mu}^2 + 2(E_{\mu\mu}E_\gamma - \vec{p}_{\mu\mu} \cdot \vec{p}_\gamma)
 \end{aligned}$$

En définissant le vecteur unitaire  $\vec{u}_\gamma$  tel que  $\vec{p}_\gamma = E_\gamma \vec{u}_\gamma$ , l'équation précédente permet d'écrire :

$$E_\gamma = \frac{m_{\mu\mu\gamma}^2 - m_{\mu\mu}^2}{2(E_{\mu\mu} - \vec{p}_{\mu\mu} \cdot \vec{u}_\gamma)} \quad (5.2)$$

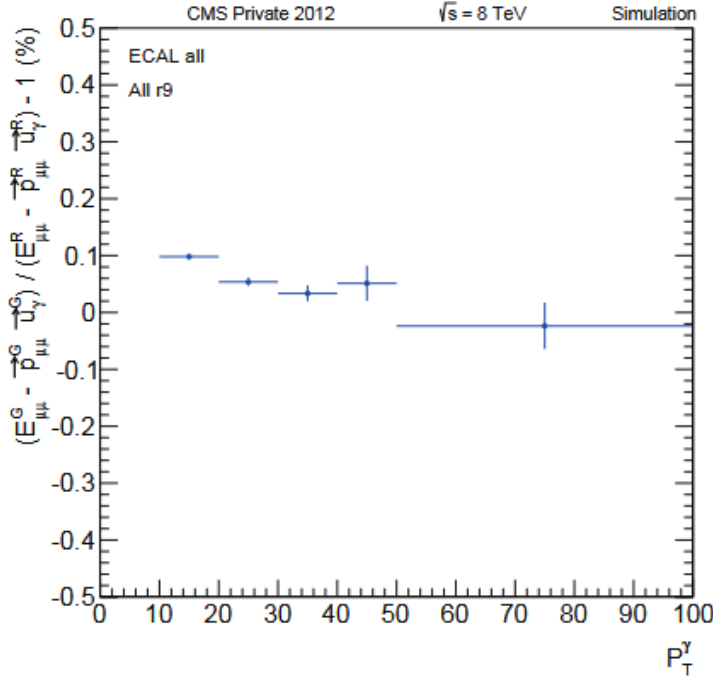


FIGURE 5.4 – Distribution de  $\frac{E_{\mu\mu,vraie} - \vec{p}_{\mu\mu,vraie} \cdot \vec{u}_{\gamma,vraie}}{E_{\mu\mu,mesure} - \vec{p}_{\mu\mu,mesure} \cdot \vec{u}_{\gamma,mesure}} - 1$  en fonction de  $p_T^\gamma$  [95]

La formule 5.2 peut être utilisée pour décrire l'énergie mesurée et l'énergie vraie des photons issues d'évènements  $Z \rightarrow \mu\mu\gamma$ , on peut donc réécrire l'expression de l'échelle d'énergie  $s$  de l'équation 5.1 comme :

$$s = \frac{m_{\mu\mu\gamma,mesure}^2 - m_{\mu\mu,mesure}^2}{m_{\mu\mu\gamma,vraie}^2 - m_{\mu\mu,vraie}^2} \times \frac{E_{\mu\mu,vraie} - \vec{p}_{\mu\mu,vraie} \cdot \vec{u}_{\gamma,vraie}}{E_{\mu\mu,mesure} - \vec{p}_{\mu\mu,mesure} \cdot \vec{u}_{\gamma,mesure}} - 1 \quad (5.3)$$

Dans l'objectif de construire un estimateur de  $s$  qui n'utilise pas les termes "vraie", on cherche à négliger et/ou remplacer ces termes par des termes pouvant être mesurés expérimentalement. Premièrement, il a été montré [95] que le terme  $\frac{E_{\mu\mu,vraie} - \vec{p}_{\mu\mu,vraie} \cdot \vec{u}_{\gamma,vraie}}{E_{\mu\mu,mesure} - \vec{p}_{\mu\mu,mesure} \cdot \vec{u}_{\gamma,mesure}}$  est très proche de 1, avec des écarts inférieurs à  $10^{-3}$ , comme on peut l'observer sur la figure 5.4. Les variations de ce terme pourront alors être négligées dans l'estimateur de  $s$  que l'on cherche à construire.

Les énergies et directions des muons étant connue avec une grande précision (avec une résolution sur le moment transverse inférieure à 2 % dans le tonneau [75] et une résolution spatiale de l'ordre d'une centaine de  $\mu\text{m}$  atteinte dans les chambres à muons) dans CMS et étant bien corrigées dans les simulations, on peut approximer  $m_{\mu\mu,vraie}$  par son équivalent reconstruit  $m_{\mu\mu,mesure}$ . On fait aussi l'hypothèse [96] que l'on peut approximer  $m_{\mu\mu\gamma,vraie}$  par la masse  $m_Z$  du boson Z tabulée dans le PDG [92], afin de définir notre estimateur de  $s$  comme :

$$s_{RECO} = \frac{m_{\mu\mu\gamma,mesure}^2 - m_{\mu\mu,mesure}^2}{m_Z^2 - m_{\mu\mu,mesure}^2} \quad (5.4)$$

L'estimateur  $s_{RECO}$  ainsi construit est bien indépendant de toute variable "vraie". Il a, de plus, été montré [95] que  $s_{RECO}$  est un bon estimateur de la variable  $s$ . On peut donc utiliser la variable  $s_{RECO}$  pour calculer l'échelle d'énergie des photons à la fois dans les données et dans



les simulations.

## 5.3 Extraction de l'échelle d'énergie

Notre but est maintenant de calculer l'estimateur de l'échelle d'énergie défini dans l'Eq. 5.4 pour les données et les simulations correspondant aux évènements  $Z \rightarrow \mu\mu\gamma$  étudiés précédemment. Les résultats présentés dans cette section utilisent les échantillons *2017 Ultra Legacy* et sont présentés selon différentes régions en  $\eta$ , soit EB et EE, et  $R_9$ ; on ne s'intéresse d'ailleurs ici qu'à de hautes valeurs de  $R_9 > 0.965$  correspondant à de très bons photons. La méthode utilisée est une méthode approuvée [97] par le *ECAL Detector Performance Group* présentée ci-après.

### 5.3.1 Méthode d'extraction

#### Ajustement

Pour extraire  $s_{RECO}$ , on commence tout simplement par calculer sa valeur pour chaque évènement et régions en  $\eta$  et  $R_9$ . On ajuste ensuite les distributions obtenues par une fonction de Voigt  $V$ , définie comme la convolution d'une fonction de Lorentz avec une fonction gaussienne :

$$V(m; \mu, \sigma, \gamma) = \int_{-\infty}^{+\infty} G(x, \sigma) L(m - x; \mu, \gamma) dx \quad (5.5)$$

avec  $G(x; \sigma) = \frac{e^{\left(\frac{-x^2}{2\sigma^2}\right)}}{\sqrt{2\pi}\sigma}$  et  $L(x; \gamma) = \frac{\gamma}{\pi((x - \mu)^2 + \gamma^2)}$

où  $\mu$  est la valeur maximisant la fonction, et  $\sigma$  et  $\gamma$  sont les paramètres de forme de la gaussienne et de la lorentzienne. L'ajustement est réalisé sur différents intervalles de  $s_{RECO}$ . En effet, on considère que  $s_{RECO}$  est originellement définie entre -0.5 et 0.5 puis l'on procède à un ajustement non binné par la fonction de Voigt en utilisant l'algorithme MINUIT [98] sur des intervalles de définition variant par pas de 1 % entre 60 % et 100 % de cet intervalle de départ. On identifie finalement la valeur  $\mu$  correspondant au pic de la fonction comme la valeur  $s_{RECO}$  obtenue pour un ajustement donné.

On obtient alors 41 ajustements différents, soit 41 valeurs de  $s_{RECO}$  pour chaque région en  $(\eta, R_9)$ , parmi lesquels on définit la notion de meilleur ajustement.

#### Choix du meilleur ajustement

##### Figure de mérite - Cas général

Le critère de qualité des ajustements, ou figure de mérite, utilisé est la p-valeur ou p-value basée sur un test de  $\chi^2$ .

En effet un test basique pour s'assurer de la qualité d'un ajustement sur des données est de calculer le  $\chi^2$  défini comme la somme sur chaque bin  $i$  de la différence entre la valeur  $d_i$  des données dans le bin considéré et l'évaluation  $a_i$  dans ce bin de l'ajustement effectué, divisé par

l'erreur statistique  $\sigma$  de la valeur  $d_i$ ; le tout au carré comme décrit ci-dessous :

$$\chi^2 = \sum_{i=1}^{\text{bins}} \frac{(d_i - a_i)^2}{\sigma^2} \quad (5.6)$$

Un ajustement est ensuite qualifié de bon si le rapport entre le  $\chi^2$  et le nombre de degrés de liberté  $ndf$  (défini comme la différence entre le nombre de bins choisi et le nombre de paramètres utilisés dans l'ajustement) est proche de 1. Ce test en lui-même ne constitue cependant pas une véritable figure de mérite [99] et il vaut mieux, dans cette optique, utiliser la p-value que l'on définit comme étant la probabilité, pour une valeur de  $\chi^2$  observée donnée, d'obtenir une valeur de  $\chi^2$  plus élevée. Cette probabilité est calculée à l'aide de la fonction gamma incomplète telle que :

$$\text{p-value}(ndf, \chi_{obs}^2) = 1 - \frac{\int_0^{\chi_{obs}^2/2} t^{ndf/2-1} e^{-t} dt}{\int_0^{+\infty} t^{ndf/2-1} e^{-t} dt} \quad (5.7)$$

Ainsi une p-value maximale de 0,5 correspond à avoir 50 % de chance d'obtenir, lors d'un nouveau test d'ajustement, une valeur de  $\chi^2$  plus élevée par rapport à celle observée, dans ce cas l'ajustement réalisé est donc excellent. Inversement une p-value minimale de 0 correspond à avoir une chance, dans un monde parfait, de 0 % d'obtenir une valeur de  $\chi^2$  plus élevée par rapport à celle observée, autrement dit la valeur observée est déjà la plus élevée possible et donc l'ajustement est mauvais.

### Meilleur ajustement

Le choix du meilleur ajustement de l'échelle d'énergie s'effectue de manière itérative pour chaque bin en  $\eta$  (EB/EE) et des valeurs de  $R_9 > 0,965$  correspondant aux très bon candidats photons :

1. On choisit des ajustements sur des intervalles entre 90 et 100 % de l'intervalle originel si les ajustements ont une p-value  $> 0,01$ . Si plus d'un ajustement satisfait ce critère, on choisit celui minimisant l'incertitude statistique sur le paramètre  $\mu$ .
2. Si aucun ajustement ne vérifie les conditions précédents, on assouplit la condition sur les intervalles en demandant, toujours avec une p-value  $> 0,01$ , des intervalles entre 80 et 90 %, puis entre 70 et 80 % et finalement entre 60 et 70 %
3. Si aucun ajustement n'est sélectionné, on recommence la procédure en relâchant la coupure sur la p-value de manière itérative ( 0,001 puis 0,0001, etc) jusqu'à obtenir un ajustement, par bin de  $(\eta, R_9)$ , alors défini comme le meilleur ajustement.

Les meilleurs ajustements choisis sont présentés dans le tableau 5.1 avec leur p-value et intervalle associé. Ils sont séparés entre EB et EE pour des valeurs  $R_9 > 0,965$  pour les données ainsi que pour les simulations MC correspondant à des événements Drell-Yan, soient des désintégrations leptoniques de boson Z. Les courbes d'ajustement correspondantes sont visibles sur la figure 5.5 où la valeur moyenne affichée correspond à la valeur de  $s_{RECO}$  coïncidente avec l'incertitude statistique associée. Il faut cependant encore calculer les erreurs systématiques

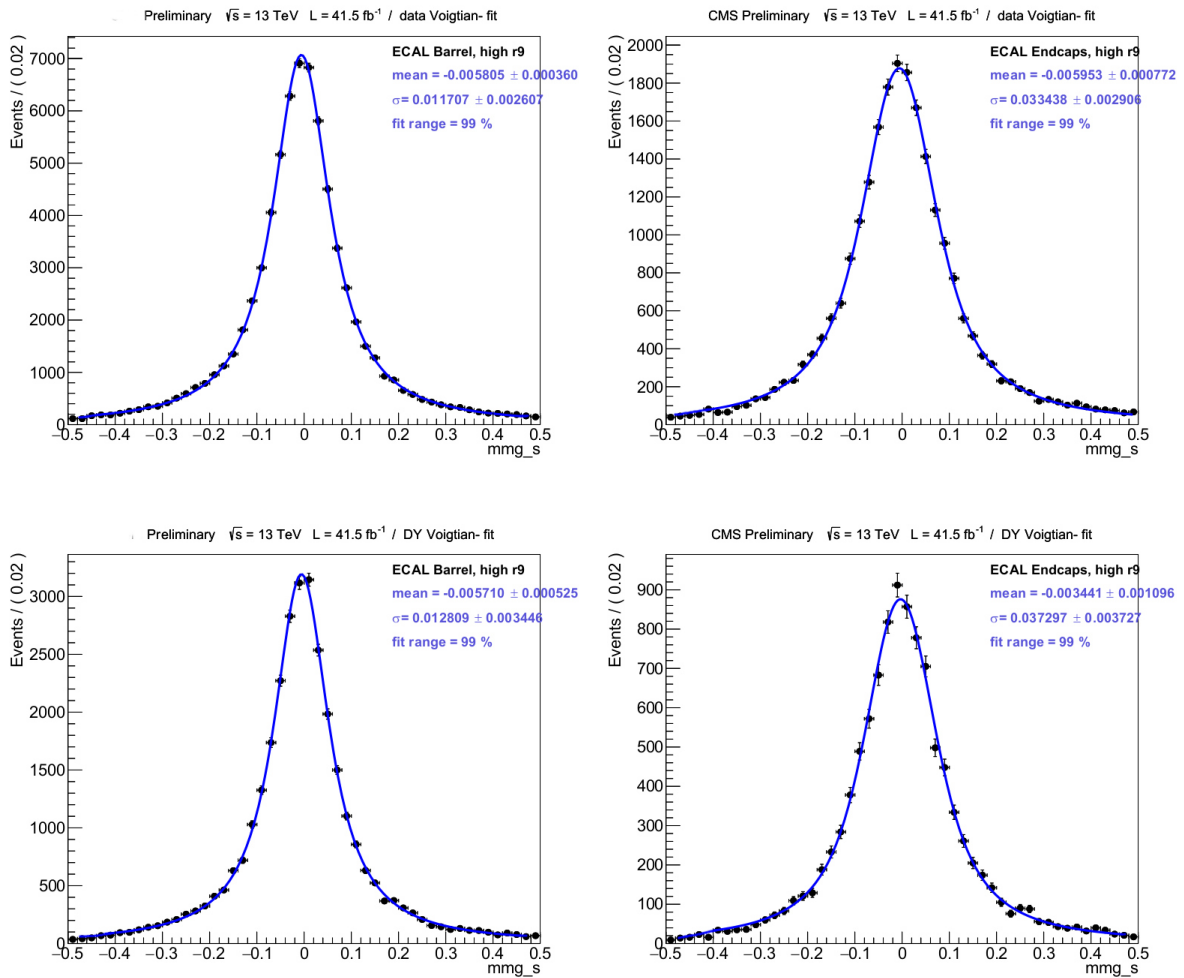
### 5.3. EXTRACTION DE L'ÉCHELLE D'ÉNERGIE

associées à la méthode utilisée pour extraire l'échelle d'énergie  $s_{RECO}$  afin d'avoir une méthode robuste.

p-value et intervalle des meilleurs ajustements				
	data (p-value \ fit range)		DY (p-value \ fit range)	
EB High $R_9$	0.153	99 %	0.189	99 %
EE High $R_9$	0.034	99 %	0.429	99 %

TABLE 5.1 – P-value et intervalle pour les meilleurs ajustements choisis dans EB et EE pour des valeurs  $R_9 > 0.965$  pour les données et les simulations MC correspondant à des événements Drell-Yan, soient des désintégrations leptoniques de boson Z.

FIGURE 5.5 – Distribution de l'échelle d'énergie  $s_{RECO}$  (points noirs) et meilleur ajustement voigtien (courbes bleues) avec  $R_9 > 0.965$  dans EB (gauche) et EE (droite) pour les données (haut) et les simulations (bas) 2017 Ultra Legacy. Sont aussi indiqués la valeur de  $s_{RECO}$  extraite (mean), la largeur de la partie gaussienne  $\sigma$  et l'intervalle utilisé pour réaliser l'ajustement.



## Estimations des incertitudes systématiques

On identifie deux sources prépondérantes d'erreurs systématiques associées à la méthode utilisée pour extraire  $s_{RECO}$  : une incertitude systématique pour prendre en compte le choix de l'intervalle choisi pour l'ajustement et une incertitude systématique pour prendre en compte le choix de la fonction choisie pour réaliser les ajustements.

### Incertitude systématique liée à l'intervalle d'ajustement

Pour prendre en compte l'effet du choix de l'intervalle d'ajustement, on assigne comme systématique la différence maximale entre la valeur  $s_{RECO}$  pour le meilleur ajustement et les valeurs de  $s_{RECO}$  correspondant à des intervalles d'ajustement situés dans une plage de valeur de 20 % centrée autour de l'intervalle du meilleur ajustement et avec une p-value  $> 0,0001$ . Par exemple, si l'intervalle du meilleur ajustement est de 95 %, on considère des valeurs de  $s_{RECO}$  extraites avec des intervalles de 80 % à 100 % ; si l'intervalle du meilleur ajustement est de 85 %, on va considérer des valeurs de  $s_{RECO}$  extraites avec des intervalles de 75 % à 95 %. L'incertitude systématique associée correspond alors à la variation maximale de  $\mu$  obtenue sur l'intervalle considéré. Les variations de  $\mu$ , l'échelle d'énergie  $s_{RECO}$  extraite des ajustements, en fonction de l'intervalle d'ajustement considéré sont visibles<sup>3</sup> sur la figure 5.6 tandis que les variations de p-value associées sont visibles sur la figure 5.7 afin d'avoir un ordre de grandeur des incertitudes considérées (environ 0,05 %).

---

3. La figure présente des intervalles allant de 60 à 100 % bien que dans le cas présent l'intervalle utilisé pour la détermination de l'incertitude va de 80 à 100 %.

### 5.3. EXTRACTION DE L'ÉCHELLE D'ÉNERGIE

FIGURE 5.6 – Valeur de l'échelle d'énergie  $s_{RECO}$  en fonction de l'intervalle d'ajustement avec  $R_9 > 0.965$  dans EB (gauche) et EE (droite) pour les données (haut) et les simulations (bas) *Ultra Legacy 2017*.

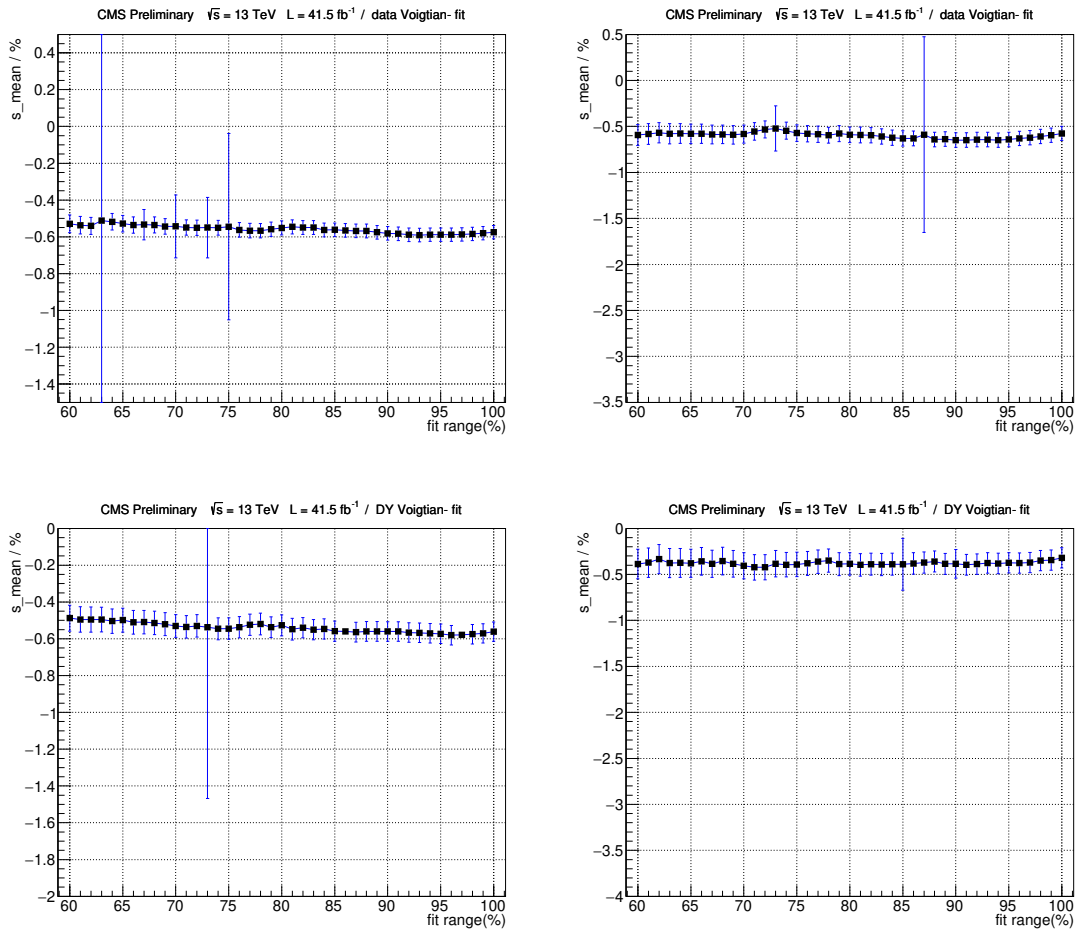
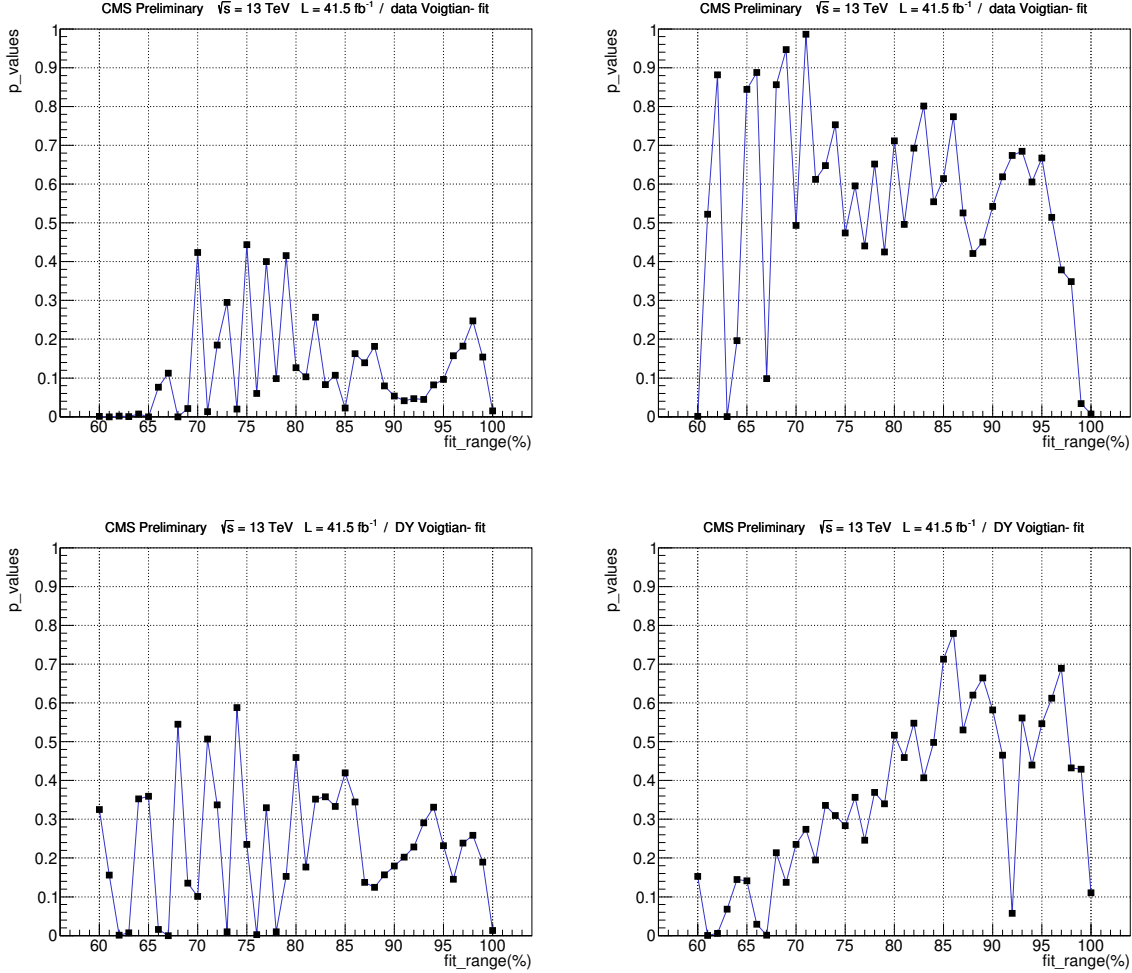


FIGURE 5.7 – Distribution la p-value associée à l'ajustement en fonction de l'intervalle d'ajustement avec  $R_9 > 0.965$  dans EB (gauche) et EE (droite) pour les données (haut) et les simulations (bas) *Ultra Legacy 2017*.



### Incertitude systématique liée au choix de la fonction d'ajustement

Pour estimer l'incertitude liée au choix de la fonction d'ajustement utilisée, on procède selon la procédure suivante :

1. On utilise la fonction de distribution voigtienne correspondant au meilleur ajustement dans chaque bin pour générer 1000 jeux de pseudo-données (MC toys) contenant chacun plus de 10 000 évènements.
2. On ajuste ces 1000 distributions générés par une fonction de Cruiff, qui donne des performances d'ajustement comparables à la fonction de Voigt utilisée, définie par la formule :

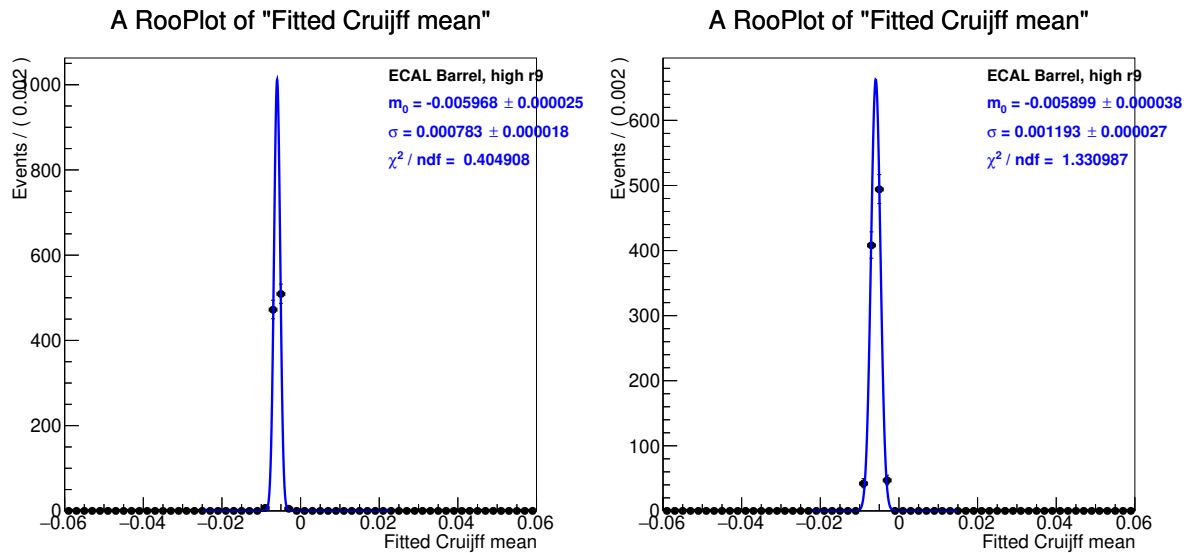
$$C(m; m_0, \sigma_L, \sigma_R, \alpha_L, \alpha_R) = \exp\left(\frac{-(m - m_0)^2}{2\sigma^2 + \alpha(m - m_0)^2}\right) \quad (5.8)$$

$$\text{avec } \sigma, \alpha = \begin{cases} \sigma_L, \alpha_L & \text{si } (m - \mu) < 0 \\ \sigma_R, \alpha_R & \text{si } (m - \mu) > 0 \end{cases}$$

3. On ajuste avec une fonction gaussienne la distribution des 1000 moyennes des fonctions de Cruiff obtenues.
4. La différence entre la valeur centrale de la gaussienne obtenue avec les fonctions de Cruiff et celle obtenue pour le meilleur ajustement avec la fonction de Voigt originelle est assignée comme source d'incertitude systématique.

Un exemple d'ajustement obtenu est visible pour la partie tonneau sur la figure 5.8 pour les données et simulations ; la valeur centrale des gaussiennes  $m_0$  y est aussi visible, une comparaison de cette valeur avec les valeurs  $s_{RECO}$  obtenues précédemment permet d'estimer l'incertitude associée au choix de la fonction d'ajustement à  $\sim 0,03$  % selon les régions considérées.

FIGURE 5.8 – Distribution des valeurs moyennes  $m_0$  des fonctions de Cruiff ajustées sur des pseudo-données (points noirs) et l'ajustement gaussien réalisé (courbes bleues) avec  $R_9 > 0.965$  dans EB pour les données (droit) et les simulations (gauche) *Ultra Legacy 2017*. Sont aussi indiqués la valeur de la moyenne  $m_0$  et la largeur  $\sigma$  de la gaussienne ainsi que le rapport du  $\chi^2$  et du nombre de degrés de liberté.



Les incertitudes systématiques totales sont alors calculées comme la somme quadratique des incertitudes liées au choix de l'intervalle d'ajustement et des incertitudes systématiques associées au choix de la fonction d'ajustement.

### 5.3.2 Résultats obtenus

#### Valeur de l'échelle d'énergie extraite

Les résultats de cette méthode sont lisibles dans le tableau 5.2 où les corrections et dégradations à l'échelle d'énergie ont été appliquées. Les valeurs de  $s_{RECO}$  y sont affichées avec leur incertitude statistique et leur incertitude systématique totale. Pour valider les corrections à l'échelle d'énergie, il faut que les valeurs de  $s_{RECO}$  extraites soient semblables, modulo les incertitudes, entre les données et les simulations. On observe que c'est clairement le cas dans

le tonneau puisque les valeurs extraites sont de  $-0,58$  pour les données et  $-0,57$  pour les simulations. Dans les bouchons, on obtient une valeur  $s_{RECO,d} = -0,59 \pm 0,08$  (stat)  $\pm 0,06$  (syst) pour les données et  $s_{RECO,s} = -0,34 \pm 0,11$  (stat)  $\pm 0,06$  (syst) donc des valeurs compatibles si l'on prend en compte les incertitudes à  $\pm 1\sigma$ .

TABLE 5.2 – Échelle d'énergie  $s_{RECO}$  obtenue à partir de la méthode d'ajustement, après avoir appliqué les corrections à l'échelle d'énergie et les dégradations de la résolution en énergie, dans le tonneau EB et les bouchons EE pour  $R_9 > 0,965$  pour les données (Data) et les simulations (DY) 2017 *Ultra Legacy* affectée de son incertitude statistique et de son incertitude systématique totale décomposée en ses différentes composantes étant l'incertitude sur le choix de l'intervalle d'ajustement et celle sur le choix de la fonction d'ajustement. Tous ces résultats sont donnés en %.

Valeur de l'échelle d'énergie, Incertitudes statistiques et systématiques avec corrections (en %)					
Data					
Catégories	$s_{RECO}$	Inc. Stat.	Inc. Intervalle	Inc. Fonction	Inc. Syst.
EB Haut $R_9$	-0.58	0.04	0.03	0.02	0.04
EE Haut $R_9$	-0.60	0.08	0.06	0.02	0.06
DY					
Catégories	$s_{RECO}$	Inc. Stat.	Inc. Intervalle	Inc. Fonction	Inc. Syst.
EB Haut $R_9$	-0.57	0.05	0.04	0.02	0.05
EE Haut $R_9$	-0.34	0.11	0.05	0.02	0.06

A noter qu'il est aussi possible de ne pas appliquer les corrections à l'échelle d'énergie et les dégradations de la résolution en énergie et ainsi de déterminer directement l'échelle d'énergie des photons pour les données et simulations par cette méthode, que l'on peut ensuite appliquer en guise de corrections. De telles valeurs sont présentées dans le tableau 5.3 et on peut facilement observer que les différences entre données et simulations sont plus importantes, surtout dans les bouchons, en absence de corrections.

TABLE 5.3 – Échelle d'énergie  $s_{RECO}$  obtenue à partir de la méthode d'ajustement, sans les corrections à l'échelle d'énergie et les dégradations de la résolution en énergie, dans le tonneau EB et les bouchons EE pour  $R_9 > 0,965$  pour les données (Data) et les simulations (DY) 2017 *Ultra Legacy* affectée de son incertitude statistique et de son incertitude systématique totale décomposée en ses différentes composantes étant l'incertitude sur le choix de l'intervalle d'ajustement et celle sur le choix de la fonction d'ajustement. Tous ces résultats sont donnés en %.

Valeur de l'échelle d'énergie, Incertitudes statistiques et systématiques sans correction (en %)					
Data					
Catégories	$s_{RECO}$	Inc. Stat.	Inc. Intervalle	Inc. Fonction	Inc. Syst.
EB High $R_9$	-0.52	0.04	0.03	0.02	0.04
EE High $R_9$	-1.02	0.07	0.06	0.04	0.07
DY					
Catégories	$s_{RECO}$	Inc. Stat.	Inc. Intervalle	Inc. Fonction	Inc. Syst.
EB Haut $R_9$	-0.57	0.05	0.03	0.02	0.04
EE Haut $R_9$	-0.43	0.11	0.05	0.02	0.05



### Test de fermeture

Afin de montrer la robustesse de la méthode utilisée, on réalise un test d'auto-fermeture qui consiste à appliquer les corrections à l'échelle d'énergie obtenues avec la méthode  $\mu\mu\gamma$  aux énergies des photons de tous les évènements et de refaire toute la procédure d'extraction de l'échelle d'énergie à partir de ces évènements corrigés. On s'attend alors à obtenir de nouvelles échelles d'énergie dont les valeurs seront compatibles, aux incertitudes près, avec zéro, démontrant alors que les corrections optimums ont été dérivées.

On observe bien dans le tableau 5.4 que les nouvelles valeurs de  $s_{RECO}^{new}$  sont statistiquement compatibles avec zéro et que notre méthode est bien robuste.

TABLE 5.4 – Tableau explicitant la nouvelle échelle d'énergie  $s_{RECO}^{new}$  obtenue à partir de la méthode d'ajustement après application des échelles d'énergie  $s_{RECO}^{old}$  obtenues précédemment. Ces résultats sont donnés dans le tonneau EB et les bouchons EE pour  $R_9 > 0,965$  pour les données (Data) et les simulations (DY) 2017 *Ultra Legacy* affectée de son incertitude statistique et de son incertitude systématique totale décomposée en ses différentes composantes étant l'incertitude sur le choix de l'intervalle d'ajustement et celle sur le choix de la fonction d'ajustement. Tous ces résultats sont donnés en %.

Valeur de l'échelle d'énergie, Incertitudes statistiques et systématiques sans correction (en %)						
Data						
Catégories	$s_{RECO}^{old}$	$s_{RECO}^{new}$	Inc. Stat.	Inc. Intervalle	Inc. Fonction	Inc. Syst.
EB Haut $R_9$	-0.52	-0.03	0.04	0.03	0.01	0.03
EE Haut $R_9$	-1.02	-0.02	0.08	0.06	0.01	0.06
DY						
Catégories	$s_{RECO}^{old}$	$s_{RECO}^{new}$	Inc. Stat.	Inc. Intervalle	Inc. Fonction	Inc. Syst.
EB Haut $R_9$	-0.57	0.01	0.05	0.04	0.01	0.04
EE Haut $R_9$	-0.43	0.07	0.11	0.05	0.01	0.05

### Conclusion

Nous avons montré dans ce chapitre l'importance des évènements  $Z \rightarrow \mu\mu\gamma$  : ces évènements ont une statistique moindre comparée à celle des évènements  $Z \rightarrow ee$  mais ils sont néanmoins une des seules sources de photons dont la cinématique est bien définie. Grâce à des coupures de sélections ingénieuses, on obtient des échantillons photons de grande pureté permettant, notamment, de calculer, et donc de valider, l'échelle d'énergie des photons.

En ce qui concerne l'échelle d'énergie des photons, une méthode de dérivation de cette quantité a été présentée, basée sur l'ajustement d'un estimateur utilisant des informations cinématiques mesurées sur des évènements  $Z \rightarrow \mu\mu\gamma$ . Cette méthode est robuste et a permis de valider les corrections à l'échelle d'énergie et les dégradations de la résolution en énergie pour les données et les simulations 2017 *Ultra Legacy* avec des différences à  $\pm 1\sigma$  pour des collisions proton-proton d'énergie dans le centre de masse de 13 TeV et pour une luminosité intégrée d'environ  $41,5 \text{ fb}^{-1}$ . Ces résultats seront incorporés dans un article présentant différentes études et calibrations du ECAL.

Ce travail a été repris par d'autres doctorants qui auront à charge de traiter les données et simulations correspondant à *Ultra Legacy 2016* ainsi qu'à préparer, configurer et améliorer la méthode d'extraction de l'échelle d'énergie pour le Run 3 du LHC.

Pour finir, ces évènements  $Z \rightarrow \mu\mu\gamma$  sont aussi utiles pour calculer les efficacités et valider différentes étapes des analyses utilisant des photons comme nous le verrons dans le chapitre suivant.

# Analyse pour la recherche de $h \rightarrow \gamma\gamma$ à basses masses

## Sommaire

---

6.1	Stratégie de l'analyse . . . . .	109
6.1.1	Stratégie générale . . . . .	109
6.1.2	Stratégie détaillée . . . . .	111
6.2	Échantillons utilisés . . . . .	113
6.2.1	Données . . . . .	113
6.2.2	Simulations . . . . .	113
	Signal . . . . .	113
	Bruit de fond . . . . .	114
6.3	Chemin de déclenchement et efficacités . . . . .	114
6.3.1	Description des chemins HLT . . . . .	114
6.3.2	Efficacités des chemins de déclenchement . . . . .	117
	Méthode Tag and Probe . . . . .	117
	Efficacités de déclenchement . . . . .	118
6.4	Sélection de l'analyse . . . . .	119
6.4.1	Critères de présélection . . . . .	119
6.4.2	Efficacités de présélection . . . . .	121
	Mesure de l'efficacité de présélection avec avec les évènements $Z \rightarrow ee$ . . . . .	121
	Mesure de l'efficacité de présélection avec avec les évènements $Z \rightarrow \mu\mu\gamma$ . . . . .	121
6.5	Identification des vertex diphoton . . . . .	123
6.5.1	BDT d'identification de vertex . . . . .	124
6.5.2	Probabilité d'estimer le bon vertex . . . . .	125
6.5.3	Validation et facteurs d'échelle . . . . .	125
6.6	Identification des photons . . . . .	126
6.6.1	Photon ID BDT . . . . .	127
6.6.2	Validation du photon ID BDT . . . . .	130

	Validation avec les évènements $Z \rightarrow ee$ . . . . .	130
	Validation avec les évènements $Z \rightarrow \mu\mu\gamma$ . . . . .	130
6.7	Classification des évènements . . . . .	<b>134</b>
6.7.1	Classificateur d'évènements : le diphoton BDT . . . . .	134
	Le diphoton BDT . . . . .	134
	Validation entre données et simulations . . . . .	138
	Validation avec les évènements $Z \rightarrow ee$ . . . . .	140
6.7.2	Catégorisations . . . . .	143
6.8	Modélisation du signal . . . . .	<b>147</b>
6.9	Modélisation du bruit de fond . . . . .	<b>156</b>
6.9.1	Modélisation de la composante Drell-Yan . . . . .	156
6.9.2	Modèle de bruit de fond total . . . . .	160
6.9.3	Incertitudes systématiques associées . . . . .	164
6.10	Incertitudes systématiques . . . . .	<b>170</b>
6.10.1	Listes des incertitudes systématiques pour le signal . . . . .	170
	Incertitudes théoriques . . . . .	171
	Incertitudes expérimentales . . . . .	171
	Incertitudes liées à la classe VBF . . . . .	172
6.10.2	Impacts des incertitudes . . . . .	173
6.11	Résultats finaux . . . . .	<b>178</b>
6.11.1	Analyse statistique . . . . .	178
	Méthode d'ajustement par maximisation de la fonction de vraisemblance	178
	Limite d'exclusion en l'absence de signal . . . . .	180
	Limites attendues et approximation asymptotique . . . . .	181
	Quantification d'un excès d'évènements . . . . .	182
6.11.2	Résultats des tests statistiques . . . . .	183
6.12	Résultats avec catégorie VBF . . . . .	<b>184</b>
6.12.1	Catégorie VBF . . . . .	184
6.12.2	Résultats finaux . . . . .	187
6.12.3	Comparaison avec ATLAS . . . . .	191

---

Dans ce chapitre est présenté l'une des parties de mon travail effectué lors de ce doctorat à travers la recherche d'un boson de Higgs de basse masse ( $<110$  GeV) se désintégrant en deux photons dans le détecteur CMS. Une telle recherche est motivée par la possible existence de bosons de Higgs additionnels dans les théories BSM qui peuvent permettre, comme décrites dans les chapitres 2 et 8, à de tels bosons d'avoir une masse inférieure à la masse du boson de Higgs du MS qui est d'environ 125 GeV.

La stratégie de l'analyse sera décrite dans la section suivante mais elle consiste brièvement, à mettre en évidence un éventuel excès d'évènements diphoton correspondant à un boson de Higgs *SM-like* à travers un bruit de fond décroissant de manière quasi-exponentielle étendant alors la méthode utilisée dans CMS pour observer et mesurer les propriétés du boson de Higgs du MS [100, 101, 102]. Cette méthode a été adaptée pour la recherche à basse masse.

La méthode est alors similaire à celle employée pour traiter les données 2012 (Run 1) et 2016 (Run 2) [103] mais de nombreuses modifications, améliorations et optimisations y ont été apportées. L'analyse présentée ici utilise les données 2017 et 2018 *ReReco* du Run 2 correspondant à une énergie dans le centre de masse proton-proton de 13 TeV et une luminosité intégrée de, respectivement, environ  $41,5 \text{ fb}^{-1}$  et  $54,4 \text{ fb}^{-1}$ . L'analyse est décrite en détail dans [104].

## 6.1 Stratégie de l'analyse

### 6.1.1 Stratégie générale

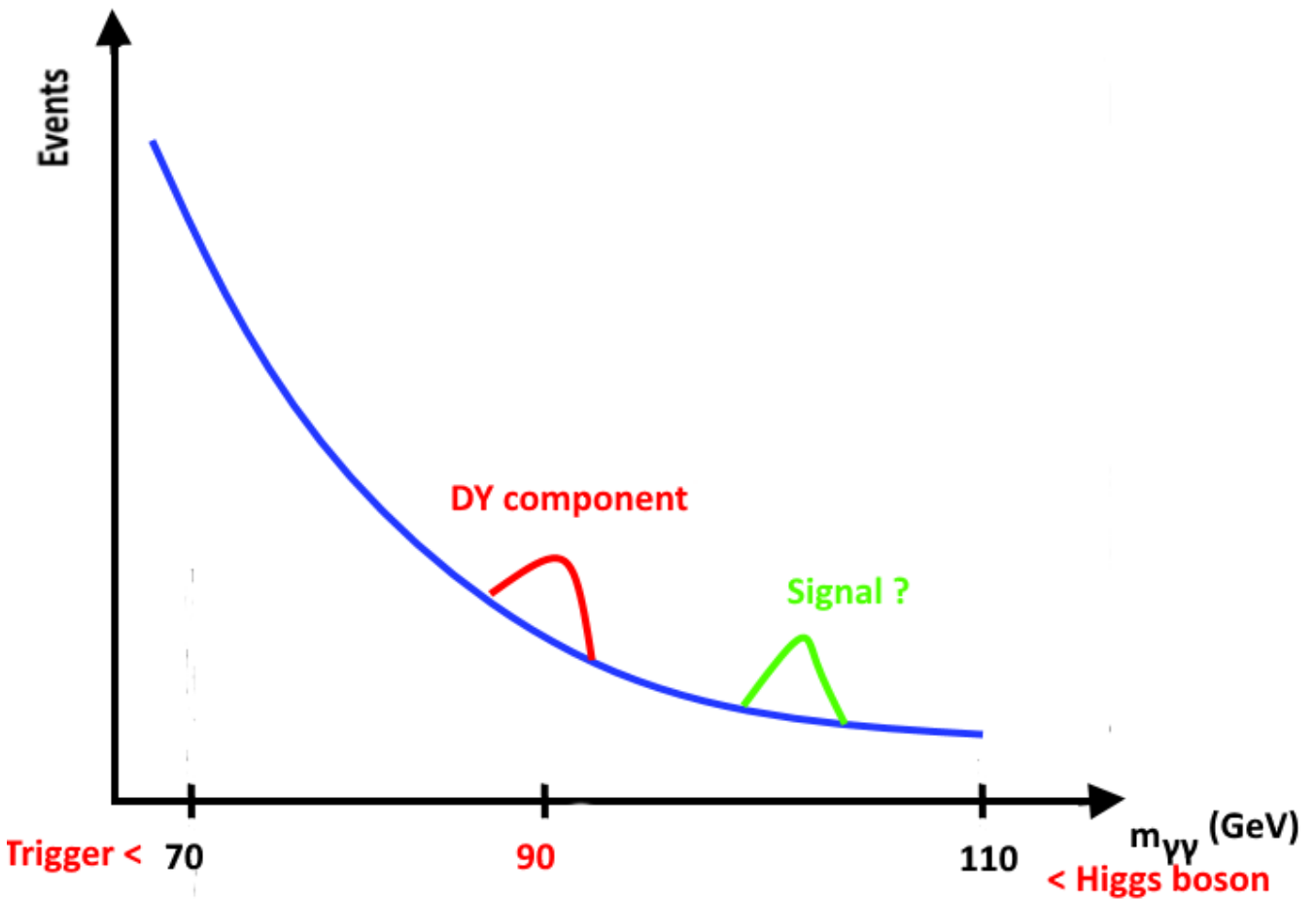
On recherche donc une résonance étroite, de largeur bien inférieure à la résolution du détecteur, correspondant à une nouvelle particule neutre qui se désintégrerait en deux photons. Si la désintégration en deux photons ne présente pas le rapport de branchement le plus élevé dans le cas MS, comme on l'observe sur la figure 1.8, elle a cependant l'avantage de présenter une signature claire de deux photons isolés de haute énergie transverse rendant son observation plus aisée. Un autre avantage est que les photons sont reconstruits avec une excellente résolution dans CMS, comme étudié dans le chapitre 5, permettant ainsi une bonne résolution de la masse invariante diphoton  $m_{\gamma\gamma} = \sqrt{E_{\gamma,1}E_{\gamma,2}(1 - \cos\Delta\alpha)}$  et donc une bonne reconstruction d'un éventuel signal observé.

La masse invariante diphoton est en effet une quantité remarquable puisque notre objectif est d'observer la présence, ou l'absence, d'un excès d'évènements diphoton, dit signal, pouvant correspondre à une nouvelle particule par rapport à un bruit de fond diphoton décroissant. Pour ce faire, il est nécessaire de paramétriser le bruit de fond par une fonction rendant compte de la décroissance observée ainsi que de modéliser la forme du signal recherché. L'hypothèse faite concernant le signal est que l'on cherche un boson de Higgs de type *SM-like*; autrement dit les caractéristiques, notamment la section-efficace et le rapport de branchement, de la particule recherchée sont ceux, présentées dans la section 1.2.2, d'un boson de Higgs du MS mais variant avec sa masse. Cela permet une modélisation relativement simple du signal tout

en étant relativement indépendant du modèle considéré puisque facilement réinterprétable pour des modèles plus spécifiques.

Le domaine actuel en masse de cette recherche est défini entre 70 GeV et 110 GeV : la présence du système de déclenchement, pour les masses inférieures à 70 GeV, et celle du boson de Higgs SM à 125 GeV, pour les masses supérieures, produisent des effets sur la forme de bruit de fond qui seraient difficiles à inclure dans la modélisation qui en est faite. Ce domaine en masse présente cependant la particularité de contenir la masse du boson Z ; les désintégrations leptoniques du boson Z, la composante Drell-Yan, vont alors devoir être prises en compte dans notre bruit de fond puisque les électrons émis par des désintégrations  $Z \rightarrow ee$  peuvent être identifiés à tort comme des photons et ainsi créer une distribution piquée à la masse du Z. Il a, par conséquent, été choisi de modéliser cette composante Drell-Yan à part entière afin de ne pas prendre le risque de l'identifier à tort comme un signal. La stratégie de l'analyse est résumée dans la figure 6.1.

FIGURE 6.1 – Résumé de la stratégie de l'analyse. En bleu, le bruit de photon diphoton, en rouge la composante Drell-Yan  $Z \rightarrow ee$  où les  $e$  sont identifiés à tort comme des photons et en vert la présence éventuel d'un pic de signal  $\gamma\gamma$  étroit.



Avec un peu plus de détails, le modèle du signal recherché, l'éventuel excès d'évènements, est construit à partir des simulations des quatre modes de productions principaux (ggH, VBFH,

ttH et VH), présentés dans la section 1.2.2, d'un boson de Higgs du Modèle Standard avec des masses variant de 65 GeV à 115 GeV.

Le continuum diphoton, le bruit de fond de l'analyse correspondant aux évènements avec deux candidats photons dans l'état final, est quant à lui composé d'une partie irréductible et d'une partie réductible dont les principales contributions sont :

- Des processus du Modèle Standard contenant deux photons prompts dans l'état final pour la partie irréductible. Ils proviennent en majorité de radiations de quarks comme observé sur la figure 6.2.
- Des processus QCD avec deux jets, et avec un photon et un jet où le jet contient un photon (méson neutre se désintégrant en deux photons très boostés donc indiscernables par exemple) contribuant à la partie réductible.
- Contribution réductible Drell-Yan  $Z \rightarrow ee$  où les électrons sont reconstruits à tort comme des photons.

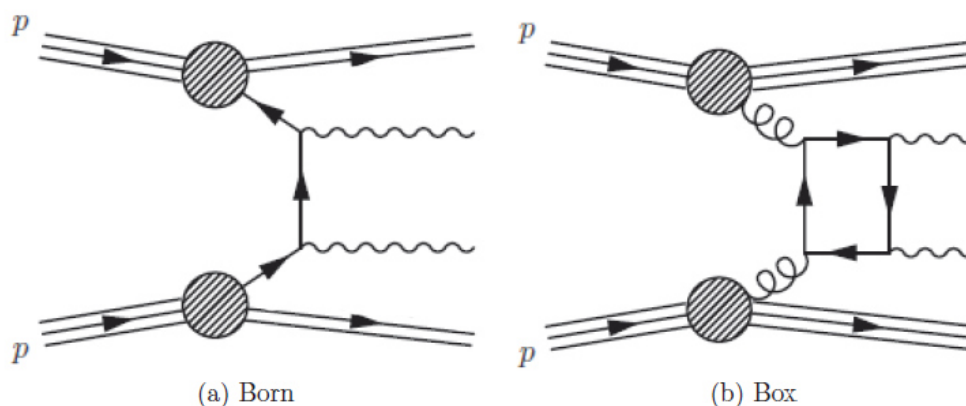


FIGURE 6.2 – Processus du Modèle Standard Born a) et Box b) contribuant à la composante de bruit de fond irréductible de l'analyse.

Afin de maximiser la sensibilité de l'analyse à la présence d'éventuels évènements de signal par rapport aux nombreux évènements de bruit de fond attendu, il est nécessaire de reconstruire les évènements diphoton de façon optimale afin de pouvoir modéliser la composante du continuum diphoton, de la composante Drell-Yan et celle du signal de façon la plus précise possible.

### 6.1.2 Stratégie détaillée

Les points de départ de l'analyse sont les échantillons de données et de simulations choisis, détaillés dans la section suivante, représentant les évènements issus de collisions proton-proton dans CMS pour un chemin de déclenchement adapté à la basse-masse et reconstruits comme des candidats photons<sup>1</sup> d'après les chapitres précédents. On applique à ces évènements quelques [corrections supplémentaires](#) relatives aux analyses  $H \rightarrow \gamma\gamma$  [105] recalculées pour

1. A cette étape, ces candidats peuvent très bien correspondre à des électrons par exemple comme étudié dans le chapitre 4.

notre analyse incluant les efficacités de déclenchement et des corrections spécifiques de forme de gerbe électromagnétique.

On applique une série de coupures, nommée **présélection**, sur différentes variables de nos évènements reconstruits afin de rejeter une première partie du bruit de fond en utilisant, par exemple, un critère d'électron veto permettant de rejeter la majeure partie des électrons de l'analyse. Cette partie concerne, en partie, le travail de collègues travaillant sur l'analyse.

Il est ensuite nécessaire de reconstruire précisément la masse invariante diphoton  $m_{\gamma\gamma}$ . Or celle-ci dépend, en plus des énergies des photons déjà corrigées, de l'angle entre les deux photons reconstruits. Il devient alors primordial de reconstruire efficacement le **vertex d'interaction diphoton** afin de connaître cet angle de manière précise. On utilise pour ce faire deux arbres de décision boostés (ou BDT présenté dans l'annexe A) permettant in-fine une meilleure identification de ce vertex.

A cette étape, il est toujours difficile de séparer les vrais photons de photons mal identifiés (jets, électrons ou photons non résolus issus de la désintégration de mésons). On utilise alors un BDT, le **Photon ID BDT**, permettant **d'identifier les photons prompts**. On applique alors à la sortie de ce BDT une coupure pour rejeter une part importante du bruit de fond, notamment du bruit de fond QCD. Ce BDT a été spécialement entraîné pour l'analyse, ce qui représente une nouveauté par rapport à [103]. Cette partie concerne principalement le travail de collègues travaillant sur l'analyse.

Les évènements sont ensuite **classifiés** selon leur note de sortie d'un nouveau BDT, le **diphoton BDT**, qui permet de discriminer les évènements. Les évènements avec une note faible sont rejetés et les autres sont répartis au sein de catégories selon leur proximité avec les évènements signal. Ce BDT a été spécialement entraîné pour l'analyse de même que la classification, ce qui représente une nouveauté par rapport à [103]. Cette partie est le fruit de mon propre travail. Une autre classification est ensuite effectuée dans le but d'étiqueter les évènements ressemblant à des évènements produits par le processus de fusion de bosons vecteurs (VBF). Cette partie concerne principalement le travail de collègues travaillant sur l'analyse et ne sera qu'abordée.

On est alors en mesure de modéliser, pour chaque catégorie ainsi définie, le **signal** en utilisant des simulations MC.

Le bruit de fond est lui aussi modélisé, pour chaque catégorie, en commençant par la paramétrisation de la composante Drell-Yan via des simulations MC. La composante ainsi modélisée est ajoutée, avec une normalisation flottante, dans un modèle de **paramétrisation générale du continuum diphoton** s'effectuant directement sur les données. On calcule finalement les **incertitudes systématiques** introduites lors de la chaîne d'analyse avant d'effectuer des **tests statistiques** permettant de statuer sur la présence ou l'absence d'un éventuel excès d'évènements signal dans les données. Les résultats alors obtenus peuvent ensuite être interprétés dans le cadre de différents modèles théoriques. Tous ces éléments de l'analyse sont le fruit de mon propre travail.



Toutes ces parties seront détaillées dans la suite. A noter cependant que la plupart des BDT utilisés dans l'analyse ont été entraînés pour les données et simulations de 2017, et que cet entraînement est aussi utilisé sur les données et simulations de 2018.

### 6.2 Échantillons utilisés

Les échantillons utilisés dans cet analyse seront brièvement décrits dans cette section. Une liste complète, mais plus technique, de ces échantillons est néanmoins visible dans l'annexe C.

#### 6.2.1 Données

Les échantillons utilisés pour les données sont les échantillons électrons/photons pour l'analyse principale, di-muon pour les études d'efficacités, et électrons pour la validation des BDT et la détermination de certaines incertitudes systématiques, correspondant aux échantillons **DoubleEG** (collection d'évènements avec au moins deux électrons/photons ou assimilés), **DoubleMuon** (collection d'évènement avec au moins deux muons) , **SingleElectron** collection d'évènement avec au moins un électron) pour les runs B à F en 2017, et **EGamma**, **DoubleMuon**, **SingleElectron** pour les runs A à D en 2018. Ces échantillons doivent satisfaire certains critères de qualité déterminés par le *Data Quality Monitoring* de CMS induisant une baisse effective de la luminosité intégrée totale par rapport à la luminosité délivrée.

Pour 2017, ces données **DoubleEG** correspondent à une luminosité intégrée de  $41,5 \text{ fb}^{-1}$ , pour une énergie de collision proton-proton dans le centre de masse de 13 TeV, pour le système de déclenchement, présenté dans la prochaine section, utilisé. Pour 2018, cette luminosité intégrée est de  $54,4 \text{ fb}^{-1}$  ; elle est inférieure aux  $59,7 \text{ fb}^{-1}$  de la prise de données totale mais le chemin de déclenchement HLT n'était pas en place au début de la prise de données résultant en une perte d'évènements alors non utilisables.

#### 6.2.2 Simulations

Les simulations sont obtenues en générant des évènements avec PYTHIA (v8) [80], MadGraph (@NLO v2.4) [79] pour le calcul des éléments de matrice interfacé ensuite avec PYTHIA, ou encore SHERPA (v2.2) [106], avant de propager ces évènements dans une simulation complète du détecteur CMS utilisant GEANT4 [81]. On applique un poids de correction aux évènements de ces simulations pour que la distribution de l'empilement corresponde à celle observée dans les données.

### Signal

Les échantillons de signal des quatre modes de production principaux du boson de Higgs pour des masses allant de 60 GeV à 115 GeV par pas de 5 GeV sont générés avec MadGraph interfacé avec PYTHIA pour les niveaux partonique et hadronique respectivement. Les sections

efficaces et rapports de branchement utilisés sont ceux recommandés par le *LHC Cross-Section Working Group* [25].

## Bruit de fond

Le bruit de fond irréductible de deux photons prompts est généré avec SHERPA et contient les processus *Born* et *Box* de la figure 6.2 à l'ordre dominant, et sa section efficace est multipliée par un facteur 1,3 pour tenir compte de corrections non présentes dans SHERPA.

Les bruits de fond réductibles QCD et  $\gamma$ +jets sont générés avec PYTHIA et sont enrichis en états finaux contenant des particules susceptibles d'être mal reconstruites ou identifiées comme des photons.

La composante réductible Drell-Yan est également générée avec MadGraph et PYTHIA.

## 6.3 Chemin de déclenchement et efficacités

### 6.3.1 Description des chemins HLT

Les chemins HLT actifs permettant de sélectionner, parmi la multitude d'interactions au sein du détecteur, les données photon de l'analyse sont HLT\_Diphoton30PV\_18PV\_R9Id\_AND\_IsoCaloId\_AND\_HE\_R9Id\_PixelVeto\_Mass55 pour 2017 et HLT\_Diphoton30\_18\_R9IdL\_AND\_HE\_AN\_IsoCaloId\_NoPixelVeto pour 2018 ; ces deux chemins ont été spécialement développés pour cette analyse à basse masse et permettent d'explorer des domaines de masses inférieures à ceux utilisés dans les analyses mesurant les propriétés du boson de Higgs du MS [90]. Le chemin de 2017 est basé sur une logique, très complexe, similaire à celui employé pour traiter les données 2016 [103] mais avec quelques simplifications permettant de garder une bande-passante raisonnable malgré l'augmentation de luminosité entre 2016 et 2017. La logique du chemin 2018 est, quant à elle, différente ; plus simple, elle permet, en gardant la même efficacité que pour 2017, de réduire la bande-passante par un facteur 2 voire 3 [104] ce qui a permis de retirer le *pixel veto*, utilisé pour rejeter les leptons, et la coupure en masse à 55 GeV du chemin HLT.

Ce dernier point est primordial puisque cette coupure en masse au sein du chemin de déclenchement est, avec les seuils en impulsion transverse, ce qui limite la borne inférieure en masse de l'analyse. Une coupure  $m_{\gamma\gamma} > 55$  GeV, comme en 2017, crée, avec les coupures en impulsion transverse, un effet d'allumage ou *turn-on* dans la distribution  $m_{\gamma\gamma}$  (un exemple est visible sur la figure 7.3 du chapitre suivant) qui est, en effet, extrêmement difficile à modéliser. On doit alors attendre la fin de ce phénomène (une petite dizaine de GeV plus tard) pour avoir une distribution en masse invariante stable. Le fait d'avoir réussi à retirer cette coupure du chemin HLT est donc une étape clé dans la recherche de nouvelles particules de basses masses ; ces implications, aussi intéressantes soient-elles, dépassent cependant le cadre de ce manuscrit.

Les chemins sont organisés en deux parties comme on peut le voir sur les figures 6.3 pour 2017 et 6.4 pour 2018. La première partie est appelée *Seeded leg*, elle sélectionne, par

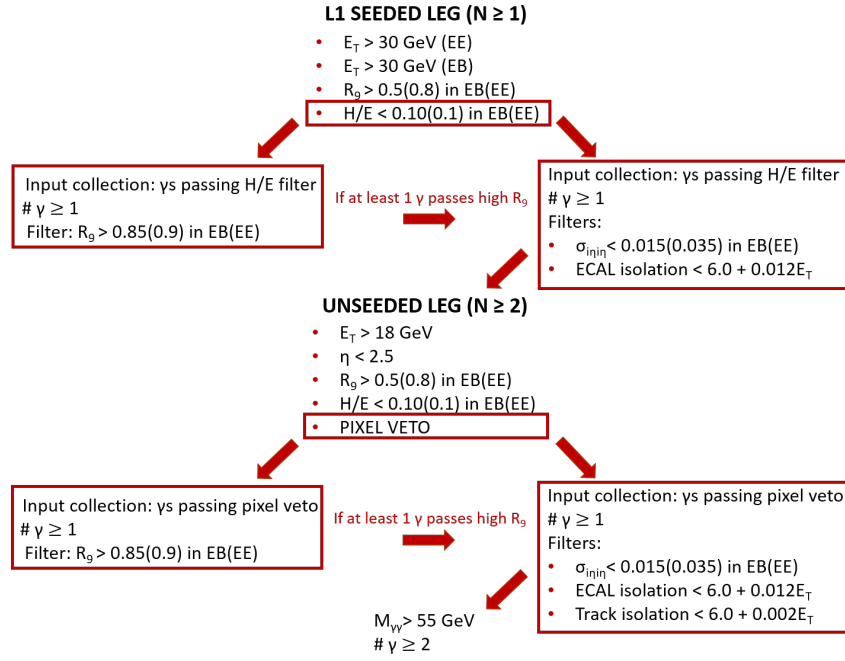


FIGURE 6.3 – Schéma illustrant la logique, position et liste des filtres du chemin HLT pour 2017. A noter [104] que le filtre  $\sigma_{i\eta i\eta}$  et isolation du ECAL n'est pris en compte que si au moins un photon, en principe ça peut être un photon différent, passe le filtre haut  $R_9$  (sinon l'évènement est rejeté).

évènement<sup>2</sup>, au moins un candidat photon avec  $E_T > 30$  GeV correspondant à un candidat électromagnétique du système de déclenchement L1 [104]. L'agrégation des objets ECAL effectuée par les algorithmes HLT à partir de la graine d'un candidat permet, à la suite, pour la seconde partie dite *Unseeded leg*, de sélectionner, par évènement, au moins deux photons. Cette sélection s'effectue, après l'agrégation avec  $E_T > 18$  GeV, sans critère de correspondance avec des candidats L1. Cela signifie qu'un des deux photons passant la partie *Unseeded leg* a de fortes chances d'être le candidat photon du L1 ayant allumé le déclenchement de la partie *Seeded leg*. Des filtres de sélections et coupures sont appliqués dans chacune de ces parties du chemin HLT sur des critères de qualité généraux et d'isolation comme résumé dans les tableaux 6.1 pour 2017 et 6.2 pour 2018. Pour 2017, ces filtres incluent aussi une coupure à haut  $R_9$ , un pixel veto permettant de rejeter des électrons qui pourraient être sélectionnés comme photons à cette étape, et une coupure sur la masse de l'objet diphoton  $m_{\gamma\gamma} > 55$  GeV.

La logique du chemin de déclenchement utilisé en 2017 est complexe notamment à cause du fait que des photons différents peuvent passer le filtre de haut  $R_9$ , et le filtre  $\sigma_{i\eta i\eta}$  et  $I_\gamma$  (voir l'annexe B pour plus de détails sur ces variables) pour la partie *Unseeded leg*. On peut ainsi avoir, par exemple, deux photons qui passent le filtre de haut  $R_9$  ainsi que le filtre  $\sigma_{i\eta i\eta}$  et  $I_\gamma$ , ou le cas où un photon passe le filtre de haut  $R_9$  et le deuxième passe uniquement le filtre  $\sigma_{i\eta i\eta}$  et  $I_\gamma$ , ou encore le cas où deux photons passent le filtre  $\sigma_{i\eta i\eta}$  et  $I_\gamma$  mais un seul passe le filtre de haut  $R_9$ . On requiert donc la présence d'au moins un photon de haut  $R_9$  en 2017 alors

2. On rappelle qu'un évènement correspond aux interactions physiques mesurées au sein du détecteur pour une collision proton-proton donnée.

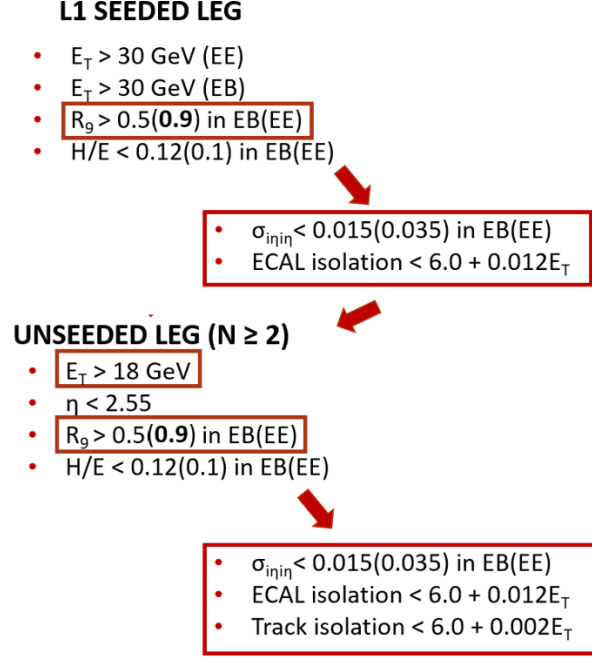


FIGURE 6.4 – Schéma illustrant la logique, position et liste des filtres du chemin HLT pour 2018.

TABLE 6.1 – Description du chemin HLT\_Diphoton30PV\_18PV\_R9Id\_AND\_IsoCaloId\_AND\_HE\_R9Id\_PixelVeto\_Mass55 utilisé en 2017 : l'objet HLT provenant d'un objet L1 se voit appliquer des coupures en  $E_T$ , puis bas  $R_9$  et  $H/E$ . La coupure en haut  $R_9$ , et les sélections en  $\sigma_{i\eta i\eta}$  (CaloId) et  $I_\gamma$  (ECAL iso) sont ensuite appliquées suivant la logique décrite sur la figure 6.3. Une fois l'agrégation des objets ECAL effectuée par les algorithmes HLT, les mêmes critères de sélection, plus une coupure sur  $I_{track}$  (Track iso), sont appliqués aux objets diphoton *unseeded*. Le pixel veto est présent dans cette deuxième partie, et une coupure sur la masse du système diphoton est finalement appliquée. Les variables utilisées sont présentées plus en détail dans l'annexe B (la variable  $R_9$  présente ici est  $R_{9,5 \times 5}$ ).

	$E_T$	Low $R_9 >$	$H/E <$	High $R_9 >$	$\sigma_{i\eta i\eta} <$	ECAL PF cluster iso $<$	Track iso $<$
<b>Seeded leg</b>							
Photons in EB	30 GeV	0.5	0.1	0.85	0.015	$6.0 + 0.12E_T$	-
Photons in EE	30 GeV	0.8	0.1	0.9	0.035	$6.0 + 0.12E_T$	-
<b>Unseeded leg</b>							
Photons in EB	18 GeV	0.5	0.1	0.85	0.015	$6.0 + 0.12E_T$	$6+0.002E_T$
Photons in EE	18 GeV	0.8	0.1	0.9	0.035	$6.0 + 0.12E_T$	$6+0.002E_T$
<b>Other trigger requirements</b>							
Pixel veto				$m_{\gamma\gamma} > 55$ GeV			

qu'en 2018 on sélectionne, de manière plus lâche, les photons ayant un bas  $R_9$  ( $R_9 > 0,5$ ), s'ils se situent dans EB, et ayant un haut  $R_9$  ( $R_9 > 0,9$ ), s'ils se situent dans EE.

A noter qu'en 2017 (et en 2016), un problème, nommé *prefiring*, dans l'alignement temporel des primitives de déclenchement L1 du ECAL dans des régions de haut  $\eta$  a fait en sorte que certaines primitives de déclenchement ont été associées à tort au croisement de faisceaux précédent. Les règles de niveau L1 ne permettant pas à deux croisements de faisceaux consécutifs de s'allumer, une partie non négligeable des événements dans les régions  $2 < |\eta| < 3$  a été

TABLE 6.2 – Description du chemin HLT\_Diphoton30\_18\_R9IdL\_AND\_HE\_AN\_IsoCaloId\_NoPixelVeto utilisé en 2018 : l’objet HLT provenant d’un objet L1 se voit appliquer des coupures en  $E_T$ , puis bas  $R_9$  et  $H/E$ . La coupure en haut  $R_9$ , et les sélections en  $\sigma_{i\eta j\eta}$  (CaloId) et  $I_\gamma$  (ECAL iso) sont ensuite appliquées. Une fois l’agrégation des objets ECAL effectuée par les algorithmes HLT, les mêmes critères de sélection, plus une coupure sur  $I_{track}$  (Track iso), sont appliqués aux objets diphoton *unseeded*. Les variables utilisées sont présentées plus en détail dans l’annexe B (le  $R_9$  présent ici est  $R_{9,5\times 5}$ ).

	$E_T$	$R_9 >$	$H/E <$	$\sigma_{i\eta j\eta} <$	ECAL iso <	Track iso <
<b>Seeded leg</b>						
Photons in EB	30 GeV	0.5	0.1	0.015	$6.0 + 0.12E_T$	-
Photons in EE	30 GeV	0.9	0.1	0.035	$6.0 + 0.12E_T$	-
<b>Unseeded leg</b>						
Photons in EB	18 GeV	0.5	0.1	0.015	$6.0 + 0.12E_T$	$6+0.002E_T$
Photons in EE	18 GeV	0.9	0.1	0.035	$6.0 + 0.12E_T$	$6+0.002E_T$

perdue. Cet effet n’étant pas décrit dans les simulations, on calcule pour ces dernières des poids associés à la probabilité pour un évènement de déclencher sur le croisement de faisceaux précédent, et une incertitude systématique associée sera considérée dans la suite. Cet effet n’affecte pas 2018.

### 6.3.2 Efficacités des chemins de déclenchement

Il est important de connaître le nombre d’évènements signal éventuellement rejeté par les chemins HLT choisis pour déterminer la sensibilité au signal de l’analyse et puisque l’objectif de ces chemins est de maximiser ce nombre d’évènements afin d’augmenter notre probabilité de détecter un nouveau boson de Higgs tout en réduisant la bande-passante. Pour ce faire, on calcule l’efficacité, idéalement définie comme le rapport du nombre d’évènements passant les filtres HLT de notre analyse sur le nombre d’évènements totaux, des chemins HLT pour des évènements proches du signal mesurant les performances de nos chemins HLT. Ce calcul est effectué grâce à une méthode de *Tag and Probe* [105] basée sur des données  $Z \rightarrow ee$  étant donné que l’on n’utilise pas de simulation du système déclenchement. Ces efficacités seront ensuite directement appliquées aux simulations MC d’évènements signal.

#### Méthode Tag and Probe

La méthode *Tag and Probe* (TnP) est une méthode générale permettant de mesurer directement des efficacités sur les données, même si elle peut tout à fait être utilisée sur des simulations. Cette méthode est basée sur l’utilisation d’une résonance en masse (comme le  $Z$ , ou les mésons  $J/\Psi$  ou  $Y$ ) pour sélectionner certaines particules et déterminer in-fine les critères de sélection qui ont été appliqués à ces particules. On nomme "Tag" l’objet (en général muon ou électron) passant une série de coupures et sélections très resserrées permettant d’isoler cet objet et de déclencher sur l’évènement correspondant. On sélectionne ensuite les "Probes" en appariant les objets, avec des critères de sélection bien plus lâches, avec des "Tags" afin d’obtenir une masse invariante combinée compatible avec celle de la résonance choisie. On élimine finalement le bruit de fond avec des techniques diverses comme un ajustement

spécifique.

La méthode TnP est ainsi très proche, bien que plus simple, de la méthode utilisant les événements  $Z \rightarrow \mu\mu\gamma$  étudiée dans le chapitre précédent. Pour les analyses  $H \rightarrow \gamma\gamma$  les événements  $Z \rightarrow ee$  sont utilisés pour la méthode TnP ; ces événements sont, en effet, proches des événements signal puisque les électrons ressemblent à des photons et la cinématique de la désintégration résonante est similaire au vu de la masse du  $Z$ .

Pour sélectionner des événements  $Z \rightarrow ee$  de haute pureté, les événements doivent être sélectionnés par un chemin HLT *Single Electron*, correspondant aux jeux de données du même nom, ne requérant explicitement qu'un seul électron afin que l'autre électron issu de la désintégration ne soit pas biaisé et puisse être utilisé pour mesurer des efficacités. Les coupures de présélection de l'analyse, explicitées dans la section 6.4, sont appliquées à la fois aux "tags" et aux "probes", en inversant le critère de veto électron afin de pouvoir sélectionner ces derniers. Le bruit de fond obtenu avec ces événements est faible et est éliminé grâce à un ajustement précis de la forme de la résonance.

La méthode TnP est notamment utilisée ici pour calculer des facteurs d'échelle, pour valider les variables d'entrée du *Photon ID BDT* (voir la section 6.6) ainsi que pour calculer les efficacités des chemins de déclenchement.

### Efficacités de déclenchement

Les efficacités de déclenchement sont calculées comme le rapport entre le nombre de "probes" associées passant les conditions HLT sur le nombre de "probes" associées total. Elles sont calculées pour différentes régions en  $\eta$  (EB et EE), en  $R_9$  et en  $p_T$  de la "probe", et séparément pour les parties "seeded" and "unseeded" avant d'être multipliées pour obtenir des valeurs par événement<sup>3</sup>. Les "tags" doivent, en plus des conditions évoquées précédemment, satisfaire  $|\eta| < 2,1$  (mais pas  $1,4442 \leq |\eta| \leq 1,566$ ) et  $p_T > 40$  GeV ; pour les "probes" on requiert  $|\eta| < 2,5$  et  $p_T > 30$  GeV, dans le cas *seeded*, ou  $p_T > 18$  GeV, dans le cas *unseeded*. Comme les électrons, provenant de désintégrations du  $Z$  utilisées dans la méthode TnP, ont des formes de gerbes électromagnétiques et des distributions angulaires différentes de celles des photons, provenant d'éventuelles désintégrations d'un boson de Higgs, on effectue une repondération des événements  $Z \rightarrow ee$  en  $R_9$  et  $\eta$  afin que ces distributions correspondent à celles de la désintégration en deux photons d'un boson de Higgs de masse 90 GeV. On compare pour cela les distributions  $R_9$  et  $\eta$  de simulations Drell-Yan  $Z \rightarrow ee$  et  $H \rightarrow \gamma\gamma$  à 90 GeV pour le mode de production ggH.

Le nombre de "probes" passant ou non les conditions HLT est calculé par une méthode de comptage pour 2017 dans laquelle on fait l'hypothèse que le bruit de fond est négligeable. Pour 2018, une méthode plus précise est utilisée qui consiste à ajuster une fonction composée

3. L'objet TnP (*sub*)*leading* reconstruit est ensuite identifié avec la partie (*un*)*seeded* pour obtenir la valeur par événement.

d'une fonction de forme spécifique convoluée avec une gaussienne pour la partie signal et d'une fonction erreur plus une exponentielle décroissante pour le bruit de fond. Des exemples d'efficacités sont visibles sur la figure 6.5 et plus de détails pourront être lus avec profit dans [104]. Ces efficacités sont données après une série de coupures effectuées après reconstruction des photons (*offline*) décrite dans la section suivante.

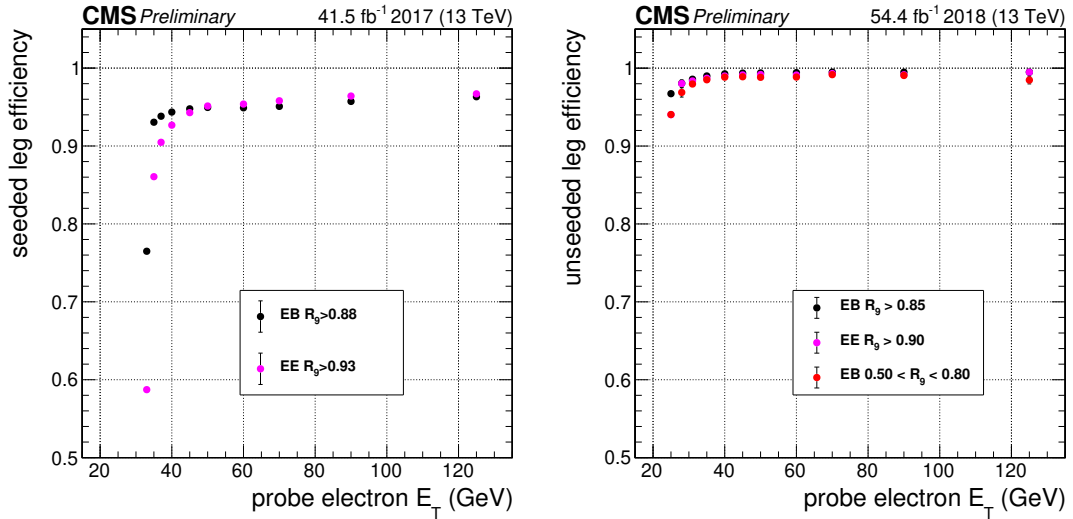


FIGURE 6.5 – Efficacités de déclenchement pour les photons mesurées pour des données  $Z \rightarrow ee$  avec la méthode TnP pour la partie *seeded* pour 2017 (gauche) et la partie *unseeded* pour 2018 (droite).

J'ai aussi contribué à ce travail sur les chemins de déclenchement en calculant les efficacités de déclenchement pour la réanalyse des données 2016 *Legacy* qui ne sera cependant pas présentée dans ce manuscrit.

## 6.4 Sélection de l'analyse

Les chemins HLT décrits dans la section précédente ne sont pas appliqués aux simulations MC. Afin d'avoir le même espace de phase entre nos données et nos simulations il faut donc appliquer une série de coupures aux simulations afin de reproduire l'effet de ces chemins. En pratique, on applique une série de coupures de sélection à tous les candidats diphoton (données et simulations) légèrement plus strictes que celles utilisées dans les chemins HLT et décrites en détail dans cette section.

### 6.4.1 Critères de présélection

Ces critères de sélection généraux à tous les événements diphoton, nommés aussi présélection, utilisent les variables relatives au photon explicitées dans l'annexe B.

Les événements doivent avoir au moins deux photons présents dans une zone géométrique d'acceptance définie par  $|\eta| < 2,5$  mais hors de l'interstice EB-EE situé entre



$$1,442 < |\eta| < 1,566.$$

Après l'assignement du vertex décrit dans la section suivante, le photon avec le  $p_T$  le plus élevé est nommé photon *leading* et doit satisfaire  $p_T > 30$  GeV. Le photon avec le  $p_T$  le deuxième plus élevé est identifié comme photon *subleading* et doit satisfaire  $p_T > 18$  GeV. Chacun de ces photons doit aussi satisfaire un critère de veto électron qui rejette l'évènement si le super-agrégat du photon considéré est associé à au moins deux coups dans le détecteur à pixel.

Les deux photons d'un évènement diphoton doivent satisfaire, pour être acceptés, soit  $R_9 > 0,80$  et  $I_{charged} < 20$  GeV, soit  $I_{charged}/p_T < 0,3$  si leur  $p_T > 14$  GeV et  $H/E < 0,15$ .

Des coupures similaires à celles utilisées par les chemins HLT sur  $H/E$ ,  $\sigma_{i\eta i\eta}$ ,  $I_\gamma$ ,  $I_{track}$  et  $R_9$ , résumées dans les tableaux pour 2017 et pour 2018, sont aussi appliquées aux candidats photons. Pour 2017, les paires de photons sélectionnées doivent avoir au moins un photon dans une région de haut  $R_9$ . Pour 2018, les photons de bas  $R_9$  sont acceptés s'ils sont situés dans la partie EB et rejetés s'ils sont dans la partie EE.

TABLE 6.3 – Critères de présélection pour 2017

	$H/E$	$\sigma_{i\eta i\eta}$	$R_9$	$I_\gamma$	$I_{track}$
EB; $R_9 > 0.85$	$< 0.08$	$< 0.015$	$> 0.5$	$< 4.0$	$< 6.0$
EB; $R_9 \leq 0.85$	$< 0.08$	$< 0.015$	$> 0.5$	$< 4.0$	$< 6.0$
EE; $R_9 > 0.90$	$< 0.08$	$< 0.035$	$> 0.8$	$< 4.0$	$< 6.0$
EE; $R_9 \leq 0.90$	$< 0.08$	$< 0.035$	$> 0.8$	$< 4.0$	$< 6.0$

TABLE 6.4 – Critères de présélection pour 2018

	$H/E$	$\sigma_{i\eta i\eta}$	$R_9$	$I_\gamma$	$I_{track}$
EB; $R_9 > 0.85$	$< 0.08$	$< 0.015$	$> 0.5$	$< 4.0$	$< 6.0$
EB; $R_9 \leq 0.85$	$< 0.08$	$< 0.015$	$> 0.5$	$< 4.0$	$< 6.0$
EE; $R_9 > 0.90$	$< 0.08$	$< 0.035$	$> 0.8$	$< 4.0$	$< 6.0$

Une coupure lâche  $< -0,9$  est ensuite appliquée sur le score de sortie du BDT utilisé pour identifier les photons, et présenté dans la section 6.6, afin d'éliminer les photons non-prompts tout en gardant  $\sim 100$  % des photons prompts.

Finalement les paires de photons doivent satisfaire  $65 < m_{\gamma\gamma} < 120$  GeV.



### 6.4.2 Efficacités de présélection

Similairement aux efficacités de déclenchement présentées précédemment, on souhaite connaître l'efficacité, sur le signal, des coupures de présélection présentées. Plus précisément, on veut comparer les efficacités obtenues pour les simulations et les données afin de calculer des facteurs d'échelle à appliquer aux simulations pour obtenir un meilleur accord avec les données.

La méthode TnP, présentée dans la section précédente, est tout à fait adaptée pour ce calcul, cependant la présélection de l'analyse requiert un veto électron rejetant les électrons. On commence donc par calculer l'efficacité de la présélection sans ce veto avec des événements  $Z \rightarrow ee$  TnP avant d'utiliser des photons provenant d'évènement  $Z \rightarrow \mu\mu\gamma$  pour calculer l'efficacité du veto.

#### Mesure de l'efficacité de présélection avec avec les évènements $Z \rightarrow ee$

Pour calculer les efficacités de la présélection sans veto électron ainsi que de la coupure appliquée au score de sortie du BDT d'identification des photons, on utilise donc des événements  $Z \rightarrow ee$  et la méthode TnP. Les critères utilisés pour ce faire sont identiques à ceux présentés dans la section 6.6.2 (sans critère explicite sur le score de sortie du BDT d'identification des photons hors coupure minimale de -0,9 pour l'efficacité concernée). Le calcul s'effectue par une simple méthode de comptage pour les simulations MC et par une méthode d'ajustement pour les données, décrites dans la section précédente pour les efficacités HLT.

L'efficacité est définie comme le rapport entre le nombre d'électrons "probes" passant les coupures choisies sur le nombre total d'électrons "probes". Les facteurs d'échelle (SF) sont définis comme le rapport entre l'efficacité dans les données et l'efficacité dans les simulations.

Les efficacités et SF de la présélection sont visibles sur les tableaux 6.5 et 6.7 pour la présélection (jusqu'au veto électron), et sur les tableaux 6.6 et 6.8 pour la coupure à -0,9 sur le score minimal du BDT d'identification de photons pour 2017 et 2018 respectivement. Les incertitudes systématiques associées prennent en compte la variation des efficacités si on utilise  $p_T > 40$  GeV pour les "tags", les choix de fonctions utilisées pour l'ajustement signal et bruit de fond dans la méthode de calcul d'efficacité basée sur l'ajustement et la propagation des incertitudes provenant de la repondération en  $(\eta, R_9)$  entre les simulations  $Z \rightarrow ee$  et  $h \rightarrow \gamma\gamma$ . Tous les SF, sauf le SF associé à la présélection 2018 dans EE, sont compatibles avec 1 en prenant en compte les incertitudes associées. Ces SF sont appliqués aux simulations MC et notamment aux événements des simulations signal, et les incertitudes sont prises en compte en tant qu'incertitudes systématiques de l'analyse, comme présenté dans la section 6.10.

#### Mesure de l'efficacité de présélection avec avec les évènements $Z \rightarrow \mu\mu\gamma$

L'efficacité du veto électron est calculée grâce au lot de photons de haute pureté accessibles en utilisant les événements  $Z \rightarrow \mu\mu\gamma$  présentés au chapitre précédent. Les critères utilisés pour ce faire sont identiques à ceux présentés dans la section 6.6.2. L'efficacité du veto électron est alors mesurée comme le rapport entre le nombre de photons passant le veto électron (en plus

	Data		Simulation		Ratio		
	Eff.	Stat. Unc.	Eff.	Stat. Unc.	Scale	Stat. Unc.	Syst. Err
Barrel; $R_9 > 0.85$	0.8699	0.0001	0.8756	0.0002	0.9934	0.0003	0.0139
Barrel; $R_9 < 0.85$	0.5986	0.0002	0.5956	0.0003	1.0050	0.0004	0.0333
Endcap; $R_9 > 0.90$	0.8475	0.0002	0.8494	0.0003	0.9977	0.0004	0.0070
Endcap; $R_9 < 0.90$	0.5739	0.0004	0.5770	0.0006	0.9946	0.0012	0.0155

 TABLE 6.5 – Efficacités de la présélection (jusqu’au veto électron) pour 2017 calculées en différentes catégories  $R_9$  et  $\eta$  avec des évènements  $Z \rightarrow ee$  TnP.

	Data		Simulation		Ratio		
	Eff.	Stat. Unc.	Eff.	Stat. Unc.	Scale	Stat. Unc.	Syst. Err
Barrel; $R_9 > 0.85$	0.9943	< 0.0001	0.9963	< 0.0001	0.9979	< 0.0001	0.0017
Barrel; $R_9 < 0.85$	0.9803	0.0001	0.9852	0.0001	0.9950	0.0001	0.0026
Endcap; $R_9 > 0.90$	0.9951	0.0001	0.9950	0.0001	1.0001	0.0001	0.0012
Endcap; $R_9 < 0.90$	0.9730	0.0002	0.9723	0.0003	1.0007	0.0004	0.0022

 TABLE 6.6 – Efficacités de la coupure en score du BDT d’identification des photons pour 2017 calculées en différentes catégories  $R_9$  et  $\eta$  avec des évènements  $Z \rightarrow ee$  TnP.

	Data		Simulation		Ratio		
	Eff.	Stat. Unc.	Eff.	Stat. Unc.	Scale	Stat. Unc.	Syst. Err
Barrel; $R_9 > 0.85$	0.8677	0.0001	0.8631	0.0002	1.0053	0.0003	0.0201
Barrel; $R_9 < 0.85$	0.8160	0.0002	0.8105	0.0003	1.0068	0.0005	0.0588
Endcap; $R_9 > 0.90$	0.7761	0.0002	0.7588	0.0004	1.0228	0.0006	0.0060

 TABLE 6.7 – Efficacités de la présélection (jusqu’au veto électron) pour 2018 calculées en différentes catégories  $R_9$  et  $\eta$  avec des évènements  $Z \rightarrow ee$  TnP.

	Data		Simulation		Ratio		
	Eff.	Stat. Unc.	Eff.	Stat. Unc.	Scale	Stat. Unc.	Syst. Err
Barrel; $R_9 > 0.85$	0.9952	< 0.0001	0.9954	< 0.0001	0.9998	< 0.0001	0.0007
Barrel; $R_9 < 0.85$	0.9841	0.0001	0.9829	0.0001	1.0012	0.0001	0.0029
Endcap; $R_9 > 0.90$	0.9939	< 0.0001	0.9930	0.0001	1.0009	0.0001	0.0006

 TABLE 6.8 – Efficacités de la coupure en score du BDT d’identification des photons pour 2018 calculées en différentes catégories  $R_9$  et  $\eta$  avec des évènements  $Z \rightarrow ee$  TnP.

de la présélection) sur le nombre total de photons présélectionnés sans veto électron.

Les efficacités et SF du veto électron sont visibles sur les tableaux 6.9 et 6.10 pour 2017 et 2018 respectivement. A nouveau ces SF sont appliqués aux simulations MC et notamment aux évènements des simulations signal, et les incertitudes sont prises en compte en tant qu'incertitudes systématiques de l'analyse.

	Data		Simulation		Scale Factor	
	Eff.	Stat. Unc.	Eff.	Stat. Unc.	Eff.	Unc.
Barrel ; $R_9 > 0.85$	0.9382	0.0010	0.9619	0.0013	0.9754	0.0017
Barrel ; $R_9 < 0.85$	0.8438	0.0026	0.8781	0.0042	0.9610	0.0054
Endcap ; $R_9 > 0.90$	0.6897	0.0034	0.7495	0.0052	0.9202	0.0078
Endcap ; $R_9 < 0.90$	0.4561	0.0081	0.4979	0.0133	0.9161	0.0293

TABLE 6.9 – Efficacités du veto électron pour 2017 calculées en différentes catégories  $R_9$  et  $\eta$  avec des évènements  $Z \rightarrow \mu\mu\gamma$ .

