Study of the Higgs boson decay into b-quarks with the ATLAS experiment - run 2

Charles Delporte

université PARIS-SACLAY

ÉCOLE DOCTORALE PHENIICS

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

NNT : 2018SACLS404

Thèse de doctorat

Study of the Higgs boson decay into b-quarks with the ATLAS experiment run 2

Thèse de doctorat de l'Université Paris-Saclay
préparée à Université Paris-Sud

Ecole doctorale n°576 Particules, Hadrons, Énergie, Noyau, Instrumentation, Imagerie, Cosmos et Simulation (PHENIICS)
Spécialité de doctorat : Physique des particules

Thèse présentée et soutenue à Orsay, le 19 Octobre 2018, par

CHARLES DELPORTE

Composition du Jury :

Achille STOCCHI Université Paris Saclay (LAL)	Président
Giovanni MARCHIORI Sorbonne Université (LPNHE)	Rapporteur
Paolo MERIDIANI Université de Rome (INFN), CERN	Rapporteur
Matteo CACCIARI Université Paris Diderot (LPTHE)	Examinateur
Frédéric DELIOT Université Paris Saclay (CEA)	Examinateur
Jean-Baptiste DE VIVIE Université Paris Saclay (LAL)	Directeur de thèse
Daniel FOURNIER Université Paris Saclay (LAL)	Invité

Synthèse

Le Modèle Standard fournit un modèle élégant à la description des particules élémentaires, leurs propriétés et leurs interactions. Les particules y sont distinguées par leur spin, qui permet de les qualifier de fermions ou bosons, ainsi que les processus auxquels elles sont susceptibles de participer. Les interactions électrofaible et forte font intervenir les bosons γ (photon), Z, et $W^{\pm}$ d'une part, et les gluons d'autre part. Ces médiateurs des forces interagissent de manière sélective avec les fermions : les leptons et les quarks. L'étude des symétries dans les propriétés des particules a conduit à modéliser leurs interactions dans un environnement mathématique prenant appui sur des groupes de symétrie. Dans ce modèle, les leptons peuvent changer de nature au sein de doublets vis-à-vis de l'interaction électrofaible, ce qui est décrit par le groupe de symétrie $SU(2)_I \times U(1)_Y$. Les quarks sont porteurs d'une charge de couleur variable par interaction avec des gluons. L'interaction forte correspondant est décrite par le groupe de symétrie $SU(3)_C$. Le Modèle Standard qui en résulte a pour structure de groupe $SU(3)_C \times SU(2)_I \times U(1)_Y$ et a permis de prédire de nombreuses des propriétés des particules élémentaires.

La symétrie de jauge du Modèle Standard prédit une masse nulle des bosons Z et W, ce qui a été invalidé par des mesures expérimentales. La solution est apportée par le mécanisme de Higgs, qui requiert la brisure spontanée de la symétrie à l'aide d'un champ additionnel et prédit l'existence d'un boson scalaire nommé boson de Higgs (symbolisé par un H). La découverte d'une particule compatible avec le boson de Higgs en 2012 par les expériences ATLAS et CMS au run 1 du LHC fournit un indice fort de la brisure spontanée de symétrie, mais nécessite de vérifier les propriétés de la particule qui a été découverte.

Au run 2 du LHC, plusieurs analyses sont conduites par les expériences ATLAS et CMS, visant à mesurer les couplages du boson de Higgs aux particules du Modèle Standard. Le LHC consiste en un accélérateur circulaire à l'intérieur duquel circulent des faisceaux de protons en sens opposé. Les protons accélérés à de très hautes énergies produisent des particules lors de leurs collisions. Dans le cadre d'une analyse donnée, les évènements d'intérêt sont qualifiés de signal, par opposition aux évènements de bruit de fond. Les données analysées dans cette thèse sont issues d'évènements reconstruits dans le détecteur ATLAS, qui permet d'identifier l'ensemble des particules qui le traversent, à l'exception des neutrinos.

Les données reconstruites dans le détecteur sont comparées à des évènements issus de programmes de simulation. L'objectif de telles simulations est de pouvoir prédire au mieux la topologie des processus qui se produisent au LHC, et ainsi exclure ou contraindre les modèles théoriques. Les modèles de simulation reposent sur les connaissances théoriques et expérimentales des particules, leurs propriétés et interactions. Le processus de simulation comprend plusieurs étapes successives. La simulation d'évènements est implémentée de façon que la production de particules, leurs interactions, et leur désintégration suivent des lois de probabilité. Les programmes de simulation d'évènements, dits simulations "Monte-Carlo", incluent différents modèles et paramètres qui permettent de comparer les différentes topologies qui en résultent. Enfin, l'interaction des particules stables avec le détecteur ATLAS est modélisée par le programme GEANT 4 dans lequel est implémenté un modèle du détecteur.

De nombreuses analyses conduites par l'expérience ATLAS ciblent des processus au cours desquels des quarks beauté (b-quark) sont produits. La désintégration des particules qui en résultent (notamment des hadrons beaux) conduit à des états finals dans lesquels des jets sont reconstruits par le détecteur au moyen des calorimètres notamment. La saveur du quark ou gluon à l'origine d'un jet lui est donnée comme label. Le label des jets issus d'évènements simulés peut être connu par identification entre quarks et jets, en se basant sur leur séparations angulaires. En revanche, le label de jets reconstruits par le détecteur ATLAS ne peut être connu, les quarks n'étant pas observés directement.

Certaines propriétés des hadrons beaux, telles que leur temps de vie ou leur masse, permettent d'identifier les jets issus de leur désintégration (b-jets). Cette technique d'identification, le b-Tagging, est mise en oeuvre par l'algorithme MV2c10. Les informations sur les jets reconstruits dans le détecteur ATLAS sont utilisées par l'algorithme afin de produire une variable discriminante permettant de mesurer la probabilité qu'un jet soit issu d'un b-quark (efficacité de b-Tagging). La valeur de la coupure est ajustée selon la pureté en b-jets souhaitée dans le lot d'évènements candidats. Un jet qui satisfait la coupure est dit b-Taggé pour l'efficacité sélectionnée. Le caractère probabiliste de cette technique présente plusieurs inconvénients. D'une part, certains jets issus de b-quarks ne sont pas identifiés comme tels. D'autre part, des jets peuvent être étiquetés par erreur comme issus de la désintégration d'un b-quark.

Des désaccords entre les nombres d'évènements et de la saveur des jets sélectionnés dans les données reconstruites par le détecteur ATLAS, et les simulations supposées les reproduire, sont généralement observés. Afin d'éviter ce biais, une correction est appliquée aux évènements issus de simulations. Pour une efficacité de b-Tagging donnée, des facteurs correctifs sont employés pour pondérer les évènements de simulation afin que leur nombre reproduise de façon plus réaliste celui des évènements issus de données réelles. Ces facteurs correctifs sont calculés à partir de lots d'évènements très purs en jets issus de b-quarks, au cours d'une procédure dite de calibration intégrée. La complexité de cette procédure ne permet de la conduire que pour quelques valeurs de coupure sur la variable MV2c10 séparément. Cette méthode conduit à qualifier un jet selon deux possibilités seulement, b-Taggé ou non, sans exploiter davantage l'information apportée par la variable MV2c10.

Une calibration plus élaborée consiste à calculer des facteurs correctifs tenant compte d'intervalles de probabilité pour qu'un jet soit issu d'un b-quark. Cette calibration, dit pseudo-continue, a été élaborée au run 1 du LHC, et mise à jour au cours de ma thèse pour les données du run 2. La même stratégie a été mise en place pour les jets issus d'autres saveurs de quarks. La calibration pseudo-continue consiste à exploiter le spectre de la variable MV2c10 avec une finesse accrue, d'où un gain d'information sur la probabilité qu'un jet soit issu d'une certaine saveur de quark. La calibration pseudo-continue du b-Tagging a été employée par différentes analyses, et devrait gagner en précision avec l'amélioration des techniques de simulation.

Le b-Tagging est employé dans la recherche du signal de désintégration du boson de Higgs en paire de quarks beauté (Hbb). Cette recherche est particulièrement sensible dans le canal de production d'un boson de Higgs en association avec un boson vecteur (VH). Les modes de désintégration leptonique des bosons W ou Z apportent des signatures dont la topologie améliore la discrimination entre évènements de signal (VHbb) et d'autres processus qui constituent le bruit du fond.

La recherche du processus VHbb dans la voie de désintégration d'un boson Z en neutrinos (canal 0-lepton) fait l'objet d'études approfondies dans ce manuscrit. Dans ce canal, les processus à l'origine de plusieurs jets (multijet) sont une source de bruit de fond importante mais dont l'estimation et la réjection ont été optimisés. L'analyse, qui consistait dans un premier temps en un ensemble de coupures destinées à discriminer les évènements de signal des évènements de bruit de fond, a été enrichie par l'utilisation de techniques d'analyse multivariée. Les Arbres de Décision (Boosted Decision Trees) consistent en une technique d'apprentissage automatisée, permettant d'affiner la sélection des évènements de signal. Le choix des variables discriminantes utilisées a été étudié, ainsi que certains paramètres de l'apprentissage de l'Arbre de Décision. Le nombre d'évènements candidats au processus de signal VH(bb) est extrait de l'ajustement de la distribution de la variable multivariée des modélisations Monte-Carlo à celles des évènements reconstruits dans le détecteur. Ces travaux ont permis d'atteindre une signification de 3.5 déviations standard pour le processus $\mathrm{pp} \to V(\ell\bar{\ell}, \nu\bar{\nu})\mathrm{H}(\mathrm{b\bar{b}})$ avec les données collectées par le détecteur ATLAS en 2015 et 2016.

Ce résultat a souligné l'importance des incertitudes systématiques liées à l'efficacité de génération d'évènements de simulation. L'espace des phases associé aux évènements de signal est insuffisament couvert par les algorithmes de simulation d'évènements de bruit de fond. Les processus de bruit de fond impliquant deux quarks tops d'une part, et un boson Z accompagné de jets (Z+jets) d'autre part, on fait l'objet d'études approfondies afin d'identifier certains critères susceptibles de réduire l'incertitude liée à la statistique des évènements de simulation. Enfin, une ré-évaluation de l'incertitude associée à la modélisation de l'évènement sous-jacent et du développement des gerbes partoniques des processus de signal a révélé une surestimation de cette incertitude. L'analyse a été menée en ajoutant les données collectées en 2017, et a conduit à l'observation non ambigüe du signal de désintégration du boson de Higgs en quarks beauté.

De nouvelles techniques ont été explorées dans la recherche du processus VH(bb). Notamment, les quarks beauté issus de la désintégration d'un boson singulet de couleur tel que le boson de Higgs, sont connectés et constituent une paire de charges de "couleur opposées". Cette connexion assure une répartition topologique de hadrons ("colourflow") qu'il est possible d'identifier dans le calorimètre et trajectographe interne du détecteur ATLAS, dans la limite de l'efficacité de reconstruction du détecteur. La plupart des processus de bruit de fond ne présente pas une telle connexion de couleur entre quarks, ce qui permettrait de les discriminer des évènements du signal. Une étude de faisabilité a démontré l'intérêt limité pour cette technique avec les logiciels actuels de reconstruction. Néamoins, le bon accord entre les différents générateurs Monte-Carlo et la possibilité de combiner cette information aux autres variables cinématiques démontrent l'intérêt du colourflow lorsque la précision de reconstruction sera améliorée.

Enfin, une étude approfondie des différences de normalisation des évènements de simulation modélisant le bruit de fond Z+jets dans les canaux $\mathrm{Z} \to \nu\bar{\nu}$ et $\mathrm{Z} \to \mathrm{ee}/\mu\mu$ a été conduite. La modélisation des différents processus relève d'un processus complexe. Des incertitudes sur la forme des variables cinématiques et sur le nombre d'évènements du processus Z+jets doivent

être introduites. Dans les canaux 0-lepton et 2-leptons, un paramètre de nuisance permet de pondérer les évènements afin d'ajuster la normalisation des évènements Z+jets. En outre, dans le canal 2-leptons, plus pur, la région de faible impulsion transverse du boson Z peut être exploitée, et un paramètre de nuisance additionnel lié à l'incertitude sur les contributions relatives des régions de faible et grande impulsion transverse du boson Z (exploitée dans les deux canaux) est considérée. La comparaison d'ajustements des données dans les canaux 0-lepton et 2-leptons séparément a permis de clarifier la dégénérescence de l'impact des paramètres de nuisance précédents dans le canal 2-leptons, et incite à certaines précautions dans la comparaison des fits réalisés dans chaque canal individuellement.

Les analyses menées dans l'expérience ATLAS reposent sur l'hypothèse de couplages du boson de Higgs aux particules du Modèle Standard dépendant de leur masse. Certaines déviations des couplages du boson de Higgs aux valeurs prédites pourraient constituer un signe de physique au-delà du Modèle Standard. La mesure du couplage du boson de Higgs au quark top et un quark léger t-q-H (avec q=u,c) est motivée par sa très faible valeur prédite dans le Modèle Standard, nulle à l'ordre des arbres, et fortement supprimée pour des processus électrofaibles d'ordre supérieur. L'existence des Changements de Saveur par Courant Neutre (Flavour Changing Neutral Currents (FCNC)) prédite dans certains modèles modifie la valeur de tels couplages de plusieurs ordres de grandeur, et leur observation constituerait une preuve directe de nouvelle physique.

La production abondante des quarks top au LHC motive la recherche de désintégrations exotiques du quark top impliquant les couplages tqH. Au run 2 du LHC, la recherche du processus t $\rightarrow$ qH avec H $\rightarrow$ $\gamma\gamma$ a été conduite en utilisant 36.1 fb^{-1} de données collectées par l'expérience ATLAS. L'analyse consiste en la sélection d'évènements compatibles avec le processus de production d'une paire de quarks top. La désintégration d'un quark top dans le canal t $\rightarrow$ bW fournit une signature du mode de production, tandis que la recherche d'un état final avec deux photons candidats au processus H $\rightarrow$ $\gamma\gamma$ et un jet permet de tester le couplage tqH. Une limite supérieure sur le taux de désintégration pour le processus t $\rightarrow$ cH est extraite de l'ajustement du nombre d'évènements candidats au processus de signal sur la distribution de la masse diphoton d'évènements issus des données réelles. Cette stratégie a permis d'extraire la limite supérieure $Br < 2.2 \times 10^{-3}$ avec un niveau de confiance de 95% en utilisant les données reconstruites en 2015 et 2016.

Dans ce manuscrit, seule la sélection ciblant le canal de désintégration hadronique W $\rightarrow$ $q\bar{q}$ du boson W a fait l'objet d'optimisations, bien que l'analyse publiée en 2017 bénéficie également de la grande sensitivité du canal leptonique.

L'analyse a évolué de façon à profiter au mieux de l'ajout des données collectées en 2017. De nouvelles catégories ont été pensées afin de sélectionner des lots d'évènements avec une plus grande pureté en signal. L'utilisation du c-Tagging permet de lever la dégénérescence entre les couplages tcH et tuH mesurés par l'analyse, de même qu'une sélection additionnelle ciblant le processus de production d'un quark top isolé.

L'utilisation d'Arbres de Décision a permis de gagner en discrimination, en exploitant les différentes topologies des évènements de signal et de bruit de fond. Afin de réduire la sensibilité de cette méthode aux incertitudes systématiques de simulation, le choix des variables cinématiques discriminantes et la valeur de coupure sur la variable multivariée ont fait l'objet d'optimisations.

L'extraction des couplages tcH et tuH repose sur l'ajustement de la distribution $m_{\gamma\gamma}$ dans les données reconstruites. Pour les processus de bruit de fond, la forme et la normalisation de cette distribution est issue d'une fonction contrainte en dehors du bin de signal 121 GeV < $m_{\gamma\gamma}$ < 129 GeV, ce qui permet de comparer les nombres d'évènements candidats attendus et observés dans ce bin. Cette stratégie peut conduire à l'observation d'un signal spurieux créé artificiellement par la modélisation approximative du bruit de fond dans le bin sensible de l'analyse. Le choix et les paramètres de la fonction modélisant le bruit de fond ont été étudiés de façon à ajuster la distribution d'évènements de bruit de fond de manière réaliste, sans toutefois accommoder la résonnance produite par les évènements de signal autour de la valeur $m_{\gamma\gamma}$ = 125 GeV.

Des résultats préliminaires sur les limites attendues sur les rapports d'embranchement $t \to uH$ et $t \to cH$ ont été produits et démontrent des améliorations potentielles vis à vis des résultats précédents. L'utilisation de régions de contrôle exploitant des critères d'isolation alternatifs a été étudiée, et pourrait permettre d'améliorer la modélisation du bruit de fond à l'avenir.

À mon grand-père ...

List of Abbreviations

EW	**E**lectro**W**eak
QED	**Q**uantum **E**lectro **D**ynamic
QCD	**Q**uantum **C**hromo **D**ynamic
SM	**S**tandard **M**odel
ATLAS	**A** **T**horoidal **L**HC **A**pparatu**s**
CMS	**C**ompact **M**uon **S**olenoid
IBL	**I**nsertable **B**-**L**ayer
ID	**I**nner **D**etector
LAr	**L**iquid **Ar**gon
LHC	**L**arge **H**adron **C**ollider
FSR	**F**inal **S**tate **R**adiations
ISR	**I**nitial **S**tate **R**adiations
LO	**L**eading **O**rder
MC	**M**onte-**C**arlo
ME	**M**atrix **E**lement
MPI	**M**ulti-**P**arton **I**nteraction
NLO	**N**ext-to **L**eading **O**rder
PDF	**P**arton **D**istribution **F**unction
PS	**P**arton **S**hower
UE	**U**nderlying **E**vents
PDF	**P**robability **D**ensity **F**unction
SF	**S**cale **F**actor
SV	**S**econdary **V**ertex
WP	**W**orking **P**oint
BDT	**B**oosted **D**ecision **T**ree
CR	**C**ontrol **R**egion
DT	**D**ecision **T**ree
HF	**H**eavy **F**lavour
MLE	**M**aximum-**L**ikelihood **E**stimate
NP	**N**uisance **P**arameter
SR	**S**ignal **R**egion
FCNC	**F**lavour **C**hanging **N**eutral **C**urrent(s)

Contents

Synthèse iii

Personal contributions 1

Introduction 2

1 Theory of the Standard Model of elementary particles 4
1.1 Algebraic framework : group theory . . . 4
1.1.1 Definitions . . . 4
1.1.2 Group representation . . . 4
1.1.3 Lie groups . . . 5
1.2 Fundamental concepts . . . 5
1.2.1 Hamiltonian mechanics . . . 6
1.2.2 Symmetries . . . 6
1.2.2.1 Gauge transformations . . . 6
1.2.2.2 Noether theorem . . . 7
1.3 The Standard Model . . . 7
1.3.1 Particles content . . . 7
1.3.2 Gauge theories of the Standard Model . . . 7
1.3.2.1 Gauge theories . . . 8
1.3.2.2 The Electroweak interaction . . . 9
1.3.2.3 Quantum Chromodynamics . . . 10
1.3.3 Spontaneous symmetry breaking . . . 11
1.4 Phenomenology of the Higgs boson . . . 12
1.4.1 Production modes . . . 12
1.4.2 Decay channels . . . 12
1.4.3 Motivations for Higgs boson searches . . . 14
1.5 Problems of the Standard Model . . . 14
1.5.1 Neutrinos . . . 14
1.5.2 Dark matter and dark energy . . . 15
1.5.3 Gravity . . . 15
1.5.4 The hierarchy problem . . . 15
1.5.5 Unification . . . 15
1.5.6 Matter-antimatter asymmetry . . . 16

2 The LHC and the ATLAS experiment 17
2.1 The Large Hadron Collider . . . 17
2.1.1 Acceleration chain . . . 17
2.1.2 Collision parameters . . . 17
2.1.3 Experiments . . . 19
2.1.4 Data taking . . . 19
2.1.4.1 Run 1 . . . 19
2.1.4.2 Run 2 . . . 19
2.2 The ATLAS detector . . . 19

2.2.1 Inner detector . . . 21
2.2.1.1 Insertable B-Layer . . . 22
2.2.1.2 Pixel detector . . . 22
2.2.1.3 SemiConductor Tracker . . . 23
2.2.1.4 Transition Radiation Tracker . . . 23
2.2.2 Calorimeters . . . 23
2.2.2.1 Electromagnetic calorimeter . . . 24
2.2.2.2 Hadronic calorimeter . . . 25
2.2.2.3 Forward calorimeter . . . 25
2.2.3 Muon spectrometer . . . 25
2.2.4 Trigger system . . . 25
2.3 Reconstruction techniques in the ATLAS experiment . . . 26
2.3.1 Tracks and vertices . . . 26
2.3.2 Photons and electrons . . . 27
2.3.3 Muons . . . 27
2.3.4 Jets . . . 27
2.3.5 Missing Transverse Energy . . . 27

3 Simulation and event generators **28**
3.1 Motivation and overview . . . 28
3.2 Structure of an event . . . 28
3.2.1 Hard process . . . 28
3.2.2 Parton shower . . . 30
3.2.3 Conciliating fixed order and parton shower simulations . . . 30
3.2.3.1 Matching . . . 31
3.2.3.2 Merging . . . 31
3.2.4 Hadronization . . . 31
3.2.4.1 The Lund String Model . . . 31
3.2.4.2 The Cluster Model . . . 32
3.2.5 Particle decays . . . 32
3.2.6 Multi-Parton Interactions . . . 32
3.3 Review of Monte-Carlo generation programs . . . 32
3.3.1 Madgraph . . . 32
3.3.2 POWHEG . . . 32
3.3.3 PYTHIA . . . 33
3.3.4 HERWIG 7 . . . 33
3.3.5 Sherpa . . . 33
3.4 Detector response simulation . . . 33
3.5 ATLAS software . . . 34

4 Continuous calibration of *b*-Tagging **35**
4.1 Introduction . . . 35
4.1.1 Motivation . . . 35
4.1.2 Performance estimators . . . 35
4.1.2.1 Label . . . 35
4.1.2.2 *b*-Tagging discriminant . . . 35
4.1.2.3 Working Points . . . 36
4.1.2.4 Scale Factors . . . 36
4.1.3 Key ingredients . . . 36
4.2 *b*-Tagging algorithms . . . 36
4.2.1 Impact parameter based algorithms . . . 36

4.2.2 Secondary vertex finding algorithms . . . 37
4.2.3 Decay chain reconstruction algorithm with JetFitter . . . 38
4.2.4 Multivariate algorithms . . . 38
4.2.5 Algorithms performances . . . 38
4.3 Integrated b-Tagging calibrations . . . 38
4.3.1 b-jets calibration . . . 38
4.3.2 c-jets calibration . . . 39
4.3.3 Light-jets calibration . . . 40
4.4 Continuous calibration of b-Tagging . . . 41
4.4.1 Methods . . . 41
4.4.1.1 Derivation from cumulative calibrations . . . 42
4.4.1.2 Likelihood calibration . . . 43
4.4.2 Systematic uncertainties . . . 44
4.4.3 Performances . . . 44
4.4.3.1 light-jets calibration . . . 44
4.4.3.2 charm-jets calibration . . . 44
4.4.3.3 b-jets calibration . . . 45
4.4.4 Use in physics analyses . . . 45
4.4.5 Prospects . . . 45

5 Search for the SM VH(bb) process **50**
5.1 Introduction . . . 50
5.1.1 Context . . . 50
5.1.2 Signal, background processes and signatures . . . 50
5.1.2.1 Signal processes . . . 51
5.1.2.2 Background processes . . . 52
5.1.3 Overview of the analysis strategy . . . 54
5.2 Datasets . . . 54
5.2.1 Experimental data . . . 54
5.2.2 Monte-Carlo samples . . . 55
5.3 Objects selection . . . 55
5.3.1 Trigger . . . 55
5.3.2 Leptons . . . 56
5.3.3 Jets . . . 57
5.3.4 b-Tagging . . . 58
5.3.5 Missing Transverse Energy . . . 58
5.3.6 Missing Transverse Momentum . . . 58
5.3.7 Corrections to Monte-Carlo predictions . . . 58
5.3.7.1 Pile-up reweighting . . . 58
5.3.7.2 Trigger efficiency Scale Factors . . . 59
5.3.7.3 b-Tagging Scale Factors . . . 59
5.4 Event selection . . . 59
5.4.1 Selection . . . 59
5.4.2 Categories . . . 60
5.4.2.1 Signal Regions . . . 60
5.4.2.2 Control Regions . . . 60
5.4.3 Signal acceptance . . . 62
5.4.4 Contamination from other Higgs processes . . . 63
5.4.4.1 Higgs production modes . . . 63
5.4.4.2 Higgs decay channels . . . 64
5.4.5 Multijet estimate . . . 64

5.4.5.1 0-lepton channel . . . 64
5.4.5.2 1-lepton channel . . . 65
5.4.5.3 2-leptons channel . . . 69
5.5 Multivariate method . . . 69
5.5.1 Boosted Decisions Trees . . . 69
5.5.2 Parameters . . . 71
5.5.3 Training strategy . . . 72
5.5.3.1 k-Folding . . . 72
5.5.3.2 Truth tagging . . . 72
5.5.3.3 The truth/direct tagging discrepancy . . . 73
5.5.4 Training variables . . . 73
5.5.4.1 Testing new variables in the 0-lepton channel . . . 73
5.5.4.2 Transformation of variables . . . 76
5.5.4.3 Process normalization scaling . . . 82
5.5.4.4 BDT enhancement from Continuous b-Tagging . . . 82
5.6 Systematic uncertainties . . . 83
5.6.1 Theory uncertainties . . . 83
5.6.2 Experimental systematics . . . 83
5.6.3 Modelling systematics . . . 83
5.7 Statistical treatment . . . 84
5.7.1 Statistical rebinning . . . 84
5.7.2 Pruning . . . 85
5.7.3 Smoothing . . . 85
5.7.4 Fit Model . . . 85
5.7.4.1 Systematics-free likelihood function . . . 85
5.7.4.2 Systematics Nuisance Parameters . . . 86
5.7.4.3 Monte-Carlo statistics Nuisance Parameters . . . 86
5.7.5 Test statistic . . . 87
5.8 Results . . . 88
5.8.1 Run 2 . . . 88
5.8.1.1 Diboson search . . . 88
5.8.1.2 VH(bb) cut-based analysis . . . 88
5.8.1.3 VH(bb) multivariate analysis . . . 89
5.8.2 Comparison with Run 1 . . . 93
5.8.3 Run 1 - Run 2 combination . . . 93
5.8.4 Comparison to CMS . . . 97
5.8.5 Observation of the $H \to b\bar{b}$ decay in ATLAS . . . 97
5.8.6 Observation of the $H \to b\bar{b}$ decay in CMS . . . 98