	Data		Simulation		Scale Factor	
	Eff.	Stat. Unc.	Eff.	Stat. Unc.	Eff.	Unc.
Barrel ; $R_9 > 0.85$	0.9343	0.0008	0.9656	0.0013	0.9676	0.0015
Barrel ; $R_9 < 0.85$	0.8465	0.0022	0.8742	0.0042	0.9683	0.0053
Endcap ; $R_9 > 0.90$	0.6711	0.0028	0.7589	0.0054	0.8844	0.0073

TABLE 6.10 – Efficacités du veto électron pour 2018 calculées en différentes catégories  $R_9$  et  $\eta$  avec des évènements  $Z \rightarrow \mu\mu\gamma$ .

## 6.5 Identification des vertex diphoton

Pour des désintégrations de bosons de Higgs en deux photons, l'état final où les deux photons ne sont pas convertis n'est pas détecté dans le trajectographe et le ECAL n'est pas segmenté longitudinalement donc il est difficile de déterminer le vertex d'interaction associé parmi tous les vertex issus de l'empilement. Par évènement, il y a en effet en moyenne environ 27 (28) choix de vertex possible pour 2017 (2018), et on peut avoir jusqu'à 75 vertex dans un évènement. Or le choix du vertex a un effet direct sur la résolution en masse invariante diphoton<sup>4</sup> et on estime [101] qu'un mauvais choix de vertex peut dégrader cette résolution d'1 GeV en moyenne. La résolution en masse étant essentielle dans ce type d'analyse, il est donc important de pouvoir déterminer le bon vertex d'interaction. Pour ce faire on utilise deux BDT communs à toutes les analyses  $H \rightarrow \gamma\gamma$  et décrits en détail dans [90] et [107]. Le travail présenté dans cette section n'est donc pas directement le mien bien que je possède la responsabilité de produire, pour toutes les analyses  $H \rightarrow \gamma\gamma$  dans CMS, les jeux de données, appelés microAOD, où le vertex d'interaction est identifié selon la méthode décrite ici.

4. Le choix du vertex a une incidence directe, par exemple, sur l'angle entre les deux photons apparaissant dans la formule de la masse invariante.

### 6.5.1 BDT d'identification de vertex

On utilise un premier BDT, le *vertex ID*, afin de construire une variable discriminant les bons vertex en utilisant les corrélations entre les traces de recul pour un vertex donné et le système diphoton, ainsi que les traces de conversion si au moins un des photons est un photon converti.

La *sortie* de ce BDT est donc une variable discriminante prenant valeur entre -1 et +1, piquant à +1 pour un signal, le bon vertex, et -1 pour du bruit de fond, un très mauvais vertex diphoton. Le vertex choisi parmi un ensemble de vertex diphoton possible, pour chaque évènement, dans l'analyse est celui qui est le plus proche d'un signal en terme de ce score de sortie donc le vertex qui a le meilleur score de sortie du *vertex ID*.

L'*entraînement* de ce BDT est effectué avec des simulations MC  $H \rightarrow \gamma\gamma$  à une masse de 125 GeV où le vrai vertex d'interaction diphoton est connu. Le *signal* est alors défini comme étant le vertex reconstruit le plus proche du vrai vertex et le *bruit de fond* comme étant n'importe quel autre mauvais vertex.

Les *variables d'entrée*, permettant la discrimination entre signal et bruit de fond, sont des variables relatives aux traces de recul par rapport au système diphoton :

- somme des  $p_T^2$ ,  $\sum_i |\vec{p}_{T,i}|^2$
- balance en impulsion transverse,  $-\sum_i \frac{\vec{p}_{T,i} \cdot \vec{p}_T^{\gamma\gamma}}{|\vec{p}_T^{\gamma\gamma}|}$
- asymétrie en impulsion transverse,  $\frac{|\sum_i \vec{p}_{T,i}| - p_T^{\gamma\gamma}}{|\sum_i \vec{p}_{T,i}| + p_T^{\gamma\gamma}}$

où on somme sur tous les candidats PF chargés associés au vertex diphoton considéré.  $\vec{p}_{T,i}$  correspond au moment transverse de la  $i$ -ème trace associée au vertex diphoton considéré et  $\vec{p}_T^{\gamma\gamma}$  est le moment transverse du système diphoton pour le vertex considéré.

Si des traces de photons convertis dans le trajectographe sont présentes, deux variables supplémentaires sont utilisées :

- le nombre de photons convertis (0, 1 ou 2)
- la différence entre la position longitudinale du vertex reconstruit considéré et celle estimée en utilisant les traces de conversion, divisée par l'incertitude sur cette dernière position [107]

Le réentraînement du *vertex ID* a été effectué pour les années 2017 et 2018 mais les performances ont été jugées similaires et seul l'entraînement 2016 a été finalement utilisé pour les trois années [105].

### 6.5.2 Probabilité d'estimer le bon vertex

Sachant que  $m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - \cos\Delta\alpha)}$ , il peut être montré [107] que la résolution de la différence angulaire  $\Delta\alpha$  entre les photons a une contribution négligeable à la résolution totale de la masse invariante diphoton lorsque le vertex identifié choisi se situe à moins d'1 cm du vrai vertex d'interaction. Un vertex choisi à plus de 1,5 cm du vrai vertex d'interaction peut ainsi dégrader la résolution en masse invariante diphoton de plus de 20 %. Il est donc crucial de pouvoir estimer la probabilité, par évènement, pour le vertex choisi grâce au *vertex ID* d'être situé à moins d'1 cm du vrai vertex.

Cette probabilité par évènement, surnommée  $p_{vertex}$ , est calculée en tant que score de sortie d'un deuxième BDT similaire à celui présenté précédemment ; elle sera ensuite utilisée en tant que variable d'entrée pour le BDT de classification des évènements présenté dans la section 6.7. Elle est paramétrée, séparément pour les photons non/convertis, par un polynôme d'ordre 4 pour les années 2016, 2017 et 2018 et en utilisant des simulations MC  $H \rightarrow \gamma\gamma$ . Les variables d'entrée utilisées pour ce faire sont :

- $p_T$  du système diphoton
- le nombre de vertex
- les valeurs de sortie du *vertex ID* pour les 3 meilleurs vertex choisis
- les distances longitudinales entre le meilleur vertex choisi, le deuxième et le troisième vertex choisis
- le nombre de photons convertis

On observe sur la figure 6.6 droite que la probabilité moyenne pour un vertex d'être à moins de 1 cm du vrai vertex d'interaction est compatible avec la fraction de vertex situé à moins d'1 cm du vrai vertex déterminée grâce aux informations vraies d'évènements  $H \rightarrow \gamma\gamma$ . De plus, un bon accord entre données et simulation pour des évènements  $Z \rightarrow \mu\mu$  est observé sur la gauche de la figure 6.6 concernant la fraction d'évènement dont le vertex identifié se situe à moins de 1 cm du vrai vertex.

### 6.5.3 Validation et facteurs d'échelle

La validation de l'algorithme d'identification de vertex décrit précédemment repose sur l'utilisation de données et de simulations provenant d'évènements  $Z \rightarrow \mu^+\mu^-$  imitant un système diphoton. On requiert des critères d'identifications resserrés [93] pour les muons ainsi qu'une masse invariante dimuon entre 70 et 110 GeV ; les traces de ces deux muons sont aussi ignorées lors du procédé d'identification de vertex. Ces traces permettent de remonter à un vertex d'interaction qui sert de vertex de référence pour la validation : si le vertex identifié par le *vertex ID* est situé à moins de 1 cm de ce vertex de référence, le vertex identifié est alors considéré comme ayant été correctement identifié. L'efficacité d'avoir assigné correctement le vertex est visible sur la partie de gauche de la figure 6.6, on peut y voir que les données et les simulations sont en accord avec une différence d'environ 2 % sur toute la gamme  $p_T^{\mu^+\mu^-}$ .

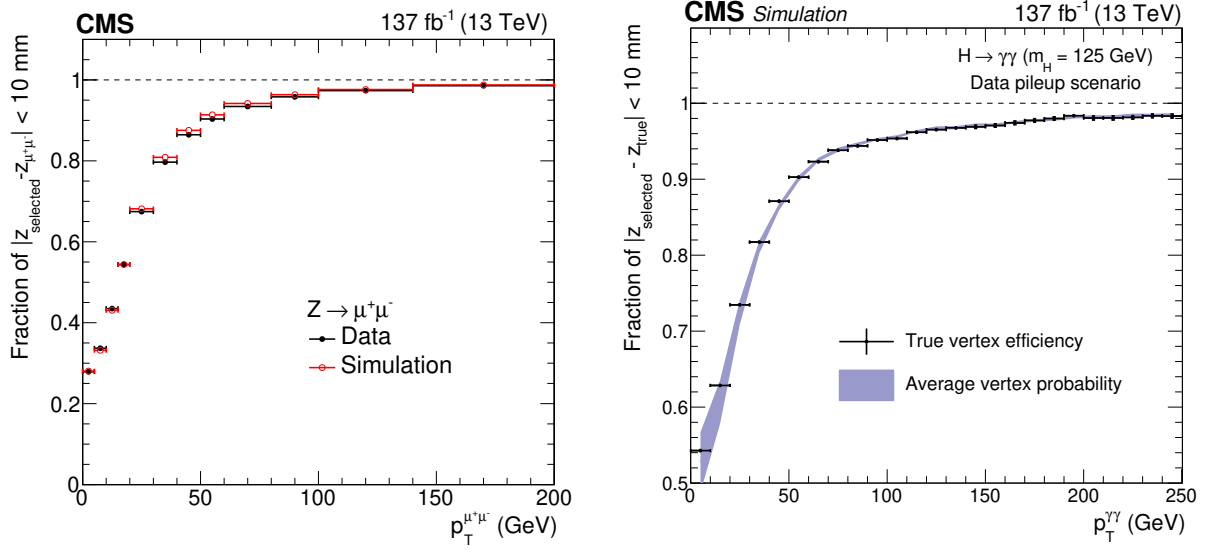


FIGURE 6.6 – A gauche, la validation de l’algorithme d’identification des vertex  $H \rightarrow \gamma\gamma$  en utilisant des évènements  $Z \rightarrow \mu^+\mu^-$  permettant de comparer la fraction d’évènement dont le vertex a été correctement assigné entre données et simulations en fonction du  $p_T$  du système dimuon. A droite, la probabilité moyenne pour un vertex choisi d’être à moins de 1 cm du vrai vertex, calculée pour les données, et l’efficacité du vraie vertex, obtenue pour les simulations, en fonction de  $p_T$  du système diphoton. Les jeux de données correspondant aux années 2016, 2017 et 2018, ainsi que les simulations correspondantes, ont été utilisées pour ces figures [90].

Pour prendre ces différences entre données et simulations MC en compte, des facteurs d’échelle sont calculés à partir de ces évènements  $Z \rightarrow \mu^+\mu^-$  et sont ensuite appliqués aux simulations  $H \rightarrow \gamma\gamma$  en fonction de l’impulsion transverse du boson de Higgs [90] pour chaque année. Une incertitude systématique supplémentaire, notamment détaillée dans la section 6.10 dédiée, est aussi implémentée pour les simulations signal de l’analyse permettant une migration entre catégorie d’évènements avec une bonne identification de vertex et entre catégorie d’évènements avec une mauvaise reconstruction de vertex.

## 6.6 Identification des photons

L’objectif principal de l’étape d’identification des photons est de distinguer les photons prompts des photons non-prompts, ou faux photons. Ces photons non-prompts viennent en partie de mésons neutres ( $\pi^0$  se désintégrant en deux photons partageant le même super-agrégat et ainsi reconstruit comme un seul photon) et de particules reconstruites à tort comme des photons. Pour séparer ces deux types de photons, on utilise un BDT, nommé *Photon ID BDT* décrit, avec ses procédés de validation, dans cette section. Ce BDT, plus performant que celui utilisé pour identifier les photons dans le cas général [74], est commun aux analyses  $H \rightarrow \gamma\gamma$  mais il a été ici spécialement réentraîné avec les conditions de l’analyse basses masses. Le travail présenté dans cette section n’est pas directement le mien.

### 6.6.1 Photon ID BDT

Le *Photon ID BDT*, ou *Photon ID MVA*, permet d'obtenir, en *sortie*, une fonction donnant un score pour les données entre -1 et 1 par photon selon si ce photon est un photon prompt, dans ce cas il aura une valeur proche de 1, ou s'il est au contraire non-prompt ou n'est pas un vrai photon, il aura dans ce cas une valeur proche de -1. On réalise ensuite une coupure très lâche sur ce score de sortie pour rejeter les événements avec des photons non-prompts ou des photons mal identifiés, et on utilisera dans la suite ce score de sortie comme variable d'entrée d'un autre BDT responsable de la classification des événements décrit dans la section 6.7.

Ce BDT est *entraîné* sur des simulations MC 2017 d'échantillons  $\gamma$ + jets décrits précédemment. Ces échantillons contiennent un photon prompt, considéré comme signal, par événement et des photon non-prompts, considérés comme bruit de fond, différenciés par leurs informations au niveau générateur. Les photons contenus dans ces échantillons sont soumis à la sélection générale utilisée dans l'analyse, et décrite dans la section 6.4, qui permet notamment d'avoir le même espace de phase entre données et simulations. Les photons considérés doivent notamment satisfaire aux conditions  $p_T > 18$  GeV et  $m_{\gamma\gamma} > 55$  GeV. Le réentraînement du *Photon ID BDT* avec ces conditions est une nouveauté par rapport à l'analyse utilisant les données 2012 et 2016 [103].

Malgré les *corrections* appliquées aux photons présentées dans le chapitre 4, il reste des différences entre données et simulations notamment pour les variables concernant les gerbes électromagnétiques et leur isolation dans le ECAL qui pourraient être propagées au score de sortie du *Photon ID BDT*. La distribution de ces variables est donc corrigée dans les simulations pour correspondre à celle obtenue dans les données à l'aide d'une technique de transformation par quantiles pour 2017 et des chaînes de régression de quantiles utilisant des BDT décrites en détail dans [105] et [108]. Ces corrections sont dérivées grâce à la méthode TnP  $Z \rightarrow ee$  dans le cas de l'analyse mesurant les propriétés du boson de Higgs du MS [90] et sont appliquées à notre analyse, et une incertitude systématique est déterminée pour couvrir les éventuelles différences restantes entre données et simulations. Ces corrections sont finalement validées pour notre analyse en comparant la distribution des variables concernées avec des événements  $Z \rightarrow ee$ , selon la méthode TnP présentée dans la section 6.3.2 qui inclut la présélection générale de l'analyse discutée dans la section 6.4, et des événements  $Z \rightarrow \mu\mu\gamma$ , selon la méthode et la sélection présentées dans la section 5.1.3 auquel on ajoute, pour le photon, la présélection générale de l'analyse.

Afin d'éviter que le *Photon ID BDT* n'utilise les différences entre la distribution en  $p_T$  des photons entre le signal et le bruit de fond lors de l'entraînement, une repondération en deux dimensions ( $\eta$ ,  $p_T$ ) des photons est effectuée sur les simulations afin que la distribution en  $p_T$  du signal corresponde à celle du bruit de fond. Les *variables d'entrée* permettant la discrimination entre signal et bruit de fond utilisées par le BDT sont alors :

- $\rho$  (densité d'énergie),  $\eta$  du super-agrégat,  $E_{raw}$  en ce qui concerne les variables cinématiques des photons



- $\sigma_{i\eta i\eta}, \sigma_{i\eta i\phi}, E_{2 \times 2}/E_{5 \times 5}, R_{9,5 \times 5}, \sigma_{\eta}$  (déviations logarithmiquement pondérées en  $\eta$  et  $\phi$ ) pour les variables relatives à la forme et l'étalement de la gerbe électromagnétique des photons considérés
- $I_{\gamma}$  et  $I_{charged}$  (calculé pour le meilleur vertex sélectionné par le *vertex ID* et le pire vertex<sup>5</sup>) pour les variables relatives à l'isolation des photons

On rappelle que toutes les variables relatives au photon sont décrites dans l'annexe B. L'entraînement du *Photon ID BDT* est réalisé séparément pour la partie tonneau et pour la partie bouchons ; pour améliorer les performances du BDT dans les bouchons on utilise deux variables d'entrée supplémentaires pour l'entraînement correspondant qui sont le rapport entre l'énergie mesurée dans le détecteur à pied de gerbe et l'énergie brute du super-agrégat  $E_{ES}/E_{raw}$  et la déviation standard de l'étalement de la gerbe dans le détecteur à pied de gerbe  $\sigma_{RR}$ .

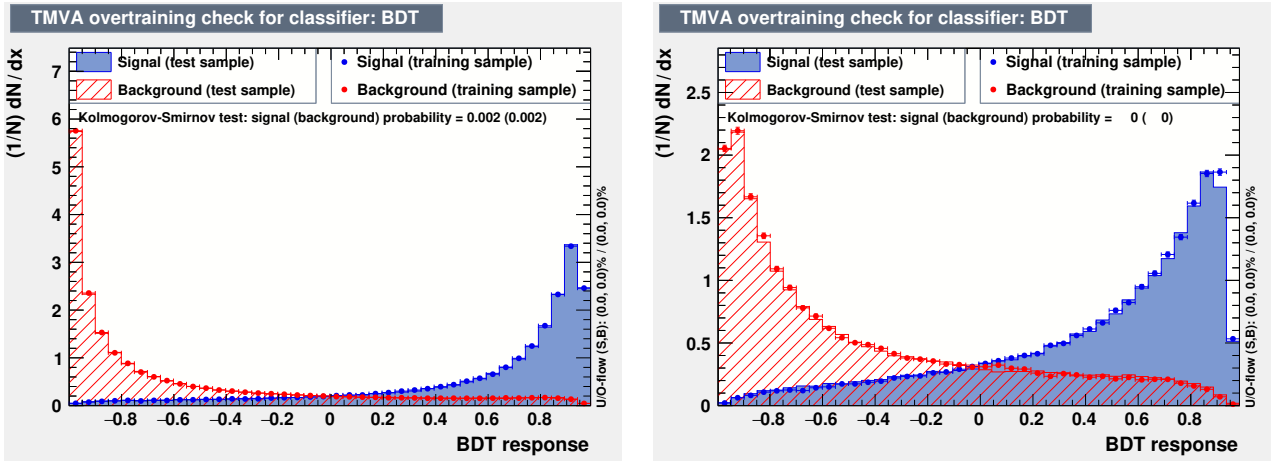


FIGURE 6.7 – Comparaison des distributions en score de sortie du *Photon ID BDT* entre les événements d'entraînement et les événements test pour les simulations 2017 du signal (en bleu) et du bruit de fond (en rouge) correspondant aux échantillons utilisés dans l'entraînement du *Photon ID BDT* ; ceci pour le tonneau (à gauche) et les bouchons (à droite).

La moitié des événements des simulation MC  $\gamma$  + jets disponibles est utilisée pour l'entraînement du *Photon ID BDT* et l'autre moitié est utilisée pour tester cet entraînement. La fonction de sortie obtenue avec les événements d'entraînement est appliquée aux événements de test afin de déterminer leur distribution en terme de score de sortie *Photon ID BDT* et la comparer à celle obtenue avec les événements d'entraînement. On observe sur la figure 6.7 que les distributions de score de sortie sont semblables entre événements d'entraînement et de test pour signal et bruit de fond ce qui signifie qu'il n'y a pas de sur-entraînement et que l'entraînement est performant.

Pour comparer, en partie, l'effet de l'entraînement entre données et simulations mais aussi valider l'utilisation des simulations 2017 sur les données de 2018, on étudie le score de sortie du

5. En général, un photon prompt est isolé par rapport à n'importe quel vertex ce qui n'est pas le cas des faux photons.



*Photon ID BDT* pour les échantillons de simulations signal ( $h \rightarrow \gamma\gamma$  à 90 GeV) et bruit de fond ( $\gamma\gamma$ ,  $\gamma$  + jets, QCD) de l'analyse ainsi que pour les données, le tout pour 2017 et 2018. Toutes ces simulations sont corrigées selon les explications données précédemment. Les simulations et données doivent, de plus, satisfaire les critères de sélections généraux de l'analyse. Le score du *Photon ID BDT* est ensuite calculé (dans les données et simulations 2017 et 2018) pour chacun des deux photons des événements diphoton, en utilisant la fonction de sortie déterminée par l'entraînement basé sur les simulations 2017 présenté dans cette section. La figure 6.8 montre les distributions en score de sortie *Photon ID BDT* pour les photons ayant le score le moins élevé parmi les deux photons du système diphoton pour chaque événement; ces distributions sont visibles pour les données (on utilise des événements provenant de gammes de masse non utilisées dans les résultats finaux) et les simulations de l'analyse pour 2017 (à gauche) et 2018 (à droite). Les accords entre les données et simulations sont bons pour 2017 et 2018. De plus, les performances de l'entraînement 2017 ont été calculées comme étant similaires entre 2017 et 2018 [104], ce qui, en addition des validations présentées dans la suite, permet de justifier l'utilisation de cet entraînement pour les données et simulations 2018.

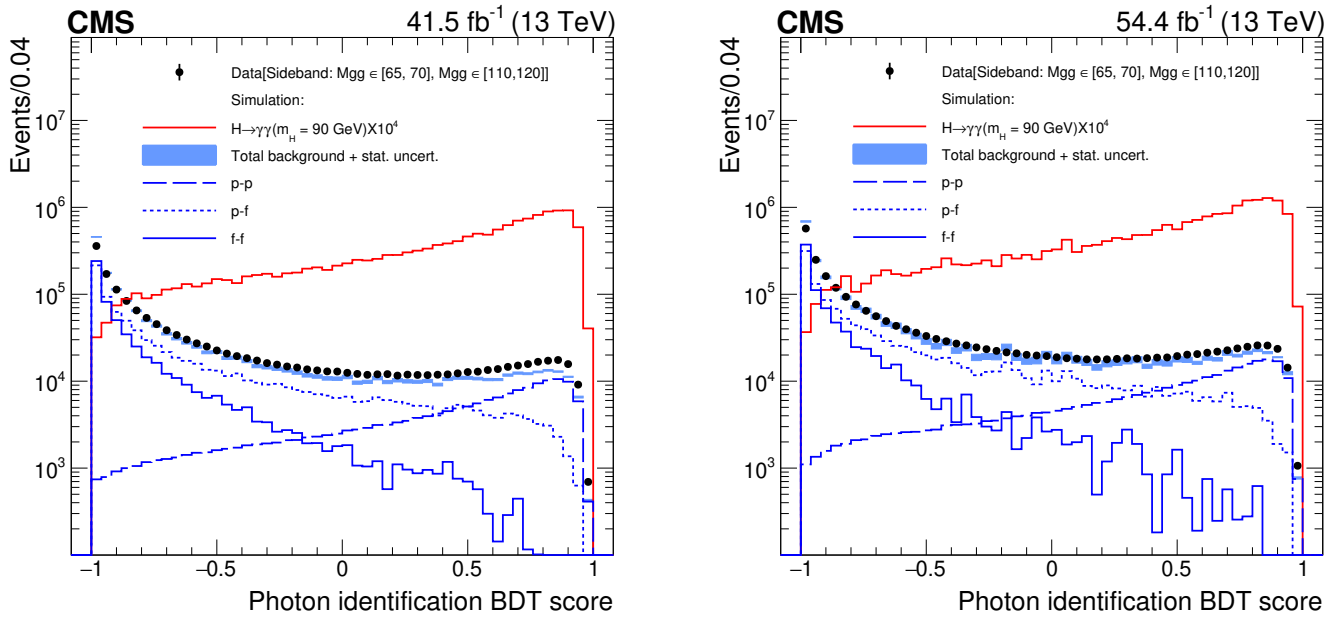


FIGURE 6.8 – Distribution du score de sortie du *Photon ID BDT* des photons avec le score le plus faible dans les paires diphoton présentes dans les événements pour les données et simulations pour 2017 (gauche) et 2018 (droite). Les données (points noirs) sont sélectionnées dans des zones de masses non utilisées dans les résultats finaux,  $m_{\gamma\gamma} \in [65, 70] \cup [110, 120]$  GeV. Les simulations du bruit de fond (en pointillé) satisfont  $m_{\gamma\gamma} \in [65, 120]$  GeV et sont séparées en trois composantes ayant deux (p-p), un (p-f) ou zéro (f-f) photons prompts. La somme des contributions des simulations du bruit de fond est multiplié par un facteur d'échelle pour correspondre, en nombre d'évènements, à la contribution des données. En rouge la distribution attendue pour un signal  $h \rightarrow \gamma\gamma$  à 90 GeV.

A partir de cette étape, on requiert systématiquement une coupure lâche sur le score de sortie du *Photon ID BDT* telle que l'on rejette, comme décrit dans la section 6.4, les événements

ayant avec un photon avec un score *Photon ID BDT*  $< -0,9$ .

Le BDT faisant appel à de nombreuses variables discriminantes, des validations supplémentaires sont nécessaires pour valider les accords entre données et simulations pour l'entraînement effectué avec les simulations 2017. Ces validations sont étudiées dans la suite avec des évènements  $Z \rightarrow ee$  et  $Z \rightarrow \mu\mu\gamma$ .

## 6.6.2 Validation du photon ID BDT

### Validation avec les évènements $Z \rightarrow ee$

Pour visualiser si les variables d'entrée et le score de sortie du *Photon ID BDT* sont bien simulées pour des évènements ressemblant au signal de l'analyse, on compare leur distribution pour les données et les simulations MC avec des évènements  $Z \rightarrow ee$  où les électrons sont reconstruits comme des photons. Cette validation s'effectue alors avec les électrons "probes" de la méthode TnP pour les désintégrations  $Z \rightarrow ee$ .

Les corrections de forme et d'isolation des gerbes électromagnétiques évoquées précédemment sont appliquées ici ainsi que la sélection générale de l'analyse (comprenant la coupure sur le score *Photon ID BDT*) où le critère de pixel veto est inversé de façon à sélectionner des électrons. On requiert de plus des coupures tels que  $p_T > 35$  GeV (18 GeV) et  $|\eta| < 2,1$  (2,5), sans la région  $1,4442 \leq |\eta| \leq 1,566$ , pour les électrons "tags" ("probes"); et des critères de qualité supplémentaires pour les "tags" avec  $R_9 > 0,8$ , score *Photon ID BDT*  $> 0,4$ ,  $I_{charged} < 20$  GeV et  $I_{charged}/p_T < 0,3$ .

Les accords entre données et simulations résultants sont corrects; pour le score de sortie il y a néanmoins quelques différences. On ajoute alors une incertitude systématique, correspondant à des variations du score *Photon ID BDT*, aux simulations de l'analyse pour couvrir ces différences entre données et simulations, et on peut alors voir sur les figures 6.9 que les accords entre données et simulations sont bien couverts par l'ajout de cette incertitude systématique.

De la même manière, une incertitude supplémentaire est prise en compte pour couvrir les différences entre les données et simulations pour la distribution de la résolution en énergie par photon  $\sigma_E/E$  comme visible sur la figure 6.10.

### Validation avec les évènements $Z \rightarrow \mu\mu\gamma$

Il est aussi important d'effectuer cette validation pour des photons. Pour cela, on sélectionne un échantillon de photons de haute pureté provenant de désintégration  $Z \rightarrow \mu\mu\gamma$ , comme décrit dans la section 5.1.3, qui doivent de plus passer les critères de sélection généraux de l'analyse, décrits dans la section 6.4 (incluant la coupure sur le score *Photon ID BDT*). Les corrections de forme et d'isolation des gerbes électromagnétiques évoquées précédemment sont aussi appliquées ici.

La distribution en masse invariante  $Z \rightarrow \mu\mu\gamma$  pour les données et les simulations est visible sur la figure 6.11.

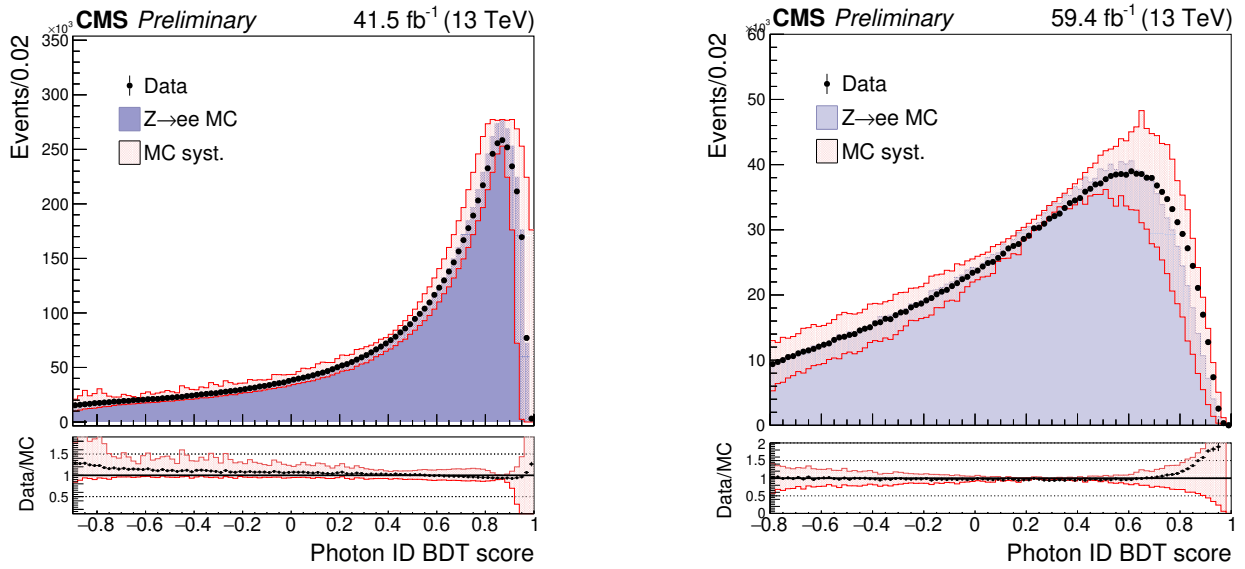


FIGURE 6.9 – Distribution du score de sortie du *Photon ID BDT* pour des évènements  $Z \rightarrow ee$  pour les données et simulations 2017 dans le tonneau (gauche) et pour les données et simulations 2018 dans les bouchons (droite). Les données sont représentées par des points noirs, les simulations, normalisées au nombre d'évènements des données, en bleu plein et l'incertitude systématique appliquée à la forme de la distribution des simulations est visible en rouge.

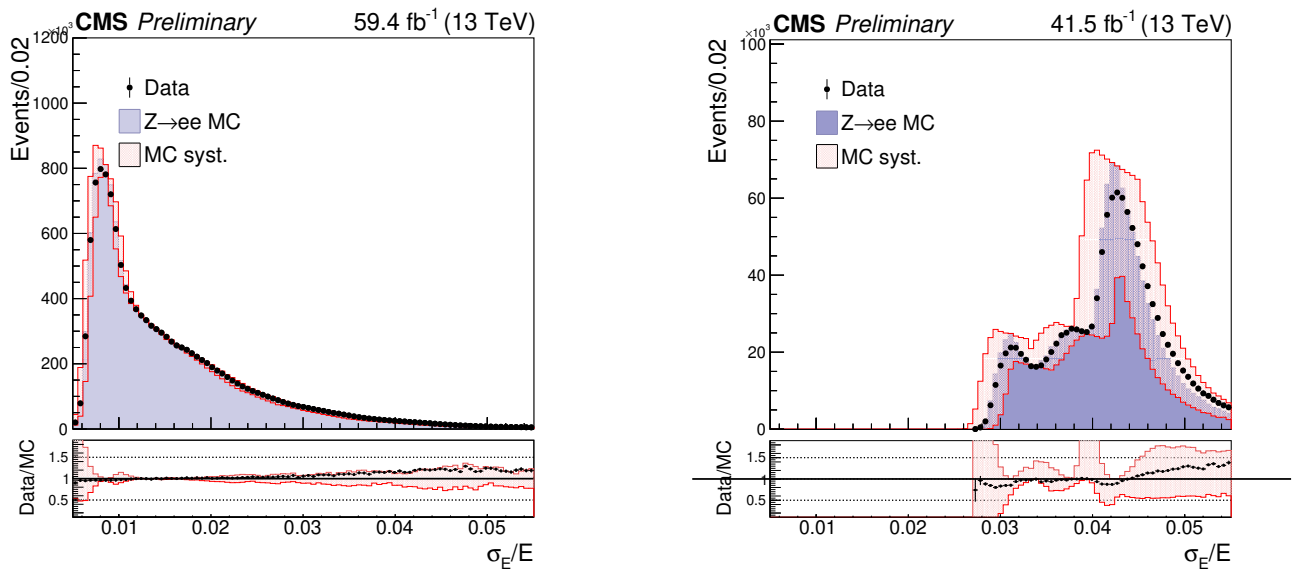


FIGURE 6.10 – Distribution  $\sigma_E/E$  pour des évènements  $Z \rightarrow ee$  pour les données et simulations 2018 EB (gauche) et 2017 EE (droite). Les données sont représentées par des points noirs, les simulations, normalisées au nombre d'évènements des données, en bleu plein et l'incertitude systématique appliquée à la forme de la distribution des simulations est visible en rouge.

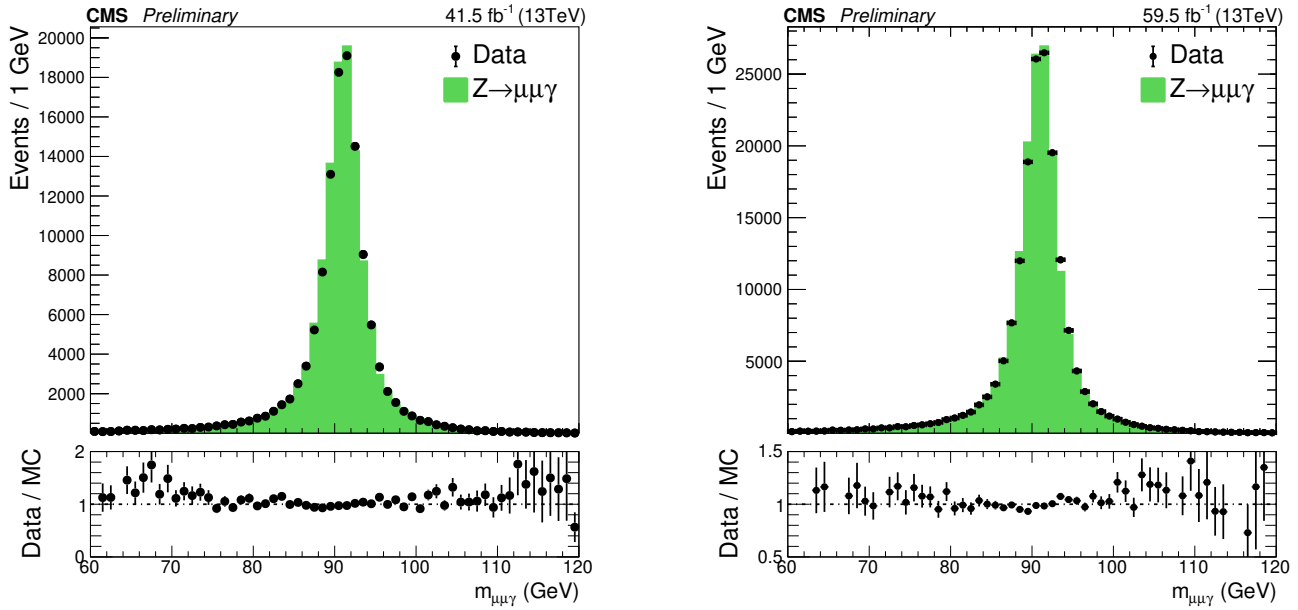


FIGURE 6.11 – Distribution de la masse invariante à trois corps pour des événements  $Z \rightarrow \mu\mu\gamma$  pour les données et simulations pour 2017 (gauche) et 2018 (droite). Les données sont représentées par des points noirs, les simulations, normalisées au nombre d'évènements des données, en vert plein.

Les accords entre données et simulations, en prenant en compte les incertitudes mentionnées précédemment, sont bons pour les distributions du score *Photon ID BDT*, visibles sur la figure 6.12, et de  $\sigma_E/E$ , visible sur la figure 6.13.

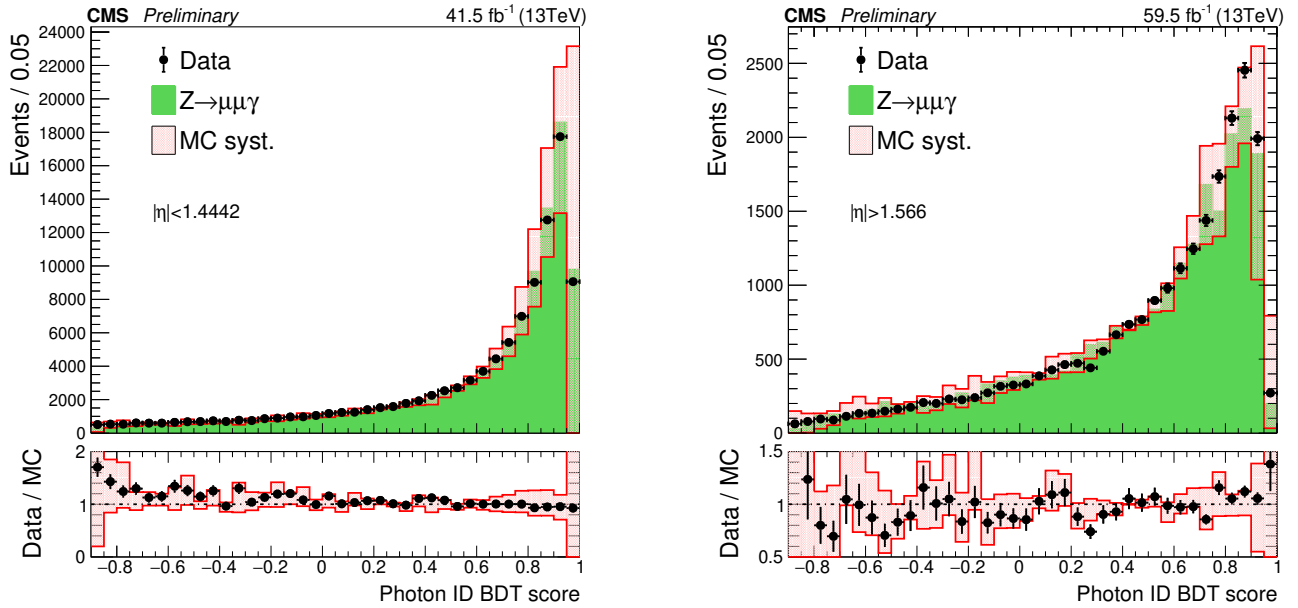


FIGURE 6.12 – Distribution du score de sortie du *Photon ID BDT* pour des évènements  $Z \rightarrow \mu\mu\gamma$  pour les données et simulations pour 2017 EB (gauche) et 2018 EE (droite). Les données sont représentées par des points noirs, les simulations, normalisées au nombre d'évènements des données, en vert plein et l'incertitude systématique appliquée à la forme de la distribution des simulations est visible en bords rouges.

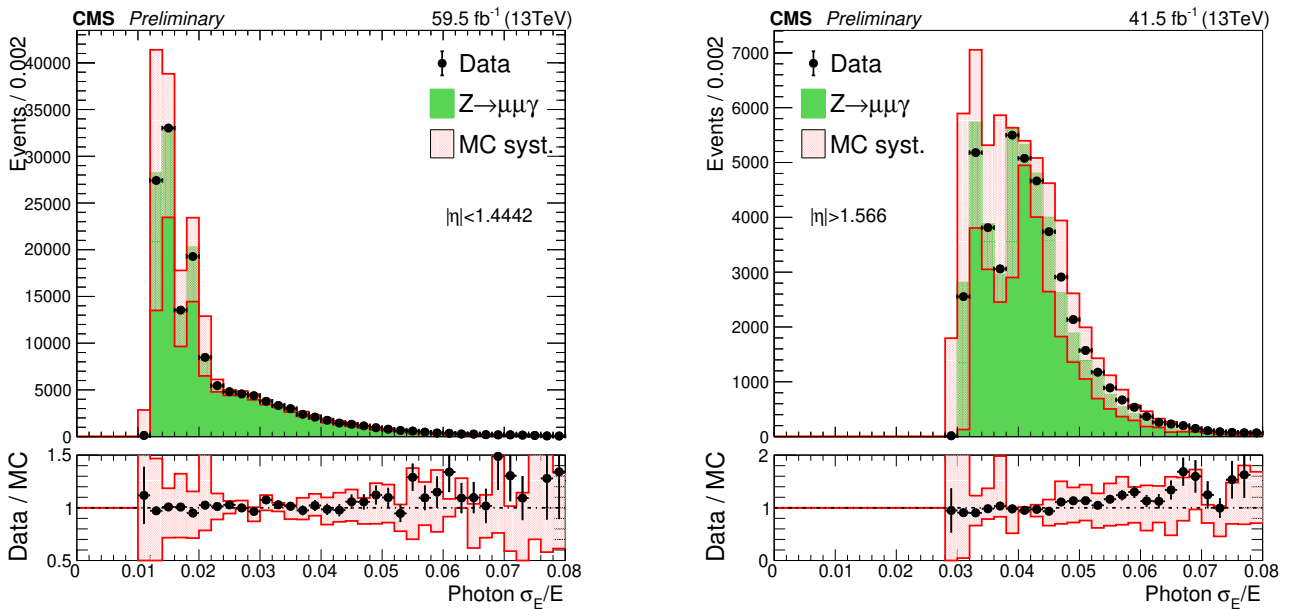


FIGURE 6.13 – Distribution de  $\sigma_E/E$  pour des évènements  $Z \rightarrow \mu\mu\gamma$  pour les données et simulations pour 2017 EB (gauche) et 2018 EE (droite). Les données sont représentées par des points noirs, les simulations, normalisées au nombre d'évènements des données, en vert plein et l'incertitude systématique appliquée à la forme de la distribution des simulations est visible en bords rouges.

## 6.7 Classification des évènements

L'analyse utilise des évènements avec deux photons passant tous les critères de présélection, en prenant en compte les facteurs d'échelle associés, de l'analyse comprenant la coupure sur le score minimal du *Photon ID BDT* auquel on ajoute des coupures de rapport entre le moment transverse du photon (*sub*)*leading* et la masse invariante diphoton  $p_T^{(1)}/m_{\gamma\gamma} > 0,47$  ( $p_T^{(2)}/m_{\gamma\gamma} > 0,28$ ) correspondant environ au rapport entre les seuils en  $p_T$  et la masse inférieure de l'analyse soit 30/65 (18/65). L'utilisation de ces coupures en  $p_T$  normalisées à la masse invariante  $m_{\gamma\gamma}$  permet d'éviter, comme dans [103], les distorsions de la partie basse du spectre de masse invariante diphoton. On requiert aussi que la masse invariante satisfasse  $65 < m_{\gamma\gamma} < 120$  GeV.

La sensibilité de l'analyse peut, à ce point, être améliorée en catégorisant les évènements diphoton selon leur résolution en masse, ratio signal sur bruit de fond, ou encore leur mode de production. Cette catégorisation s'effectue à l'aide de BDT; on présente dans cette section le *diphoton BDT*, et la catégorisation associée permettant de classer les évènements selon leur ressemblance avec le signal ou le bruit de fond de l'analyse. Les travaux présentés dans cette section sont essentiellement le fait de mon travail.

### 6.7.1 Classificateur d'évènements : le diphoton BDT

#### Le diphoton BDT

Ce procédé de classification d'évènements basé sur un BDT, nommé *diphoton BDT*, a pour objectif la catégorisation des évènements, où les catégories les plus performantes sont caractérisées par un haut rapport signal sur bruit de fond et une bonne résolution en masse. Ainsi le *diphoton BDT* va assigner des bons scores à des évènements dont les candidats photons ressemblent à des photons provenant d'une désintégration d'un boson de Higgs de basse masse; notamment des évènements avec des caractéristiques cinématiques proches de celles recherchées pour un signal, une bonne résolution en masse et un bon score de sortie du *Photon ID BDT*. Les évènements obtenant un score de sortie trop faible sont alors rejetés de l'analyse. Le réentraînement, uniquement pour 2017, et la validation du *diphoton BDT* pour les basses masses est une nouveauté par rapport à [103].

Similairement au *Photon ID BDT*, la *sortie* du *diphoton BDT* est un score prenant valeur entre -1 et 1, lié au ratio signal sur bruit de fond. Les évènements ressemblant le plus au signal  $h \rightarrow \gamma\gamma$  de l'analyse vont avoir un score proche de 1 alors que les évènements bruit de fond vont avoir une note plus proche de -1. Techniquement, ce score de sortie est proche d'un rapport de fonction de densités de probabilité normalisées entre signal et bruit de fond [107]. Un exemple de la distribution obtenue est visible sur la figure 6.18.

Le *diphoton BDT* est entraîné sur les simulations MC signal  $h \rightarrow \gamma\gamma$  pour les quatre modes de production principaux et pour une masse de 90 GeV pour le signal de l'entraînement. Le bruit de fond utilisé pour l'entraînement correspond aux bruits de fond QCD,  $\gamma$ +jets et  $\gamma$ + $\gamma$  de l'analyse. Les échantillons utilisés correspondent aux simulation 2017 passant toutes les

coupures décrites dans la section 6.4.

Les **variables d'entrée** permettant de discriminer entre le signal et le bruit de fond (correspondant, dans ce cas, au signal et bruit de fond de l'analyse) sont le score de sortie des BDT présentés précédemment, des variables cinématiques ainsi que des variables relatives à la résolution en masse transformées (normalisées par  $m_{\gamma\gamma}$  en général) de sorte que le BDT ne sélectionne pas les évènements en fonction de la masse du système diphoton afin de ne pas biaiser l'entraînement en fonction de la masse de l'échantillon signal utilisé. Ces variables d'entrée du *diphoton BDT* sont alors :

- $p_{vertex}$  le score de sortie du BDT présenté dans la section 6.5.2 correspondant à la probabilité pour le vertex identifié d'être situé à moins de 1 cm du vrai vertex
- le score de sortie du *Photon ID BDT* présenté dans la section précédente pour chaque photon (*lead* et *sublead*) de la paire diphoton
- $\eta^{(1)}$  et  $\eta^{(2)}$  pseudo-rapacité du photon *lead* et *sublead*
- $\cos(\Delta\phi)$  cosinus de l'angle  $\phi$  dans le plan transverse entre les deux photons
- $p_T^{(1)}/m_{\gamma\gamma}$  et  $p_T^{(2)}/m_{\gamma\gamma}$  impulsion transverse normalisée par la masse invariante diphoton du photon *lead* et *sublead*
- $\sigma^{rv}$  résolution en masse invariante diphoton dans l'hypothèse que le vertex identifié est le vrai vertex
- $\sigma^{wv}$  résolution en masse invariante diphoton dans l'hypothèse que le vertex identifié n'est pas le vrai vertex

La résolution en masse dans l'hypothèse de vertex correct se calcule simplement, sous l'hypothèse de distribution gaussienne, comme une incertitude sur la masse invariante en négligeant la résolution angulaire :

$$\sigma^{rv} = \frac{\sigma_m^{right}}{m_{\gamma\gamma}} = \frac{1}{2} \sqrt{\left(\frac{\sigma_{E1}}{E_1}\right)^2 + \left(\frac{\sigma_{E2}}{E_2}\right)^2} \quad (6.1)$$

où  $\sigma_E$  est la résolution en énergie pour un photon, discutée en détail dans le chapitre 4. Dans l'hypothèse de vertex incorrect, il faut prendre en compte la résolution angulaire entre les deux photons et on fait l'hypothèse que cette dernière est déterminée par l'intervalle de distance entre le vertex sélectionné, et le vrai vertex, que l'on considère égal à  $\sqrt{2}$  fois la longueur à mi-hauteur du *beamspot* faisceau [107] et modélisée par un terme  $\sigma_m^{vtx}$ . La résolution en masse pour le cas du vertex incorrect donne alors :

$$\sigma^{wv} = \frac{\sigma_m^{wrong}}{m_{\gamma\gamma}} = \sqrt{\left(\frac{\sigma_m^{right}}{m_{\gamma\gamma}}\right)^2 + \left(\frac{\sigma_m^{vtx}}{m_{\gamma\gamma}}\right)^2} \quad (6.2)$$

A noter que, lors de l'entraînement, les évènements signal sont pondérés de façon à assigner un meilleur score de sortie *diphoton BDT* aux évènements avec une bonne résolution en masse. Ce facteur de pondération  $w_{sig}$  est donné par la formule :

$$w_{sig} = \frac{p_{vertex}}{\sigma^{rv}} + \frac{1 - p_{vertex}}{\sigma^{wv}} \quad (6.3)$$



A cause du faible nombre d'évènements QCD, leur poids relatif dans l'entraînement est réduit d'un facteur 1/25 pour ne pas qu'ils perturbent l'entraînement.

Les distributions des variables d'entrée du BDT sont visibles sur la figure 6.14 pour le signal et le bruit de fond pour les simulations 2017 utilisées pour l'entraînement. On constate que ces distributions sont, en général, différentes entre signal et bruit de fond et sont donc discriminantes.

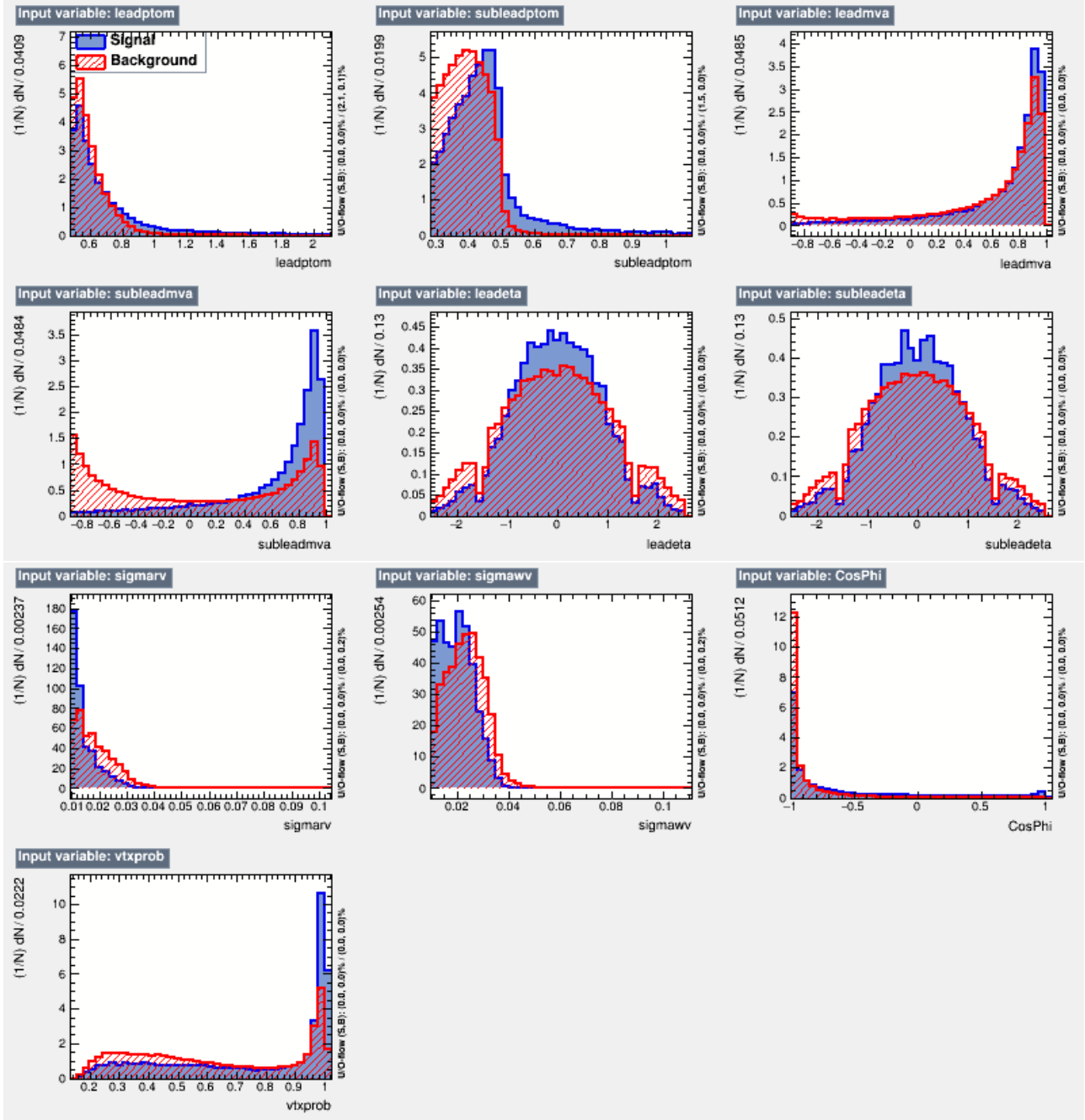


FIGURE 6.14 – Comparaison entre signal  $h \rightarrow \gamma\gamma$  (en bleu) et bruit de fond diphoton (rouge) des distributions des variables d'entrée du *diphoton BDT* pour les simulations 2017 pour l'impulsion transverse normalisée à la masse invariante diphoton ((sub)leadptom), le score de sortie du *Photon ID BDT* ((sub)leadmva),  $\eta$  ((sub)leadeta),  $\sigma^{(rv/wv)}$  (sigmarv/wv), le cos de l'angle  $\Phi$  (CosPhi) et  $p_{vertex}$  (vtxprob).



L'entraînement est effectué sur un quart des simulations 2017 et les trois quarts des évènements sont utilisés pour tester l'entraînement obtenu. Comme pour le *Photon ID BDT*, la fonction de sortie obtenue avec les évènements d'entraînement est appliquée aux évènements de test afin de déterminer leur distribution en terme de score de sortie *diphoton BDT* et la comparer à celle obtenue avec les évènements d'entraînement. La performance de ce score de sortie en terme d'efficacité de signal gardé et de bruit de fond rejeté est lisible sur les courbes dites "ROC", l'aire sous ce genre de courbe permet alors de décrire le degré de séparabilité entre les différents types d'évènements ce qui est utile pour comparer deux entraînements de nature différente. On peut ainsi voir sur la figure 6.15 (à gauche) que les courbes ROC sont quasiment identiques pour les évènements d'entraînement et pour les évènements de test, montrant qu'il n'y a pas de sur-entraînement et que l'entraînement réalisé est performant. Sur la droite de cette figure, on observe que la courbe ROC obtenue en utilisant des échantillons de signal  $h \rightarrow \gamma\gamma$  de masse 70 à 110 GeV par pas de 10 GeV pour l'entraînement est très similaire à celle obtenue avec un signal à 90 GeV pour les mêmes évènements de test, ce qui justifie le choix de n'utiliser que les échantillons de signal  $h \rightarrow \gamma\gamma$  à 90 GeV pour l'entraînement.

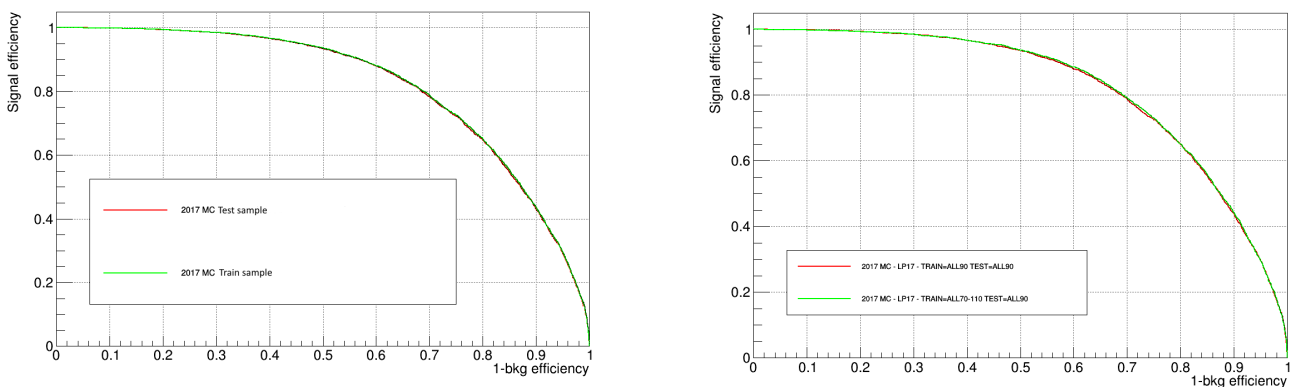


FIGURE 6.15 – Courbes ROC du *diphoton BDT* présentant l'efficacité du signal en fonction du rejet du bruit de fond pour les simulations 2017. A gauche, une comparaison des courbes ROC calculées avec les évènements d'entraînement (vert) et les évènements test (rouge). A droite, une comparaison de courbes ROC calculées sur des évènements de test où le signal utilisé à une masse de 90 GeV entre un entraînement réalisé avec un signal de masse 90 GeV (rouge) et un entraînement avec un signal avec des masses allant de 70 à 110 GeV par pas de 10 GeV (vert).

Comme mentionné tout au long de cette section, l'entraînement est basé sur les simulations 2017 et c'est cet entraînement qui est utilisé pour les données et simulations de 2017 ainsi que pour les données et simulations de 2018. Autrement dit, le *diphoton BDT* n'est pas réentraîné pour 2018. On peut justifier ce choix grâce à la figure 6.16 qui montre que les distributions de signal et de bruit de fond utilisées pour l'entraînement sont similaires, avec la même normalisation, entre 2017 et 2018 utilisant l'entraînement 2017 (gauche); les courbes ROC correspondantes pour 2017 et 2018 obtenues avec l'entraînement 2017 (droite) sont aussi très proches avec une différence de valeur de l'aire sous la courbe de l'ordre de 2 % ce qui est acceptable. Une justification supplémentaire sera apportée avec les validations données/simulations présentées dans la suite.

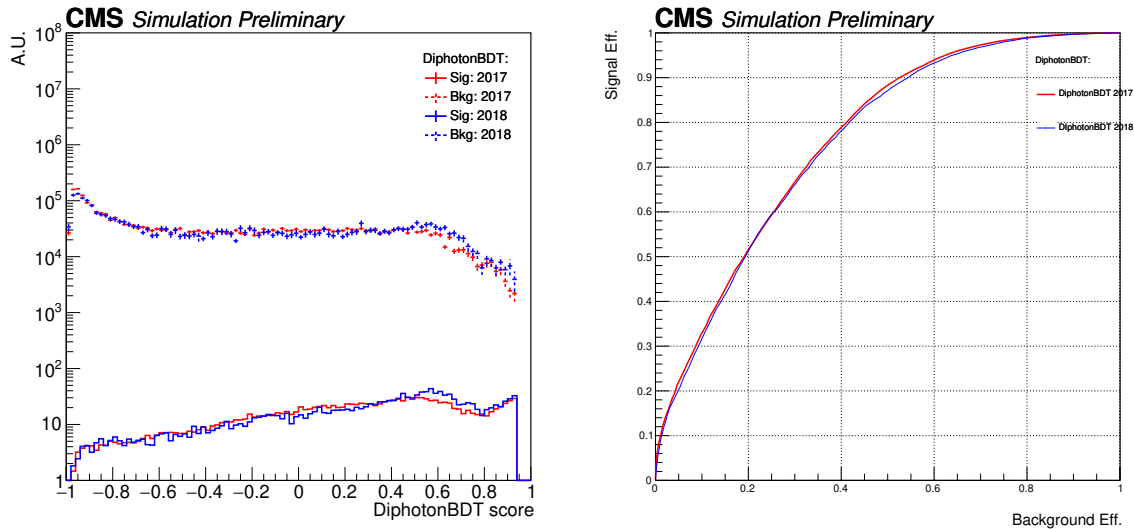


FIGURE 6.16 – Comparaisons des distributions signal et bruit de fond *diphoton BDT* (gauche) et des courbes ROC (droit) entre les simulations 2017 (rouge) obtenues avec l’entraînement 2017 présenté ici et les simulations 2018 (bleu) obtenus avec ce même entraînement. Les distributions de la figure droite sont normalisées à la même luminosité afin de rendre la comparaison plus aisée.

En effet, l’entraînement 2017 étudié ici est ensuite utilisé pour calculer le score de sortie *diphoton BDT* pour les données et simulations 2017 et 2018, on souhaite donc vérifier nos accords données/simulations pour ce score de sortie ainsi que pour les variables d’entrée du *diphoton BDT*. Ces validations sont effectuées par une comparaison directe entre les distributions de données (on utilise des zones de contrôles pour respecter la politique d’analyse à l’aveugle et éviter les biais éventuels) ainsi que des évènements  $Z \rightarrow ee$  dont la cinématique est proche de celle d’une désintégration  $h \rightarrow \gamma\gamma$ . Les évènements d’intérêt étant des évènements diphoton, à la différence du *Photon ID BDT*, il faut nécessairement des évènements contenant deux photons liés par un vertex pour la validation et on ne peut donc pas utiliser les évènements  $Z \rightarrow \mu\mu\gamma$  pour cette tâche.

### Validation entre données et simulations

On peut donc effectuer une première validation entre données et simulations pour les variables d’entrée et le score de sortie du *diphoton BDT* en comparant directement les distributions correspondantes. Cette comparaison est effectuée pour des zones de contrôle de données prises dans des gammes de masse qui ne sont pas utilisées pour l’obtention des résultats finaux, et en séparant le bruit de fond de l’analyse en composantes avec deux photons prompts, un photon prompt et un faux photon, et deux faux photons.

La gamme de masse de l’analyse est définie entre 65 et 120 GeV mais seules les masses de 70 à 110 GeV sont utilisées<sup>6</sup> pour dériver les résultats finaux, on peut donc utiliser les évènements avec une masse comprise entre 65 et 70 GeV et 110 et 120 GeV comme éléments de contrôle pour les données afin de ne pas risquer de biaiser l’analyse en manipulant les données complètes.

6. Cet élément sera développé au chapitre suivant.

La somme des contributions du bruit de fond est globalement normalisée au nombre total d'évènements dans les données. A noter qu'un échantillon QCD est manquant pour 2018<sup>7</sup>, dégradant de fait les comparaisons entre données et simulations.

La comparaison des distributions en masse invariante diphoton est visible sur la figure 6.17, les points de données correspondant aux régions sensibles de l'analyse sont cependant cachés afin d'éviter tout risque de biais. La masse invariante diphoton n'étant pas directement une variable d'entrée du *diphoton BDT*, cette comparaison données/simulations est donnée pour référence.

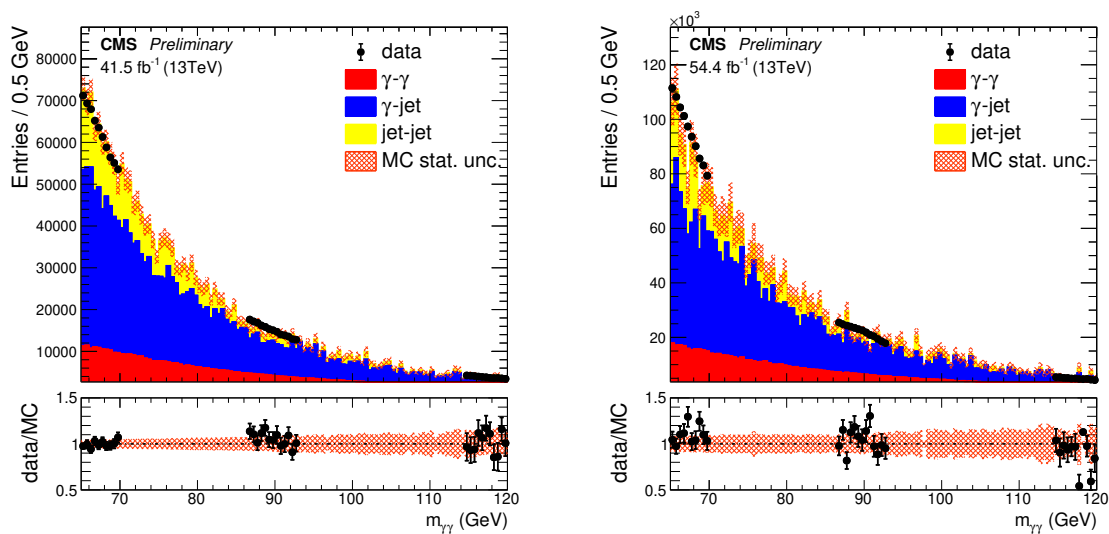


FIGURE 6.17 – Distribution de la masse invariante diphoton pour les données (points noirs), et les simulations séparées en différentes composantes comprenant les incertitudes statistiques associées (en hachuré) pour 2017 (gauche) et 2018 (droite). Les évènements données correspondant à des fenêtres de masse  $70 \text{ GeV} < m_{\gamma\gamma} < 87 \text{ GeV}$  et  $93 \text{ GeV} < m_{\gamma\gamma} < 115 \text{ GeV}$  sont masqués. Les parties basses des sous-figures présentent la distribution des rapports entre données et simulations.

La validation du score de sortie du BDT peut être observée sur la figure 6.18. La validation de quelques variables d'entrée est aussi visible sur la figure 6.19, la liste exhaustive des figures de validation est présentée, afin de conserver une certaine fluidité, en annexe D. Les accords entre données et simulations sont corrects dans tous les cas.

7. L'impact de cet échantillon sur le score du *diphoton BDT* et les catégories associées est faible.

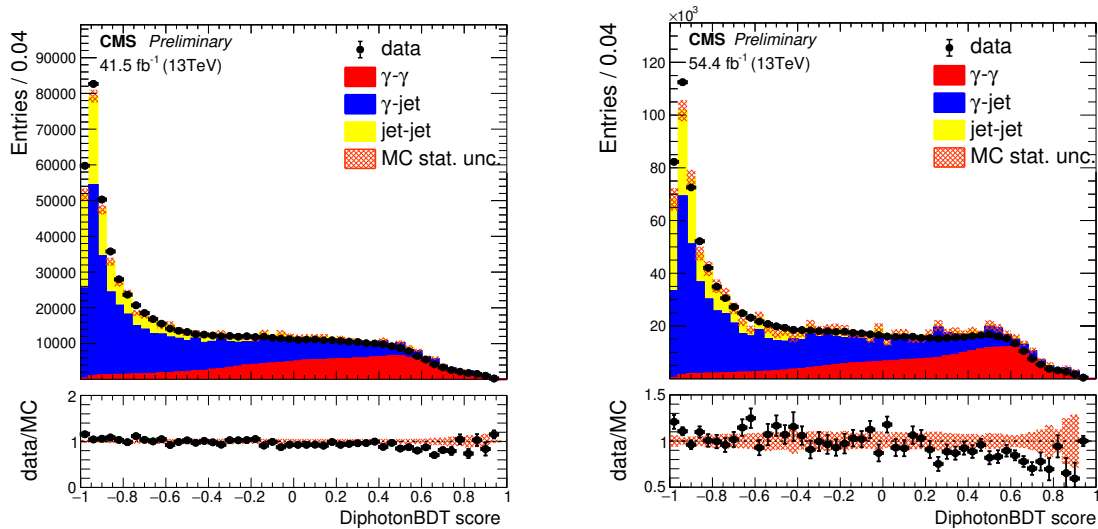


FIGURE 6.18 – Distribution en score de sortie du *diphoton BDT* pour les données (points noirs), et les simulations séparées en différentes composantes comprenant les incertitudes statistiques associées (en hachuré). Les données sont prises dans des zones de contrôle avec  $65 \text{ GeV} < m_{\gamma\gamma} < 70 \text{ GeV}$  et  $110 \text{ GeV} < m_{\gamma\gamma} < 120 \text{ GeV}$  pour 2017 (gauche) et 2018 (droite). Les parties basses des figures présentent la distribution des rapports entre données et simulations. A noter que l'échelle des ordonnées de la figure basse pour 2017 est différente de celle utilisée pour 2018.

### Validation avec les événements $Z \rightarrow ee$

La capacité de la simulation MC du signal à reproduire correctement l'éventuelle présence d'un vrai signal dans les données peut être validée en utilisant des électrons reconstruits comme des photons dans des événements  $Z \rightarrow ee$ . La comparaison entre la forme des variables pour des événements  $Z \rightarrow ee$  pour les données et les simulations représente donc une étape importante pour vérifier si les variables d'entrée du BDT sont suffisamment bien décrites dans les simulations pour que la méthode utilisée ici soit viable. Néanmoins, ce test présente quelques limites puisque le boson  $Z$  est intrinsèquement différent du signal  $h$  recherché : le boson  $Z$  est de spin 1 donc la cinématique sera notamment différente par rapport à une désintégration d'un boson de Higgs.

On s'intéresse ici aux deux photons du vertex diphoton donc on ne souhaite pas utiliser la méthode TnP. Les échantillons de données utilisées sont les données usuelles de l'analyse (double  $e/\gamma$ ) et on utilise les échantillons Drell-Yan (désintégration leptonique du  $Z$ ) pour les simulations. On sélectionne nos électrons, reconstruits comme des photons, en utilisant les mêmes sélections et corrections que celles décrites dans tout ce chapitre, en appliquant notamment la présélection de l'analyse mais en inversant le critère de veto électron de façon à sélectionner les événements diélectron qui ont alors des propriétés proches de celles des événements diphoton sélectionnés. Si le chemin HLT de l'analyse est utilisé pour les données de 2018, on est forcé à utiliser un autre chemin HLT pour les données de 2017 puisque ce chemin HLT inclut un veto électron en son sein qui rejette la majorité des électrons ; on utilise alors un chemin HLT similaire mais sans ce veto :

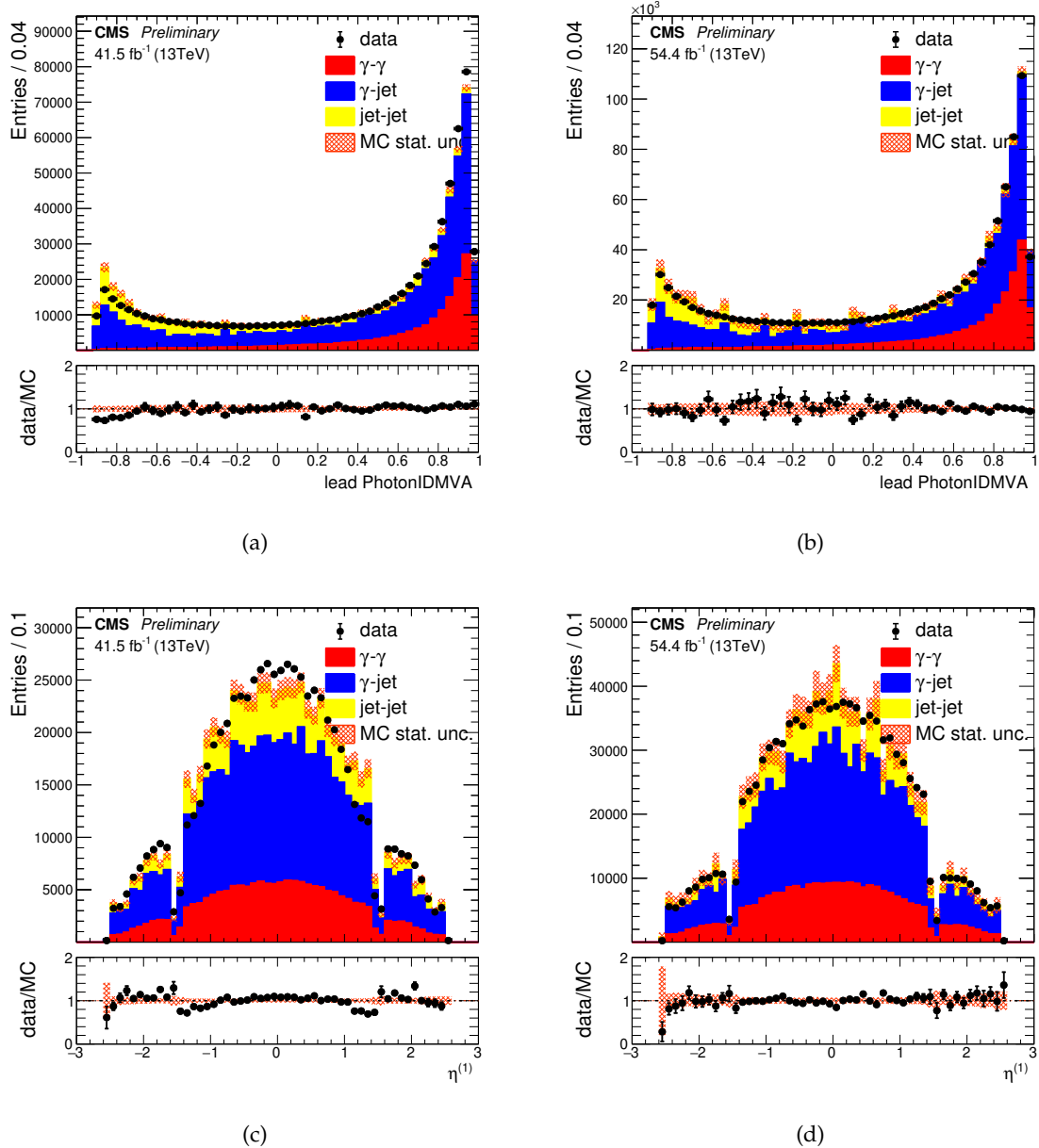


FIGURE 6.19 – Comparaisons données/simulations des variables d'entrée du *diphoton BDT* pour le score de sortie du *Photon ID BDT* du photon *lead* de plus haute impulsion transverse pour a) 2017, b) 2018, et pseudo-rapidité du photon de plus haute impulsion transverse pour c) 2017, d) 2018.

HLT\_Diphoton30PV\_18PV\_R9Id\_AND\_IsoCaloId\_AND\_HE\_R9Id\_NoPixelVeto\_Mass55; ce chemin a la particularité d'être "prescalé" d'un facteur d'environ 18 (un évènements sur  $\sim 18$  est enregistré). On requiert de plus que les évènements aient une masse proche de celle du Z soit  $80 < m_{ee} < 100$  GeV.

Pour prendre en compte la cinématique du boson Z, une repondération des évènements provenant des simulations est effectuée en bin de  $p_{T,ee}$  afin que la distribution de cette variable soit en accord avec celle des données, comme visible sur la figure 6.20. Cette repondération améliore l'accord données/simulations pour des variables corrélées avec  $p_{T,ee}$  soit la  $p_{vertex}$ , les  $p_T/m$  ou encore le cosinus de l'angle du système diphoton. On normalise aussi ces simulations, par un facteur global, afin d'obtenir le même nombre d'évènements que dans les données.

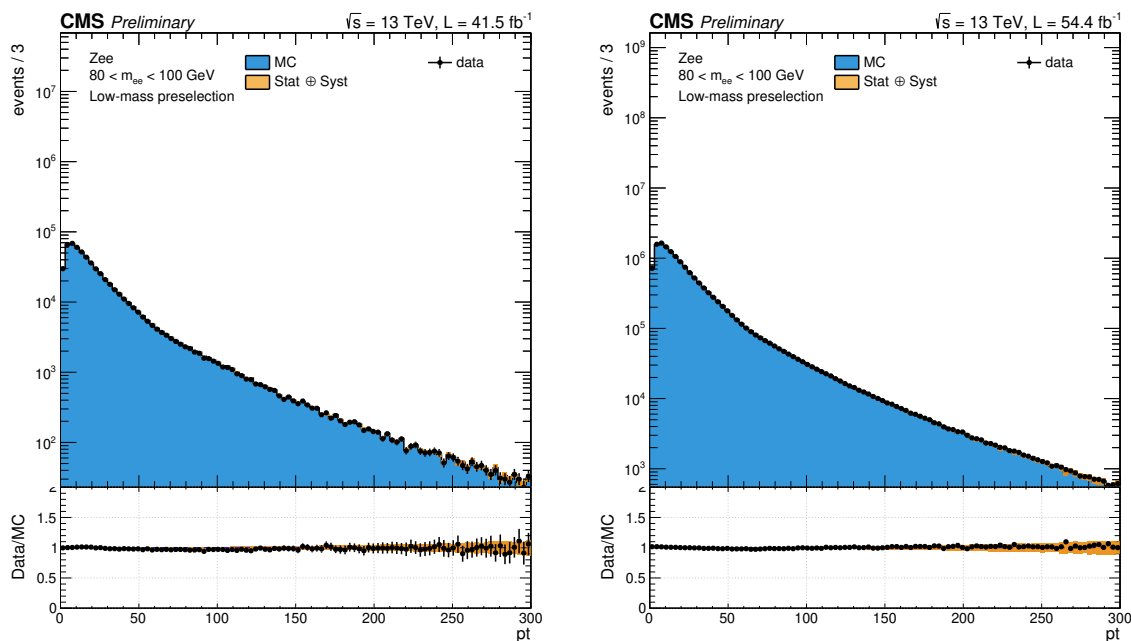


FIGURE 6.20 – Comparaison données/simulations après repondération des simulations pour 2017 (gauche) et 2018 (droite) de la distribution  $p_T$  du système diélectron où les électrons sont reconstruits comme des photons. La comparaison entre données (points noirs) et simulations (bleu plein) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.

Les deux incertitudes systématiques calculées dans la section précédente sont prises en compte ici : l'incertitude sur le score de sortie du *Photon ID BDT* pour chacun des deux photons utilisée dans les deux variables correspondantes, et l'incertitude sur la résolution en énergie par photon utilisé dans les calculs de  $\sigma^{FV}$  et  $\sigma^{WV}$ . Ces incertitudes sont propagées jusqu'au score de sortie du *diphoton BDT* et permettent de couvrir les accords entre données et simulations comme on peut le voir sur la figure 6.21. On observe aussi une différence de forme de la distribution du score de sortie du BDT entre 2017 et 2018 pour les évènements  $Z \rightarrow ee$ . Évidemment les chemins HLT et la présélection sont différents entre ces deux années ce qui ne permet pas une comparaison parfaite; notamment, on autorise des évènements de bas  $R_9$  en EE pour 2017 qui

sont des électrons de mauvaise qualité; le  $R_9$  pour des événements  $Z \rightarrow ee$  est aussi décalé vers des valeurs plus élevées pour 2018 ce qui résulte en un même léger décalage du score *Photon ID BDT* pour les électrons, on ajoute à cela une meilleure résolution en masse pour 2018 que l'on propage jusqu'au score *diphoton BDT* et on peut ainsi expliquer ce décalage vers de meilleurs scores.

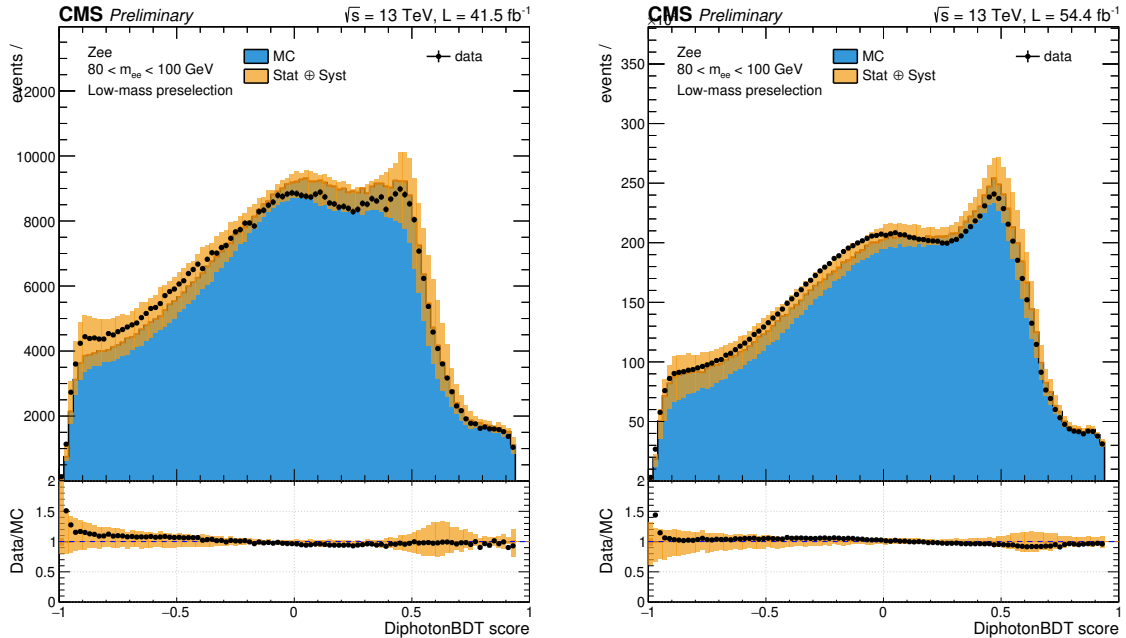


FIGURE 6.21 – Comparaison données/simulations pour 2017 (gauche) et 2018 (droite) de la distribution de score de sortie *Combined MVA* du système diélectron où les électrons sont reconstruits comme des photons. La comparaison entre données (points noirs) et simulations (couleur pleine) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.

La validation de quelques variables d'entrée est aussi visible sur la figure 6.22, la liste exhaustive des figures de validation est présentée, afin de conserver une certaine fluidité, en annexe D. Les accords entre données et simulations sont bons et compris dans les incertitudes associées. L'entraînement du *diphoton BDT* est donc fiable, aussi bien sur les données 2017 que sur les données 2018.

### 6.7.2 Catégorisations

Afin de maximiser la sensibilité de l'analyse, on souhaite maintenant catégoriser les événements données et simulations en fonction de leur ressemblance avec du signal ou du bruit de fond en utilisant le score de sortie du *diphoton BDT*. On cherche alors à obtenir des classes avec des événements de type signal et un ratio signal sur bruit de fond élevé mais avec une faible statique globale, et des classes avec des événements avec de moins bons scores de sortie qui auront donc plus de statistiques. Le nombre ainsi que la taille/frontières de ces classes sont optimisés par un algorithme dédié [107, 109] utilisant les simulations MC.

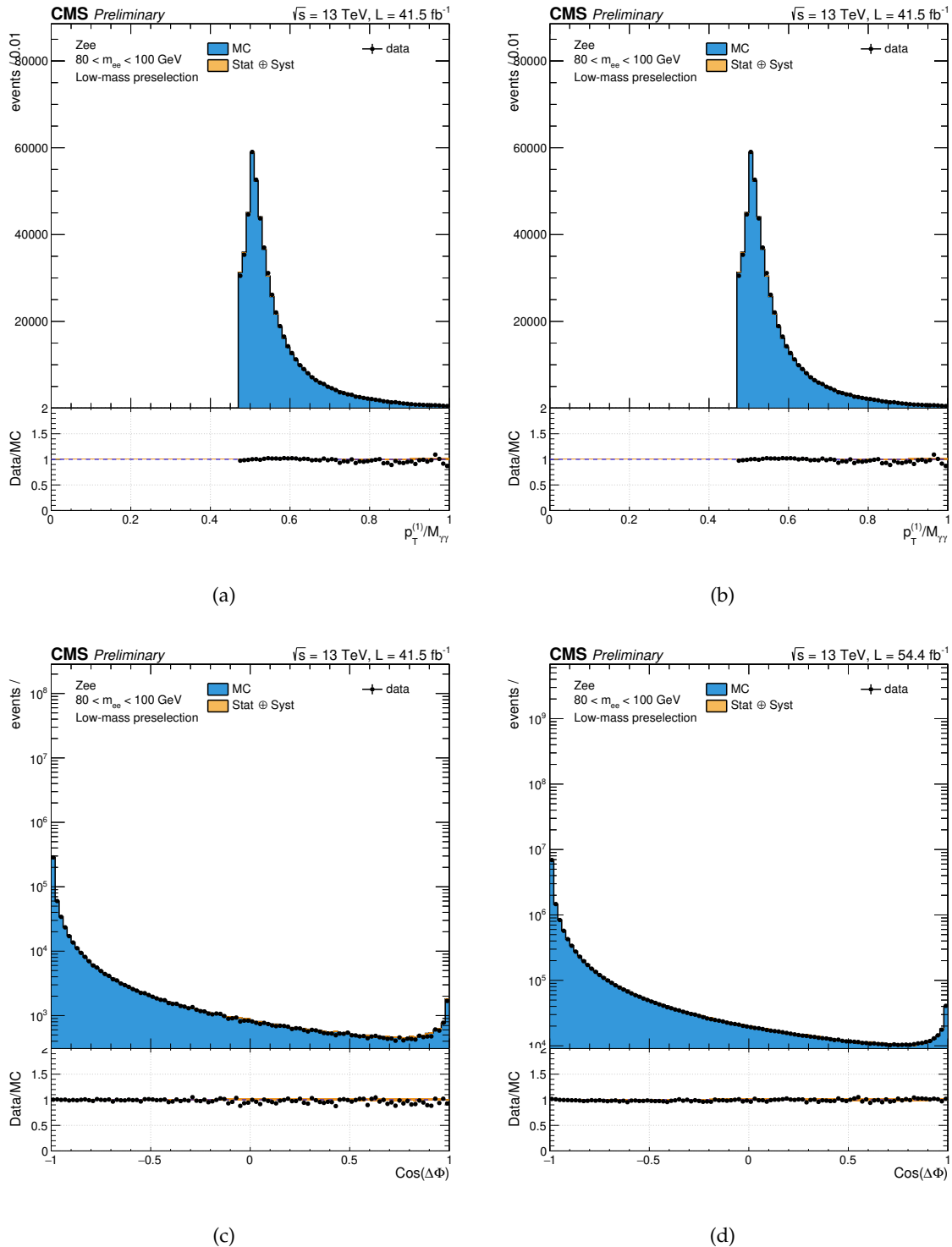


FIGURE 6.22 – Comparaison données/simulations de la distribution de l’impulsion transverse du photon dominant normalisée par la masse invariante diphoton pour a) 2017, b) 2018, et du cosinus de l’angle du système diphoton pour c) 2017, d) 2018, pour des électrons reconstruits comme des photons. La comparaison entre données (points noirs) et simulations (bleu plein) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.



Initialement,  $n + 1$  frontières sont placées sur la distribution de score de sortie à une distance égale l'une de l'autre en terme d'évènements signal, 2 des frontières étant placées aux valeurs -1 et 1 du score de sortie. Les positions des frontières sont ensuite libres de varier et sont ajustées une à une afin de minimiser une fonction de vraisemblance (selon une méthode proche de la p-value attendue décrite dans la section 6.11 et semblable à celle utilisée dans [107]) combinée rendant compte de la sensibilité de l'analyse. Cette figure de mérite est obtenue en générant des pseudo-données dans chaque catégorie de frontières variables à partir de distribution simplifiée, afin de faciliter le calcul, de signal et de bruit de fond. Ainsi le signal est ajusté par une somme de deux gaussiennes, une dans l'hypothèse de bonne et une dans l'hypothèse mauvaise identification du vertex et le bruit de fond est ajusté par une exponentielle décroissante.

Les échantillons utilisés sont les mêmes que ceux utilisés pour l'entraînement *diphoton BDT* décrits plus hauts et les mêmes conditions, le bruit de fond étant normalisé au nombre total d'évènements des données. L'échantillon Drell-Yan n'est pas inclus dans ce bruit de fond par simplification. En raison du faible nombre d'évènements normalisés attendus pour cet échantillon, les résultats de l'algorithme d'utilisation ne dépendent de toutes façons quasiment pas de sa présence au sein des échantillons bruit de fond.

La détermination de la forme de la distribution en masse du bruit de fond  $Z \rightarrow ee$  de l'analyse étant effectuée à partir de simulations MC, il est nécessaire, pour cette analyse, d'assurer que les catégories obtenues possèdent assez d'évènements DY  $Z \rightarrow ee$  afin que les ajustements réalisés par catégorie dans la section 6.9 soient corrects et que les valeurs des paramètres d'ajustement obtenues soient raisonnables et bien définies par rapport à leur incertitude statistique. Pour être certain de satisfaire à cette condition, une taille minimale en terme de score BDT a été implémentée pour chaque classe. La classe avec les évènements les plus performants, appelée classe 0, c'est-à-dire les évènements avec les meilleurs ratio signal sur bruit donc les meilleurs score de sortie *diphoton BDT*, étant celle avec le moins d'évènements en général, il est suffisant d'uniquement appliquer cette taille minimale à cette classe 0.

Pour déterminer cette taille minimale, les ajustements des fonctions DY  $Z \rightarrow ee$  par une fonction dCB+exp sont réalisés pour différentes tailles de la classe 0. La taille minimale est alors choisie comme la différence entre 1 et le score *diphoton BDT* permettant d'obtenir des paramètres bien définis et situés sur un plateau. Comme on peut le voir sur la figure 6.23 pour les simulations 2017, tous les paramètres sont bien définis et les valeurs des paramètres, notamment le  $\sigma$ , sont relativement stables pour des valeurs, conservatrices, de score *diphoton BDT* inférieures (ou proches) à 0,75. Cette valeur d'environ 0,75 est donc choisie comme frontière inférieure de la classe 0 de taille minimale, ce qui correspond à une taille minimale de  $1-0,75=0,25$  en terme de score *diphoton BDT* pour la classe 0.

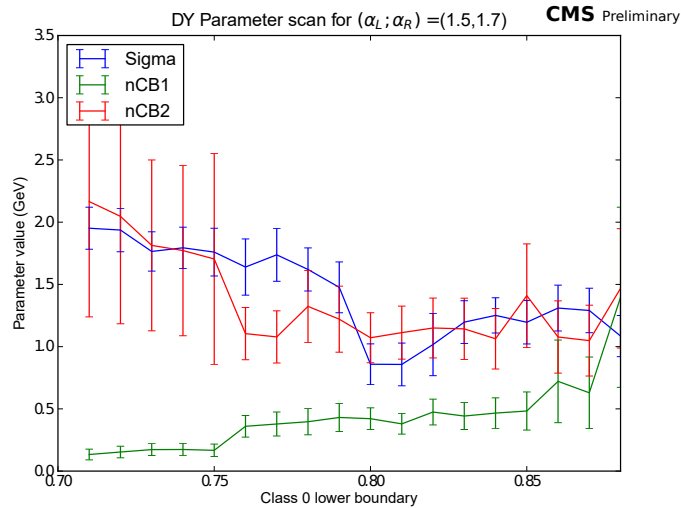


FIGURE 6.23 – Valeur de quelques uns des paramètres d’ajustement de la distribution en masse d’évènements  $DY Z \rightarrow ee$  simulés en fonction de la frontière inférieure, en terme de score de sortie *diphoton BDT*, de la classe 0 qui contient les évènements les plus performants.

La taille minimale pour la classe 0 est donc incorporée dans la procédure d’optimisation des catégories décrites précédemment. L’optimisation renvoie une figure de mérite correspondant à la sensibilité attendue en fonction du nombre de catégories à utiliser dans l’analyse. Cette figure de mérite ainsi que la valeur des frontières associée est visible sur la figure 6.24. La figure de mérite ne varie que très peu en passant de  $n=3$  à  $n=4$  voire  $n=5$  catégories ; j’ai par ailleurs étudié que ces différences dans la figure de mérite ne donneront pas lieu à d’amélioration significative des résultats finaux. On peut donc raisonnablement choisir  $n=3$  catégories dont les frontières associées sont [ 1,000 ; 0,753 ; 0,334 ; -0,364 ] en terme de score de sortie de *diphoton BDT* fournies par la figure 6.24 (droite). Ainsi la classe 0 a des évènements avec des score de sortie entre 1 et 0,753, la classe 1 entre 0,753 et 0,334, et la classe 2 entre 0,334 et -0,364. Les évènements avec un score  $< -0,364$  sont rejetés de l’analyse, ils sont, à ce point, considérés comme trop semblables à des évènements bruit de fond.

Ces catégories et frontières associées sont estimées sur des simulations 2017 et appliquées aux données et simulations 2017 ainsi que 2018. L’optimisation des catégories a aussi été réalisée pour 2018 mais les résultats finaux obtenus avec ces nouvelles catégories n’étaient pas significativement différents de ceux obtenus avec les catégories 2017 ; il a donc été choisi de conserver les catégories de 2017 pour 2018. Cette étude, que j’ai aussi réalisée, sera présentée plus en détail dans le chapitre suivant. Ces catégories seront utilisées dans toute la suite.

Aucune incertitude directe n’est dérivée pour la classification des évènements, cependant les incertitudes systématiques sur le score du *diphoton BDT* (venant des incertitudes systématiques sur le score *Photon ID BDT* et de la résolution en énergie) vont provoquer des fluctuations de ce score pour les évènements ce qui va créer une migration d’évènements entre les catégories définies ici. Les incertitudes systématiques décrites dans les parties précédentes seront aussi propagées par catégorie.

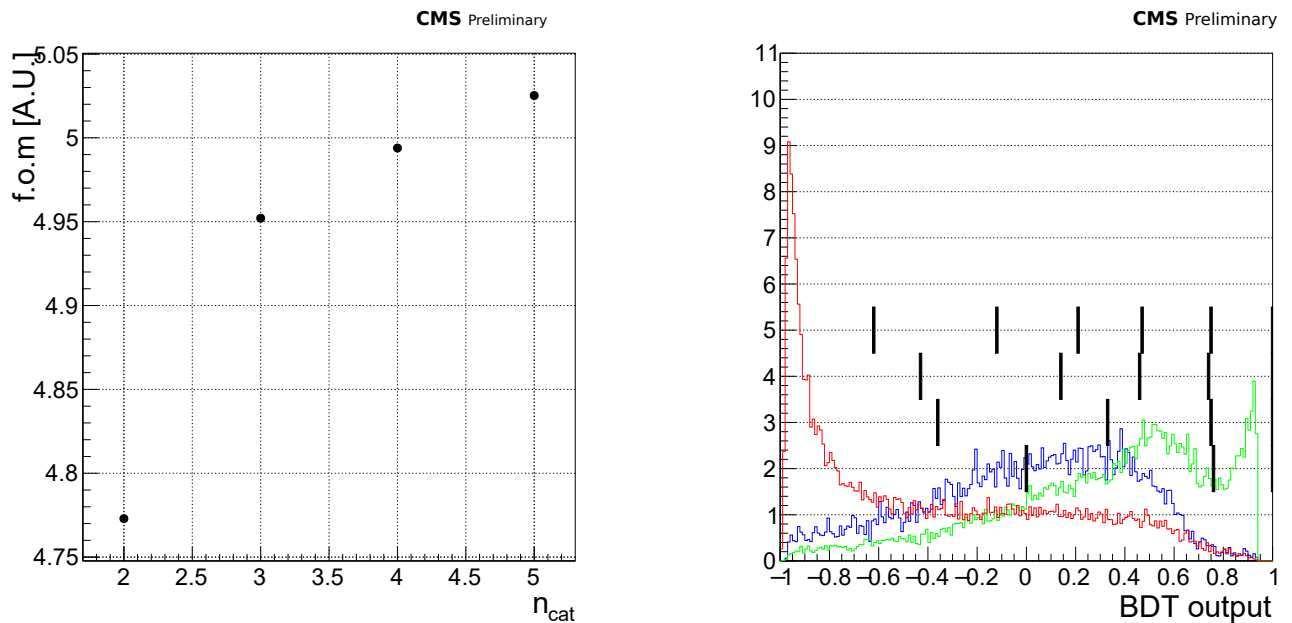


FIGURE 6.24 – Figure de mérite, correspondant à une  $p$ -value, obtenue à la suite de l’optimisation des catégories de l’analyse basée sur des simulations 2017 en fonction du nombre de catégories  $n_{cat}$  (gauche). Les frontières associées (barres noires verticales) sont visibles sur la figure de droite avec les distributions du score de sortie *diphoton BDT* pour les simulations signal dans l’hypothèse de vertex correct (vert) et de mauvais vertex (bleu) et bruit de fond (rouge).

## 6.8 Modélisation du signal

Afin de caractériser la forme de l’excès recherché dans les données, il est nécessaire de construire un modèle de signal à partir de simulations MC ; ce modèle sera ensuite utilisé dans les tests statistiques permettant d’obtenir les résultats finaux en terme d’exclusion ou de potentiel de découverte. Il est donc important que nos simulations soient en accord avec nos données, d’où l’utilisation de diverses corrections, facteurs d’échelle et incertitudes systématiques qui sont propagés aux simulations signal utilisées ici. L’objectif est de paramétrer le signal de façon continue en  $m_h$  en utilisant les simulations de données  $h \rightarrow \gamma\gamma$  pour les quatre modes de production (ggH, ttH, VBF et VH) pour des échantillons avec des masses allant de 70 à 115 GeV par pas de 5 GeV. La méthode utilisée, développée dans [107], est légèrement différente de celle utilisée dans [103]. Les études présentées dans cette section sont essentiellement le fait de mon travail.

On souhaite avoir un modèle paramétrique qui sera ajusté sur la masse invariante diphoton des simulations de signal pour chaque point de masse, chaque mode de production du Higgs, chaque classe en terme de score *diphoton BDT*, et pour les cas de bonne (moins de 1 cm en  $z$  du vrai vertex) et mauvaise (plus de 1 cm en  $z$  du vrai vertex) identification du vertex, par une somme de gaussiennes.

Le nombre de gaussiennes utilisées pour l’ajustement est donné par le résultat d’un F-test qui consiste à incrémenter le nombre  $N$  de gaussiennes jusqu’à ce que l’amélioration

Année	2017						2018					
	bon vertex			mauvais vertex			bon vertex			mauvais vertex		
Vertex	0	1	2	0	1	2	0	1	2	0	1	2
Mode de production \ Classe	0	1	2	0	1	2	0	1	2	0	1	2
ggH	4	4	4	3	3	3	5	4	4	2	3	3
VBF	3	3	3	3	3	2	4	4	3	2	3	3
WH	3	4	4	2	3	3	4	4	4	2	3	3
ZH	4	4	3	2	3	3	4	4	4	3	3	3
ttH	4	4	3	2	2	2	4	4	3	2	2	2

TABLE 6.11 – Nombre de gaussiennes utilisées pour la modélisation du signal par classe, mode de production et cas bonne et mauvaise identification du vertex pour 2017 et 2018.

de la qualité de l’ajustement de la masse invariante diphoton ne soit plus significative. Ce nombre est déterminé séparément pour les cas de simulations signal pour chaque mode de production du Higgs, chaque classe en terme de score *diphoton BDT*, et pour les cas de bonne et mauvaise identification du vertex, mais uniquement pour une masse de 90 GeV. Plus précisément, on calcule une différence de vraisemblances associées à l’ajustement de  $m_{\gamma\gamma}$  à l’ordre N+1 et N qui est alors distribuée selon une loi de  $\chi^2$  selon le théorème de Wilks [110] et on impose un critère de qualité sur la p-value obtenue (voir la sous-section 5.3.1 pour plus de détails) avec ce  $\chi^2$ , correspondant à la probabilité d’obtenir un meilleur  $\chi^2$  en utilisant N+1 gaussiennes pour l’ajustement. Considérons une fonction de log-vraisemblance  $LL_N = -\ln(m_{\gamma\gamma}|G_N)$  minimisant l’ajustement de la masse invariante diphoton d’un échantillon signal avec une somme de gaussiennes d’ordre N. La différence entre l’ordre N+1 et N est alors donnée par le  $\chi^2$  correspondant :

$$\chi_{N \rightarrow N+1}^2 = 2 \times (LL_N - LL_{N+1}) \quad (6.4)$$

On considère alors que l’amélioration de la qualité d’ajustement en passant de l’ordre N à l’ordre N + 1 n’est plus significative si la probabilité d’obtenir une valeur de  $\chi^2$  supérieure à valeur de ce  $\chi^2$  observé est inférieure à 5 %, soit en terme de p-value :

$$\text{p-value}(\chi^2 > \chi_{N \rightarrow N+1}^2) < 0,05 \quad (6.5)$$

Si cette condition est satisfaite, on choisit alors N comme nombre de gaussiennes à utiliser pour les ajustements. Un exemple d’ajustement réalisé durant le F-test est visible sur la figure 6.25. Le nombre de gaussiennes N obtenu se situe, selon les modes de production et les classes, entre 3 et 5 pour le cas bon vertex et seulement entre 2 et 3 pour le cas mauvais vertex puisque la distribution en masse y est plus étalée<sup>8</sup>, pour 2017 et 2018 comme visible sur le tableau 6.11.

Les paramètres de la fonction analytique, la somme de N gaussiennes, utilisée pour la paramétrisation, et définie de façon continue en masse invariante diphoton, sont des polynômes en  $m_H$ , masse du boson de Higgs considérée. Ces polynômes sont des polynômes d’ordre 1 en  $m_H$ ; cet ordre a été montré [107] comme étant suffisant pour cette modélisation. On ajuste cette fonction simultanément sur tous les points de masse par mode de production, classe et cas bon/mauvais vertex. Les contributions bon/mauvais vertex sont ensuite sommées avec une

8. Comme étudié dans la section précédente, la résolution en masse n’est plus dominée par la résolution en énergie.

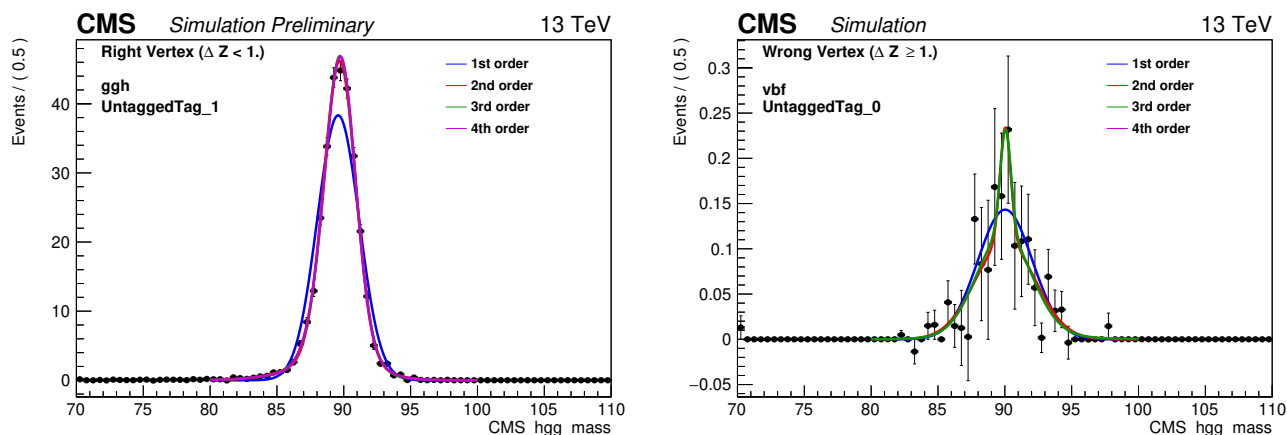


FIGURE 6.25 – Ajustement de la masse invariante diphoton pour des simulations de signal à 90 GeV par des sommes de gaussiennes d’ordre différent au sein du F-test. A gauche, le mode de production ggH pour 2017 dans le cas d’une bonne identification du vertex et pour la classe 1 de la classification *diphoton BDT*. A droite, le mode de production VBF pour 2017 dans le cas d’une mauvaise identification du vertex et pour la classe 0 de la classification *diphoton BDT*.

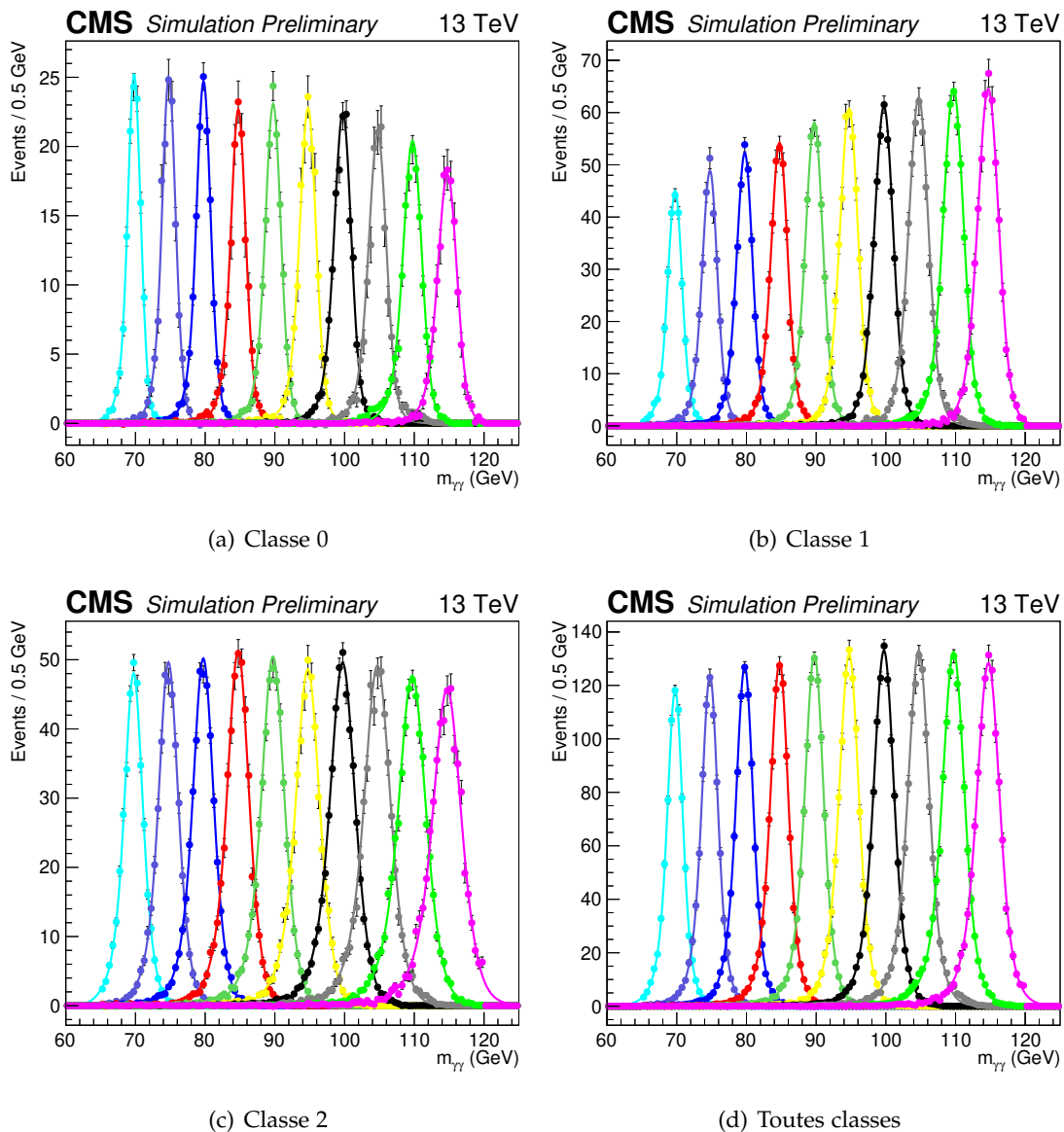
fraction adéquate. On obtient alors un modèle décrivant le signal pour une gamme de masse continue et par mode de production.

Les figures 6.26 à 6.29 montrent respectivement les ajustements pour chaque échantillon de masse, en guise d’illustration, et les modèles analytiques de signal obtenus interpolés pour les simulations 2017 et 2018 pour chaque classe. Les contributions des différents mode de production ont été sommées en fonction de leur section efficace. Le modèle paramétrique final du signal évalué à une masse de 90 GeV est représenté sur les figures 6.30 et 6.31 pour 2017 et 2018, la résolution en masse est inférieure à 1,5 GeV pour les meilleures classes. Cette méthode de modélisation est plus stable que la méthode utilisée dans [103] qui repose sur un simple interpolation linéaire des paramètres d’ajustement pour chaque échantillon.

Des incertitudes systématiques supplémentaires, pouvant modifier la forme du spectre de masse invariante, correspondant aux échelles d’énergie des photons, aux fractions d’évènements où le vrai vertex a été correctement identifié et à des corrections ECAL sont incorporées dans ce modèle de signal, plus de détails peuvent être trouvés dans la section 6.10. On prend aussi en compte une incertitude systématique pour le phénomène de *prefiring* du système de déclenchement, expliqué dans la section 6.3. On peut voir sur le tableau 6.12 que ce phénomène est responsable d’une réduction pouvant aller jusqu’à 5 % de la population des évènements signal.

TABLE 6.12 – Pourcentage d'évènements de simulation du signal perdus à cause du *prefiring* par classe et mode de production pour une masse de 90 GeV.

Tableau d'évènements signal à 90 GeV perdus à cause du <i>prefiring</i> du système de déclenchement par classe			
Mode de production	Classe 0 (% perdu)	Classe 1 (% perdu)	Classe 2 (% perdu)
ggH	$2.35 \pm 0.06$	$0.61 \pm 0.01$	$0.95 \pm 0.01$
VBF	$4.8 \pm 0.2$	$2.51 \pm 0.09$	$3.2 \pm 0.2$
WH	$1.43 \pm 0.02$	$1.38 \pm 0.02$	$2.44 \pm 0.03$
ZH	$1.24 \pm 0.03$	$1.23 \pm 0.02$	$2.23 \pm 0.04$
tth	$2.53 \pm 0.03$	$2.36 \pm 0.03$	$3.23 \pm 0.07$


 FIGURE 6.26 – Ajustement des échantillons signal 2017 par classe pour une luminosité de  $41,5 \text{ fb}^{-1}$ . Les contributions bonne et mauvaise identification de vertex sont sommées selon leur fraction respective, et les contributions des modes de production sont sommées selon leur section efficace.

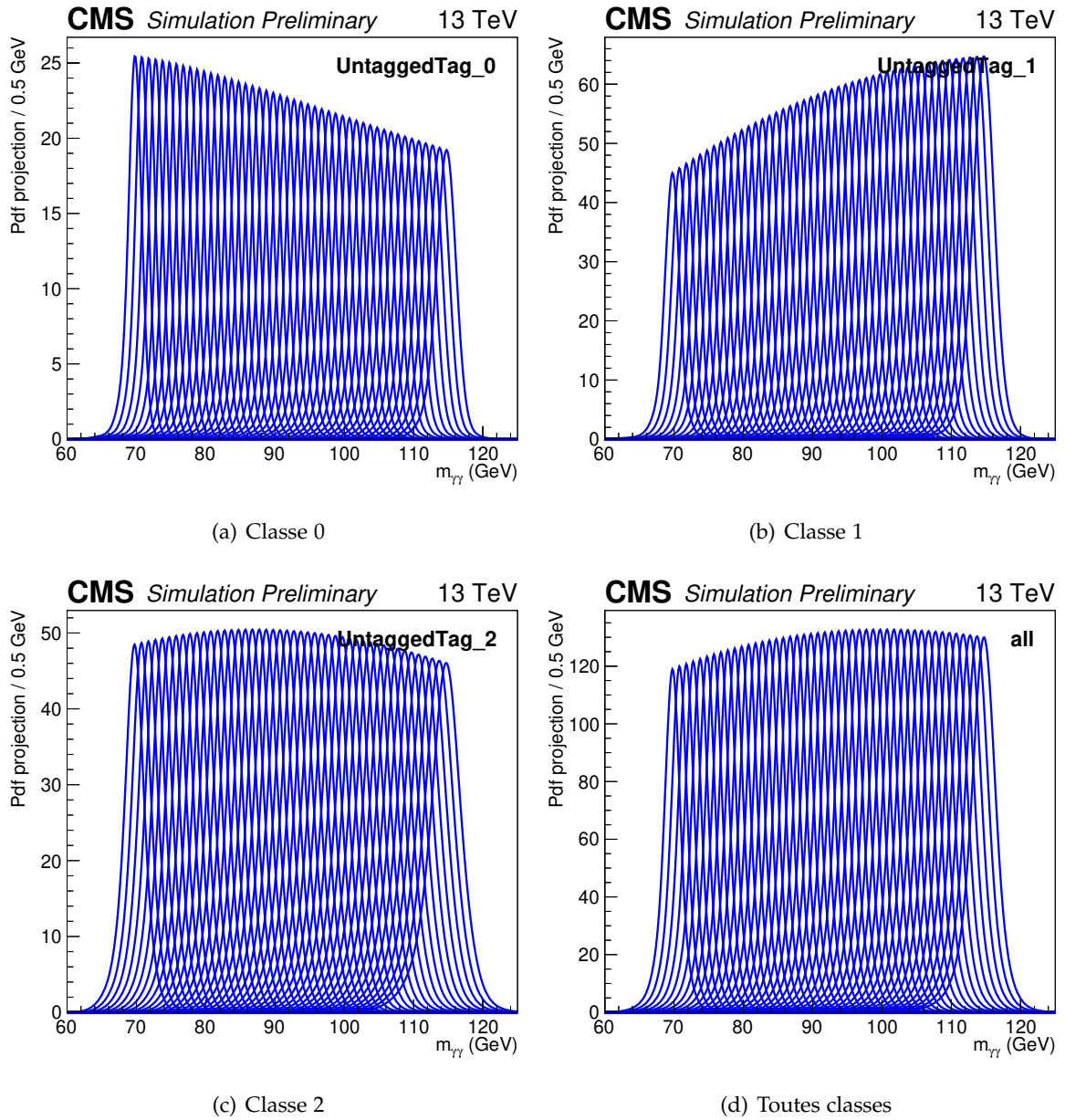


FIGURE 6.27 – Modèle paramétrique du signal 2017 pour une luminosité de  $41,5 \text{ fb}^{-1}$  illustré ici avec des intervalles de masse de 1 GeV par classe. Les contributions bonne et mauvaise identification de vertex sont sommées selon leur fraction respective, et les contributions des modes de production sont sommées selon leur section efficace.



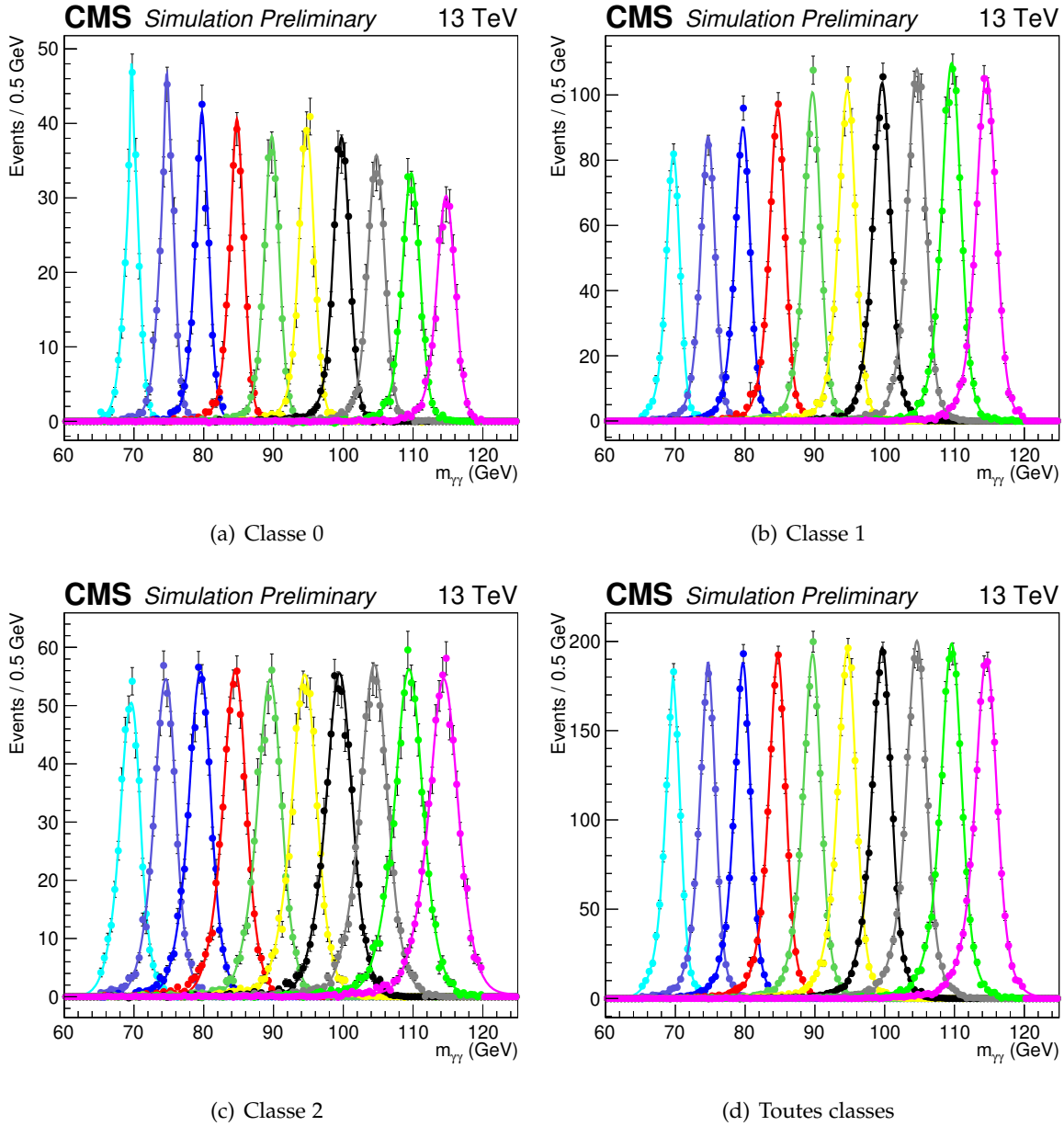


FIGURE 6.28 – Ajustement simultané des échantillons signal 2018 pour une luminosité de  $54,4 \text{ fb}^{-1}$  par classe. Les contributions bonne et mauvaise identification de vertex sont sommées selon leur fraction respective, et les contributions des modes de production sont sommées selon leur section efficace.



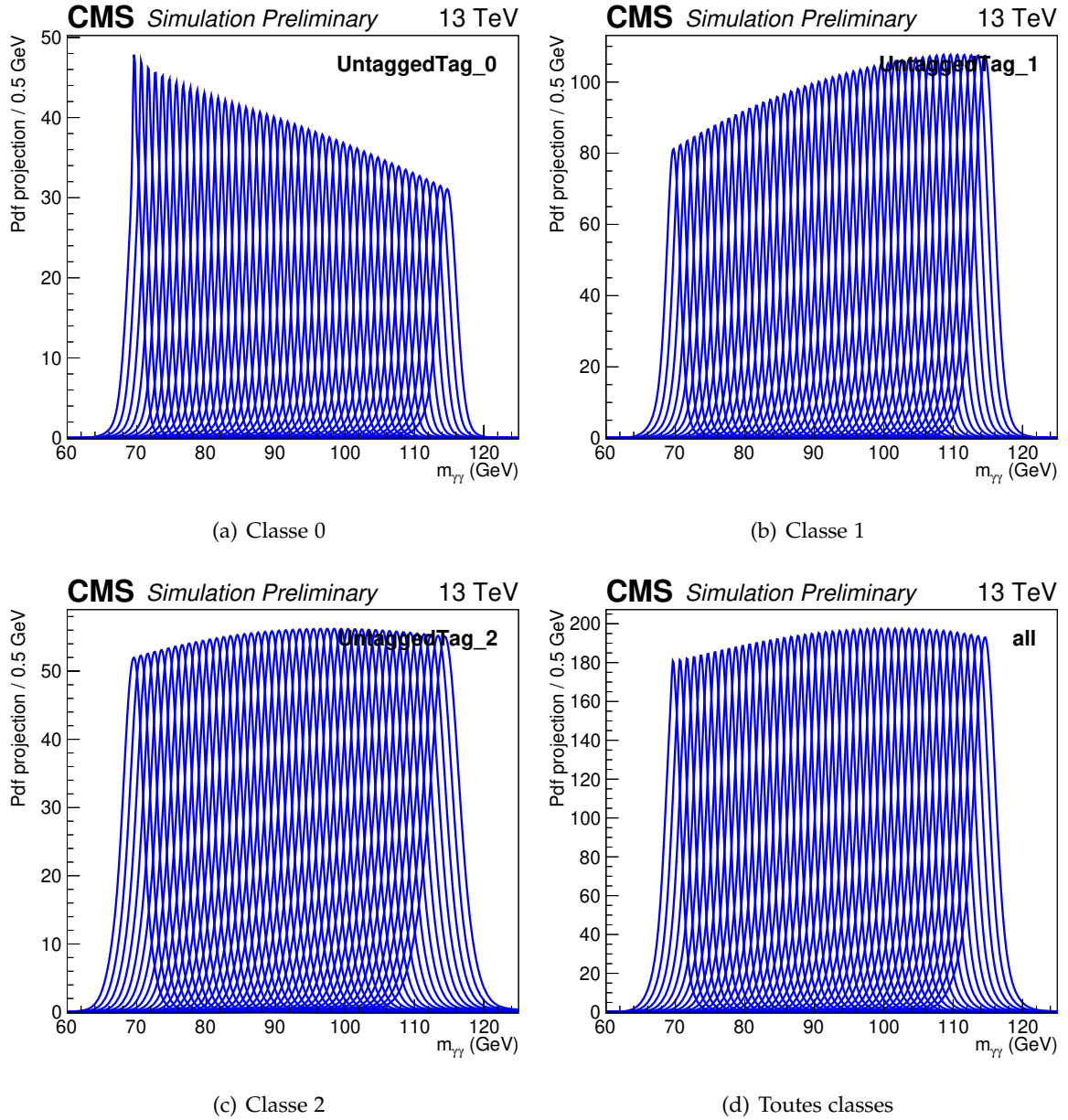


FIGURE 6.29 – Modèle paramétrique du signal 2018 pour une luminosité de  $54,4 \text{ fb}^{-1}$  illustré ici avec des intervalles de masse de 1 GeV par classe. Les contributions bonne et mauvaise identification de vertex sont sommées selon leur fraction respective, et les contributions des modes de production sont sommées selon leur section efficace.

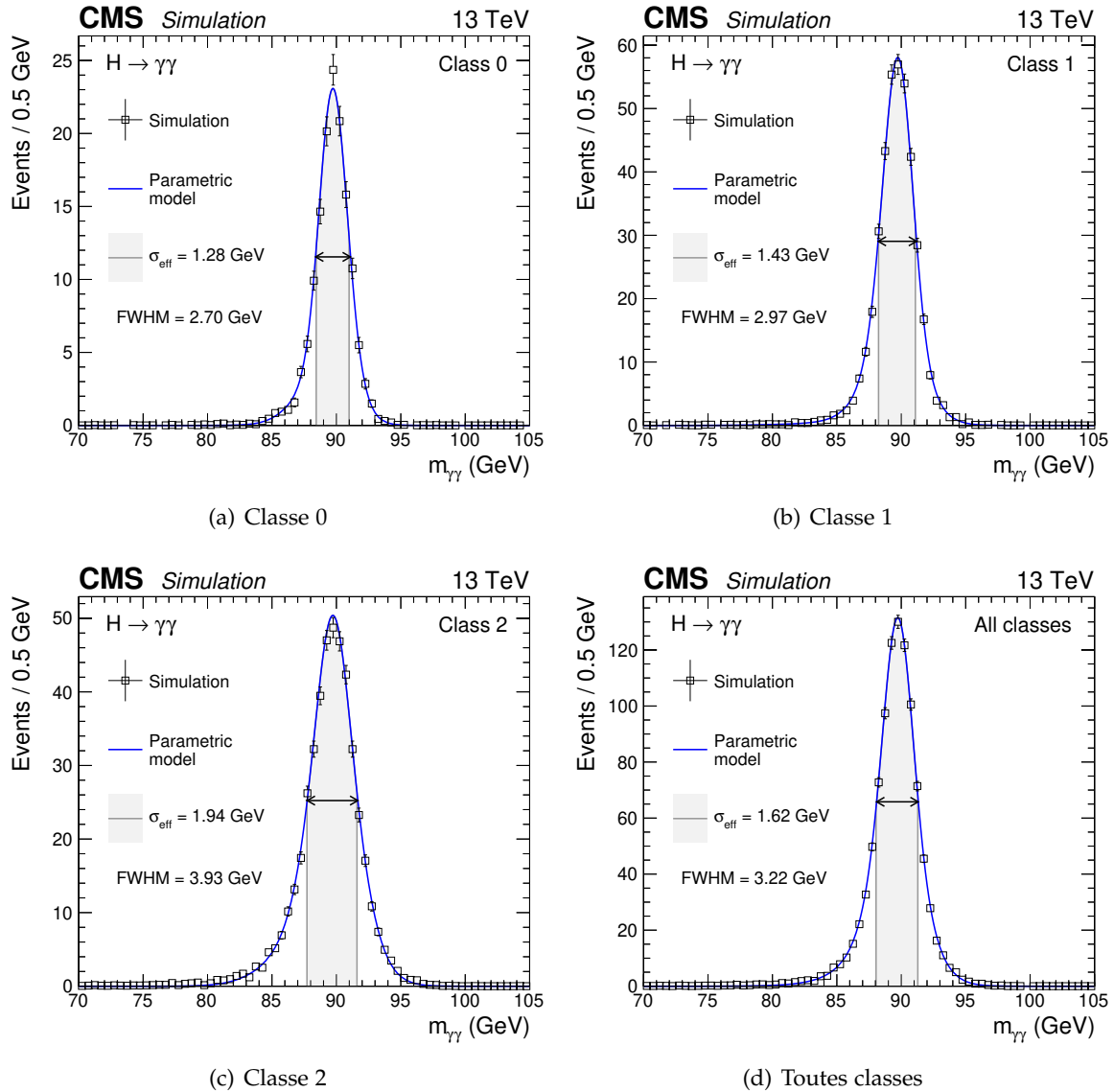


FIGURE 6.30 – Modèle paramétrique du signal 2017 pour une luminosité de  $41,5 \text{ fb}^{-1}$ , évalué à une masse de 90 GeV, par classe. Les contributions bonne et mauvaise identification de vertex sont sommées selon leur fraction respective, et les contributions des modes de production sont sommées selon leur section efficace. Les points blancs correspondent aux simulations et la courbe bleue au modèle paramétrique. L'écart type  $\sigma_{\text{eff}}$  défini comme étant la largeur centrée contenant 68 % des événements de la distribution, correspondant aussi à la résolution en masse, et la largeur à mi-hauteur FWHM sont aussi indiqués.

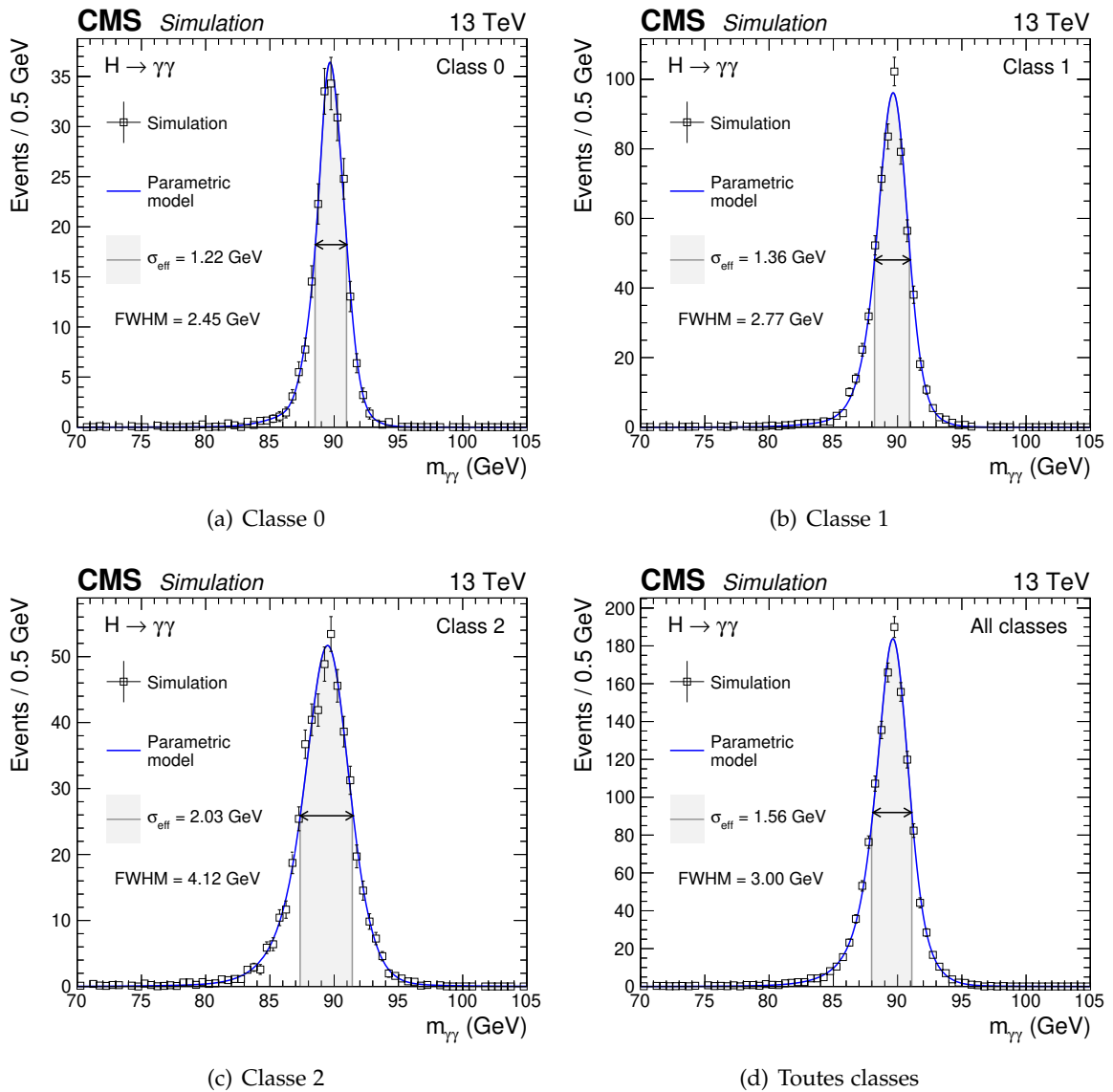


FIGURE 6.31 – Modèle paramétrique du signal 2018 pour une luminosité de  $54,4 \text{ fb}^{-1}$ , évalué à une masse de 90 GeV, par classe. Les contributions bonne et mauvaise identification de vertex sont sommées selon leur fraction respective, et les contributions des modes de production sont sommées selon leur section efficace. Les points blancs correspondent aux simulations et la courbe bleue au modèle paramétrique. L'écart type  $\sigma_{eff}$  défini comme étant la largeur contenant 68 % des événements de la distribution, correspondant aussi à la résolution en masse, et la largeur à mi-hauteur FWHM sont aussi indiqués.

On souhaite finalement connaître le pourcentage d'évènements signal conservé en prenant en compte les séries de sélection effectuées dans l'analyse, l'efficacité, et les caractères non parfaits du détecteur, l'acceptance, qui induisent en effet une perte d'évènements signal. Cette quantité joue, notamment un rôle important, dans l'interprétation des tests statistiques utilisés pour obtenir les résultats finaux de l'analyse, comme abordé dans la section 6.11. Le produit de l'efficacité par l'acceptance en fonction de la masse d'un boson de Higgs correspondant au signal est visible sur la figure 6.32; ce produit est légèrement plus faible pour 2017 dû au *prefiring*.

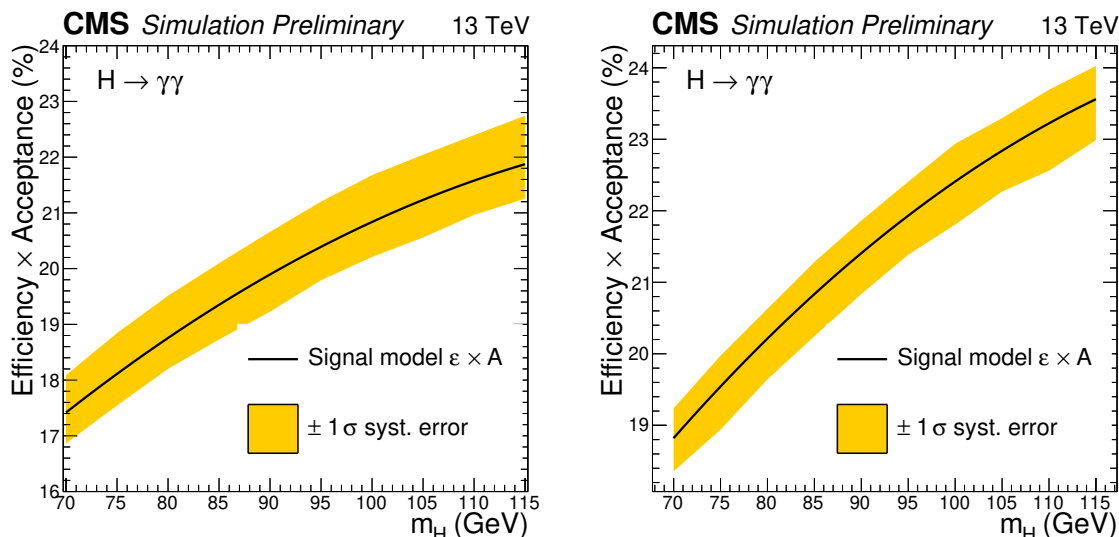


FIGURE 6.32 – Efficacité de l’analyse multiplié par l’acceptance du détecteur en fonction de la masse d’un boson de Higgs *SM-like* pour 2017 (gauche) et 2018 (droite). Les bandes jaunes correspondent à  $\pm 1\sigma$  de l’incertitude systématique sur le nombre d’évènements signal incorporant les incertitudes systématiques présentées dans cette section.

## 6.9 Modélisation du bruit de fond

On veut maintenant paramétrer le bruit de fond dans le même esprit, soit en décrivant sa masse invariante diphoton par un modèle analytique. Cette modélisation s’effectue en deux étapes. La première étape consiste à modéliser la partie Drell-Yan  $Z \rightarrow ee$  à partir des simulations. La seconde consiste à modéliser le continuum diphoton directement sur les données et en ajoutant la paramétrisation DY obtenue dans l’étape précédente. La méthode utilisée est légèrement différente de celle utilisée dans [103]. Les études présentées dans cette section sont essentiellement le fait de mon travail.

### 6.9.1 Modélisation de la composante Drell-Yan

La composante Drell-Yan du bruit de fond correspond, on le rappelle, à des événements  $Z \rightarrow ee$  où les électrons passent tous les critères de l’analyse présentés précédemment (toutes les coupures et le veto électron notamment) et sont donc reconstruits à tort comme des photons, on parle alors d’évènements *double-fake*. Cette composante est naturellement présente dans notre analyse au vu de la gamme de masse considérée, c’est aussi une des spécificités, voire difficultés, de l’analyse. Il est difficile de séparer la composante DY du continuum diphoton dans les données et la zone proche du pic du Z peut aussi être une zone de signal. On modélise donc la composante DY sur des simulations MC pour 2017 et 2018.

La fonction paramétrique utilisée pour la modélisation de la masse invariante diphoton est une fonction *double-sided Crystall Ball*, déjà rencontrée dans le chapitre 4, à laquelle on ajoute une fonction exponentielle. L’ajout de cette fonction est, entre autre, une nouveauté par rapport à [103] qui permet l’amélioration de l’ajustement réalisé. L’étude montrant les effets de

L'ajout de cette fonction sera abordé dans le chapitre suivant. La fonction dCB est une fonction gaussienne avec des queues en loi de puissances asymétriques, elle possède six paramètres :  $\mu$  et  $\sigma$  qui ont la même signification que pour une gaussienne,  $\alpha_L$  ( $\alpha_R$ ) et  $n_L$  ( $n_R$ ) décrivent eux le comportement de la queue gauche (droite),  $\alpha_{L,R}$  contrôlent le début des queues et  $n_{L,R}$  décrivent la puissance décroissante des queues. La fonction exponentielle est définie selon sa fraction relative *frac* par rapport à la fonction dCB, et son coefficient *p1*. La fonction totale est donnée, avec un facteur de normalisation *N*, par :

$$dCB(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma}} \left(\frac{n_L}{|\alpha_L|}\right)^{n_L} \exp\left(-\frac{|\alpha_L|^2}{2}\right) \left(\frac{n_L}{|\alpha_L|} - |\alpha_L| - \frac{x-\mu}{\sigma}\right)^{-n_L} & , \text{ si } \frac{x-\mu}{\sigma} < -\alpha_L \\ \frac{1}{\sqrt{2\pi\sigma}} \left(\frac{n_R}{|\alpha_R|}\right)^{n_R} \exp\left(-\frac{|\alpha_R|^2}{2}\right) \left(\frac{n_R}{|\alpha_R|} - |\alpha_R| + \frac{x-\mu}{\sigma}\right)^{-n_R} & , \text{ si } \frac{x-\mu}{\sigma} > \alpha_R \\ \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & , \text{ autrement} \end{cases} \quad (6.6)$$

$$N \times (dCB + exp)(x) = N \times ((1 - frac) \times dCB(x) + frac \times \exp(-p1 \times x)) \quad (6.7)$$

Cette fonction comprenant 8 paramètres libres, on fixe les valeurs de  $\alpha_{L/R}$  afin de faciliter l'ajustement. Pour déterminer ces valeurs de  $\alpha_{L/R}$ , on utilise des événements *single-fake* qui sont des événements contenant un électron passant un critère de veto électron inversé (donc un électron avec des coups dans le détecteur à pixel) et un électron, reconstruit comme un photon, passant le critère de veto électron usuel (tous les autres critères de sélection de l'analyse étant satisfaits). On souhaite ajuster les distributions en masse invariante diphoton de ces événements pour des données *single-fake* pour 2018. Pour 2017, le chemin HLT contient un pixel veto qui réduirait drastiquement la population *single-fake* des données, on utilise donc le même chemin prescalé que pour la validation  $Z \rightarrow ee$  des éléments *diphoton* BDT abordés dans la section 6.7.1. L'utilisation de ce chemin réduit cependant la statistique de cet échantillon par le facteur de prescaling ; pour choisir les valeurs de  $\alpha_{L/R}$  à utiliser pour les *double-fake* on préférera utiliser des simulations MC incorporant les échantillons bruit de fond DY,  $\gamma + \text{jets}$  et  $\gamma + \gamma$  (on omet QCD) avec la sélection *single-fake*.

Les ajustements de la distribution en masse invariante diphoton pour les événements *single-fake* (données pour 2018 et simulation bruit de fond total pour 2017) sont ensuite réalisés par des fonctions dCB+exp pour différentes valeurs de  $\alpha_{L/R}$  variant de 0,1 à 2 par pas de 0,1. On choisit les valeurs de  $\alpha_{L/R}$  minimisant la distribution  $\chi^2/ndf - 1$ . Ce processus est illustré sur la figure 6.33 pour 2018. Les valeurs de  $\alpha_L$  et  $\alpha_R$  retenues pour chaque classe sont visibles sur les figures 6.34 pour 2017 et 6.35 pour 2018.

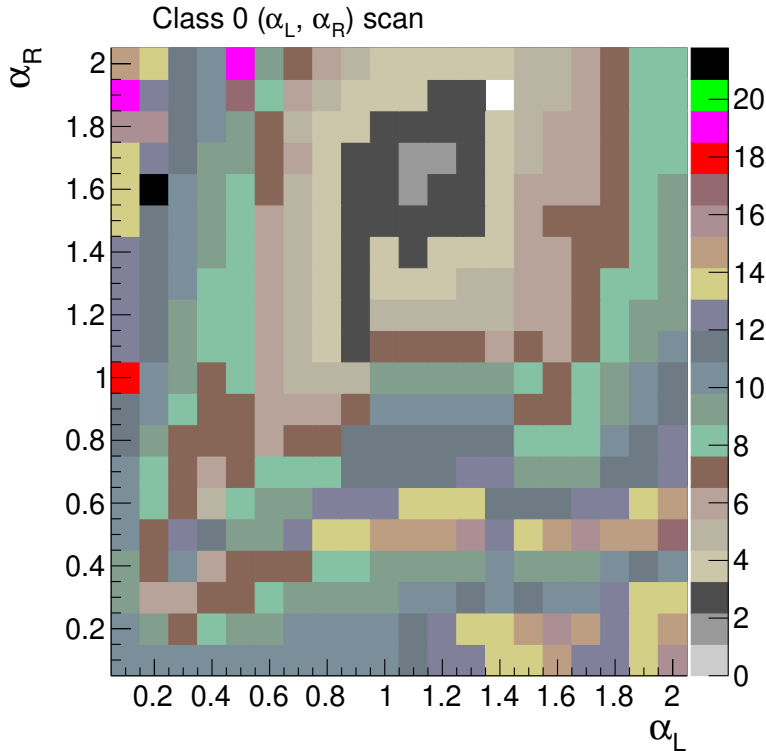


FIGURE 6.33 – Optimisation des valeurs  $\alpha_{L/R}$  obtenues à partir d’ajustement sur des données *single-fake* 2018 et pour la classe 0. La barre de couleur représente la valeur  $\chi^2/ndf$  obtenue pour chaque ajustement réalisé pour différentes valeurs de  $(\alpha_L, \alpha_R)$ .

Les paramètres de la fonction dCB+exp sont finalement obtenus par ajustement de la distribution de la masse invariante diphoton des événements simulations *double-fake* par classe, en utilisant les valeurs  $(\alpha_L, \alpha_R)$  déterminées précédemment. Les valeurs de ces paramètres sont visibles sur les figures 6.34 pour 2017 et 6.35 pour 2018. On remarque notamment qu’il y a environ deux fois plus d’évènements *double-fake* en 2018 par rapport à 2017 alors qu’on attend un facteur de 1,3 correspondant à l’augmentation de la luminosité. Cet excès d’évènements en 2018 est compris et est abordé en détail dans le chapitre 7 : il est dû à des inefficacités dans certaines zones du détecteur à pixels laissant passer de très bons candidats électrons qui sont ensuite identifiés comme des photons.

On injecte ensuite cette modélisation de la composante Drell-Yan au modèle total du bruit de fond. La modélisation de la composante DY étant effectuée sur les simulations MC, il est nécessaire de considérer des incertitudes systématiques prenant en compte les différences entre la composante Drell-Yan *double-fake* obtenue sur les simulations et sa correspondance dans les données. Ces incertitudes seront discutées dans la suite.

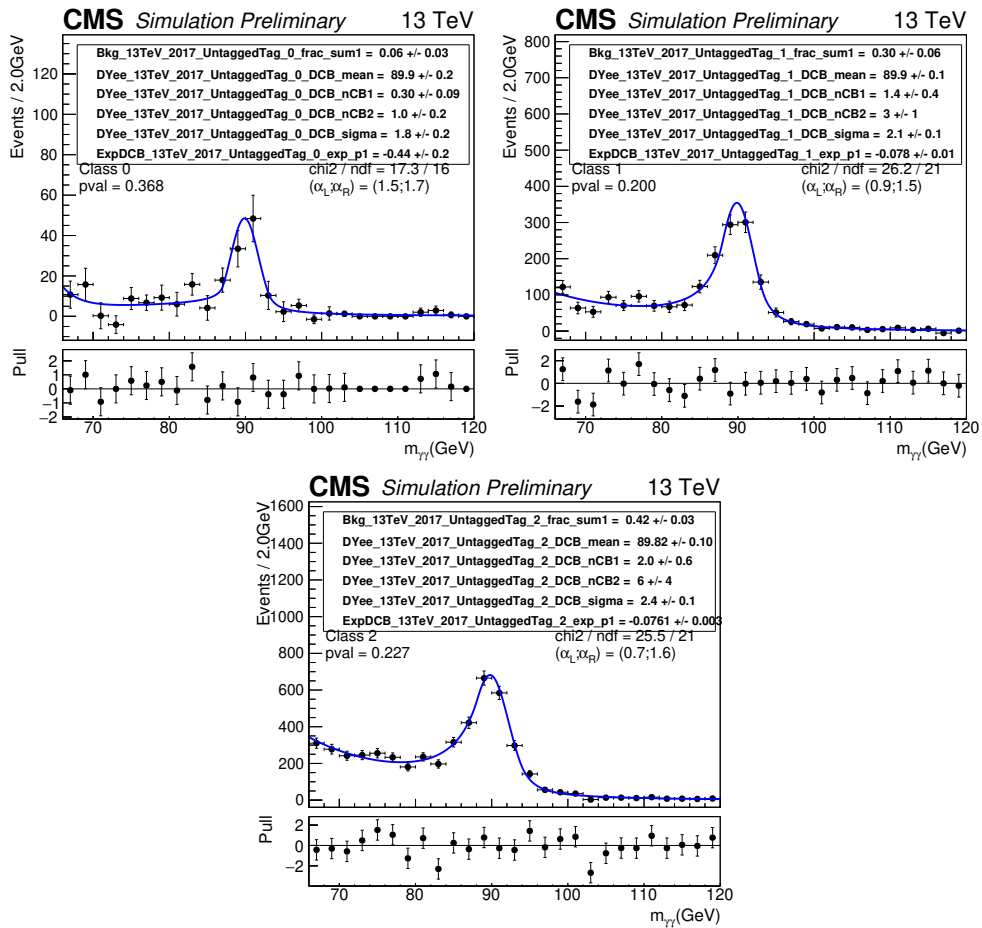


FIGURE 6.34 – Ajustement de la distribution de masse invariante diphoton par une fonction dCB+exp pour des simulations d'événements Drell-Yan *double-fake* par classe et pour 2017 pour une luminosité de  $41,5 \text{ fb}^{-1}$ . Les points noirs correspondant aux simulations, l'ajustement est visible en bleu, en bas la différence entre simulations et ajustement normalisée à l'erreur statistique. Le paramètre  $nCB1(2)$  correspond à  $n_{L/R}$ . A noter que le bin à 73 GeV de la classe 0 a un contenu négatif.

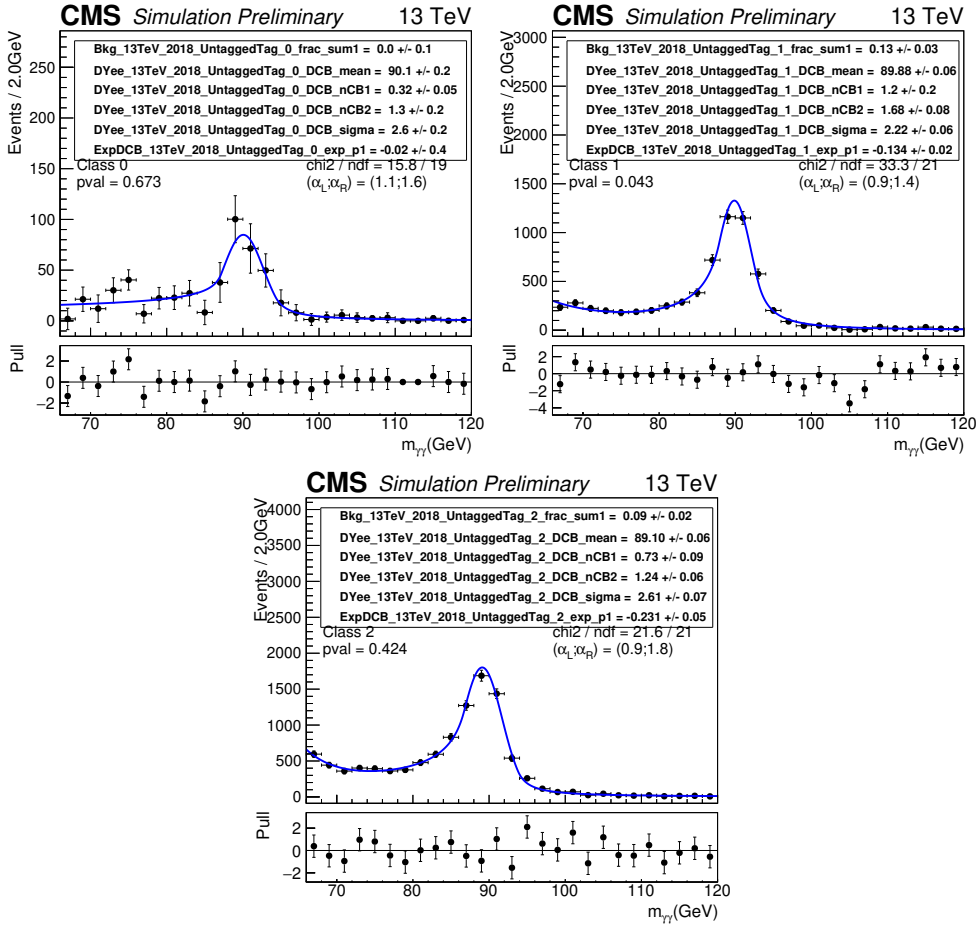


FIGURE 6.35 – Ajustement de la distribution de masse invariante diphoton par une fonction dCB+exp pour des simulations d'évènements Drell-Yan *double-fake* par classe et pour 2018 pour une luminosité de  $54,4 \text{ fb}^{-1}$ . Les points noirs correspondant aux simulations, l'ajustement est visible en bleu, en bas la différence entre simulations et ajustement normalisée à l'erreur statistique. Le paramètre  $nCB1(2)$  correspond à  $n_{L/R}$ .

## 6.9.2 Modèle de bruit de fond total

Pour obtenir le modèle final du bruit de fond, on cherche à décrire le continuum diphoton directement sur les données par une fonction analytique à laquelle on ajouterait la fonction dCB+exp obtenue précédemment. Dans les faits, on n'utilise pas une mais plusieurs fonctions analytiques par la méthode de profilage discret [111] ou méthode de l'enveloppe. L'idée est de sélectionner un ensemble de fonctions parmi des sommes de fonctions exponentielles, des lois de puissances, des polynômes de Laurent et des polynômes de Bernstein pour former une enveloppe décrivant la distribution en masse du bruit de fond. Le choix final de fonction utilisée est alors susceptible de varier pour prendre en compte le fait qu'on ne connaît pas a priori la meilleure fonction ajustant la distribution du bruit de fond, permettant donc d'associer une incertitude systématique au choix de la fonction d'ajustement. Comme l'ajustement est réalisé directement sur les données il est important que le pic d'un éventuel signal recherché soit étroit afin qu'il ne perturbe pas l'ajustement, soulignant l'importance d'avoir une très bonne résolution en masse. L'ajustement est réalisé par classe.



Les fonctions considérées pour l'ajustement sont présentées dans l'Eq. 6.8 avec  $N$  un facteur de normalisation globale,  $M$  l'ordre de la somme considérée,  $\beta_i$  des paramètres libres et  $f$  les fractions associées aux différentes familles de fonction. On rappelle que la normalisation de la composante Drell-Yan est laissée flottante et que ses paramètres sont ici fixés à leur valeur centrale (cette hypothèse est discutée dans le chapitre 7).

Somme d'exponentielles :

$$N(f_E \text{Exp}_M(x) + (1 - f_E)(dCB + \text{exp})(x)) = N \times \left( f_E \sum_{i=0}^M \beta_{2i} e^{\beta_{2i+1} x} + (1 - f_E) \times (dCB + \text{exp})(x) \right)$$

Somme de lois de puissance :

$$N(f_P \text{Pow}_M(x) + (1 - f_P)(dCB + \text{exp})(x)) = N \times \left( f_P \sum_{i=0}^M \beta_{2i} x^{\beta_{2i+1}} + (1 - f_P) \times (dCB + \text{exp})(x) \right)$$

Polynômes de Bernstein :

$$N(f_B \text{Ber}_M(x) + (1 - f_B)(dCB + \text{exp})(x)) = N \times \left( f_B \sum_{i=0}^M \beta_i b_{i,M} + (1 - f_B) \times (dCB + \text{exp})(x) \right)$$

avec  $b_{i,M} = \binom{M}{i} x^i (1-x)^{M-i}$  avec  $\binom{M}{i} = \frac{m!}{i!(M-i)!}$

Polynômes de Laurent :

$$N(f_L \text{Lau}_M(x) + (1 - f_L)(dCB + \text{exp})(x)) = N \times \left( f_L \sum_{i=0}^M \beta_i x^{-4 + \sum_{j=0}^i (-1)^j} + (1 - f_L) \times (dCB + \text{exp})(x) \right) \quad (6.8)$$

La sélection des fonctions au sein de l'enveloppe s'effectue en deux étapes. La première étape consiste à déterminer les fonctions "vérités" dont le nombre  $(m+1)^9$  de paramètres libres passe les critères d'un F-test tel que décrit dans la section précédente. On admet alors dans l'enveloppe les fonctions vérités vérifiant les conditions sur la p-value associée à ce F-test qui sont :

$$\text{p-value}(\chi^2 > \chi_{M \rightarrow M+1}^2) < 0,05 \quad \text{et} \quad m < 7 \quad (6.9)$$

On admet aussi dans l'enveloppe, en plus de ces fonctions vérités, les fonctions dont l'ajustement présente une figure de mérite acceptable, définie avec une probabilité d'obtenir un  $\chi^2$  supérieur au  $\chi_M^2$  mesuré qui doit être supérieur à 1 % (voir la section 5.3.1 pour plus de détails). Pour éviter un ajustement trop fin qui provoquerait des distorsions dans nos résultats finaux,  $m+1$  est limité à 4 [103]. Pour entrer dans l'enveloppe les fonctions doivent alors satisfaire :

$$\text{p-value}(\chi^2 > \chi_{M \rightarrow M+1}^2) < 0,10 \quad \text{et} \quad (\text{p-value}(\chi^2 > \chi_M^2) > 0,01 \quad \text{ou} \quad M = M_{\text{verite}}) \quad \text{et} \quad m < 4 \quad (6.10)$$

Finalement, on identifie au sein de l'enveloppe la fonction présentant la meilleure qualité d'ajustement comme la fonction minimisant la quantité  $LL_M + 0,5N_p$  où  $LL_M$  est la fonction

9. Les familles de fonction n'ont pas le même nombre de paramètres libres selon l'indice de sommation;  $m$  peut donc être différent de  $M$ .

log-vraisemblance négative associée à l'ajustement et  $N_p$  est une pénalité égale au nombre de paramètres de la fonction considérée.

La composition des enveloppes ainsi que les meilleures fonctions identifiées sont visibles par classe sur les figures 6.36 pour 2017 et 6.37 pour 2018. En adéquation avec la politique d'analyse à l'aveugle, seules les régions non-sensibles correspondant à des gammes de masses entre  $65 < m_{\gamma\gamma} < 70$  GeV,  $87 < m_{\gamma\gamma} < 93$  GeV et  $115 < m_{\gamma\gamma} < 120$  GeV sont visibles, bien que toute la gamme entre 65 et 120 GeV soit utilisée pour l'ajustement. Comme indiqué, le chiffre donné dans le nom de la fonction correspond au nombre  $m + 1$  de paramètres libres correspondant : ainsi *exp\_3* correspond à la somme obtenue avec  $Exp_1 = \beta_0 e^{\beta_1} + \beta_2 e^{\beta_3}$  et *lau\_2* correspond à la somme obtenue avec  $Lau_2 = \beta_0 x^{-4} + \beta_1 x^{-5} + \beta_2 x^{-3}$ . La fraction dCB+exp varie entre 0,3 % et 4 % selon les fonctions et les classes. On observe de plus que l'excès d'évènements Drell-Yan évoqué précédemment est bien présent dans les données.

Dans l'analyse statistique présentée dans les sections suivantes, les paramètres de la fonction dCB+exp seront libres de flotter dans un intervalle de  $\pm 4\sigma_{tot}$ , correspondant à leur incertitude totale, autour de leur valeur centrale et la fraction dCB+exp associée sera libre de varier. Les paramètres des fonctions de l'enveloppe seront aussi libres de complètement flotter et la meilleure fonction d'ajustement sera libre de changer.

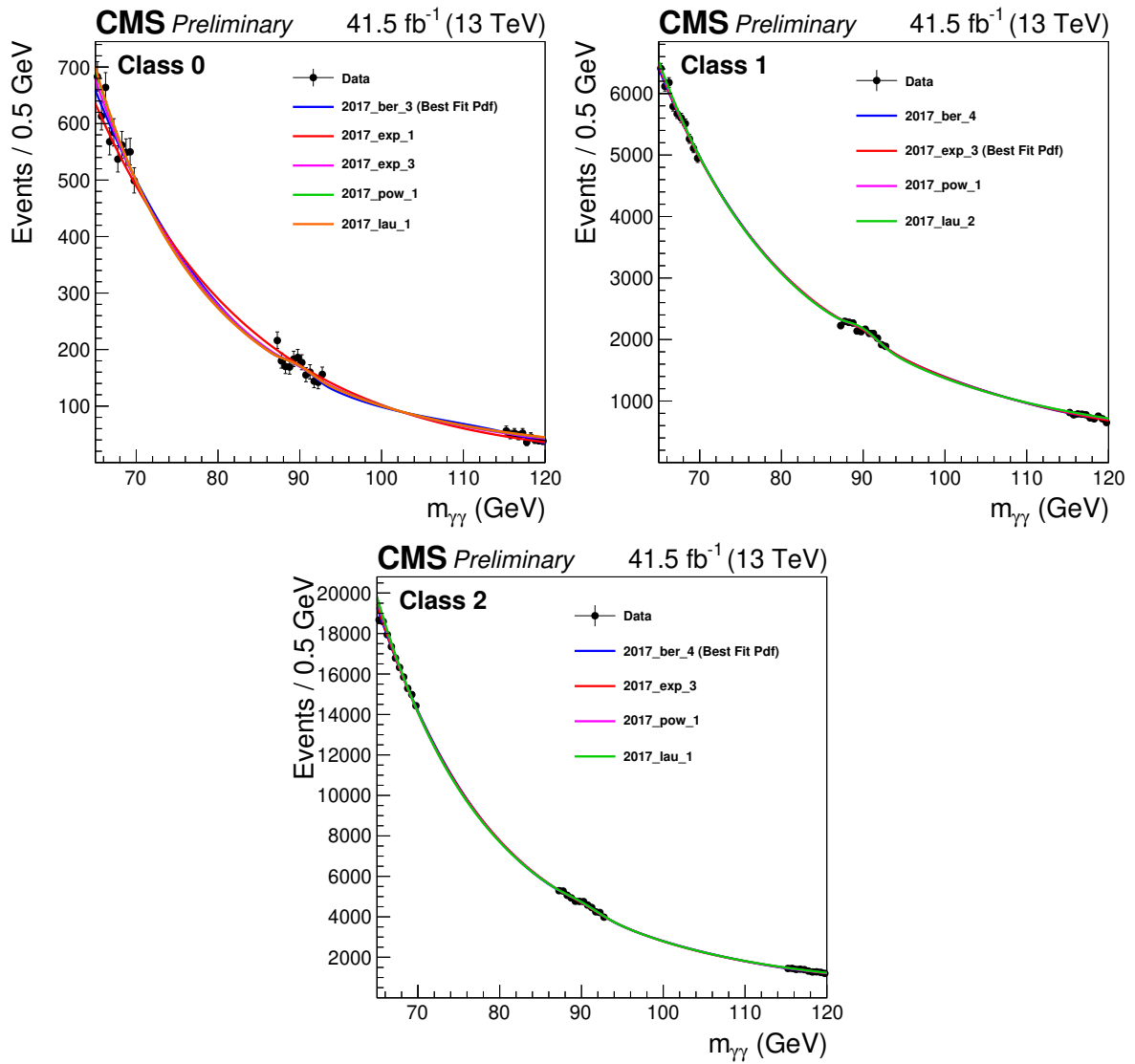


FIGURE 6.36 – Ajustement de la masse invariante diphoton des données (points noirs) par les fonctions composant l'enveloppe pour 2017. Le chiffre donné dans le nom de la fonction correspond au nombre  $m + 1$  de paramètres libres de cette fonction. La fonction de meilleur ajustement est nommée "Best Fit Pdf".

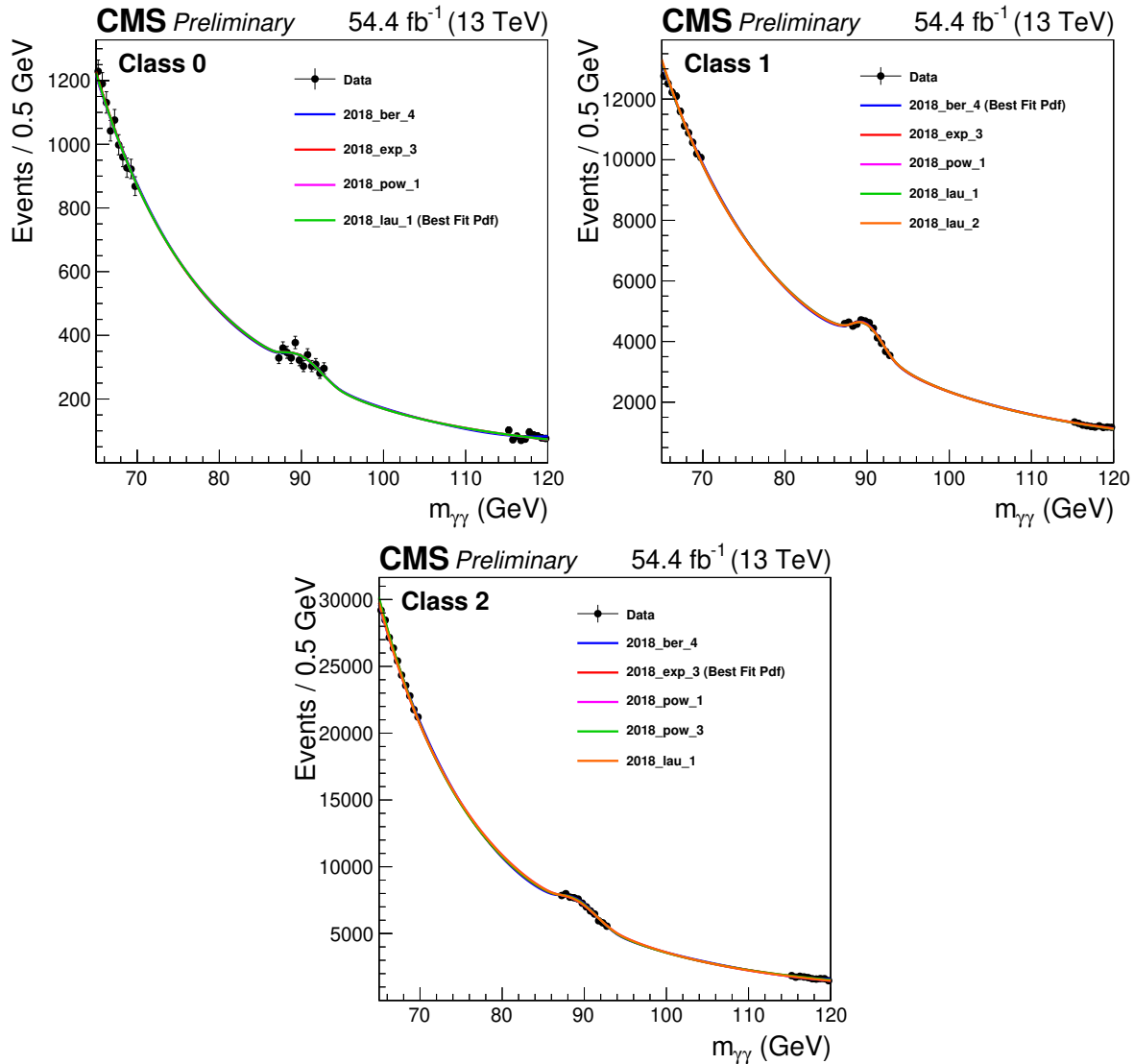


FIGURE 6.37 – Ajustement de la masse invariante diphoton des données (points noirs) par les fonctions composant l’enveloppe pour 2018. Le chiffre donné dans le nom de la fonction correspond au nombre  $m + 1$  de paramètres libres de cette fonction. La fonction de meilleur ajustement est nommée "Best Fit Pdf".

### 6.9.3 Incertitudes systématiques associées

Comme mentionné précédemment, une incertitude systématique est associée à la méthode de l’enveloppe (profilage discret) prenant en compte le fait que chaque fonction de l’enveloppe peut ajuster la distribution de masse invariante diphoton. Il faut cependant aussi prendre en compte des incertitudes systématiques pour la modélisation directe de la composante DY. Cette dernière étant, en effet, ajustée sur les simulations MC, il faut pouvoir estimer les différences de paramètres pour des ajustements dCB+exp entre simulations et données. Pour ce faire on utilise les échantillons *single-fake*, présentés précédemment, qui contiennent des événements avec deux candidats photon dont l’un des deux est un électron passant le veto électron et l’autre ne passant pas ce veto.

L'idée ici est de calculer les différences de forme de la fonction dCB+exp entre les données et les simulations *single-fake* avant de propager ces différences pour les *double-fake* suivant [103]. On ajuste donc les distributions en masse invariante diphoton des évènements *single-fake* pour les données (data) et les simulations. Pour les simulations on considère, avec la sélection *single-fake*, le cas (MC<sub>DY</sub>) où seul l'échantillon Drell-Yan est utilisé, comme c'est le cas pour l'ajustement *double-fake*, mais aussi le cas (MC<sub>All</sub>) où tous les échantillons bruit de fond (sauf QCD) de l'analyse sont utilisés ; ce dernier cas est en effet plus représentatif de la distribution en masse des données (de la même manière que l'on utilise l'ensemble des échantillons bruit de fond pour comparer la distribution en masse des données). On cherche alors à comparer la distribution en masse des évènements *single-fake* entre data et MC<sub>All</sub>, et aussi, pour mieux rendre compte des différences données/simulations de la composante DY *double-fake*, entre MC<sub>All</sub> et MC<sub>DY</sub>. Les ajustements par classe dans les cas données, DY et bruit de fond total sans QCD sont visibles sur les figures 6.38 à 6.43 pour 2017 et 2018 respectivement. Les valeurs des paramètres  $\alpha$  sont les mêmes que celles utilisées lors de l'ajustement des *double-fake*.

Les paramètres de la fonction dCB n'étant pas indépendants entre eux, on fait l'hypothèse qu'il est suffisant<sup>10</sup> de déterminer uniquement une incertitude systématique pour les paramètres  $\mu$  et  $\sigma$  de la dCB. Les valeurs de ces paramètres après ajustements des évènements *single-fake* sont lisibles dans les tableaux 6.44 pour 2017 et 6.45 pour 2018. On calcule les différences data/MC<sub>All</sub> et MC<sub>All</sub>/MC<sub>DY</sub> entre les différentes fonctions dCB+exp obtenues par ajustement des évènements *single-fake* si ces différences sont statistiquement significatives. On multiplie ensuite ces différences par un facteur 2 pour prendre en compte le fait qu'il y ait deux électrons au sein des évènements *double-fake*, et on les ajoute en quadrature à l'incertitude statistique des ajustements dCB+exp *double-fake* afin d'obtenir les incertitudes totales  $\Delta\mu_{tot}$  et  $\Delta\sigma_{tot}$  sur les paramètres  $\mu$  et  $\sigma$  de la fonction dCB+exp *double-fake*.

Pour résumer les incertitudes totales sur les paramètres  $\mu$  et  $\sigma$  de la fonction dCB+exp *double-fake* sont données par :

$$\Delta\mu_{tot} = \sqrt{(\Delta\mu_{stat})^2 + (2\Delta\mu_{data,MC_{All}})^2 + (2\Delta\mu_{MC_{All},MC_{DY}})^2} \quad (6.11)$$

$$\Delta\sigma_{tot} = \sqrt{(\Delta\sigma_{stat})^2 + (2\Delta\sigma_{data,MC_{All}})^2 + (2\Delta\sigma_{MC_{All},MC_{DY}})^2} \quad (6.12)$$

où  $\Delta\mu_{stat}$  est l'incertitude statistique des paramètres pour l'ajustement DY *double-fake*, utilisé dans la modélisation du bruit de fond, et  $\Delta\mu_{X,Y}$  est la différence pour les ajustements *single-fake* data/MC<sub>All</sub> et MC<sub>All</sub>/MC<sub>DY</sub> donnée par :

$$\Delta\mu_{data,MC_{All}} = |\mu_{data} - \mu_{MC_{All}}| \quad \text{si } > \sqrt{(\Delta\mu_{data}^{stat})^2 + \Delta\mu_{MC_{All}}^{stat}} \quad \text{sinon non applicable (N.A.)} \quad (6.13)$$

$$\Delta\mu_{MC_{All},MC_{DY}} = |\mu_{MC_{All}} - \mu_{MC_{DY}}| \quad \text{si } > \sqrt{(\Delta\mu_{MC_{All}}^{stat})^2 + \Delta\mu_{MC_{DY}}^{stat}} \quad \text{sinon non applicable (N.A.)} \quad (6.14)$$

Et de même pour  $\sigma$ .

10. Cette hypothèse sera, elle aussi, discutée dans le chapitre suivant.

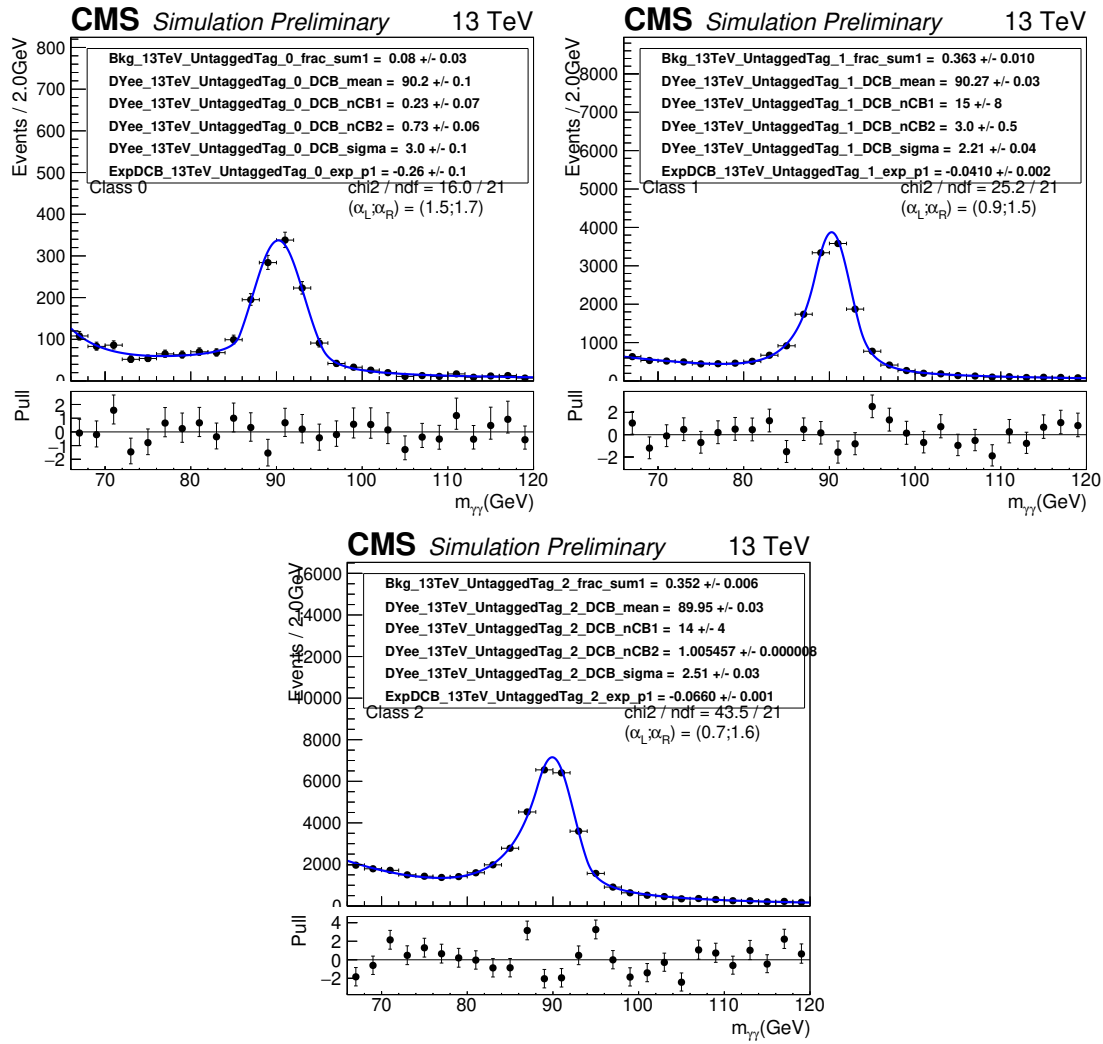


FIGURE 6.38 – Ajustement par classe de la masse invariante diphoton des évènements *single-fake* par la fonction dCB+exp pour les données (data) de 2017.

Les valeurs de ces incertitudes, ainsi que celles de ses différents termes, sont données dans les tableaux 6.46 pour 2017 et 6.47; ces valeurs peuvent aller jusqu'à 2,5 fois la valeur des incertitudes statistiques associées.

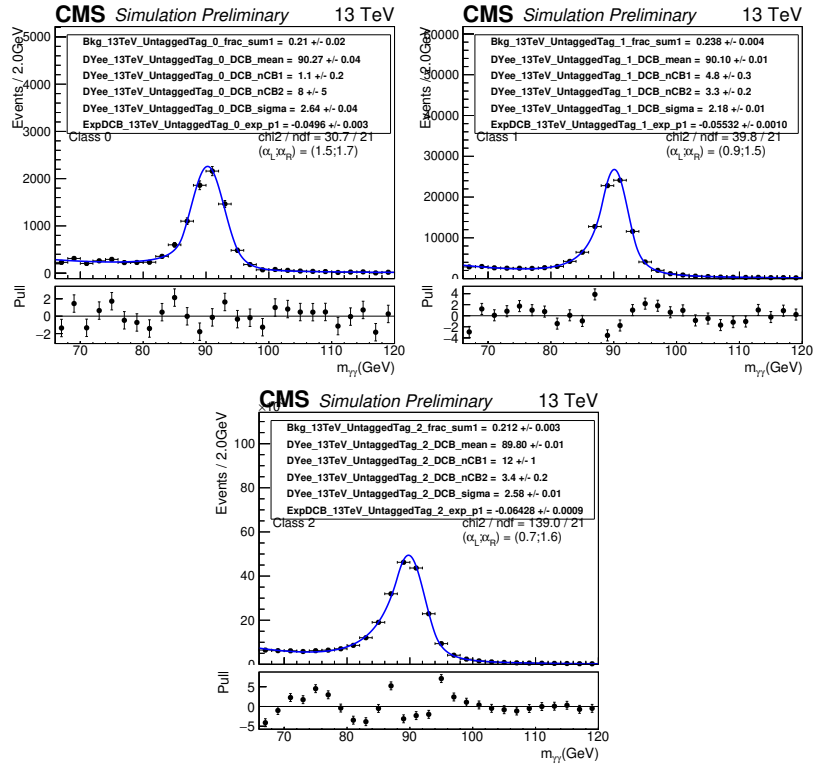


FIGURE 6.39 – Ajustement par classe de la masse invariante diphoton des évènements *single-fake* par la fonction dCB+exp pour les simulation DY ( $MC_{DY}$ ) de 2017.

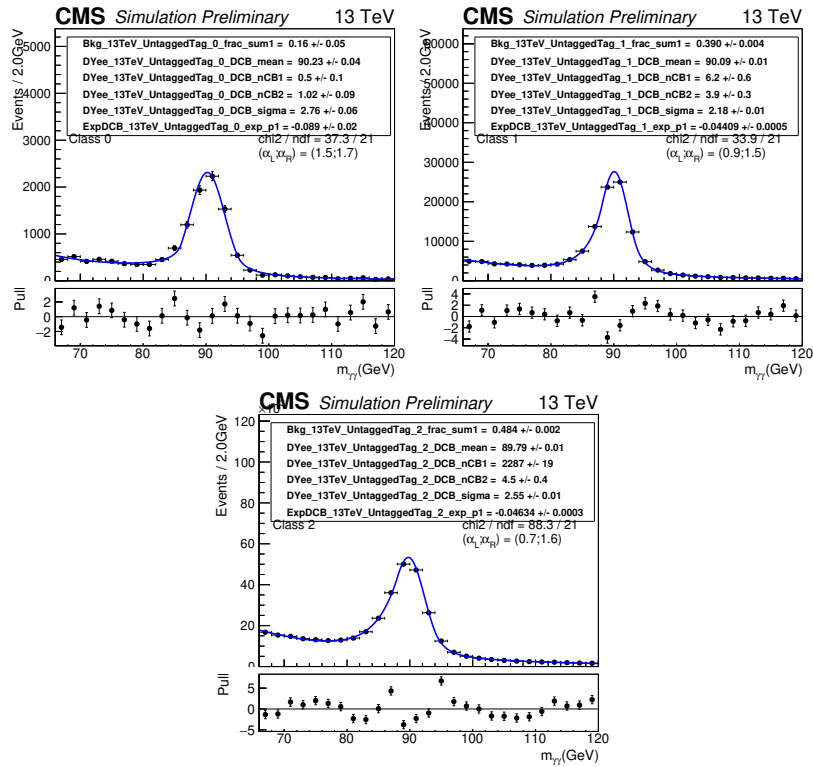
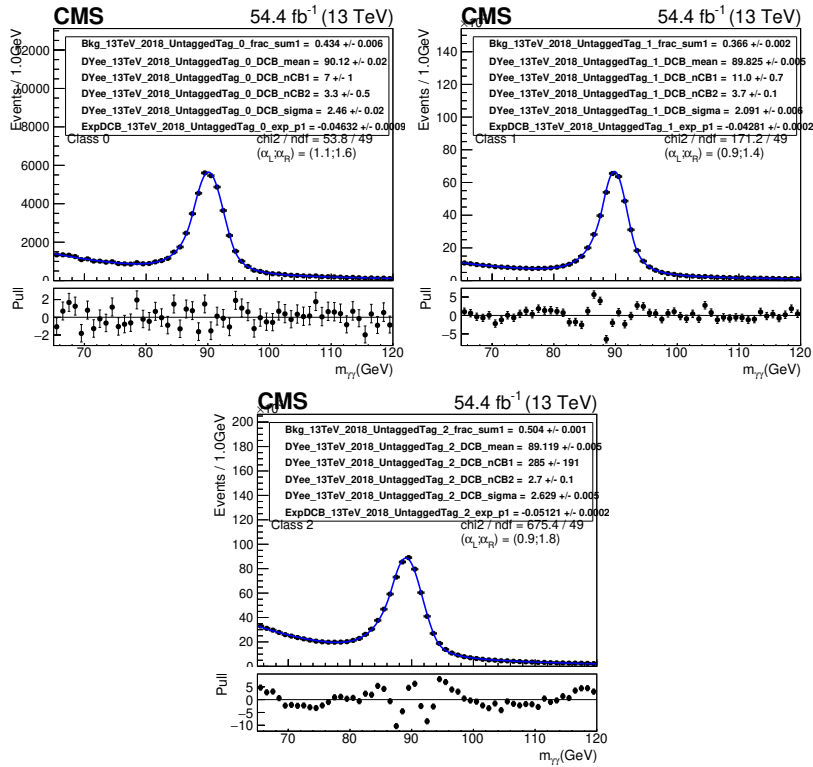
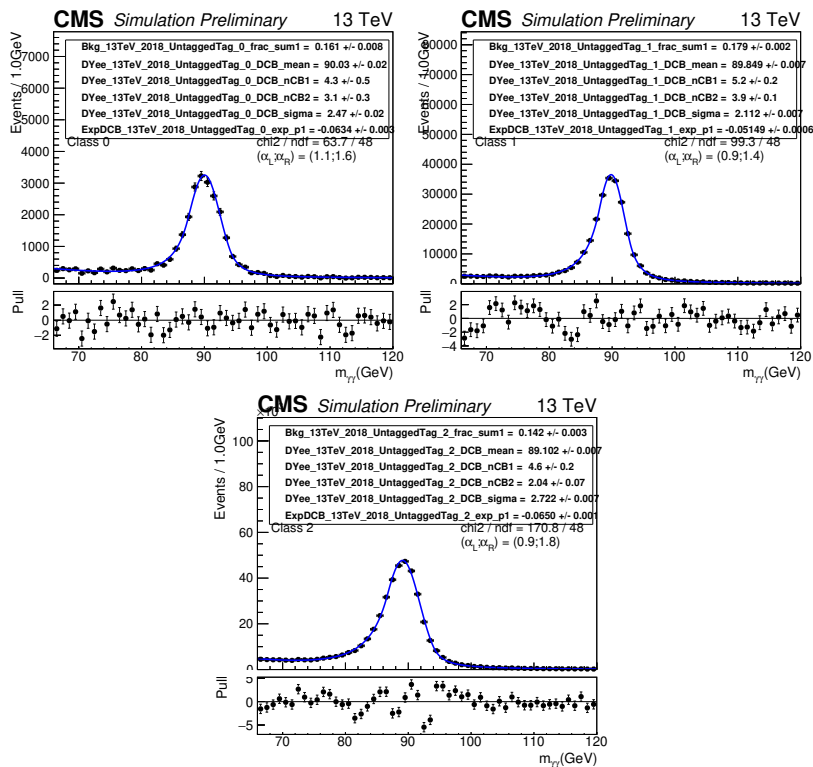


FIGURE 6.40 – Ajustement par classe de la masse invariante diphoton des évènements *single-fake* par la fonction dCB+exp pour les simulation DY,  $\gamma + \gamma$  et  $\gamma + \text{jets}$  ( $MC_{All}$ ) de 2017.


 FIGURE 6.41 – Ajustement par classe de la masse invariante diphoton des évènements *single-fake* par la fonction dCB+exp pour les données (data) de 2018.

 FIGURE 6.42 – Ajustement par classe de la masse invariante diphoton des évènements *single-fake* par la simulation DY (MC<sub>DY</sub>) de 2018.



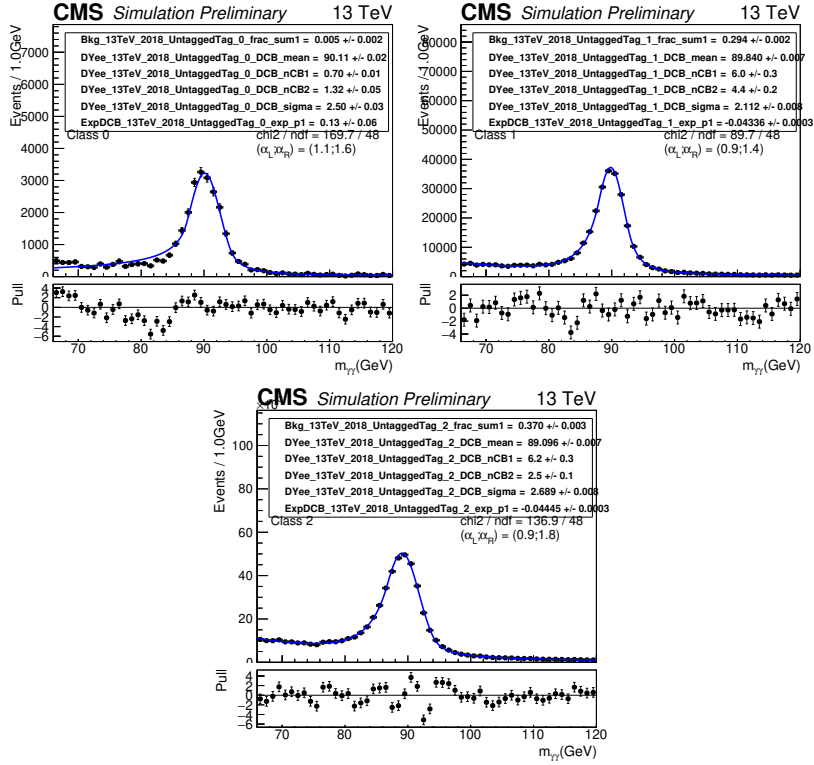


FIGURE 6.43 – Ajustement par classe de la masse invariante diphoton des évènements *single-fake* par la fonction dCB+exp pour les simulation DY,  $\gamma + \gamma$  et  $\gamma + \text{jets}$  ( $MC_{All}$ ) de 2018.