6 Additional studies on the SM VH(bb) analysis **99**
6.1 Studies dedicated to better understanding the published result . . . 99
6.1.1 The Z+Jets normalization issue . . . 99
6.1.2 Parton Shower systematic on signal samples . . . 105
6.2 Optimisation of Monte-Carlo filters . . . 106
6.2.1 Z+jets . . . 106
6.2.2 ttbar . . . 108
6.3 Studies of the ttbar background . . . 110
6.3.1 Acceptance for ttbar events . . . 110
6.3.1.1 Fully hadronic ttbar decays . . . 110
6.3.1.2 Semi-leptonic ttbar decays . . . 110
6.3.1.3 Dileptonic ttbar decays . . . 112

6.3.2 $t\bar{t}$ decays in sensitive multivariate discriminant bins . . . 113
6.3.3 Reconstructed τ-jets in $t\bar{t}$ events . . . 113
6.3.4 Flavour composition . . . 114
6.3.5 Discrimination of $t\bar{t}$ events using track-jets and MET soft term . . . 114
6.4 Colourflow . . . 116
6.4.1 Colourflow observables . . . 117
6.4.2 TruthAnalysis with a single-event . . . 117
6.4.2.1 Analysis setup . . . 117
6.4.2.2 Results . . . 118
6.4.3 TruthAnalysis with ATLAS MC samples . . . 119
6.4.4 Analysis at reconstructed level . . . 119

7 Search for top quark decays $t \to qH$ with $H \to \gamma\gamma$ 125
7.1 Introduction . . . 125
7.1.1 Motivation . . . 125
7.1.2 Signal, background processes and signatures . . . 126
7.1.2.1 Signal processes . . . 127
7.1.2.2 Background processes . . . 128
7.1.2.3 Expected signatures . . . 129
7.1.2.4 Overview of the analysis strategy . . . 129
7.1.3 The Run 2 publication . . . 130
7.1.3.1 Selection and categories . . . 130
7.1.3.2 Fitting procedure . . . 130
7.1.3.3 Results . . . 132
7.2 Datasets . . . 132
7.2.1 Experimental data . . . 132
7.2.2 Monte-Carlo samples . . . 133
7.3 Objects selection . . . 134
7.3.1 Trigger . . . 134
7.3.2 Photons . . . 134
7.3.3 Jets . . . 134
7.3.4 Electrons . . . 135
7.3.5 Muons . . . 135
7.3.6 Overlap removal . . . 135
7.3.7 Missing Transverse Energy . . . 135
7.4 Events selection . . . 135
7.4.1 Selection . . . 135
7.4.2 Categories . . . 137
7.4.2.1 Kinematic categories . . . 137
7.4.2.2 c-Tagging categorisation . . . 137
7.4.3 Improvement of the selection in the hadronic channel . . . 138
7.5 Study of events kinematic . . . 140
7.6 Multivariate method . . . 144
7.7 Statistical analysis . . . 153
7.7.1 Fit model . . . 153
7.8 Prospects . . . 155
7.9 Update of the background shape parameterisation . . . 157
7.9.1 Preliminary estimate using the procedure from the 2017 publication . . . 157
7.9.2 Update of the procedure for spurious signal estimates . . . 157
7.9.2.1 Design of Control Regions . . . 157
7.9.2.2 Statistical issues in toys . . . 160

7.10 Conclusion 160

Conclusion **163**

Remerciements **164**

Bibliography **165**

Personal contributions

A list of my personal contributions documented in this manuscript is presented below.

1. Pseudo-continuous calibration of b-tagging, Chapter 4
    - Creation of a simpler software with respect to the existing codes.
    - Production and updates of calibration files with the pseudo-continuous calibration of b-tagging.
    - Tests for the use of pseudo-continuous b-tagging in the CxAODReader framework.
2. Search for the SM VH(bb) process, Chapter 5
    - Pre-fit estimates of the signal and background (non-resonant, SM Higgs processes) yields and distributions.
    - Studies of the multijet background contamination in the 0-lepton channel using Monte-Carlo predictions, validation of the shape used in the fit to data.
    - Training and tests for the use of multivariate discriminants in the 0-lepton channel : nominal training, use of truth tagging and investigations on the observed bias, tests of new training variables and transformations, optimization of the set of training variables, test of normalisation scaling.
3. Additional studies on the SM VH(bb) analysis, Chapter 6
    - The Z+Jets normalization issue.
    - Investigations on the Parton Shower systematic on signal samples.
    - Optimization of Monte-Carlo filters for the Z+jets and ttbar sample processes.
    - Studies of the topology of ttbar background.
    - Feasibility studies for the use of colourflow in the VH(bb) search.
4. Search for top quark decays t $\rightarrow q$H with H $\rightarrow \gamma\gamma$, , Chapter 7
    - Participation to the update of the event selection and design of the new categorization using c-tagging.
    - Optimization of the use of the combinatoric approach in the hadronic channel of the analysis, based on kinematic studies performed on signal samples.
    - Training and optimization of a multivariate discriminant.
    - Prospect estimates of the sensitivity with the updated selection and categorization.
    - Preliminary estimates of the spurious signal yields.
    - First look at the use of Control Regions using reverted photon isolation and identification requirements.

Introduction

In the field of High Energy Physics, the Standard Model of particles is a framework based on gauge theories which encompass the description of processes involving elementary particles. A part of the Standard Model, the electroweak gauge theory, was a priori in contradiction with the experimental fact that weak bosons are very massive. The issue can be solved assuming the existence of a scalar field whose none-zero vacuum expectation value generates the spontaneous symmetry breaking of the SU(2)×U(1) gauge symmetry associated to the electroweak interaction. This is denoted as the Higgs mechanism, and a search for the corresponding scalar boson was operated in a variety of experiments.

The search for the Higgs boson was performed over the last decades, in particular at the LEP and Fermilab Tevatron colliders. Since 2008, the Large Hadron Collider (LHC) accelerates beams of protons which produce collisions recorded by four experiments. In 2012, the ATLAS and CMS collaborations reported the observation of a new particle at a mass about 125 GeV with properties compatible with that predicted by the Standard Model for the Higgs boson. After that discovery, a major aim of the ATLAS and CMS experiments is the search for deviations in experimental measurements with respect to Standard Model predictions, which could reveal the existence of additional particles and rare processes. This requires to study the properties of the Higgs boson with a very demanding level of precision, and constituted an important effort of the collaborations at the run 1 of the LHC between 2008 and 2013.

My implication in the ATLAS collaboration started at the beginning of the run 2 of the LHC, after a period intended to upgrade the accelerator and detectors. The energy of accelerated protons was increased from a center-of-mass energy of 8 TeV to 13 TeV, which allows to improve our understanding of the Standard Model and further constrain hypotheses predicting the existence of additional particles and rare processes. At the same time, some detection devices were upgraded in order to improve the sensitivity of searches for additional production and decay modes of the Higgs boson.

The first part of this manuscript deals with the theoretical background leading to the Standard Model construction. A special attention is given to relate symmetries of the Standard Model to particles interactions. Also, a short description of the limits of the model is reported.

Second, the experimental apparatus is presented, from the accelerator complex to detectors. Especially, the ATLAS detector is described with more details. A brief overview of its subsystems and data taking process is given.

The Standard Model predictions are tested using artificial data, denoted as Monte-Carlo simulations, which are compared to events recorded in detectors. The third section reviews the key aspects of Monte-Carlo simulation in high energy physics, and the main features of simulation programs used by the ATLAS collaboration.

The upgrade of the ATLAS detector is expected to allow improvements of the identification performances of particles produced in collisions of protons. Especially, the decay of b-quarks provides signatures, analysed in the context of b-jet tagging, that are exploited by several searches performed at the run 2 of the LHC. A brief description of the b-Tagging techniques is presented in this manuscript. The pseudo-continuous calibration of b-Tagging discriminants is an extension of these techniques. The update of pseudo-continuous b-Tagging is described.

The search for the Higgs boson decay to beauty quarks is primordial given the dominant branching ratio for this process. In its production mode in association with a vector boson, clear signatures allow to separate the process $\mathrm{pp} \to \mathrm{VH}(\mathrm{H} \to \mathrm{b\bar{b}})$ from most other processes

predicted by the Standard Model occuring at the LHC. At the run 1 of the LHC, the ATLAS and CMS experiments reported a statistical significance of respectively 1.4 and 2.1 standard deviations for this search. This search was updated, and many ideas intending to optimise the analysis have been explored.

Last, a search for the decay of a top quark into an up-type (q = u, c) quark and a Higgs boson decaying into two photons is presented. The same search was performed previously and the result was published recently. Ideas for improvements are tested and implemented in the perspective of exploiting the full set of data collected at the run 2 of the LHC. Prospect results are presented in this manuscript.

Chapter 1

Theory of the Standard Model of elementary particles

1.1 Algebraic framework : group theory

This section describes some aspects of the mathematical background of particles physics. Group theory provides a framework to particles individual properties and interactions rules. Little mathematical parameterization is required in order to achieve powerful physics predictions of a theory. Some of these underlying mathematical aspects are briefly explained in this section.

1.1.1 Definitions

A group G is defined as a set of elements g_i related by operation rules denoted $\star$ and meeting the following requirements :

- closure : $\forall g_i, g_j \in G, g_i \star g_j \in G$
- associativity : $\forall g_i, g_j, g_k \in G, (g_i \star g_j) \star g_k = g_i \star (g_j \star g_k)$
- identity : $\exists e \in G$ so that $\forall g_i \in G, e \star g_i = g_i \star e = g_i$
- inverse : $\forall g_i \in G, \exists g_j \in G$ so that $g_i \star g_j = g_j \star g_i = e$

The elements of a group should be seen as abstract transformation operations. The order of a group defines the number of elements it contains, which can be either finite or infinite. Also, the elements of a group can form a discrete or continuous set of transformations. Last, the action of a group on objects describes the particular way elements of the group perform transformations on these objects[1].

1.1.2 Group representation

Any group can be illustrated with respect to different group representations, which provide equivalent mathematical formulations of the way elements of the group relate to each other. For instance, the set of integers modulo 3 related by addition operation ($\star = +$) with the association $\{e = 0, g_1 = 1, g_2 = 2\}$ is a representation of the cyclic group of order 3. An equivalent representation of this group consists of attaching complex numbers to the elements of the group $\{e = 1, g_1 = e^{i\frac{2\pi}{3}}, g_2 = e^{i\frac{4\pi}{3}}\}$ related by a multiplicative law ($\star = \times$). In the following, some fixed representation of a group will be denoted $D(G)$, and $D(g_i)$ for a specific element of the group.

A remarkable feature of group theory is that, although elements of any group should be seen as abstract objects, they can always be represented in the form of matrices with normal matrix

[1] the objects which undergo the action of the group are abstract as well

multiplication to relate the elements of the group in their matrix form. Yet, many equivalent representations of the group can be formulated with matrices of different dimensions. Especially, the representation of a group of order n in terms of $n \times n$ matrices denotes its regular representation[2]. If a set of matrices with dimension $m < n$ can (not) be found, the dimension is said to be reducible (irreducible). Also, any matrix representation of dimension n can be associated the space spanned by n orthonormal unit vectors as objects which undergo the action of the group. The one-to-one association of elements of the group to base vectors $g_i \rightarrow |g_i\rangle$ with operation law $D(g_i)|g_j\rangle = |g_i \star g_j\rangle$ allows writing explicitly the matrices associated to elements of the group :

$$[D(g_k)]_{ij} = \langle g_i|D(g_k)|g_j\rangle \tag{1.1}$$

1.1.3 Lie groups

A Lie group characterizes a finite group which has its elements parameterized by one or several continuous parameters α_i, meaning that transformations induced by the elements of the group $g(\alpha_i)$ are infinitely differentiable. This allows to write the Taylor expansion of an infinitesimal transformation in some representation as :

$$D(g(\delta\alpha_i)) = 1 + \delta\alpha_i \frac{\partial D(g(\alpha_i))}{\partial \alpha_i}\Big|_{\alpha_i=0} + ... \tag{1.2}$$

From the previous equation, the generators of the group are defined as :

$$X_i = -i\frac{\partial D(g(\alpha_i))}{\partial \alpha_i}\Big|_{\alpha_i=0} \tag{1.3}$$

Which allow to write the previous infinitesimal transformation as :

$$D(g(\delta\alpha_i)) = 1 + i\delta\alpha_i X_i + ... \tag{1.4}$$

Following the previous result, associativity allows to write any finite transformation of the group as an expansion of an infinite number of infinitesimal transformations :

$$\lim_{N\rightarrow\infty}(1 + i\frac{\alpha_i}{N}X_i)^N = e^{i\alpha_i X_i} \tag{1.5}$$

In any representation of a Lie group, a subset of M generators mutually commuting can be written in the form of diagonal matrices. These generators are denoted as Cartan generators of the group, and can be associated eigenvectors and eigenvalues. Furthermore, linear combinations of the non-Cartan generators can be written so that their action is to interchange eigenvectors of the Cartan generators amongst themselves. A common example to this is the SU(2) group with Pauli matrices $\tau^{1,2,3}$ as generators of the group, in the view of quantum mechanical spin for fermions of spin $1/2$. The diagonal Cartan generator τ^3 can be associated two eigenvectors with eigenvalues $\pm 1/2$ which interchange one into each other when multiplied by linear combinations of τ^1 and τ^2. Similar principles will apply to the Standard Model of particles described in section 1.3.