FIGURE 6.44 – Valeurs des paramètres dCB  $\mu$  et  $\sigma$  obtenues après ajustement en masse des évènements *single-fake* pour 2017.

class	$\mu_{data}$	$\mu_{MC_{All}}$	$\mu_{MC_{DY}}$	$\sigma_{data}$	$\sigma_{MC_{All}}$	$\sigma_{MC_{DY}}$
0	$90.20 \pm 0.10$	$90.23 \pm 0.03$	$90.27 \pm 0.04$	$3.0 \pm 0.1$	$2.76 \pm 0.06$	$2.64 \pm 0.04$
1	$90.27 \pm 0.03$	$90.09 \pm 0.01$	$90.10 \pm 0.01$	$2.21 \pm 0.04$	$2.18 \pm 0.01$	$2.18 \pm 0.01$
2	$89.95 \pm 0.03$	$89.79 \pm 0.01$	$89.80 \pm 0.01$	$2.51 \pm 0.03$	$2.55 \pm 0.01$	$2.58 \pm 0.01$

FIGURE 6.45 – Valeurs des paramètres dCB  $\mu$  et  $\sigma$  obtenues après ajustement en masse des évènements *single-fake* pour 2018.

class	$\mu_{data}$	$\mu_{MC_{All}}$	$\mu_{MC_{DY}}$	$\sigma_{data}$	$\sigma_{MC_{All}}$	$\sigma_{MC_{DY}}$
0	$90.12 \pm 0.02$	$90.11 \pm 0.02$	$90.03 \pm 0.02$	$2.46 \pm 0.02$	$2.50 \pm 0.03$	$2.47 \pm 0.02$
1	$89.82 \pm 0.01$	$89.84 \pm 0.01$	$89.85 \pm 0.01$	$2.09 \pm 0.01$	$2.11 \pm 0.01$	$2.11 \pm 0.01$
2	$89.12 \pm 0.01$	$89.10 \pm 0.01$	$89.10 \pm 0.01$	$2.63 \pm 0.01$	$2.69 \pm 0.01$	$2.72 \pm 0.01$

FIGURE 6.46 – Incertitudes totales par classe sur les paramètres dCB  $\mu$  et  $\sigma$  utilisés dans l’ajustement de la masse des évènements *double-fake* décomposées en leurs différentes composantes pour 2017. "N.A." signifie que la différence dans les ajustements *single-fake* n’est pas statistiquement significative.

class	$\mu_{DY}^{2F}$	$\Delta\mu_{stat}^{2F}$	$\Delta\mu_{data,MC_{All}}$	$\Delta\mu_{MC_{All},MC_{DY}}$	$\Delta\mu_{tot}$	$\sigma_{DY}^{2F}$	$\Delta\sigma_{stat}^{2F}$	$\Delta\sigma_{data,MC_{All}}$	$\Delta\sigma_{MC_{All},MC_{DY}}$	$\Delta\sigma_{tot}$
0	89.90	0.24	N.A.	N.A.	0.24	1.80	0.23	0.24	0.12	0.58
1	89.90	0.13	0.18	N.A.	0.38	2.10	0.14	N.A.	N.A.	0.14
2	89.82	0.10	0.16	N.A.	0.33	2.40	0.10	0.04	0.03	0.14

FIGURE 6.47 – Incertitudes totales par classe sur les paramètres dCB  $\mu$  et  $\sigma$  utilisés dans l’ajustement de la masse des évènements *double-fake* décomposées en leurs différentes composantes pour 2018. "N.A." signifie que la différence dans les ajustements *single-fake* n’est pas statistiquement significative.

class	$\mu_{DY}^{2F}$	$\Delta\mu_{stat}^{2F}$	$\Delta\mu_{data,MC_{All}}$	$\Delta\mu_{MC_{All},MC_{DY}}$	$\Delta\mu_{tot}$	$\sigma_{DY}^{2F}$	$\Delta\sigma_{stat}^{2F}$	$\Delta\sigma_{data,MC_{All}}$	$\Delta\sigma_{MC_{All},MC_{DY}}$	$\Delta\sigma_{tot}$
0	90.1	0.19	N.A.	0.08	0.25	2.60	0.19	0.04	N.A.	0.21
1	89.88	0.06	0.02	N.A.	0.07	2.22	0.06	0.02	N.A.	0.07
2	89.10	0.06	0.02	N.A.	0.07	2.61	0.07	0.06	0.03	0.15

Lors de l’analyse statistique permettant de dériver les résultats finaux, les paramètres de la fonction dCB seront alors libres<sup>11</sup> de flotter autour de leur valeur centrale dans un intervalle de  $4\pm$  fois la valeur de leur incertitude totale mais avec une normalisation totalement flottante.

## 6.10 Incertitudes systématiques

### 6.10.1 Listes des incertitudes systématiques pour le signal

Les incertitudes systématiques associées à la modélisation de bruit de fond ont été décrites dans la section précédente. On récapitule ici toutes les incertitudes systématiques, dont certaines ont déjà été explicitées à travers ce chapitre, affectant notre modèle de signal. Ces incertitudes peuvent être des incertitudes modifiant directement le nombre d’évènements signal attendu, elle correspondent alors souvent à des incertitudes théoriques modifiant l’acceptance du signal ou venant des différents BDT utilisés pour catégoriser les évènements, ou des incertitudes affectant directement la forme de la distribution de masse invariante diphoton, typiquement des incertitudes expérimentales liées à l’énergie des photons individuels. Ces incertitudes sont calculées séparément pour chaque classe et mode de production. La méthode et les incertitudes considérées ici sont similaires (hors bruit de fond) à celles utilisées dans [90]. Les études présentées dans cette section sont essentiellement le fait de mon travail.

11. Les déviations par rapport à la valeur centrale des ces paramètres sont pénalisées dans la méthode de vraisemblance profilée.

### Incertitudes théoriques

Les incertitudes théoriques sont des incertitudes venant des calculs et des simplifications effectués pour décrire les sections efficaces des différents mode de production du signal utilisés ainsi que de leur propriétés cinématiques. Elles consistent en :

- **Incertitudes d'échelle de renormalisation et de factorisation** : variations de l'échelle de renormalisation et de factorisation lors du calcul des sections efficaces du MS attendues pour prendre en compte des corrections de termes d'ordre supérieurs calculées d'après les recommandations de [25]. Ces incertitudes sont dénommées *CMS\_hgg\_scaleWeight*.
- **Incertitudes sur les fonctions de distribution des partons** : prenant en compte notre connaissance imparfaite de la composition des protons et calculées suivant les recommandations PDF4LHC\_100 émises dans [112, 113, 114, 115]. L'impact des ces incertitudes est quasiment négligeable ici. Ces incertitudes sont dénommées *CMS\_hgg\_pdfWeight*.
- **Incertitudes sur la constante de couplage de l'interaction forte** : prenant en compte les modifications des propriétés cinématiques induites par cette variation. Ces incertitudes sont dénommées *pdf\_alphaS* et *QCDscale*.
- **Incertitudes sur le rapport de branchement  $H \rightarrow \gamma\gamma$**  : portant sur le calcul du rapport de branchement, incluses dans les calculs où ce rapport intervient, estimées à 2 % [25].
- **Incertitudes sur les évènements sous-jacents et la gerbe des partons** : pour prendre en compte ces phénomènes, intervenant notamment dans des évènements avec des jets, dans les simulations de signal, calculées avec des échantillons adaptées (on utilise les mêmes échantillons que dans [90]). Ces incertitudes sont dénommées *UnderlyingEvent\_norm* et *PartonShower\_norm*.

### Incertitudes expérimentales

Les incertitudes expérimentales prennent en compte les incertitudes correspondant aux effets de détecteur imparfaitement reproduits dans les simulations et celles provenant des différentes méthodes et techniques utilisées dans l'analyse. On commence par expliciter les incertitudes modifiant la forme du spectre en  $m_{\gamma\gamma}$  :

- **Incertitudes sur l'échelle d'énergie et résolution des photons** : associées aux corrections appliquées dans les données et la dégradation de la résolution dans les simulations en utilisant des évènements  $Z \rightarrow ee$ . Ces incertitudes sont dénommées *CMS\_hgg\_nuisance\_HighR9EE\_13TeVscale* et *CMS\_hgg\_nuisance\_HighR9EE\_13TeVsmear* selon la limite du  $R_9$  (haut ou bas) considéré et la zone du ECAL (EB ou EE).
- **Incertitudes sur la non-linéarité de l'échelle d'énergie des photons** : sources supplémentaires prenant en compte les différences entre données et simulations sur le facteur non-linéaire de l'échelle d'énergie des photons. Ces incertitudes sont dénommées *CMS\_hgg\_nuisance\_NonLinearity\_13TeVscale*.
- **Incertitudes sur les corrections de forme de gerbe électromagnétique** : pour prendre en compte la modélisation imparfaite de la forme de gerbe dans les simulations. Ces incertitudes sont dénommées *CMS\_hgg\_nuisance\_ShowerShapeHighR9EE\_scale* selon la

limite du  $R_9$  (haut ou bas) considéré et la zone du ECAL (EB ou EE) et *CMS\_hgg\_nuisance\_Geant4\_13TeVscale*.

- **Incertitudes sur la non-uniformité de la lumière collectée** : associées à la modélisation de la lumière collectée dans un cristal ECAL en fonction du type d'émission [74]. Ces incertitudes sont dénommées *CMS\_hgg\_nuisance\_FNUFEE\_scale* selon la région du ECAL considérée (EB ou EE).
- **Incertitudes sur la modélisation du budget matériel du ECAL** : prenant en compte les imperfections dans les simulations de la modélisation du matériel en amont du ECAL. Ces incertitudes sont dénommées *CMS\_hgg\_nuisance\_MaterialForward\_scale* selon la région considérée.
- **Incertitudes sur l'assignement de vertex** : variations de la fraction d'évènements où le vertex identifié est situé à moins de 1 cm du vrai vertex, calculées en partie avec des évènements  $Z \rightarrow \mu\mu$ . Ces incertitudes sont dénommées *CMS\_hgg\_nuisance\_deltafracright*.

Les incertitudes expérimentales modifiant directement le nombre d'évènements signal par classe sont :

- **Incertitudes sur la luminosité intégrée** : mesurées comme étant de 2,3 % [116] et de 2,5 % [117] pour 2017 et 2018 respectivement. Ces incertitudes sont en partie corrélées. Ces incertitudes sont dénommées *lumi\_13TeV*.
- **Incertitudes sur les SF de la présélection** : prenant en compte les incertitudes associées à la détermination des facteurs d'échelle de la présélection appliqués aux simulations MC pour les rendre comparables aux données. On dénomme plusieurs types d'incertitude selon les SF considérés *CMS\_hgg\_PreselSF*, *CMS\_hgg\_LooseMvaSF* et *CMS\_hgg\_electronVetoSF*.
- **Incertitudes sur le score de sortie du BDT identification de photons** : pour prendre en compte les différences données et simulations. Ces incertitudes sont dénommées *CMS\_hgg\_phoIdMva*.
- **Incertitudes sur la résolution en énergie par photon** : pour prendre en compte les différences données et simulations. Ces incertitudes sont dénommées *CMS\_hgg\_SigmaEOverEShift*.
- **Incertitudes sur les efficacités de déclenchement** : venant de la mesure de l'efficacité avec des évènements  $Z \rightarrow ee$  et la méthode TnP. Ces incertitudes sont dénommées *CMS\_hgg\_TriggerWeight*.
- **Incertitudes sur le phénomène de *prefiring* pour 2017** : pour prendre en compte l'effet associé. Ces incertitudes sont dénommées *CMS\_hgg\_prefireWeight*.

### Incertitudes liées à la classe VBF

Si l'on considère une classe VBF additionnelle, voir la section 6.12.1 dédiée, des incertitudes supplémentaires, voir [104] pour plus de détails, doivent être incorporées :

- **Incertitudes sur l'échelle et la résolution en énergie des jets** : dénommées *CMS\_hgg\_JER\_migration* et *CMS\_hgg\_JEC\_migration*.

- **Incertainces sur l'identification de jets dus à l'empilement** : obtenues avec des comparaisons données/simulations. Ces incertainces sont dénommées *CMS\_hgg\_PUJIDShift\_migration0*.
- **Incertainces sur la modélisation des radiations de jets supplémentaires** : pour prendre en compte des effets non physiques présents dans les simulations des événements VBF avec PYTHIA. Ces incertainces sont dénommées *CMS\_hgg\_JetHEM* et *DipoleRecoil*.

### 6.10.2 Impacts des incertainces

Afin de visualiser, de façon concrète, l'importance relative de ces incertainces, on propose de visualiser leur impact sur la force du signal recherché (qui correspond au rapport entre la section efficace fois le rapport de branchement  $h \rightarrow \gamma\gamma$  mesuré dans l'analyse sur cette même quantité mais pour le Modèle Standard). On présente ici l'effet de ces systématiques sur un signal attendu de force  $r = 1$  : on génère des jeux de pseudo-données à partir du modèle de bruit de fond obtenu précédemment dans lequel on injecte notre modèle de signal de force 1 à différentes masses, on ajuste ensuite une fonction de vraisemblance<sup>12</sup> que l'on minimise pour des variations de  $\pm 1\sigma$  de l'incertitude systématique pour chaque paramètre considéré afin d'observer l'effet de cette variation sur la variation  $\Delta r$  de la force du signal. Les paramètres liés à une incertitude systématique sont appelés paramètres de nuisance. Notons qu'en pratique, la génération de pseudo-données est trop coûteuse, on préfère donc utiliser un *Asimov dataset* [118], généré ici à partir des valeurs nominales des paramètres de nuisance, utilisant une hypothèse asymptotique où les fluctuations statistiques sont supprimées.

Les impacts des 30 incertainces systématiques les plus importantes sont visibles pour des masses de signal injecté de 70 GeV, 90 GeV, 93 GeV et 110 GeV pour 2017 et 2018 respectivement sur les figures 6.48 à 6.52 : le *pull* de ces incertainces y est aussi représenté, il correspond, dans le cas signal+bruit de fond présenté ici, comme la différence entre la valeur du paramètre associé à l'incertitude considérée après injection du signal (soit *post-fit*) et la valeur nominale du paramètre (avant injection du signal, soit *pre-fit*) divisée par la valeur *pre-fit* de l'incertitude associée ; les barres d'erreur asymétriques associées représentent la valeur de l'incertitude *post-fit* divisée par sa valeur *pre-fit*. Ces résultats incluent les paramètres de nuisance correspondant à la modélisation du bruit de fond ainsi que la classe additionnelle VBF. Les paramètres de la fonction dCB+exp apparaissent sous les noms *DYee\_13TeV\_year\_UntaggedTag\_class\_DCB\_para* où *year* est l'année, *class* à la catégorie considérée et *para* le paramètre correspondant. La variation  $\Delta r$  induite par les paramètres de nuisance est déterminée par la méthode de vraisemblance profilée qui sera décrite dans la section suivante. Cette variation est obtenue indépendamment pour chaque paramètre de nuisance en fixant les autres.

On observe que, pour ces points de masse, la force du signal est bien définie et n'est pas dominée par des variations induites par des variations des paramètres de nuisance ; ces zones sont donc sensibles à la présence d'un signal. On peut aussi remarquer que les paramètres de nuisance associés aux paramètres de la fonction dCB+exp utilisée pour modéliser la composante

12. Cette méthode est d'avantage explicitée dans la section suivante.

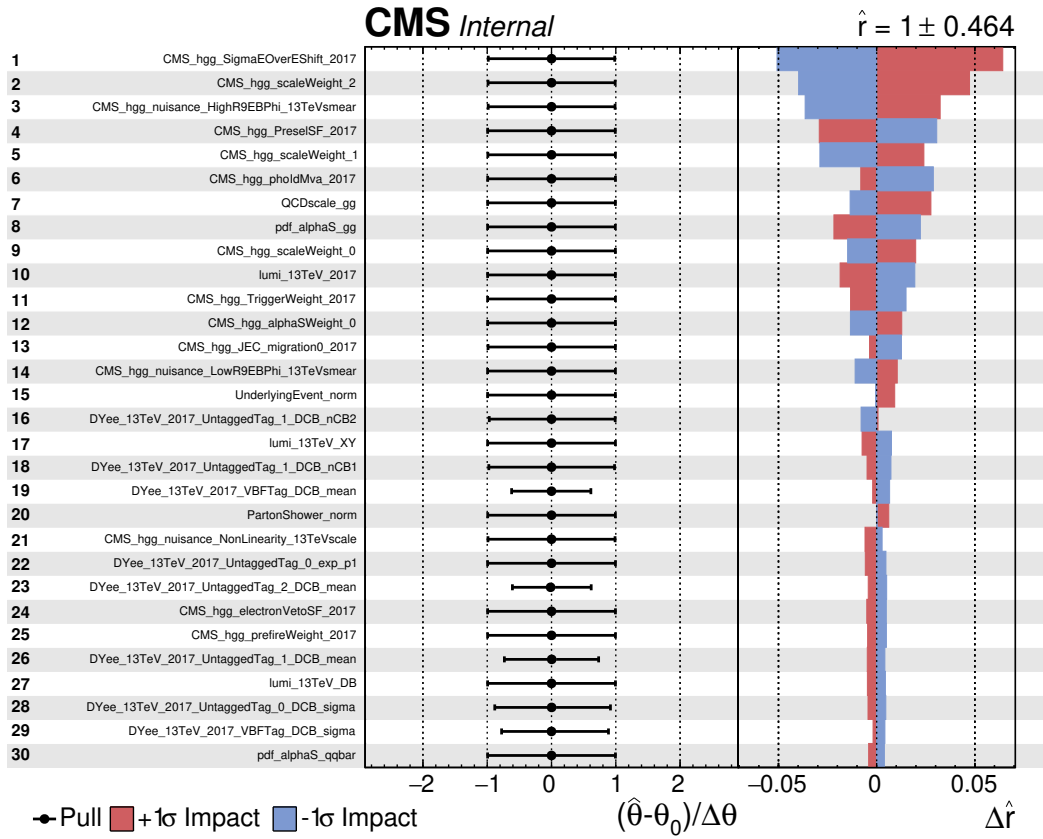


FIGURE 6.48 – Impacts des paramètres de nuisance sur la force du signal pour un signal attendu  $r = 1$  de masse  $m_H = 70$  GeV en 2017.

Drell-Yan sont dominants pour des masses proches du pic du Z, soit 90 et 93 GeV, ce qui est cohérent.

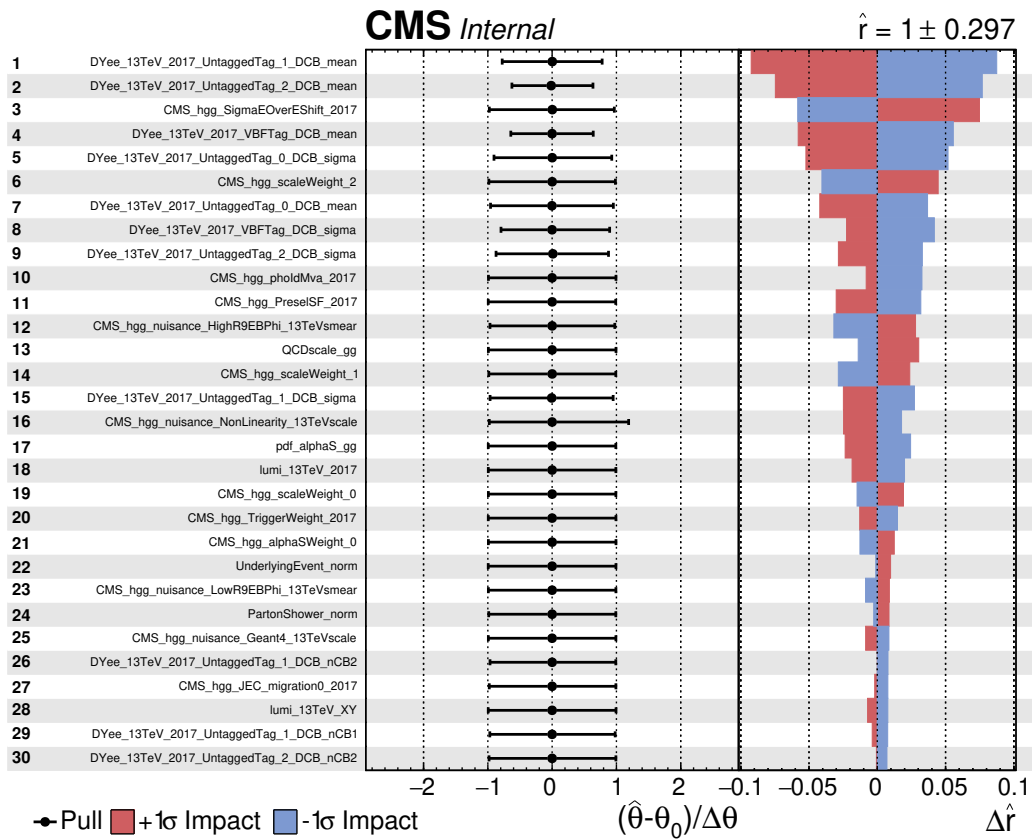
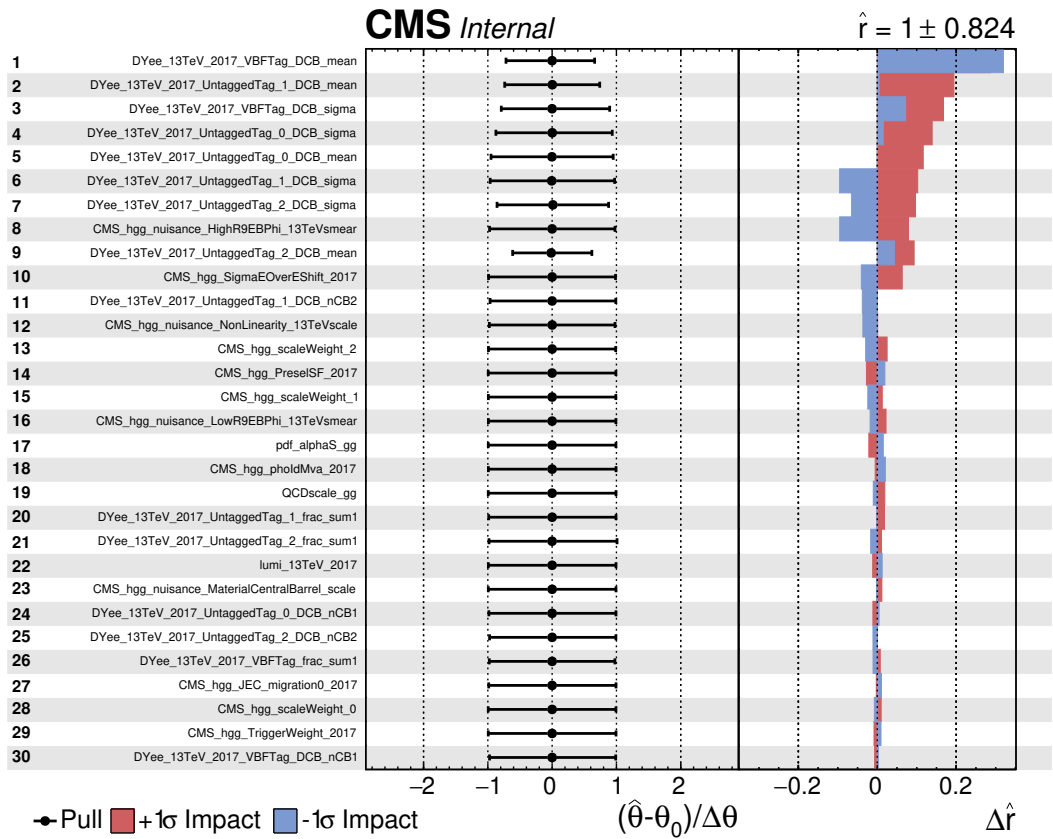


FIGURE 6.49 – Impacts des paramètres de nuisance sur la force du signal pour un signal attendu  $r = 1$  de masse  $m_H=90$  (haut) et 93 (bas) GeV en 2017.

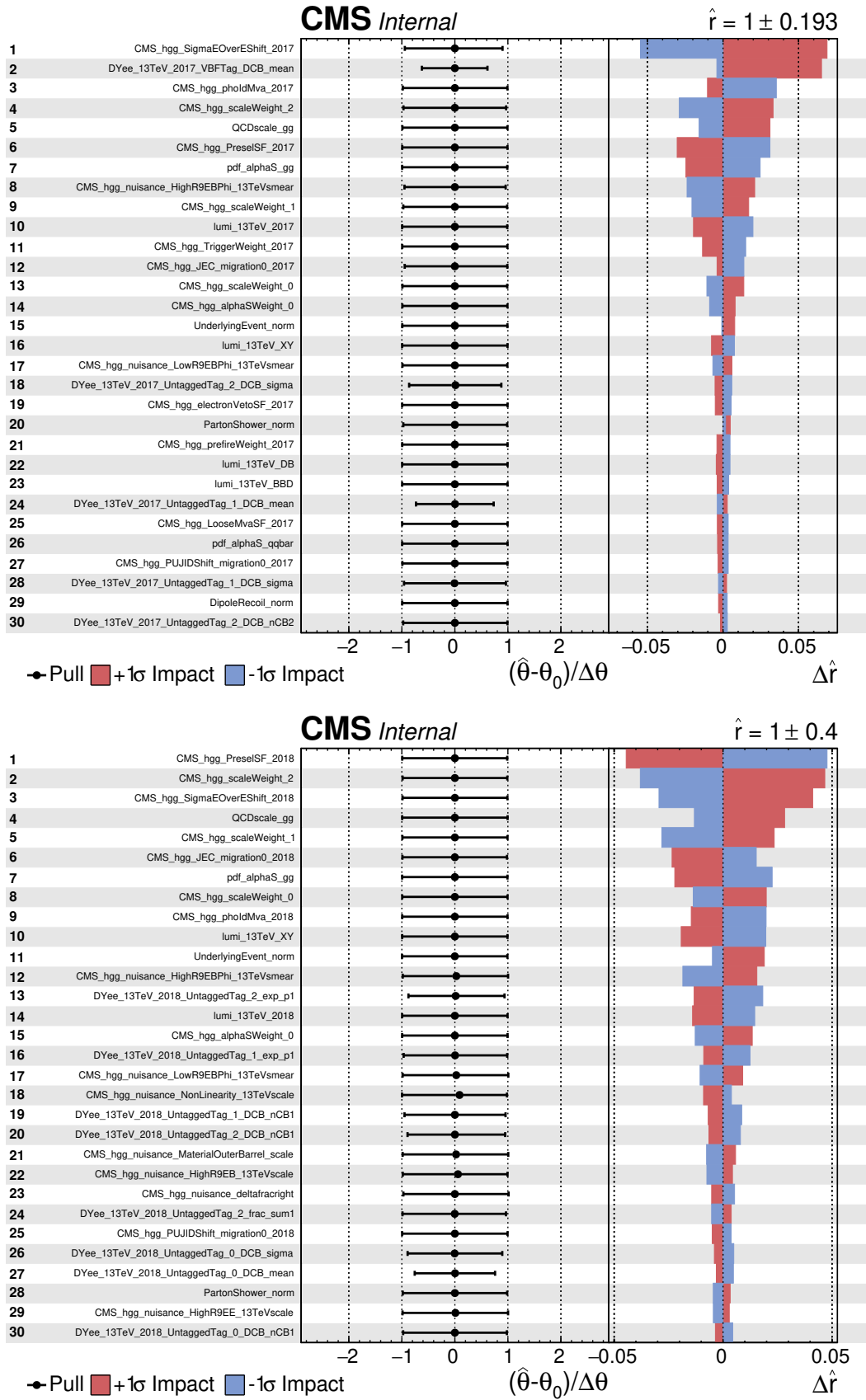


FIGURE 6.50 – Impacts des paramètres de nuisance sur la force du signal pour un signal attendu  $r = 1$  de masse  $m_H = 110$  GeV en 2017 (haut) et  $m_H = 70$  GeV en 2018 (bas).



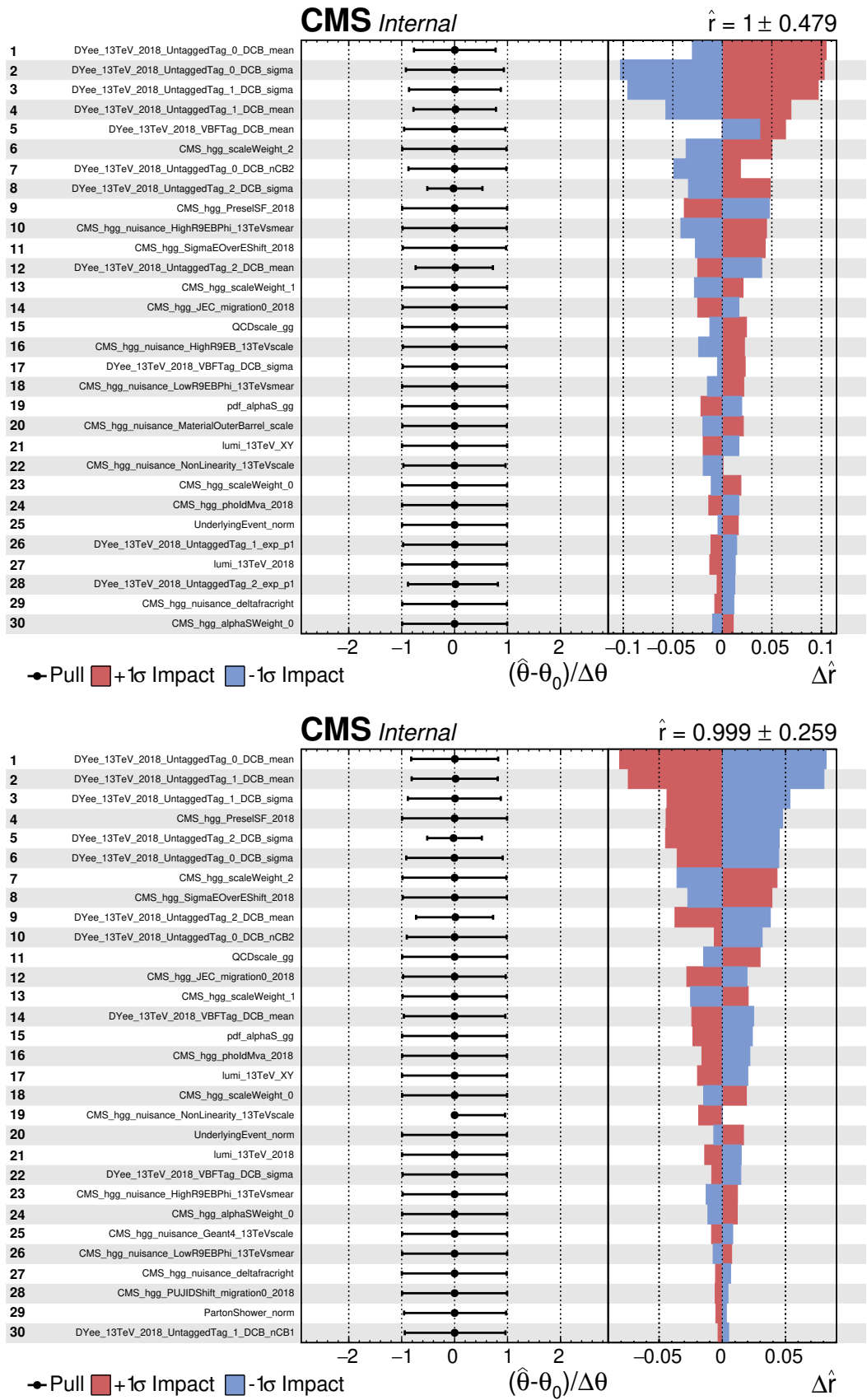


FIGURE 6.51 – Impacts des paramètres de nuisance sur la force du signal pour un signal attendu  $r = 1$  de masse  $m_H=90$  (haut) et  $93$  (bas) GeV en 2018.

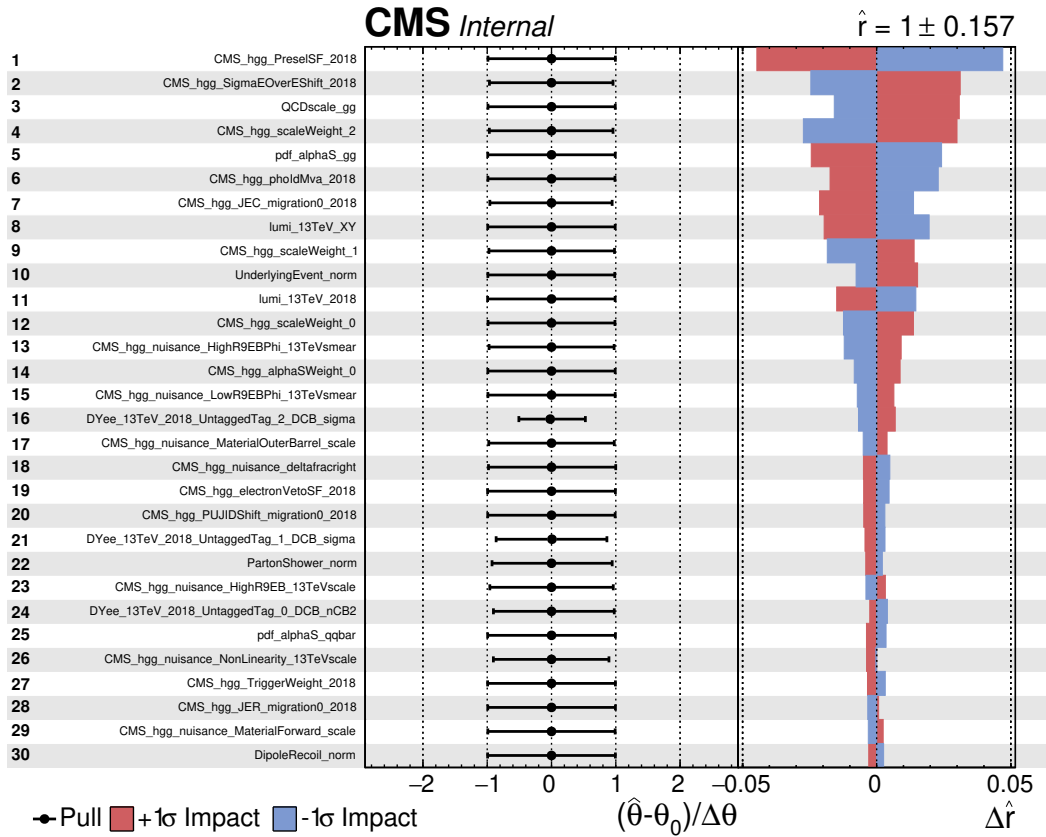


FIGURE 6.52 – Impacts des paramètres de nuisance sur la force du signal pour un signal attendu  $r = 1$  de masse  $m_H = 110$  GeV en 2018.

## 6.11 Résultats finaux

L'objectif de l'analyse est de calculer les **limites d'exclusion** sur le produit section efficace fois rapport de branchement en deux photons pour un nouveau boson de Higgs en absence de signal observé et la probabilité (p-value), si l'on est en présence d'un excès d'évènements, que ce signal ne soit qu'un simple excès d'évènements imputable à une fluctuation statistique du bruit de fond afin de statuer sur l'éventualité d'une découverte. Les méthodes utilisées pour dériver ces résultats sont les mêmes que celles utilisées dans [103]. Les résultats présentés dans cette partie sont le fruit de mon travail.

### 6.11.1 Analyse statistique

#### Méthode d'ajustement par maximisation de la fonction de vraisemblance

Le calcul de ces quantités se base sur une méthode d'ajustement par maximisation d'une fonction de vraisemblance binnée [28] reposant sur une méthode fréquentiste modifiée [119]. Concrètement, on réalise un ajustement simultané de la distribution  $m_{\gamma\gamma}$  des modèles signal et bruit de fond sur les données par classe, réalisé par une maximisation d'une fonction de vraisemblance calculée par bin de taille 0,1 GeV pour des masses  $m_{\gamma\gamma}$  de 70 à 110 GeV. Cette

méthode est décrite en détail dans la suite.

On commence par définir nos paramètres d'intérêt (POI) correspondant aux quantités que l'on souhaite mesurées. Ici, on ne considère qu'un seul POI qui est la force du signal (on peut aussi définir  $\mu$  par classe ou encore par mode de production, noté  $\mu$  (noté  $r$  dans la section précédente), et défini comme la valeur par laquelle on doit multiplier le produit section efficace fois rapport de branchement en deux photons d'un boson de Higgs SM-like pour avoir ce même produit pour un nouveau boson de Higgs de même masse (sans notion d'un modèle particulier), soit :

$$\mu = \frac{\sigma(h) \times BR(h \rightarrow \gamma\gamma)}{\sigma(h_{SM}) \times BR(h_{SM} \rightarrow \gamma\gamma)} \quad (6.15)$$

On définit aussi l'ensemble des paramètres de nuisance (NP), noté  $\vec{\theta}$ , qui sont des paramètres qui n'ont pas d'intérêt direct mais qui influent sur les résultats observés. Comme abordé dans la partie précédente, les NP sont des paramètres auxquels sont assignés une incertitude. Ces NP sont contraints, selon les valeurs des incertitudes associées, autour de leur valeur centrale  $\vec{\tilde{\theta}}$ .

La fonction de vraisemblance  $\mathcal{L}$  utilisée dans cette méthode dépend du POI, des NP et aussi de la masse  $m_H$  du signal considérée, et est donnée comme le produit des fonctions de vraisemblance par classe selon :

$$\mathcal{L}(data|\mu, m_H, \vec{\theta}) = \prod_c^{N_{cat}} \mathcal{L}_c(data|\mu, m_H, \vec{\theta}) \quad (6.16)$$

où les fonctions de vraisemblance profilées par classe  $\mathcal{L}_c$  sont données par une distribution de Poisson binnée selon la formule :

$$\mathcal{L}_c(data|\mu, m_H, \vec{\theta}) = \prod_i^{N_{bin}} Poisson(d_i|\mu \times s_i(m_H, \vec{\theta}) + b_i(\vec{\theta})) p(\vec{\theta}|\vec{\tilde{\theta}}) \quad (6.17)$$

où  $N_{bin}$  est le nombre de bins utilisé, *Poisson* correspond à une fonction de poisson évaluée avec le nombre d'évènements observés dans les données pour le  $i$ -ème bin  $d_i$  et vaut  $\frac{(\mu \times s_i + b_i)^{d_i}}{d_i!} e^{-\mu \times s_i - b_i}$ ,  $s_i$  et  $b_i$  correspondent au nombre d'évènements dans le modèle signal et bruit de fond pour le  $i$ -ème bin, et avec  $p(\vec{\theta}|\vec{\tilde{\theta}})$  le produit des densités de probabilité des incertitudes systématiques, qui correspondent en général à des distributions gaussiennes et log-normales ici, évaluées en  $\vec{\theta} = \vec{\tilde{\theta}}$  en les valeurs centrales des NP. Ce dernier terme est un terme qui pénalise les déviations des NP du signal et du bruit de fond par rapport à leur valeur centrale.

La fonction de vraisemblance  $\mathcal{L}$  est alors maximisée; en pratique, on minimise plutôt la fonction log-vraisemblance associée  $-\ln(\mathcal{L})$ . Les paramètres libres de l'ajustement sont le POI  $\mu$  et  $m_H$ ; les NP peuvent aussi varier mais sont contraints par le terme décrit dans le paragraphe précédent. Les paramètres obtenus, après minimisation, sont notés avec le symbole " $\hat{\mu}$ ".

Une approche fréquentiste est ensuite utilisée pour extraire la valeur de meilleur ajustement du POI ainsi que l'incertitude associée. On définit pour cela le rapport de vraisemblance profilé, qui quantifie une différence de fonctions log-vraisemblance, qui tient lieu de quantité discriminante<sup>13</sup> :

$$\tilde{q}_\mu = -2 \ln \frac{L(\text{data}|\mu, \hat{m}_{H,\mu}, \hat{\theta}_\mu)}{L(\text{data}|\hat{\mu}, \hat{m}_H, \hat{\theta})} \quad (6.18)$$

où  $\hat{m}_{H,\mu}$  et  $\hat{\theta}_\mu$  sont les valeurs de meilleur ajustement à  $\mu$  donné, on a par exemple  $\forall \mu$ ,  $\frac{\partial L(d|\mu, \hat{m}_{H,\mu}, \hat{\theta}_\mu)}{\partial \theta} = 0$ , et  $\hat{\mu}$ ,  $\hat{m}_H$  et  $\hat{\theta}$  sont les valeurs de meilleur ajustement global pour le POI, la masse du signal, et les NP.  $\tilde{q}_\mu$  est ainsi défini comme étant positif. En règle générale, des valeurs de  $\tilde{q}_\mu$  proches de 0 indiquent un accord avec la valeur de  $\mu$  testée et des valeurs élevées indiquent un désaccord.

Le rapport de vraisemblance  $\tilde{q}_\mu$  est utilisé pour comparer deux hypothèses : l'hypothèse où les données correspondent à du seul bruit de fond  $b$  et l'hypothèse où les données sont en accord avec la présence de signal en plus du bruit de fond  $s + b$ . Pour effectuer cette comparaison, on commence par calculer  $\tilde{q}_\mu^{obs}$  qui est la valeur du rapport de vraisemblance obtenue pour les données observées pour une valeur de  $\mu$  donnée. Ceci permet de définir respectivement le degré de confiance associée à l'hypothèse d'être en présence de bruit de fond seulement et de signal + bruit de fond qui sont :

$$1 - p_b = CL_b = P(\tilde{q}_\mu > \tilde{q}_\mu^{obs} | b) = \int_{\tilde{q}_\mu^{obs}}^{+\infty} f(\tilde{q}_\mu | 0, \hat{m}_{H,0}^{obs}, \hat{\theta}_0^{obs}) d\tilde{q}_\mu \quad (6.19)$$

et

$$p_\mu = P(\tilde{q}_\mu > \tilde{q}_\mu^{obs} | s + b) = CL_{s+b} = \int_{\tilde{q}_\mu^{obs}}^{+\infty} f(\tilde{q}_\mu | \mu, \hat{m}_{H,\mu}^{obs}, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu \quad (6.20)$$

où  $f(q|s+b)$  et  $f(q|b)$  sont les densités de probabilité de  $\tilde{q}_\mu$  associées aux hypothèses bruit de fond seul et signal+bruit de fond obtenues pour les valeurs de  $m_H$  et des NP maximisant les fonctions de vraisemblance dans le cas de l'hypothèse considérée. Elles sont en principe calculées en générant des jeux de pseudo-données, on verra cependant dans la suite que ce n'est pas ce qui est utilisé dans la pratique à cause de la consommation CPU demandée. D'après leur définition,  $p_\mu$  et  $p_b$  sont des p-values correspondant à la probabilité d'obtenir, par chance, un  $\tilde{q}_\mu$  supérieur (inférieur) à celui mesuré dans l'hypothèse  $s+b$  ( $b$ ). Un schéma représentatif de ces quantités est visible sur la figure 6.53 en guise d'illustration.

### Limite d'exclusion en l'absence de signal

Un critère d'exclusion peut finalement être calculé avec le degré de confiance associé à la présence d'un signal de force  $\mu$  en utilisant une prescription fréquentiste modifiée [120] :

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_\mu}{1 - p_b} \quad (6.21)$$

13. L'approche suivante est semblable à celle étudiée précédemment avec les lois de  $\chi^2$ .

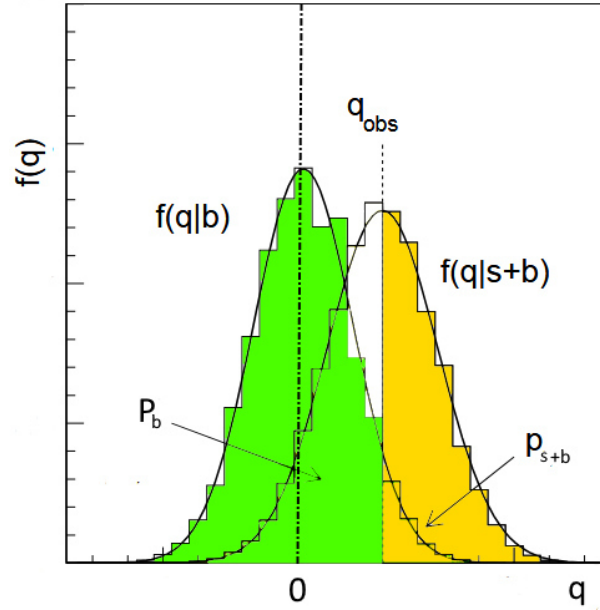


FIGURE 6.53 – Exemple de distribution de  $q$  pour les hypothèse  $s+b$  (signal plus bruit de fond) et  $b$  (bruit de fond seul) montrant les  $p$ -valeurs associées. Les distributions en histogramme sont construites à partir de jeux de pseudo-données, les courbes solides sont obtenues dans le cadre de l'approximation asymptotique [118].

Cette approche permet notamment de réduire les taux de fausses exclusions et d'augmenter le taux de couverture, en terme de paramètres, de l'analyse par rapport à l'utilisation de la prescription  $CL_{s+b}$  [120].

Pour exclure la présence d'un signal avec un niveau de confiance  $CL$  de 95%, on trouve la valeur du POI  $\mu$  permettant d'obtenir une valeur de  $CL_s$  proche de 0,05. Cela permet alors de fixer la valeur de  $\mu$ . On peut ensuite exclure avec un intervalle de confiance à 95 % les produits de section efficace  $\sigma(h)$  fois rapport de branchement en deux photons  $BR(h \rightarrow \gamma\gamma)$  tels que :

$$\sigma(h) \times BR(h \rightarrow \gamma\gamma) < \mu^{95\%} \times (\epsilon A)^{-1} \times \sigma(h_{SM}) \times BR(h_{SM} \rightarrow \gamma\gamma) \quad (6.22)$$

où  $\epsilon A$  est le produit de l'efficacité de l'analyse et de l'acceptance du détecteur, à noter ici que tous les termes dépendent de la masse  $m_H$  de la particule recherchée.

### Limites attendues et approximation asymptotique

En pratique, les limites d'exclusion présentées plus haut sont les limites dites "observées" sur le produit de la section efficace, multiplié par le rapport de branchement en deux photons, de production d'un nouveau boson de Higgs. Ces limites se comparent avec les limites attendues sur la section efficace en absence de signal obtenues en appliquant la méthode précédente (calcul de  $CL_s$  et de  $\mu^{95\%}$ ) non pas directement aux données mais à un jeu de pseudo-données de bruit de fond uniquement (hypothèse  $b$ ) qui correspond aux courbes en pointillé "Expected" présentées ci-après. En utilisant de nombreux jeux de pseudo-données, on peut obtenir une distribution de  $\mu^{95\%}$  pour le bruit de fond uniquement ce qui permet de calculer les bandes

d'incertitude à  $\pm 1\sigma$  et  $\pm 2\sigma$  pour les limites attendues sur la section efficace en absence de signal, ces dernières étant les bandes vert et jaune visible par exemple sur la figure 6.54. Ces valeurs correspondent respectivement à des valeurs de  $\mu^{95\%}$  qui permettent d'obtenir, toujours respectivement, 68 % et 95 % des valeurs dans la distribution de  $\mu^{95\%}$  en absence de signal.

Comme vu dans la section précédente, la génération des pseudo-données, largement utilisée dans les méthodes présentées dans cette section, est très coûteuse en temps CPU et n'est donc pas utilisée dans la pratique. On utilise à la place une approximation asymptotique (c'est-à-dire qu'on considère que le nombre d'évènements est très élevé) permettant la génération d'*Asimov dataset* à partir de la distribution de probabilité du bruit de fond seul. Cet échantillon est un échantillon "parfait" c'est-à-dire qu'il n'y contient pas de fluctuation statistique, il permet en outre un calcul pratique et rapide des quantités présentées ici par des formules analytiques obtenues grâce à l'approximation asymptotique. Se référer à [118] pour plus de détails.

### Quantification d'un excès d'évènements

Pour quantifier la présence d'un éventuel signal dans les données, on calcule la p-value associée à l'hypothèse de bruit de fond seul qui correspond alors à la probabilité pour le bruit de fond de fluctuer et créer un excès au moins aussi grand que celui observé dans les données. Plus cette p-value est petite, plus on a des chances que l'excès observé corresponde bien à du signal. Elle est définie en évaluant l'Eq. 6.20 en  $\mu=0$  comme :

$$p_0 = \int_{q_0^{obs}}^{\infty} f(\tilde{q}_0 | 0, \hat{\theta}_0^{obs}) dq_0 \quad (6.23)$$

Il est usuel d'exprimer la valeur de cette p-value en terme de signficance  $Z$  correspondant au nombre de déviations standards  $\sigma$  d'une loi gaussienne selon :

$$p_0 = \int_Z^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \quad (6.24)$$

On parle de découverte lorsqu'un signal est observé avec une p-value en absence de signal associée inférieure à  $p_0 = 2,8 \times 10^{-7}$ , soit, en terme de déviations standards, supérieure à  $5\sigma$ .

A nouveau cette définition demande de générer des pseudo-données gourmandes en CPU. En pratique, on calcule donc la p-value dans la limite asymptotique, fournissant d'excellentes approximations [119], et donnée par :

$$p_0 = \frac{1}{2} \left[ 1 - \text{erf}\left(\sqrt{q_0^{obs}/2}\right) \right] \quad (6.25)$$

où *erf* est la fonction erreur.

La p-value définie ici est évaluée pour une masse  $m_H$  de signal donnée, on parle alors de p-value "locale". Il faut cependant prendre en compte le fait que l'on calcule cette p-value pour tout un set de masse  $m_H$  puisqu'on ne sait pas a priori où le signal éventuel se situe dans la

gamme de masse. Ce phénomène est appelé *look-elsewhere effect* et est pris en compte par le calcul d'une p-value "globale". Sans entrer dans les détails, ce calcul s'effectue en comparant le nombre de jeux de pseudo-données qui ont une signification, dans l'hypothèse bruit de fond seul, supérieure à la signification maximale observée dans les données [119].

Dans la suite on présente la p-value attendue ("Expected") qui est calculée en générant un *Asimov dataset* dans l'hypothèse d'un signal attendu de force  $\mu = 1$ .

### 6.11.2 Résultats des tests statistiques

On présente ici les résultats attendus, obtenus avec les méthodes décrites précédemment dans l'approximation asymptotique, pour le cas où l'on utilise les trois classes *diphoton BDT* et pour des masses entre 70 et 110 GeV.

Les limites d'exclusion attendues, pour un niveau de confiance de 95 %, du produit de la section efficace d'un nouveau boson de Higgs fois son rapport de branchement en deux photons, normalisé aux valeurs du MS [25], sont données sur la figure 6.54 pour 2017 et 2018. Les bandes vertes et jaunes correspondent aux bandes d'erreur à  $\pm 1\sigma$  et  $\pm 2\sigma$  introduites dans la sous-section précédente. Une dégradation de la sensibilité à une masse proche du pic du Z peut y être observée, elle est induite par la composante Drell-Yan  $Z \rightarrow ee$  et les incertitudes associées. On s'attend donc à exclure un signal de force supérieure à 1 (0,9) à 70 GeV et 0,35 (0,30) à 110 GeV pour 2017 (2018). Le sensibilité est meilleure en 2018 par rapport à celle de 2017; ceci s'explique par l'augmentation de luminosité et de l'efficacité fois acceptance en 2018 résultant en un meilleur ratio signal sur bruit de fond.

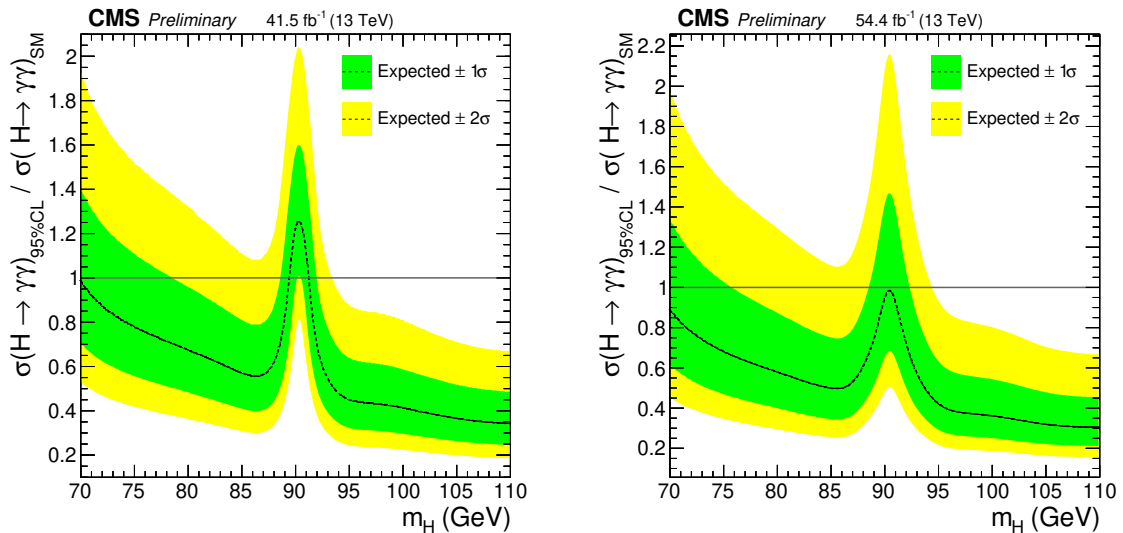


FIGURE 6.54 – Limite d'exclusion attendue à 95 % sur le produit de la section efficace de production par le rapport de branchement en deux photons d'un nouveau boson de Higgs, normalisé par sa valeur du MS, en fonction de la masse d'un tel boson pour 2017 (gauche) et 2018 (droite), et bandes d'incertitude associées.

La p-value attendue, calculée dans le cas d'un signal de force 1, est quant à elle visible sur

la figure 6.55. Outre la dégradation autour du pic du Z, on peut observer que la significativité est supérieure à  $5\sigma$  pour des masses supérieures à 100 (96) GeV pour 2017 (2018), on pourrait donc statuer sur la découverte d'une nouvelle particule dans ces gammes de masse si la force du signal correspondante est au moins égale à 1.

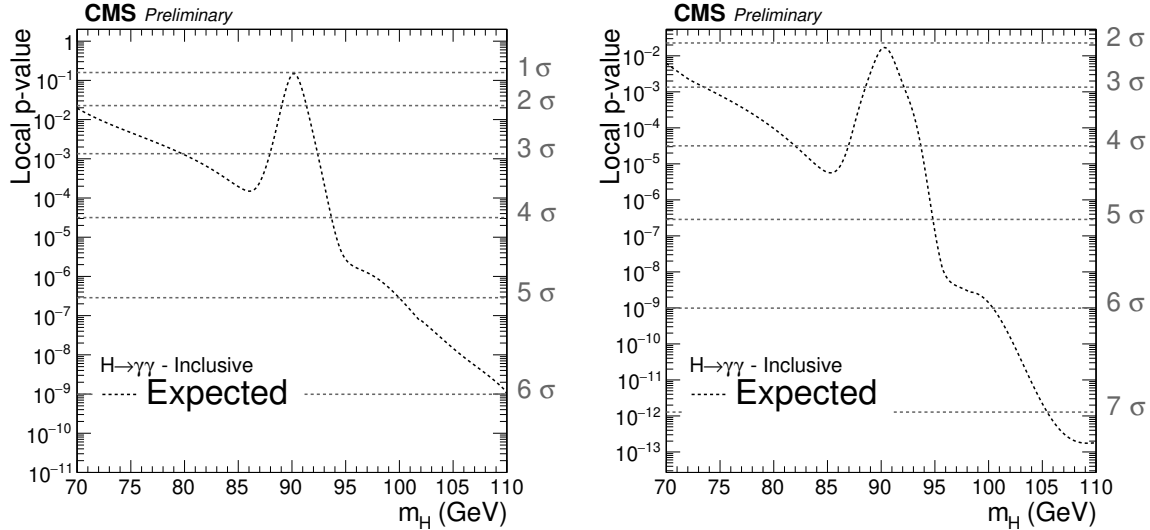


FIGURE 6.55 – P-value locale attendue en fonction de la masse d'un nouveau boson de Higgs pour 2017 (gauche) et 2018 (droite). La significativité correspondante en terme de déviations standards d'une loi gaussienne est donnée en ordonnée.

## 6.12 Résultats avec catégorie VBF

### 6.12.1 Catégorie VBF

Afin d'augmenter la sensibilité de notre analyse aux événements signal  $h \rightarrow \gamma\gamma$  produits par le processus VBF, on ajoute une classe supplémentaire VBF ayant pour objectif d'étiqueter de tels événements. La figure 6.56 présente un exemple de diagramme de Feynman pour un événement VBF. Une augmentation de la sensibilité VBF permet alors de mieux contraindre les modèles théoriques privilégiant ce mode de production comme celui présenté dans la section 8.2. Ce travail n'étant pas le mien, il ne sera que brièvement décrit ici, plus de détails sont néanmoins disponibles dans [104]. La méthode utilisée est similaire à celle employée dans [107].

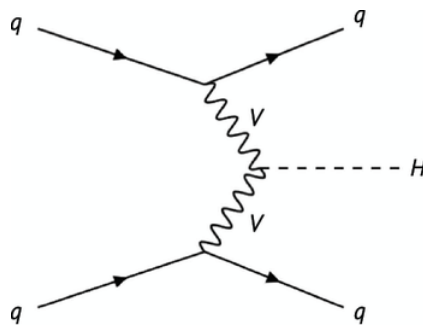


FIGURE 6.56 – Diagramme de production d'un boson de Higgs par fusion de bosons vecteurs



Pour cela on commence par sélectionner les évènements ayant au moins deux jets dans l'état final en addition des deux photons usuels, en choisissant les deux jets ayant le plus haut  $p_T$ . Ces deux jets doivent satisfaire les conditions suivantes :

- $p_T^{j_1} > 30$  GeV et  $p_T^{j_2} > 20$  GeV
- $m_{j_1, j_2} > 100$  GeV
- 2 jets passant un algorithme d'identification de jet d'empilement dédié [90]

Pour distinguer les évènements correspondant au mode de production VBF, on utilise deux BDT discriminants : le *Dijet MVA* et le *Combined MVA*.

Le *Dijet MVA* est un BDT utilisé pour distinguer les évènements VBF des autres évènements grâce à des variables d'entrée discriminantes liées aux variables cinématiques. L'entraînement s'effectue sur des simulations. Le signal utilisé pour l'entraînement est le signal de l'analyse pour une production VBF avec une masse de 60 à 115 GeV par pas de 5 GeV. Le bruit de fond utilisé pour l'entraînement est le bruit de fond de l'analyse, passant les coupures sur les jets précédentes, sans les échantillons DY et QCD, ainsi que des échantillons signal ggH avec une masse de 65 à 115 GeV par pas de 5 GeV afin de permettre une bonne différenciation entre les différents évènements signal dijet. La sortie du *Dijet MVA* est un score discriminant, basé sur les propriétés cinématiques, les évènements VBF des autres.

Pour améliorer la discrimination des évènements VBF, on utilise ensuite le *Combined MVA* dont les variables d'entrée sont le score de sortie *Dijet MVA*, le score de sortie du *diphoton BDT* ainsi que la quantité  $p_T^{\gamma\gamma} / m_{\gamma\gamma}$ . L'entraînement utilise les mêmes échantillons VBF pour le signal et les échantillons de bruit de fond standard de l'analyse, toujours sans QCD. La sortie du *Combined MVA* est un score qui permet de discriminer les évènements VBF des autres évènements. Ce score permet ensuite de définir une catégorisation des évènements VBF.

Ces deux BDT sont entraînés pour 2017 et cet entraînement est utilisé pour les données et simulations 2017 et 2018, comme la plupart des BDT présentés dans ce chapitre. L'entraînement est validé avec des évènements  $Z \rightarrow ee + 2$  jets selon la méthode présentée dans la section 6.7.1. Les incertitudes systématiques correspondant aux évènements VBF-like, décrites dans la section 6.10, sont ajoutées ici, permettant alors de couvrir les différences entre données et simulations. La figure 6.57 présente la validation données/simulations avec des éléments  $Z \rightarrow ee + 2$  jets pour le score de sortie du *Combined MVA*.

On utilise donc ce score de sortie pour classifier les évènements VBF-like. Le nombre d'évènements obtenus passant la sélection définie plus haut étant faible, on choisit de n'utiliser qu'une classe VBF dédiée. Pour choisir la coupure en terme de score *Combined MVA*, on utilise la même méthode que celle présentée dans la sous-section 6.7.2 pour le *Diphoton BDT* c'est-à-dire qu'on calcule, avec des simulations, la p-value attendue associée à un modèle simple où le bruit de fond est décrit par une exponentielle et le signal, constitué de simulations de signal VBF à 90 GeV, par une somme de gaussienne en fonction de la coupure utilisée pour former la catégorie VBF. Cette p-value, notée  $Z_0$ , est visible sur la figure 6.58. Il est aussi nécessaire que cette catégorie contienne assez d'évènements DY *double-fake* afin de réaliser l'ajustement

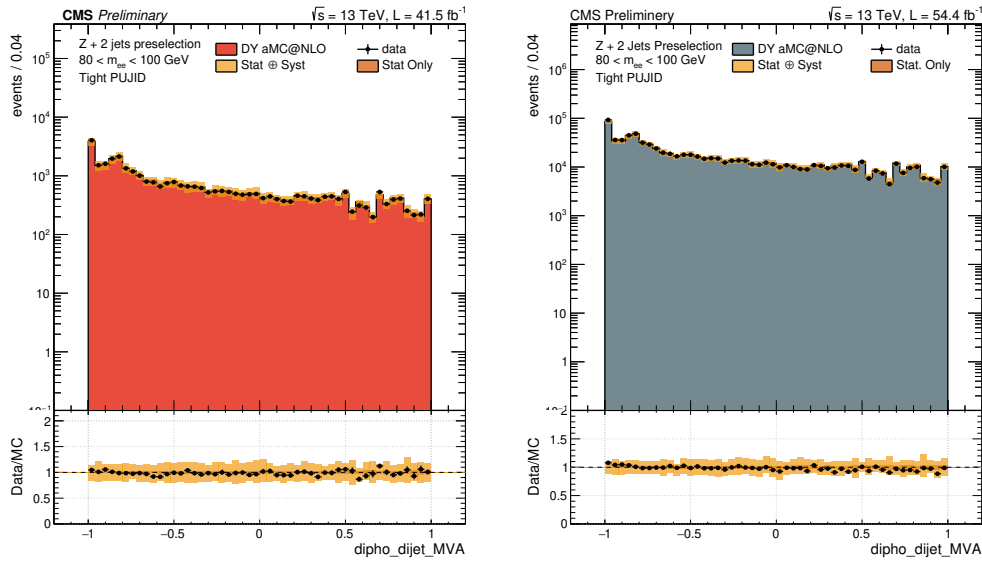


FIGURE 6.57 – Comparaison entre données et simulations pour 2017 (gauche) et 2018 (droite) de la distribution du score de sortie *Combined MVA* du système diélectron + 2 jets où les électrons sont reconstruits comme des photons. La comparaison entre données (points noirs) et simulations (bleu plein) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.

de la composante DY pour cette classe. On choisit pour cela une coupure de 0,8 sur le score *Combined MVA* pour définir la classe VBF pour 2017 et 2018<sup>14</sup>.

Les évènements satisfaisant les critères de coupures précédents et dont le score de sortie *Combined MVA* est supérieure à 0,8 sont donc catégorisés dans la classe VBF dédiée; les autres évènements vont ensuite peupler les trois classes<sup>15</sup> du *diphoton BDT* décrites précédemment.

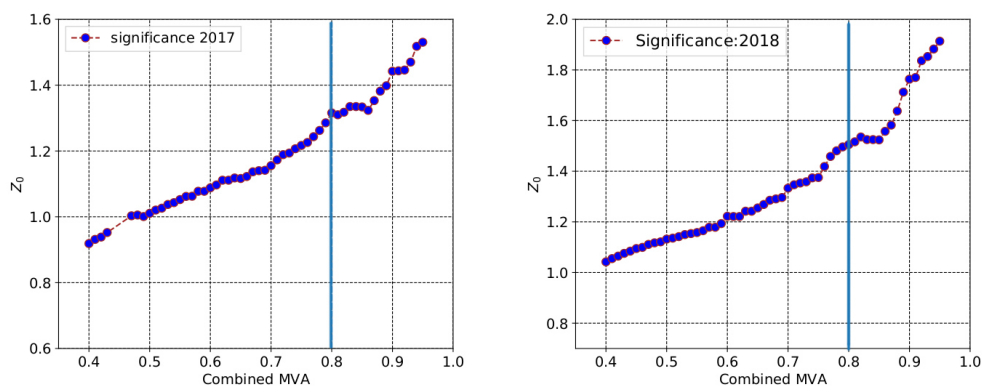


FIGURE 6.58 – P-value, notée  $Z_0$ , attendue pour un signal VBF, simplifiée et exprimée en terme de signficance en fonction de la coupure en terme de score de sortie *Combined MVA* définissant la classe VBF pour 2017 (gauche) et 2018 (droite). La ligne bleu verticale présente la valeur retenue.

14. Une réoptimisation de cette frontière a été effectuée et les résultats obtenus ne sont pas significativement différents.

15. Il y aura donc un peu moins d'évènements dans ces classes par rapport au cas sans classe VBF.

Finalement, les étapes de modélisation du signal et du bruit de fond sont refaites avec ces 3 classes *diphoton* BDT+ 1 classe VBF. On peut alors présenter les résultats finaux de l'analyse incluant la classe VBF.

### 6.12.2 Résultats finaux

On montre finalement ici les résultats attendus en incluant la classe VBF et les NP associés. Les limites d'exclusion normalisées pour 2017 et 2018 sont présentées, séparées en différents modes de production de signal, ainsi que les limites d'exclusion absolues calculées en multipliant les limites précédentes par le produit section efficace fois rapport de branchement en deux photons pour le cas d'un boson de Higgs SM-like. Ce produit, et ses incertitudes systématiques, est d'ailleurs visible<sup>16</sup> sur les figures associées.

Les figures 6.59 et 6.60 présentent ainsi les limites d'exclusion normalisées et absolues pour 2017 et 2018 respectivement. Ces limites sont globalement meilleures que celles obtenues sans la classe VBF (surtout pour 2018).

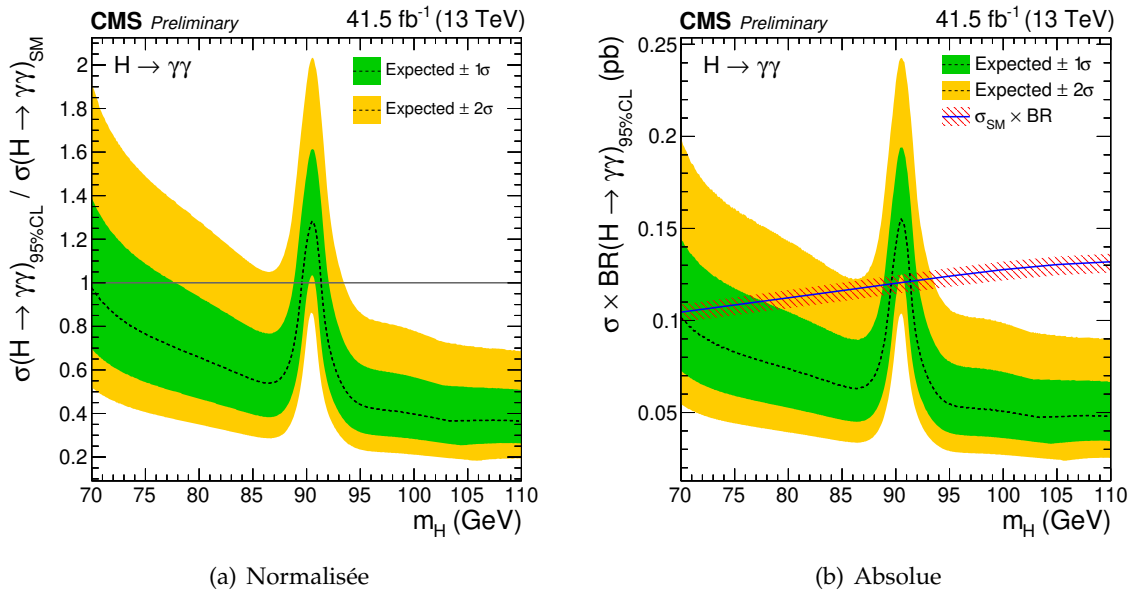


FIGURE 6.59 – Limite d'exclusion attendue à 95 % sur le produit de la section efficace de production par le rapport de branchement en deux photons d'un nouveau boson de Higgs pour 2017 dans le cas normalisé par sa valeur du MS a) et dans le cas absolu b) en fonction de la masse d'un tel boson, et bandes d'incertitude associées.

Les limites d'exclusion absolues dans l'hypothèse où le signal est uniquement produit par VBF sont visibles sur la figure 6.61 pour 2017 et 2018. On observe aussi, sur cette même figure, que ces limites d'exclusion sont plus strictes d'environ  $\sim 40\%$  dans le cas où on utilise la classe VBF et les trois classes *diphoton* BDT par rapport au cas où l'on a seulement les trois classes *diphoton* BDT. On obtient donc une meilleure capacité d'exclusion, en ajoutant la classe VBF,

16. Une comparaison équitable n'est possible qu'en prenant en compte le produit efficacité fois acceptation.

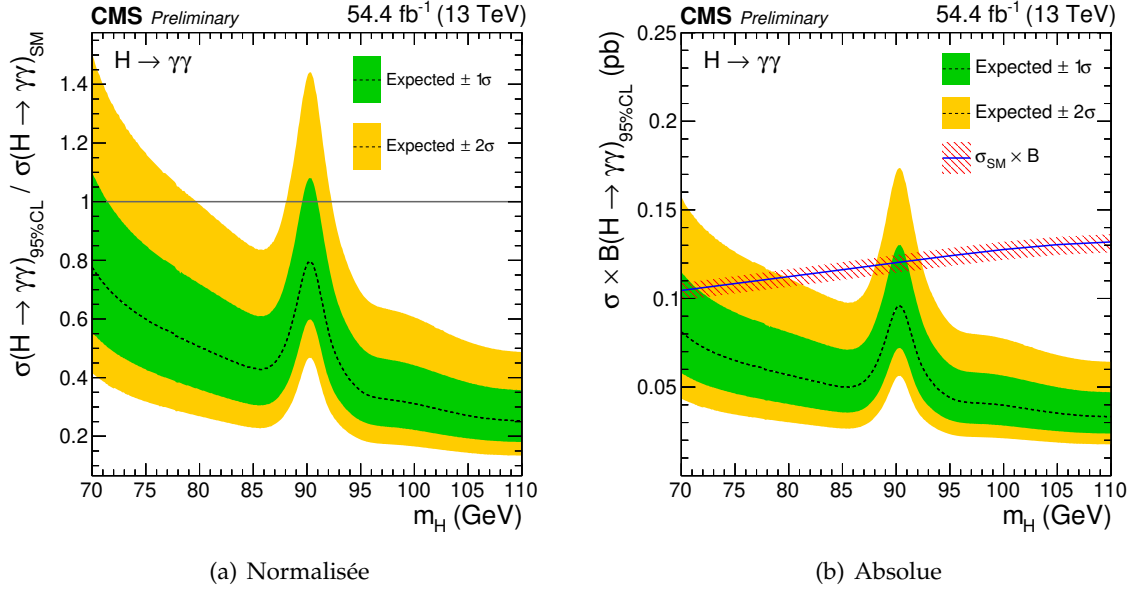


FIGURE 6.60 – Limite d’exclusion attendue à 95 % sur le produit de la section efficace de production par le rapport de branchement en deux photons d’un nouveau boson de Higgs pour 2018 dans le cas normalisé par sa valeur du MS a) et dans le cas absolu b) en fonction de la masse d’un tel boson, et bandes d’incertitude associées.

dans l’hypothèse d’une production de signal 100 % VBF justifiant ainsi l’ajout de cette classe.

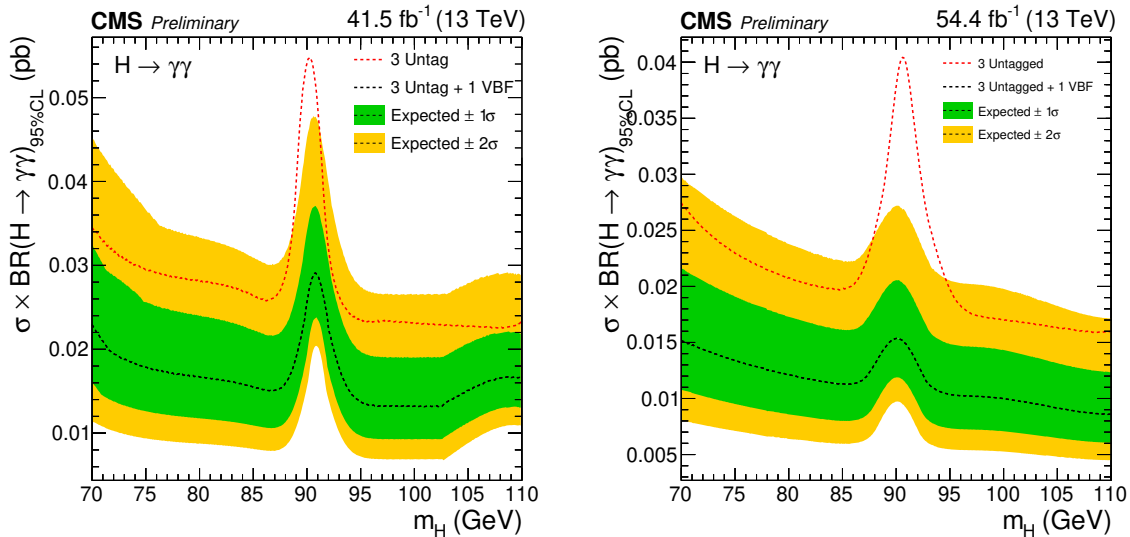


FIGURE 6.61 – Limite d’exclusion attendue à 95 % sur le produit de la section efficace de production par le rapport de branchement en deux photons d’un nouveau boson de Higgs dans l’hypothèse d’une production signal 100 % VBF en fonction de la masse d’un tel boson pour 2017 (gauche) et 2018 (droite). Le cas 3 classes + 1 classe VBF, et les incertitudes associées est en pointillé noirs, le cas 3 classes est visible en pointillé rouges.

De plus, les limites d’exclusion attendues dans l’hypothèse de mode de production

$ggH + ttH$  et  $VBF + VH$  sont visibles sur la figure 6.62 pour 2017 et 2018. Les deux modes de production, dans chaque cas, sont combinés selon leur fraction relative prédite dans le MS.

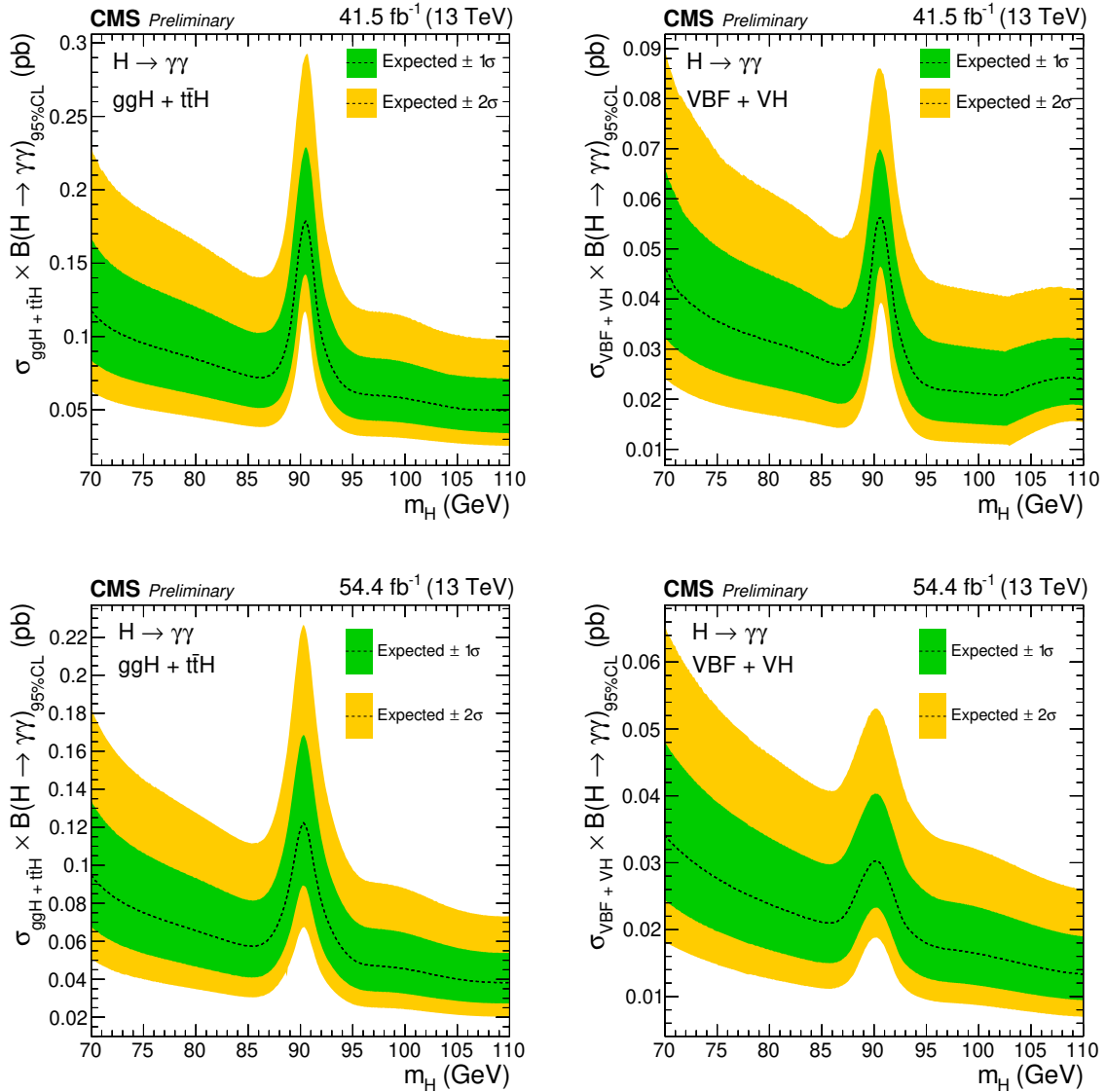


FIGURE 6.62 – Limite d'exclusion attendue à 95 % sur le produit de la section efficace de production par le rapport de branchement en deux photons d'un nouveau boson de Higgs dans l'hypothèse d'une production  $ggH + ttH$  (gauche) et  $VBF + VH$  (droite) en fonction de la masse d'un tel boson pour 2017 (haut) et 2018 (bas) et bandes d'incertitude associées.

Les résultats obtenus pour 2017 et 2018 peuvent ensuite être statistiquement combinés. On utilise les mêmes méthodes que celles décrites dans la section précédente, en considérant le produit des fonctions de vraisemblances et assumant que les incertitudes systématiques sont non corrélées entre les deux années. La figure 6.63 montre les limites d'exclusion alors obtenues dans le cas normalisé et non normalisé. Le minimum (maximum) de limite d'exclusion à 95 % est de 26 (87) fb correspondant à une masse hypothétique de 110,0 (90,4) GeV.

On combine ensuite les résultats obtenus pour le Run 1 (2012) à 8 TeV [121], ceux obtenus

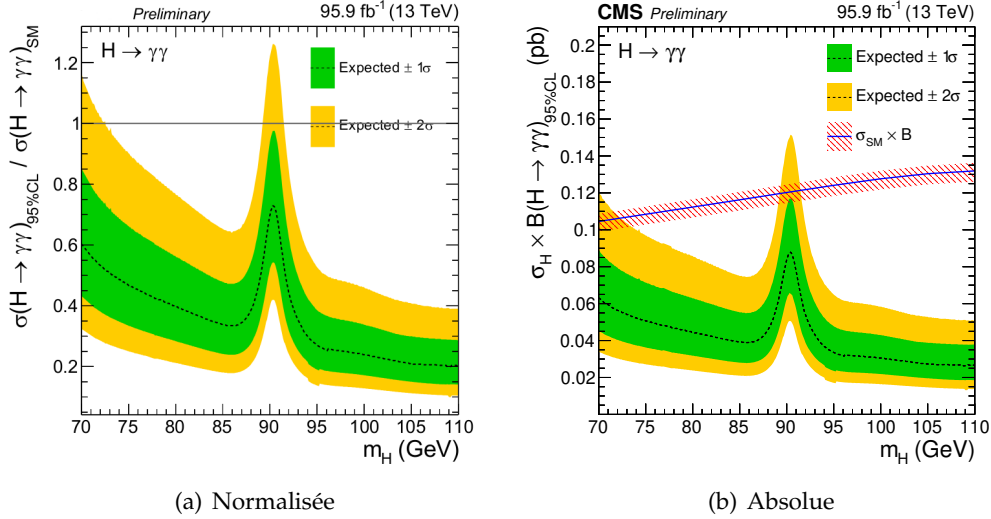


FIGURE 6.63 – Limite d’exclusion attendue à 95 % sur le produit de la section efficace de production par le rapport de branchement en deux photons d’un nouveau boson de Higgs pour la combinaison de 2017 et 2018 dans le cas normalisé par sa valeur SM a) et dans le cas absolu b) en fonction de la masse d’un tel boson et bandes d’incertitude associées.

en 2016 [103] et les résultats 2017 et 2018 obtenus ici dans une gamme de masse commune soit  $80 < m_{\gamma\gamma} < 110$  GeV. Les incertitudes systématiques expérimentales ainsi que la plupart des incertitudes théoriques sont supposées non corrélées mais les incertitudes théoriques dues à la variation de l’échelle QCD et aux calculs des sections efficaces à des énergie de centre de masse de 8 et 13 TeV sont supposées pleinement corrélées.

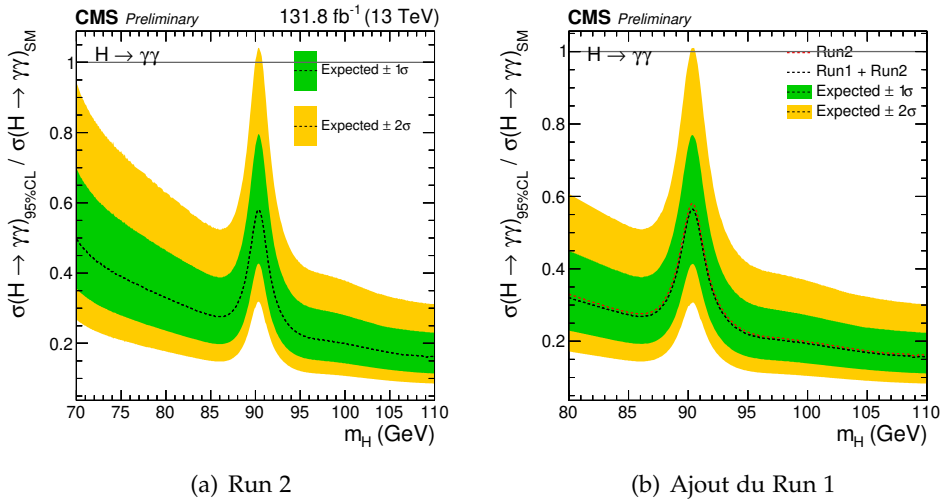


FIGURE 6.64 – Limite d’exclusion attendue à 95 % , et bandes d’incertitude associées, sur le produit de la section efficace de production par le rapport de branchement en deux photons d’un nouveau boson de Higgs normalisé par sa valeur du MS dans le cas de la combinaison de tout le Run 2 a) et avec l’ajout du Run 1 ainsi qu’une comparaison de résultats obtenus grâce à ces combinaison b).

La figure 6.64 montre les limites d'exclusion sur le produit de la section efficace fois rapport de branchement en deux photons pour un boson de Higgs additionnel normalisé pour la combinaison de tout le Run 2 et du Run 1 + Run 2. Le minimum (maximum) de limite d'exclusion à 95 % normalisée pour la combinaison de Run 2 est de 0,16 (0,58) correspondant à une masse hypothétique de 110,0 (90,3) GeV. L'ajout du Run 1 à la combinaison totale n'a qu'un faible impact sur ces limites. Cette combinaison permet d'obtenir des limites d'exclusion qui sont plus de 40 % plus sensibles que celles obtenues dans [103].

Les p-values attendues dans toutes les configurations présentées ici sont visibles sur la figure 6.65. On peut observer que la signification est supérieure à  $5\sigma$  sur la majeure partie de la gamme de masse sauf aux alentours directs du pic de Z, on pourrait donc statuer quant à la découverte d'une nouvelle particule dans ces gammes de masse si la force du signal correspondante est au moins égale à 1.

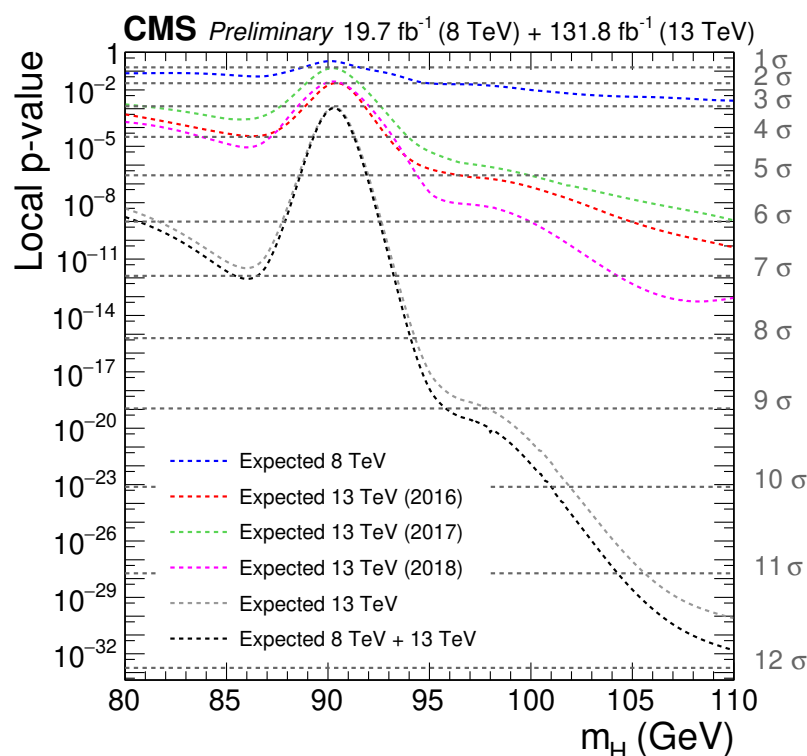


FIGURE 6.65 – P-value locale attendue en fonction de la masse d'un nouveau boson de Higgs pour chaque année du Run 1 et 2 ainsi que leurs combinaisons. La signification correspondante en terme de déviations standards d'une loi gaussienne est donnée en ordonnée.

### 6.12.3 Comparaison avec ATLAS

Une comparaison sommaire avec les résultats obtenus par la collaboration ATLAS [122] en utilisant les données 2016+2017 correspondant à une luminosité intégrée de 80 fb<sup>-1</sup> est présentée ici.

Les limites d'exclusion à 95 % obtenues pour un produit de section efficace fiducielle fois rapport de branchement en deux photons  $\sigma_{fid} \cdot \mathcal{B}$  sont visibles sur la figure 6.66 pour une



masse invariante diphoton allant de 65 à 110 GeV. Ces limites vont de 30 à 101 fb environ et aucun excès significatif par rapport au MS n'a été observé.

Afin de permettre une comparaison équitable, il faut d'abord estimer les résultats d'ATLAS dans l'espace de paramètre total ; pour cela, on divise leurs limites  $\sigma_{fid} \cdot \mathcal{B}$  par leur acceptation  $A_X$ , aussi visible sur la figure 6.66. Ce faisant, leurs limites attendues sont d'environ 185 fb à 70 GeV et 73 fb à 110 GeV. Ainsi juste en prenant les résultats [103] à  $35,9 \text{ fb}^{-1}$  nos résultats sont meilleurs que les derniers résultats publics d'ATLAS ; en effet, on obtient des limites attendues de 93/102/82 fb pour 2016/2017/2018 à 70 GeV et 38/39/34 fb à 110 GeV pour chaque année prise séparément. Ainsi, en combinant les années 2016+2017 pour avoir une luminosité intégrée proche de celle obtenue par ATLAS, on obtient des limites attendues de 65 fb à 70 GeV et 25 fb à 110 GeV soient des limites d'exclusion attendues bien meilleures d'un facteur proche de 3.

Ces différences dans les résultats peuvent être notamment expliquées par les différences dans la stratégie d'analyse utilisée et optimisation associée. ATLAS n'emploie en effet pas de BDT pour catégoriser les événements ni même de BDT pour identifier les photons ; ATLAS utilise à la place des catégories selon l'état converti ou non du photon (soit 3 catégories différentes pour les événements diphoton) qui sont semblables à celles qu'ATLAS utilisait pour le Run 1. Leur méthode d'estimation des incertitudes liées à la modélisation du bruit de fond est aussi très conservatrice [122].

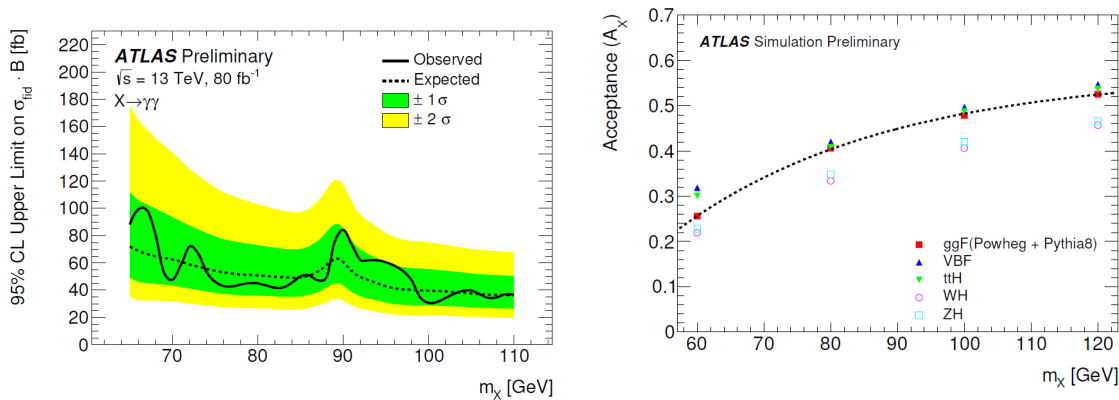


FIGURE 6.66 – Résultats d'ATLAS [122] : à gauche limite observée (ligne noire) et limite attendue (ligne noire en pointillé) sur la section efficace fiducielle fois rapport de branchement en deux photons en fonction de la masse du signal  $m_X$ , et bandes d'incertitudes associées. À droite courbe de l'acceptance dans le volume fiduciel pour différents mode de production du signal en fonction de  $m_X$ , la courbe en pointillé représente la paramétrisation de l'acceptance du mode ggH.