1.2 Fundamental concepts

This section presents a short review of the physical background connecting particle physics and groups theory as previously described. It relies on concepts first developed in the context of classical physics, and reformulated in the perspective of quantum field theories.

[2] but more generally, the dimension of the representation can be very different from the order of the group

1.2.1 Hamiltonian mechanics

The mathematical support of particles physics relies on the description of fields denoted φ and defined at any point x of space and time by scalar, vectors or tensors. Fields preclude the existence of a "$true$" vacuum, and their excitations in space-time embody point-like particles. The evolution of systems of particles can be derived from the Lagrangian quantity $\mathcal{L}(\varphi, \partial\varphi, x)$ which is defined at any point of space and time and allows to compute the action :

$$S = \int \mathcal{L}(\varphi, \partial\varphi, x)\, \mathrm{d}^3x \tag{1.6}$$

Hamilton's principle [1] states that the evolution of particles is enforced at stationary points of the action functional. The Euler-Lagrange equation allows to derive the corresponding equations of motion for the particle system.

1.2.2 Symmetries

Local (global) symmetries assert the invariability of observables under transformations of the physics fields that (do not) depend on space-time coordinates. Symmetries play a fundamental role in the construction of the Standard Model of particles. Basis concepts are presented in the following, while the interplay with group theory and particles physics will be developed in Section 1.3.2.

1.2.2.1 Gauge transformations

The invariance of physics laws with respect to local symmetries enforces the local gauge principle. This criterion ensures the consistency of quantum mechanics under arbitrary phase transformations in order to guarantee the existence of a conserved probability current. Local transformations of physics fields require introducing new degrees of freedoms via gauge transformations of their Lagrangian. For instance, the Lagrangian for a free propagating particle with spin 1/2, mass m and electric charge q_e reads :

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \tag{1.7}$$

where ψ denotes the Dirac spinor of the particle's field. The gauge transformation under local U(1) complex phase transformations is defined by :

$$\psi(x) \longrightarrow e^{iq_e\alpha(x)}\psi(x) \tag{1.8}$$

where $\alpha(x)$ can be any arbitrary real function of x. The local gauge principle requires to add a massless gauge field term A_μ that transforms as :

$$A_\mu \longrightarrow A_\mu + \frac{1}{q_e}\partial_\mu\alpha(x) \tag{1.9}$$

In this process, the Quantum Electrodynamic (QED) Lagrangian becomes :

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi + e\bar{\psi}\gamma^\mu\psi A_\mu \tag{1.10}$$

The previous equation factorizes using the covariant derivative, written explicitly here in the context of QED :

$$D_\mu = \partial_\mu - ieA_\mu \tag{1.11}$$

The last term in equation 1.10 translates the interaction of photons with Dirac particles, namely charged fermions, which cannot exist as free particles and observe the U(1) local phase

symmetry at the same time. Such terms play a fundamental role in Quantum Field Theory (QFT), as they implement elementary particles interactions in the form of products of the corresponding fields in the Lagrangian. In the previous example, QED can be said to be a field theory of the gauge group U(1). The field A_μ which was introduced is able to propagate as well, and the lagrangian requires an additional kinetic term :

$$\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.12}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength tensor.

Last, this formalism allows deriving Feynman rules, which provide a pictorial framework for computing the probability for such interactions to take place.

1.2.2.2 Noether theorem

Noether's theorem [2] states that any variational transformation of the Lagrangian of a physical system that leaves the action invariant can be associated to a conservation law. For instance, the differentiable symmetry of the action under translations in space and time enforces the conservation of momentum and energy respectively. This theorem extends to QFT and relates for instance the invariance of QED under U(1) local phase transformations of Dirac fields with the conservation of electric charge.

The application of Hamilton's principle to field theories which observe local symmetries demonstrates that the action functional can be invariant in non-measurable local transformations of the fields, and yet involve the conservation of physics observable as stated by Noether's theorem. In analogy to the example of QED, this logic will be extended to other field theories describing the Standard Model of elementary particles with respect to appropriate symmetries in the framework of group theory.

1.3 The Standard Model

This section describes the particle content of the universe and their interactions in the current knowledge of the Standard Model. First, an overview of particles properties is proposed, while the inner relation with the respective gauge theories is presented in a second part.

1.3.1 Particles content

The Standard Model encompasses gauge bosons and fermions with integer and half-integer spin values respectively. Three generations of fermions which behave equally with respect to the SM gauge group transformations are needed to describe the observations. At first sight, each generation includes a quark doublet, one charged lepton and an electrically null charge neutrino. Generations prove to be consistently ordered in mass of their particle content[3]. These generations present further similarities which arise from the gauge theory framework of the Standard Model, as described in the next section. These properties relate to the coupling of fermion particles to bosons, which are the mediators of fundamental interactions. Figure 1.1 presents fermions and bosons described in the Standard Model, with a schematic encapsulation which depicts their possible interactions.

1.3.2 Gauge theories of the Standard Model

This section describes the inner properties of gauge theories which enable specific interactions of fermions with bosons. The Standard Model Lagrangian is invariant with respect to local

[3] It is still uncertain if this logic applies to neutrinos.

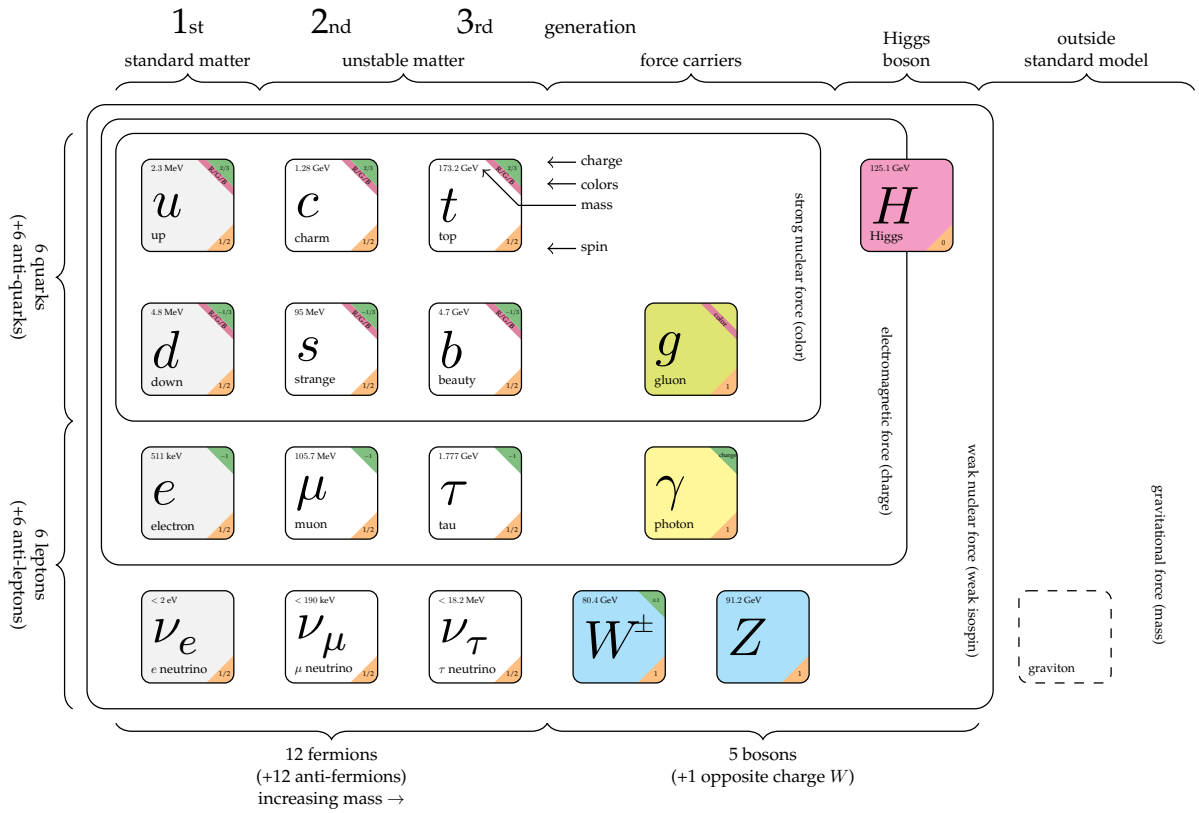


FIGURE 1.1: The particle content of the Standard Model.

symmetries of the group $G_{SM} = SU(3)_C \times SU(2)_I \times U(1)_Y$, in which QCD colour charge C, weak isospin I and hypercharge Y are conserved respectively. In the Lagrangian expression, fermion fields are written in the form of Dirac spinors ψ which multiply gauge field terms which stand for bosons.

1.3.2.1 Gauge theories

Gauge theories collectively refer to physics field theories whose Lagrangian is invariant under specific Lie groups local symmetries. A major achievement of group theory consists of the description of strong (QCD) and and electro-weak (EW) fundamental interactions as field theories of Lie groups. In that framework, bosons embody the action of the group generators, and interact solely with fields charged under the group. The action of generators can either preserve (Cartan generators) or change the charge of the fields (linear combinations of the non-Cartan generators), which corresponds to maintaining or transforming particles involved in the interaction respectively. Particles unchanged in the interaction with a gauge boson embody the eigenvectors of the Cartan generators of the group, while corresponding eigenvalues conform to the physically measurable charges. As an example, photons and Z-bosons allow transferring momentum but do not change the particle charge in the interaction. Conversely, non-Cartan generators shift the charge of the fields in the corresponding group in addition to momentum transfers, which transcribes the action of gluons and W-bosons on quarks for instance. In this picture, the number of possible charge values particles can be associated correspond to the dimension of the group in the appropriate representation.

1.3.2.2 The Electroweak interaction

The Electroweak model unifies the electromagnetic and weak interactions as theories of the $SU(2)_I \times U(1)_Y$ group with gauge fields $W_\mu^{1,2,3}$ and B_μ [3, 4, 5]. The related covariant derivative highlights the association to their respective group generators $\tau^{1,2,3}$ (Pauli matrices) and Y[4] acting on doublets of Dirac fields :

$$D_\mu = \partial_\mu - i[gW_\mu^a \frac{\tau^a}{2} + g' B_\mu \frac{Y}{2}] \tag{1.13}$$

Here, g and g' are coupling constants which will balance the probability for the different interaction processes. The corresponding Lagrangian splits into charged and neutral components with respect to the (non-) Cartan generators of the group and allows writing force carrying bosons as a mixing of the group gauge fields :

$$W_\mu^+ = \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2) \tag{1.14}$$

$$W_\mu^- = \frac{1}{\sqrt{2}}(W_\mu^1 + iW_\mu^2) \tag{1.15}$$

$$Z_\mu = cos(\theta_W)W_\mu^3 - sin(\theta_W)B_\mu \tag{1.16}$$

$$A_\mu = sin(\theta_W)W_\mu^3 + cos(\theta_W)B_\mu \tag{1.17}$$

Anticipating on the next sections, the weak mixing angle $\theta_W = tan^{-1}(g'/g)$ which rotates the W_μ^3 and B_μ fields to generate Z and γ bosons, arise from the spontaneous symmetry breaking in the Higgs mechanism which grants their mass to the $W^\pm$ and Z bosons :

$$m_Z = \frac{v}{2}\sqrt{g^2 + g'^2} \tag{1.18}$$

$$m_W = \frac{gv}{2} \tag{1.19}$$

Experimental results proved parity violation of the weak interaction. Chiral projection operators P_L and P_R allow writing left helicity doublets χ_L and right helicity charged lepton singlets ψ_R. The Gell-Mann–Nishijima formula [6, 7] encodes the relationship of the $SU(2)_I \times U(1)_Y$ group charges (weak isospin[5] and hypercharge), with the electrical charged of fermions and force-carrying mediators.

$$Q = I_3 + \frac{1}{2}Y \tag{1.20}$$

Quarks are not enforced to stand within their original doublet in weak interaction processes, as quark interaction eigenstates do not write as a one-to-one relationship with respect to quarks mass eigenstates. This generates a flavour mixing amongst the three generations of quarks in weak interaction processes. The mixing probabilities of quarks q_i and q_j is governed by the Cabibbo-Kobayashi-Maskawa (CKM) unitary matrix V_{ij}^{CKM} [8, 9]. The covariant derivative

[4] Y denotes both the diagonal generator of U(1) and its associated eigenvalues.

[5] Weak isospin projection is most commonly denoted as the eigenvalue I_3 of the Cartan generator related to W^3

defined in equation 1.13 allows to write the Electroweak component of the Standard Model Lagrangian as [6] :

$$\mathcal{L}_{EW} = i\overline{\chi}^i_L \gamma^\mu D_\mu \chi^i_L + i\overline{\psi}^i_R \gamma^\mu D_\mu \psi^i_R - \frac{1}{4} W^a_{\mu\nu} W^{\mu\nu}_a - \frac{1}{4} B^a_{\mu\nu} B^{\mu\nu}_a \tag{1.21}$$

The symmetry breaking in the Higgs mechanism allows to write the neutral and charged components of the electroweak interaction of fermions and gauge bosons, denoted $\mathcal{L}^N_{EW}$ and $\mathcal{L}^C_{EW}$ respectively. The correspondence of electroweak bosons to ladder operators of $SU(2)_I \times U(1)_Y$ which can either preserve (A_μ, Z_μ) or lower and rise (W^+_μ, W^-_μ) the weak isospin charge of fermions electroweak doublets is now explicit :

$$\begin{aligned} \mathcal{L}^N_{EW} = {} & g \sin(\theta_W) \sum_f q_f \overline{f} \gamma_\mu f A^\mu + \\ & \frac{g}{\cos(\theta_W)} \Big[\sum_f I^3_f \overline{f} \gamma_\mu P_L f - \sin^2(\theta_W) \sum_f q_f \overline{f} \gamma_\mu f \Big] Z^\mu \end{aligned} \tag{1.22}$$

$$\mathcal{L}^C_{EW} = -\frac{g}{\sqrt{2}} \Big[\overline{u_i} \gamma^\mu P_L V^{CKM}_{ij} d_j + \overline{\nu}_i \gamma^\mu P_L e_i \Big] W^+_\mu + h.c. \tag{1.23}$$

In the neutral component $\mathcal{L}^N_{EW}$, fermion fields denoted f with non-zero electric charge q_f are able to interact with the photon field A^μ. The coupling to Z bosons also relates to the weak isospin charge I_3.

On the other hand, the charged Lagrangian part $\mathcal{L}^C_{EW}$ has ν_i and e_i denoting leptons fields with flavour i, which are the two possible eigenstates of $SU(2)_I \times U(1)_Y$ with isospin charge +1/2 and -1/2 respectively. The same applies to quark electroweak doublets u_i and d_j, with additional mixing probabilities encoded in the CKM matrix V^{CKM}_{ij}.

1.3.2.3 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the gauge theory of the strong interaction mediated by gluon fields G^a_μ. Is it described in the $SU(3)_C$ formalism with three colour charges and eight generators which are the Gell-Mann matrices λ^a in the corresponding covariant derivative :

$$D_\mu = \partial_\mu - i g_S G^a_\mu \frac{\lambda^a}{2} \tag{1.24}$$

The strong running coupling constant $g_S(\mu^2)$ is observed to decrease with increasing energy scale, which causes the strength of the QCD force to drop at high energies carried by gluons. Quarks are the only fermions charged with respect to QCD, and their Dirac spinor $\psi = (\psi^a, \psi^b, \psi^c)$ transforms as colour triplets of $SU(3)_C$. The non-abelian gauge structure of the strong force allows gluons self-interactions. The QCD Lagrangian reads [7]

$$\mathcal{L}_{QCD} = i\overline{\psi} \gamma^\mu D_\mu \psi - m\overline{\psi}\psi - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \tag{1.25}$$

[6] The field strength tensors for the non-abelian sectors are defined similarly to the one of QED, but with an additional term, e.g. for $SU(2)_I, W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g f^{abc} W_{b,\mu} W_{c,\nu}$, which induces triple and quartic interactions between gauge bosons.

[7] Another term exists, which does not enter the classical equations of motion but has real physics effects and induces CP violation in the strong sector : $-\frac{\theta_{QCD}}{16\pi^2} G^a_{\mu\nu} \tilde{G}^{\mu\nu}_a$ with $\tilde{G}^{\mu\nu}_a = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} G^a_{\rho\sigma}$, and $\epsilon^{\mu\nu\rho\sigma}$ is the Levi-Civita tensor.

1.3.3 Spontaneous symmetry breaking

Local symmetries of the electroweak model require fermions and vector bosons to be massless particles; for the fermion a mass term $-m\overline{\psi}\psi$ can be written as $-m(\overline{\chi}_L\psi_R + \overline{\psi}_R\chi_L)$, and given that χ_L and ψ_R transform differently under the gauge transformations, it is not invariant; for a vector boson, a mass term $m^2 A^\mu A_\mu$ is not invariant under 1.9. This prediction is in contradiction with experimental results which proved that the $W^\pm$ and Z bosons are very massive, while the photon is indeed observed as a massless particle. The solution arises from the spontaneous EW symmetry breaking, thanks to an additional field ϕ, induced by the Brout-Englert-Higgs (BEH) mechanism[8] [10, 11]. The simplest possibility for the Higgs field consists in a weak isospin doublet of complex fields with unit hypercharge value. Its Lagrangian involves the Higgs potential, which reads :

$$\mathcal{L}_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) \quad \text{with} \quad V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{1.26}$$

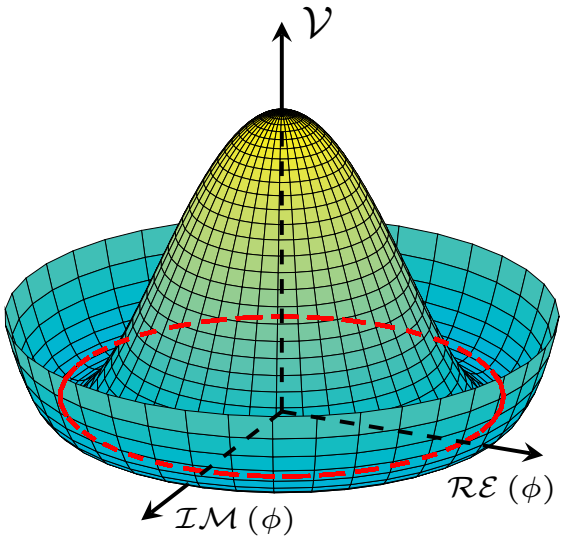


FIGURE 1.2: Illustration of the Higgs potential in the SM. The spontaneous symmetry breaking arise from the vacuum value at the minimum of the potential, which occurs for $\phi \neq 0$ along the red line.

The form of this potential is illustrated in Fig. 1.2. The Vacuum Expectation Value of the Higgs field is $v = \sqrt{-\mu^2/\lambda}$. Gauge invariance with respect to $SU(2)_I \times U(1)_Y$ local phase transformations allows to develop the Higgs field from its ground state, picked along the real component :

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix} \quad \longrightarrow \quad \phi(x) = \frac{e^{i\tau_a\theta_a(x)/v}}{\sqrt{2}}\begin{pmatrix}0\\v+h(x)\end{pmatrix} \tag{1.27}$$

In the latter equation, the Higgs field is embodied in $h(x)$ with mass value $m_H = \sqrt{-2\mu^2}$. Three pseudo-scalar Goldstone bosons denoted θ_a arise from the broken generators τ_a. The exponential is eliminated in the Lagrangian because of local phase invariance. The same mechanism allows granting fermions a mass term proportional to the vacuum expectation of the Higgs field. This is often mentioned as the Yukawa interaction which describes the coupling

[8] Higgs mechanism for short.

between the Higgs field and massless Dirac fields. The Higgs mechanism generates coupling constants G proportional to fermions mass, and bosons squared mass. It reads with respect to the lepton (quark) left helicity isospin doublets L_L^i (Q_L^i) and right helicity isospin singlets l_R^j (up-type u_R^j, down-type quarks d_R^j).

$$\mathcal{L}_{Yukawa} = -G_l^{ij}\overline{L}_L^i\phi l_R^j - G_d^{ij}\overline{Q}_L^i\phi d_R^j - G_u^{ij}\overline{Q}_L^i\tilde{\phi}u_R^j + h.c. \tag{1.28}$$

The mass term in $\mathcal{L}_{QCD}$ is now naturally generated in $\mathcal{L}_{Yukawa}$, and the complete Lagrangian of the Standard Model reads :

$$\mathcal{L} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \tag{1.29}$$

1.4 Phenomenology of the Higgs boson

1.4.1 Production modes

Higgs boson production at the LHC originates from gluons and quarks parton interactions in proton collisions. The main processes are illustrated with Feynman diagrams at Leading Order in the following. Their respective production cross-sections depend on the Higgs boson mass and center-of-mass energy of the collision as shown in Fig. 1.4a, and are given in the following for m_H = 125 GeV and $\sqrt{s}$ = 13 TeV [12].

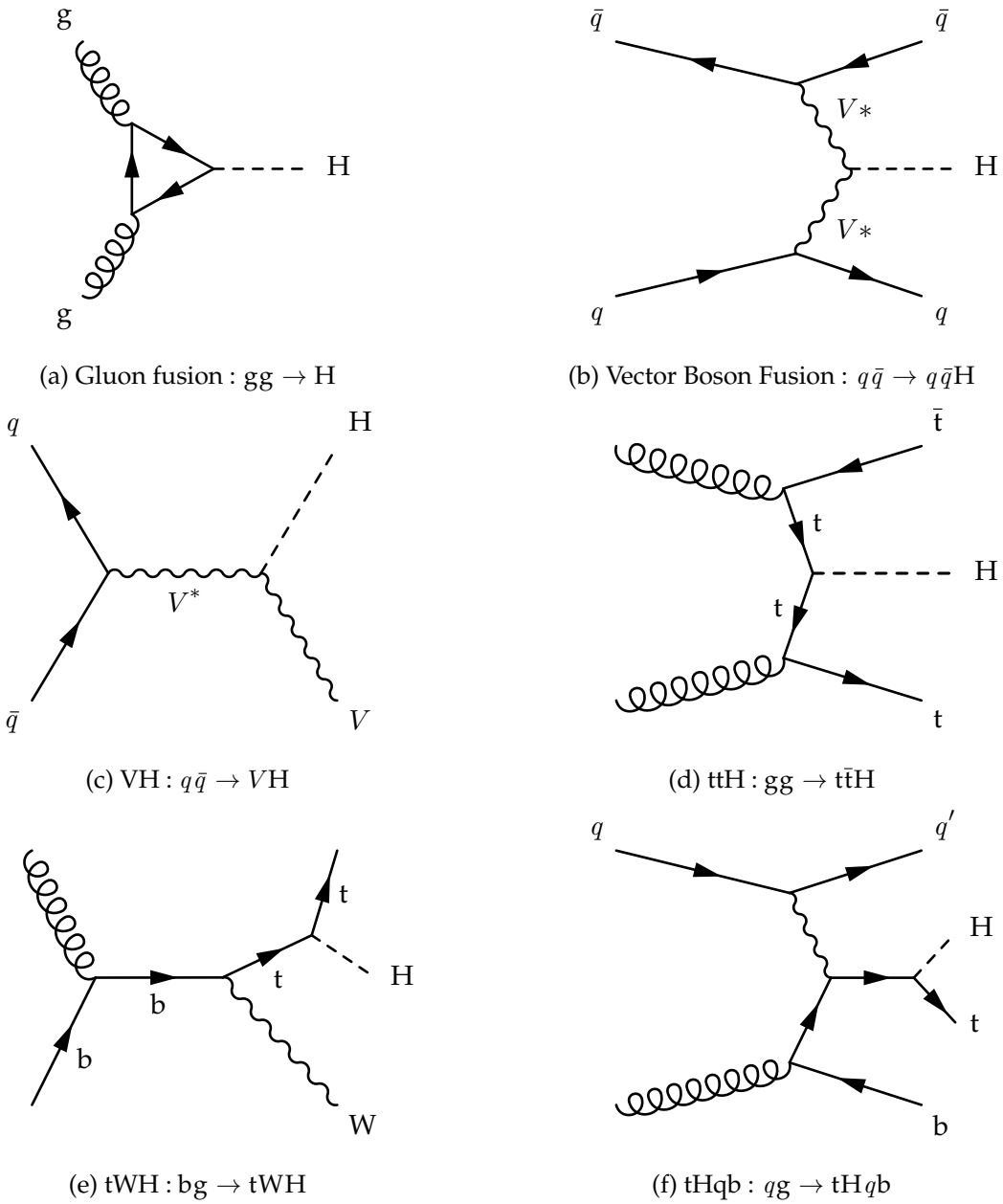
- The gluon fusion induced Higgs boson production mode has the largest predicted cross-section σ_{ggH} = 48.58 pb (N3LO QCD) at the LHC. This process is enabled by the intermediate quark loop out of massless gluons. Especially, it benefits from the large coupling of top quarks to the Higgs field.
- Vector Bosons Fusion gather processes in which two quarks radiate massive weak bosons which fuse into a Higgs boson. The quarks remnants generally produce jets of particles in opposite directions along the beam axis. The cross-section for this process is currently predicted to σ_{VBFH} = 3.782 pb (NNLO QCD + NLO EW).
- Vector Boson associated production consist of the radiation of a Higgs boson from a massive vector boson[9]. The low cross-sections for this process σ_{WH} = 1.373 pb (NNLO QCD + NLO EW) and σ_{ZH} = 0.8839 pb[10] (NNLO QCD + NLO EW) is compensated by clear signatures provided by the vector boson leptonic decay products.
- Top quarks associated production is triggered by the emission of a Higgs boson in $t\bar{t}$ events. This channel is very challenging due to the small cross-section σ_{ttH} = 0.5071 pb (NLO QCD + NLO EW) and profusion of jets and leptons produced in top quarks decays.