## Conclusion

L'analyse pour la recherche d'un boson de Higgs additionnel se désintégrant en deux photons utilisant les données 2017 et 2018 correspondant à des collisions proton-proton à une énergie de 13 TeV dans CMS a donc été présentée. La grande majorité des étapes de l'analyse a



été explicitée et les résultats attendus sur les limites d'exclusion sur la production de la section efficace d'un nouveau boson de Higgs fois rapport de branchement en deux photons ainsi que sur les p-values en absence de signal ont aussi été présentés avec ces mêmes jeux de données ainsi que leur combinaison avec les données 2012 et 2016.

Il est cependant crucial, afin de statuer sur une éventuelle découverte ou même de contraindre les modèles théoriques, de pouvoir étudier les résultats "Observés" obtenus directement sur les données. J'ai d'ailleurs aussi travaillé à l'obtention et la compréhension de ces résultats qui ne sont cependant pas publics à l'heure où ces lignes sont écrites.

Malgré les excellentes performances montrées dans ce chapitre, il existe des pistes d'amélioration, d'optimisation et quelques points de l'analyse peuvent être discutés. Ces perspectives, discussions et quelques autres études réalisées sont présentées dans le chapitre suivant.



# Études et perspectives d'amélioration

## Sommaire

---

7.1	Relâchement des coupures sur la gamme de masse . . . . .	196
7.1.1	Extension de la gamme de masse . . . . .	196
7.1.2	Relâchement des coupures d'impulsion transverse normalisée . . . . .	198
7.2	Réoptimisation des catégories <i>diphoton BDT</i> pour 2018 . . . . .	203
7.3	Études de la modélisation de la composante Drell-Yan . . . . .	210
7.3.1	Ajout de la composante exponentielle . . . . .	210
7.3.2	Accords données/simulations pour les évènements Drell-Yan <i>double-fake</i>	212
	Composante Drell-Yan non fixée dans la méthode de l'enveloppe . . . . .	212
	Incertitudes sur les paramètres de queues de la dCB . . . . .	221
	Choix de la fonction d'ajustement de la composante Drell-Yan . . . . .	228
7.3.3	Régions d'inefficacité en $\phi$ . . . . .	230

---

On présente dans cette partie une liste non exhaustive de différentes études, pistes d'amélioration et critiques de l'analyse  $h \rightarrow \gamma\gamma$  présentée précédemment. On commencera par une brève description de l'effet apporté par un relâchement des coupures en masse et en impulsion transverse dans l'analyse. On détaillera ensuite le résultat de l'optimisation des catégories *diphoton BDT* pour 2018, avant d'étudier en profondeur la façon dont on modélise la composante Drell-Yan *double-fake*. Ces différentes études ont été effectuées à des étapes différentes de l'avancement de l'analyse et les résultats obtenus ne sont pas directement comparables aux résultats finaux présentés ci-avant ; ils sont cependant tous dérivés à partir du cas où seules les trois classes *diphoton BDT* sont utilisées. Toutes les études présentées dans ce chapitre sont le fruit de mon travail.

## 7.1 Relâchement des coupures sur la gamme de masse

### 7.1.1 Extension de la gamme de masse

On a vu précédemment que la gamme de masse de l'analyse est définie entre 65 et 120 GeV, contrainte dans la partie inférieure par la présence de seuils en énergie transverse appliqués (et en masse pour les données de 2017) à chacun des photons du chemin de déclenchement et dans la partie supérieure par la présence du boson de Higgs du MS. Les résultats finaux sont cependant calculés pour des masses entre 70 et 110 GeV, en accord avec le domaine de masse initial utilisé pour la modélisation du signal [103], ce qui permet notamment de garder des régions de données pour les études de validation. On peut alors légitimement étudier la possibilité d'étendre la gamme de masse.

On peut commencer par essayer d'étendre la gamme de masse utilisée pour les résultats finaux à une gamme supérieure à l'intervalle [70,110] GeV. Grâce aux modifications que j'ai apportées à la modélisation du signal par rapport à l'analyse utilisant les données de 2016 [103], le modèle signal est maintenant défini jusqu'à 115 GeV ce qui nous permet d'étendre la gamme de masse jusqu'à cette valeur. La figure 7.1 montre les limites sur le produit section efficace fois rapport de branchement  $h \rightarrow \gamma\gamma$  normalisé et la p-value locale obtenues dans le cas attendu. On y observe quelques légères discontinuités, liées à des changements de meilleures fonctions d'ajustement au sein de l'enveloppe en fonction de la masse, et quelques ondulations, caractéristiques du polynôme Bernstein d'ordre 4 en tant que meilleure fonction d'ajustement, restant cependant acceptables. On peut donc raisonnablement augmenter la gamme de masse jusqu'à 115 GeV.

On peut ensuite se demander ce qu'il se passe si on réalise une extrapolation du signal pour des masses inférieures à 70 GeV et supérieures à 115 GeV. La figure 7.2 permet de partiellement répondre à cette question puisqu'elle montre les limites sur le produit section efficace fois rapport de branchement  $h \rightarrow \gamma\gamma$  normalisé pour des masses allant jusqu'à 118 GeV. Outre les remarques précédentes, on observe une forte dégradation de la sensibilité pour des masses supérieures à 115 GeV imputable à des effets de bord<sup>1</sup> sur la modélisation du signal. Ce

1. On se trouve dans la queue de la distribution du signal pour des masses  $\sim 118$  GeV et il y a donc un déficit

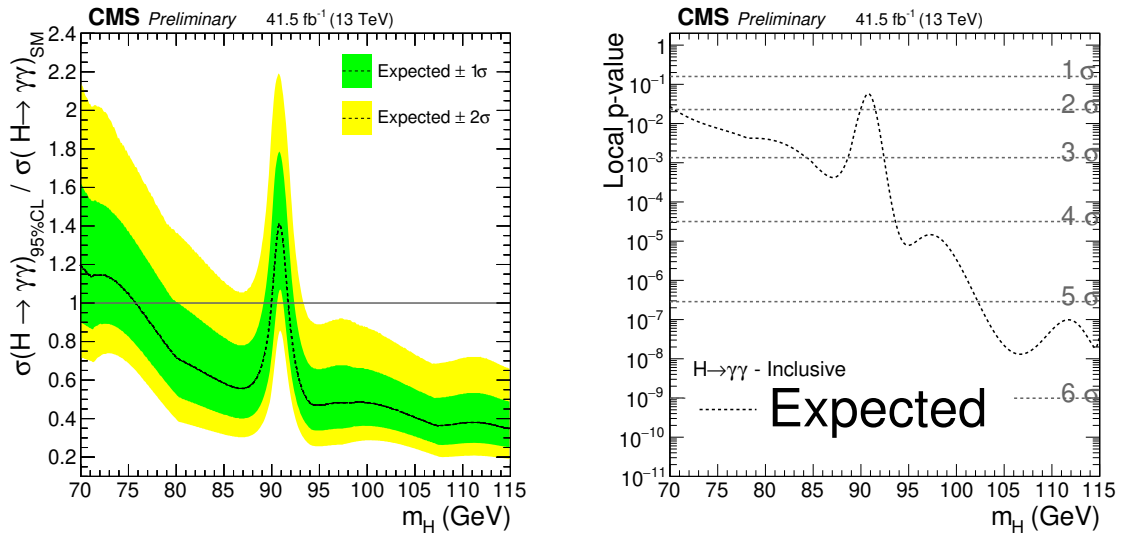


FIGURE 7.1 – Limite d’exclusion attendue sur le produit section efficace fois rapport de branchement en deux photons normalisé pour un boson de Higgs additionnel (gauche) et p-value locale attendue (droite) en fonction de la masse d’un tel boson jusqu’à 115 GeV pour 2017.

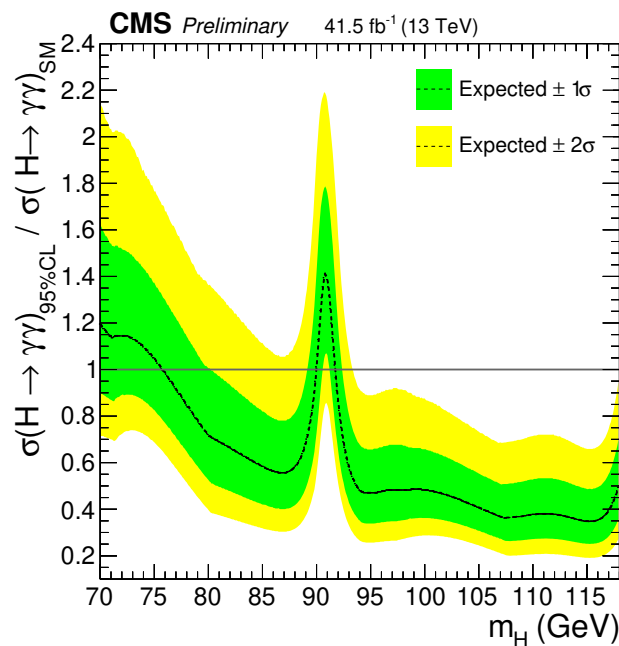


FIGURE 7.2 – Limite d’exclusion attendue sur le produit section efficace fois rapport de branchement en deux photons normalisé pour un boson de Higgs additionnel en fonction de la masse d’un tel boson définie de 70 à 118 GeV pour 2017.

problème peut être simplement résolu en ajoutant un échantillon de signal de masse 120 GeV à notre méthode de modélisation du signal et en interpolant jusqu’à cette masse, ou par une simple extrapolation du modèle de signal; ceci nécessite cependant de repousser la limite supérieure en masse de l’analyse au-delà de 120 GeV, au moins pour la modélisation du signal.

d’événements signal par rapport aux autres points de masse.

On peut de la même manière descendre à des masses de 67-68 GeV. Une option nettement plus intéressante pour sonder des gammes de masses inférieures à 70 GeV est de tirer parti de l'absence de coupure en masse dans le chemin de déclenchement pour 2018. Des travaux sont actuellement en cours pour étudier la faisabilité d'explorer des gammes de masses plus faibles. Les résultats sont encourageants et il semblerait qu'il soit possible de descendre jusqu'à des masses d'environ 50 GeV et potentiellement plus faibles. Ces études permettant de scanner des domaines de masse plus faibles seront implémentées dans l'analyse pour le Run 3 du LHC.

### 7.1.2 Relâchement des coupures d'impulsion transverse normalisée

Une autre façon d'augmenter la gamme de masse de l'analyse est de modifier la valeur des coupures en impulsion transverse normalisée  $p_T^{(1,2)}/m_{\gamma\gamma}$ . En effet ces coupures sculptent la distribution de masse aux alentours de la zone de *turn-on* : on peut observer sur la figure 7.3 que la distribution de masse invariante diphoton est plus piquée vers des masses faibles lorsque l'on augmente la valeur des coupures en impulsion transverse normalisée (figure gauche) permettant alors de commencer l'ajustement du bruit de fond correspondant au continuum diphoton à des masses légèrement plus faibles.

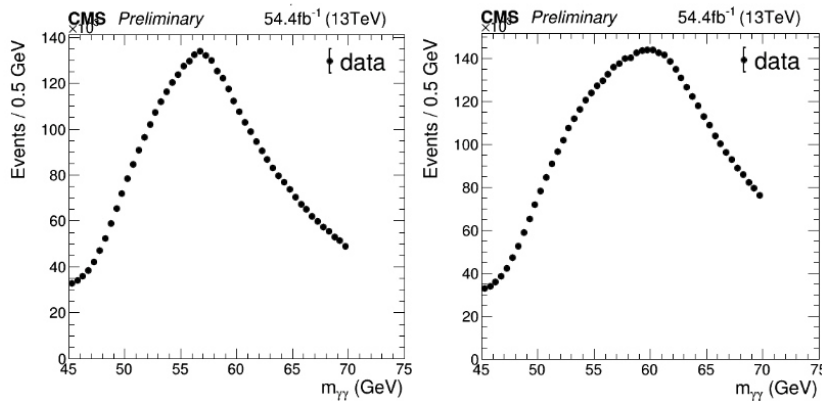


FIGURE 7.3 – Distribution de masse invariante diphoton pour les données 2018 pour des valeurs de coupures en impulsion transverse normalisée  $p_T^{(1,2)}/m_{\gamma\gamma}=(0,53;0,32)$  (gauche) et  $p_T^{(1,2)}/m_{\gamma\gamma}=(0,49;0,30)$  (droite).

L'inconvénient majeur d'une telle augmentation des coupures en impulsion transverse normalisée est qu'elle résulte en une diminution de l'efficacité du signal, et cela même pour des échantillons de signal ayant des masses éloignées de celle du *turn-on*. On observe ainsi sur la figure 7.4 qu'une partie non négligeable des événements signal, pour un mode de production ggH, est située à la frontière des coupures en impulsion transverse normalisée actuelles pour des masses de 70 GeV mais aussi de 110 GeV : en cas d'une augmentation de ces coupures ces événements seraient rejetés ce qui n'est pas souhaitable. Mieux, il pourrait être intéressant d'étudier si un relâchement de ces coupures ne pourrait pas permettre d'obtenir une meilleure sensibilité.

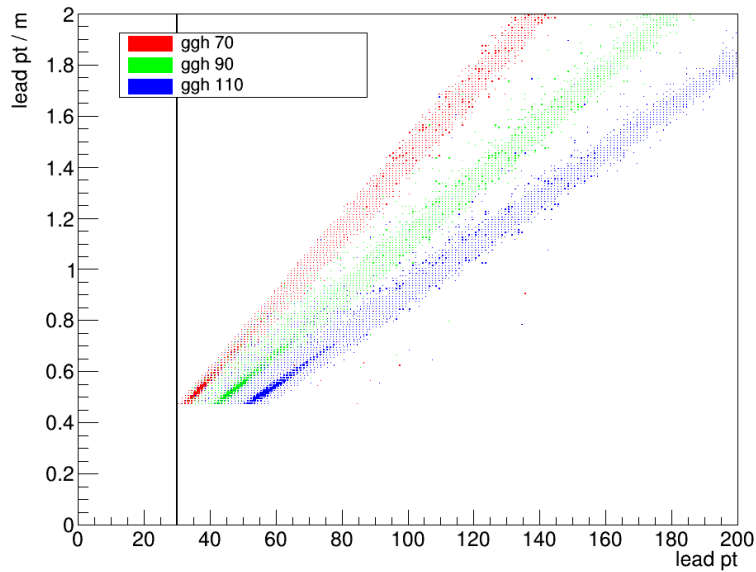


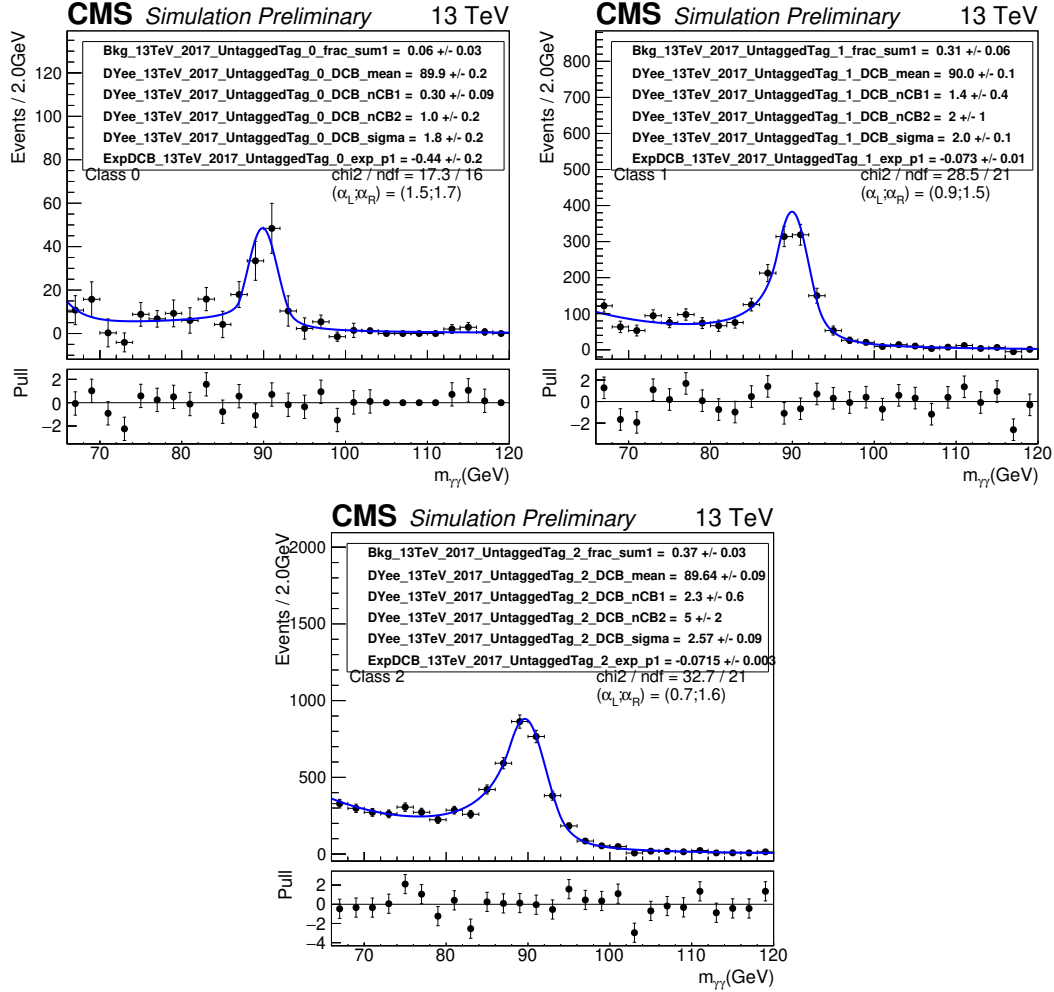
FIGURE 7.4 – Distribution du nombre d'évènements signal ggH pour des simulations 2018 à différentes masses en fonction de l'impulsion transverse du photon *leading* normalisé par la masse invariante diphoton et en fonction de l'impulsion transverse du photon *leading*.

Les coupures en l'impulsion transverse normalisée ayant été instituées, lors du Run 1, pour éviter que le bruit de fond soit modélisé par des fonctions d'ordre trop élevé, on cherche à savoir si l'on arrive à garder la forme actuelle du bruit de fond en relâchant ces coupures afin d'augmenter l'efficacité du signal et ainsi la sensibilité globale de l'analyse. L'efficacité du signal étant surtout dépendante de la coupure en l'impulsion transverse normalisée du photon *leading*, on relâche totalement cette coupure dans la suite de l'étude (ainsi l'on applique uniquement la coupure  $p_T^{(2)}/m_{\gamma\gamma} > 0,28$  usuelle). On pourrait alors régler finement la coupure  $p_T^{(2)}/m_{\gamma\gamma}$  restante pour avoir accès à des valeurs de masse plus faibles (potentiellement jusqu'à 60 GeV).

Pour ce faire, on reproduit les modélisations du signal et du bruit de fond pour l'analyse 2017 sans la coupure sur  $p_T^{(1)}/m_{\gamma\gamma}$ . Notons que les BDT de l'analyse ne sont pas spécialement réentraînés pour cette étude; notamment le *diphoton BDT* n'est pas réentraîné alors que les impulsions transverses normalisées font partie de ses variables d'entrée<sup>2</sup>. Ainsi la répartition des évènements supplémentaires, arrivant suite au relâchement total de la coupure, dans les classes *diphoton BDT* n'est pas forcément optimale. On observe dans le tableau 7.1 qu'il n'y a aucun évènement supplémentaire pour la classe 0 sans coupure ce qui est tout de même rassurant : la résolution en énergie relative étant meilleure à haute énergie, on s'attend à ce que les évènements avec une meilleure résolution en masse, c'est-à-dire les évènements de classe 0, soient situés dans des domaines de haute énergie donc peu affectés par la coupure. Il y a par contre environ 24 % plus d'évènements dans la classe 2.

2. Les coupures en impulsion transverse normalisée sont donc appliquées pour l'entraînement utilisé.

Nombre d'évènements dans les données		
	Sans coupure $p_T^{(1)}/m_{\gamma\gamma}$	Avec coupure usuelle
Classe 0	22943	22943
Classe 1	274135	260409
Classe 2	799303	645617

 TABLE 7.1 – Tableau du nombre d'évènements par classe *diphoton* BDT avec ou sans coupure  $p_T^{(1)}/m_{\gamma\gamma}$  pour les données 2017

 FIGURE 7.5 – Ajustement de la distribution de masse invariante diphoton par une fonction dCB+exp pour des simulations d'évènements Drell-Yan *double-fake* par classe et pour 2017 sans coupure sur  $p_T^{(1)}/m_{\gamma\gamma}$ .

Les figures 7.5 et 7.6 présentent respectivement la composante DY *double-fake* et l'enveloppe finale du bruit de fond obtenues sans coupure  $p_T^{(1)}/m_{\gamma\gamma}$ . Le contenu de l'enveloppe est différent de celui obtenu avec la coupure  $p_T^{(1)}/m_{\gamma\gamma}$  usuel (notamment avec l'apparition de  $Ber_3$ ) mais on n'observe pas de distorsion majeure dans la distribution du bruit de fond. Il faut néanmoins rappeler que l'on a choisi de se limiter à un nombre de paramètres inférieur à 5 (cf Eq. 6.10) pour les fonctions d'ajustement du bruit de fond<sup>3</sup>.

3. Dans le cas général, j'ai montré qu'autoriser les fonctions  $N < 5$  ou  $N < 7$  dans l'enveloppe n'amenait pas de



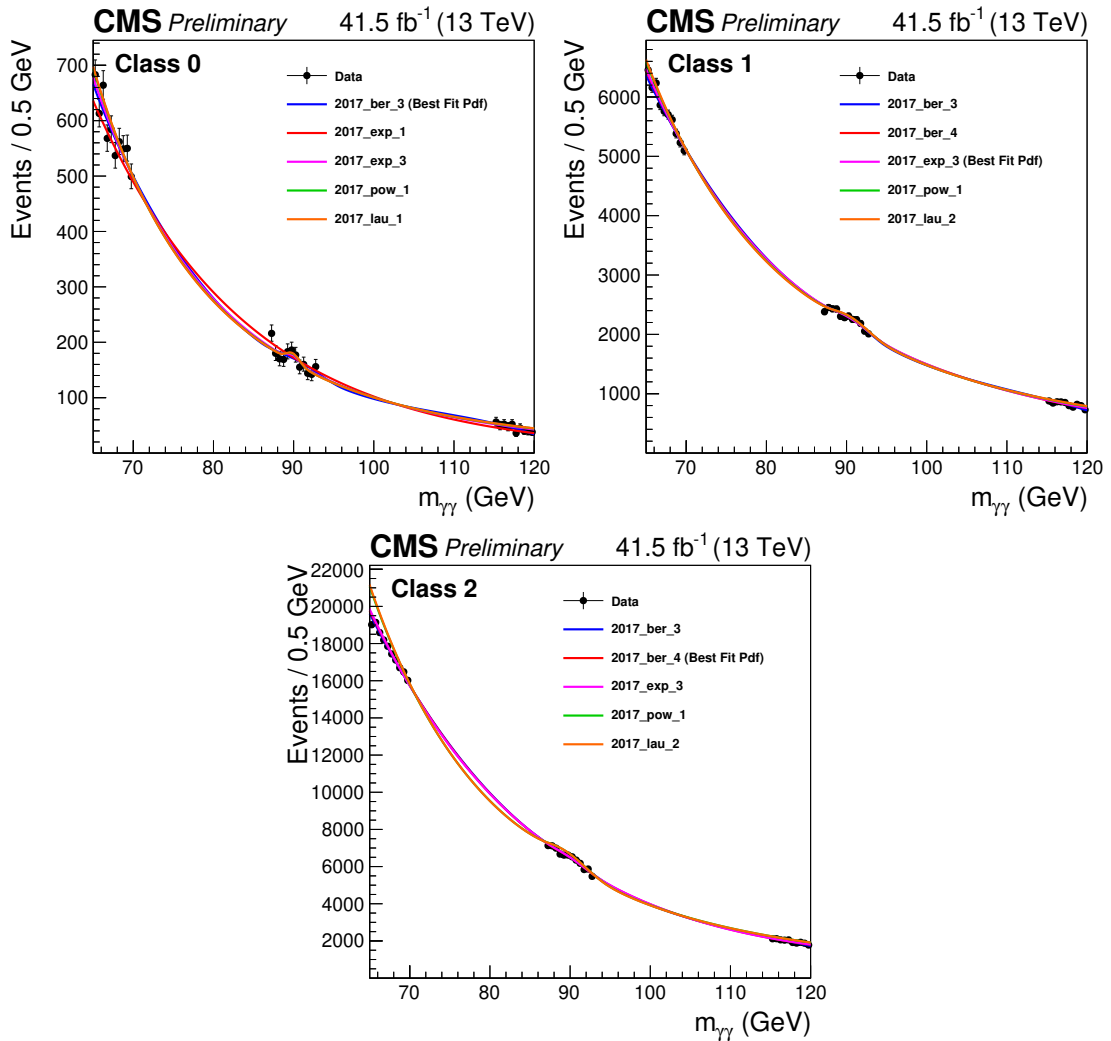


FIGURE 7.6 – Ajustement de la masse invariante diphoton des données (points noirs) par les fonctions composant l’enveloppe pour 2017 sans coupure sur  $p_T^{(1)}/m_{\gamma\gamma}$ .

Le nouveau modèle de signal obtenu est, lui, visible sur la figure 7.7. Comme attendu, on observe une augmentation du nombre d’évènements signal dans les classes 1 et 2, et dans les régions de masses plus élevées. Ceci est aussi visible sur la figure 7.8 qui présente l’efficacité fois acceptance obtenue sans la coupure  $p_T^{(1)}/m_{\gamma\gamma}$  puisque l’on voit que l’efficacité fois acceptance est, relativement, environ 4,6 % plus importante à 70 GeV et environ 10 % plus importante à 115 GeV dans le cas où la coupure n’est pas présente. Si cela peut paraître à priori contre-intuitif, la figure 7.4 montre que pour des signaux de basse masse, par exemple à 70 GeV, une bonne partie des évènements a déjà été rejetée par la coupure en impulsion transverse à 30 GeV ce qui n’est pas le cas pour une masse de 110 GeV, il y a donc plus d’évènements signal à récupérer à haute masse.

différence significative dans les limites d’exclusion attendues.

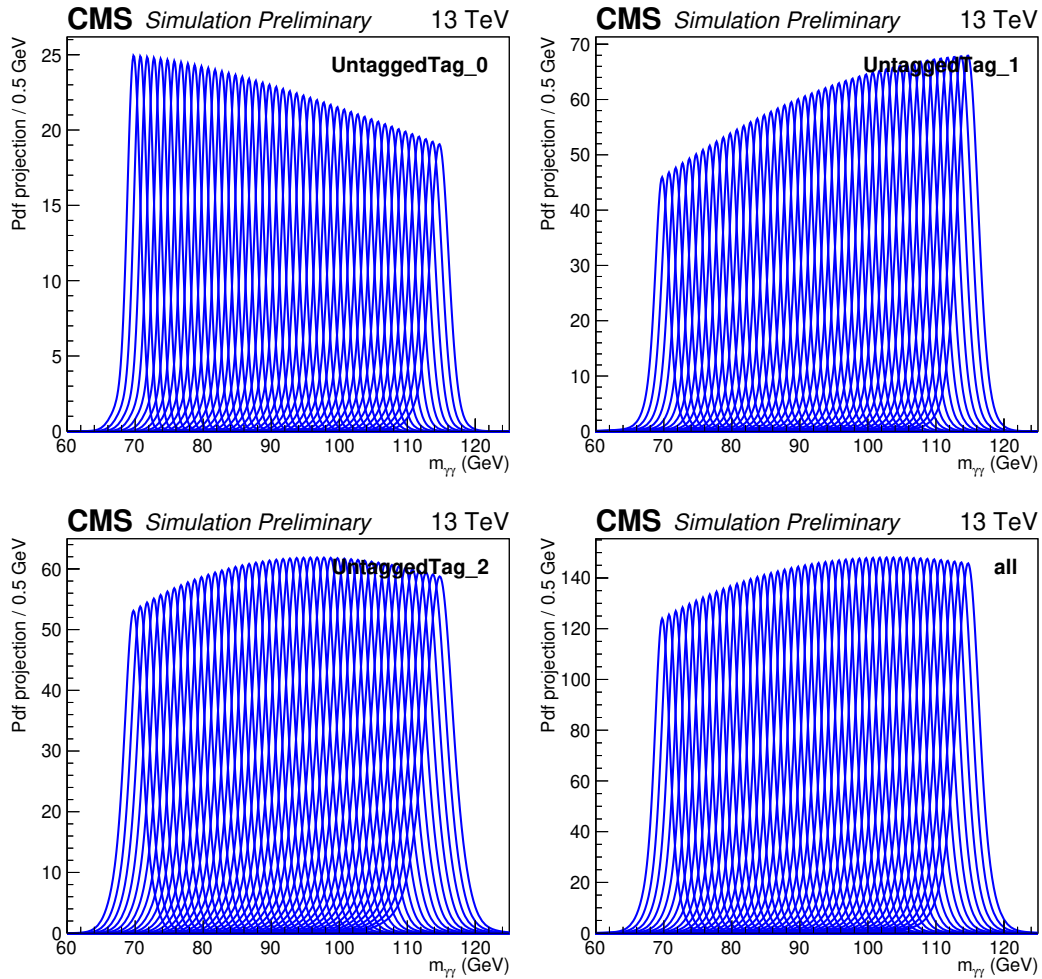


FIGURE 7.7 – Modèle paramétrique du signal 2017 illustré ici avec des intervalles de masse de 1 GeV par classe sans coupure sur  $p_T^{(1)}/m_{\gamma\gamma}$ .

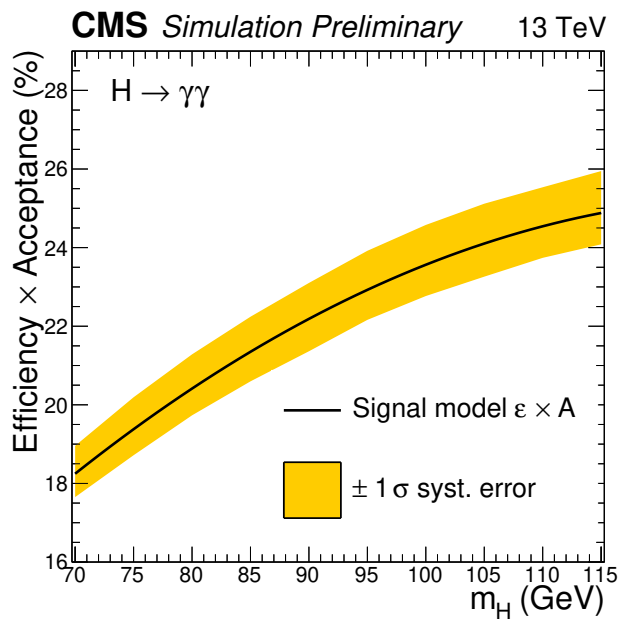


FIGURE 7.8 – Efficacité de l'analyse multipliée par l'acceptance du détecteur en fonction de la masse d'un boson de Higgs *SM-like* pour 2017 sans coupure sur  $p_T^{(1)}/m_{\gamma\gamma}$ .

Pour finir, les résultats finaux attendus obtenus sans coupure  $p_T^{(1)}/m_{\gamma\gamma}$  sont visibles sur la figure 7.9, on voit que les limites d'exclusion sur le produit section efficace fois rapport de branchement en deux photons pour un boson de Higgs additionnel normalisé sont meilleures, jusqu'à une dizaine de pourcent (relativement) à haute masse, lorsque cette coupure n'est pas présente. De même, la p-value attendue dans l'hypothèse bruit de fond seul est nettement meilleure, notamment à haute masse, et une signification de  $5\sigma$  est attendue, dans le cas d'un signal de force 1, à des masses inférieures à 98,2 GeV sans coupure alors que cette signification est atteinte pour des masses inférieures à 99,8 GeV dans le cas avec coupure. A noter que les incertitudes systématiques associées à la composante DY n'ont pas été recalculées ici.

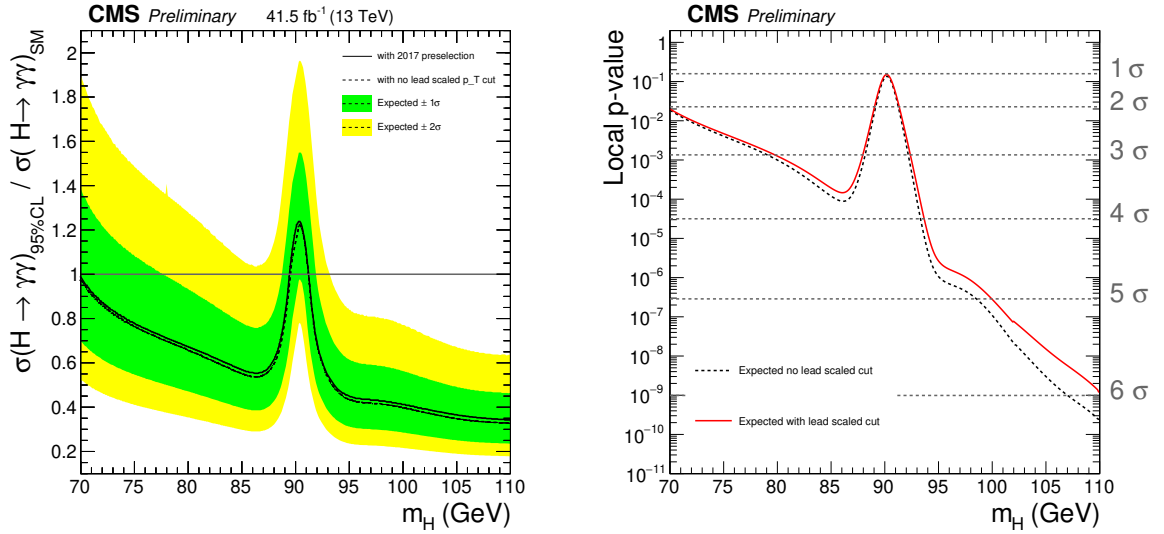


FIGURE 7.9 – Limites d'exclusion attendues sur le produit section efficace fois rapport de branchement en deux photons normalisé pour un boson de Higgs additionnel (gauche) et p-value locale attendue (droite) en fonction de la masse d'un tel boson jusqu'à 110 GeV pour 2017 sans coupure sur  $p_T^{(1)}/m_{\gamma\gamma}$ .

Le relâchement total de la coupure en impulsion transverse normalisée par la masse invariante diphoton pour le photon *leading* augmente donc l'acceptance du signal et permet d'obtenir une meilleure sensibilité attendue sans provoquer de distorsion apparente dans les distributions et résultats obtenus. Ces résultats peuvent potentiellement concerner toutes les analyses  $H \rightarrow \gamma\gamma$  impliquant de telles coupures et pourraient être une piste d'amélioration pour le Run 3. Il faut cependant être prudent puisque les distorsions mentionnées ci-dessus pourraient apparaître dans les limites et p-value observées et il faudrait, dans ce cas, réentraîner le *diphoton BDT*.

## 7.2 Réoptimisation des catégories *diphoton BDT* pour 2018

On présente ici la procédure de réoptimisation du nombre de catégories et frontières *diphoton BDT* associées pour les données de 2018 ainsi que la dérivation des résultats obtenus avec

ces nouvelles catégories. Les résultats obtenus ici peuvent directement être comparés à ceux présentés pour les données 2018 dans le chapitre précédent.

On commence à nouveau par déterminer la taille minimale de classe 0 qui doit contenir assez d'évènements MC pour obtenir des ajustements de la composante Drell-Yan *double-fake* acceptables. La figure 7.10 montre ainsi que pour une taille minimale de classe 0 de  $\sim 0,80$ , les paramètres de la fonction d'ajustement dCB sont bien définis et sur un plateau. On choisit donc cette valeur comme frontière basse minimale de notre classe 0.

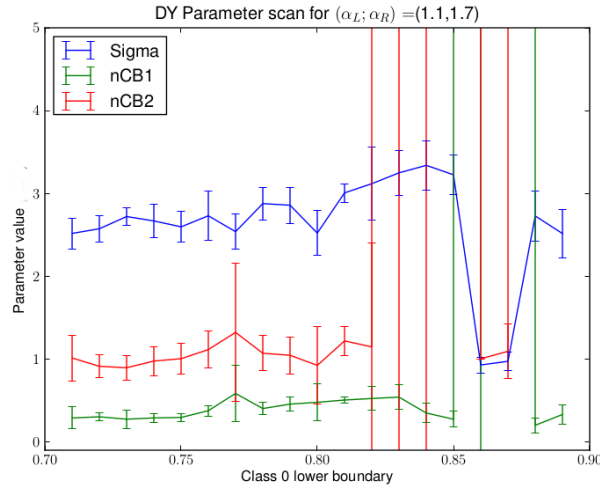


FIGURE 7.10 – Valeur de quelques uns des paramètres d'ajustement d'évènements DY  $Z \rightarrow ee$  simulés pour l'année 2018 en fonction de la frontière inférieure, en terme de score de sortie *diphoton BDT*, de la classe 0 qui contient les évènements les plus performants.

On détermine ensuite le nombre de catégories et les frontières associées par la même méthode que celle présentée dans la sous-section 6.7.2 : on calcule une p-value attendue en fonction du nombre de classes et en fonction de leurs frontières pour un simple modèle utilisant des simulations où le bruit de fond est paramétré par une exponentielle et le signal, de masse 90 GeV, par une somme de deux gaussiennes. Les résultats obtenus sont visibles sur la figure 7.11. A la vue des faibles différences entre les figures de mérite, on choisit à nouveau d'utiliser 3 classes ; les frontières associées sont alors  $[1,000; 0,792; 0,334; -0,466]$ . Ces nouvelles frontières sont proches de celles obtenues pour 2017, la coupure de rejet des évènements avec un mauvais score étant cependant plus lâche ici (on rejette donc moins d'évènements).

On reproduit alors les modèles de bruit de fond et signal avec ces nouvelles frontières. Les figures 7.12 et 7.13 présentent respectivement la composante DY *double-fake* et l'enveloppe finale du bruit de fond obtenues. On observe notamment que les valeurs de  $\chi^2/ndf$  obtenues pour ces nouveaux ajustements de la composante Drell-Yan sont moins bonnes que celles obtenues avec l'ancienne classification, visibles sur la figure 6.35 ; ceci s'explique par le changement de catégories et, surtout, par des valeurs différentes de  $\alpha_{R/L}$  fixés dans les ajustements DY. De la même manière, le contenu des enveloppes est légèrement différent par rapport à celui

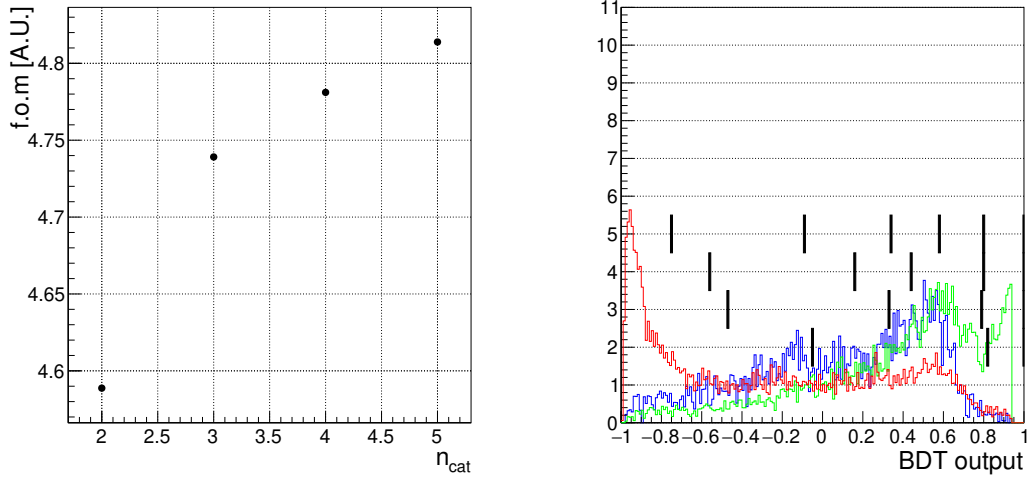


FIGURE 7.11 – Figure de mérite obtenue à la suite de l’optimisation des catégories de l’analyse basée sur des simulations 2018 en fonction du nombre de catégories  $n_{cat}$  (gauche). Les frontières associées (barres noires verticales) sont visibles sur la figure de droite avec les distributions des simulations signal dans l’hypothèse de vertex correct (vert) et de mauvais vertex (bleu) et bruit de fond (rouge) en fonction du score de sortie *diphoton* BDT.

obtenu pour 2018 avec la classification de 2017 visible sur la figure 6.37 : pour la classe 0 la fonction  $Ber_3$  est maintenant présente au sein de l’enveloppe et  $Ber_4$  est maintenant la meilleure fonction d’ajustement à la place de  $Lau_1$ , et pour la classe 2 la fonction  $Pow_1$  est maintenant acceptée dans l’enveloppe.

Le nouveau modèle de signal obtenu est, lui, visible sur la figure 7.14. Comme attendu, on observe une augmentation du nombre d’évènements signal dans les classes 1 et 2, ainsi qu’une diminution du nombre d’évènements dans la classe 0, comparé au modèle obtenu avec la classification de 2017 visible sur la figure 6.29. Étant donné que la coupure de rejet des évènements avec un mauvais score est plus lâche ici, l’acceptance du signal obtenue est un peu meilleure, comme visible sur la figure 7.15, avec une augmentation relative, par rapport à la figure 6.32 droite, de 3,2 % à 70 GeV et de 2,1 % à 115 GeV. Les incertitudes systématiques associées à la modélisation de la composante Drell-Yan sont aussi recalculées et sont visibles sur la tableau 7.16, elles sont comparables avec celles obtenues en 2018 avec la classification de 2017.

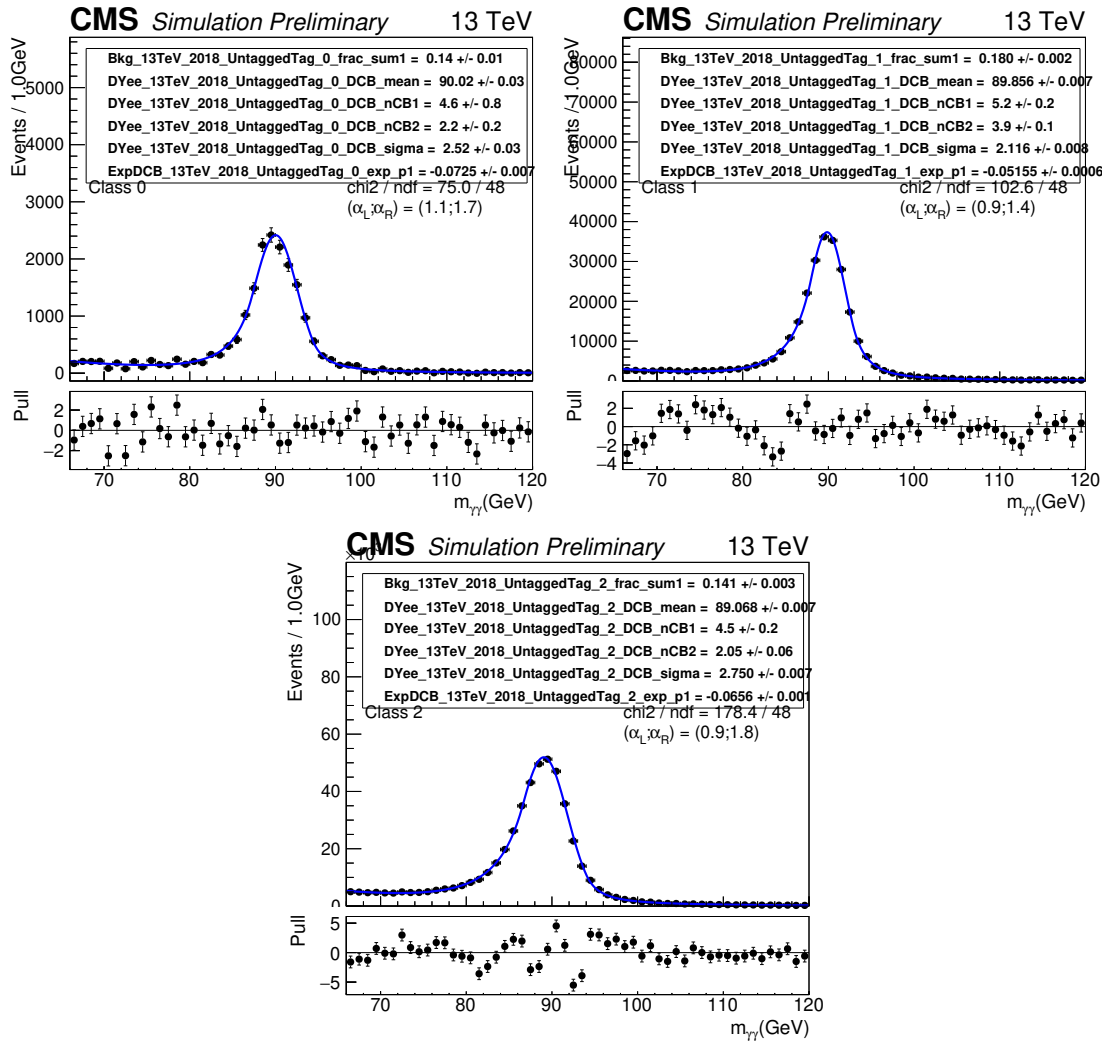


FIGURE 7.12 – Ajustement de la distribution de masse invariante diphoton par une fonction dCB+exp pour des simulations d'évènements Drell-Yan *double-fake* par classe pour la nouvelle classification 2018.

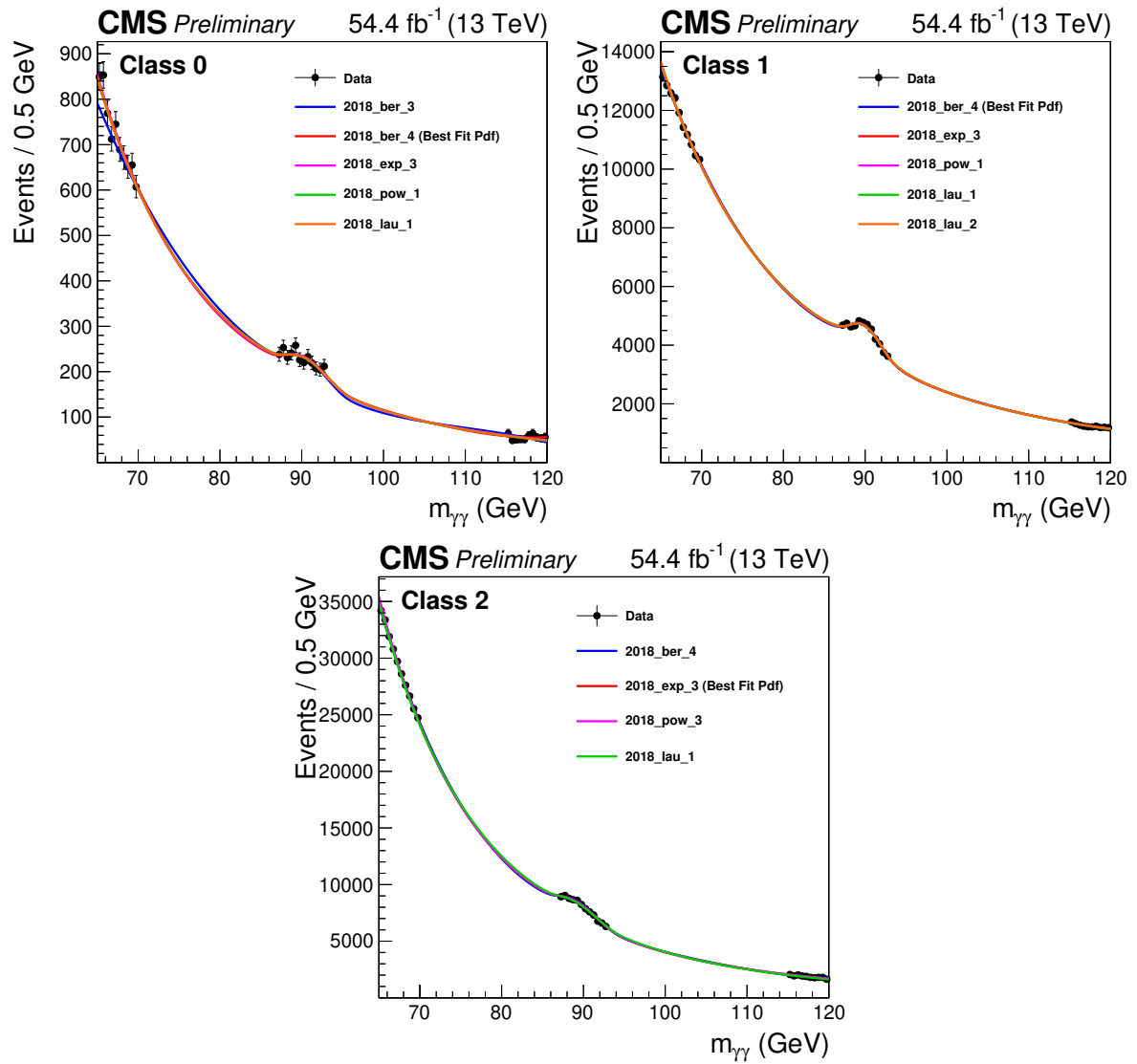


FIGURE 7.13 – Ajustement de la masse invariante diphoton des données (points noirs) par les fonctions composant l’enveloppe pour la nouvelle classification 2018.

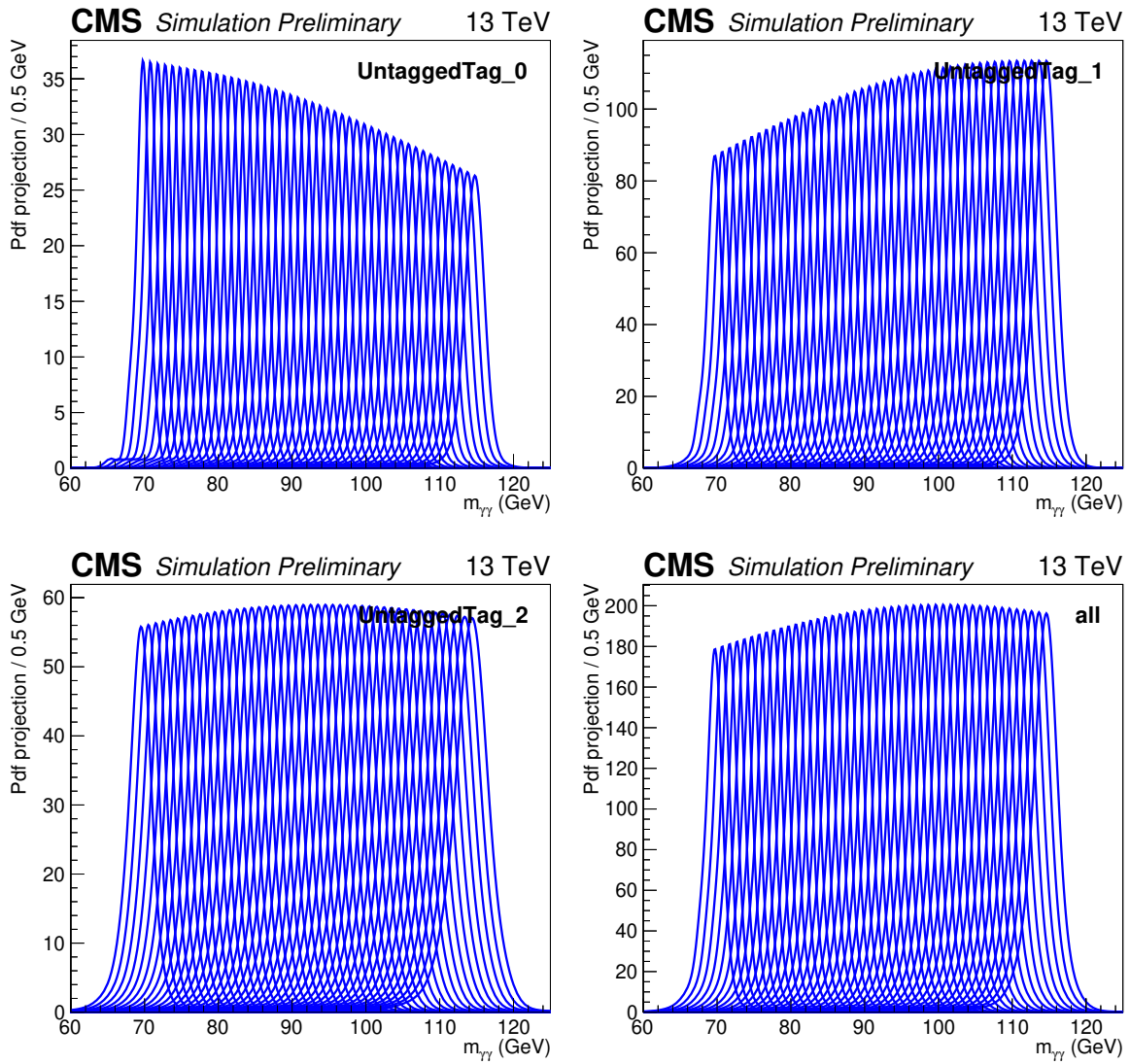


FIGURE 7.14 – Modèle paramétrique du signal 2018 illustré ici avec des intervalles de masse de 1 GeV par classe pour la nouvelle classification 2018.



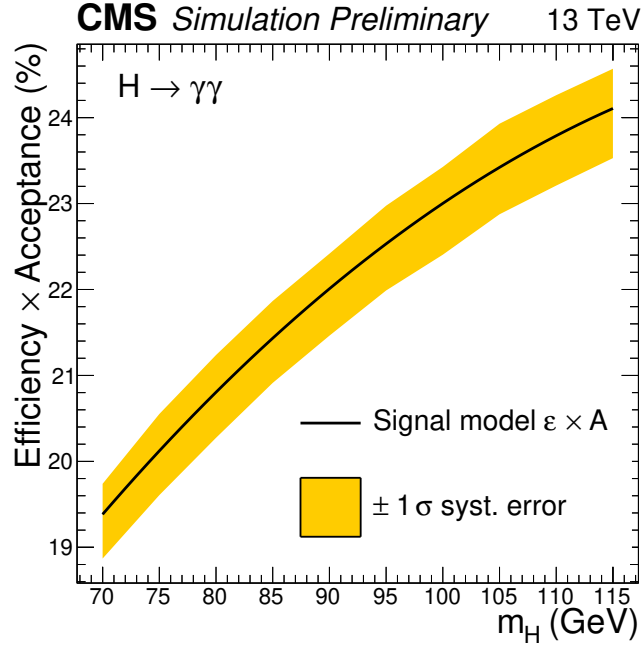


FIGURE 7.15 – Efficacité de l’analyse multipliée par l’acceptance du détecteur en fonction de la masse d’un boson de Higgs *SM-like* pour la nouvelle classification 2018.

FIGURE 7.16 – Incertitudes totales par classe sur les paramètres dCB  $\mu$  et  $\sigma$  utilisés dans l’ajustement de la masse des évènements *double-fake* décomposées en leurs différentes composantes pour la nouvelle classification 2018. "N.A." signifie que la différence dans les ajustements *single-fake* n’est pas statistiquement significative.

class	$\Delta\mu_{stat}$	$\Delta\mu_{data,MC_{All}}$	$\Delta\mu_{MC_{All},MC_{DY}}$	$\Delta\sigma_{stat}$	$\Delta\sigma_{data,MC_{All}}$	$\Delta\sigma_{MC_{All},MC_{DY}}$	$\Delta\mu_{tot}$	$\Delta\sigma_{tot}$
0	0.21	0.08	0.08	0.21	0.06	N.A.	0.31	0.24
1	0.06	0.14	0.11	0.06	0.06	0.08	0.36	0.21
2	0.06	0.05	N.A.	0.06	N.A.	0.03	0.12	0.08

Pour finir, les résultats finaux attendus obtenus avec cette nouvelle catégorisation sont visibles sur la figure 7.17, on voit que les limites d’exclusion sur le produit section efficace fois rapport de branchement en deux photons pour un boson de Higgs additionnel normalisé sont légèrement meilleures pour des masses inférieures à 86 GeV et supérieures à 103 GeV par rapport à la classification dérivée pour 2017. Pour la p-value attendue dans l’hypothèse bruit de fond seul on observe aussi un léger gain de sensibilité avec cette nouvelle catégorisation entre 70 et 76 GeV et entre 103 et 106 GeV ; de manière globale la p-value obtenue avec ces nouvelles classes n’est pas meilleure que celle obtenue précédemment. On constate aussi des instabilités sur ces résultats dues à la présence de la fonction  $Ber_4$  en tant que meilleure fonction d’ajustement dans l’enveloppe. Les résultats n’étant pas strictement meilleurs que ceux obtenus avec la catégorisation 2017, on préfère donc garder cette catégorisation 2017 pour 2018.

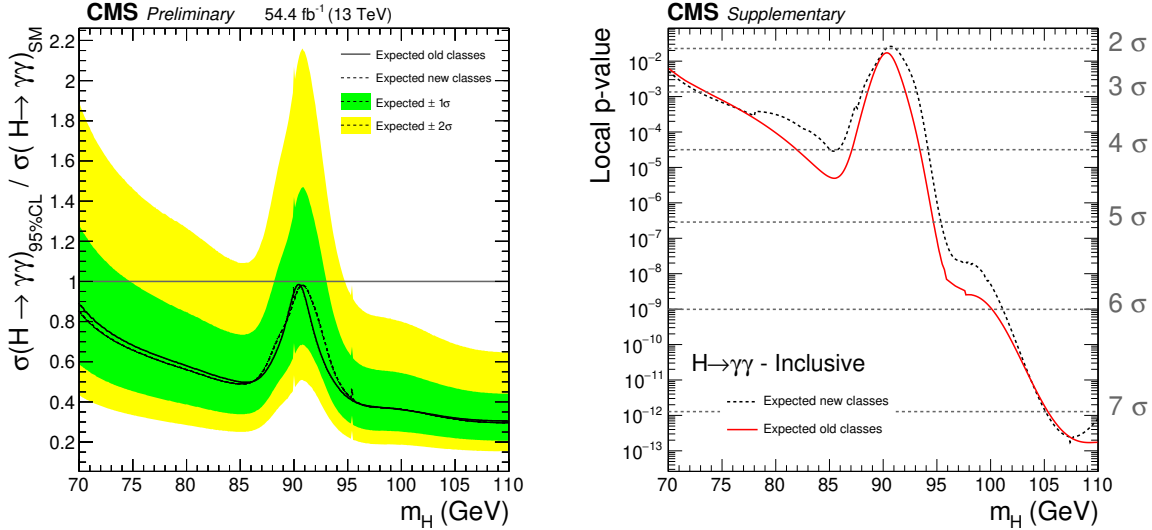


FIGURE 7.17 – Limites d'exclusion attendues sur le produit section efficace fois rapport de branchement en deux photons normalisé pour un boson de Higgs additionnel (gauche) et p-value locale attendue (droite) en fonction de la masse d'un tel boson pour la nouvelle classification 2018.

### 7.3 Études de la modélisation de la composante Drell-Yan

La composante Drell-Yan du bruit de fond, la façon dont elle est modélisée, est exclusive à l'analyse  $h \rightarrow \gamma\gamma$  à basses masses. Elle est un élément essentiel de l'analyse qu'il est nécessaire de comprendre en détail, d'autant plus que le léger excès d'événements observé dans les données 2012 et 2016 [103] se situe proche du pic du Z. Il est donc important d'étudier la façon dont cette composante est modélisée sur les simulations MC, et comment cela se propage à la modélisation du bruit de fond dans les données ainsi que dans les tests statistiques finaux.

#### 7.3.1 Ajout de la composante exponentielle

Les événements simulés DY *double-fake* contiennent une large contribution de bruit de fond qui n'est pas bien couvert par un ajustement avec une fonction dCB seule comme utilisé dans [103]. On ajoute donc une fonction exponentielle additionnelle avec deux nouveaux paramètres libres pour prendre en compte cette contribution supplémentaire.

Les ajustements obtenus sont visibles sur la figure 7.18. On observe que les figures de mérite obtenues sont meilleures lors de l'ajout de l'exponentielle, notamment pour les classes 1 et 2. Les valeurs de  $\mu$  et  $\sigma$  sont quasiment identiques (sauf pour le  $\sigma$  de classe 2) lors de l'ajout de cette fonction. Les enveloppes obtenues avec cette fonction sont visibles sur la figure 7.19. Le seul changement notable est le changement de meilleure fonction d'ajustement pour la classe 1 qui passe de  $\text{Ber}_4$  à  $\text{Exp}_3$ .

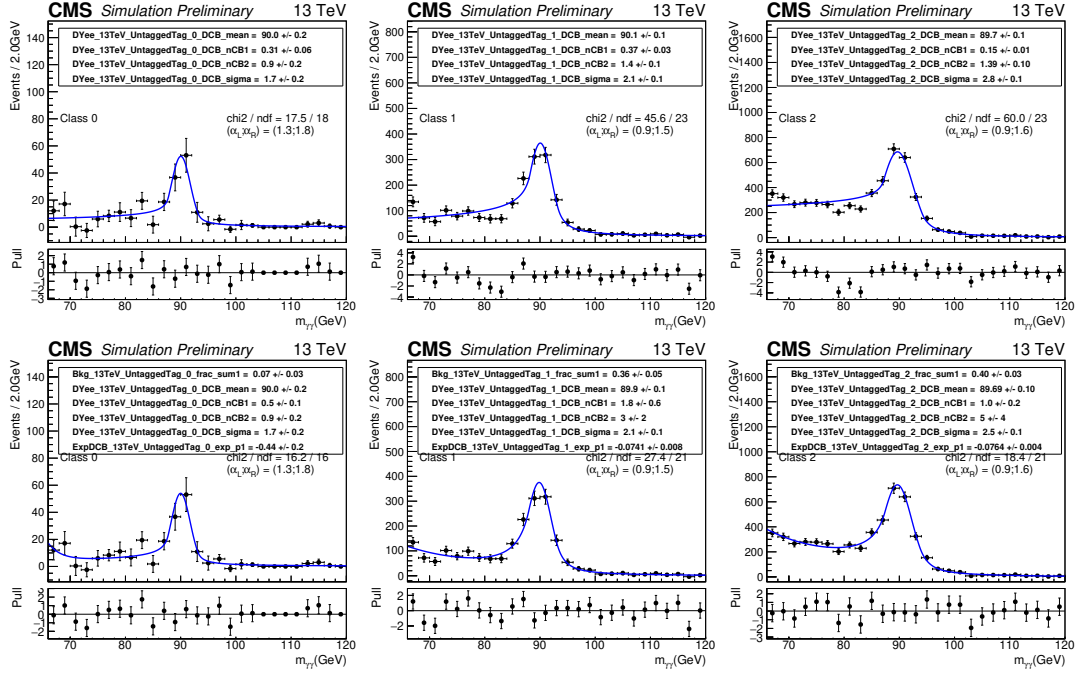


FIGURE 7.18 – Ajustement par une fonction dCB seul (haut) et dCB+exp (bas) des simulations *double-fake* DY pour 2017 par classe. A noter que le bin à 73 GeV de la classe 0 a un contenu négatif.

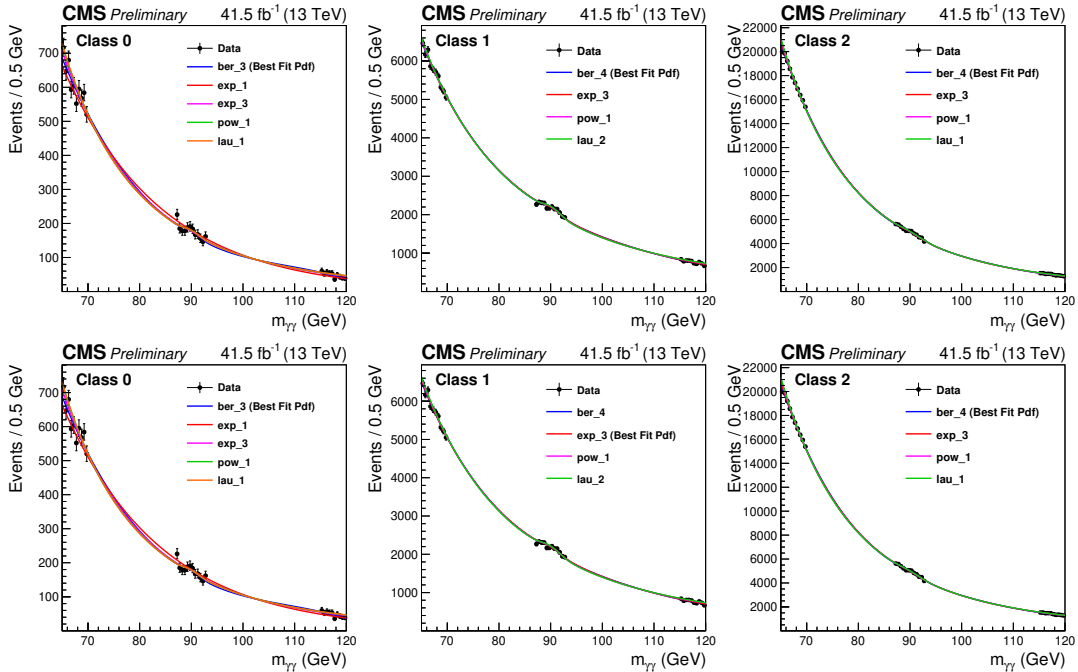


FIGURE 7.19 – Ajustement de la masse invariante diphoton des données (points noirs) par les fonctions composant l'enveloppe pour 2017 dans le cas où la composante DY est ajustée par une fonction dCB seule (haut) et dCB+exp (bas).

La figure 7.20 présente finalement les limites et p-value obtenues avec la fonction exponentielle dans la modélisation de la composante DY. Ces résultats sont plus lisses et légèrement meilleurs à hautes masses que ceux obtenus avec la fonction dCB seule. Ces résultats et les

meilleures figures de mérite obtenues précédemment justifient l'utilisation de cette fonction exponentielle supplémentaire pour les ajustements DY *double-fake* dans l'analyse. Cette étude est une contribution importante à l'analyse puisque la fonction dCB+exp est maintenant la fonction utilisée par défaut dans celle-ci.

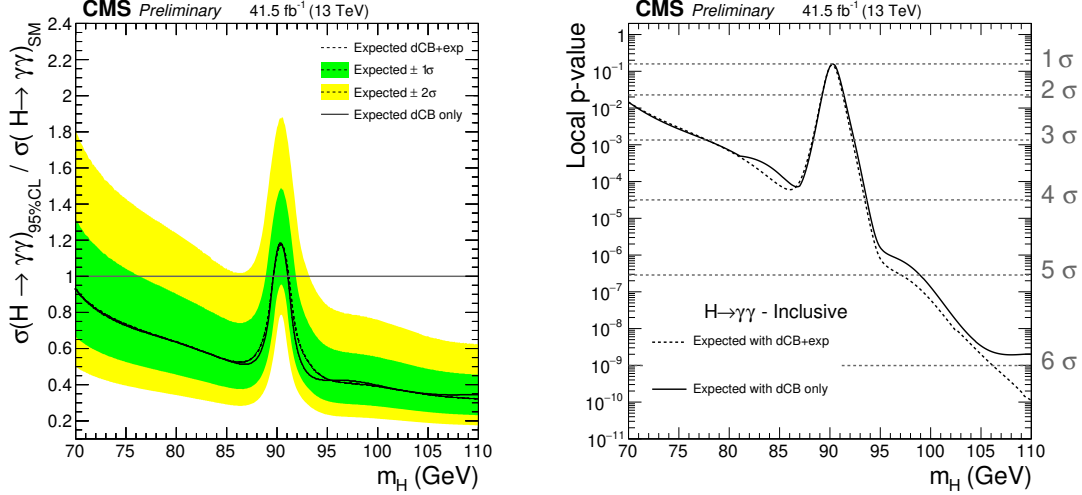


FIGURE 7.20 – Comparaison des limites d'exclusion attendues sur le produit section efficace fois rapport de branchement en deux photons normalisé pour un boson de Higgs additionnel (gauche) et p-value locale attendue (droite) en fonction de la masse d'un tel boson dans le cas d'un ajustement DY *double-fake* dCB seule et dCB+exp pour 2017.

### 7.3.2 Accords données/simulations pour les événements Drell-Yan *double-fake*

Actuellement, dans l'analyse, les paramètres des fonctions dCB+exp pour la composante Drell-Yan sont obtenus par des ajustements sur des événements DY *double-fake* simulés et les incertitudes systématiques censées prendre en compte les différences entre données et simulations sont estimées en utilisant des événements *single-fake* qui n'ont pas nécessairement le même comportement que les événements *double-fake*.

Il est alors possible que notre modélisation de la composante Drell-Yan ne permette pas une description tout à fait robuste des données, ce qui est d'autant plus problématique que le léger excès d'événements observé dans [103] se situe à une masse de  $\sim 95$  GeV soit proche du pic du Z.

On essaie donc ici de quantifier l'accord entre l'ajustement des simulations DY *double-fake* et la distribution en masse invariante diphoton correspondante dans les données. Cette étude est effectuée pour les données et simulations de 2018.

### Composante Drell-Yan non fixée dans la méthode de l'enveloppe

Dans la méthode de l'enveloppe, utilisée pour l'ajustement du continuum diphoton sur les données, les valeurs des paramètres de la fonction dCB+exp utilisée pour ajuster la composante Drell-Yan sont fixées aux valeurs centrales obtenus après ajustement sur les simulations *double-fake*. Cela permet de faciliter la convergence des différents ajustements mais cela ne

permet pas de rendre compte du comportement observé dans les données à cette étape. Étant donné que l'on souhaite ici étudier le comportement de cette composante au sein des données, on va donc laisser les paramètres de la fonction dCB+exp libres de flotter<sup>4</sup>, dans des gammes définies ci-après, pour la méthode de l'enveloppe. On va ensuite chercher à comparer la valeur des paramètres obtenue dans ce cas avec la valeur obtenue avec les simulations *double-fake*, ainsi que celle obtenue en sortie des analyses statistiques permettant d'obtenir les résultats finaux dans l'hypothèse bruit de fond seul.

A noter que dans les analyses statistiques finales, les valeurs de ces paramètres dCB+exp sont libres de flotter, avec une pénalité en terme de vraisemblance selon les écarts avec la valeur centrale, dans un domaine défini autour de  $\pm 4\sigma_{tot}$  des valeurs centrales obtenues via ajustement sur les simulations, où  $\sigma_{tot}$  est l'incertitude totale associée au paramètre considéré. On rappelle que pour les paramètres qui ne sont pas  $\mu$  ou  $\sigma$  de la fonction dCB, l'incertitude systématique est égale à l'incertitude statistique obtenue<sup>5</sup> lors des ajustements DY *double-fake* sur les simulations.

Pour les paramètres dCB+exp dans la méthode de l'enveloppe, on considère alors 5 possibilités :

- 1) les paramètres sont fixés aux valeurs obtenues lors des ajustements DY *double-fake* sur les simulations (cas utilisé actuellement dans l'analyse)
- 2) les paramètres sont libres de flotter dans un domaine  $\pm 1\sigma_{stat}$  de leur incertitude statistique
- 3) les paramètres sont libres de flotter dans un domaine  $\pm 4\sigma_{stat}$  de leur incertitude statistique
- 4) les paramètres sont libres de flotter dans un domaine  $\pm 1\sigma_{tot}$  de leur incertitude systématique totale
- 5) les paramètres sont libres de flotter dans un domaine  $\pm 4\sigma_{tot}$  de leur incertitude systématique totale

La composition des enveloppes et les meilleures fonctions d'ajustement obtenues pour ces possibilités présentent quelques différences et sont visibles sur la figure 7.21 : pour la classe 0, les fonctions  $Ber_3$  et  $Exp_1$  sont présentes en plus dans l'enveloppe pour les possibilités 3) et 5), et la meilleure fonction d'ajustement passe de  $Lau_1$  à  $Pow_1$  pour la possibilité 5); pour la classe 1, la fonction  $Lau_2$  n'est présente que pour la possibilité 1) et la fonction  $Pow_3$  uniquement dans la possibilité 2); enfin, pour la classe 2, la meilleure fonction d'ajustement passe de  $Exp_3$  à  $Ber_4$  pour les possibilités 3) et 5).

---

4. Aucune pénalité concernant les déviations par rapport aux valeurs centrales n'est incluse à cette étape.

5. Cette hypothèse sera brièvement discutée dans la suite.

Les valeurs des paramètres dCB+exp obtenues après les ajustements de la méthode de l'enveloppe sont elles visibles sur le tableau 7.2. On observe que dans les cas  $\pm 1\sigma_{stat/tot}$  de nombreux paramètres sont à la limite de leur domaine de définition, ces paramètres à la limite apparaissent en bleu dans le tableau. De façon plus générale, on observe que les valeurs des paramètres peuvent s'éloigner à plusieurs  $\sigma_{stat/tot}$  de leur valeur centrale lorsque leur domaine de définition les autorise. Cependant, plusieurs paramètres, notamment le paramètre sigma de la dCB pour les classes 0 et 1, sont toujours à la limite de leur domaine de définition, même lorsque celui-ci est défini à  $\pm 4\sigma_{tot}$ . Ceci est problématique puisque synonyme de désaccords entre données et simulations (dans l'hypothèse où les données ne contiennent pas de signal) non couverts pas les incertitudes systématiques. Le fait que les contenus des enveloppes changent selon la possibilité envisagée est de même problématique.

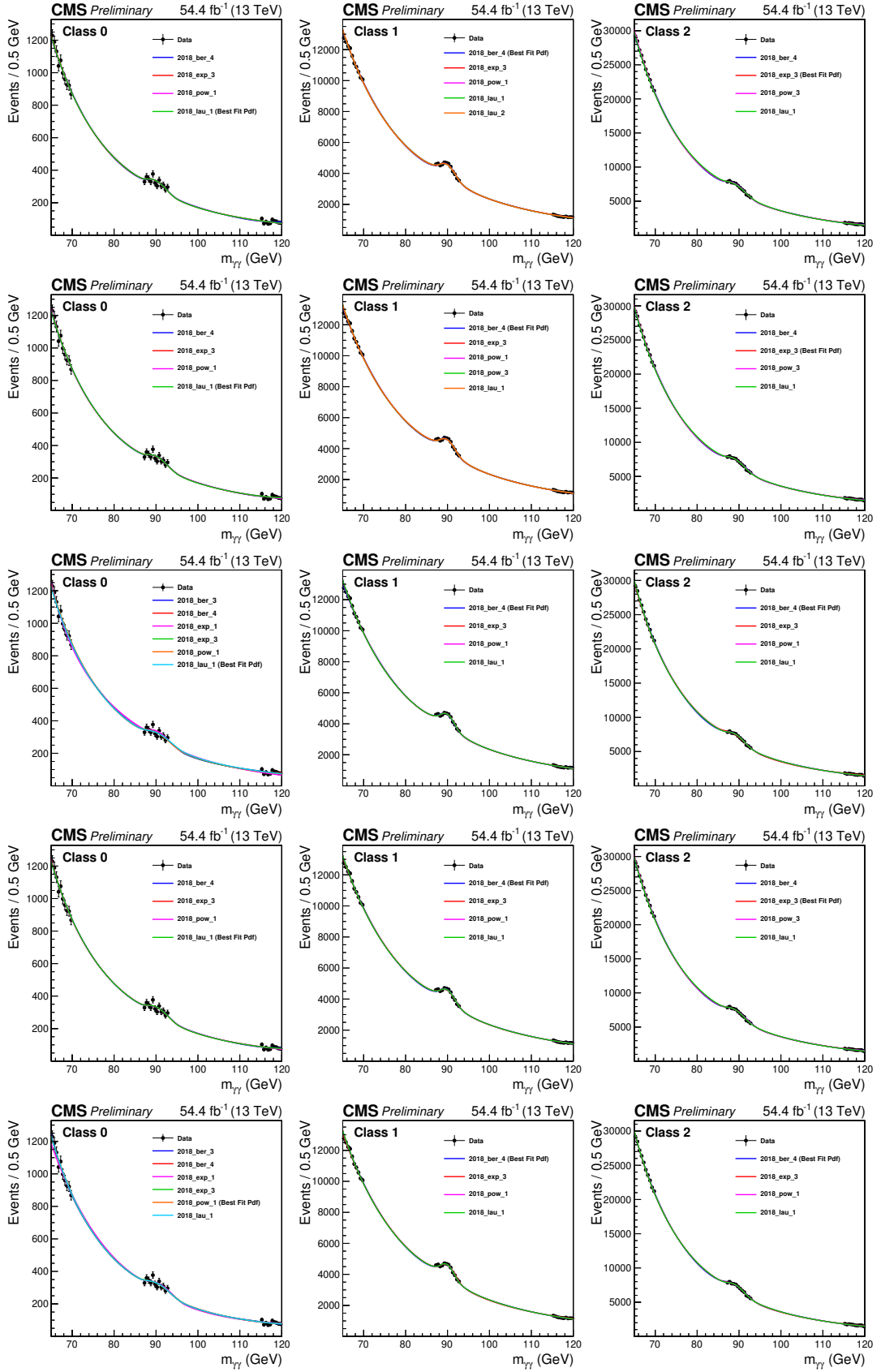


FIGURE 7.21 – Ajustement de la masse invariante diphoton des données 2018 par les fonctions composant l’enveloppe pour les possibilités 1) à 5) de définition des paramètres de la fonction dCB+exp de haut en bas.

	Definition	1) fix	2) $1\sigma$ stat	3) $4\sigma$ stat	4) $1\sigma$ total	5) $4\sigma$ total
Class 0	Best fit pdf	Lau <sub>1</sub>	Lau <sub>1</sub>	Lau <sub>1</sub>	Lau <sub>1</sub>	Pow <sub>1</sub>
	DCB frac	0.04	0.05	0.04	0.05	0.04
	mean : $90.07 \pm (0.18)$ 0.27	90.07	<b>90.25</b>	90.41	<b>90.34</b>	90.42
	sigma : $2.60 \pm (0.19)$ 0.19	2.60	<b>2.79</b>	<b>3.36</b>	<b>2.79</b>	<b>3.36</b>
	nCB1 : $0.31 \pm 0.06$	0.31	<b>0.37</b>	0.53	<b>0.37</b>	0.53
	nCB2 : $1.26 \pm 0.22$	1.26	1.26	<b>2.13</b>	1.03	1.17
	Exp frac : $0.01 \pm 0.12$	0.01	0.02	0.03	0.02	0.04
Exp coeff : $-0.02 \pm 0.38$	-0.02	0.35	0.32	0.35	0.27	
Class 1	Best fit pdf	Ber <sub>4</sub>	Ber <sub>4</sub>	Ber <sub>4</sub>	Ber <sub>4</sub>	Ber <sub>4</sub>
	DCB frac	0.05	0.05	0.04	0.05	0.035
	mean : $89.86 \pm (0.06)$ 0.06	89.86	89.82	89.89	<b>89.80</b>	89.96
	sigma : $2.20 \pm (0.06)$ 0.07	2.20	<b>2.14</b>	2.09	<b>2.27</b>	<b>1.92</b>
	nCB1 : $1.19 \pm 0.16$	1.19	<b>1.36</b>	<b>1.82</b>	<b>1.35</b>	<b>1.82</b>
	nCB2 : $1.68 \pm 0.08$	1.68	<b>1.60</b>	1.35	<b>1.60</b>	1.93
	Exp frac : $0.14 \pm 0.03$	0.14	0.16	0.23	0.16	0.20
Exp coeff : $-0.13 \pm 0.02$	-0.13	<b>-0.15</b>	-0.04	<b>-0.11</b>	-0.04	
Class 2	Best fit pdf	Exp <sub>3</sub>	Exp <sub>3</sub>	Ber <sub>4</sub>	Exp <sub>3</sub>	Ber <sub>4</sub>
	DCB frac	0.04	0.04	0.05	0.04	0.03
	mean : $89.09 \pm (0.06)$ 0.07	89.09	<b>89.03</b>	<b>88.85</b>	<b>89.02</b>	89.09
	sigma : $2.62 \pm (0.07)$ 0.80	2.62	<b>2.69</b>	<b>2.90</b>	2.69	3.31
	nCB1 : $0.76 \pm 0.12$	0.76	<b>0.65</b>	<b>1.23</b>	<b>0.88</b>	<b>1.23</b>
	nCB2 : $1.25 \pm 0.06$	1.25	<b>1.20</b>	<b>1.02</b>	<b>1.20</b>	1.29
	Exp frac : $0.10 \pm 0.03$	0.10	<b>0.13</b>	<b>0.21</b>	<b>0.13</b>	0.11
Exp coeff : $-0.21 \pm 0.05$	-0.21	-0.15	-0.25	<b>-0.15</b>	-0.26	

TABLE 7.2 – Valeurs des paramètres dCB+exp par classe obtenues après méthode de l'enveloppe pour les possibilités 1) à 5). A gauche, les valeurs obtenues suite à l'ajustement sur les simulations DY *double-fake* avec leur incertitude (statistique) totale. En bleu sont visibles les paramètres dont la valeur est à la limite de leur domaine de définition.

Afin de visualiser si ces changements de contenu des enveloppes ont un effet important sur les résultats finaux, on calcule les limites d'exclusion attendues sur le produit section efficace fois rapport de branchement en deux photons normalisé et la p-value attendue en absence de signal pour chacune des 5 possibilités précédentes, en utilisant les valeurs centrales obtenues par défaut (soit la possibilité 1)) mais avec des compositions d'enveloppe différentes. Ces résultats sont respectivement visibles sur les figures 7.22 et 7.23.

Les différences obtenues entre les différentes possibilités sont faibles et surtout dues aux changements de meilleures fonctions d'ajustement. Ces différences ne sont pas significatives.



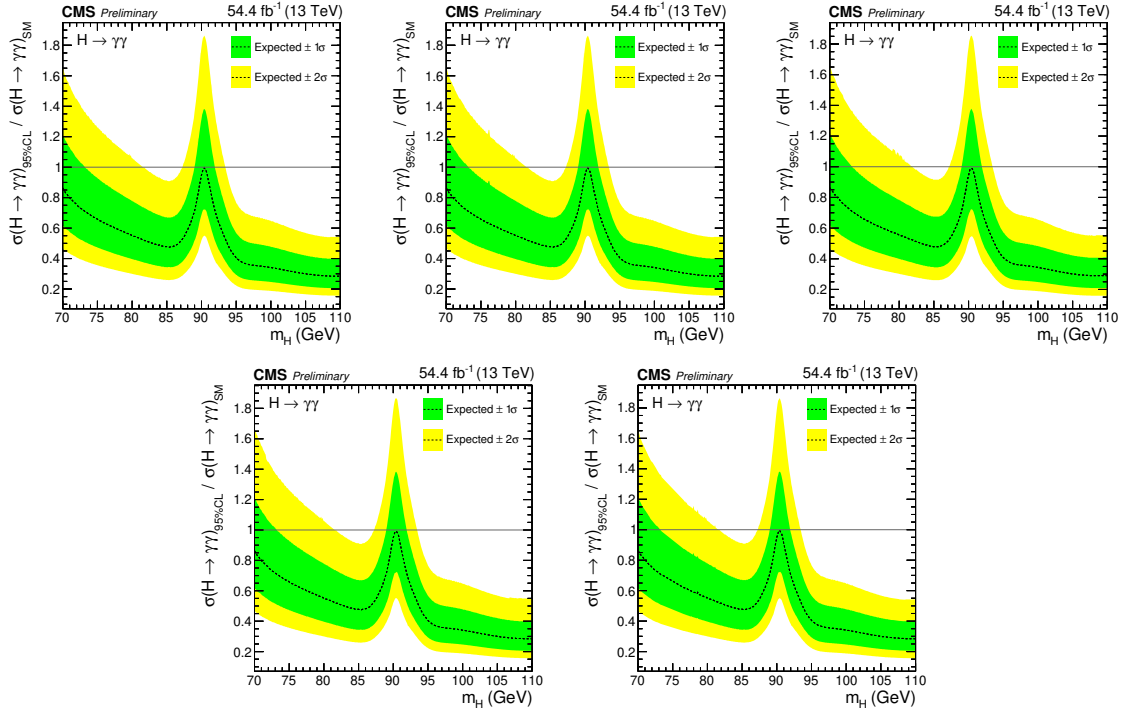


FIGURE 7.22 – Limites d’exclusion attendues sur le produit section efficace fois rapport de branchement en deux photons normalisé pour un boson de Higgs additionnel pour les possibilités 1), 2), 3) (haut), et 4) et 5) (bas) pour 2018.

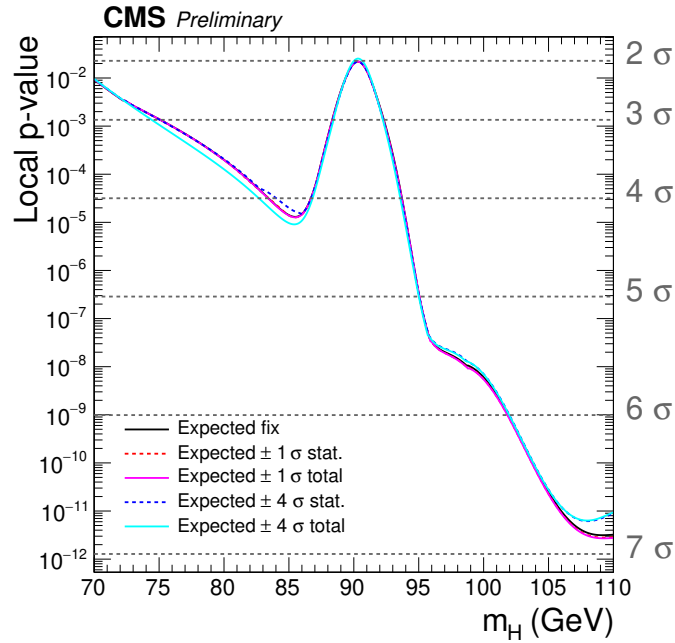


FIGURE 7.23 – P-value locale attendue en fonction de la masse d’un boson de Higgs additionnel pour les possibilités 1) à 5) (bas) pour 2018.

On cherche maintenant à comparer les valeurs des paramètres dCB+exp obtenues après la méthode de l’enveloppe avec celles obtenues en sortie des tests statistiques finaux, selon la méthode décrite à la section 6.11.1, que l’on dénommera avec le terme "Combine" du nom du

logiciel utilisé. On s'intéresse aux valeurs Combine correspondant aux données dans l'hypothèse de **bruit de fond seul** qui devraient être comparables à celles obtenues après la méthode de l'enveloppe. On rappelle que les paramètres dCB+exp sont définis autour de  $\pm 4\sigma_{tot}$  des valeurs centrales considérées dans Combine.

La figure 7.24 présente à gauche le *pull* obtenu pour le cas par défaut (soit la possibilité 1)) défini comme la différence entre la valeur Combine des paramètres observée obtenue, à partir des valeurs centrales de la possibilité 1), après ajustement du bruit de fond seul et la valeur correspondant à celle obtenue lors de l'ajustement dCB+exp sur les simulations DY divisée par l'incertitude totale associée au paramètre considéré. On observe que cette différence est contenue dans  $\pm 1\sigma_{tot}$  de l'erreur systématique (à part pour le  $\mu$  de la classe 2) ce qui est donc différent des résultats obtenus après la méthode de l'enveloppe où, pour la possibilité 5) par exemple, on avait des valeurs atteignant jusqu'à 3 ou plus  $\sigma_{tot}$ . Ceci s'explique par le fait qu'il est très coûteux en terme de fonction de vraisemblance pour ces paramètres de nuisance de s'éloigner de leur valeur centrale dans l'analyse statistique Combine, comme on l'a vu dans la section 6.11.1. La figure 7.24 droite présente la simple différence entre ces mêmes valeurs Combine et les valeurs obtenues après la méthode de l'enveloppe pour la possibilité 5), les conclusions sont similaires : Combine ne parvient pas à reproduire les valeurs obtenus pour la possibilité 5) car le processus de pénalisation des déviations des NP par rapport à leur valeur centrale rend ces déviations trop coûteuses en terme de vraisemblance. Autrement dit, avec la méthode actuellement utilisée dans l'analyse, on ne parvient pas à reproduire les résultats obtenus lorsque les paramètres sont, relativement, libres de flotter dans la méthode de paramétrisation du bruit de fond sur les données.

Si l'on prend maintenant les valeurs obtenues après la méthode de l'enveloppe pour la possibilité 5) comme valeurs centrales par défaut pour nos paramètres dCB+exp, on peut voir sur la tableau 7.3 que les valeurs Combine obtenues sont très similaires aux valeurs obtenues après la méthode de l'enveloppe pour la possibilité 5), ce qui est cohérent mais souligne la forte dépendance de Combine envers les valeurs centrales.

On cherche ensuite à déterminer s'il est possible de retrouver les valeurs de la méthode de l'enveloppe pour la possibilité 5) avec Combine dans le cas où les paramètres sont libres de flotter, sans pénalité. Le tableau 7.4 présente alors les valeurs Combine des paramètres dCB+exp obtenues dans le cas<sup>6</sup> où les incertitudes totales associées sont très grandes, cas ainsi analogue à celui où les paramètres seraient complètement libres de flotter. Les valeurs obtenues dans ce cas sont proches de celles obtenues avec la possibilité 5), ce qui est cohérent, à part pour les paramètres  $n_{CB1/2}$  et pour la classe 2 qui est ici pathologique puisque artificiellement dominée par la partie exponentielle. Ces écarts sur les paramètres modélisant la forme des queues de la dCB invitent à considérer des incertitudes systématiques sur ces paramètres ; nous aborderons ce problème dans la suite. Ces grandes variations de  $n_{CB1/2}$  permettent de plus d'atteindre un nouveau minimum pour le paramètre sigma de la dCB pour les classes 0 et

6. Les résultats sont similaires si les valeurs centrales de la possibilité 1) sont utilisées en place de celles de la possibilité 5).