1.4.2 Decay channels

- The H $\to f\bar{f}$ decay is dominant because of mass restrictions which strongly suppress the branching ratio for the Higgs to decay into bosons. Especially, the coupling of the Higgs field proportional to fermion mass favors the H $\to b\bar{b}$ decay branching ratio B(H $\to b\bar{b}$) = 58.24% [13]. Second is the H $\to \tau^-\tau^+$ decay with branching fraction 6.272%. The decay probabilities for H $\to c\bar{c}$ and H $\to \mu^-\mu^+$ respectively amount to 2.891% and

[9] this channel is sometimes referred to as Higgs-strahlung.

[10] The photon induced contribution ($\sim$ 1% of σ_{ZH}) is not accounted here.



(a) Gluon fusion : $gg \to H$

(b) Vector Boson Fusion : $q\bar{q} \to q\bar{q}H$

(c) VH : $q\bar{q} \to VH$

(d) ttH : $gg \to t\bar{t}H$

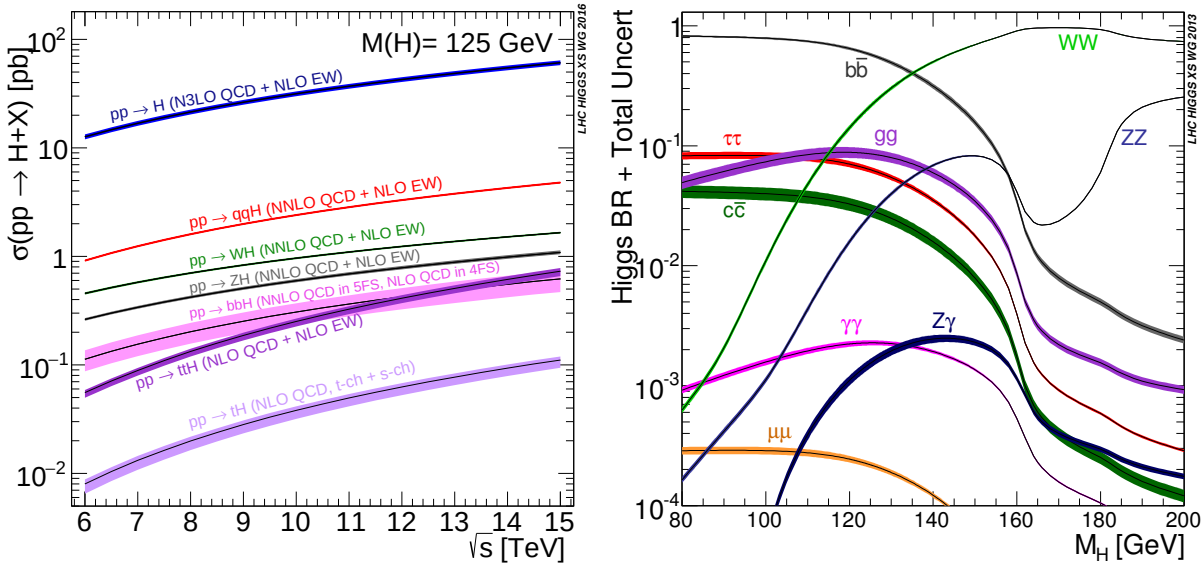
(e) tWH : $bg \to tWH$

(f) tHqb : $qg \to tHqb$

FIGURE 1.3: Feynman diagrams for the Higgs boson production processes. The tWH and tHqb production modes have very small cross-section values, and are generally not presented. They are considered in the FCNC search presented in Chapter 7. In Fig. 1.3e and 1.3f, the Higgs boson leg can be attached either to the top-, b-quark, or W boson line.

0.2176% and will most likely undergo extensive studies in the High-Luminosity LHC data taking period.

- The Higgs boson can decay into massive bosons at tree level. The $H \to WW^*$ and $H \to ZZ^*$ decay channels have respective decay probabilities 21.37% and 2.619%. The decay into gluons and photons is induced by loops, with inclusive branching fractions 8.187% and 0.227% respectively.



(a) (b)

FIGURE 1.4: Figure A shows the evolution of the Standard Model Higgs boson production cross sections as a function of the centre-of-mass-energy for $m_H = 125$ GeV. Figure B illustrates the Standard Model Higgs boson decay branching ratios as a function of its mass.

1.4.3 Motivations for Higgs boson searches

From the perspective of experiments, the Standard Model was incomplete until 2012 and the discovery of a particle with properties compatible with that predicted for the Higgs boson. Further studies have to be performed in order to measure the coupling of particles to the Higgs field. Especially, searches for the Higgs boson decaying into fermions should allow to measure the appropriate Yukawa couplings and test the predictability of the Standard Model. Deviations and inconsistencies with respect to the SM predictions are likely to provide hints for new particles and processes which would allow addressing problems of the Standard Model described in Section 1.5. Especially, the large branching ratio for the $H \to b\bar{b}$ decay mode allows testing Standard Model predictions at the run 2 of the LHC with proton collisions at $\sqrt{s}$ = 13 TeV. Important background sources complicate the analysis which benefits from recent improvements in jets flavour identification techniques especially.

1.5 Problems of the Standard Model

The Standard Model allows for a great diversity of physics predictions with a good accuracy. Yet, in our current knowledge of particles physics, some experimental observations are explained with separate models and cannot be addressed by the SM. Furthermore, all symmetries and parameters of the SM are selected with no other motivation than providing physics predictions in agreement with experimental measurements. This arbitrariness could be the manifestation of a more fundamental theory we do not perceive. These issues are briefly discussed in the following.

1.5.1 Neutrinos

Experimental observations of the oscillation of neutrino flavour [14, 15] were explained in the formalism of neutrino mass eigenstates mixing, from which arise neutrino weak-interaction

eigenstates [16]. Yet, the exact relationship between neutrinos and anti-neutrinos is still controversial. Dirac's model predicts that they are distinct particles, while they are identical in the theory of Majorana neutrinos. The Standard Model predicts that right-helicity neutrinos (left-helicity anti-neutrinos), if they exist, only interact through Yukawa couplings with the Higgs boson. This degeneracy could be lifted by the MAJORANA experiment for instance, looking for neutrinoless double beta decay ($\beta\beta0\nu$) with neutrino self-absorption.

1.5.2 Dark matter and dark energy

Astrophysical experiments enlightened the existence of matter contributing to the gravitational potential in the vicinity of galaxies. This particular form of matter which does not appear to interact electromagnetically is denoted as dark matter [17], and is predicted to constitute about 26% of the universe energy density. Further measures showed the acceleration of universe expansion [18], which could be explained with the existence of dark energy filling the universe with uniform density. While Standard Model neutrinos are massive and do not interact with photons, the current limits on their mass prevent them from being dark matter candidates. Instead, Weakly Interacting Massive Particles, axions [19] and sterile neutrino models [20] provide candidates to the dark matter. On the other hand, dark energy, which could be explained by the modified gravity models, is most commonly admitted as an ad-hoc property of space encoded in the cosmological constant.

1.5.3 Gravity

The formalism of General Relativity allows for an accurate description of classical gravity. Yet, gravity is not unified with the other fundamental interactions described in the Standard Model in the form of quantum theories. Kazula-Klein and supergravity models allow for non-renormalizable quantum theories of gravity, i.e. divergent interaction probability terms. Superstring theories account for renormalizability, but imply a form of supersymmetry, at least at the highest energy scales (e.g. Planck scale), which was not observed yet, but could manifest at the TeV scale.

An additional inconsistency arise from the coupling of gravity to the vacuum energy of the Higgs potential, which contributes to the cosmological constant. The difference between predicted and measured values of the cosmological constant cannot be explained apart from incredibly fine-tuned cancellations allowed by additional constant terms in the scalar potential.

1.5.4 The hierarchy problem

The hierarchy problem relates to the inconsistent magnitudes of the weak and gravitational forces, which involve Fermi's and Newton's constants. As a consequence of quantum corrections, the Higgs bare mass receives quadratically divergent contributions from loop diagrams. In the hypothesis of finite corrections in the limit of new physics energy scale Λ, the Higgs boson should still acquire a mass in the scale of Λ. Experimental measurements contradict these predictions, unless there is a delicate cancellation of the bare mass and the quantum corrections.

Supersymmetry provides a solution to this issue, with cancellation of the fermion and boson loops corrections. Alternative extensions of the Standard Model allow achieving similar results, such as Little Higgs and Twin-Higgs models.

1.5.5 Unification

The Standard Model is a gauge theory of the complicated direct product $SU(3)_C \times SU(2)_I \times U(1)_Y$ with different gauge couplings of the subgroups. In overall, the dynamic of the SM

Lagrangian relies on 19 parameters whose values are not predicted by the theory and need to be measured in experiments :

- Fermion masses ($m_{e,\mu,\tau}$ and $m_{d,u,s,c,b,t}$);
- CKM mixing angles (θ_{12}, θ_{23}, θ_{13}) and CP violation phase δ;
- $U(1)_Y, SU(2)_I$ and $SU(3)$ gauge couplings (g, g' and g_S);
- QCD vacuum angle θ_{QCD};
- Higgs vacuum expectation value v;
- Higgs boson mass m_H.

The spontaneous symmetry breaking of the $SU(2)_I \times U(1)_Y$ gauge symmetry grants the mass of electroweak bosons. Yet, it is achieved requiring a special form of the Higgs field potential. The chirality of the weak-interaction solely looks just as another capricious feature of the model. Also, the number of generations of fermions is not fixed in the Standard Model, and is explicitly written in the Lagrangian. Last, the quantization of particles charge as multiples of $|e|/3$ is unexplained, though it is fundamental in atoms neutrality.

These arguments motivate the search for a more fundamental theory which would allow to derive the Standard Model groups and parameters from fewer hypotheses. For instance, Grand Unification Theory (GUT) assumes that SM gauge interactions merge into a single force at high energy. The distinct coupling constants of the strong, electromagnetic and weak forces result from the decoupling of the gauge fields in the energy regime where we can measure their values. Superstring theory provides an alternative solution to these problems, as a string theory of fermions and bosons, and accounting for gravity as a manifestation of supersymmetric particles.

1.5.6 Matter-antimatter asymmetry

The matter-antimatter asymmetry problem arises from the observation that the universe is essentially filled with matter[11]. The Big Bang is supposedly responsible for the creation of matter and anti-matter in equal quantities. Baryon number is conserved in perturbative interaction processes of the Standard Model, while CP-violation in weak-interaction processes and non-perturbative SM sphalerons allows for a much smaller asymmetry than what is observed. Some mechanisms are proposed to explain how this asymmetry was enhanced, and could be tested using the electric dipole moments of fundamental particles for instance.

[11] convention for the sake of simplicity

Chapter 2

The LHC and the ATLAS experiment

This chapter presents the experimental apparatus of the LHC. This corresponds to the setup and conditions for data taking on which rely physics analyses presented in other parts of this document. The first section deals with the LHC device used for particle acceleration and collisions. The next part presents the ATLAS apparatus used to detect particles produced in LHC collisions.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a circular collider with 27 km circumference. Its construction started in 2000 in the tunnel formerly housing the Large Electron Positron (LEP) accelerator at the border between France and Switzerland. The LHC is mainly used for proton-proton collisions, although heavy ions collisions are being performed as well. Several experiments with various physics purposes are located along the LHC tunnel to exploit events produced in LHC collisions. Objectives of the LHC included the discovery of the Higgs boson, predicted in the Standard Model of particle physics. In the next sections, a brief overview of the LHC system is presented, together with experiments exploiting LHC collisions.

2.1.1 Acceleration chain

The LHC is designed to rise the energy of particles pre-accelerated by a chain of linear and circular accelerators. Protons are produced in the ionisation of a Hydrogen gas, and first accelerated up to a 50 MeV energy by the Linac 2 (Linear Accelerator 2), which is 30 m long. This beam is fed to the Proton Synchrotron Booster (PSB), a circular accelerator with 157 m circumference, which brings the energy of protons to 1.4 GeV. These are then successively injected to the Proton Synchroton (PS) and Super Proton Synchrotron (SPS), which are circular accelerators with respectively 628 m and 6.0 km circumference. Protons accelerated by the SPS have an energy about 450 GeV before being injected into the LHC tunnel. Details of LHC acceleration conditions for protons collisions are given in section 2.1.4.

Protons accelerated in LHC beams are assembled within bunches, separated by a time period corresponding to the bunch crossing time. Bunches of protons are themselves gathered into bunch trains. This structure of LHC beams allows for an efficient acceleration of protons, but presents an important drawback : a single bunch crossing is likely to produce several interactions of protons. This phenomenon is called pile-up, and constrains to the separation of multiple events occurring simultaneously from detector energy measurements.

2.1.2 Collision parameters

Collision conditions at the LHC determine the efficiency of experiment searches or measurements, according to two important parameters. First, the center-of-mass energy $\sqrt{s}$ which defines the energy available for producing new particles out of proton collisions. Higher values

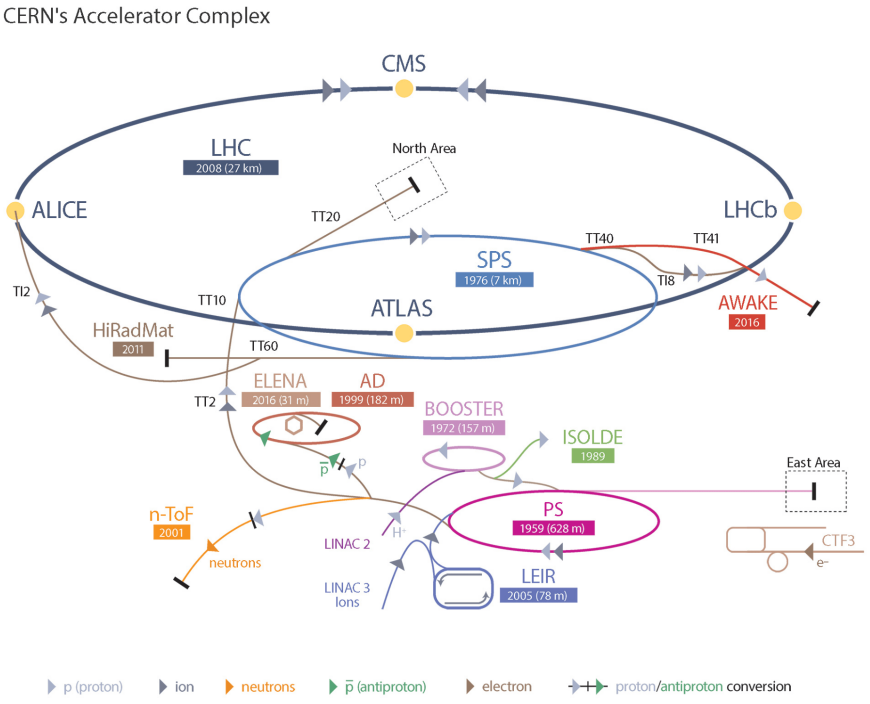


FIGURE 2.1: The LHC accelerator system.

of $\sqrt{s}$ allow for producing heavier particles, which enables parsing higher energy scales for discoveries or constraining many physics models. Also, the instantaneous luminosity $\mathcal{L}$ is an important parameter, as it determines the available statistic (number of events collected over a time period dt), which is often an important source of uncertainty :

$$\mathrm{N} = \int_t \mathcal{L}\, dt\, \sigma \tag{2.1}$$

Where σ is the cross-section of a certain process leading to the production of N events. The luminosity can be computed from beam parameters :

$$\mathcal{L} = \frac{n_1 n_2 f \mathrm{N}_b \mathrm{F}}{4\pi\sigma_x\sigma_y} \tag{2.2}$$

With :

- n_1 and n_2 the number of protons per bunch in opposite beams ($\sim 1.2 \times 10^{11}$);
- N_b the number of protons bunches per beam (~ 2500);
- f the revolution frequency of bunches (11.1 kHz);
- F a geometrical factor accounting for the collision angle of beams (~ 0.8);
- σ_x and σ_y the transverse size of beams at the collision point ($\sim 12 \times 10^{-4}$ cm).

The design instantaneous luminosity is $\mathcal{L} = 10^{-34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and the maximal luminosity reached in 2017 and 2018 was a factor two above desgin.

2.1.3 Experiments

Particles produced in LHC collisions are studied by a few experiments with different physics purposes :

- ATLAS (A Thoroidal LHC Apparatus) and CMS (Compact Muon Solenoid) are general purpose experiments with physics programs oriented toward the search for the Higgs boson, new physics, and study of the Standard Model.
- LHCb is a precision measurement experiment dedicated to the study of CP violation and the search for rare decays and Flavour Changing Neutral Currents in the context of B-physics.
- ALICE (A Large Ion Collider Experiment) is an experiment motivated by the study of quark-gluon plasma produced in heavy ions collisions.
- LHCf (LHC forward) is a forward physics (close to LHC beams) program aiming at the study of the origin of ultra-high energy cosmic rays.
- TOTEM (TOTal Elastic and diffractive cross section Measurement) is an experiment aiming at the measurement of the total cross-section, elastic scattering, and diffractive processes in proton-proton collisions.

2.1.4 Data taking

The LHC was operational for physics purpose protons collisions in 2010 at very low instantaneous luminosity. The data taking period from 2011 to 2012 is referred to as Run 1 of the LHC, while Run 2 includes years 2015 to 2018. The period between Run 1 and Run 2 was intended to LHC upgrades, in order to increase the center-of-mass energy and detector improvements. As collisions conditions evolved in order to progressively increase the instantaneous luminosity, the average number of interactions per bunch crossing increased with time, as presented in Fig 2.2 and 2.3. This impacts the goodness of events reconstruction due to parasitic objects caused by pile-up, and is consistently accounted for in Monte-Carlo simulations.

2.1.4.1 Run 1

The integrated luminosity of data recorded during the Run 1 of the LHC amounts to L $= 5 fb^{-1}$ at $\sqrt{s} = 7$ TeV in 2011 and about L $= 20 fb^{-1}$ at $\sqrt{s} = 8$ TeV in 2012, and is represented in Fig. 2.2.

2.1.4.2 Run 2

Since 2015, the Run 2 of the LHC allows for protons collisions at $\sqrt{s} = 13$ TeV. The integrated luminosity delivered by the LHC until the end of 2017 is L $= 92.9 fb^{-1}$ amongst which 80.1 fb^{-1} were recorded and satisfy sufficient data quality criteria to be used for physics analyses, as shown in Fig. 2.3.

2.2 The ATLAS detector

ATLAS [21] is a general purpose detector, aiming at the discovery of the Higgs boson, new physics, and addressing the study of the Standard Model. Its cylindrical structure allows for a near complete coverage of particles emitted from the collision point at its center. This feature,

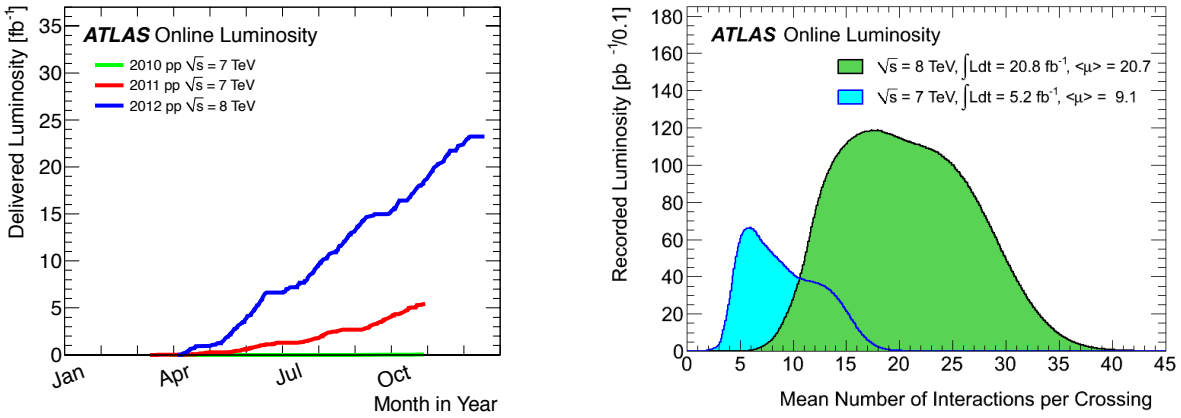


FIGURE 2.2: Profile of integrated luminosity and average number of interactions per bunch crossing in Run 1.

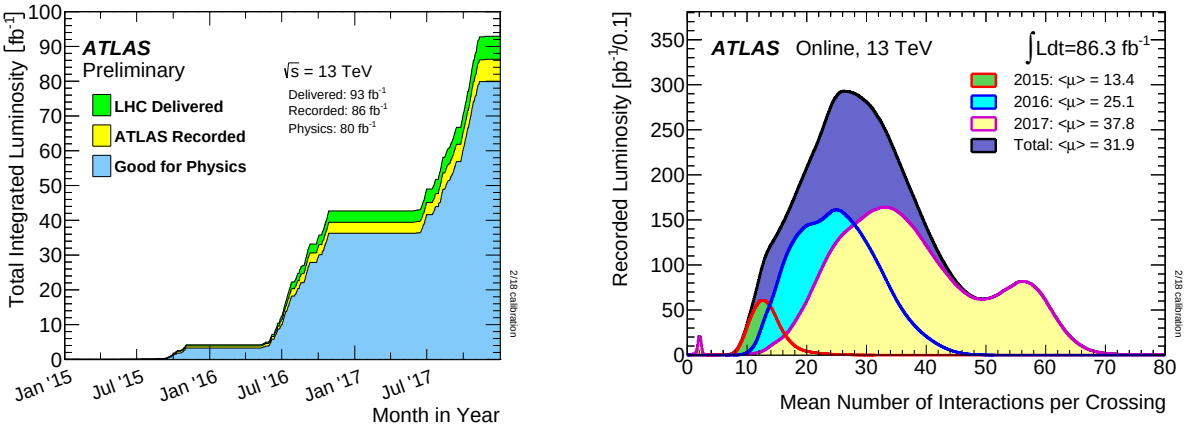


FIGURE 2.3: Profile of integrated luminosity and average number of interactions per bunch crossing in Run 2.

together with the high-density of detector materials, allows for an efficient reconstruction of events produced in LHC collisions.

The inner momentum of partons inside protons prevents us from knowing the momentum of scattering particles along the beam axis. On the contrary, the fraction of parton energy in the transverse plane is negligible. This latter hypothesis has major importance, since energy conservation in the transverse plane allows to infer the vector sum of undetected particle energy, such as neutrinos.

Due to the detector's geometry, cylindrical coordinates are used in the transverse plane. A radius is defined from the distance to the detector center, and is completed by the azimutal angle ϕ. The polar angle θ, defined with respect to the beam axis, is expressed in term of pseudo-rapidity η[1]. In this coordinate system, the angular distance between two objects with respect to the collision point reads :

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{2.3}$$

Central and forward parts of the detector are defined along the longitudinal axis, depending on the pseudo-rapidity[2]. This latter segmentation allows for a more efficient reconstruction of

[1] $\eta = -ln(\tan\frac{\theta}{2})$

[2] The transverse plane at the detector center is defined by $\eta = 0$

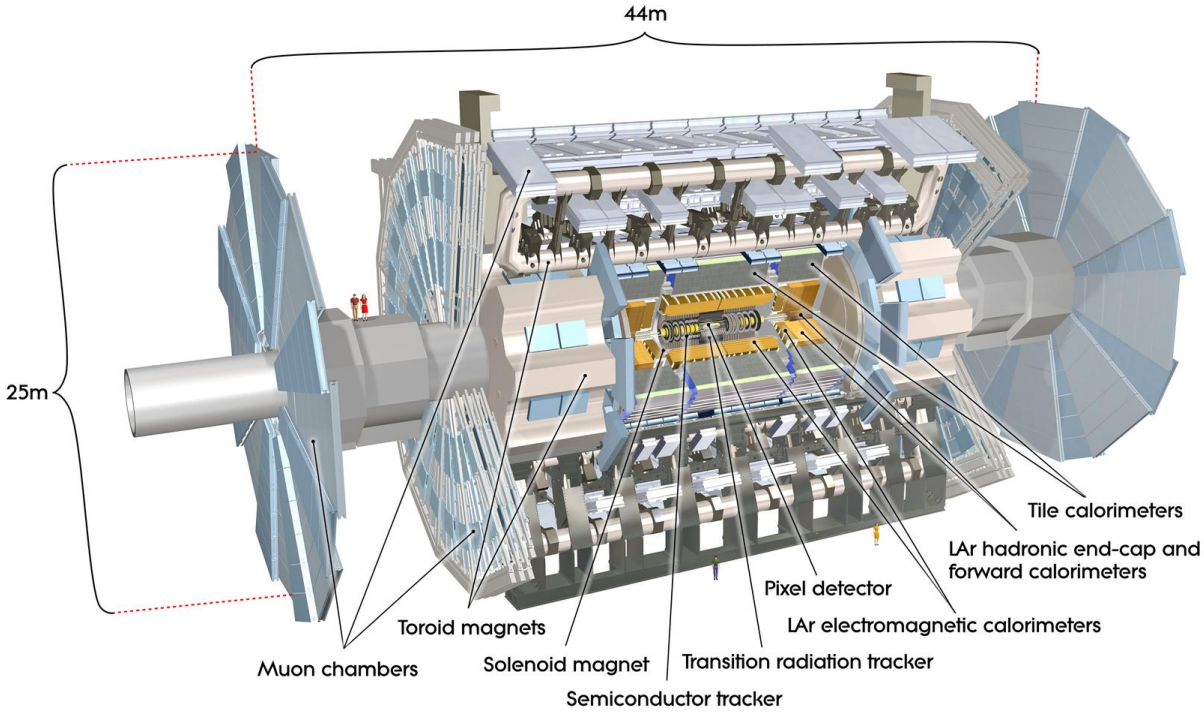


FIGURE 2.4: Schematic view of the ATLAS detector.

forward particles, for which dedicated detectors were designed.