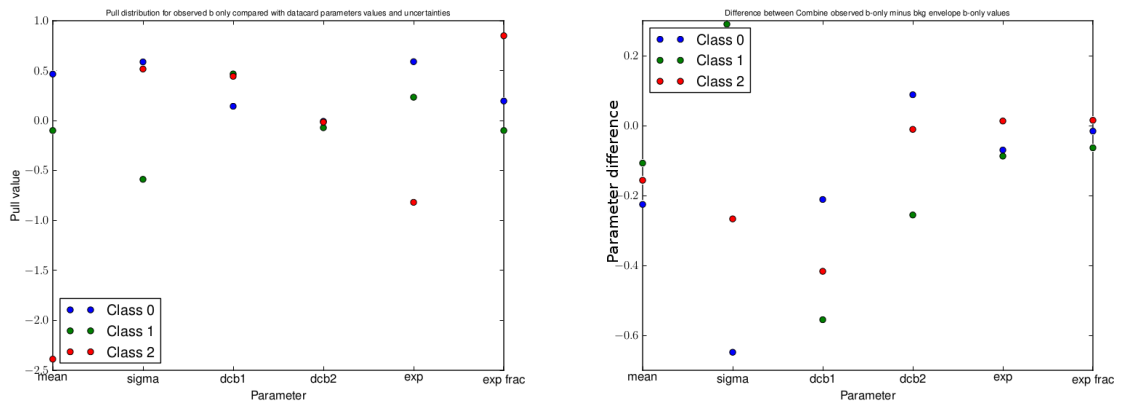


FIGURE 7.24 – A gauche, le **pull** obtenu pour le cas par défaut défini comme la différence entre la valeur Combine des paramètres observée après ajustement du bruit de fond seul et la valeur correspondant à celle obtenue lors de l’ajustement dCB+exp sur les simulations DY divisée par l’incertitude totale associée au paramètre considéré. A droite, la **différence** entre ces mêmes valeurs Combine et les valeurs obtenues après la méthode de l’enveloppe pour la possibilité 5) pour les différents paramètres de la fonction dCB+exp.

	Definition	1) fix bkg envelope	5) 4σ total bkg envelope	Combine with 1) central values	Combine with 5) central values
Class 0	mean :90.07 ± (0.18) 0.27	90.07	90.42	90.20	90.42
	sigma :2.60 ± (0.19) 0.19	2.60	3.36	2.71	3.37
	nCB1 :0.31 ± 0.06	0.31	0.53	0.32	0.53
	nCB2 :1.26 ± 0.22	1.26	1.17	1.26	1.17
	Exp frac :0.01 ± 0.12	0.01	0.04	0.025	0.04
	Exp coeff :-0.02 ± 0.38	-0.02	0.27	0.20	0.27
Class 1	mean :89.86 ± (0.06) 0.06	89.86	89.96	89.85	89.92
	sigma :2.20 ± (0.06) 0.07	2.20	1.92	2.16	1.96
	nCB1 :1.19 ± 0.16	1.19	1.82	1.26	1.85
	nCB2 :1.68 ± 0.08	1.68	1.93	1.67	1.93
	Exp frac :0.14 ± 0.03	0.14	0.20	0.14	0.2
	Exp coeff :-0.13 ± 0.02	-0.13	-0.04	-0.13	-0.04
Class 2	mean :89.09 ± (0.06) 0.07	89.09	89.09	88.93	88.94
	sigma :2.62 ± (0.07) 0.80	2.62	3.31	3.04	3.04
	nCB1 :0.76 ± 0.12	0.76	1.23	0.81	1.25
	nCB2 :1.25 ± 0.06	1.25	1.29	1.25	1.26
	Exp frac :0.10 ± 0.03	0.10	0.11	0.125	0.13
	Exp coeff :-0.21 ± 0.05	-0.21	-0.26	-0.25	-0.28

TABLE 7.3 – Valeurs des paramètres dCB+exp par classe obtenus après méthode de l’enveloppe pour les possibilités 1) et 5) suivies des valeurs Combine obtenues à partir des valeurs 1) comme valeurs centrales et des valeurs 5) comme valeurs centrales.

	Definition	5) $4\sigma$ total bkg envelope	Combine with 5) central values	Combine with 5) central values and very big uncert. ( $\sim$ free)
Class 0	mean : $90.07 \pm (0.18) 0.27$	90.42	90.42	90.48
	sigma : $2.60 \pm (0.19) 0.19$	3.36	3.37	3.05
	nCB1 : $0.31 \pm 0.06$	0.53	0.53	1.98
	nCB2 : $1.26 \pm 0.22$	1.17	1.17	2.69
	Exp frac : $0.01 \pm 0.12$	0.04	0.04	0.04
	Exp coeff : $-0.02 \pm 0.38$	0.27	0.27	0.32
Class 1	mean : $89.86 \pm (0.06) 0.06$	89.96	89.92	89.87
	sigma : $2.20 \pm (0.06) 0.07$	1.92	1.96	2.12
	nCB1 : $1.19 \pm 0.16$	1.82	1.85	7.64
	nCB2 : $1.68 \pm 0.08$	1.93	1.93	1.01
	Exp frac : $0.14 \pm 0.03$	0.20	0.2	0.14
	Exp coeff : $-0.13 \pm 0.02$	-0.04	-0.04	-0.06
Class 2	mean : $89.09 \pm (0.06) 0.07$	89.09	88.94	88.93
	sigma : $2.62 \pm (0.07) 0.80$	3.31	3.04	2.73
	nCB1 : $0.76 \pm 0.12$	1.23	1.25	28.04
	nCB2 : $1.25 \pm 0.06$	1.29	1.26	1.11
	Exp frac : $0.10 \pm 0.03$	0.11	0.13	0.95
	Exp coeff : $-0.21 \pm 0.05$	-0.26	-0.28	-0.09

TABLE 7.4 – Valeurs des paramètres dCB+exp par classe obtenus après méthode de l'enveloppe pour les possibilités 5) suivies des valeurs Combine obtenues à partir des valeurs 5) comme valeurs centrales et avec des incertitudes systématiques très élevées.

1 et ce paramètre n'est alors plus à la limite. Ceci montre une corrélation entre les différents paramètres de la fonction dCB+exp.

Que conclure de ces différentes études ? La question est complexe. On a bien observé des désaccords entre nos données et nos simulations pour les valeurs des paramètres dCB+exp pouvant aller jusqu'à plusieurs  $\sigma_{tot}$  après l'ajustement par la méthode de l'enveloppe ; ces désaccords sont moins prononcés avec Combine, du fait des pénalités en terme de vraisemblance pour les NP, mais les résultats obtenus sont très dépendants des valeurs centrales utilisées.

On pourrait considérer des incertitudes systématiques supplémentaires prenant par exemple en compte les différences observées entre données et simulations en prenant la différence entre les valeurs Combine dans le cas où les paramètres sont quasiment libres de flotter et les valeurs obtenues sur les simulations DY. Le problème est que les résultats obtenus sur les données peuvent être biaisés.

Notamment la présence d'un éventuel signal proche du pic du Z comme le léger excès observé dans [103] peut provoquer un élargissement et un déplacement de la dCB vers des masses plus élevées pour absorber une partie de l'excès. On peut par exemple voir sur la figure 7.25 que la moyenne et la largeur de la fonction dCB obtenues après ajustement sur les simulations DY *double-fake* augmente lorsqu'on ajoute un signal  $h \rightarrow \gamma\gamma$  de force 1 à une masse de 95 GeV. Les valeurs obtenues pour les données ne sont donc pas fiables au sens où elles peuvent potentiellement être modifiées par la présence d'un signal ce qui n'est pas le cas des valeurs obtenues sur les simulations.

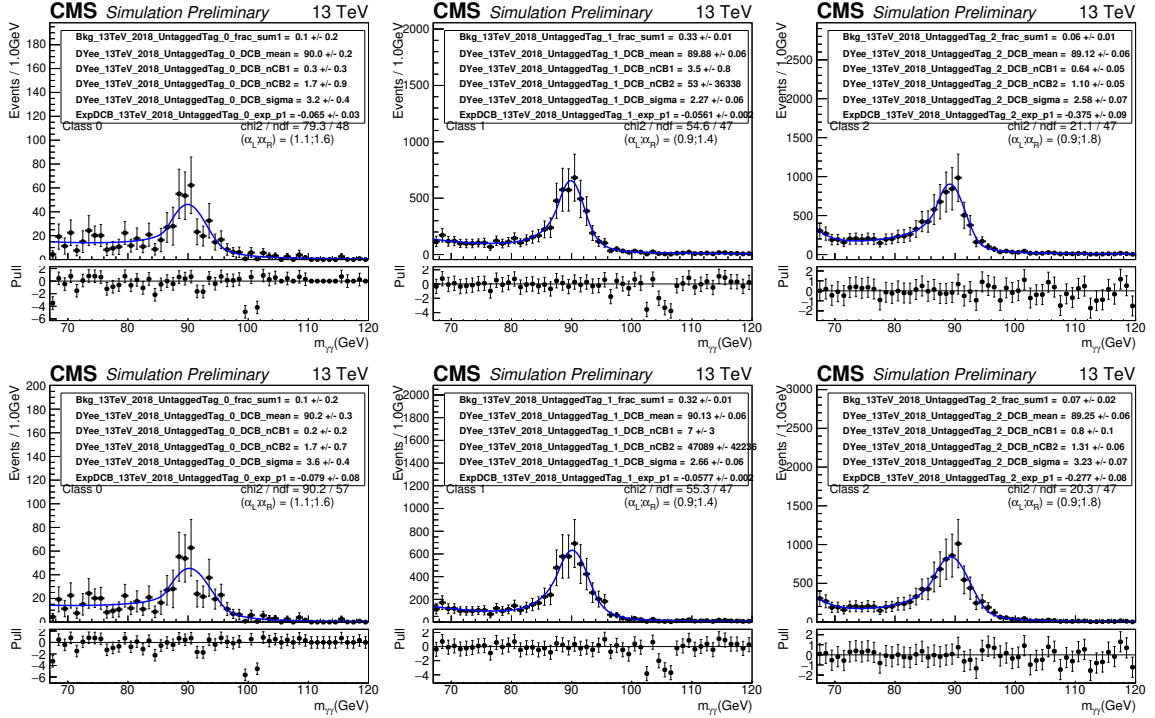


FIGURE 7.25 – Exemples d’ajustement par une fonction dCB+exp des simulations DY pour 2018 par classe. En haut, seule la composante DY *double-fake* est présente alors que l’on a artificiellement ajouté un signal  $h \rightarrow \gamma\gamma$  à une masse de 95 GeV et de force 1 en bas.

### Incertitudes sur les paramètres de queues de la dCB

On a vu précédemment que l’on pouvait avoir d’importantes variations des paramètres définissant les queues de la fonction dCB notamment dans les données, or l’erreur systématique associée à ces paramètres n’est qu’une incertitude statistique obtenue après ajustement sur les simulations DY *double-fake*. De plus, aucune incertitude n’est actuellement prise en compte sur les paramètres  $\alpha_{R/L}$  qui sont fixés en utilisant les échantillons *single-fake*. On peut cependant observer sur la figure 7.26 que ces paramètres, lorsqu’autorisés à varier dans un intervalle défini, peuvent s’éloigner des valeurs fixées à l’étape Combine.

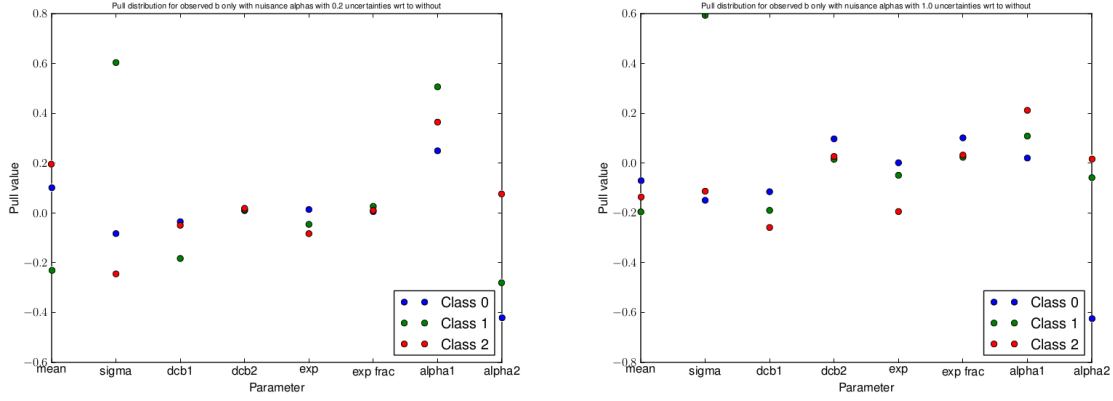


FIGURE 7.26 – Valeur du *pull* défini comme la différence entre la valeur Combine obtenue dans l’hypothèse de bruit de fond seul et la valeur centrale obtenue après ajustement de la fonction dCB+exp sur les simulations DY divisée par l’incertitude systématique associée pour les paramètres de la fonction dCB+exp. Les incertitudes systématiques considérées pour  $\alpha_{R/L}$  sont 0,2 à gauche et 1,0 à droite pour 2018.

Bien que l’on ait vu que les valeurs obtenues par un ajustement direct sur les données peuvent être biaisées, il semble pertinent de prendre en compte une incertitude systématique sur les valeurs  $\alpha_{R/L}$ . On considère deux façons de faire : la première consiste en une simple incertitude sur les valeurs de  $\alpha_{R/L}$  choisies en minimisant la distribution  $\chi^2/ndf - 1$  pour les événements *single-fake*. D’après les figures 7.28, 7.30 et 7.32, une incertitude de 0,2 sur les valeurs de  $\alpha_{R/L}$  permet de couvrir les zones ayant un  $\chi^2/ndf - 1$  proche du minimum. La figure 7.27 droite montre les limites d’exclusion attendues avec cette incertitude supplémentaire sur les valeurs de  $\alpha_{R/L}$ , la forme de ces limites est alors différente et elles sont dégradées autour du pic du Z vers 85 GeV et, plus légèrement, vers 96 GeV.

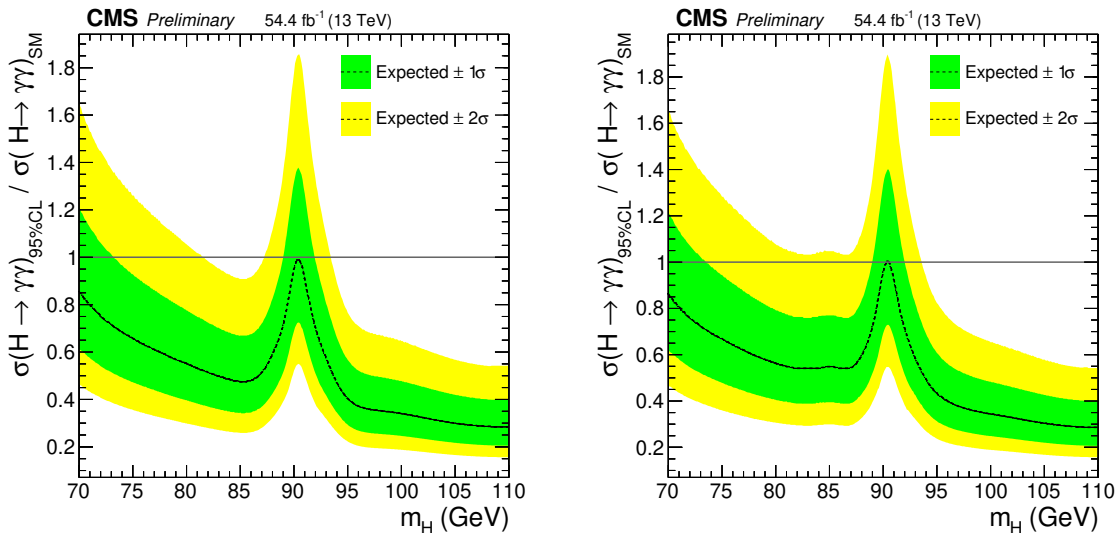


FIGURE 7.27 – Limite d’exclusion attendue sur le produit section efficace fois rapport de branchement en deux photons normalisée pour un boson de Higgs additionnel avec (droite) et sans (gauche) incertitude systématique supplémentaire sur les paramètres  $\alpha_{R/L}$  pour 2018.

Si cette méthode permet bien de prendre en compte une incertitude systématique sur la

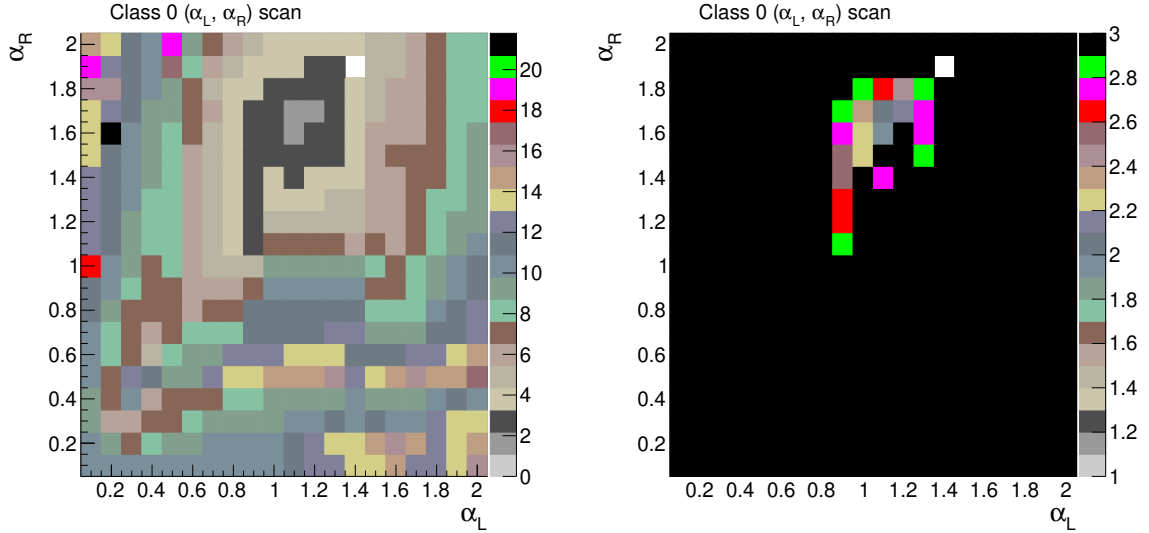


FIGURE 7.28 – Optimisation des valeurs  $\alpha_{L/R}$  obtenues à partir d’ajustement sur des données *single-fake* 2018 et pour la classe 0. La barre de couleur représente la valeur  $\chi^2/ndf$  obtenue pour chaque ajustement réalisé pour différentes valeurs de  $\alpha_{L/R}$ . La figure de droite présente une échelle de couleur plus fine.

forme des queues de la fonction dCB, ses paramètres étant corrélés entre eux, les variations des autres paramètres induites par ces incertitudes pourraient être trop contraintes dans Combine par leur propre incertitude totale qui est, à part pour le  $\mu$  et le  $\sigma$ , égale à l’incertitude statistique associée à l’ajustement des simulations DY; ces variations des autres paramètres pourraient être de même contraintes par leur intervalle de définition.

La deuxième façon de faire, plus conservatrice, consiste à considérer des valeurs de  $\alpha_{R/L}$  proches de celle minimisant la distribution  $\chi^2/ndf - 1$  des événements *single-fake* et ajouter de manière quadratique la différence maximale obtenue entre les valeurs des paramètres obtenues avec ces nouvelles valeurs  $\alpha_{R/L}$  et les valeurs usuelles obtenues avec les valeurs  $\alpha_{R/L}$  minimisant le  $\chi^2/ndf - 1$ .

Pour la classe 0, la distribution  $\chi^2/ndf$  pour les événements données *single-fake* permettant de choisir les  $\alpha_{R/L}$  est visible sur la figure 7.28, les meilleurs  $\alpha_{R/L}$  sont toujours (1,1;1,6) mais d’autres valeurs avec un  $\chi^2/ndf$  proche sont notamment (1,1;1,7), (1,2;1,7) ou encore (1,2;1,8); les ajustements DY *double-fake* correspondant sont visibles sur la figure 7.29. On observe bien des différences pour les valeurs centrales des paramètres dCB+exp pour différentes valeurs de  $\alpha_{R/L}$  cependant, à cause de la faible statistique de la classe 0, les incertitudes statistiques associées sont trop importantes pour permettre de prendre en compte ces différences.

De même pour la classe 1, la distribution  $\chi^2/ndf$  pour les événements données *single-fake* permettant de choisir les  $\alpha_{R/L}$  est visible sur la figure 7.30, les meilleurs  $\alpha_{R/L}$  sont toujours (0,9;1,4) mais d’autres valeurs avec un  $\chi^2/ndf$  proche sont notamment (1,0;1,4), (0,9;1,3), (0,9;1,5) ou des valeurs un peu plus éloignées comme (0,8;1,6) et (1,0;1,6); les ajustements DY *double-fake* correspondant sont visibles sur la figure 7.31. On observe bien des différences statistiquement significatives pour les valeurs centrales de quelques un des paramètres dCB+exp pour

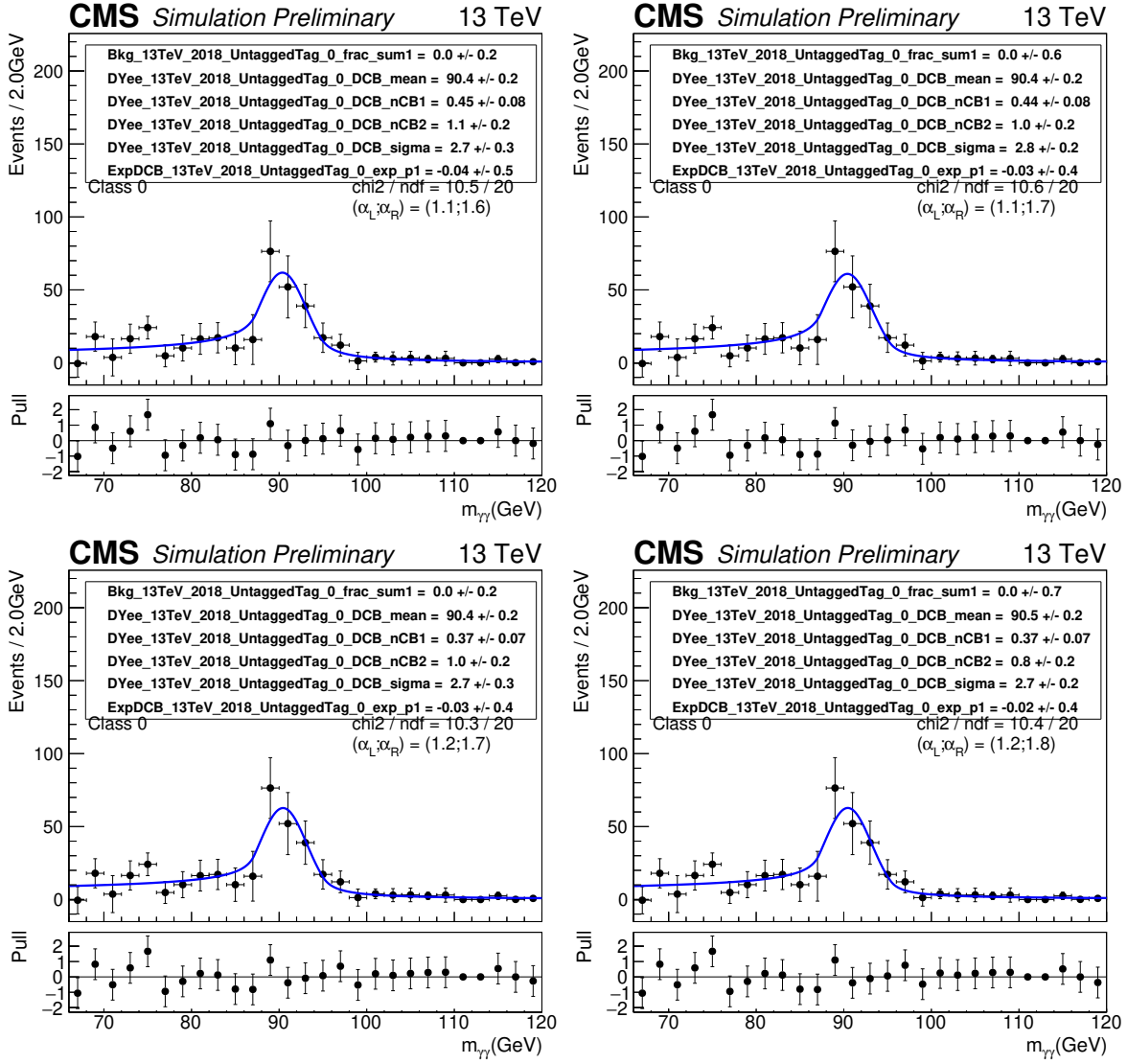


FIGURE 7.29 – Ajustement de la distribution de masse invariante diphoton par une fonction dCB+exp pour des simulations d'évènements Drell-Yan *double-fake* 2018 pour différentes valeurs de  $\alpha_{L/R}$  pour la classe 0. A noter que les valeurs de  $\chi^2$  ne sont ici pas correctes mais permettent néanmoins des comparaisons relatives.

différentes valeurs de  $\alpha_{R/L}$  pour la classe 1. A noter que pour des valeurs  $\alpha_{R/L}$  éloignées des valeurs choisies dans l'analyse, le paramètre  $n_{CB2}$  devient très grand et la nature de l'ajustement change.

Finalement pour la classe 2, la distribution  $\chi^2 / ndf$  pour les évènements données *single-fake* permettant de choisir les  $\alpha_{R/L}$  est visible sur la figure 7.32, les meilleurs  $\alpha_{R/L}$  sont toujours (0,9;1,8) mais d'autres valeurs avec un  $\chi^2 / ndf$  proche sont notamment (0,8;1,5), (1,1;1,8) ou encore (1,0;1,9); les ajustements DY *double-fake* correspondant sont visibles sur la figure 7.33. On observe bien des différences statistiquement significatives pour les valeurs centrales de quelques uns des paramètres dCB+exp pour différentes valeurs de  $\alpha_{R/L}$  pour la classe 2 mais ces différences sont faibles.

Assigner une incertitude systématique supplémentaire sur les paramètres de la fonction



dCB+exp de cette façon est donc possible pour les classes 1 et 2 mais semble difficile pour la classe 0 où les incertitudes statistiques sont grandes. Cette classe étant la plus sensible, cette méthode n'est donc pas optimale.

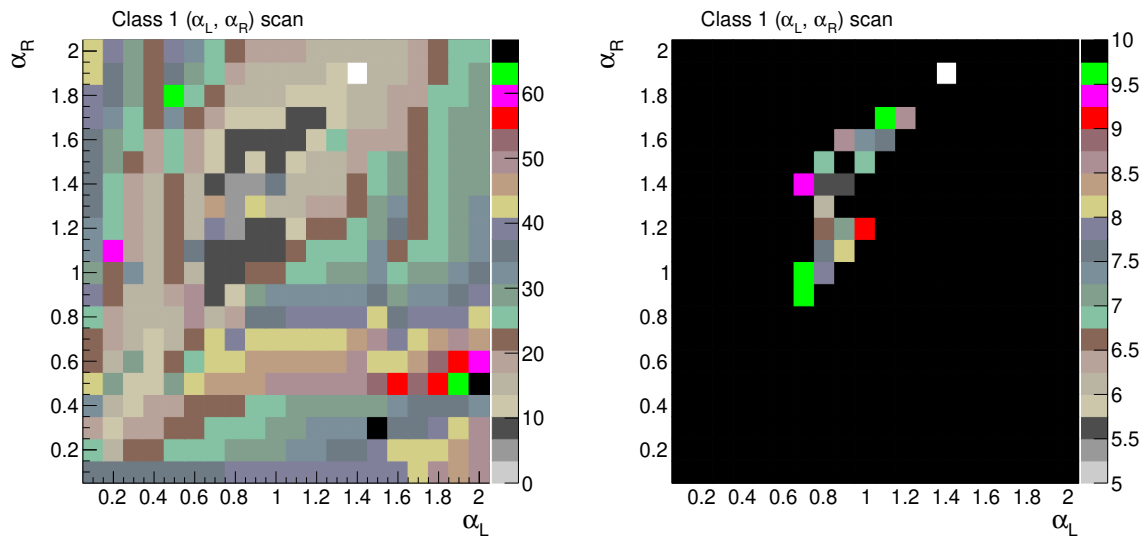


FIGURE 7.30 – Optimisation des valeurs  $\alpha_{L/R}$  obtenues à partir d'ajustement sur des données *single-fake* 2018 et pour la classe 1. La barre de couleur représente la valeur  $\chi^2/ndf$  obtenue pour chaque ajustement réalisé pour différentes valeurs de  $\alpha_{L/R}$ . La figure de droite présente une échelle de couleur plus fine.

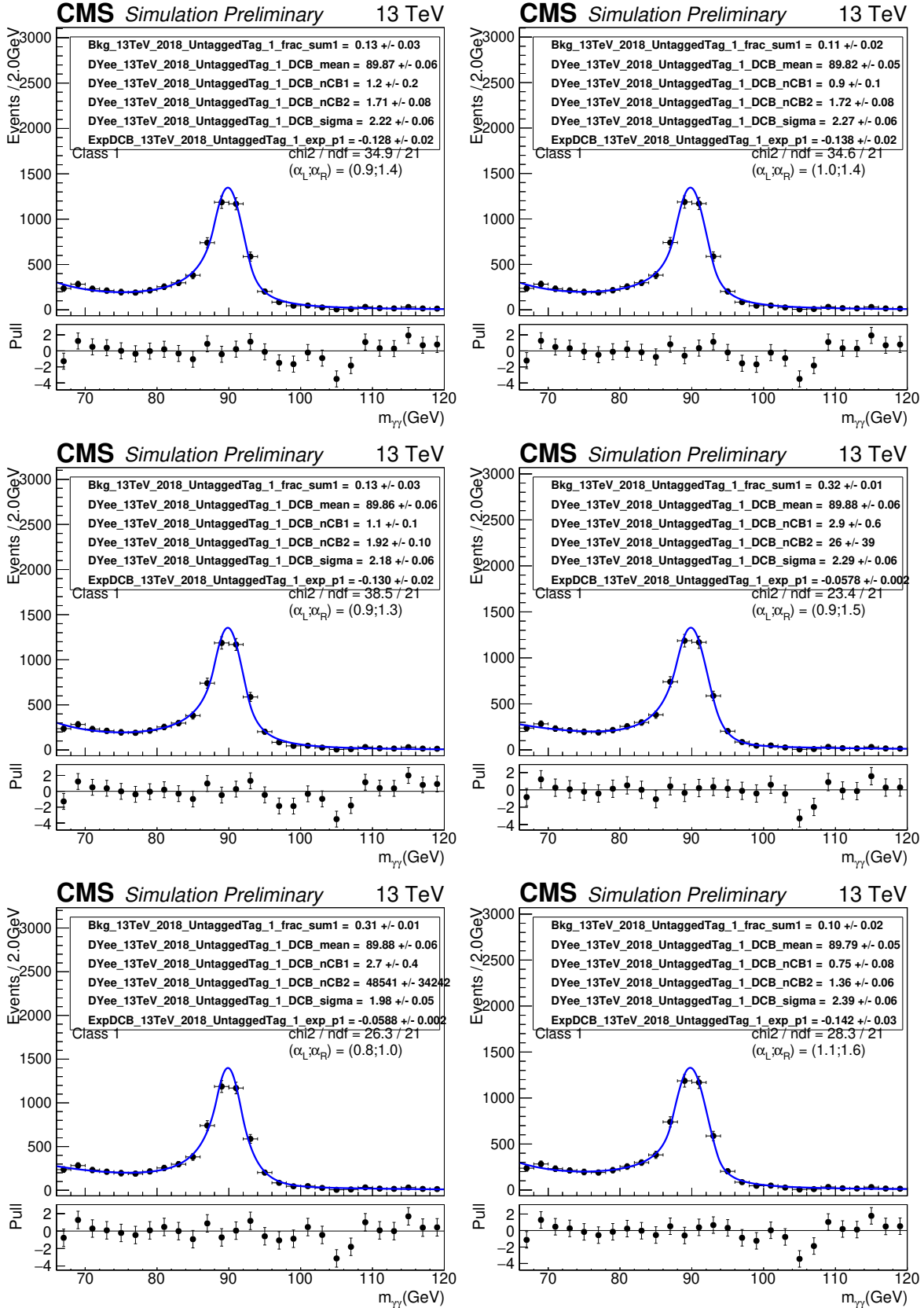


FIGURE 7.31 – Ajustement de la distribution de masse invariante diphoton par une fonction dCB+exp pour des simulations d'évènements Drell-Yan *double-fake* 2018 pour différentes valeurs de  $\alpha_{L/R}$  pour la classe 1. A noter que les valeurs de  $\chi^2$  ne sont ici pas correctes mais permettent néanmoins des comparaisons relatives.

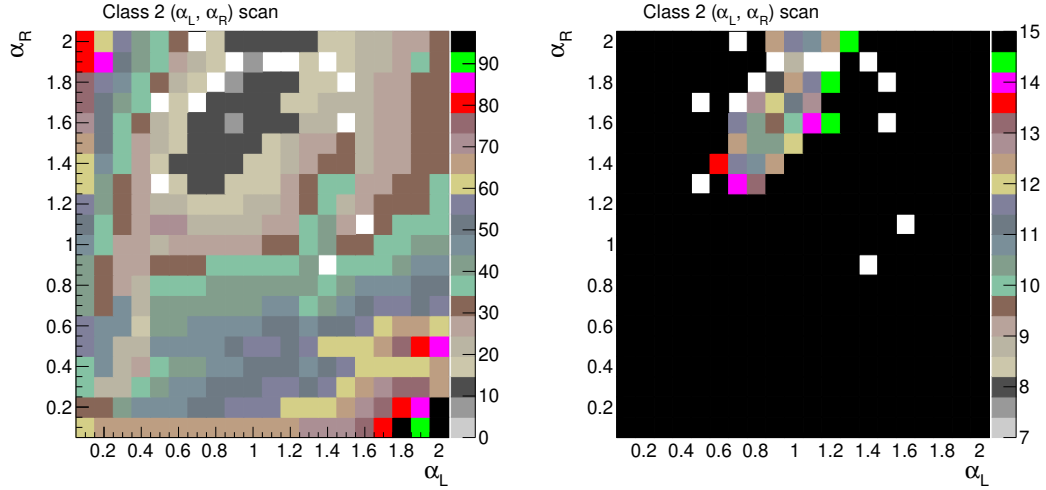


FIGURE 7.32 – Optimisation des valeurs  $\alpha_{L/R}$  obtenues à partir d’ajustement sur des données *single-fake* 2018 et pour la classe 2. La barre de couleur représente la valeur  $\chi^2/ndf$  obtenue pour chaque ajustement réalisé pour différentes valeurs de  $\alpha_{L/R}$ . La figure de droite présente une échelle de couleur plus fine.

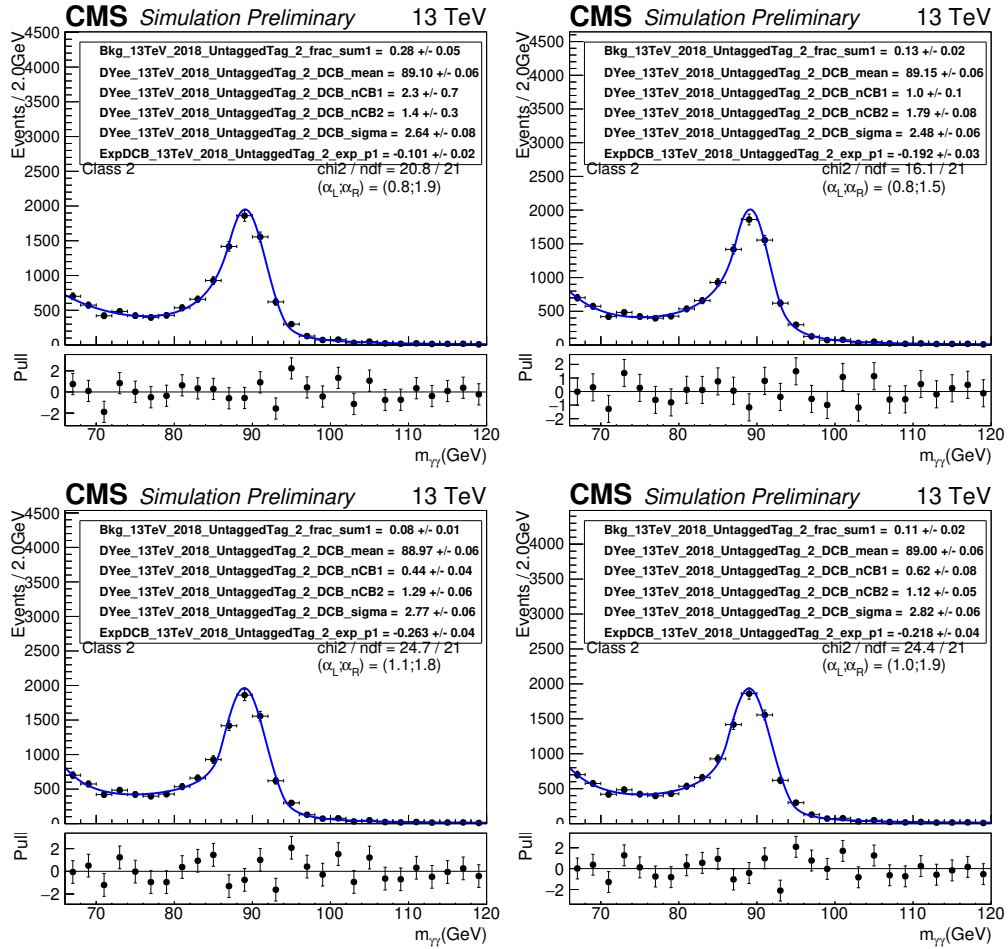


FIGURE 7.33 – Ajustement de la distribution de masse invariante diphoton par une fonction dCB+exp pour des simulations d’événements Drell-Yan *double-fake* 2018 pour différentes valeurs de  $\alpha_{L/R}$  pour la classe 2. A noter que les valeurs de  $\chi^2$  ne sont ici pas correctes mais permettent néanmoins des comparaisons relatives.

## Choix de la fonction d'ajustement de la composante Drell-Yan

Le choix de la fonction dCB pour ajuster les événements DY *double-fake* n'étant à priori pas ad hoc, il est intéressant de réaliser cet ajustement avec d'autres fonctions : on considère ici une fonction voigtienne et une fonction de Cruiff décrites dans le chapitre 5 qui ont une forme semblable à celle de la dCB.

Les ajustements réalisés avec la fonction voigtienne sont visibles sur la figure 7.34. On observe que le rapport de  $\chi^2/ndf$  est assez large bien qu'il soit acceptable pour la classe 0. De manière générale, les ajustements des simulations d'évènements Drell-Yan *double-fake* réalisés avec la fonction voigtienne ne sont pas bons.

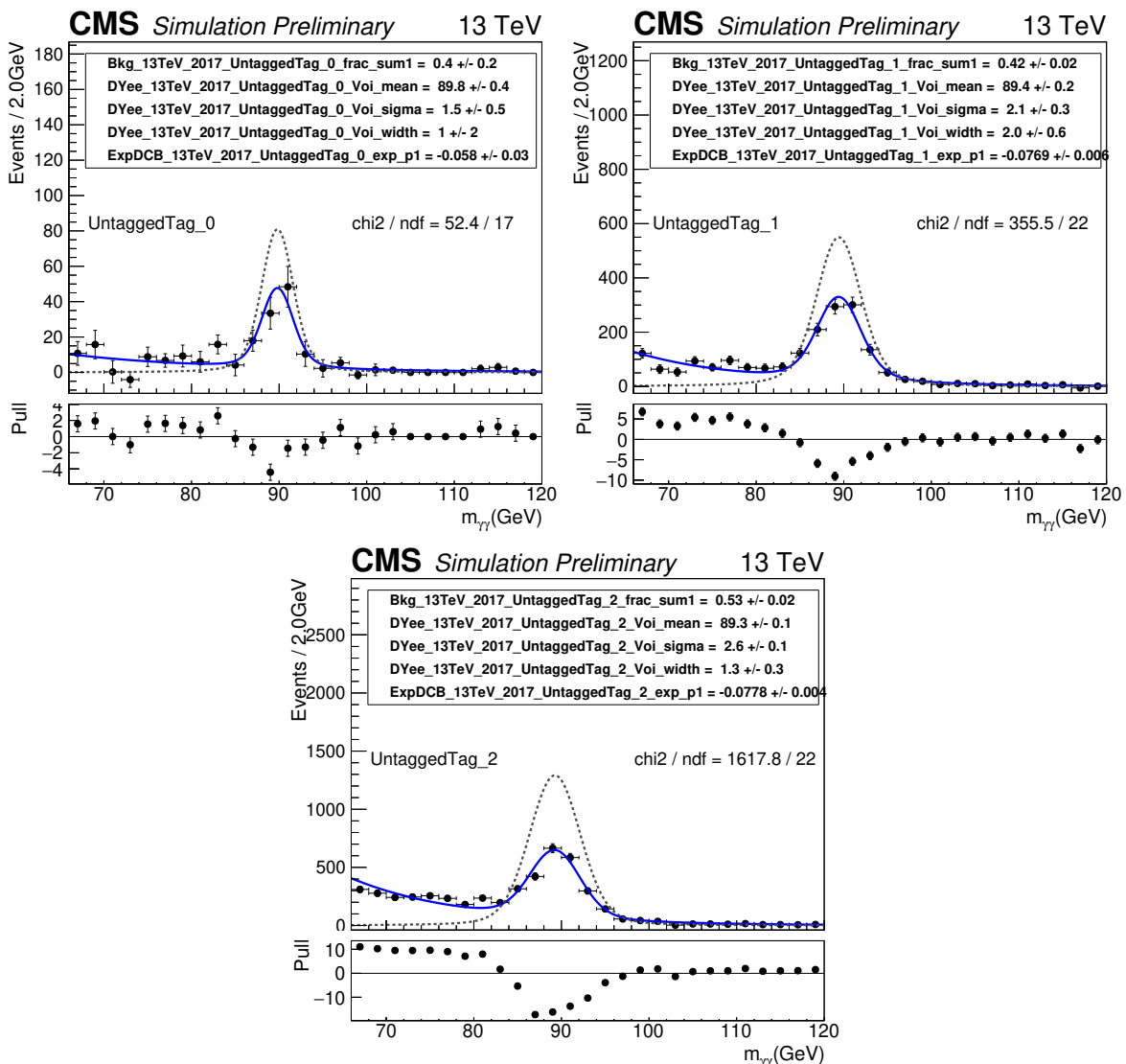


FIGURE 7.34 – Ajustement de la distribution de masse invariante diphoton par une fonction de Voigt+Exp pour des simulations d'évènements Drell-Yan *double-fake* 2017 par classe. En pointillé gris, la composante de la fonction Voigt seule.

Avec la fonction de Cruiff, les ajustements sont meilleurs comme on peut l'observer sur

la figure 7.35 où les valeurs  $\chi^2/ndf$  sont moins élevées et même proche de 1 pour la classe 0. Bien que ces figures de mérite ne justifient pas le remplacement de la fonction dCB, la fonction de Cruiff reste une alternative intéressante.

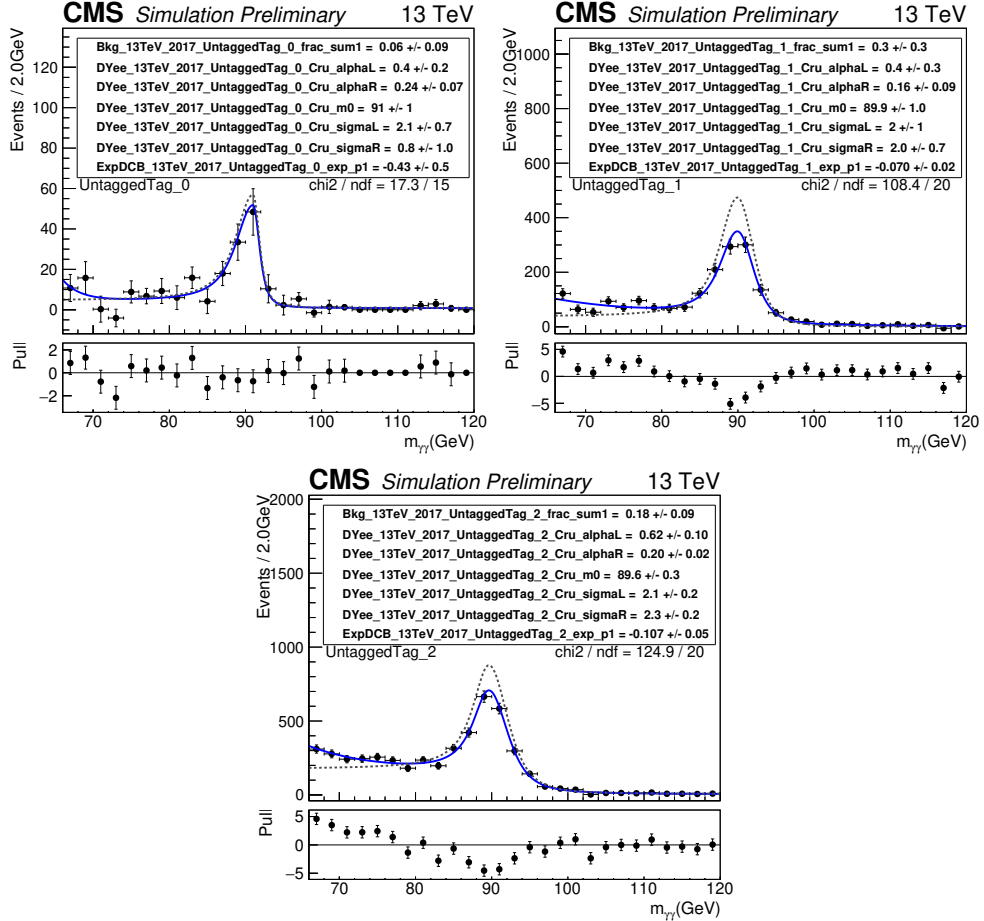


FIGURE 7.35 – Ajustement de la distribution de masse invariante diphoton par une fonction de Cruiff+Exp pour des simulations d’évènements Drell-Yan *double-fake* 2017 par classe. En pointillé gris, la composante de la fonction de Cruiff seule.

On pourrait finalement utiliser ces fonctions alternatives pour assigner une source d’incertitude systématique supplémentaire sur les paramètres de la fonction dCB+exp (mettons  $\mu$  et le  $\sigma$ ) prenant en compte le choix de la fonction d’ajustement. Cette incertitude serait alors définie comme la différence maximale observée sur un ajustement avec une fonction de Cruiff et celle observée avec une fonction dCB.

## Conclusion

On a donc étudié en quoi il était difficile d’estimer les accords entre données et simulations pour nos ajustements des évènements Drell-Yan *double-fake*. On peut néanmoins ajouter de nouvelles incertitudes systématiques pour prendre en compte la forme des queues de la dCB ou encore le choix de la fonction d’ajustement.

Une meilleure manière de faire, non abordée, ici serait de réaliser cet ajustement par une méthode basée directement sur les données. Un projet en cours est en effet d'utiliser une méthode de régression semblable à celle, présentée dans la section 4.3.1, permettant de corriger l'énergie des photons afin de modifier la distribution d'évènements Drell-Yan *zero-fake* (où aucun des deux électrons ne passe le critère de veto électron) afin qu'ils ressemblent aux évènements Drell-Yan *double-fake*.

### 7.3.3 Régions d'inefficacité en $\phi$

Comme on l'a vu dans la section 6.9 et comme on peut à nouveau l'observer sur la figure 7.36, il y a environ deux fois plus d'évènements Drell-Yan *double-fake* dans les simulations et données pour 2018 par rapport aux simulations et données 2017 alors que la facteur attendu dû à l'augmentation de la luminosité est d'environ 1,3. La différence de présélection ne peut expliquer qu'une partie de cette différence du nombre d'évènements, il faut donc comprendre d'où en provient la majorité.

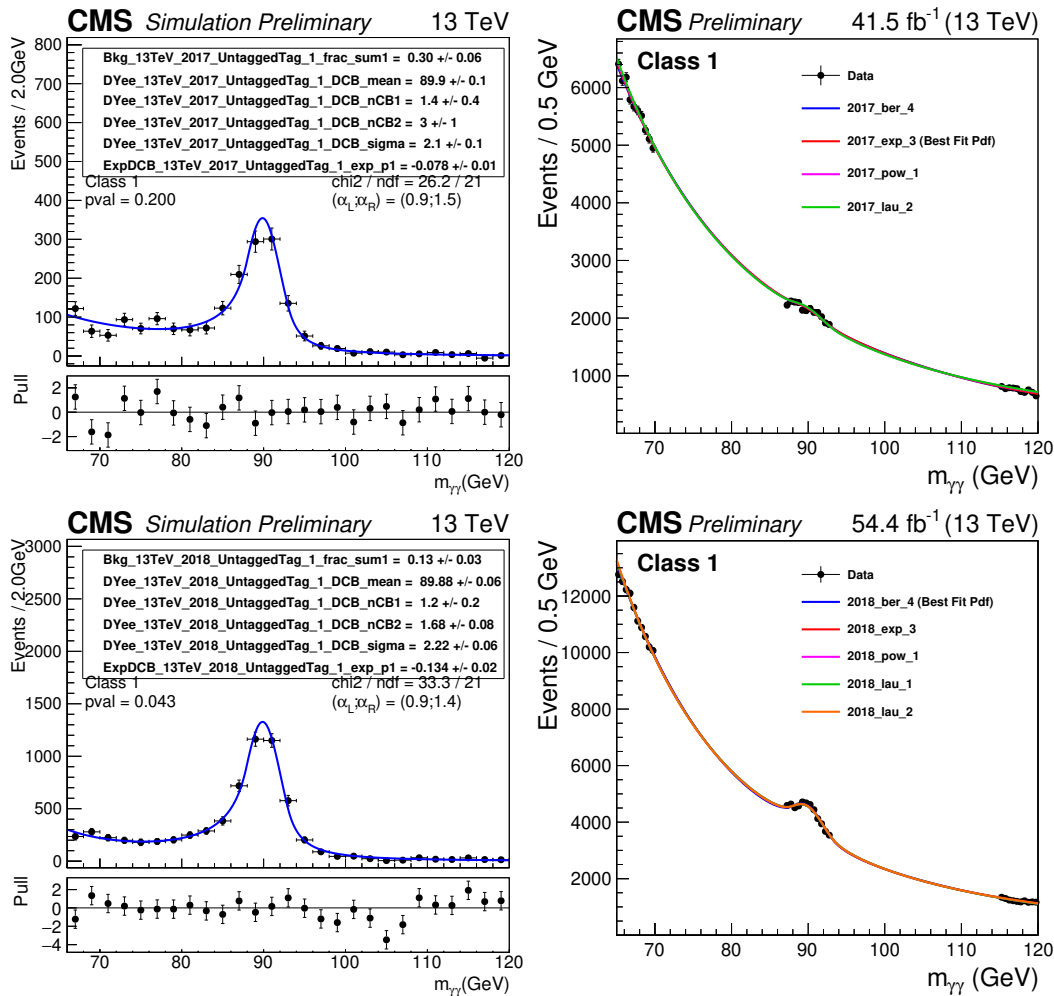


FIGURE 7.36 – Distribution et ajustement du bruit de fond total incluant la composante Drell-Yan pour 2017 (haut) et 2018 (bas) pour la classe 1.

Pour comprendre d'où vient cette différence dans le nombre d'évènements *DY double-fake*, on étudie les distributions angulaires de ces évènements. La figure 7.37 présente ainsi les distributions en  $\eta$  et  $\phi$  pour 2017 et 2018 pour des simulations *DY double-fake*. On observe que la forme des distributions en  $\eta$  est similaire entre 2017 et 2018 mais que la distribution en  $\phi$  est très différente entre 2017 et 2018. Notamment pour 2018 on observe un gros excès d'évènements à  $\phi \sim 0,7$  (et  $\phi \sim -2,4$ ) qui semble alors correspondre à des zones d'inefficacité du détecteur à pixels provoquant une baisse d'efficacité du veto électron.

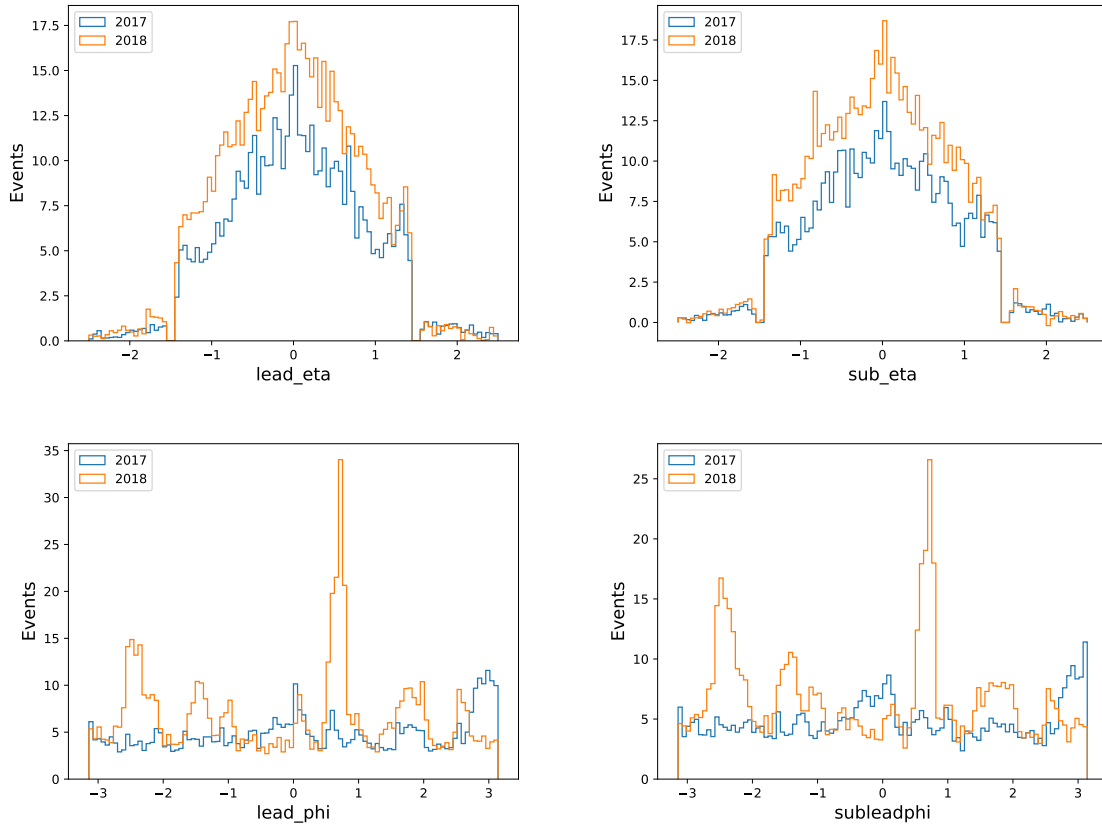


FIGURE 7.37 – Distributions en  $\eta$  (haut) et  $\phi$  (bas) pour le photon *lead* (gauche) et *sublead* (droite) pour 2017 et 2018 pour des simulations *DY double-fake*.

Pour étudier ce phénomène, on utilise des évènements  $Z \rightarrow ee$  et la méthode TnP pour visualiser la distribution en  $\phi$  de l'électron "probe", visible sur la figure 7.38, pour les données et les simulations, pour 2017 et 2018. On observe à nouveau une zone d'inefficacité en  $\phi$  pour 2018 correspondant à une baisse du nombre de vrais électrons vers  $\phi \sim 0,7$ . Ces résultats sont en accord avec les résultats officiels [123].

Ces zones d'inefficacité sont le fait de problèmes d'alimentation dans deux couches du détecteur à pixels. Une partie des électrons interagissant avec ces couches ne laisse ainsi pas de coup dans les pixels et sont donc reconstruits à tort comme n'étant pas des électrons. Notamment, dans notre analyse, ces électrons viennent peupler les populations *single-fake* et *double-fake* puisqu'ils apparaissent comme des électrons sans coup dans le détecteur à pixels

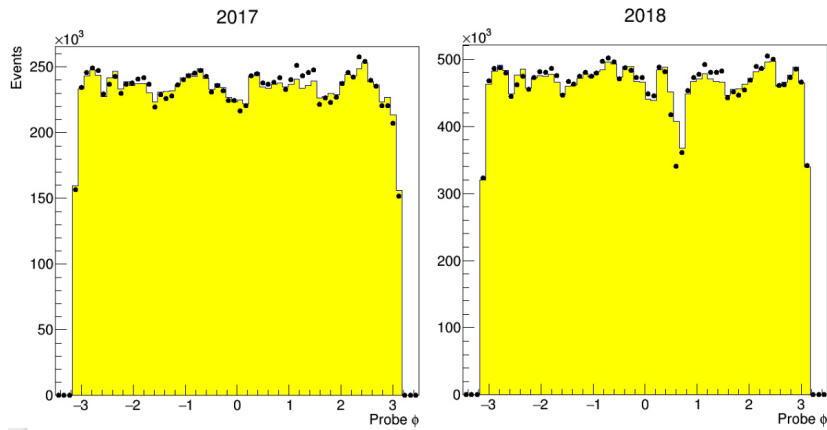


FIGURE 7.38 – Distribution en  $\phi$  pour des électrons "probes" obtenue avec la méthode Tag and Probe pour 2017 (gauche) et 2018 (droite). La distribution pour les données est représentée par des points noirs, celle des simulations est en jaune plein.

donc des électrons passant le veto électron.

La population de vrais électrons, ou *zero-fake*, étant bien plus importante que celle des *double-fake*, une contamination, même faible, de ces événements peut tout à fait mener à une augmentation du nombre d'évènements *double-fake* par un facteur d'environ 2. Ainsi une baisse d'efficacité  $\sim 0,8\%$  du détecteur à pixels conduit à une augmentation du nombre d'évènement DY par un facteur 1,8 expliquant donc l'augmentation du nombre d'évènement *double-fake* en 2018 par rapport à 2017.

Ces inefficacités en  $\phi$  sont bien simulées dans les MC et leur correspondance avec les données est bonne. Néanmoins, elles autorisent de vrais électrons à contaminer notre analyse ce qui est problématique. J'ai néanmoins effectué une étude montrant que la zone d'inefficacité  $\phi \sim 0,7$  n'induit pas un excès d'évènements ressemblant à du signal. Cependant les variations locales d'efficacité du détecteur à pixels peuvent aussi induire des contaminations, plus légères, de l'analyse par de vrais électrons. Une étude est alors en cours pour essayer de rejeter spécifiquement les électrons, incluant les électrons *double-fake*, de l'analyse en utilisant une méthode d'analyse multivariée.

## Conclusion

On a donc présenté ici quelques études sur la modélisation des composantes Drell-Yan. Ces études ont été effectuées dans le cas bruit de fond seul, or il est indispensable de pouvoir procéder à des études s+b après ajustement afin d'aboutir à une conclusion. Différentes études ont ainsi été réalisées avec les résultats observés *post-fit* mais elles ne seront pas abordées ici.

On a vu que les accords entre les simulations et les données pour les événements *double-fake* sont quelque peu difficiles à établir, bien qu'on puisse considérer des incertitudes systématiques supplémentaires. De plus, on a aussi étudié que de vrais électrons pouvaient contaminer notre



analyse. Une solution à ces problèmes serait alors d'étudier la faisabilité d'un rejet total des électrons de l'analyse incluant les événements *double-fake* et ainsi ne plus à avoir à modéliser la composante DY quitte à restreindre les classes du *diphoton BDT*.



# Études phénoménologiques

## Sommaire

---

8.1	Recherche de désintégrations exotiques au LHC $h \rightarrow Za$ . . . . .	236
8.1.1	Introduction . . . . .	236
8.1.2	Désintégration $a \rightarrow \gamma\gamma$ . . . . .	237
8.1.3	Désintégration $a \rightarrow \mu\mu$ . . . . .	240
8.1.4	Interprétations en terme d'un modèle de particules axion-like . . . . .	241
	Cadre du modèle . . . . .	241
	Cas $C_f = 0$ . . . . .	242
	Cas $C_\gamma = 0$ . . . . .	243
8.2	Recherche de nouveaux scalaires au LHC par fusion de bosons vecteur . . . . .	245
8.2.1	Introduction . . . . .	245
8.2.2	Modèle effectif . . . . .	246
8.2.3	Distributions cinématiques simulées . . . . .	247
8.3	Recherche d'une nouvelle particule via un couplage $h\eta\gamma$ au HL-LHC . . . . .	250
8.3.1	Cadre du modèle . . . . .	250
8.3.2	Proto-analyse avec simulations . . . . .	251
8.3.3	Cas des collisionneurs $e^+e^-$ . . . . .	254

---

Les résultats de l'analyse étant dérivés dans le cas d'un signal *SM-like*, il est possible de les réinterpréter dans le cadre de différents modèles théoriques BSM présentant des bosons de Higgs additionnels. J'ai eu l'occasion de travailler, en collaboration avec différents théoriciens et théoriciennes, sur des études phénoménologiques de tels modèles présentant une nouvelle particule similaire à celle recherchée dans l'analyse décrite dans cette thèse mais dans un cadre plus général.

## 8.1 Recherche de désintégrations exotiques au LHC $h \rightarrow Za$

On étudie ici la désintégration exotique du boson de Higgs à 125 GeV en un pseudo-scalaire BSM léger  $a$ . Cette désintégration est motivée par un grand nombre de modèles physiques, des modèles d'axions aux modèles avec des bosons de Higgs supplémentaires. On étudie dans cette section la sensibilité du LHC à détecter ce type de désintégration pour différentes désintégrations du pseudo-scalaire  $a$  notamment  $a \rightarrow \gamma\gamma$  et  $a \rightarrow \mu\mu$ . On propose aussi une interprétation des résultats obtenus en terme d'un modèle d'axion. Ce travail est issu d'une collaboration [124] débutée lors de l'atelier "Physics at TeV Colliders" organisé à l'École de Physique des Houches.

### 8.1.1 Introduction

S'il est toujours possible de mettre en évidence des désintégrations exotiques du boson de Higgs par des mesures très précises de ses propriétés, les recherches directes ciblant des désintégrations BSM du boson de Higgs de masse 125 GeV donnent potentiellement accès à des rapports de branchement bien plus faibles et permettent donc de découvrir de la physique BSM à basse masse. On s'intéresse ici à une désintégration exotique du boson de Higgs en un pseudo-scalaire léger soit  $h \rightarrow Za$  (étant ici dans le cadre théorie BSM le boson de Higgs à une masse de 125 GeV est noté  $h$ ).

Ce type de désintégration est naturellement présent dans des extensions du MS comme les modèles de particules de type axion (ALP) [125, 126, 127, 128, 129], ou encore avec des secteurs de Higgs non-minimaux contenant de nouveaux (pseudo)scalaires [130, 131, 132]. La capacité du LHC à mettre en évidence la désintégration  $h \rightarrow Za$  a été étudiée par Bauer et al. dans [127] en considérant  $a$  comme une ALP se désintégrant en deux photons ou deux électrons. Dans cette continuité, on présente ici pour la première fois une analyse détaillée de la sensibilité du LHC à cette désintégration exotique  $h \rightarrow Za$  où l'état final est  $Z \rightarrow \mu\mu$  et  $a \rightarrow \gamma\gamma$  pour lequel aucune analyse de recherche directe d'ATLAS ou de CMS n'existe (bien que cet état final fut mesuré dans des collisions à 8 TeV [133] mais sans contrainte sur la masse du boson de Higgs). Les désintégrations leptoniques  $a \rightarrow \mu\mu$  et  $a \rightarrow \tau\tau$  (cette dernière ne sera cependant pas présentée ici) ont aussi été considérées et les limites sur les rapports de branchement correspondants ont été dérivées en réinterprétant des analyses ATLAS et CMS déjà existantes.

Dans tous les cas, on se limite à des masses  $m_a < 34$  GeV pour lesquelles le boson de Higgs se désintègre *on-shell*. On présentera finalement les limites d'exclusion des rapports

de branchement  $BR(h \rightarrow Za) \times BR(a \rightarrow \gamma\gamma)$ ,  $BR(h \rightarrow Za) \times BR(a \rightarrow \mu\mu)$  et  $BR(h \rightarrow Za) \times BR(a \rightarrow \tau\tau)$  ainsi qu'une interprétation de ces résultats dans un modèle ALP effectif.

### 8.1.2 Désintégration $a \rightarrow \gamma\gamma$

La désintégration de  $a$  en deux photons permet un état final avec une signature claire et propre en combinaison avec les désintégrations leptoniques venant du boson  $Z$  pouvant d'ailleurs être utilisées, d'un point de vue expérimental, pour le déclenchement. Selon la masse  $m_a$  du pseudo-scalaire et de la cinématique des évènements, les deux photons peuvent apparaître comme distincts ou bien n'être reconstruits que comme un seul objet<sup>1</sup>. Cette distinction dans l'état final obtenu s'effectue via une valeur minimale de la séparation angulaire  $\Delta R$  déterminée par la résolution angulaire du ECAL. Si l'on considère que la majorité des photons se situe dans le tonneau, en prenant en compte le fait que plus de 97 % de l'énergie des photons est contenue dans une matrice  $5 \times 5$  et en estimant la résolution en  $\eta/\phi$  à environ 0,0175, on peut simplement estimer cette valeur minimale de  $\Delta R \sim 5 \times 0,0175 \simeq 0,09$ .

Pour simplifier, on considère<sup>2</sup> alors que si  $\Delta R_{\gamma,\gamma} < 0,1$  les super-agrégats ECAL sont indiscernables et on ne considère alors qu'il n'y a qu'un seul photon reconstruit dans l'état final, bien qu'il soit expérimentalement possible de distinguer les photons en étudiant les formes de gerbes associées. Une valeur minimale similaire sur la séparation angulaire permettant de distinguer deux photons peut être estimée pour ATLAS [134].

On commence donc notre analyse en étudiant la séparation angulaire entre les photons pour le processus  $pp \rightarrow h \rightarrow Za, a \rightarrow \gamma\gamma$  où le boson de Higgs est produit par le mode  $ggH$  à une énergie dans le centre de masse de 13 TeV en fonction de  $m_a$ . La distribution obtenue est visible sur la figure 8.1 pour des masses  $m_a \in [1, 30]$  GeV. Pour  $m_a \lesssim 1$  GeV, la plupart des évènements ont une valeur de  $\Delta R_{\gamma\gamma} < 0,1$  et ne sont pas considérés comme des évènements avec un état final diphoton.

Pour des masses  $m_a \ll 1$  GeV, on considère de la même façon que les deux photons sont reconstruits en tant qu'unique photon, ce qui nous permet de réinterpréter les limites d'exclusion du rapport de branchement  $h \rightarrow Z\gamma$  des analyses ATLAS et CMS à 13 TeV existantes [135, 136]. On utilise pour ce faire l'analyse ATLAS possédant une luminosité intégrée de  $63,1 \text{ fb}^{-1}$  qui nous permet de poser la contrainte  $BR(h \rightarrow Za) \times BR(a \rightarrow \gamma\gamma) < 0,01$ . L'extrapolation de cette limite pour des masses proches de 1 GeV est faite par une simple pondération dépendant de la fraction d'évènements signal satisfaisant la condition  $\Delta R_{\gamma\gamma} < 0,1$  en fonction de  $m_a$  (par exemple pour  $m_a = 500$  MeV la fraction est de 0,83, alors qu'elle est de 0,008 pour  $m_a = 2$  GeV). Une méthode plus réaliste, mais hors de portée des travaux présentés ici, serait de prendre en compte les procédés de reconstruction et d'identification des photons pour ATLAS et CMS.

En ce qui concerne maintenant l'analyse  $pp \rightarrow h \rightarrow Za, a \rightarrow \gamma\gamma$  où les deux photons sont résolus, on simule des évènements de signal en utilisant un modèle Universal FeynRules Output (UFO) adapté de [127] à 13 TeV avec MadGraph, PYTHIA et DELPHES [137] pour

1. Cela arrive notamment lorsque  $a$  est cinématiquement boosté.

2. Pour le cas de photons convertis, la valeur minimale à considérer doit sûrement être plus élevée.

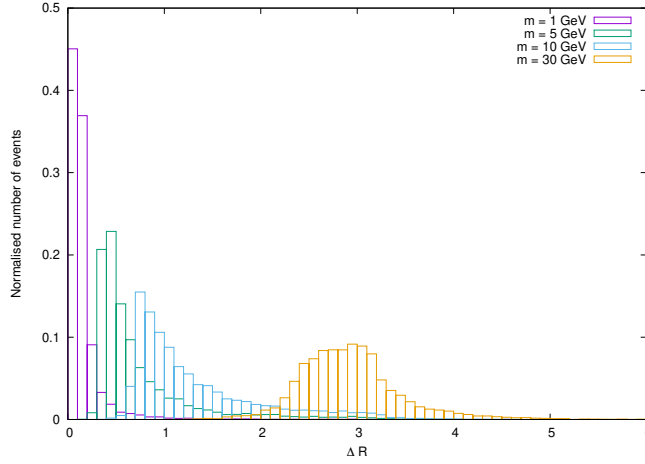


FIGURE 8.1 – Distribution de la séparation angulaire entre les deux photons  $\Delta R_{\gamma,\gamma}$  de l'état final pour les événements simulés  $pp \rightarrow h \rightarrow Za, a \rightarrow \gamma\gamma$  en fonction de la masse  $m_a$ .

la simulation du détecteur<sup>3</sup>. La valeur de section efficace pour la production du boson de Higgs est la valeur N<sup>3</sup>LO venant de [138] correspondant à une valeur de 48,58 pb pour le mode de production ggH dans le cas MS. Dans DELPHES, on utilise une isolation angulaire de  $\Delta R = 0,5$  pour les photons qui cause alors une forte dégradation de l'efficacité du signal pour des masses  $m_a \lesssim 5$  GeV, comme visible sur la figure 8.1. On considère donc des masses situées dans l'intervalle  $m_a \in [5, 34]$  GeV.

Pour le bruit de fond principal du MS, on considère le bruit de fond irréductible  $pp \rightarrow \mu\mu\gamma\gamma$  (dont la section efficace à 13 TeV est de 0,109 pb) et le bruit de fond réductible, mais important,  $pp \rightarrow jj\mu\mu$  venant notamment de processus  $Z$ + jets où les deux jets sont identifiés à tort comme des photons (désintégrations de mésons neutres par exemple). On applique des facteurs prenant en compte le taux de fausse identification  $jet \rightarrow \gamma$  estimés<sup>4</sup> en fonction du  $p_T$  du jet dans [139].

On réalise ensuite une série de coupure afin de sélectionner un maximum d'événements signal et un minimum d'événements bruit de fond :

- On requiert deux photons isolés avec  $p_T^\gamma > 10$  GeV et  $|\eta_\gamma| < 2,5$ , ainsi que deux muons de charge opposée avec  $p_T^\mu > 10$  GeV et  $|\eta_\mu| < 2,4$ .
- La masse invariante de deux muons doit satisfaire  $m_{\mu\mu} \in [75, 105]$  GeV. Et de la même manière, les deux photons et les deux muons doivent reconstruire la masse du boson de Higgs soit  $m_{\mu\mu\gamma\gamma} \in [115, 135]$  GeV.
- Les deux photons doivent satisfaire  $m_{\gamma\gamma} + p_T^{\gamma\gamma} < m_h/2$ . Cette condition permet de supprimer la majeure partie du bruit de fond réductible  $Z$ + jets tout en maintenant une haute efficacité pour le signal.

3. On choisit une "saveur" correspondant au détecteur CMS.

4. On peut souligner que pour des jets avec une impulsion transverse faible  $p_T \lesssim 30$  GeV, l'incertitude sur les taux  $jet \rightarrow \gamma$  devient très importante. Elle n'est cependant pas prise en compte alors que cette région de faible  $p_T$  est une région d'intérêt de l'analyse présentée ici, les limites présentées ci-après sont donc des résultats optimistes.

On définit ensuite une région de signal concernant la variable  $m_{\gamma\gamma}$ , pour les évènements qui passent la sélection, dans laquelle on cherche un excès de signal dans un bruit de fond défini dans une fenêtre de  $\pm 5$  GeV centrée autour de la masse  $m_a$  considérée.

On calcule finalement le nombre d'évènements signal et bruit de fond MS attendus (pour  $BR(h \rightarrow Za) \times BR(a \rightarrow \gamma\gamma) = 1$ ) pour une luminosité de  $36.1 \text{ fb}^{-1}$  dans la région signal en fonction de la masse du signal  $m_a$  considérée. A partir de ce nombre d'évènements attendus, on calcule les limites d'exclusion à 95 % du produit  $BR(h \rightarrow Za) \times BR(a \rightarrow \gamma\gamma)$  en fonction de la masse  $m_a$  pour l'analyse décrite précédemment avec les deux photons résolus. On utilise pour cela une méthode basée sur un rapport de vraisemblance qui correspond à la méthode présentée dans la section 6.11 où le POI est ici  $\mu = BR(h \rightarrow Za) \times BR(a \rightarrow \gamma\gamma)$ .

Les limites obtenues sont visibles en vert sur la figure 8.2. Il est cependant important de noter que ces résultats ne prennent aucune systématique (ou NP) en compte ; en particulier il faudrait prendre en compte des incertitudes liées à la faible statistique de nos évènements bruit de fond et aux facteurs de fausses identifications  $j \rightarrow \gamma$ . Ces limites sont donc une première approximation qu'il faudrait développer.

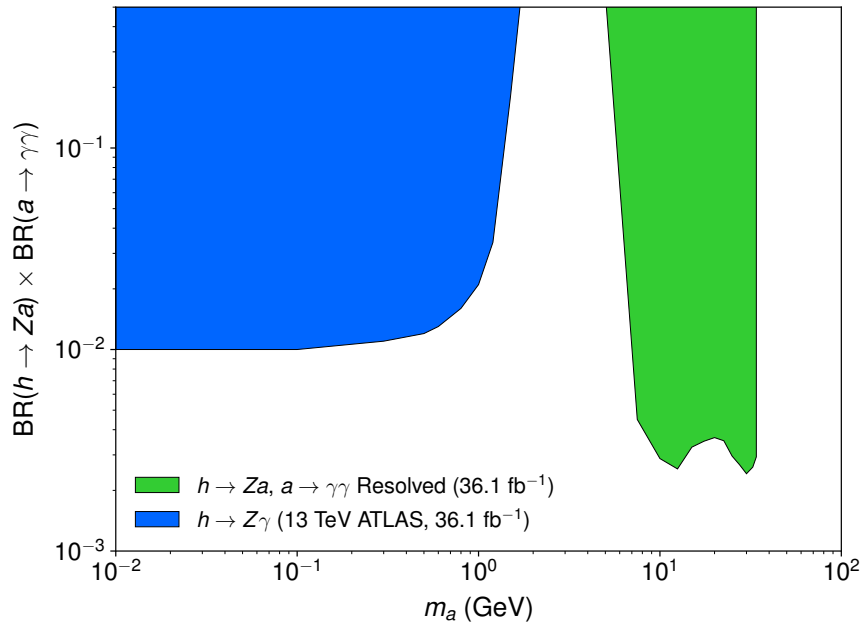


FIGURE 8.2 – Limites d'exclusion à 95 % sur  $BR(h \rightarrow Za) \times BR(a \rightarrow \gamma\gamma)$  dans notre analyse diphotons résolus (en vert) et dans le cas  $h \rightarrow Z\gamma$  où les deux photons ne sont pas discernables (en bleu) au LHC à 13 TeV pour une luminosité de  $36.1 \text{ fb}^{-1}$ .

La figure 8.2 montre aussi les limites d'exclusion attendues sur  $BR(h \rightarrow Za) \times BR(a \rightarrow \gamma\gamma)$  en fonction de  $m_a$  venant de la réinterprétation des résultats  $h \rightarrow Z\gamma$  de ATLAS [135] en bleu. On montre ainsi que, pour  $BR(a \rightarrow \gamma\gamma) = 1$ , des valeurs de rapport du branchement exotique du boson de Higgs peuvent être mesurées au LHC jusqu'à  $BR(h \rightarrow Za) \sim 0,01$  pour des  $m_a < 1$  GeV alors que pour le cas deux photons résolus,  $10 \text{ GeV} \lesssim m_a \lesssim 35$ , la sensibilité atteint une valeur  $BR(h \rightarrow Za) \sim 0,003$ . On rappelle finalement que le creux de masse  $m_a \in [1, 10]$  GeV pourrait être significativement réduit en abaissant à une valeur de  $\Delta R \sim 0,2$  le critère d'isolation angulaire des photons utilisé ici.

### 8.1.3 Désintégration $a \rightarrow \mu\mu$

Contrairement au cas précédent, il existe des analyses ATLAS et CMS ciblant les états finaux avec désintégrations leptoniques  $a \rightarrow \ell\ell$  ( $\ell = e, \mu$ ) et il est donc possible d'obtenir des limites d'exclusion sur  $BR(h \rightarrow Za) \times BR(a \rightarrow \ell\ell)$  en réinterprétant ces analyses. Dans la suite on se concentre sur la désintégration  $a \rightarrow \mu\mu$ .

Parmi la multitude d'analyses existante avec cet état final, l'analyse d'ATLAS à 13 TeV qui recherche des désintégrations exotiques du boson de Higgs dans un canal  $h \rightarrow ZZ_D \rightarrow 4\ell$  (où  $Z_D$  est un boson de jauge BSM) [140] a l'avantage de demander un boson  $Z$  reconstruit en plus d'une résonance de masse variable. Les autres analyses considèrent, au contraire, en général deux résonances de même masse et leur sélection en masse rejeterait donc une bonne partie de notre signal.

On réinterprète donc dans la suite ces travaux d'ATLAS [140] pour dériver la sensibilité du LHC à des processus  $pp \rightarrow h \rightarrow Za$  ( $Z \rightarrow \ell\ell, a \rightarrow \mu\mu$ ). Pour valider notre analyse et notamment comparer le nombre d'évènements signal attendu, on considère des simulations partoniques<sup>5</sup> de bruit de fond le plus important  $h \rightarrow ZZ^* \rightarrow 4\ell$  que l'on souhaite comparer à la distribution de variables cinématiques  $m_{34}$  publiées par ATLAS (figure 2 de [140]); on requiert pour cela une série de coupures :

- Deux paires de leptons de même saveur, de signes opposés,  $\ell_i$ .
- $p_T(\ell_1, \ell_2, \ell_3) > 20, 15, 10$  GeV.
- $\Delta R(\ell, \ell') > 0, 1$  (0,2) pour des leptons de même saveur ou de saveur différente.
- $115 \text{ GeV} < m_{4\ell} < 130 \text{ GeV}$ .
- $m_{i,j} > 5 \text{ GeV}$ .
- $50 \text{ GeV} < m_{1,2} < 106 \text{ GeV}$  et  $12 \text{ GeV} < m_{3,4} < 115 \text{ GeV}$ .

avec  $i, j = 1, 2, 3, 4$  les indices des leptons. Les indices 1 et 2 correspondent aux indices utilisés pour les leptons qui reconstruisent la masse du boson  $Z$ , et les indices 3 et 4 correspondent à la paire de leptons qui ne reconstruit pas la masse du boson  $Z$ .

Après ces coupures, l'analyse ATLAS montre la distribution de masse  $m_{34}$  pour des gammes  $m_{34} \in [11, 59]$  GeV. La comparaison obtenue avec nos simulations est présentée à gauche de la figure 8.3 où l'on a appliqué une simple efficacité de 0,9 aux leptons reconstruits pour tenir compte des effets de détecteur non inclus. Les résultats obtenus sont raisonnablement semblables.

Pour déterminer les limites sur  $BR(h \rightarrow Za) \times BR(a \rightarrow \mu\mu)$ , on extrait les limites d'exclusion à  $2\sigma$  de l'analyse ATLAS que l'on pondère par la fraction signal attendue dans notre analyse par rapport à celle attendue dans l'analyse ATLAS. Le signal est simulé de la même manière que dans la section précédente. Les limites d'exclusion à 95 % obtenues sur  $BR(h \rightarrow Za) \times BR(a \rightarrow$

5. Dans l'idéal il faudrait évidemment considérer une simulation totale des évènements physiques incluant hadronisation et détecteur mais cela est hors de portée de l'étude présentée ici.



$\mu\mu$ ) avec une luminosité intégrée de  $36,1 \text{ fb}^{-1}$  en fonction de  $m_a$  sont visibles sur la figure droite de 8.3. Il est en principe possible de couvrir des masses  $m_a > 35 \text{ GeV}$  ici mais, comme précédemment évoqué, on se restreint à des désintégrations *on-shell* de  $a$ .

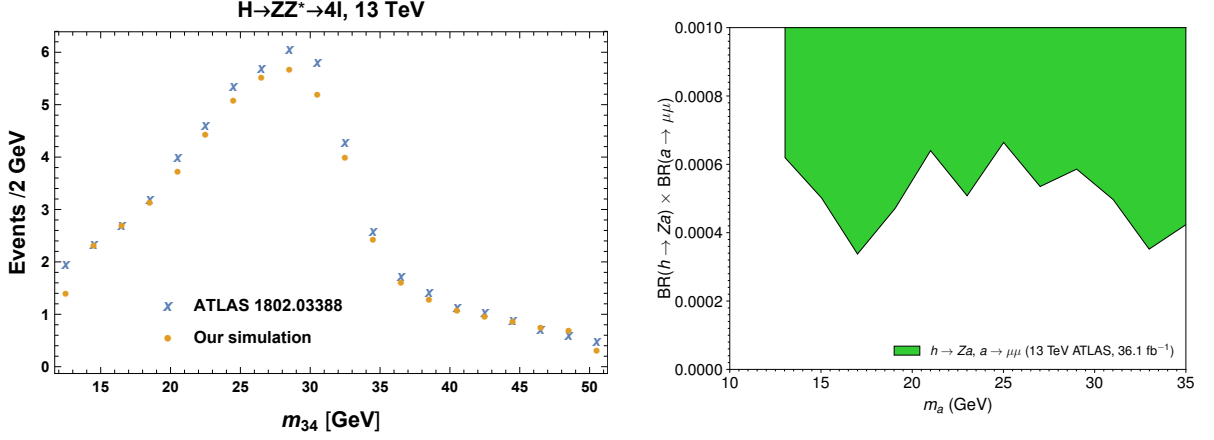


FIGURE 8.3 – A gauche, comparaison de la distribution de la variable  $m_{34}$  du bruit de fond MS pour des processus  $H \rightarrow ZZ^* \rightarrow 4\ell$  entre ATLAS [140] (croix bleues) et nos simulations (cercles oranges). A droite, limites d'exclusion à 95 % sur les valeurs de  $BR(h \rightarrow Za) \times BR(a \rightarrow \mu\mu)$  en fonction de  $m_a$ .

### 8.1.4 Interprétations en terme d'un modèle de particule axion-like

#### Cadre du modèle

Les études présentées précédemment étant relativement indépendantes du modèle, il peut être intéressant de les interpréter en terme d'une théorie BSM particulière afin de permettre de contraindre les différents paramètres de ces modèles. On considère ainsi dans la suite une simple interprétation en terme d'un modèle de théorie de champs effective (EFT) en considérant le cas MS étendu par l'hypothèse d'une particule *axion-like*  $a$  [125, 126, 127, 129]. Ce type d'extension ALP est inspiré des modèles d'axions qui visent notamment à expliquer l'asymétrie CP en QCD et à présenter un candidat au problème de matière noire froide.

En toutes généralités, les champs ALP sont supposés se coupler à la fois aux bosons de jauge et aux fermions du MS. On modifie alors le lagrangien du MS, en incluant uniquement jusqu'aux termes d'ordre 5 dans un développement EFT linéaire, comme suit :

$$L_{eff}^{d \leq 5} \supset \mathcal{L}_{SM} + \frac{1}{2} \partial_\mu a \partial^\mu a - \frac{1}{2} m_a^2 a^2 + \sum_f \frac{m_f}{\Lambda} C_f a \bar{f} \gamma_5 f - \frac{e^2}{4\Lambda} C_\gamma a F_{\mu\nu} \tilde{F}_{\mu\nu}. \quad (8.1)$$

où  $a$  est le champ associé à la particule *axion-like*,  $f$  dénote les fermions du MS et les champs associés,  $F_{\mu\nu}$  est le terme usuel de jauge pour les photons,  $\Lambda$  (de l'ordre de 1 TeV ici) dénote l'échelle d'énergie de la symétrie considérée, et  $C_\gamma$  est le couplage de l'ALP au photon et  $C_f$  son couplage avec les fermions. En principe l'ALP peut aussi se coupler aux autres bosons de jauge selon des couplages  $aVV$  (ceci est notamment discuté dans [141]) mais ces termes ne sont pas pertinents pour les études présentées ici.

Notons que les couplages de l'ALP avec le boson de Higgs ne sont pas présents dans le

lagrangien précédent puisque ces termes de couplage peuvent être réécrits [125, 126, 127] en fonction des opérateurs ALP-fermions en utilisant les équations du mouvement dans le développement EFT à l'ordre 5 présenté ici. Ces termes de couplage, nécessaires pour l'étape de la désintégration  $h \rightarrow Za$ , apparaissent de deux façons différentes<sup>6</sup> dans le développement EFT ; la première consiste à faire intervenir des termes d'ordres supérieurs prenant en compte l'interaction effective dans le lagrangien :

$$L_{eff}^{d=7} \supset \frac{1}{\Lambda^3} C_{Zh}^{(7)} (\partial^\mu a) \left( \phi^\dagger i \overleftrightarrow{D}_\mu \phi + \text{h.c.} \right) \phi^\dagger \phi + \dots, \quad (8.2)$$

où  $C_{Zh}^{(7)}$  est le couplage effectif entre l'ALP, le boson de Higgs et le boson Z [128]. La seconde façon de faire apparaître ce terme est de prendre en compte les termes de boucles, notamment venant du quark top, comme détaillé dans [128]. Le couplage effectif total entre l'ALP, le boson de Higgs et le boson Z peut alors être écrit sous la forme :

$$C_{Zh}^{eff} = -\frac{N_c y_t^2}{16\pi^2} C_t F + \frac{v^2}{2\Lambda^2} C_{Zh}^{(7)}. \quad (8.3)$$

avec  $C_t$  le couplage entre  $a$  et le top,  $N_c$  étant le nombre de couleur considéré et  $F$  est une fonction [128] qui dépend des masses du top, du boson de Higgs, du boson Z et de l'ALP. On note que pour  $\Lambda = 1$  TeV la valeur des coefficients  $C_t$  et  $C_{Zh}^{(7)}$  est comparable.

On peut maintenant exprimer la largeur de désintégration du boson de Higgs en  $Za$ , pour  $m_a < m_h - m_Z$ , en fonction de ce couplage effectif  $C_{Zh}^{eff}$  :

$$\Gamma_{h \rightarrow Za} = \frac{m_h^3}{16\pi\Lambda^2} \left| C_{Zh}^{eff} \right|^2 \lambda^{3/2} \left( \frac{m_Z^2}{m_h^2}, \frac{m_a^2}{m_h^2} \right). \quad (8.4)$$

avec  $\lambda(x, y) = (1 - x - y)^2 - 4xy$ .

Maintenant que l'expression de la largeur de désintégration  $h \rightarrow Za$  a été établie, on veut déterminer les différentes désintégrations possibles de  $a$  dans ce modèle. Étant donné que l'on considère ici des ALPs à une échelle d'énergie en dessous de l'échelle électrofaible (et même en dessous de celle du Z), les désintégrations ALP *on-shell* en bosons électrofaibles ne sont pas permises ; il reste donc des désintégrations en paire de fermions, photons et gluons.

Dans la suite, on définit alors deux modèles de référence pour lesquels on souhaite déterminer la sensibilité attendue au LHC. Dans le premier, on néglige le couplage entre  $a$  et les fermions en posant  $C_f = 0$ . Dans le second, on choisit la situation opposée où l'on considère que le couplage entre l'ALP et les photons est nul  $C_\gamma = 0$ .

### Cas $C_f = 0$

Dans le cas  $C_f = 0$ , on considère que, dans la gamme cinématique considérée, le rapport de branchement du ALP en photons est de 100 %. Par conséquent, les limites, déterminées précédemment, pour l'état final  $h \rightarrow Za \rightarrow \gamma\gamma$  permettent directement de contraindre le coefficient de Wilson  $C_{Zh}^{eff}$  qui vaut  $v^2/(2\Lambda^2) C_{Zh}^{(7)}$  ici. Notons que la seule hypothèse faite sur

6. Pour d'autres méthodes de développement EFT, comme par exemple le cas chiral, le couplage ALP- $h$ -Z apparaît à des ordres de développement inférieurs [126].

$C_\gamma$  est qu'il doit être suffisamment élevé afin que  $a$  se désintègre à l'intérieur du détecteur. Les résultats obtenus sont visibles sur la figure 8.4. On peut exclure des valeurs de  $C_{Zh}^{\text{eff}}$  jusqu'à  $\sim 10^{-1}$  pour des faibles valeurs de  $m_a$  correspondant au cas où les deux photons ne sont pas résolus, et jusqu'à  $\sim 5 \cdot 10^{-2}$  pour  $m_a = 10$  GeV dans le cas où les deux photons sont résolus, sous l'hypothèse  $\Lambda = 1$  TeV. Notons que le fait que  $C_\gamma$  ne soit pas fixé constitue une différence avec [128].

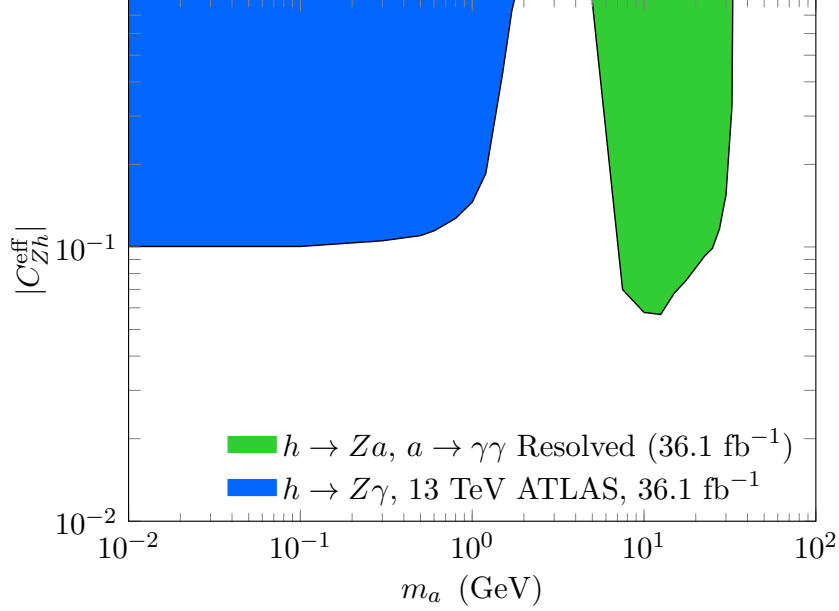


FIGURE 8.4 – Limites d'exclusion à 95% sur le coefficient de Wilson  $|C_{Zh}^{\text{eff}}|$  pour  $C_f = 0$  obtenues à partir des limites  $BR(h \rightarrow Za) \times BR(a \rightarrow \gamma\gamma)$  déterminées précédemment où les deux photons sont résolus (vert) et dans le cas  $h \rightarrow Z\gamma$  où les photons ne peuvent être séparés dans le détecteur (bleu) pour une luminosité intégrée de  $36,1 \text{ fb}^{-1}$  au LHC et  $\Lambda = 1$  TeV.

### Cas $C_\gamma = 0$

Le second cas abordé est celui où le couplage avec les bosons de jauge disparaît. Dans ce cas, la largeur de désintégration de  $a$  en leptons est donnée par [128] :

$$\Gamma(a \rightarrow \ell^+ \ell^-) = \frac{m_a m_\ell^2}{8\pi\Lambda^2} |C_\ell^{\text{eff}}|^2 \sqrt{1 - \frac{4m_\ell^2}{m_a^2}}. \quad (8.5)$$

On considère alors deux possibilités : a) le cas où tous les  $C_f$  sont égaux et b) le cas où seuls les couplages aux leptons muon ou tau sont présents, le couplage des autres fermions avec l'ALP étant nul.

Dans le scénario a), on fait l'hypothèse que la désintégration  $h \rightarrow Za$  s'effectue par une boucle de tops et qu'alors on a  $C_{Zh}^{(7)} = 0$ . Pour calculer le rapport de branchement pour la désintégration de  $a$ , il faut alors prendre en compte des modes de désintégration supplémentaires, notamment en prenant en compte les désintégrations en hadrons et en quarks lourds. Ces taux de désintégration sont donnés en fonction de  $C_f$  dans [128].

Dans le scénario b), on fait l'hypothèse que la contribution  $C_{Zh}^{(7)}$  dans la largeur de désintégration  $h \rightarrow Za$  est dominante et que le rapport de branchement de  $a$  en muons ou taus est de

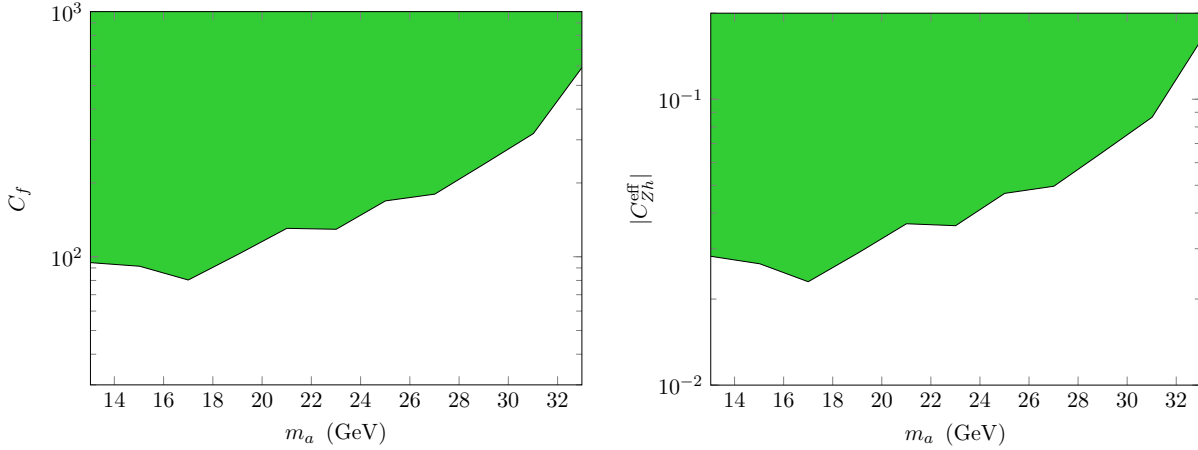


FIGURE 8.5 – Limites d’exclusion à 95% sur le coefficient de Wilson  $C_f$  sous l’hypothèse d’un couplage fermionique universel (gauche) et  $|C_{Zh}^{\text{eff}}|$  sous l’hypothèse d’une désintégration totalement muonique (droit) pour  $C_\gamma = 0$  obtenues à partir des limites  $BR(h \rightarrow Za) \times BR(a \rightarrow \mu\mu)$  pour  $\Lambda = 1$  TeV.

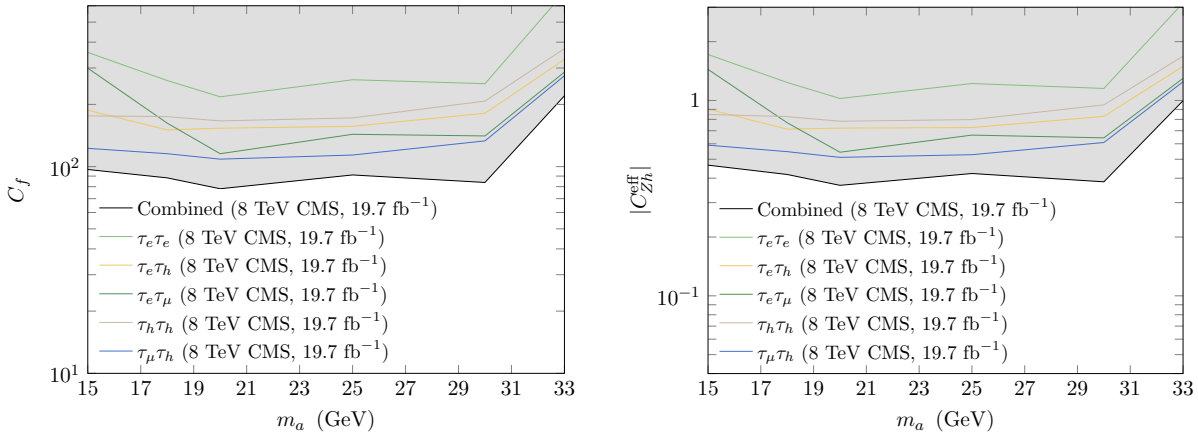


FIGURE 8.6 – Limites d’exclusion à 95% sur le coefficient de Wilson  $C_f$  sous l’hypothèse d’un couplage fermionique universel (gauche) et  $|C_{Zh}^{\text{eff}}|$  sous l’hypothèse d’une désintégration en tau totale (droit) pour  $C_\gamma = 0$  pour  $\Lambda = 1$  TeV d’après les limites d’exclusion pour des désintégrations  $a \rightarrow \tau\tau$  déterminées dans [124].

100 %. A nouveau, la seule hypothèse sur  $C_\mu$  et  $C_\tau$  est qu’ils doivent être assez importants afin de permettre une désintégration de  $a$  au sein du détecteur.

Les résultats obtenus sont présentés sur les figures 8.5 et 8.6. On observe que dans le cas a) la plus petite valeur exclue est  $C_f \sim 80$  pour des masses  $\sim 15 - 30$  GeV pour une désintégration en muons et en taus avec  $\Lambda = 1$  TeV. Dans le cas b) la contrainte sur  $C_{Zh}^{\text{eff}}$  est plus forte pour l’état final dimuon ( $\sim 2 \times 10^{-2}$ ) que pour l’état final ditau ( $\sim 4 \times 10^{-1}$ ). De manière générale, on observe que les couplages  $C_f$  ne sont pas contraints pour des valeurs faibles, ou inversement que de faibles échelles d’énergie sont toujours largement autorisées.

Pour conclure, on a étudié une désintégration exotique du boson de Higgs de masse 125 GeV en un boson Z et un pseudo-scalaire  $a$  BSM pour différents modes de désintégration de  $a$ . Pour

le canal de désintégration diphoton, on a pu contraindre  $\text{BR}(h \rightarrow Za) \times \text{BR}(a \rightarrow \gamma\gamma) \sim 0,01$  pour des masses inférieures à 1 GeV, où les deux photons sont très collimatés et où l'on considère qu'ils apparaissent comme un unique photon dans le détecteur, et  $\text{BR}(h \rightarrow Za) \times \text{BR}(a \rightarrow \mu\mu)$  à une gamme  $\sim (4 - 6) \times 10^{-4}$  pour  $m_a \in [10, 35]$  GeV pour des photons résolus. Pour le canal de désintégration dimuon, une réinterprétation d'une analyse d'ATLAS permet de contraindre  $\text{BR}(h \rightarrow Za) \times \text{BR}(a \rightarrow \mu\mu)$  à une gamme  $\sim (4 - 6) \times 10^{-4}$  pour  $m_a \in [10, 35]$  GeV.

Ces contraintes sur le rapport de branchement de cette désintégration exotique ont ensuite été interprétées pour un modèle d'ALP. En considérant la seule désintégration  $a \rightarrow \gamma\gamma$ , on contraint le paramètre effectif  $C_{Zh}^{\text{eff}} \lesssim 0,1 \text{ TeV}^{-1}$  pour une luminosité de  $36,1 \text{ fb}^{-1}$  au LHC à 13 TeV. Pour des désintégrations  $a \rightarrow \mu\mu$  ( $a \rightarrow \tau\tau$ ) seules, on pose les contraintes  $C_{Zh}^{\text{eff}} \lesssim 0,03 \text{ TeV}^{-1}$  ( $0,4 \text{ TeV}^{-1}$ ).

## 8.2 Recherche de nouveaux scalaires au LHC par fusion de bosons vecteur

On étudie ici la production d'une particule scalaire  $\phi$  par le mode VBF se désintégrant en deux photons. Ce mode de production présente en effet l'avantage d'avoir deux jets supplémentaires dans l'état final permettant d'aider à l'identification du signal et de considérer des désintégrations de particule  $\phi$  à basse masse avec un déclenchement sur ces deux jets. A nouveau l'existence de ce type de particule est motivée par un grand nombre de modèles et peut correspondre à un boson de Higgs supplémentaire, un radion, un dilaton ou bien un axion. On étudie dans cette section la possibilité de différencier ce type de (pseudo)scalaire additionnel du bruit de fond du MS au LHC. Ce travail est aussi issu d'une collaboration [124] débutée lors de l'atelier "Physics at TeV Colliders" organisé à l'École de Physique des Houches.

### 8.2.1 Introduction

Pour le cas d'un boson de Higgs *SM-like*, le second mode de production en importance, après le mode par fusion de gluons, est le mode de production par fusion de bosons vecteurs VBF au LHC pour une énergie dans le centre de masse de 13 TeV pour des masses hypothétiques allant de 100 GeV à 1 TeV. Le mode de production VBF est donc aussi un mode de production crucial pour la recherche de particule (pseudo)scalaire. De plus, selon le modèle considéré pour cette particule, la contribution du photon au processus VBF total peut être compétitive avec celle des bosons Z et  $W^\pm$  puisqu'un couplage entre le scalaire additionnel et les photons (ainsi qu'avec les bosons Z et  $W^\pm$ ) peut apparaître au niveau de l'arbre<sup>7</sup>. Une telle contribution permet en général d'augmenter la section efficace VBF par rapport à la section efficace par fusion de gluons. Il est aussi possible que le couplage Yukawa du top soit réduit par rapport au MS ce qui favoriserait le mode de production VBF par rapport à ggF.

On cherche alors à établir la possibilité pour une particule scalaire générique CP-impair/paire produite par le mode VBF se désintégrant en deux photons d'être détectée au LHC à 13 TeV. On

7. C'est le cas par exemple pour le radion en dimensions supplémentaires [142, 143].

essaie pour cela de trouver des coupures cinématiques discriminantes en tirant profit des deux jets supplémentaires qui permettent notamment l'accès à des masses diphoton plus faibles : on sait en effet que pour des masses  $< 65$  GeV les canaux d'observation privilégiés que sont la désintégration en deux  $Z$  et la désintégration en deux photons sont cinétiquement fermés pour la première et difficile à analyser à cause, notamment, du système de déclenchement, comme expliqué dans la section 6.3, pour le deuxième.

### 8.2.2 Modèle effectif

Comme étudié auparavant, la production de la particule  $\phi$  considérée ici s'effectue par double radiation de photon,  $Z$ , ou  $W^\pm$  à partir de quarks. On utilise donc une approche d'EFT prenant en compte ces différents couplages. Les interactions entre un scalaire neutre CP-pair ou CP-impair  $\phi$  et les bosons de jauge électrofaibles sont décrites, avant brisure, par ce lagrangien effectif de dimension 5 :

$$\mathcal{L}_{eff} \supset \phi \left( \frac{1}{f_W} W^{\mu\nu b} W_{\mu\nu}^b + \frac{1}{f_B} B^{\mu\nu} B_{\mu\nu} + \frac{1}{f_H} |D^\mu H|^2 \right) \quad (8.6)$$

$$\mathcal{L}_{eff} \supset \phi \left( \frac{1}{\tilde{f}_W} W^{\mu\nu b} \tilde{W}_{\mu\nu}^b + \frac{1}{\tilde{f}_B} B^{\mu\nu} \tilde{B}_{\mu\nu} \right) \quad (8.7)$$

où  $\tilde{V}^{\mu\nu} = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} V_{\rho\sigma}$ ,  $H$  représente le doublet de Higgs du MS avant brisure,  $f_V$  ( $\tilde{f}_V$ ) sont les constantes liées aux échelle d'énergie considérées pour le cas CP-pair (CP-impair) et les autres notations sont usuelles.

Après brisure de symétrie, le lagrangien contient notamment des interactions  $\phi ZZ$  comme par exemple dans le cas CP-pair :

$$L_{eff} \supset \frac{1}{f_Z} \phi (Z_{\mu\nu})^2 + \frac{m_Z^2}{2f_H} \phi (Z_\mu)^2, \quad (8.8)$$

où  $f_Z^{-1} = s_w^2 f_B^{-1} + c_w^2 f_W^{-1}$  et  $s_w^2 \equiv \sin^2(\theta_w) \approx 0,23$ . La théorie effective est valide tant que les  $f$  (par exemple reliés à l'échelle de masse des théories Kaluza-Klein) sont plus importantes que les énergies typiquement considérées dans les vertex. Le couplage de  $\phi$  avec le boson du Higgs du MS est supposé faible afin que les couplages du Higgs soient compatibles avec les mesures effectuées au LHC.

Le cas d'un champ scalaire CP-pair peut être interprété comme un radion dans un modèle avec une dimension supplémentaire enroulée sur elle-même [144]. Notons que si le terme cinétique associé à la *brane* électrofaible est négligeable dans ce type de modèle, on obtient la condition  $f_W = f_B$  qui permet de faire disparaître le terme de couplage  $\phi F^{\mu\nu} Z_{\mu\nu}$  [145] ce qui est utile pour discriminer entre différents modèles [146].

Le cas d'un champ scalaire CP-impair peut correspondre à un pseudo boson de Nambu-Goldstone (voir par exemple [147] ou [148]), comme dans les modèles de Higgs composites.

### 8.2.3 Distributions cinématiques simulées

On commence par présenter la section efficace, calculée en utilisant MadGraph à partir d'un fichier UFO écrit via FeynRules [149], pour des productions VBF d'un scalaire additionnel  $pp \rightarrow \phi qq$  au LHC à 13 TeV. La variation de la section efficace en fonction de  $m_\phi$  est visible sur la figure 8.7 pour des valeurs d'échelle usuelles  $f_H = f_B = f_W = 1$  TeV ( $\tilde{f}_B = \tilde{f}_W = 1$  TeV) pour le cas CP-pair (impair).

On voit par exemple que pour une masse  $m_\phi = 200$  GeV, le nombre d'évènements attendu pour une luminosité intégrée de  $36,1 \text{ fb}^{-1}$  ( $100 \text{ fb}^{-1}$ ) est d'environ  $1,2 \times 10^5$  ( $3 \times 10^5$ ) ce qui est prometteur pour la recherche d'un tel signal, même avec un rapport de branchement de désintégration faible pour  $\phi$ .

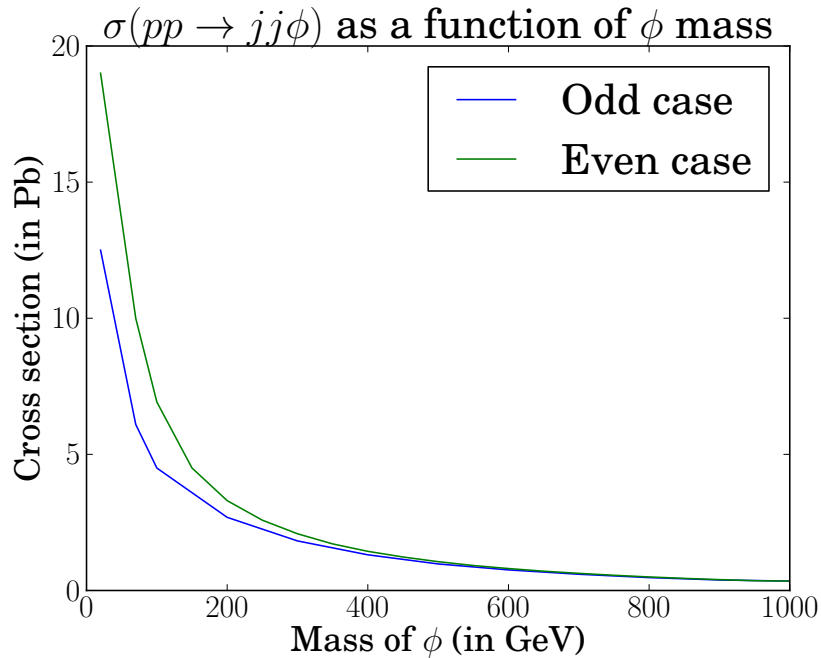


FIGURE 8.7 – Section efficace (en pb) pour la production par mode VBF d'un scalaire générique  $pp \rightarrow \phi qq$  au LHC à 13 TeV en fonction de la masse  $m_\phi$  (en GeV). Les valeurs de paramètres considérées pour le cas CP-pair (CP-impair) sont  $f_H = f_B = f_W = 1$  TeV ( $\tilde{f}_B = \tilde{f}_W = 1$  TeV).

Dans la continuité de l'analyse préliminaire présentée dans [150], on souhaite savoir s'il est possible de distinguer un signal CP-pair de CP-impair, et de distinguer un signal par rapport au bruit de fond. Pour ce faire, on simule les évènements, à partir d'un modèle UFO construit avec FeynRules [149], en utilisant MadGraph pour la génération au niveau partonique propagé à PYTHIA pour hadronisation et à Delphes (toujours avec la saveur "CMS") pour une reconstruction au niveau détecteur. MadAnalysis [151] est utilisé pour visualiser les distributions obtenues. On considère des processus  $pp \rightarrow \phi qq$ ,  $\phi \rightarrow \gamma\gamma$  CP-pairs et CP-impairs à des masses  $m_\phi = 20$  et  $70$  GeV pour le signal; pour le bruit de fond on considère le bruit de fond dominant  $pp \rightarrow \gamma\gamma qq$ . Les choix de paramètre utilisés pour ses simulations sont :



- pour le cas d'un scalaire CP-pair, deux possibilités :
  - possibilité nommée CP-pair<sub>1/f<sub>V</sub></sub> où  $f_B = f_W = 1$  TeV et  $f_H \rightarrow \infty$ , ce qui correspond à un couplage de  $\phi$  avec deux bosons  $Z^0$  via  $(Z_{\mu\nu})^2$ , deux photons  $\gamma$  via  $(F_{\mu\nu})^2$  et deux bosons  $W^\pm$  via  $(W_{\mu\nu})^2$
  - possibilité nommée CP-pair<sub>1/f<sub>H</sub></sub> où  $f_H = 1$  TeV et  $f_{B,W} \rightarrow \infty$ , ce qui correspond à un couplage de  $\phi$  avec deux bosons  $Z^0$  via  $(Z_\mu)^2$ , deux photons  $\gamma$  via  $(A_\mu)^2$  et deux bosons  $W^\pm$  via  $(W_\mu)^2$
- pour le cas CP-impair<sub>1/f<sub>V</sub></sub>, on utilise  $\tilde{f}_B = \tilde{f}_W = 1$  TeV correspondant à un couplage avec deux  $Z^0$  bosons via la structure de Lorentz  $Z_{\mu\nu}\tilde{Z}^{\mu\nu}$ , avec deux photons  $\gamma$  via  $F_{\mu\nu}\tilde{F}^{\mu\nu}$  et avec deux  $W^\pm$  via  $W_{\mu\nu}\tilde{W}^{\mu\nu}$

Les figures 8.8-8.11, obtenues avec MadAnalysis [151], montrent quelques unes des distributions cinématiques pour  $m_\phi = 20$  et 70 GeV pour les cas CP-pair<sub>1/f<sub>V</sub></sub>, CP-pair<sub>1/f<sub>H</sub></sub> et CP-impair<sub>1/f<sub>V</sub></sub> et le bruit de fond sans sélection aucune. Ces distributions sont normalisées à l'unité. On peut observer qu'il semble difficile de différencier les différents types de signaux, qu'il s'agisse de discriminer entre un scalaire CP-pair ou impair ou entre les hypothèses CP-pair<sub>1/f<sub>V</sub></sub> et CP-pair<sub>1/f<sub>H</sub></sub>. Certaines distributions semblent néanmoins permettre une discrimination efficace entre le signal et le bruit de fond. Reste cependant à déterminer et optimiser des coupures afin de réduire le nombre d'événements bruit de fond tout en maximisant le nombre d'événements signal ; étude qui ne sera pas abordée ici.

Cette étude devra donc être poursuivie et elle pourrait permettre de réinterpréter les résultats de l'analyse  $h \rightarrow \gamma\gamma$  lorsqu'ils seront publics. Le modèle générique considéré ici offre, de plus, l'opportunité d'un couplage à l'arbre entre les gluons et  $\phi$  qui permettrait une augmentation significative de la section efficace du signal ; ce mode de production pourrait ainsi, lui aussi, être étudié ultérieurement.

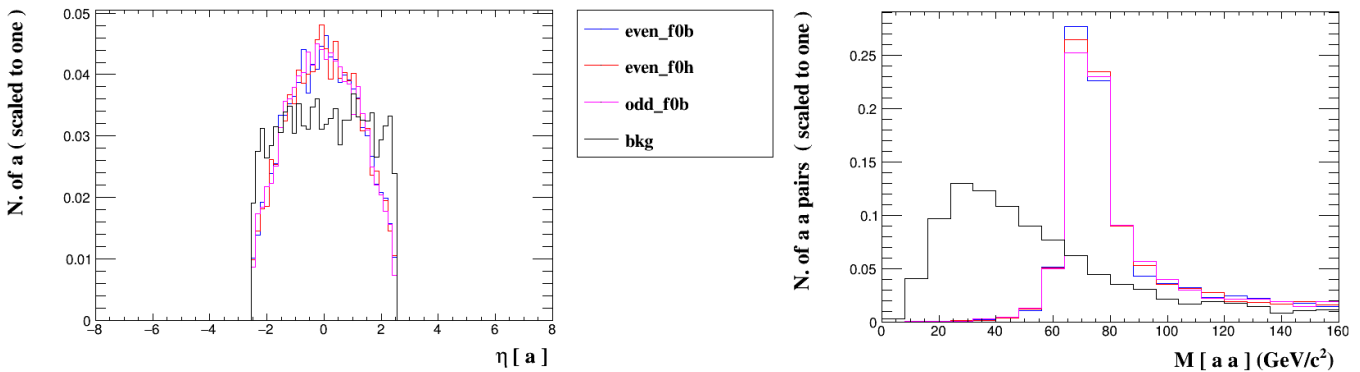


FIGURE 8.8 – Distribution normalisée du  $\eta$  des photons (gauche) et de la masse invariante diphoton (droite) pour une particule  $\phi$  de masse  $m_\phi = 70$  GeV se désintégrant en deux photons pour les cas CP-pair<sub>1/f<sub>V</sub></sub> (bleu), CP-pair<sub>1/f<sub>H</sub></sub> (rouge) et CP-impair<sub>1/f<sub>V</sub></sub> (rose) ainsi que pour le bruit de fond MS (noir).



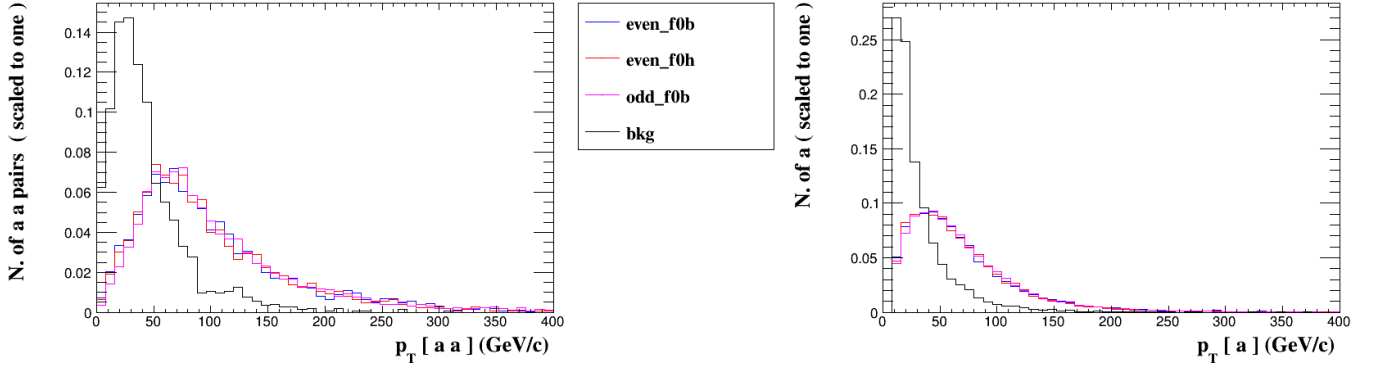


FIGURE 8.9 – Distribution normalisée du moment transverse du système diphoton (gauche) et du moment transverse des photons (droite) pour une particule  $\phi$  de masse  $m_\phi = 20$  GeV se désintégrant en deux photons pour les cas CP-pair $_{1/f_V}$  (bleu), CP-pair $_{1/f_H}$  (rouge) et CP-impair $_{1/\tilde{f}_V}$  (rose) ainsi que pour le bruit de fond MS (noir).

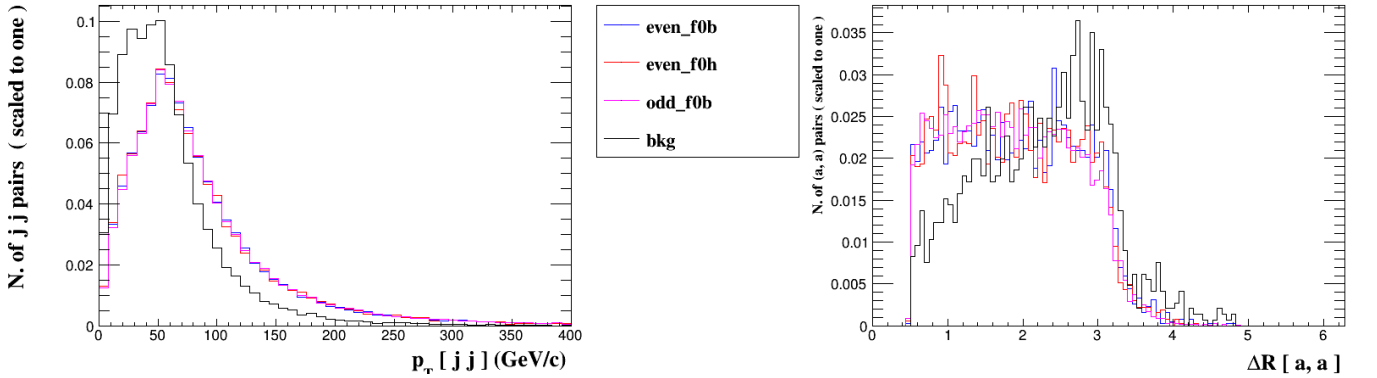


FIGURE 8.10 – Distribution normalisée du moment transverse du système dijet (gauche) et de la séparation angulaire entre les photons (droite) pour une particule  $\phi$  de masse  $m_\phi = 20$  GeV se désintégrant en deux photons pour les cas CP-pair $_{1/f_V}$  (bleu), CP-pair $_{1/f_H}$  (rouge) et CP-impair $_{1/\tilde{f}_V}$  (rose) ainsi que pour le bruit de fond MS (noir).

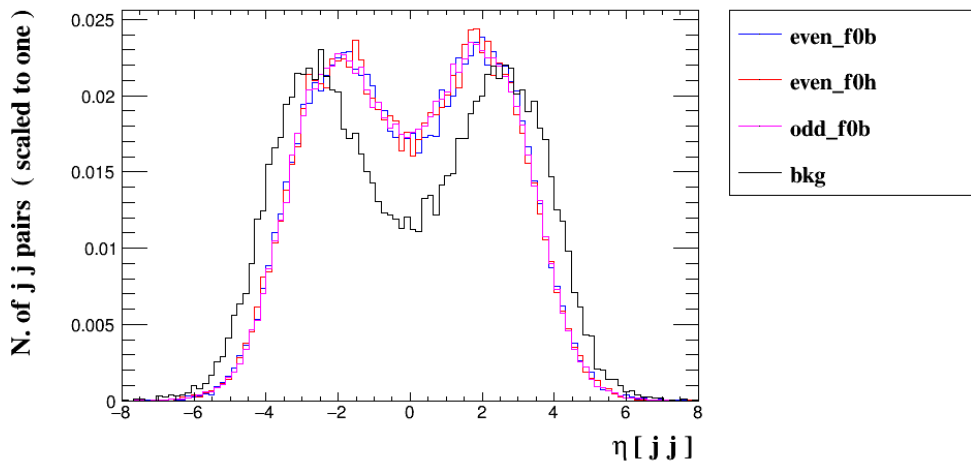


FIGURE 8.11 – Distribution normalisée de la différence d'angle  $\eta$  entre deux jets pour une particule  $\phi$  de masse  $m_\phi = 20$  GeV se désintégrant en deux photons pour les cas CP-pair $_{1/f_V}$  (bleu), CP-pair $_{1/f_H}$  (rouge) et CP-impair $_{1/\tilde{f}_V}$  (rose) ainsi que pour le bruit de fond MS (noir).

### 8.3 Recherche d'une nouvelle particule via un couplage $h\eta\gamma$ au HL-LHC

On étudie ici la possibilité de détecter un couplage  $h\eta\gamma$ , avec  $h$  boson de Higgs à 125 GeV et  $\eta$  un pseudo-scalaire BSM, motivé par des modèles de Higgs composites. On souhaite déterminer dans cette section la sensibilité du LHC à haute luminosité (HL-LHC) à détecter ce type de particule pour un mode de production  $pp \rightarrow \gamma^* \rightarrow h\eta$  et  $\eta \rightarrow Z\gamma$ . Les études présentées, résumées dans [152], ici sont le fruit d'une collaboration, toujours active, avec G. Cacciapaglia, N. Manglani et K. Sridhar.

#### 8.3.1 Cadre du modèle

On considère un modèle de Higgs composites [53, 153] basé sur un modèle de technicouleur [154] utilisant la classe  $SU(4)/Sp(4)$  qui présente un candidat  $\phi$  à la matière noire. Dans la limite Higgs composite, on obtient deux pseudo-bosons de Nambu-Goldstone dont l'un peut être identifié comme le boson de Higgs du MS  $h$  et l'autre est un pseudo-scalaire noté  $\eta$  :

$$\phi = \frac{h + i\eta}{\sqrt{2}} \quad (8.9)$$

Les limites entre la théorie technicouleur et Higgs composites sont représentées par un angle de mixage défini tel que :

$$\sin \theta = \frac{v_{SM}}{f} \quad (8.10)$$

où  $v_{SM} = 246$  GeV est l'échelle électrofaible et  $f$  est la constante de désintégration<sup>8</sup> des pNGB.

Dans la limite technicouleur, le terme d'interaction de  $\phi$  peut s'écrire [155, 156] comme :

$$\mathcal{L} \supset ie \frac{d_B}{\Lambda^2} \left( \phi^* \overleftrightarrow{\partial}_\mu \phi \right) \partial_\nu F^{\mu\nu} \quad (8.11)$$

où  $d_B$  est un paramètre libre,  $F^{\mu\nu}$  le tenseur de Faraday et  $\Lambda$  l'échelle de la brisure de symétrie considérée (de l'ordre du TeV).

Ce terme peut être réécrit dans la limite Higgs composite pour faire apparaître un couplage entre  $\eta$ , le boson de Higgs du MS et les photons :

$$\mathcal{L} \supset ie \frac{d_B}{\Lambda^2} \sin \theta \left( \eta \overleftrightarrow{\partial}_\mu h \right) \partial_\nu F^{\mu\nu} \quad (8.12)$$

où le facteur  $\sin \theta$  permet d'être sûr que le couplage disparaît lorsque la symétrie électrofaible est restaurée.

Le couplage de l'Eq. 8.12 est supprimé pour des photons *on-shell*, ce qui empêche les désintégrations  $\eta \rightarrow h\gamma$  où  $h \rightarrow \eta\gamma$ , il est cependant possible d'avoir des processus de type  $pp \rightarrow \gamma^* \rightarrow h\eta$  que l'on cherche à étudier ici.

Le pseudo-scalaire  $\eta$  peut être plus léger ou plus lourd que le boson de Higgs du MS, il

8. On rappelle que cette constante est analogue à celle utilisée pour la désintégration des pions.

interagit avec les bosons de jauge permettant les désintégrations [53, 153]  $\eta \rightarrow Z\gamma$ ,  $\eta \rightarrow ZZ$  et  $\eta \rightarrow W^+W^-$ , le canal  $\gamma\gamma$  n'étant pas autorisé. Si  $m_\eta < 2m_W$  alors la seule désintégration autorisée est  $\eta \rightarrow Z\gamma$  (ou  $\eta \rightarrow Z^*\gamma$ ). On considère donc le processus :

$$pp \rightarrow \gamma^* \rightarrow h\eta \rightarrow hZ\gamma \quad (8.13)$$

soit une production associée à un boson de Higgs, un boson Z et un photon dur. Cette signature très particulière pourrait permettre une identification aisée du signal par rapport au bruit de fond.

### 8.3.2 Proto-analyse avec simulations

La section efficace du processus 8.13 n'étant pas très élevée il est nécessaire de considérer une production au HL-LHC à 14 TeV afin d'avoir assez d'évènements signal attendus. On s'intéresse à des désintégrations du boson de Higgs  $h \rightarrow b\bar{b}$  présentant une signature expérimentale claire grâce au *b-tagging*<sup>9</sup>. Pour le boson Z, on considère la désintégration  $Z \rightarrow l^+l^-$ .

On simule les évènements, à partir d'un modèle UFO construit avec FeynRules [149], en utilisant MadGraph pour la génération au niveau partonique propagé à PYTHIA pour hadronisation et on utilise FastJet [157] pour la reconstruction des jets (dans l'hypothèse d'un détecteur primaire). On simule les évènements signal  $pp \rightarrow \gamma^* \rightarrow h\eta \rightarrow hZ\gamma$  à une masse  $m_\eta=100$  et 150 GeV ; on utilise pour ce faire des valeurs  $d_B \sin \theta \sim 0,1$  et  $\Lambda \sim 3$  TeV<sup>10</sup>. Pour le bruit de fond du MS, on considère le bruit de fond irréductible  $pp \rightarrow b\bar{b}l^+l^- \gamma$ , incluant le bruit de fond  $Zh\gamma$ , et le bruit de fond réductible  $pp \rightarrow jjl^+l^- \gamma$  où les jets sont identifiés à tort comme étant des b-jets. La figure 8.12 montre quelques exemples de diagrammes de Feynman pour le signal et le bruit de fond irréductible considérés.

On s'attend à obtenir un boson de Higgs  $h$  très boosté avec une haute impulsion transverse pour le signal, de fait les quarks b issus de sa désintégration seront très collimatés avec un  $\Delta R$  faible comme visible sur la figure 8.13. On reconstruit alors des jets en utilisant l'algorithme Cambridge-Aachen [158] via MadAnalysis interfacé avec FastJet pour lequel on assume une efficacité de 100 %. On demande alors la présence d'au moins un gros<sup>11</sup> jet dans l'état final associé à deux sous-jet collimatés de moment transverse supérieur à 50 GeV, ce qui permet de rejeter une partie importante du bruit de fond.

9. Les hadrons associés aux jets  $b$  présentent des durées de vie relativement longues et un vertex correspondant légèrement déplacé.

10. Les valeurs  $\Lambda \lesssim 2$  TeV sont déjà fortement contraintes [156].

11. On demande un rayon de jet de  $R=1,2$  dans l'algorithme de reconstruction.

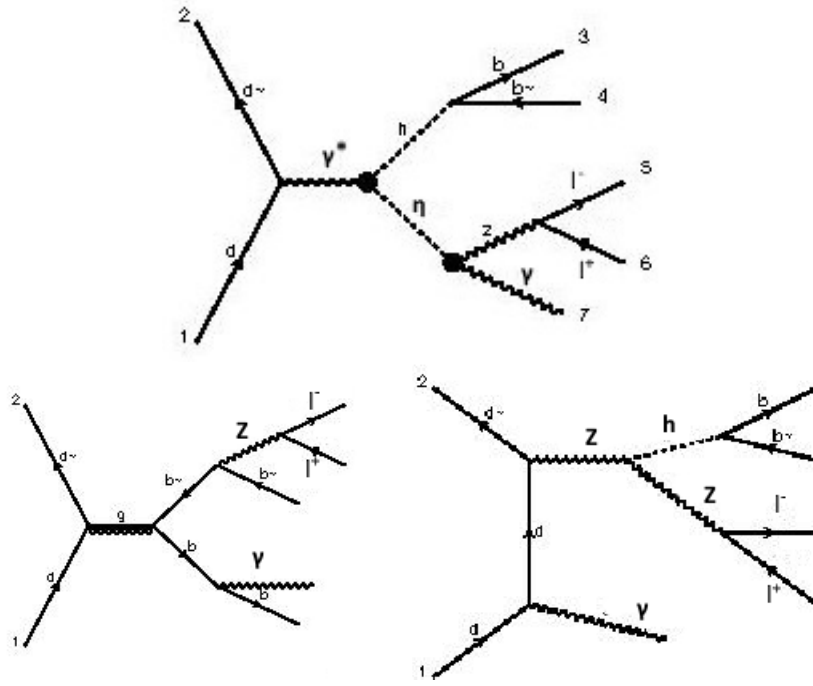


FIGURE 8.12 – Exemples de diagrammes de Feynman pour le signal (en haut) et le bruit de fond (en bas) pour une désintégration invisible du Z.

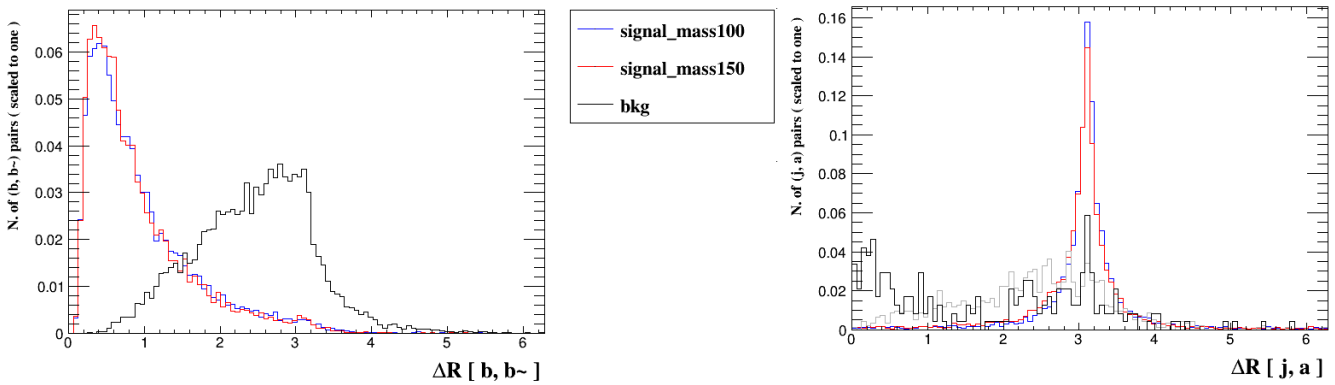


FIGURE 8.13 – Distribution normalisée de la séparation angulaire entre les deux quarks  $b$  dans l'état final (gauche) et entre le jet et le photon (droite) pour un signal de masse  $m_\eta=100$  GeV (bleu) et  $m_\eta=150$  GeV (rouge), ainsi que le bruit de fond (noir) au LHC à 14 TeV.

On effectue des coupures de sélection supplémentaires afin de maximiser le ratio signal sur bruit de fond en utilisant notamment les différences des distributions signal et bruit de fond visibles sur les figures 8.13 et 8.14 :

- On demande que la masse du jet reconstruite soit proche de celle du boson de Higgs soit  $100 < m_j < 170$  GeV.
- On requiert que le jet possède une impulsion transverse élevée soit  $p_T^j > 260$  GeV.
- Le photon doit être aussi boosté soit  $p_T^\gamma > 85$  GeV.
- Les produits de désintégrations du Z doivent reconstruire sa masse invariante soit  $86 < m_{ll} < 96$  GeV.

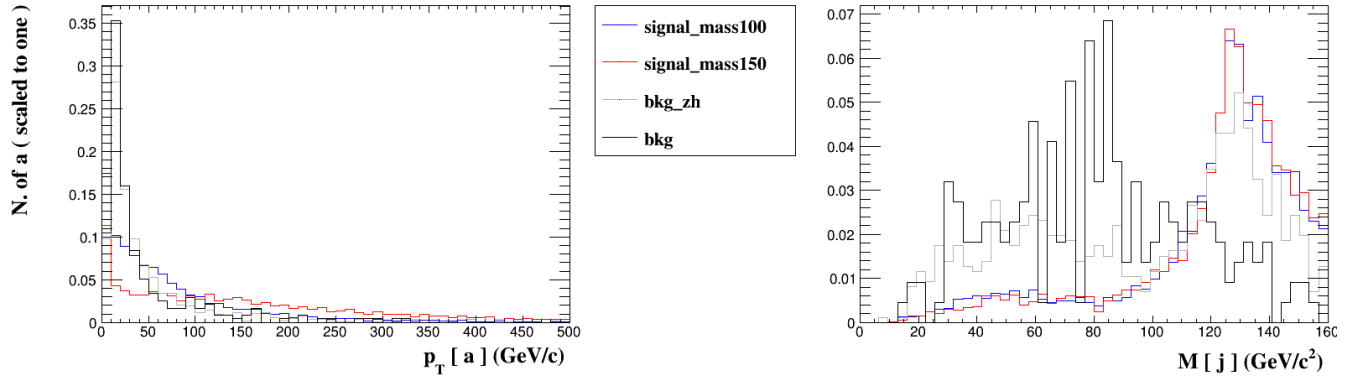


FIGURE 8.14 – Distribution normalisée de l'impulsion transverse du photon (gauche) et de la masse du jet reconstruit (droite) pour un signal de masse  $m_\eta=100$  GeV (bleu) et  $m_\eta=150$  GeV (rouge), ainsi que le bruit de fond (noir) et sa composante  $zh$  (en gris) au LHC à 14 TeV.

- Le photon doit être angulairement proche des produits de désintégration du  $Z$  soit  $\Delta R(l, \gamma) < 1,2$ .
- Le photon doit être angulairement éloigné des produits de désintégration du  $h$  soit  $\Delta R(j, \gamma) > 2,0$ .
- La masse invariante à trois corps  $m_{ll\gamma}$  doit reconstruire la masse du signal dans une fenêtre de 5 GeV.

On requiert de plus que les deux sous-jets soient étiquetés comme des  $b$ -jets en utilisant une efficacité, très simplifiée, pour le  $b$ -tagging de 70 %, ainsi qu'une probabilité de mauvaise étiquetage des quarks  $c$  de 20 % et de 1 % pour les autres jets légers.

Après ces coupures, il reste, avec les paramètres considérés et en prévision du HL-LHC avec une luminosité intégrée de  $3 \text{ ab}^{-1}$ , 0,20 évènements signal (de section efficace totale 1,23 ab et pour un signal de masse 150 GeV), 0,0012 évènements de bruit de fond irréductible (de section efficace totale 93 ab) et environ  $1,2 \times 10^{-5}$  évènements de bruit de fond réductible (de section efficace totale 96 ab). Le bruit de fond peut donc être significativement réduit comparé au signal. Cependant, à cause de la faible section efficace de production, aucun évènement signal ne survit cette sélection pour le cas HL-LHC.

On construit alors la sensibilité de la recherche avec une simple figure de mérite  $Z$  en fonction de la luminosité délivrée  $L$  et de l'échelle  $\Lambda$  considérée, permettant donc d'aborder des cas plus généraux que celui considéré ici, comme :

$$Z(L, \Lambda) = \frac{S}{\sqrt{S + B}} \quad (8.14)$$

où  $S$  et  $B$  sont le nombre d'évènements de signal (à une masse de 150 GeV) et bruit de fond attendu.

On utilise cette figure de mérite pour calculer, de manière simple, la possibilité de découverte à  $3\sigma$  et  $5\sigma$  en fonction de luminosité et l'échelle  $\Lambda$  pour le cas du HL-LHC, le cas du LHC à haute énergie HE-LHC, et le cas hypothétique d'un collisionneur hadronique avec une énergie

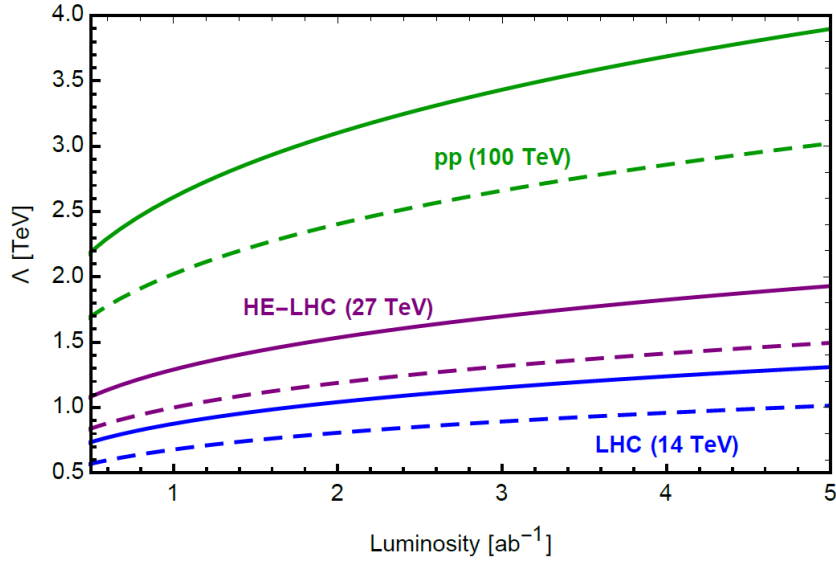


FIGURE 8.15 – Sensibilité de recherche obtenue par une simple figure de mérite en fonction de l'échelle  $\Lambda$  (avec  $d_B \sin \theta = 0, 1$ ) et de la luminosité intégrée attendue pour la proto-analyse considérée ici (avec  $m_\eta = 150$  GeV) pour différentes énergies dans le centre de masse dans des collisionneurs hadroniques. Les lignes pleines correspondent à une valeur de  $3\sigma$  en terme de figure de mérite et les lignes en pointillé correspondent à une valeur de  $5\sigma$ .

de 100 TeV dans le centre de masse. Les résultats obtenus sont visibles sur la figure 8.15. On observe que, sous ces conditions, seuls les cas avec des échelles  $\Lambda$  proches du TeV peuvent être découverts au HL-LHC ; pour le HE-LHC, les cas possédant une échelle  $\Lambda$  jusqu'à 1,5 TeV peuvent être découverts alors que le cas avec une énergie dans le centre de masse de 100 TeV permettrait un potentiel de découverte jusqu'à des échelles de 3 TeV.

La description du détecteur étant ici minimaliste et aucune incertitude n'ayant été prise en compte, cette analyse nécessite donc d'avantage d'optimisation afin de statuer sur la potentialité d'observer une nouvelle particule semblable à  $\eta$  au HL-LHC, ou dans des collisionneurs hadroniques d'énergie plus importante, bien que les résultats obtenus soient encourageants. Une méthode, en cours d'étude, afin d'augmenter la sensibilité de détection est d'utiliser des méthodes d'apprentissage machine pour discriminer le signal du bruit de fond.

Afin d'observer une telle particule, une possibilité serait aussi de s'intéresser à des modes de désintégrations plus propres afin de réduire le bruit de fond : un choix intéressant serait alors les collisionneurs  $e^+e^-$ .

### 8.3.3 Cas des collisionneurs $e^+e^-$

Les collisionneurs  $e^+e^-$  sont en effet des outils permettant de minimiser les interactions QCD synonymes de bruit de fond difficile à reconstruire comme les jets. L'inconvénient de ces collisionneurs est cependant leur énergie dans le centre de masse qui est largement inférieure à celle obtenue dans les collisions protons-protons.

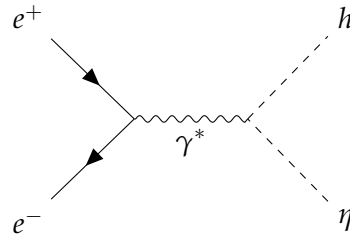


FIGURE 8.16 – Diagramme de Feynman pour la production d'une particule  $\eta$  dans les collisionneurs  $e^+e^-$

On s'intéresse ainsi d'abord à la production d'une particule pseudo-scalaire  $\eta$  visible sur la figure 8.16 au sein du Futur Circular Collider (FCC)  $ee$  [159]. On s'intéresse au point de fonctionnement dédié à la production du quark top qui permettra d'atteindre une énergie dans le centre de masse de  $\sim 350$  GeV. La figure 8.17 montre la section efficace obtenue dans ces conditions avec une valeur de  $\Lambda=1$  TeV. Cette valeur de  $\Lambda$  n'est pas physiquement réaliste mais permet un *rescaling* rapide sachant que la section efficace est proportionnelle à  $\Lambda^4$ . Ainsi on observe que la section efficace est de l'ordre de  $\sim 500$  ab pour une masse autour de 110 GeV. Pour  $\Lambda = 3$  TeV, cette valeur descend à  $\sim 5$  ab sans prendre en compte les rapports de branchement associés aux désintégrations du boson de Higgs et de  $\eta$ . La section efficace associée à la production d'une particule pseudo-scalaire  $\eta$  est donc trop faible au FCC-ee et ne permettrait ainsi pas d'obtenir suffisamment d'évènements de signal dans ce cas.

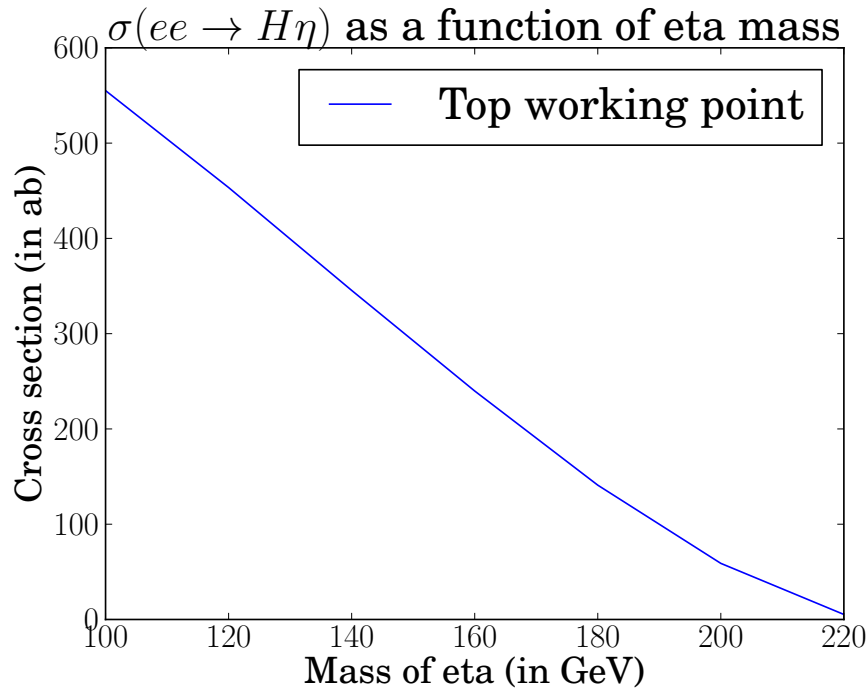


FIGURE 8.17 – Section efficace de la production d'une particule  $\eta$  associée à un boson de Higgs au FCC-ee pour le point de fonctionnement du quark top pour  $\Lambda=1$  TeV.

Il peut alors être intéressant d'étudier ce cas pour d'autres collisionneurs  $e^+e^-$  : la figure 8.18 montre ainsi les valeurs de section efficace obtenues pour différents collisionneurs  $e^+e^-$  ; ainsi

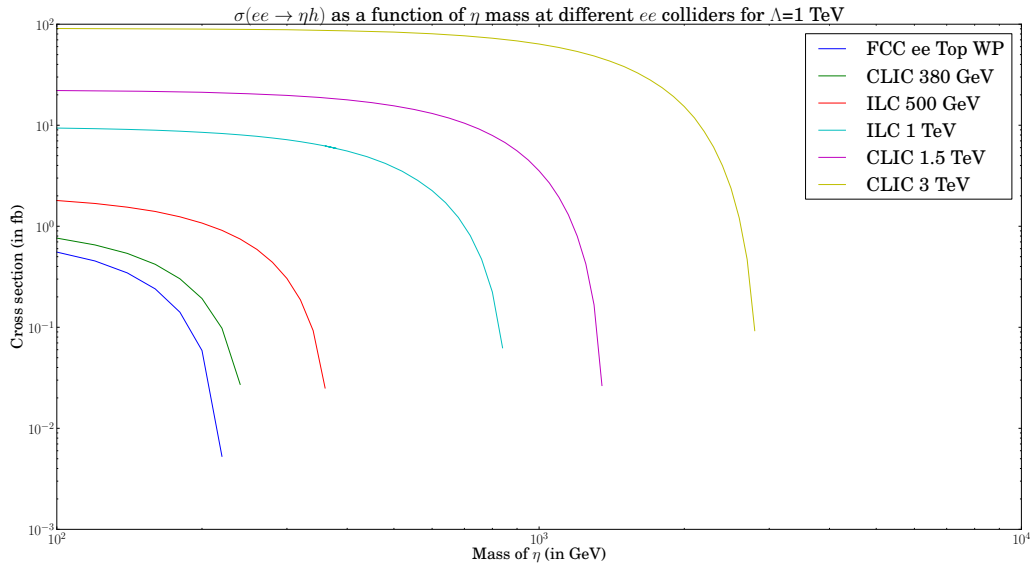


FIGURE 8.18 – Section efficace de la production d’une particule  $\eta$  associée à un boson de Higgs au FCC-ee pour le point de fonctionnement du quark top, et pour les collisionneurs linéaires ILC [161] et CLIC [160] pour  $\Lambda=1$  TeV.

le collisionneur CLIC [160] permettrait une augmentation de la section efficace d’un facteur 10 ou plus de 20 selon le point de fonctionnement considéré et permettrait ainsi l’obtention d’un nombre d’évènements signal attendu suffisant pour effectuer une analyse pour la recherche d’une nouvelle particule de ce type.

Une autre solution si l’on souhaite étudier ce type de processus au FCC-ee serait de considérer des modèles basés sur des classes plus larges comme  $SU(5)/SO(5)$  qui autorisent les couplages  $\eta\gamma\gamma$ <sup>12</sup> et permettent ainsi une production de  $\eta$  par fusion de photon de section efficace potentiellement plus élevée.

## Conclusion

On a donc exposé ici plusieurs études phénoménologiques, présentant souvent quelques hypothèses simplificatrices, invoquant de nouvelles particules (pseudo)scalaires dont l’état final de désintégration contient un ou plusieurs photons qui sont susceptibles d’être détectés au LHC ou dans de futurs collisionneurs de particules.

On a vu qu’il était possible de contraindre des modèles théoriques, notamment des modèles avec des particules de type axion, contenant un nouveau champ scalaire pouvant se désintégrer en photons ou en leptons via des désintégrations exotiques du boson de Higgs au LHC.

12. Ce type de désintégration peut aussi être considérée pour le cas HL-LHC afin d’avoir une signature plus nette.



On a aussi brièvement étudié les propriétés et intérêts d'un mode de production par fusion de bosons vecteurs d'un nouveau scalaire au LHC, en essayant notamment de discriminer une composante CP-paire d'une composante CP-impair.

On a finalement détaillé le potentiel de découverte d'une nouvelle particule en lien avec les modèles de Higgs composites et pouvant être interprétée comme un candidat matière noire associé à un état final original contenant un boson de Higgs et un boson  $Z$  au HL-LHC. Ces études sont encourageantes et ont été publiées dans une lettre dédiée [152]; toutefois le mode de production d'une telle particule via  $e^+e^-$  dans des collisionneurs associés est une piste qui semble aussi prometteuse.

Étant donné que les résultats observés pour le Run 2 de l'analyse  $h \rightarrow \gamma\gamma$  n'étaient pas encore publics lors du déroulement de cette thèse, il n'a pas été possible de les réinterpréter en terme de différents modèles théoriques et proposer ainsi des contraintes et exclusions. Ce travail pourra cependant être réalisé ultérieurement.



# Conclusion Générale

Nous avons vu au cours de ce manuscrit la façon dont a été construit le Modèle Standard de la physique des particules décrivant les interactions électromagnétique, faible et forte grâce à la théorie quantique des champs et l'invariance sous transformation de jauge locale. Cette théorie a prédit l'existence du boson de Higgs découvert au LHC en 2012 mais nous avons aussi vu qu'elle possède des limites, comme les problèmes de naturalité, et qu'il existe des phénomènes qu'elle ne décrit pas comme la matière noire ou encore l'asymétrie entre matière et antimatière. Nous avons ensuite présenté quelques modèles théoriques, les modèles à deux doublets de Higgs, les modèles de Higgs composites et les modèles supersymétriques, permettant de résoudre plusieurs des difficultés rencontrées dans le Modèle Standard. Ces nouveaux modèles prédisent l'existence de plusieurs nouvelles particules dont de nouveaux bosons de Higgs mais aucune de ces particules n'a cependant été découverte jusqu'à présent. Ces différents modèles sont autant de motivations pour rechercher la présence d'une nouvelle particule et notamment d'un boson de Higgs additionnel pouvant être plus léger que la particule découverte en 2012. On a alors choisi de se concentrer sur la recherche de bosons de Higgs additionnels de masse inférieure à 125 GeV qui se désintégreraient en deux photons.

La qualité du détecteur CMS, et en particulier de son calorimètre électromagnétique, permet en effet d'obtenir une bonne résolution en énergie des photons et la performance des algorithmes de reconstruction associés assure une bonne efficacité de reconstruction et d'identification pour les photons. Nous avons aussi étudié les différentes corrections à apporter à l'énergie des photons lors de leur reconstruction afin de prendre en compte, par exemple, l'énergie déposée dans les interstices du calorimètre électromagnétique. Une de ces corrections correspond à celle apportée à l'échelle et à la résolution en énergie des photons qui est calculée en utilisant des électrons provenant de désintégrations  $Z \rightarrow ee$  et doit donc être validée sur des photons.

Nous avons présenté une méthode permettant d'extraire cette échelle d'énergie grâce à la cinématique d'évènements  $Z \rightarrow \mu\mu\gamma$ . Après ajustements et détermination des incertitudes systématiques associées, nous avons observé un bon accord, compris dans un intervalle de  $\pm 1\sigma$  des incertitudes totales, de la valeur de l'échelle d'énergie des photons entre les données et simulations *2017 Ultra Legacy* correspondant à une luminosité intégrée d'environ  $41,5 \text{ fb}^{-1}$ .

Ces connaissances sur la reconstruction des photons au sein du détecteur CMS, nous ont

ensuite permis d'étudier les différentes parties de la chaîne d'analyse visant à rechercher un boson de Higgs additionnel de basse masse dans le canal diphoton avec les données 2017 et 2018 du Run 2 du LHC, correspondant à une luminosité respective de  $41,5 \text{ fb}^{-1}$  et  $54,4 \text{ fb}^{-1}$ . Les optimisations effectuées et le réentraînement des arbres de décision boostés utilisés ont eux-même permis d'obtenir une analyse robuste ainsi que dotée d'une excellente sensibilité. La combinaison des analyses 2017 et 2018 conduit en effet en un minimum (maximum) de limites d'exclusion à 95 % attendues sur le produit de section efficace fois rapport de branchement en deux photons pour un boson de Higgs additionnel de 26 (87) fb correspondant à une masse de signal hypothétique de 110,0 (90,4) GeV et un minimum (maximum) des limites d'exclusion à 95 % normalisées pour la combinaison supplémentaire de l'analyse 2016 est de 0,16 (0,58) correspondant à une masse de signal hypothétique de 110,0 (90,3) GeV. Dans ce second cas la p-value attendue en absence de signal est supérieure à  $5 \sigma$  en terme de déviation standard pour toute la gamme de masse, sauf entre 88 et 92 GeV, signifiant alors qu'il serait possible de découvrir une nouvelle particule dans cette gamme de masse si cette dernière possède une force de signal au moins égale à 1.

Il est cependant crucial, afin de statuer sur une éventuelle découverte ou même de contraindre les modèles théoriques, de pouvoir étudier les résultats "Observés" obtenus directement sur les données. Ces résultats ne sont malheureusement pas, à l'heure où sont écrits ces mots, publics.

Suite à cela, nous avons montré quelques études ayant été effectuée qui montre qu'il existe des pistes d'amélioration, notamment en ce qui concerne la modélisation de la composante Drell-Yan du bruit de fond. Une piste en cours d'exploration serait alors d'utiliser une méthode d'analyse multivariée pour rejeter la majorité de ces événements et se passer ainsi de cette modélisation. Des études sont en tout cas en cours et des modifications seront, au moins, apportées pour l'analyse des données du Run 3 du LHC.

Finalement, nous avons présenté, de manière phénoménologique, différents modèles explicites présentant une nouvelle particule, compatible avec celle recherchée ici, se désintégrant dans un état final contenant au moins un photon, ou ouvrant la porte à de nouvelles recherches. Des proto-analyses, basées sur des coupures de sélection, utilisant des simulations ont été présentées et des contraintes sur les paramètres associés ont été, dans certains cas, dérivées. Les résultats obtenus - bien que préliminaires - sont encourageants notamment en ce qui concerne la recherche de nouvelles particules dans les futurs collisionneurs de particules comme le HL-LHC ou le FCC e-e.

Le LHC et ses détecteurs sont en effet actuellement en arrêt technique, en vue d'améliorations, pour préparer le Run 3 qui permettra potentiellement d'atteindre une énergie dans le centre de masse de 14 TeV pour les collisions proton-proton ainsi qu'une luminosité intégrée deux fois supérieure à celle délivrée pendant le Run 2. Après cela et un autre temps d'arrêt, le LHC à haute luminosité, prévu pour 2027, permettra d'atteindre des luminosités intégrées

de l'ordre de 3000 à 4000  $\text{fb}^{-1}$ , ouvrant d'avantage les portes de découverte d'une nouvelle physique.



## Les arbres de décision boostés

Les arbres de décision boostés (BDT) sont des outils statistiques d'apprentissage automatisé, très utilisés dans le domaine de la physique des particules expérimentale, servant à classifier les événements pour discriminer le signal d'un bruit de fond. Ces BDT jouent, par exemple, un rôle prépondérant dans la classification utilisée dans l'analyse  $h \rightarrow \gamma\gamma$  présentée dans les chapitres précédents.

On a aussi présenté, dans le chapitre 4, une technique de régression utilisant des arbres de régression boostés qui fonctionnent selon le même principe que les BDT mais avec un objectif différent puisqu'on cherche dans ce cas à prédire la valeur d'une variable cible en sortie.

Les BDT et les algorithmes associés requièrent en entrée un jeu de variables d'entrée  $\vec{x}$  choisies pour leur caractère discriminant entre le signal recherché et le bruit de fond considéré, des hyper-paramètres qui permettent la configuration des algorithmes d'apprentissage (taille maximale d'un arbre par exemple) et ils retournent en sortie une note, un score, qui permet la discrimination entre signal et bruit de fond ; en général le signal correspond à un score de sortie de +1 et le bruit de fond de -1 (ou 0).

Comme tout outil d'apprentissage automatisé, il est nécessaire d'entraîner les BDT sur un jeu de données où le type de processus (signal ou bruit de fond) est connu afin qu'ils puissent apprendre à discriminer les événements. Des simulations MC, où les vrais processus sont connus, sont donc en général utilisées pour l'entraînement.

Un BDT est un ensemble d'arbre de décision (DT) qui réalisent, de manière récursive, des coupures de séparation, choisies de manière à minimiser la dispersion dans les régions obtenues, sur les valeurs des variables d'entrée créant deux nouvelles régions d'espace des phases distinctes appelées "noeuds". A chaque noeud, une unique variable d'entrée discriminante est utilisée et l'un des deux noeuds résultats est alors plus *signal-like* alors que l'autre est plus proche du bruit de fond. Le processus est répété dans ces nouveaux noeuds jusqu'à satisfaire un critère d'arrêt. Ce critère peut correspondre à un nombre maximal de noeuds dans l'arbre, un niveau suffisant de pureté de signal dans le noeud considéré ou encore la présence d'une

statistique trop faible dans le noeud considéré. Un noeud satisfaisant le critère d'arrêt ne se segmente plus et il devient une "feuille", une extrémité, de l'arbre. Le chemin logique menant à une feuille est appelé "branche". Une valeur en terme de score de sortie est alors assignée à chaque chemin en fonction, notamment, du contenu final de la feuille associée c'est-à-dire du nombre d'évènements signal et bruit de fond qu'elle contient. Le but de l'entraînement est alors d'optimiser les coupures sur les variables d'entrée afin d'avoir une bonne discrimination entre le signal et le bruit de fond considérés. Un schéma récapitulatif du principe de fonctionnement d'un DT est visible sur la figure A.1.

La carte de coupures et les valeurs de sortie des feuilles correspondantes pour l'entraînement est alors enregistrée et est ensuite utilisée sur des échantillons, en général sur des données, où la nature des processus en présence (signal ou bruit de fond) n'est pas connue afin de discriminer les évènements.

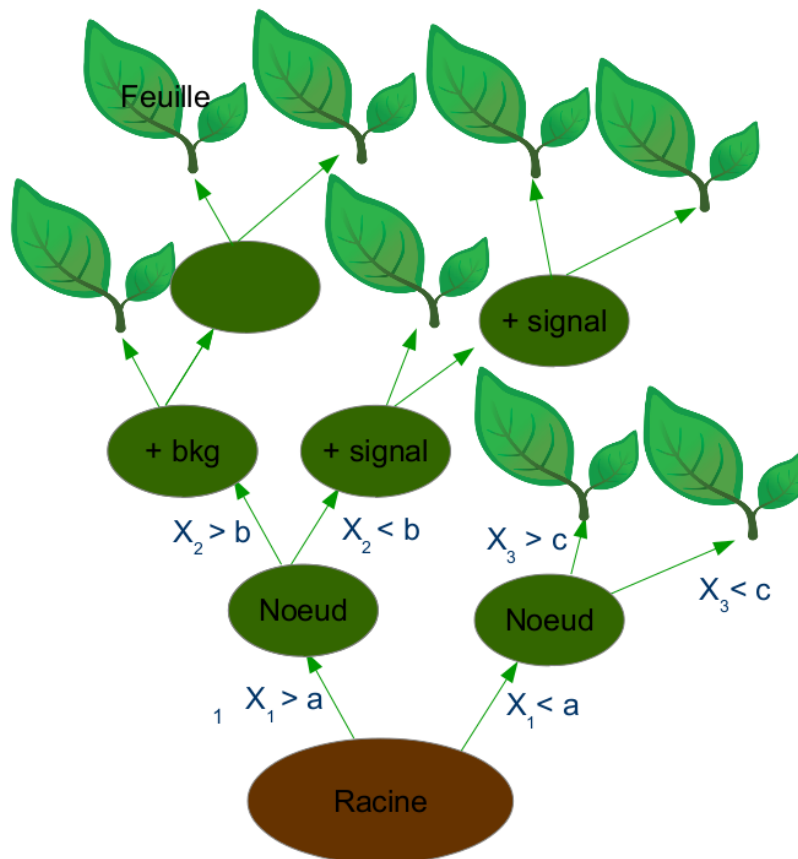


FIGURE A.1 – Schéma récapitulatif d'un arbre de décision individuel. Les noeuds sont obtenus après des coupures  $a, b, c, \dots$  sur les variables d'entrée  $X_i$ .

Le terme "boosté" signifie que l'on combine plusieurs DT pour obtenir un meilleur potentiel d'apprentissage. Un processus d'entraînement itératif est en effet utilisé pour améliorer les performances et optimiser les coupures en permettant à chaque nouvel arbre de corriger l'entraînement précédent créant ainsi une "forêt". La fonction de sortie du BDT final est alors obtenue en prenant la somme pondérée des DT individuels [162].

En pratique ici, deux algorithmes du *package* TMVA [162] sont utilisés pour le boosting. Le



---

premier est AdaBoost [163] qui permet aux évènements signal de l'échantillon d'entraînement répartis dans une feuille ayant une population plus proche du bruit de fond d'avoir un poids plus important que ceux étant répartis dans une feuille *signal-like*. On effectue ainsi une repondération de l'échantillon d'entraînement qui permet alors de donner naissance à un nouveau DT. Ce processus est effectué en moyenne de 100 à 500 fois et on obtient donc au final une forêt contenant autant d'arbres de décision.

Le deuxième est l'algorithme de Gradient Boosting [164] qui calcule la fonction de sortie du BDT par une approche de développement fonctionnelle dont les termes sont les DT individuels et les paramètres associés sont déterminés par la minimisation d'une fonction de perte.

Les performances d'un entraînement de BDT donné sont évaluées sur un échantillon indépendant de celui utilisé pour l'entraînement afin d'être sûr que le comportement appris par le BDT puisse être généralisé au-delà de l'échantillon utilisé pour l'entraînement. Si la performance du BDT est meilleure sur l'échantillon d'entraînement que sur l'échantillon dit de test, on parle de sur-entraînement ou sur-ajustement, signifiant que l'entraînement accorde trop d'importance aux fluctuations statistiques de l'échantillon d'entraînement qu'on ne retrouve pas dans d'autres échantillons d'évènements semblables. Les hyper-paramètres affectent notamment le sur-entraînement et il est donc nécessaire de bien les optimiser.

Comme l'entraînement s'effectue en général sur des simulations, il est important de valider le comportement des variables d'entrée et de sortie sur des données proches du signal considéré.



# Les variables du photon

Différents types de variables sont utilisés par les algorithmes de reconstruction et d'identification des photons. On présente ici une liste, non exhaustive, de ces variables décrivant le comportement des photons. Ces variables sont aussi utilisées dans les analyses impliquant des photons comme celle étudiée dans cette thèse, notamment dans la partie présélection dont le rôle est de sélectionner les meilleurs candidats photons. Cette annexe présente la liste de variable usuellement utilisées pour décrire le comportement des photons.

## B.1 Les variables cinématiques

- $p_T, E_T$  : moment et énergie transverse du photon. On utilise indifféremment l'un ou l'autre du fait que le photon a une masse nulle.
- $E_{raw}$  : énergie brute du photon, correspondant à la somme de l'énergie des cristaux composant le super-agrégats sans correction.
- $E_{vraie}=E_{true}$  : énergie 'vraie' du photon qui correspond à l'énergie idéale totale du photon si tous les effets et corrections possibles étaient pris en compte.
- $\eta, \phi$  : coordonnées angulaires du photon évaluées au centre de l'agrégat graine dans le cas d'un photon non converti ou au barycentre, pondéré par le logarithme des énergies, des différents cristaux regroupés dans le cas d'un photon converti.
- $E_{PS}$  : énergie du photon déposée dans le détecteur à pied de gerbe et en accord avec la position du super-agrégat correspondant.
- $\rho$  : densité d'énergie médiane par unité d'aire. Cette variable permet de soustraire la contribution en énergie de l'empilement dans les variables d'isolation.

## B.2 Les variables de forme de gerbe

Le rôle des variables de forme de gerbe est généralement de discriminer entre les vrais photons, ou photon *prompts*, et les faux photons, et plus particulièrement des photons non résolus issus de désintégration de pions neutres.

- $R_{9,5 \times 5} = \frac{E_{5 \times 5}}{E_{raw}}$ ,  $R_{9,3 \times 3} = \frac{E_{3 \times 3}}{E_{raw}}$  rapport entre l'énergie contenue dans une matrice  $5 \times 5$  ( $3 \times 3$ ) autour du cristal graine et l'énergie brute du super-agrégat correspondant.  $R_9$  est proche de 1 pour des photons *prompts*, de bons candidats photons non convertis, et permet aussi de discriminer de faux photons.
- $E_{2 \times 2} / E_{5 \times 5}$  : le rapport entre l'énergie dans une matrice  $2 \times 2$  contenant les deux cristaux dont le dépôt d'énergie est le plus important et l'énergie contenue dans une matrice  $5 \times 5$  autour du cristal graine du super-agrégat considéré.
- $\sigma_\eta, \sigma_\phi$  : étalement standard de la gerbe en  $\phi$  ( $\eta$ ) au sein du super-agrégat obtenu par une pondération logarithmique en énergie de la position des cristaux au sein du super-agrégat.
- $\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_j^{5 \times 5} w_j (i\eta_j - i\eta_{graine})^2}{\sum_j^{5 \times 5} w_j}}$  : moyenne quadratique pondérée de la position  $i\eta$  d'un cristal en  $\eta$  par rapport à la position du cristal graine dans une matrice  $5 \times 5$ , ce qui correspond, en d'autres termes, à l'étalement standard en  $\eta$  de la gerbe autour du cristal graine dans une matrice  $5 \times 5$ . Les poids  $w_j$  sont directement liés au rapport d'énergie dans le cristal  $j$  à une calibration  $b$  du bruit de fond près,  $w_j = b + \ln \frac{E_j}{E_{5 \times 5}}$ . Cette variable permet aussi de discriminer entre photons *prompts*, dont la gerbe est plutôt étalée dans le direction  $\phi$ , et, notamment, les pions neutres.
- $\sigma_{i\phi i\phi}$  : même définition mais pour l'angle  $\phi$ .
- $\sigma_{i\eta i\phi}$  : matrice de covariance présentant les corrélations entre  $\phi$  et  $\eta$
- $\sigma_{RR}$  : déviation standard de l'étalement de la gerbe dans les plans  $x$  et  $y$  dans détecteur à pied de gerbe.

### B.3 Les variables d'isolation

Le rôle des variables d'isolation est généralement de réduire les mauvaises identifications, en tant que photons, de mésons neutres produits dans les jets et autres objets.

- $H/E$  : le rapport entre l'énergie déposée dans le HCAL et l'énergie déposée dans le ECAL dans un cône  $\Delta R$  défini (en général 0,3 ou 0,15) qui permet ainsi de discriminer les objets hadroniques de ceux purement électromagnétiques.
- $I_\gamma$  : variable d'isolation des photons définie comme étant la somme des moments transverses des particules identifiées comme photons à l'intérieur d'un cône  $\Delta R=0,3$  autour du candidat photon considéré. Les corrections  $\rho$  des effets d'empilement sont appliquées.
- $I_{track}$  : variable d'isolation des traces définie comme étant la somme des moments transverses de toutes les traces à l'intérieur d'un cône compris entre  $\Delta R=0,04$ , afin d'exclure les produits de conversion, et 0,3 autour du candidat photon considéré. Les corrections  $\rho$  sont appliquées.
- $I_{charged}$  : variable d'isolation des hadrons chargés définie comme étant la somme des moments transverses de tous les hadrons chargés à l'intérieur d'un cône  $\Delta R=0,3$  autour du candidat photon considéré. Les corrections  $\rho$  sont appliquées.

# Liste des échantillons utilisés

## C.1 Données

Dataset name	Run-range	$\mathcal{L}$ [ $\text{fb}^{-1}$ ]
/DoubleEG/Run2017B-31Mar2018-v1/MINIAOD <sup>1</sup>	297047–299329	4.79
/DoubleEG/Run2017C-31Mar2018-v1/MINIAOD <sup>1</sup>	299368–302029	9.63
/DoubleEG/Run2017D-31Mar2018-v1/MINIAOD <sup>1</sup>	302030–302663	4.25
/DoubleEG/Run2017E-31Mar2018-v1/MINIAOD <sup>1</sup>	303818–304797	9.31
/DoubleEG/Run2017F-31Mar2018-v1/MINIAOD <sup>1</sup>	305040–306460	13.54
/EGamma/Run2018A-17Sep2018-v2/MINIAOD	314472–325175	54.4 <sup>1</sup>
/EGamma/Run2018B-17Sep2018-v1/MINIAOD		
/EGamma/Run2018C-17Sep2018-v1/MINIAOD		
/EGamma/Run2018D-22Jan2019-v2/MINIAOD		
/DoubleMuon/Run2017B-31Mar2018-v1/MINIAOD <sup>1</sup>	297047–299329	4.79
/DoubleMuon/Run2017C-31Mar2018-v1/MINIAOD <sup>1</sup>	299368–302029	9.63
/DoubleMuon/Run2017D-31Mar2018-v1/MINIAOD <sup>1</sup>	302031–302663	4.25
/DoubleMuon/Run2017E-31Mar2018-v1/MINIAOD <sup>1</sup>	303824–304797	9.31
/DoubleMuon/Run2017F-31Mar2018-v1/MINIAOD <sup>1</sup>	305040–306462	13.54
/DoubleMuon/Run2018A-17Sep2018-v2/MINIAOD	314472–325175	59.5
/DoubleMuon/Run2018B-17Sep2018-v1/MINIAOD		
/DoubleMuon/Run2018C-17Sep2018-v1/MINIAOD		
/DoubleMuon/Run2018D-PromptReco-v2/MINIAOD		
/SingleElectron/Run2017B-31Mar2018-v1/MINIAOD <sup>1</sup>	297047–299329	4.79
/SingleElectron/Run2017C-31Mar2018-v1/MINIAOD <sup>1</sup>	299368–302029	9.63
/SingleElectron/Run2017D-31Mar2018-v1/MINIAOD <sup>1</sup>	302030–302663	4.25
/SingleElectron/Run2017E-31Mar2018-v1/MINIAOD <sup>1</sup>	303818–304797	9.31
/SingleElectron/Run2017F-31Mar2018-v1/MINIAOD <sup>1</sup>	305040–306460	13.54

<sup>1</sup> 2018 HLT path “HLT\_Diphoton30\_18\_R9IdL\_AND\_HE\_AN\_IsoCaloId\_NoPixelVeto” not included at the beginning of 2018 data taking

TABLE C.1 – Échantillons de données utilisés dans l’analyse avec la luminosité correspondante par période.

## C.2 Simulations - Signal

TABLE C.2 – Échantillons de signal GGH 2017 utilisés dans l'analyse avec la section-efficace ( $\text{fb}$ ) et nombre d'évènements correspondant.

Mass [GeV]	Samples	Events	$\sigma$
60	/GluGluHToGG_M60_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	298308	82.724
65	/GluGluHToGG_M65_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	118169	86.596
70	/GluGluHToGG_M70_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	296573	90.354
75	/GluGluHToGG_M75_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	118504	93.968
80	/GluGluHToGG_M80_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	296107	97.531
85	/GluGluHToGG_M85_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	118705	101.059
90	/GluGluHToGG_M90_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	296579	104.576
95	/GluGluHToGG_M95_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	119570	107.965
100	/GluGluHToGG_M100_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	298757	111.019
105	/GluGluHToGG_M105_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	120391	113.381
110	/GluGluHToGG_M110_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	299329	114.7000
115	/GluGluHToGG_M115_13TeV_amcatnloFXFX_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	119797	113.846

## C.2. SIMULATIONS - SIGNAL

TABLE C.3 – Échantillons de signal VBF 2017 utilisés dans l'analyse avec la section-efficace (fb) et nombre d'évènements correspondant.

Mass [GeV]	Samples	Events	$\sigma$
60	/VBFHToGG_M60_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	15000	3.357
65	/VBFHToGG_M65_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	280352	3.832
70	/VBFHToGG_M70_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	150000	4.329
75	/VBFHToGG_M75_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	286818	4.848
80	/VBFHToGG_M80_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	100000	5.385
85	/VBFHToGG_M85_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	600000	5.947
90	/VBFHToGG_M90_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	100000	6.530
95	/VBFHToGG_M95_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	300000	7.123
100	/VBFHToGG_M100_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	300000	7.725
105	/VBFHToGG_M105_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	300000	8.294
110	/VBFHToGG_M110_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	300000	8.784
115	/VBFHToGG_M115_13TeV_amcatnlo_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	600000	9.123

ANNEXE C. LISTE DES ÉCHANTILLONS UTILISÉS

TABLE C.4 – Échantillons de signal VH 2017 utilisés dans l'analyse avec la section-efficace ( $\text{fb}$ ) et nombre d'évènements correspondant.

Mass [GeV]	Samples	Events	$\sigma$
60	/VHToGG_M60_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	628248	8.670
65	/VHToGG_M65_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	572646	8.530
70	/VHToGG_M70_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	599899	8.382
75	/VHToGG_M75_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	523397	8.222
80	/VHToGG_M80_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	571917	8.068
85	/VHToGG_M85_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	567668	7.902
90	/VHToGG_M90_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	517625	7.748
95	/VHToGG_M95_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	493042	7.584
100	/VHToGG_M100_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	541555	7.417
105	/VHToGG_M105_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	571235	7.214
110	/VHToGG_M110_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	530837	6.939
115	/VHToGG_M115_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	492721	6.586



TABLE C.5 – Échantillons de signal ttH 2017 utilisés dans l'analyse avec la section-efficace ( $\sigma$ ) et nombre d'évènements correspondant.

Mass [GeV]	Samples	Events	$\sigma$
60	/ttHJetToGG_M60_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	356102	1.485
65	/ttHJetToGG_M65_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	407477	1.474
70	/ttHJetToGG_M70_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	347928	1.452
75	/ttHJetToGG_M75_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	408064	1.427
80	/ttHJetToGG_M80_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	487949	1.406
85	/ttHJetToGG_M85_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	409862	1.398
90	/ttHJetToGG_M90_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	513669	1.406
95	/ttHJetToGG_M95_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	383203	1.429
100	/ttHJetToGG_M100_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	399489	1.465
105	/ttHJetToGG_M105_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	411852	1.495
110	/ttHJetToGG_M110_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	392737	1.495
115	/ttHJetToGG_M115_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1 /MINIAODSIM	824462	1.431

TABLE C.6 – Échantillons de signal GGH 2018 utilisés dans l’analyse avec la section-efficace (fb) et nombre d’évènements correspondant.

Mass [GeV]	Samples	Events	$\sigma$
65	/GluGluHToGG_M65_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	119610	86.596
70	/GluGluHToGG_M70_TuneCP5_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	104906	90.354
75	/GluGluHToGG_M75_TuneCP5_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	120390	93.968
80	/GluGluHToGG_M80_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	99627	97.531
85	/GluGluHToGG_M85_TuneCP5_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	121406	101.059
90	/GluGluHToGG_M90_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	99941	104.576
95	/GluGluHToGG_M95_TuneCP5_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	111864	107.965
100	/GluGluHToGG_M100_TuneCP5_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	105225	111.019
105	/GluGluHToGG_M105_TuneCP5_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	121041	113.381
110	/GluGluHToGG_M110_TuneCP5_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	97783	114.7000
115	/GluGluHToGG_M115_TuneCP5_13TeV_amcatnloFXFX_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	120057	113.846

TABLE C.7 – Échantillons de signal VBF 2018 utilisés dans l'analyse avec la section-efficace (fb) et nombre d'évènements correspondant.

Mass [GeV]	Samples	Events	$\sigma$
65	/VBFHToGG_M65_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	290000	3.832
70	/VBFHToGG_M70_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	150000	4.329
75	/VBFHToGG_M75_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	300000	4.848
80	/VBFHToGG_M80_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	100000	5.385
85	/VBFHToGG_M85_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	294000	5.947
90	/VBFHToGG_M90_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	100000	6.530
95	/VBFHToGG_M95_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	300000	7.123
100	/VBFHToGG_M100_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	300000	7.725
105	/VBFHToGG_M105_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	300000	8.294
110	/VBFHToGG_M110_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	300000	8.784
115	/VBFHToGG_M115_13TeV_amcatnlo_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	300000	9.123

TABLE C.8 – Échantillons de signal VH 2018 utilisés dans l’analyse avec la section-efficace ( $\text{fb}$ ) et nombre d’évènements correspondant.

Mass [GeV]	Samples	Events	$\sigma$
65	/VHToGG_M65_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	561203	8.530
70	/VHToGG_M70_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	547752	8.382
75	/VHToGG_M75_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	576990	8.222
80	/VHToGG_M80_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	575762	8.068
85	/VHToGG_M85_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	536249	7.902
90	/VHToGG_M90_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	520301	7.748
95	/VHToGG_M95_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	555516	7.584
100	/VHToGG_M100_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	517337	7.417
105	/VHToGG_M105_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	486739	7.214
110	/VHToGG_M110_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	545128	6.939
115	/VHToGG_M115_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	460135	6.586

TABLE C.9 – Échantillons de signal ttH 2018 utilisés dans l'analyse avec la section-efficace ( $\text{fb}$ ) et nombre d'évènements correspondant.

Mass [GeV]	Samples	Events	$\sigma$
65	/ttHJetToGG_M65_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	396886	1.474
70	/ttHJetToGG_M70_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	195497	1.452
75	/ttHJetToGG_M75_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	394128	1.427
80	/ttHJetToGG_M80_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	447020	1.406
85	/ttHJetToGG_M85_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	405784	1.398
90	/ttHJetToGG_M90_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	475989	1.406
95	/ttHJetToGG_M95_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	398321	1.429
100	/ttHJetToGG_M100_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	414518	1.465
105	/ttHJetToGG_M105_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	381260	1.495
110	/ttHJetToGG_M110_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	365974	1.495
115	/ttHJetToGG_M115_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/ RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1 /MINIAODSIM	411605	1.431

### C.3 Simulations - Bruit de fond

TABLE C.10 – Échantillons bruit de fond utilisés en 2017 ("RunIIFall17MiniAOD") et 2018 ("RunIIAutumn18MiniAOD") présent avec leur section efficace (pb), leur nombre d'évènement et la luminosité équivalente ( $\text{pb}^{-1}$ ) correspondant.

2017 Samples	Events	$\sigma$	Equiv Lumi
/GJet_Pt-20to40_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	17058859	220.0	77540.2
/GJet_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM	77049626	850.8	90561.4
/GJet_Pt-20toInf_DoubleEMEnriched_MGG-40to80_TuneCP5_13TeV_Pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	42042387	3216.0	13072.9
/QCD_Pt-30to40_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	14501828	22110.0	655.9
/QCD_Pt-30toInf_DoubleEMEnriched_MGG-40to80_TuneCP5_13TeV_Pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	41954304	241400.0	173.8
/QCD_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	20341160	118100.0	172.2
/DiPhotonJetsBox_MGG-40to80_13TeV-Sherpa/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM	3849864	303.2	12697.4
/DiPhotonJetsBox_MGG-80toInf_13TeV-Sherpa/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM	49605719	84.4	587745.5
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFFFX-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	27413121	5765.4	3222.1
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFFFX-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14_ext1-v1/MINIAODSIM	182100467	5765.4	21405.1
2018 Samples*	Events	$\sigma$	Equiv Lumi
/GJet_Pt-20to40_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15_ext1-v2/MINIAODSIM	14366641	220.0	65302.9
/GJet_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	10205533	850.8	11995.2
/GJet_Pt-20toInf_DoubleEMEnriched_MGG-40to80_TuneCP5_13TeV_Pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15_ext1-v2/MINIAODSIM	6072002	3216.0	1888.1
/QCD_Pt-30toInf_DoubleEMEnriched_MGG-40to80_TuneCP5_13TeV_Pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v2/MINIAODSIM	10912544	241400.0	45.2
/QCD_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	10895375	118100.0	92.3
/DiPhotonJetsBox_M40_80-Sherpa/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	4881986	303.2	16101.5
/DiPhotonJetsBox_MGG-80toInf_13TeV-Sherpa/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	14857013	84.4	176031.0
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFFFX-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15_ext2-v1/MINIAODSIM	193215674	5765.4	22711.5

\* "/QCD\_Pt-30to40\_DoubleEMEnriched\_MGG-80toInf\_TuneCP5\_13TeV\_Pythia8" is not available



# Figures de validation du diphoton BDT

## D.1 Validations zones de contrôle de données et simulations

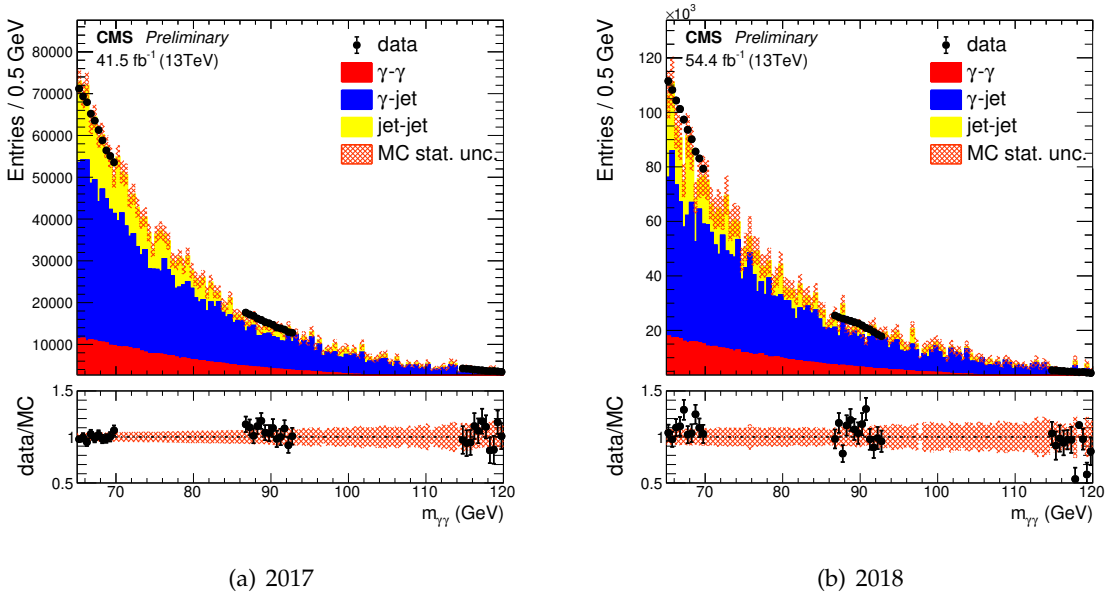


FIGURE D.1 – Distributions de masse invariante diphoton pour a) 2017 et b) 2018. Les évènements dans des fenêtres de masse entre  $70 \text{ GeV} < m_{\gamma\gamma} < 87 \text{ GeV}$  et  $93 \text{ GeV} < m_{\gamma\gamma} < 115 \text{ GeV}$  sont cachés. Les simulations sont normalisées au nombre d'évènements dans les données.

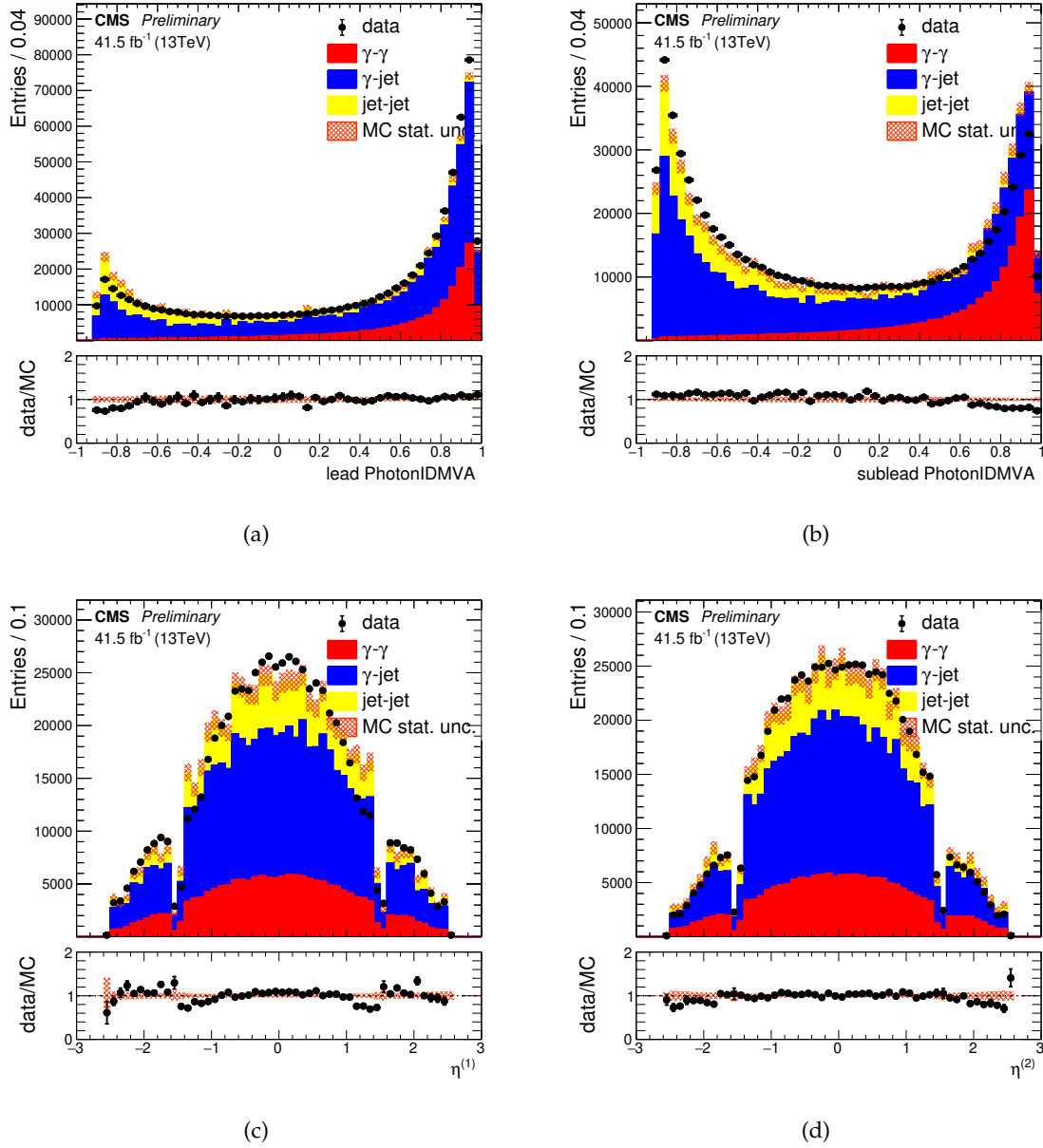


FIGURE D.2 – Comparaisons données/simulations pour 2017 des variables d'entrée du *diphoton BDT* a) score *Photon ID BDT* pour le photon *lead*, b) score *Photon ID BDT* pour le photon *sublead*, c)  $\eta$  du photon *lead* et d)  $\eta$  du photon *sublead*. Les simulations sont normalisées au nombre d'évènements dans les données.



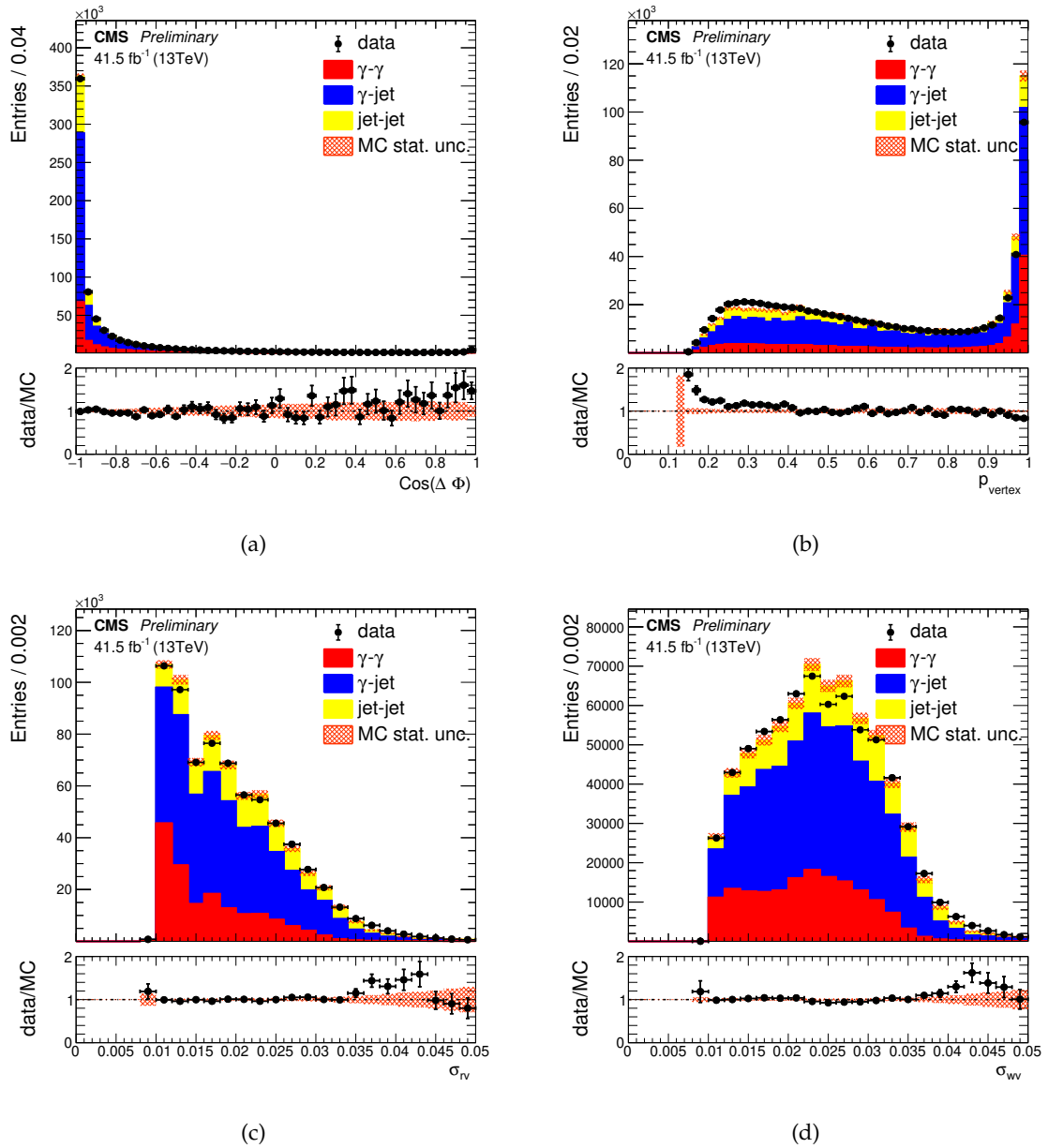


FIGURE D.3 – Comparaisons données/simulations pour 2017 des variables d'entrée du *diphoton* BDT a) cosinus de l'angle entre les photons, b) score de sortie du BDT pour la probabilité d'avoir un vertex situé à moins de 1 cm du vrai vertex, c) résolution en masse dans le cas bon vertex et d) résolution en masse dans le cas mauvais vertex. Les simulations sont normalisées au nombre d'évènements dans les données.

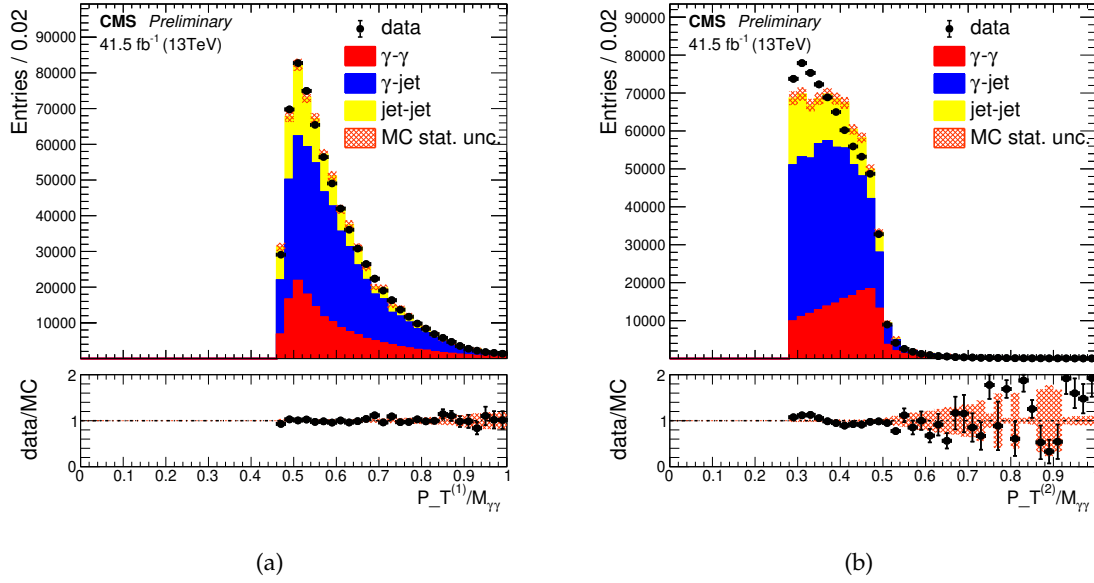


FIGURE D.4 – Comparaisons données/simulations pour 2017 des variables d'entrée du *diphoton BDT* a) impulsion transverse normalisée par la masse invariante pour le photon *lead* et b) impulsion transverse normalisée par la masse invariante pour le photon *sublead*. Les simulations sont normalisées au nombre d'évènements dans les données.

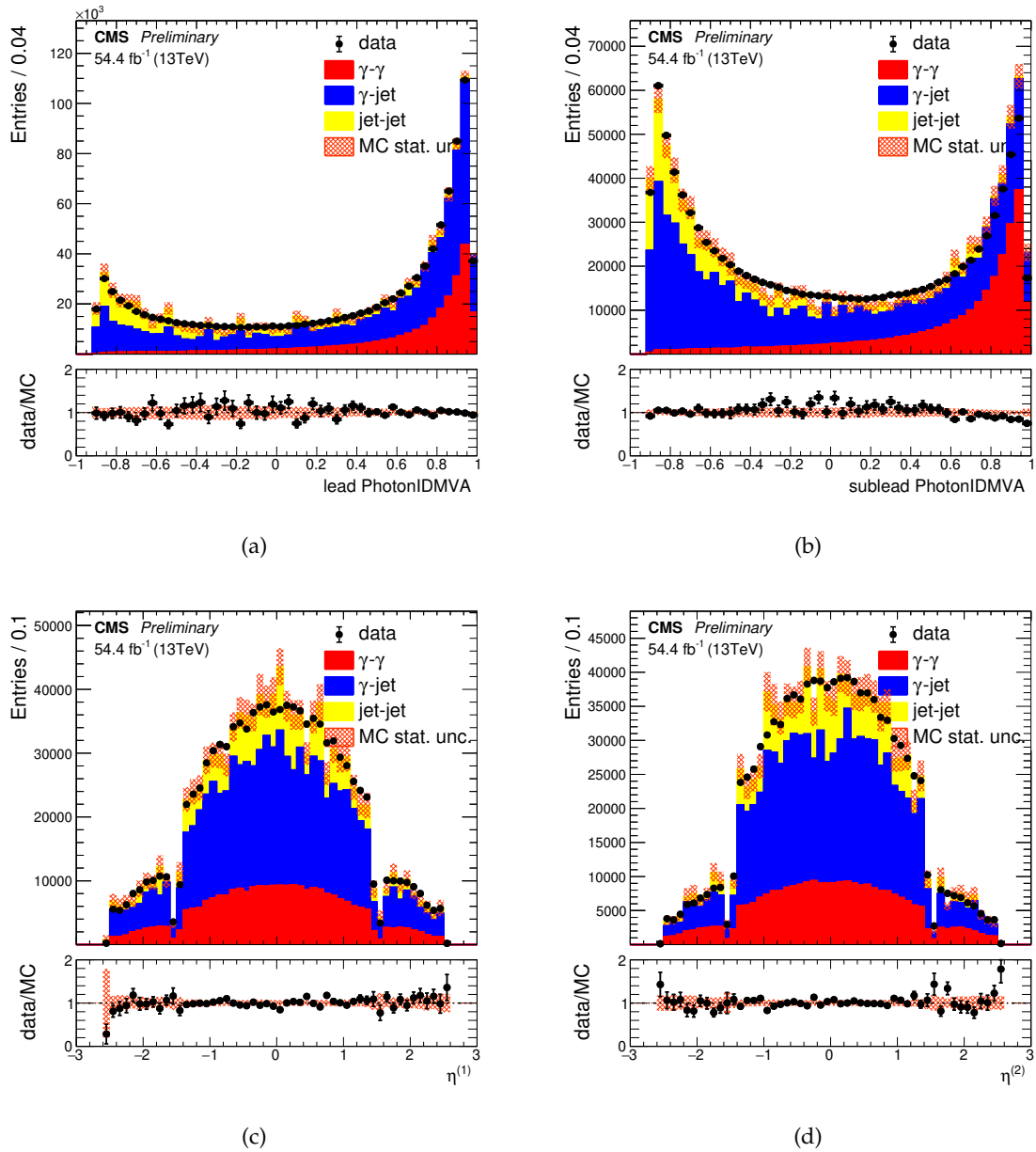


FIGURE D.5 – Comparaisons données/simulations pour 2018 des variables d'entrée du *diphoton BDT* a) score *Photon ID BDT* pour le photon *lead*, b) score *Photon ID BDT* pour le photon *sublead*, c)  $\eta$  du photon *lead* et d)  $\eta$  du photon *sublead*. Les simulations sont normalisées au nombre d'évènements dans les données.

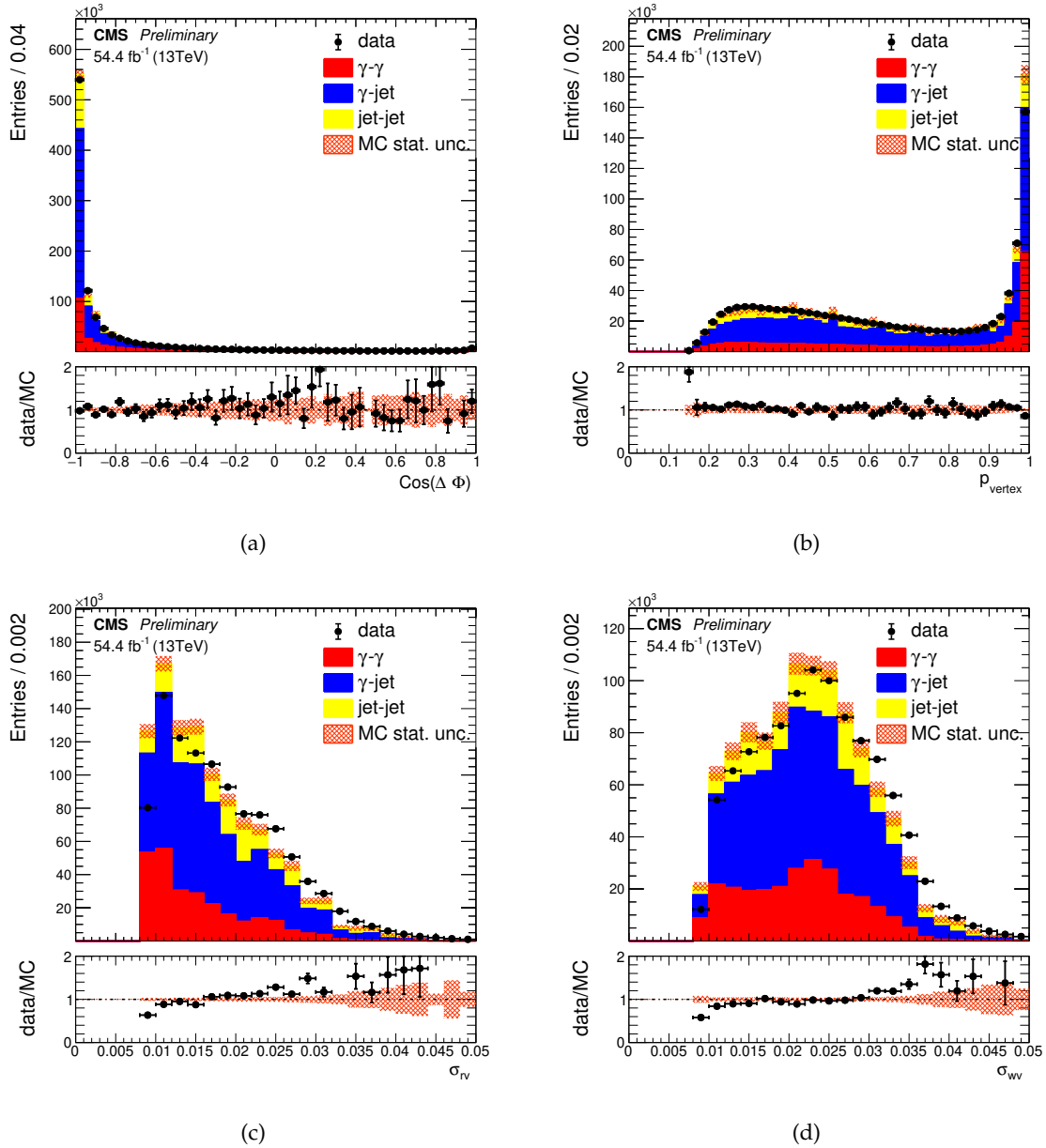


FIGURE D.6 – Comparaisons données/simulations pour 2017 des variables d'entrée du *diphoton BDT* a) cosinus de l'angle entre les photons, b) score de sortie du BDT pour la probabilité d'avoir un vertex situé à moins de 1 cm du vrai vertex, c) résolution en masse dans le cas bon vertex et d) résolution en masse dans le cas mauvais vertex. Les simulations sont normalisées au nombre d'évènements dans les données.

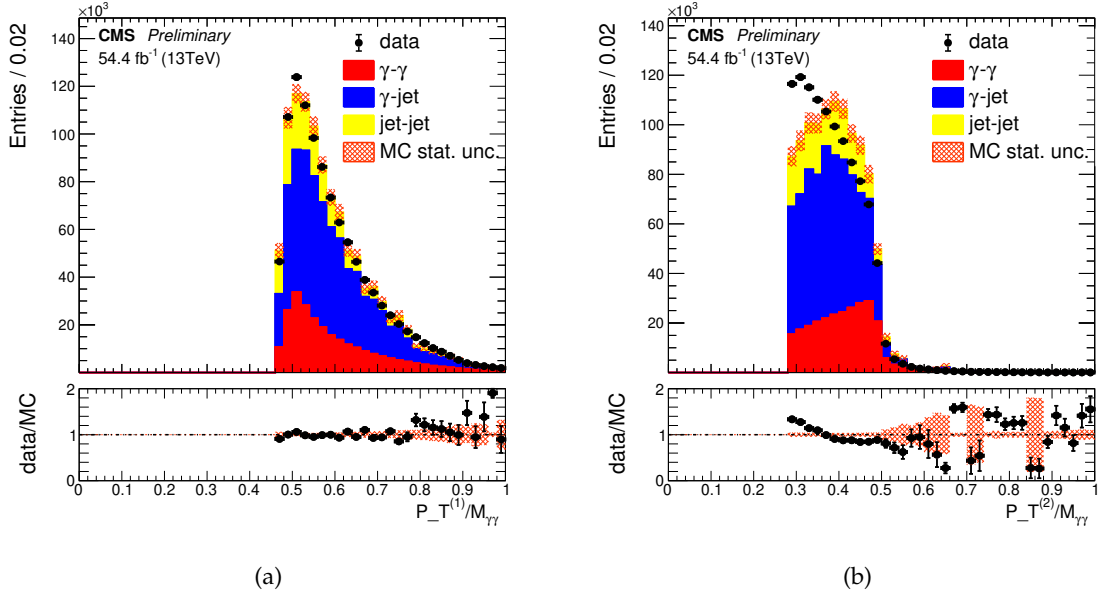


FIGURE D.7 – Comparaisons données/simulations pour 2017 des variables d'entrée du *diphoton BDT* a) impulsion transverse normalisée par la masse invariante pour le photon *lead* et b) impulsion transverse normalisée par la masse invariante pour le photon *sublead*. Les simulations sont normalisées au nombre d'évènements dans les données.

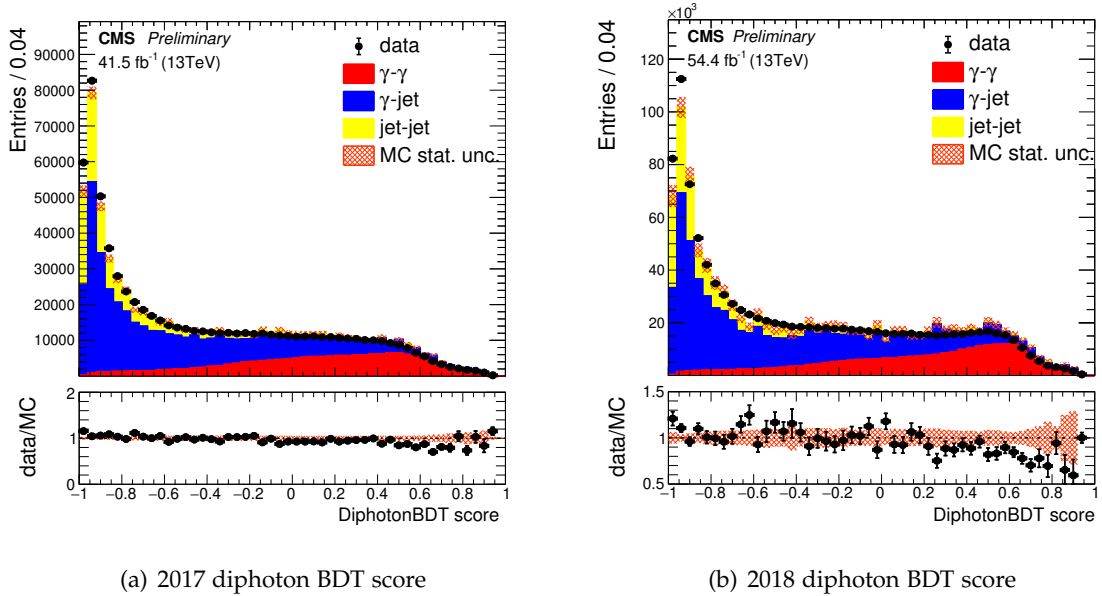


FIGURE D.8 – Distribution en score de sortie du *diphoton BDT* pour les données (points noirs), et les simulations séparées en différentes composantes comprenant les incertitudes statistiques associées (en hachuré). Les données sont prises dans des zones de contrôles avec  $65 \text{ GeV} < m_{\gamma\gamma} < 70 \text{ GeV}$  et  $110 \text{ GeV} < m_{\gamma\gamma} < 120 \text{ GeV}$  pour 2017 (gauche) et 2018 (droit). Les parties basses des sous-figures présentent la distribution des rapports entre données et simulations. A noter que l'échelle des ordonnées de la figure basse pour 2017 est différente de celle utilisée pour 2018.

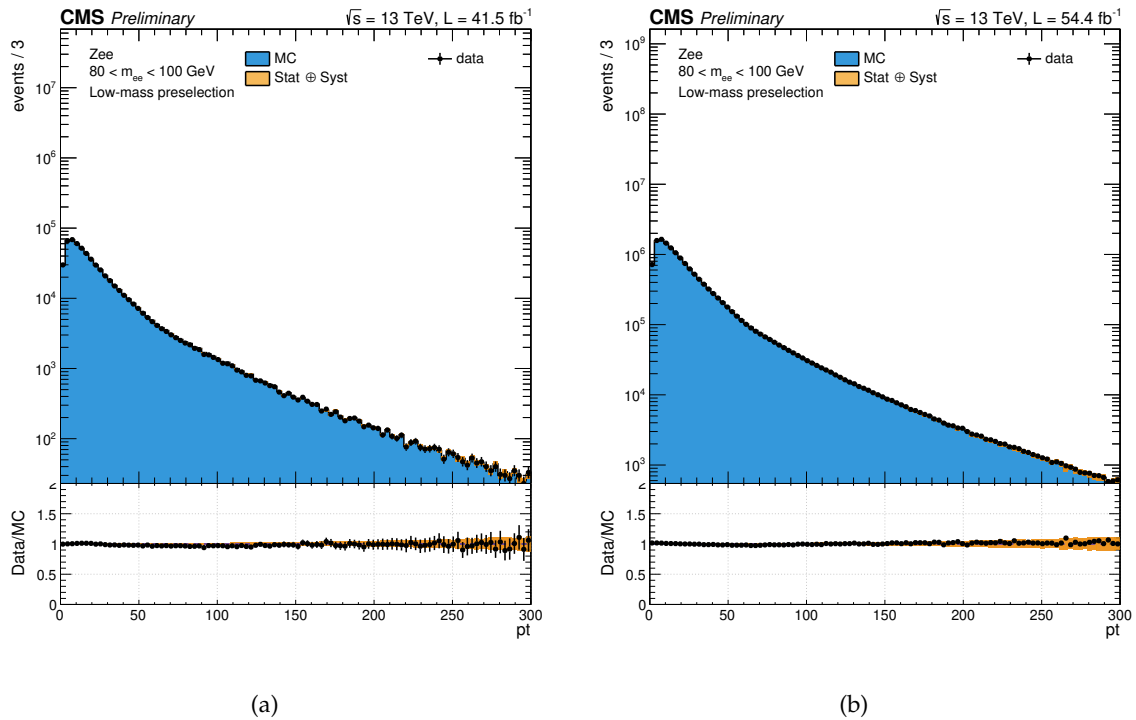
D.2 Validation  $Z \rightarrow ee$ 

FIGURE D.9 – Comparaison données/simulations pour 2017 (gauche) et 2018 (droite) de la distribution  $p_T$  du système diélectron où les électrons sont reconstruits comme des photons. La comparaison entre données (points noirs) et simulations (bleu plein) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.

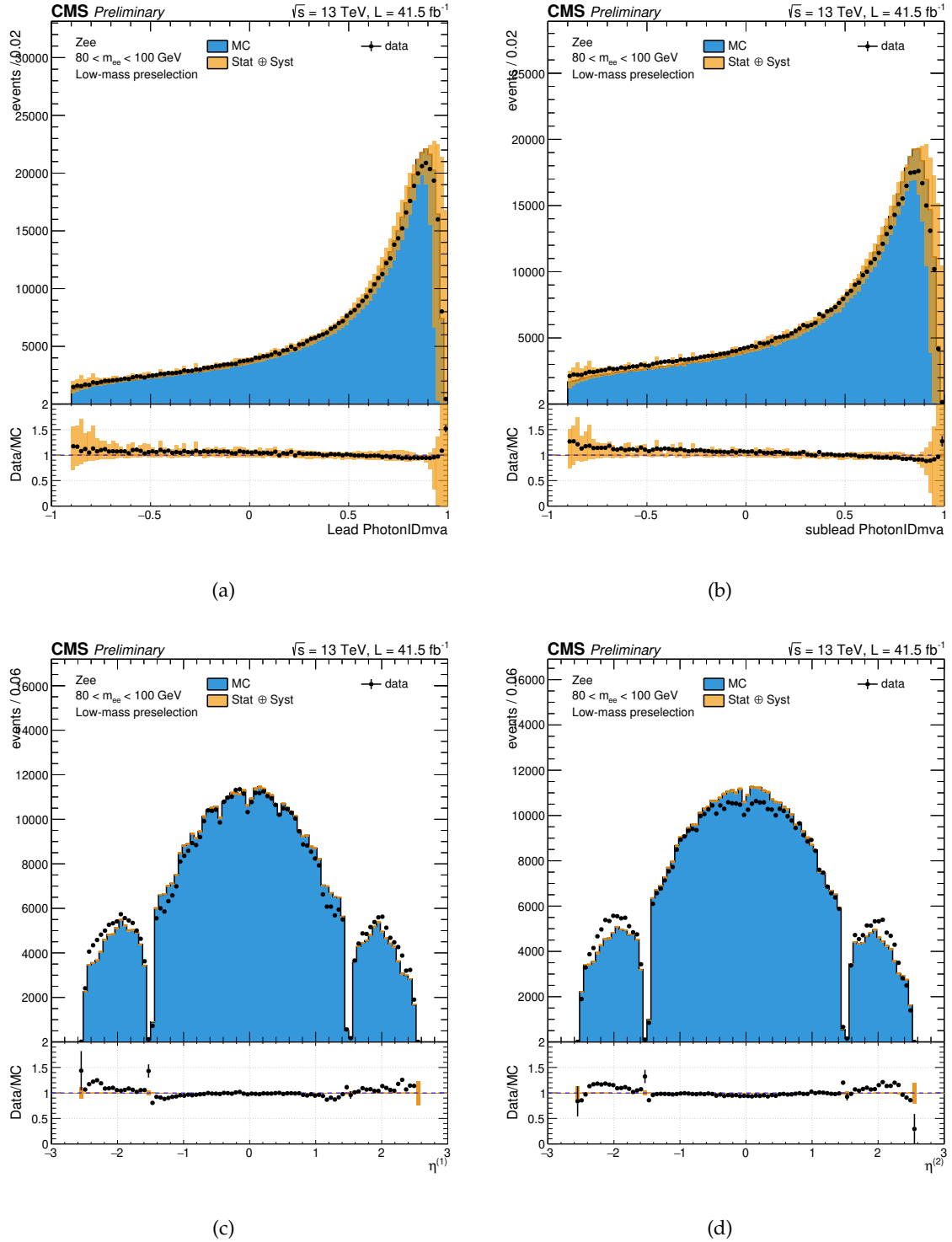


FIGURE D.10 – Comparaison données/simulations 2017 de la distribution pour des électrons reconstruits comme des photons des variables d’entrée *diphoton BDT* a) score *Photon ID BDT* pour le photon *lead*, b) score *Photon ID BDT* pour le photon *sublead*, c)  $\eta$  du photon *lead* et d)  $\eta$  du photon *sublead*. La comparaison entre données (points noirs) et simulations (bleu plein) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.

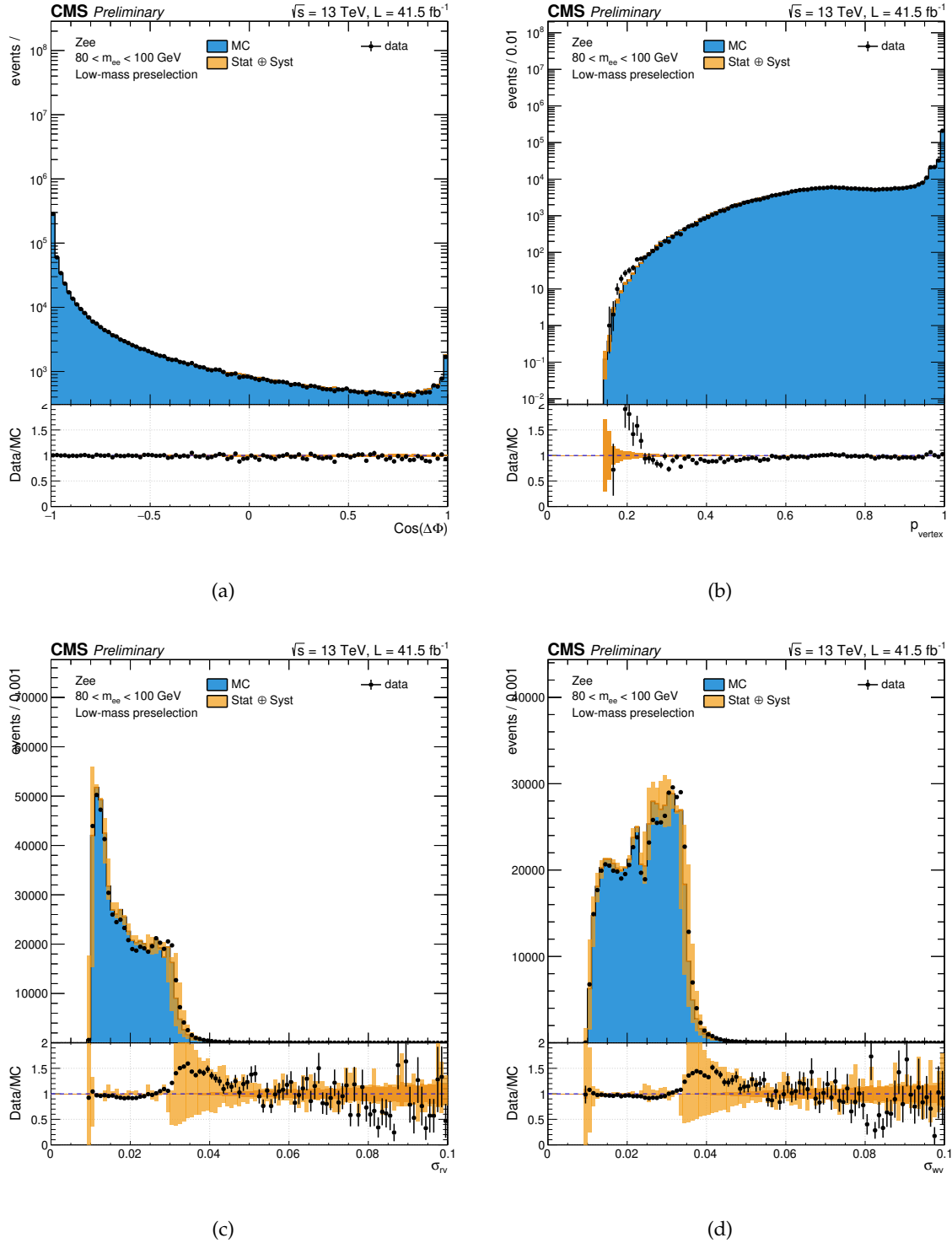


FIGURE D.11 – Comparaison données/simulations 2017 de la distribution pour des électrons reconstruits comme des photons des variables d'entrée *diphoton BDT* a) cosinus de l'angle entre les photons, b) score de sortie du BDT pour la probabilité d'avoir un vertex situé à moins de 1 cm du vrai vertex, c) résolution en masse dans le cas bon vertex et d) résolution en masse dans le cas mauvais vertex. La comparaison entre données (points noirs) et simulations (bleu plein) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.



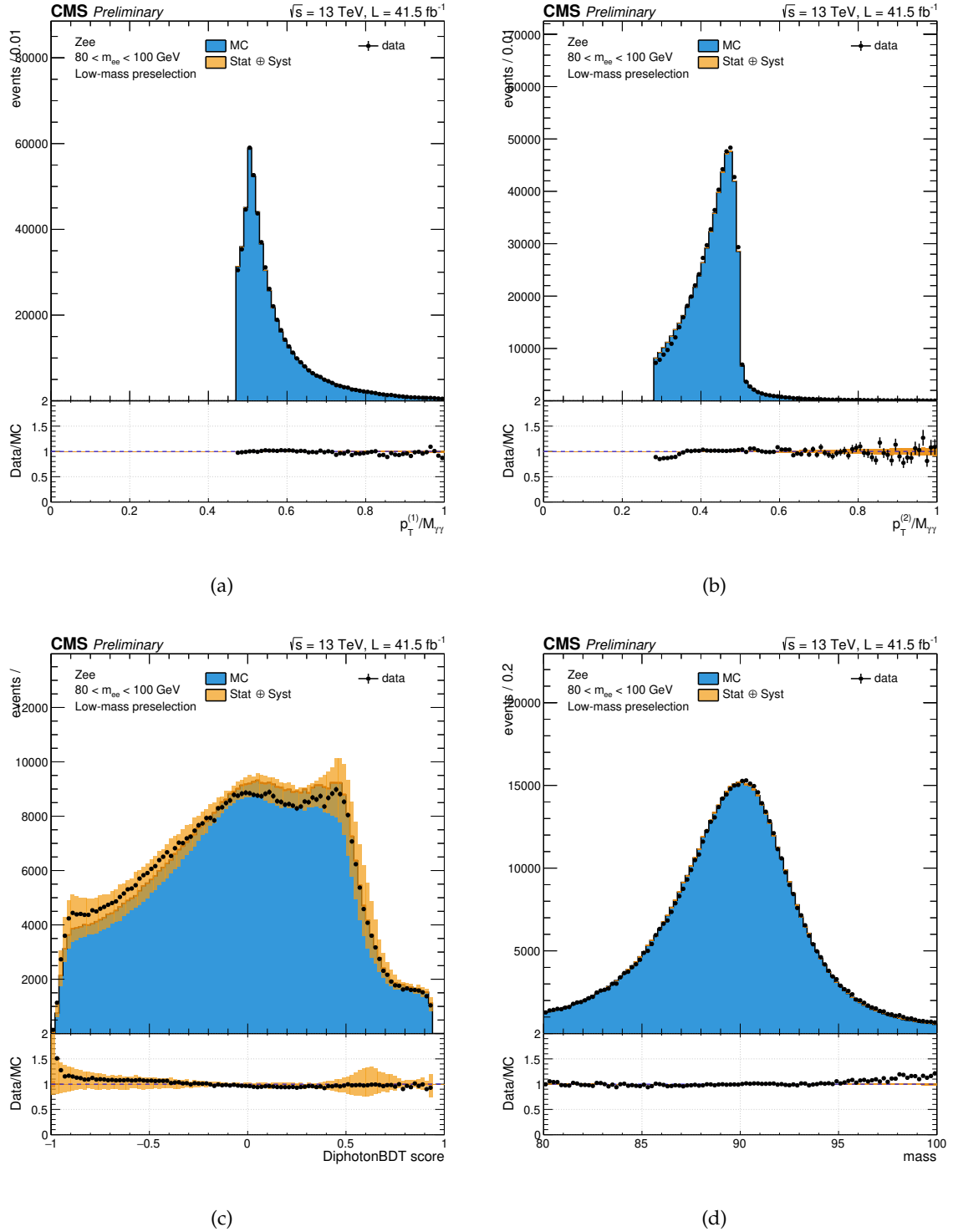


FIGURE D.12 – Comparaison données/simulations 2017 de la distribution pour des électrons reconstruits comme des photons pour a) impulsion transverse normalisée par la masse invariante pour le photon *lead* et b) impulsion transverse normalisée par la masse invariante pour le photon *sublead*, c) score de sortie du *diphoton BDT* et d) masse invariante du système. La comparaison entre données (points noirs) et simulations (bleu plein) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.

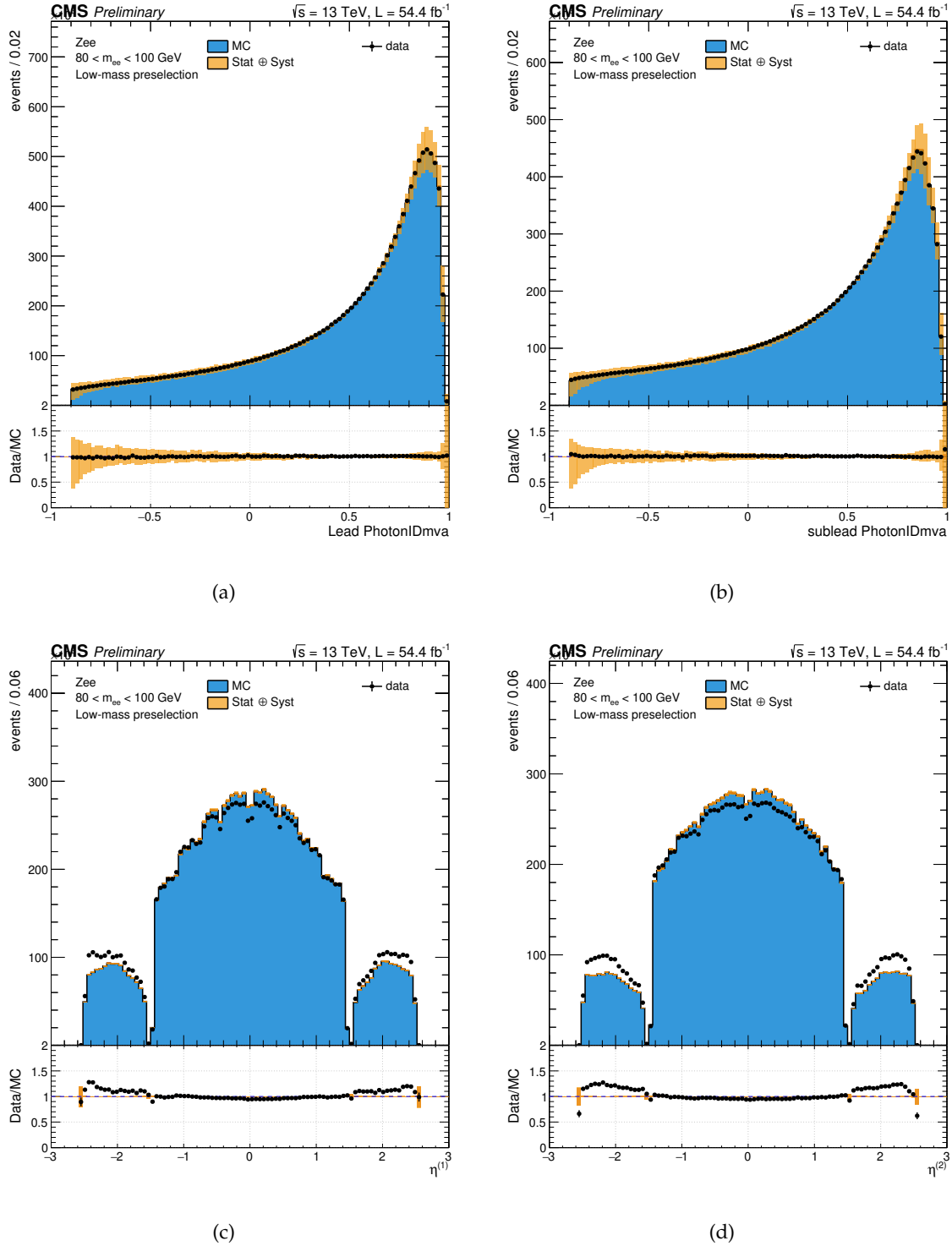


FIGURE D.13 – Comparaison données/simulations 2018 de la distribution pour des électrons reconstruits comme des photons des variables d’entrée *diphoton BDT* a) score *Photon ID BDT* pour le photon *lead*, b) score *Photon ID BDT* pour le photon *sublead*, c)  $\eta$  du photon *lead* et d)  $\eta$  du photon *sublead*. La comparaison entre données (points noirs) et simulations (bleu plein) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.

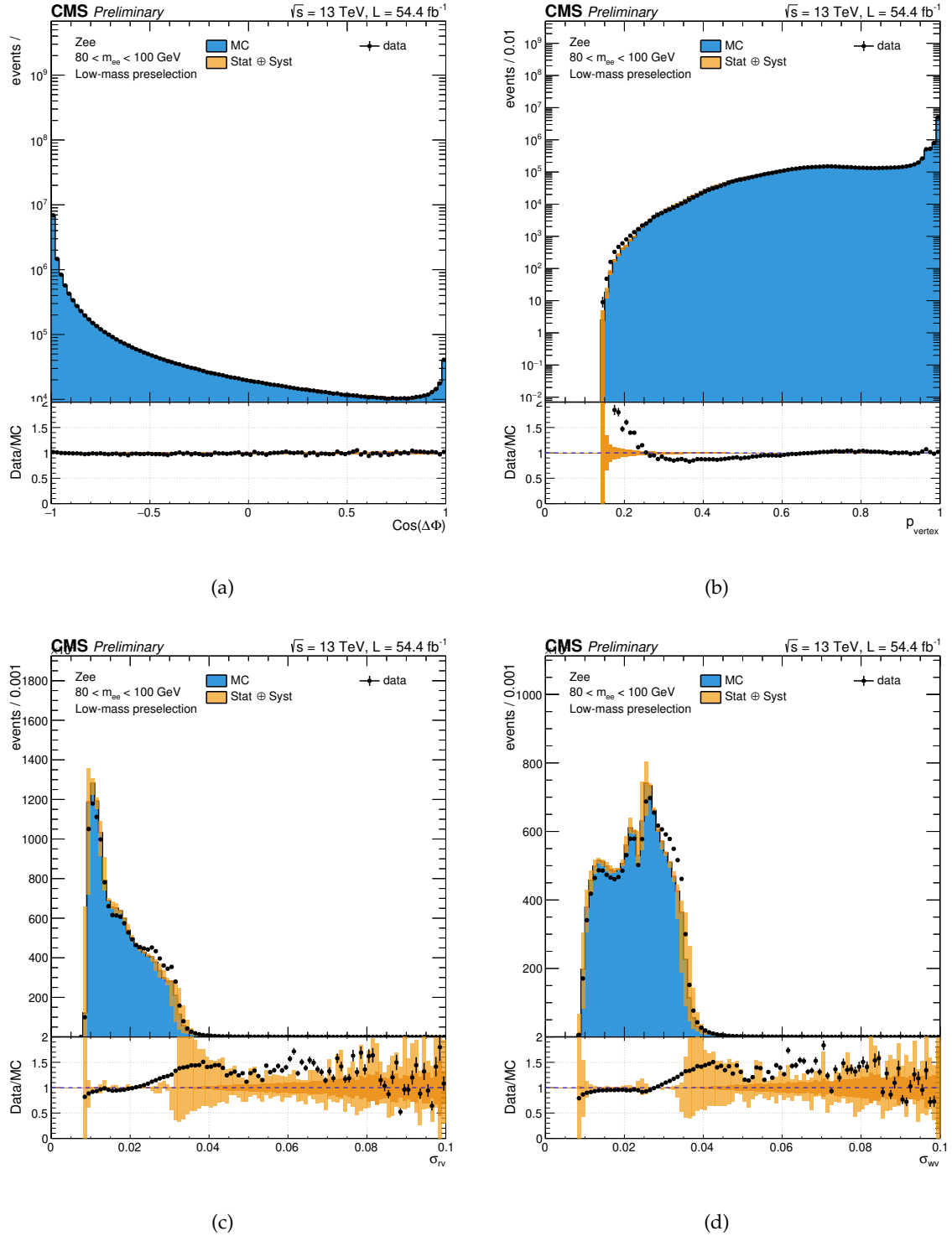


FIGURE D.14 – Comparaison données/simulations 2018 de la distribution pour des électrons reconstruits comme des photons des variables d’entrée *diphoton BDT* a) cosinus de l’angle entre les photons, b) score de sortie du BDT pour la probabilité d’avoir un vertex situé à moins de 1 cm du vrai vertex, c) résolution en masse dans le cas bon vertex et d) résolution en masse dans le cas mauvais vertex. La comparaison entre données (points noirs) et simulations (bleu plein) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.

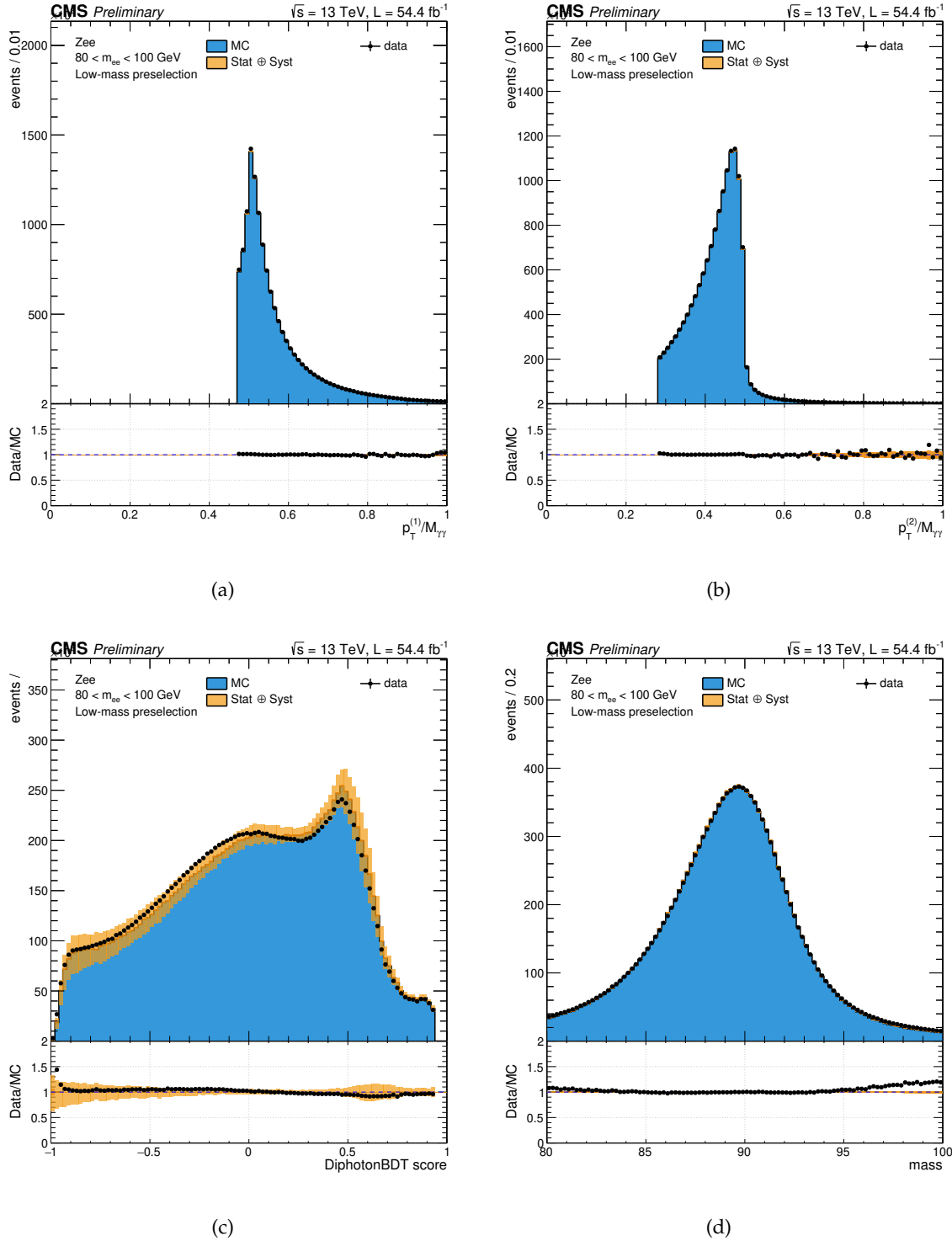


FIGURE D.15 – Comparaison données/simulations 2018 de la distribution pour des électrons reconstruits comme des photons pour a) impulsion transverse normalisée par la masse invariante pour le photon *lead* et b) impulsion transverse normalisée par la masse invariante pour le photon *sublead*, c) score de sortie du *diphoton BDT* et d) masse invariante du système. La comparaison entre données (points noirs) et simulations (bleu plein) accompagnées de leurs incertitudes (orange plein) est aussi visible sur une figure présentant le rapport de ces deux distributions.

# Bibliographie

- [1] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19 :1264–1266, Nov 1967. <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [2] Sheldon L. Glashow. The renormalizability of vector meson interactions. *Nuclear Physics*, 10 :107–117, 1959. <https://www.sciencedirect.com/science/article/pii/0029558259901968>.
- [3] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519 :367–377, 1968.
- [4] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10 :531–533, Jun 1963. <https://link.aps.org/doi/10.1103/PhysRevLett.10.531>.
- [5] Makoto Kobayashi and Toshihide Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Progress of Theoretical Physics*, 49(2) :652–657, 02 1973. <https://doi.org/10.1143/PTP.49.652>.
- [6] Gerard 't Hooft and M. J. G. Veltman. Regularization and Renormalization of Gauge Fields. *Nucl. Phys. B*, 44 :189–213, 1972.
- [7] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *PRL*, 13(9) :321–323, August 1964.
- [8] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *PRL*, 13(16) :508–509, October 1964.
- [9] G. S. Guralnik, C. R. Hagen, and T. W. Kibble. Global Conservation Laws and Massless Particles. *PRL*, 13(20) :585–587, November 1964.
- [10] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.
- [11] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716 :1–29, 2012.
- [12] CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1) :30–61, Sep 2012. <http://dx.doi.org/10.1016/j.physletb.2012.08.021>.
- [13] Lyndon Evans and Philip Bryant. LHC machine. *Journal of Instrumentation*, 3(08) :S08001–S08001, aug 2008. <https://doi.org/10.1088/1748-0221/3/08/s08001>.

- [14] *LEP design report*. CERN, Geneva, 1984. <http://cds.cern.ch/record/102083>.
- [15] Christiane Lefèvre. The CERN accelerator complex. Complexe des accélérateurs du CERN. <https://cds.cern.ch/record/1260465>, Dec 2008.
- [16] CMS Collaboration. Precision luminosity measurement in proton-proton collisions at  $\sqrt{s} = 13$  tev in 2015 and 2016 at cms, 2021.
- [17] CMS Collaboration. CMS Luminosity - Public Results. <http://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResult>.
- [18] ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3 :S08002, 2008.
- [19] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3 :S08003, 2008.
- [20] LHCb Collaboration. The LHCb Detector at the LHC. *JINST*, 3 :S08005, 2008.
- [21] Le LHC en bref, LHC France. <https://www.lhc-france.fr/spip.php?rubrique1>.
- [22] G. Anelli et al. The TOTEM experiment at the CERN Large Hadron Collider. *JINST*, 3 :S08007, 2008.
- [23] Eugenio Berti. The LHCf experiment : present status and physics results. 10 2017.
- [24] MoEDAL Collaboration. The Physics Programme Of The MoEDAL Experiment At The LHC. *Int. J. Mod. Phys. A*, 29 :1430050, 2014.
- [25] LHC Cross Section Working Group. *Handbook of LHC Higgs Cross Sections : 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports : Monographs. Oct 2016. <https://cds.cern.ch/record/2227475>.
- [26] CMS Collaboration. Observation of higgs boson decay to bottom quarks. *Physical Review Letters*, 121(12), Sep 2018. <http://dx.doi.org/10.1103/PhysRevLett.121.121801>.
- [27] ATLAS Collaboration. Observation of  $t\bar{t}h$  production. *Physical Review Letters*, 120(23), Jun 2018. <http://dx.doi.org/10.1103/PhysRevLett.120.231801>.
- [28] CMS ATLAS. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at  $\sqrt{s} = 7$  and 8 TeV. *JHEP*, 08 :045, 2016.
- [29] CMS collaboration. A measurement of the higgs boson mass in the diphoton decay channel. *Physics Letters B*, 805 :135425, Jun 2020. <http://dx.doi.org/10.1016/j.physletb.2020.135425>.
- [30] Combined Higgs boson production and decay measurements with up to  $137 \text{ fb}^{-1}$  of proton-proton collision data at  $\sqrt{s} = 13$  TeV. Technical report, CERN, Geneva, 2020. CMS-PAS-HIG-19-005 <http://cds.cern.ch/record/2706103>.
- [31] C. N. Yang. Selection rules for the dematerialization of a particle into two photons. *Phys. Rev.*, 77 :242–245, Jan 1950. <https://link.aps.org/doi/10.1103/PhysRev.77.242>.

- [32] ATLAS Collaboration. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. *Eur. Phys. J. C*, 75(10) :476, 2015. [Erratum : *Eur.Phys.J.C* 76, 152 (2016)].
- [33] CMS Collaboration. Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. *Phys. Rev. D*, 92(1) :012004, 2015.
- [34] B. Abi, T. Albahri, S. Al-Kilani, D. Allspach, L. P. Alonzi, A. Anastasi, A. Anisenkov, F. Azfar, K. Badgley, S. Baeßler, I. Bailey, V. A. Baranov, E. Barlas-Yucel, T. Barrett, E. Barzi, A. Basti, F. Bedeschi, A. Behnke, M. Berz, M. Bhattacharya, H. P. Binney, R. Bjorkquist, P. Bloom, J. Bono, E. Bottalico, T. Bowcock, D. Boyden, G. Cantatore, R. M. Carey, J. Carroll, B. C. K. Casey, D. Cauz, S. Ceravolo, R. Chakraborty, S. P. Chang, A. Chapelain, S. Chappa, S. Charity, R. Chislett, J. Choi, Z. Chu, T. E. Chupp, M. E. Convery, A. Conway, G. Corradi, S. Corrodi, L. Cotrozzi, J. D. Crnkovic, S. Dabagov, P. M. De Lurgio, P. T. Debevec, S. Di Falco, P. Di Meo, G. Di Sciascio, R. Di Stefano, B. Drendel, A. Driutti, V. N. Duginov, M. Eads, N. Eggert, A. Epps, J. Esquivel, M. Farooq, R. Fatemi, C. Ferrari, M. Fertl, A. Fiedler, A. T. Fienberg, A. Fioretti, D. Flay, S. B. Foster, H. Friedrich, E. Frlež, N. S. Froemming, J. Fry, C. Fu, C. Gabbanini, M. D. Galati, S. Ganguly, A. Garcia, D. E. Gastler, J. George, L. K. Gibbons, A. Gioiosa, K. L. Giovanetti, P. Girotti, W. Gohn, T. Gorringer, J. Grange, S. Grant, F. Gray, S. Haciomeroglu, D. Hahn, T. Halewood-Leagas, D. Hampai, F. Han, E. Hazen, J. Hempstead, S. Henry, A. T. Herrod, D. W. Hertzog, G. Hesketh, A. Hibbert, Z. Hodge, J. L. Holzbauer, K. W. Hong, R. Hong, M. Iacovacci, M. Incagli, C. Johnstone, J. A. Johnstone, P. Kammel, M. Kargiantoulakis, M. Karuza, J. Kaspar, D. Kawall, L. Kelton, A. Keshavarzi, D. Kessler, K. S. Khaw, Z. Khechadorian, N. V. Khomutov, B. Kiburg, M. Kiburg, O. Kim, S. C. Kim, Y. I. Kim, B. King, N. Kinnaird, M. Korostelev, I. Kourbanis, E. Kraegelo, V. A. Krylov, A. Kuchibhotla, N. A. Kuchinskiy, K. R. Labe, J. LaBounty, M. Lancaster, M. J. Lee, S. Lee, S. Leo, B. Li, D. Li, L. Li, I. Logashenko, A. Lorente Campos, A. Lucà, G. Lukicov, G. Luo, A. Lusiani, A. L. Lyon, B. MacCoy, R. Madrak, K. Makino, F. Marignetti, S. Mastroianni, S. Maxfield, M. McEvoy, W. Merritt, A. A. Mikhailichenko, J. P. Miller, S. Miozzi, J. P. Morgan, W. M. Morse, J. Mott, E. Motuk, A. Nath, D. Newton, H. Nguyen, M. Oberling, R. Osofsky, J.-F. Ostiguy, S. Park, G. Pauletta, G. M. Piacentino, R. N. Pilato, K. T. Pitts, B. Plaster, D. Počanić, N. Pohlman, C. C. Polly, M. Popovic, J. Price, B. Quinn, N. Raha, S. Ramachandran, E. Ramberg, N. T. Rider, J. L. Ritchie, B. L. Roberts, D. L. Rubin, L. Santi, D. Sathyan, H. Schellman, C. Schlesier, A. Schreckenberger, Y. K. Semertzidis, Y. M. Shatunov, D. Shemyakin, M. Shenk, D. Sim, M. W. Smith, A. Smith, A. K. Soha, M. Sorbara, D. Stöckinger, J. Stapleton, D. Still, C. Stoughton, D. Stratakis, C. Strohman, T. Stuttard, H. E. Swanson, G. Sweetmore, D. A. Sweigart, M. J. Syphers, D. A. Tarazona, T. Teubner, A. E. Tewsley-Booth, K. Thomson, V. Tishchenko, N. H. Tran, W. Turner, E. Valetov, D. Vasilkova, G. Venanzoni, V. P. Volnykh, T. Walton, M. Warren, A. Weisskopf, L. Welty-Rieger, M. Whitley, P. Winter, A. Wolski, M. Wormald, W. Wu, and C. Yoshikawa. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.*, 126 :141801, Apr 2021. <https://link.aps.org/doi/10.1103/PhysRevLett.126.141801>.
- [35] LHCb collaboration. Test of lepton universality in beauty-quark decays, 2021.

- [36] D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolte, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. First-year wilkinson microwave anisotropy probe ( WMAP ) observations : Determination of cosmological parameters. *The Astrophysical Journal Supplement Series*, 148(1) :175–194, sep 2003. <https://doi.org/10.1086/377226>.
- [37] Andrei D Sakharov. Baryon asymmetry of the universe. *Soviet Physics Uspekhi*, 34(5) :417, 1991. <http://stacks.iop.org/0038-5670/34/i=5/a=A15>.
- [38] Y. Fukuda, T. Hayakawa, E. Ichihara, K. Inoue, K. Ishihara, H. Ishino, Y. Itow, T. Kajita, J. Kameda, S. Kasuga, K. Kobayashi, Y. Kobayashi, Y. Koshio, K. Martens, M. Miura, M. Nakahata, S. Nakayama, A. Okada, M. Oketa, K. Okumura, M. Ota, N. Sakurai, M. Shiozawa, Y. Suzuki, Y. Takeuchi, Y. Totsuka, S. Yamada, M. Earl, A. Habig, J. T. Hong, E. Kearns, S. B. Kim, M. Masuzawa, M. D. Messier, K. Scholberg, J. L. Stone, L. R. Sulak, C. W. Walter, M. Goldhaber, T. Barszczak, W. Gajewski, P. G. Halverson, J. Hsu, W. R. Kropp, L. R. Price, F. Reines, H. W. Sobel, M. R. Vagins, K. S. Ganezer, W. E. Keig, R. W. Ellsworth, S. Tasaka, J. W. Flanagan, A. Kibayashi, J. G. Learned, S. Matsuno, V. Stenger, D. Takemori, T. Ishii, J. Kanzaki, T. Kobayashi, K. Nakamura, K. Nishikawa, Y. Oyama, A. Sakai, M. Sakuda, O. Sasaki, S. Echigo, M. Kohama, A. T. Suzuki, T. J. Haines, E. Blaufuss, R. Sanford, R. Svoboda, M. L. Chen, Z. Conner, J. A. Goodman, G. W. Sullivan, M. Mori, J. Hill, C. K. Jung, C. Mauger, C. McGrew, E. Sharkey, B. Viren, C. Yanagisawa, W. Doki, T. Ishizuka, Y. Kitaguchi, H. Koga, K. Miyano, H. Okazawa, C. Saji, M. Takahata, A. Kusano, Y. Nagashima, M. Takita, T. Yamaguchi, M. Yoshida, M. Etoh, K. Fujita, A. Hasegawa, T. Hasegawa, S. Hatakeyama, T. Iwamoto, T. Kinebuchi, M. Koga, T. Maruyama, H. Ogawa, A. Suzuki, F. Tsushima, M. Koshiya, M. Nemoto, K. Nishijima, T. Futagami, Y. Hayato, Y. Kanaya, K. Kaneyuki, Y. Watanabe, D. Kielczewska, R. Doyle, J. George, A. Stachyra, L. Wai, J. Wilkes, and K. Young. Measurements of the solar neutrino flux from super-kamiokande’s first 300 days. *Phys. Rev. Lett.*, 81 :1158–1162, Aug 1998. <https://link.aps.org/doi/10.1103/PhysRevLett.81.1158>.
- [39] Giacomo Cacciapaglia, Aldo Deandrea, Suzanne Gascon-Shotkin, Solène Le Corre, Morgan Lethuillier, and Junqian Tao. Search for a lighter Higgs boson in Two Higgs Doublet Models. *Journal of High Energy Physics*, 2016(12), Dec 2016. [http://dx.doi.org/10.1007/JHEP12\(2016\)068](http://dx.doi.org/10.1007/JHEP12(2016)068).
- [40] F. Wilczek. Problem of strong P and T invariance in the presence of instantons. *PRL*, 40(5) :279–282, January 1978.
- [41] Lars Fromme, Stephan J. Huber, and Michael Seniuch. Baryogenesis in the two-Higgs doublet model. *JHEP*, 11 :038, 2006.
- [42] John F. Gunion and Howard E. Haber. Conditions for CP-violation in the general two-Higgs-doublet model. *Phys. Rev. D*, 72 :095002, 2005.
- [43] Antonio Pich and Paula Tuzon. Yukawa Alignment in the Two-Higgs-Doublet Model. *Phys. Rev. D*, 80 :091702, 2009.



- [44] Le Corre, Solen. *Higgs boson phenomenology beyond the Standard Model*. PhD thesis, 2018. <http://www.theses.fr/2018LYSE1028>.
- [45] Sebastian Bruggisser, Benedict Von Harling, Oleksii Matsedonskyi, and Géraldine Servant. Baryon Asymmetry from a Composite Higgs Boson. *Phys. Rev. Lett.*, 121(13) :131801, 2018.
- [46] R. Sekhar Chivukula. Lectures on technicolor and compositeness. In *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2000) : Flavor Physics for the Millennium*, 6 2000.
- [47] Francesco Sannino. Conformal Dynamics for TeV Physics and Cosmology. *Acta Phys. Polon. B*, 40 :3533–3743, 2009.
- [48] David B. Kaplan. Flavor at ssc energies : A new mechanism for dynamically generated fermion masses. *Nuclear Physics B*, 365(2) :259–278, 1991. <https://www.sciencedirect.com/science/article/pii/S0550321305800215>.
- [49] Ari Hietanen, Randy Lewis, Claudio Pica, and Francesco Sannino. Fundamental Composite Higgs Dynamics on the Lattice : SU(2) with Two Flavors. *JHEP*, 07 :116, 2014.
- [50] Ben Gripaios, Alex Pomarol, Francesco Riva, and Javi Serra. Beyond the Minimal Composite Higgs Model. *JHEP*, 04 :070, 2009.
- [51] Sidney R. Coleman, J. Wess, and Bruno Zumino. Structure of phenomenological Lagrangians. 1. *Phys. Rev.*, 177 :2239–2247, 1969.
- [52] Curtis G. Callan, Jr., Sidney R. Coleman, J. Wess, and Bruno Zumino. Structure of phenomenological Lagrangians. 2. *Phys. Rev.*, 177 :2247–2250, 1969.
- [53] Giacomo Cacciapaglia and Francesco Sannino. Fundamental composite (goldstone) higgs dynamics. *Journal of High Energy Physics*, 2014(4), Apr 2014. [http://dx.doi.org/10.1007/JHEP04\(2014\)111](http://dx.doi.org/10.1007/JHEP04(2014)111).
- [54] Stephen P. Martin. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.*, 18 :1–98, 1998.
- [55] Delaunay, Cédric. *Brisure de symétrie électrofaible : origine et conséquences*. PhD thesis, 2008. <http://www.theses.fr/2008PA112353>.
- [56] Ulrich Ellwanger. Higgs Bosons in the Next-to-Minimal Supersymmetric Standard Model at the LHC. *Eur. Phys. J. C*, 71 :1782, 2011.
- [57] Lisa Randall and Raman Sundrum. Large Mass Hierarchy from a Small Extra Dimension. *PRL*, 83(17) :3370–3373, October 1999.
- [58] CMS detector and collaboration. <https://cms.cern/>.
- [59] V.I. Klyukhin, A. Ball, F. Bergsma, D. Campi, B. Cure, A. Gaddi, H. Gerwig, A. Herve, J. Korienek, F. Linde, and et al. Measurement of the CMS magnetic field. *IEEE Transactions on Applied Superconductivity*, 18(2) :395–398, Jun 2008. <http://dx.doi.org/10.1109/TASC.2008.921242>.

- [60] CMS Collaboration. *The CMS magnet project : Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. <http://cds.cern.ch/record/331056>.
- [61] CMS Collaboration. Precision measurement of the structure of the CMS inner tracking system using nuclear interactions. *Journal of Instrumentation*, 13(10) :P10034, Oct 2018. <http://dx.doi.org/10.1088/1748-0221/13/10/P10034>.
- [62] Martin Lipinski. The Phase-1 Upgrade of the CMS Pixel Detector. Technical report, CERN, Geneva, May 2017. <https://cds.cern.ch/record/2265423>.
- [63] CMS Collaboration. The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment. *JINST*, V3 :S08004. 361 p, 2008. <https://cds.cern.ch/record/1129810>.
- [64] CMS ECAL Response to Laser Light. Mar 2019. <https://cds.cern.ch/record/2668200>.
- [65] Cristina Biino. The CMS electromagnetic calorimeter : overview, lessons learned during Run 1 and future projections. *Journal of Physics : Conference Series*, 587 :012001, feb 2015. <https://doi.org/10.1088/1742-6596/587/1/012001>.
- [66] CMS Collaboration. Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV. *JINST*, 10(08) :P08010, 2015.
- [67] CMS Collaboration. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *Journal of Instrumentation*, 2(04) :P04004–P04004, Apr 2007. <https://doi.org/10.1088/1748-0221/2/4/P04004>.
- [68] H. Seo. RPC hit contribution to CMS muon reconstruction at LHC. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, 718 :437 – 439, 2013. Proceedings of the 12th Pisa Meeting on Advanced Detectors.
- [69] J. G. Layter. *The CMS muon project : Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. <http://cds.cern.ch/record/343814>.
- [70] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10) :P10003–P10003, Oct 2017. <http://dx.doi.org/10.1088/1748-0221/12/10/P10003>.
- [71] T Speer, W Adam, R Frühwirth, A Strandlie, T Todorov, and M Winkler. Track Reconstruction in the CMS Tracker. Technical Report CMS-CR-2005-014, CERN, Geneva, Jul 2005. <http://cds.cern.ch/record/884424>.
- [72] Paolo Azzurri. Track reconstruction performance in cms. *Nuclear Physics B - Proceedings Supplements*, 197(1) :275–278, Dec 2009. <http://dx.doi.org/10.1016/j.nuclphysbps.2009.10.084>.
- [73] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(10) :P10009–P10009, Oct 2014. <http://dx.doi.org/10.1088/1748-0221/9/10/P10009>.

- [74] CMS Collaboration. Electron and photon reconstruction and identification with the cms experiment at the cern lhc. *Journal of Instrumentation*, 16(05) :P05014, May 2021. <http://dx.doi.org/10.1088/1748-0221/16/05/P05014>.
- [75] The CMS collaboration. Performance of CMS muon reconstruction in pp collision events at  $\sqrt{s} = 7$  TeV. *Journal of Instrumentation*, 7(10) :P10002–P10002, Oct 2012. <http://dx.doi.org/10.1088/1748-0221/7/10/P10002>.
- [76] W Adam, R Frühwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G : Nuclear and Particle Physics*, 31(9) :N9–N20, Jul 2005. <http://dx.doi.org/10.1088/0954-3899/31/9/N01>.
- [77] Reconstruction and identification of  $\tau$  lepton decays to hadrons and  $\nu_\tau$  CMS. *Journal of Instrumentation*, 11(01) :P01019–P01019, Jan 2016. <http://dx.doi.org/10.1088/1748-0221/11/01/P01019>.
- [78] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04) :063–063, Apr 2008. <http://dx.doi.org/10.1088/1126-6708/2008/04/063>.
- [79] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), Jul 2014. [http://dx.doi.org/10.1007/JHEP07\(2014\)079](http://dx.doi.org/10.1007/JHEP07(2014)079).
- [80] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11) :852–867, Jun 2008. <http://dx.doi.org/10.1016/j.cpc.2008.01.036>.
- [81] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506 :250–303, 2003.
- [82] S. Banerjee. CMS simulation software. *J. Phys. Conf. Ser.*, 396 :022003, 2012.
- [83] Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at  $\sqrt{s} = 7$  TeV. *Journal of Instrumentation*, 8(09) :P09009–P09009, sep 2013. <https://doi.org/10.1088/1748-0221/8/9/P09009>.
- [84] CMS Collaboration. Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter. *Eur. Phys. J. C*, 46S1 :23–35, 2006.
- [85] CMS Collaboration. Reconstruction of signal amplitudes in the CMS electromagnetic calorimeter in the presence of overlapping proton-proton interactions. *Journal of Instrumentation*, 15(10) :P10002–P10002, Oct 2020. <http://dx.doi.org/10.1088/1748-0221/15/10/P10002>.
- [86] CMS Collaboration. CMS ECAL with 2017 data. *CMS-DP-2018-015; CERN-CMS-DP-2018-015*, May 2018. <https://cds.cern.ch/record/2319285>.

- [87] Benoit Courbon. *Reconstruction des photons et recherche d'un second boson de Higgs dans le canal di-photon au sein de l'expérience CMS au LHC*. PhD thesis, 2016. <http://www.theses.fr/2016LYSE1187>.
- [88] CMS Collaboration. Observation of the diphoton decay of the higgs boson and measurement of its properties. *The European Physical Journal C*, 74(10), Oct 2014. <http://dx.doi.org/10.1140/epjc/s10052-014-3076-z>.
- [89] Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at  $\sqrt{s}=8$  TeV. *Journal of Instrumentation*, 10(06):P06005–P06005, Jun 2015. <https://doi.org/10.1088/1748-0221/10/06/P06005>.
- [90] CMS Collaboration. Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at  $\sqrt{s}=13$  TeV. *Journal of High Energy Physics*, 2021(7), Jul 2021. [http://dx.doi.org/10.1007/JHEP07\(2021\)027](http://dx.doi.org/10.1007/JHEP07(2021)027).
- [91] CMS collaboration. Measurements of the production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel. *Physical Review Letters*, 125(6), Aug 2020. <http://dx.doi.org/10.1103/PhysRevLett.125.061801>.
- [92] Particle Data Group and al. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. <https://doi.org/10.1093/ptep/ptaa104>.
- [93] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s}=13$  TeV. *Journal of Instrumentation*, 13(06):P06015–P06015, Jun 2018. <http://dx.doi.org/10.1088/1748-0221/13/06/P06015>.
- [94] Camille Camen. *Recherche d'un second boson de Higgs de masse  $m_H < 110$  GeV dans le canal di-photon au sein de l'expérience CMS au LHC*. PhD thesis, 2021. <http://www.theses.fr/s200755>.
- [95] Louis Sgandurra. *Calibration du calorimètre électromagnétique à l'aide des événements  $Z \rightarrow \mu\mu\gamma$  et recherches de bosons de Higgs additionnels dans le canal  $H \rightarrow \gamma\gamma$  dans l'expérience CMS au LHC*. PhD thesis, 2014. <http://www.theses.fr/2014LYO10135>.
- [96] Olivier Bondu. *Étude des désintégrations radiatives  $Z \rightarrow \mu\mu\gamma$  et recherches du boson de Higgs dans le canal  $H \rightarrow \gamma\gamma$  dans l'expérience CMS au LHC (CERN)*. PhD thesis, 2012. <http://www.theses.fr/2012LYO10198>.
- [97] ECAL Detector Performance, 2011 Data. 5 2012.
- [98] F. James and M. Roos. Minuit - a system for function minimization and analysis of the parameter errors and correlations. *Computer Physics Communications*, 10(6):343–367, 1975. <https://www.sciencedirect.com/science/article/pii/0010465575900399>.
- [99] M Williams. How good are your fits? unbinned multivariate goodness-of-fit tests in high energy physics. *Journal of Instrumentation*, 5(09):P09004–P09004, Sep 2010. <http://dx.doi.org/10.1088/1748-0221/5/09/P09004>.

- [100] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, M. Friedl, and et al. Observation of the diphoton decay of the higgs boson and measurement of its properties. *The European Physical Journal C*, 74(10), Oct 2014. <http://dx.doi.org/10.1140/epjc/s10052-014-3076-z>.
- [101] First results on Higgs  $\rightarrow \gamma\gamma$  at 13 TeV. Technical Report CMS-PAS-HIG-15-005, CERN, Geneva, 2016. <https://cds.cern.ch/record/2140979>.
- [102] Updated measurements of Higgs boson production in the diphoton decay channel at  $\sqrt{s} = 13$  TeV in pp collisions at CMS. Technical Report CMS-PAS-HIG-16-020, CERN, Geneva, 2016. <https://cds.cern.ch/record/2205275>.
- [103] CMS Collaboration. Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at  $\sqrt{s}=8$  and 13 TeV. *Physics Letters B*, 793 :320–347, Jun 2019. <http://dx.doi.org/10.1016/j.physletb.2019.03.064>.
- [104] The CMS Collaboration. Search for low mass resonances in the diphoton final state in pp collisions at  $\sqrt{s} = 13$  TeV with the 2017 and 2018 dataset. *CMS Analysis note*, 2018/249, 2020.
- [105]  $H \rightarrow \gamma\gamma$  working group. Common tools for analyses of Higgs boson decay in the diphoton final state. *CMS Analysis note*, 2019/149, 2019.
- [106] Enrico Bothmann, Gurpreet Singh Chahal, Stefan Höche, Johannes Krause, Frank Krauss, Silvan Kuttimalai, Sebastian Liebschner, Davide Napoletano, Marek Schönherr, Holger Schulz, and et al. Event generation with sherpa 2.2. *SciPost Physics*, 7(3), Sep 2019. <http://dx.doi.org/10.21468/SciPostPhys.7.3.034>.
- [107] The CMS Collaboration. Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Journal of High Energy Physics*, 2018(11), Nov 2018. [http://dx.doi.org/10.1007/JHEP11\(2018\)185](http://dx.doi.org/10.1007/JHEP11(2018)185).
- [108] Eleftherios Spyromitros-Xioufis, Grigorios Tsoumakas, William Groves, and Ioannis Vlahavas. Multi-target regression via input space expansion : treating targets as inputs. *Machine Learning*, 104(1) :55–98, Feb 2016. <http://dx.doi.org/10.1007/s10994-016-5546-z>.
- [109] Edward John Titman Scott. Measurements of Higgs boson properties using the diphoton decay channel with the Compact Muon Solenoid experiment, 2019. <http://cds.cern.ch/record/2686716>.
- [110] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1) :60 – 62, 1938. <https://doi.org/10.1214/aoms/1177732360>.
- [111] P.D. Dauncey, M. Kenzie, N. Wardle, and G.J. Davies. Handling uncertainties in background shapes : the discrete profiling method. *Journal of Instrumentation*, 10(04) :P04015–P04015, Apr 2015. <http://dx.doi.org/10.1088/1748-0221/10/04/P04015>.



- [112] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04 :040, 2015.
- [113] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43 :023001, 2016.
- [114] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C.-P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Physical Review D*, 93(3), Feb 2016. <http://dx.doi.org/10.1103/PhysRevD.93.033006>.
- [115] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the lhc era : Mmht 2014 pdfs. *The European Physical Journal C*, 75(5), May 2015. <http://dx.doi.org/10.1140/epjc/s10052-015-3397-6>.
- [116] CMS luminosity measurement for the 2017 data-taking period at  $\sqrt{s} = 13$  TeV. Technical report, CERN, Geneva, 2018. <https://cds.cern.ch/record/2621960>.
- [117] CMS luminosity measurement for the 2018 data-taking period at  $\sqrt{s} = 13$  TeV. Technical report, CERN, Geneva, 2019. <https://cds.cern.ch/record/2676164>.
- [118] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71 :1554, 2011. [Erratum : *Eur.Phys.J.C* 73, 2501 (2013)].
- [119] Procedure for the LHC Higgs boson search combination in Summer 2011. Technical report, CERN, Geneva, Aug 2011. <https://cds.cern.ch/record/1379837>.
- [120] Alexander L. Read. Modified frequentist analysis of search results (The CL(s) method). In *Workshop on Confidence Limits*, 8 2000.
- [121] Search for new resonances in the diphoton final state in the mass range between 80 and 110 GeV in pp collisions at  $\sqrt{s} = 8$  TeV. Technical report, CERN, Geneva, 2015. <https://cds.cern.ch/record/2063739>.
- [122] Search for resonances in the 65 to 110 GeV diphoton invariant mass range using 80 fb<sup>-1</sup> of pp collisions collected at  $\sqrt{s} = 13$  TeV with the ATLAS detector. Technical report, CERN, Geneva, Jul 2018. <https://cds.cern.ch/record/2628760>.
- [123] CMS egamma results. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsEGM>.
- [124] G. Brooijmans, A. Buckley, S. Caron, A. Falkowski, B. Fuks, A. Gilbert, W. J. Murray, M. Nardecchia, J. M. No, R. Torre, T. You, G. Zevi Della Porta, G. Alguero, J. Y. Araz, S. Banerjee, G. Bélanger, T. Berger-Hryn'ova, J. Bernigaud, A. Bharucha, D. Buttazzo, J. M. Butterworth, G. Cacciapaglia, A. Coccaro, L. Corpe, N. Desai, A. De Simone, J. Dutta, A. Farbin, S. Fichet, T. Flacke, S. Ganguly, S. Gascon-Shotkin, T. Gonzalo, A. Goudelis, P. Gras, R. S. Gupta, M. Habedank, J. Heisig, L. Hendriks, S. Jain, A. Jueid, D. Kar, S. Kraml, A. Kvellestad, K. D. Lane, J. Langford, A. Leinweber, A. Lesauvage, A. Lessa, J. Mamuzic, D. Marzocca, E. Merényi, A. Mjallal, K. A. Mohan, A. Morandini, G. Moreau, S. Mutzel, C. Nellist, S. Otten, P. Pani, L. Panizzi, M. Pierini, G. Polesello, W. Porod, H. B.

- Prosper, A. Pukhov, J. Quevillon, A. Raklev, H. Reyes-González, R. Ruiz, R. Ruiz de Austri, J. Schouwenberg, S. Sekmen, D. Sengupta, C. Smith, M. Son, D. Sperka, B. Stienen, K. Tobioka, M. van Beekveld, E. Venturini, N. Vignaroli, R. Vilalta, W. Waltenberger, N. Wardle, G. Watt, M. White, S. L. Williamson, D. Yallup, and J. Zurita. Les houches 2019 physics at tev colliders : New physics working group report, 2020.
- [125] Howard Georgi, David B. Kaplan, and Lisa Randall. Manifesting the Invisible Axion at Low-energies. *Phys. Lett.*, 169B :73–78, 1986.
- [126] I. Brivio, M. B. Gavela, L. Merlo, K. Mimasu, J. M. No, R. del Rey, and V. Sanz. ALPs Effective Field Theory and Collider Signatures. *Eur. Phys. J.*, C77(8) :572, 2017.
- [127] Martin Bauer, Matthias Neubert, and Andrea Thamm. Collider Probes of Axion-Like Particles. *JHEP*, 12 :044, 2017.
- [128] Martin Bauer, Matthias Neubert, and Andrea Thamm. LHC as an Axion Factory : Probing an Axion Explanation for  $(g - 2)_\mu$  with Exotic Higgs Decays. *Phys. Rev. Lett.*, 119(3) :031802, 2017.
- [129] Jérémie Quevillon and Christopher Smith. Axions are blind to anomalies. *Eur. Phys. J.*, C79(10) :822, 2019.
- [130] Andrew S. Chisholm, Silvan Kuttimalai, Konstantinos Nikolopoulos, and Michael Spannowsky. Measuring rare and exclusive Higgs boson decays into light resonances. *Eur. Phys. J.*, C76(9) :501, 2016.
- [131] Sebastian Baum and Nausheen R. Shah. Two Higgs Doublets and a Complex Singlet : Disentangling the Decay Topologies and Associated Phenomenology. 2018. [JHEP12,044(2018)].
- [132] Jose Miguel No. Looking through the pseudoscalar portal into dark matter : Novel mono-Higgs and mono-Z signatures at the LHC. *Phys. Rev.*, D93(3) :031701, 2016.
- [133] Georges Aad et al. Measurements of  $Z\gamma$  and  $Z\gamma\gamma$  production in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *Phys. Rev.*, D93(11) :112002, 2016.
- [134] Search for a Higgs boson decaying to four photons through light CP-odd scalar coupling using 7 TeV  $pp$  collision data taken with ATLAS detector at the LHC. 2012.
- [135] M. Aaboud et al. Searches for the  $Z\gamma$  decay mode of the Higgs boson and for new high-mass resonances in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *JHEP*, 10 :112, 2017.
- [136] Albert M. Sirunyan et al. Search for the decay of a Higgs boson in the  $\ell\ell\gamma$  channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP*, 11 :152, 2018.
- [137] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02 :057, 2014.

- [138] Charalampos Anastasiou, Claude Duhr, Falko Dulat, Elisabetta Furlan, Thomas Gehrmann, Franz Herzog, Achilleas Lazopoulos, and Bernhard Mistlberger. High precision determination of the gluon fusion Higgs boson cross-section at the LHC. *JHEP*, 05 :058, 2016.
- [139] C. Uzziel Perez. Endcap Photon Fake Rate Study : Background Estimation for ADD and RS Graviton Searches. 2015.
- [140] Morad Aaboud et al. Search for Higgs boson decays to beyond-the-Standard-Model light bosons in four-lepton events with the ATLAS detector at  $\sqrt{s} = 13$  TeV. *JHEP*, 06 :166, 2018.
- [141] G. Alonso-Álvarez, M. B. Gavela, and P. Quilez. Axion couplings to electroweak gauge bosons. *Eur. Phys. J.*, C79(3) :223, 2019.
- [142] Aqeel Ahmed, Barry M. Dillon, Bohdan Grzadkowski, John F. Gunion, and Yun Jiang. Implications of the absence of high-mass radion signals. *Phys. Rev.*, D95(9) :095019, 2017.
- [143] Andrei Angelescu, Grégory Moreau, and François Richard. Scalar production in association with a Z boson at the LHC and ILC : The mixed Higgs-radion case of warped models. *Phys. Rev.*, D96(1) :015019, 2017.
- [144] Sylvain Fichet and Gero von Gersdorff. Anomalous gauge couplings from composite Higgs and warped extra dimensions. *JHEP*, 03 :102, 2014.
- [145] S. Fichet, G. von Gersdorff, O. Kepka, B. Lenzi, C. Royon, and M. Saimpert. Probing new physics in diphoton production with proton tagging at the Large Hadron Collider. *Phys. Rev.*, D89 :114004, 2014.
- [146] Cristian Baldenegro, Sylvain Fichet, Gero von Gersdorff, and Christophe Royon. Probing the anomalous  $\gamma\gamma\gamma Z$  coupling at the LHC with proton tagging. *JHEP*, 06 :142, 2017.
- [147] Alexander Belyaev, Giacomo Cacciapaglia, Haiying Cai, Gabriele Ferretti, Thomas Flacke, Alberto Parolini, and Hugo Serodio. Di-boson signatures as Standard Candles for Partial Compositeness. *JHEP*, 01 :094, 2017.
- [148] Sylvain Fichet, Gero von Gersdorff, Eduardo Ponton, and Rogerio Rosenfeld. The Excitation of the Global Symmetry-Breaking Vacuum in Composite Higgs Models. *JHEP*, 09 :158, 2016.
- [149] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks. Feynrules 2.0 — a complete toolbox for tree-level phenomenology. *Computer Physics Communications*, 185(8) :2250–2300, Aug 2014. <http://dx.doi.org/10.1016/j.cpc.2014.04.012>.
- [150] G. Brooijmans et al. Les Houches 2017 : Physics at TeV Colliders New Physics Working Group Report. In *Les Houches 2017 : Physics at TeV Colliders Standard Model Working Group Report*, 2018.
- [151] Eric Conte, Benjamin Fuks, and Guillaume Serret. Madanalysis 5, a user-friendly framework for collider phenomenology. *Computer Physics Communications*, 184(1) :222–256, Jan 2013. <http://dx.doi.org/10.1016/j.cpc.2012.09.009>.



- [152] G. Cacciapaglia, S. Gascon-Shotkin, A. Lesauvage, N. Manglani, and K. Sridhar. Testing charge-radius coupling of the composite Higgs boson at hadron colliders, 2021. <https://arxiv.org/abs/2108.03005>.
- [153] Jamison Galloway, Jared A. Evans, Markus A. Luty, and Ruggero Altair Tacchi. Minimal Conformal Technicolor and Precision Electroweak Tests. *JHEP*, 10 :086, 2010.
- [154] Thomas A. Ryttov and Francesco Sannino. Ultraminimal technicolor and its dark matter technicolor interacting massive particles. *Physical Review D*, 78(11), Dec 2008. <http://dx.doi.org/10.1103/PhysRevD.78.115010>.
- [155] Roshan Foadi, Mads T. Frandsen, and Francesco Sannino. Technicolor Dark Matter. *Phys. Rev. D*, 80 :037702, 2009.
- [156] Ari Hietanen, Randy Lewis, Claudio Pica, and Francesco Sannino. Composite Goldstone Dark Matter : Experimental Predictions from the Lattice. *JHEP*, 12 :130, 2014.
- [157] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. Fastjet user manual. *The European Physical Journal C*, 72(3), Mar 2012. <http://dx.doi.org/10.1140/epjc/s10052-012-1896-2>.
- [158] A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging. 7 2009.
- [159] FCC Collaboration. FCC-ee : The Lepton Collider.
- [160] T. K. Charles et al. The Compact Linear Collider (CLIC) - 2018 Summary Report. 2/2018, 12 2018.
- [161] Hiroaki Aihara, Jonathan Bagger, Philip Bambade, Barry Barish, Ties Behnke, Alain Bellerive, Mikael Berggren, James Brau, Martin Breidenbach, Ivanka Bozovic-Jelisavcic, Philip Burrows, Massimo Caccia, Paul Colas, Dmitri Denisov, Gerald Eigen, Lyn Evans, Angeles Faus-Golfe, Brian Foster, Keisuke Fujii, Juan Fuster, Frank Gaede, Jie Gao, Paul Grannis, Christophe Grojean, Andrew Hutton, Marek Idzik, Andrea Jeremie, Kiyotomo Kawagoe, Sachio Komamiya, Tadeusz Lesiak, Aharon Levy, Benno List, Jenny List, Shinichiro Michizono, Akiya Miyamoto, Joachim Mnich, Hugh Montgomery, Hitoshi Murayama, Olivier Napoly, Yasuhiro Okada, Carlo Pagani, Michael Peskin, Roman Poeschl, Francois Richard, Aidan Robson, Thomas Schoerner-Sadenius, Marcel Stanitzki, Steinar Stapnes, Jan Strube, Atsuto Suzuki, Junping Tian, Maksym Titovx, Marcel Vos, Nicholas Walkerx, Hans Weise, Andrew White, Graham Wilson, Marc Winter, Sakue Yamada, Akira Yamamoto, Hitoshi Yamamoto, and Satoru Yamashita. The international linear collider. a global project, 2019.
- [162] Andreas Hocker et al. TMVA - Toolkit for Multivariate Data Analysis. 3 2007.
- [163] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1) :119–139, 1997. <https://www.sciencedirect.com/science/article/pii/S002200009791504X>.
- [164] Jerome H Friedman. Greedy function approximation : a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.