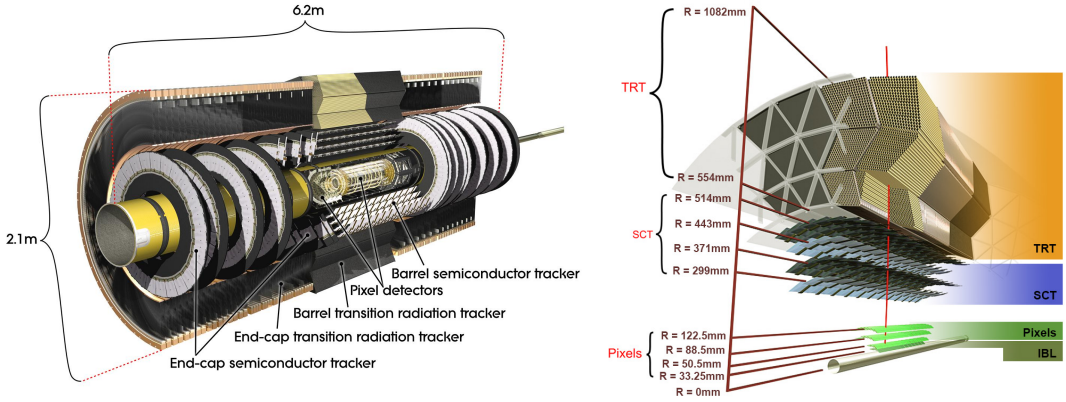
The ATLAS detector is segmented into several subdetectors organized as different layers in the transverse plane, as illustrated in Fig. 2.4. In the following list, they are sorted by increasing distance with respect to the collision point :

- The Inner Detector addresses the reconstruction of charged particles tracks and momentum measurement, and identification of primary and secondary vertices.
- Calorimeters, aiming at the energy measurement of jets of particles and Missing Transverse Energy.
- The muon spectrometer, which is complementary to the inner tracker for muon reconstruction.

Each of these layers is addressed to the detection of specific particles or in a well-defined interaction mode, so that layers have complementary roles in the reconstruction of all particles (except for neutrino). The successive subdetectors are presented with further details in the next sections.

2.2.1 Inner detector

The Inner Detector (ID) consists of succession of layers of pixel detectors, SemiConductor Trackers (SCTs), and Transition Radiation Trackers (TRTs), as illustrated in Fig. 2.5. It has pseudo-rapidity acceptance $|\eta| < 2.5$. The inner detector is immersed into a solenoid generating a 2T axial magnetic field intended to bend the trajectory of electrically charged particles.



(a) Schematic view of the inner detector. (b) Scheme of the inner detector barrel.

FIGURE 2.5: Overview of the inner detector.

This allows to identify the sign and reconstruct tracks momentum from charged particles close to the collision point. The expected relative resolution reads[3] :

$$\frac{\sigma_{p_T}}{p_T} = 0,05\% \; p_T \; [GeV] \oplus 1\% \tag{2.4}$$

Where slight dependencies on particle pseudo-rapidity are not written. The reconstruction efficiency of tracks originating from a primary vertex is about 90% in the barrel and 70% in the end-caps. The impact of pile-up on track reconstruction efficiency is estimated to be less than 1%. The reconstruction of tracks with good efficiency and precision also enables the reconstruction of events primary vertex and secondary vertices from particles decays, which is essential to physics analyses.

2.2.1.1 Insertable B-Layer

The Insertable B-Layer (IBL) [22] is the additional innermost layer of silicon pixel detectors installed during the first shutdown of the LHC, between Run 1 and Run 2. It was inserted inside the Pixel detector of ATLAS in order to compensate performance losses due to radiations, and improve charged particles track reconstruction precision. This is especially useful for long-lived particles such as B-hadrons, which have their decay vertex better reconstructed. The IBL is responsible for important improvements in the efficiency of b-jets identification techniques, as presented in Chapter 4. It uses planar and 3D silicon sensors and a new front-end chip developed in 130 nm CMOS technology intended to cope with high radiation and occupancy levels.

2.2.1.2 Pixel detector

The pixel detector surrounds the IBL. It consists of three cylindrical pixels layers in the barrel region. In the end-caps, three disks were installed around the beam pipe to reconstruct forward particles track up to $|\eta| = 2.5$. Pixels are arranged in order to ensure a good hermiticity of the detector, and provide three interaction points of a particle with the detector in average.

Detector modules are made of silicon pixels with dimensions $r\phi \times z = 50\mu m \times 400\mu m$. Charged particles passing through the silicon material cause the apparition of electron-hole pairs. The voltage applied makes electrons to drift, which induces an electric current collected

[3] $a \oplus b = \sqrt{a^2 + b^2}$

subsequently. Current intensity values larger than a defined threshold make the particle's passage to be detected. At Run 1, the track resolution in the barrel (end-caps) was about 150 μm along the z-axis (R-axis) and 10 μm along ϕ.

2.2.1.3 SemiConductor Tracker

The SemiConductor Tracker (SCT) surrounds the pixel detector, and generally provides eight measure points. It consists of a superposition of silicon strip layers, relying on the same technology as pixel detector modules. In the barrel, four layers of silicon strips modules ensure a coverage up to $|\eta| < 1.1$. Each module has two layers of silicon slightly rotated against each other, which allows to determine the track's position along the strips. In the end-caps, nine layers are arranged in disks up to $|\eta| < 2.5$. Track resolution in the barrel (end-caps) is about 17 μm in ϕ and 580 μm along the z-axis (R-axis) in the barrel (end-caps).

2.2.1.4 Transition Radiation Tracker

The Transition Radiation Tracker (TRT) surrounds the SCT, and is the outermost layer of the Inner Detector [23]. It consists of a superposition of layers of drift tubes parallel (normal) to the beam axis in the barrel (end-caps). Drift tubes have the form of straws and are filled with a mixture of gases, with 70 % Xenon, 20 % methane and 10 % CO2. Polypropylene fibers surrounding the drift tubes are used as transition radiation material in the TRT.

The propagation of a charged particle at the transition between two materials with different dielectric constants goes with the emission of transition radiation photons, which potentially ionize Xenon molecules contained in TRT straw tubes. The transition radiation emitted depends on the particle mass as $\beta = E/m$, which allows particle identification. This is particularly helpful to discriminate electron tracks from that of charged pions in the range 1 GeV < E < 100 GeV. The propagation of charged particles naturally ionize the gas as well, but does not provide any particle identification information. Electrons originating from the ionization of gas molecules are collected onto a wire located at the center of each straw tube, which acts as an anode and helps to reconstruct the particle trajectory.

Generally, tracks of charged particles are associated thirty measure points in the TRT, with track resolution about 130 μm in the R-ϕ plane.

2.2.2 Calorimeters

The calorimeter system of the ATLAS detector surrounds the Inner Detector. It is intended to measure the energy of electrons, photons and hadrons. It consists of an electromagnetic calorimeter mostly addressing the energy measurement of photons and electrons, and a hadronic calorimeter mainly dedicated to reconstruction of jets of hadrons, as presented in Fig. 2.6. Both are sampling calorimeters, designed as an alternation of layers of sensitive and absorber material. This allows using very efficient materials separately for absorption and detection, but requires correcting energy measurements for the energy loss in non-sensitive layers of the detector. Also, both electromagnetic and hadronic calorimeters are non-compensating, meaning that hadron energy loss in the detector is smaller than that of electrons and photons for identical particle energy. The non-compensating calorimeter response is corrected in order to improve particle energy and Missing Transverse Energy measurement and resolution. The calorimeter system of ATLAS allows a good containment of electromagnetic and hadronic showers not to degrade muon spectrometer performances (punch-through). It also provides a pseudo-rapidity coverage up to $|\eta| < 4.9$ which allows good reconstruction of Missing Transverse Energy, which is an important signature to many physics analyses.

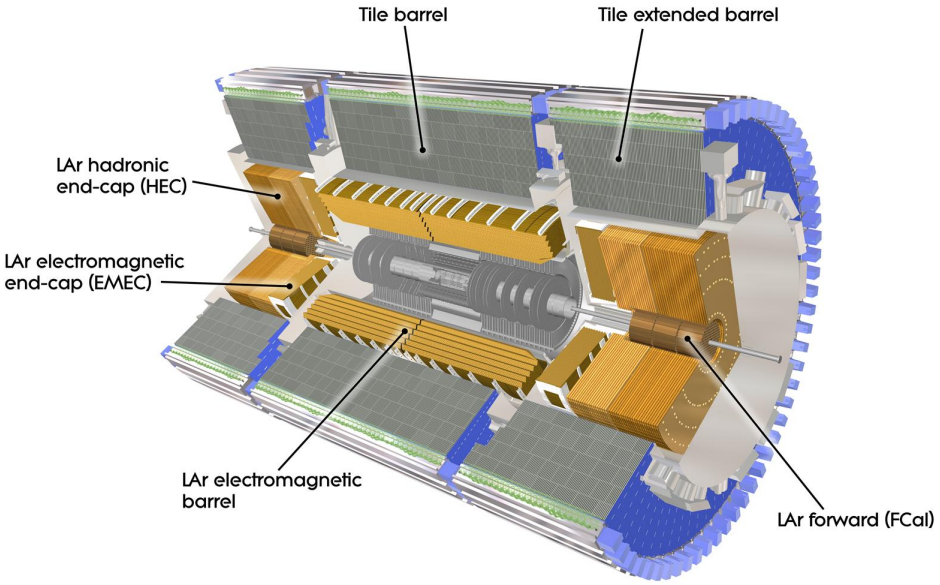


FIGURE 2.6: Schematic view of the electromagnetic and hadronic calorimeters.

2.2.2.1 Electromagnetic calorimeter

The ATLAS electromagnetic calorimeter is divided into a barrel part for $|\eta| < 1.475$ completed by forward components in the end-caps ($1.375 < |\eta| < 3.2$). Sensitive regions of the detector consist of liquid argon material whose ionisation induces charges to drift toward Kapton electrodes. Layers of lead are used as absorber material in order to enhance the development of electromagnetic showers and increase particle energy loss in the calorimeter. The Liquid Argon (LAr) calorimeter of ATLAS is divided into three layers in the transverse plane, with decreasing granularity as electromagnetic showers develop from the collision point. This design allows particle identification, for instance to discriminate isolated photons from those originating from pion decays. Each of these layers is disposed in accordion structure, which enables a full ϕ coverage without azimutal crack.

The electromagnetic calorimeter provides a relative energy resolution :

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.5}$$

Where :

- E is the reconstructed energy.
- a defines the sampling term coefficient, which represents the statistical shower development uncertainty. Its value at η = 0 is 10% with E in GeV.
- b defines the noise term, and is especially important at low energies. It describes pile-up and electronic noise, and its value is about 300 MeV for an electron cluster in the run 1 pile-up conditions.
- c is the constant term, which dominates the energy resolution at high energies. It accounts for mechanical, material defects, and radiation damages, responsible for non-uniform response of the detector. Its value is about 0.7% in the barrel.

2.2.2.2 Hadronic calorimeter

The hadronic calorimeter surrounds the Liquid Argon electromagnetic calorimeter. It consists of scintillating tiles and steel absorber layers in the barrel region, up to $|\eta| < 1.7$. In the end-caps ($1.5 < |\eta| < 3.2$), the Hadronic End-Caps (HEC) use liquid argon as sensitive material and copper as absorber.

The relative energy resolution for single pions reconstructed in the hadronic calorimeter is :

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E\ [\mathrm{GeV}]}} \oplus 3\% \tag{2.6}$$

2.2.2.3 Forward calorimeter

Calorimeters are disposed in the forward region ($3.1 < |\eta| < 4.9$) to enhance the pseudo-rapidity coverage of the detector. This allows reconstructing jets of particles at small angles, which is an important signature of many processes, and to improve Missing Transverse Energy reconstruction. The Forward Calorimeter (FCal) is arranged as three disk layers of Liquid Argon detector material. The first layer of absorber material consists of copper, which allows a good resolution on electromagnetic measurements. The next two absorber layers are made of Tungsten, optimized for hadronic measurements.

2.2.3 Muon spectrometer

The muon spectrometer is the outermost detector system of ATLAS. Muons are the only particles able to propagate to this device, except for neutrinos. It is dedicated to the reconstruction, charge identification and momentum measurement of muons. It relies on the bending of muons track by a 1T (0.5T) magnetic field in the barrel (end-caps) region generated by thoroidal magnets both in the barrel and end-caps. This field is inhomogeneous due to magnetic materials and overlapping fields : its value is known with a 5% uncertainty along the spectrometer.

Several technologies are used for muon reconstruction and triggering, as presented in Fig. 2.7 :

- Monitored Drift Tubes (MDTs) provide a precise measurement of tracks coordinates in most of the $|\eta|$ range.
- Cathode Strip Chambers (CSCs) consist of multiwire proportional chambers. CSCs are well suited to muon reconstruction in forward regions ($2 < |\eta| < 2.7$), as they can sustain demanding rate and background conditions.
- Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) are used respectively in the barrel and end-cap regions to enable muon triggering. Trigger chambers also determine the track coordinate in the direction orthogonal to that measured by precision chambers (non-bending direction).

2.2.4 Trigger system

The nominal event-rate from protons bunch-crossing at the LHC is 40 MHz. This rate has to be reduced in order to be able to record events. Hence, a two-steps selection is performed in order to keep only events relevant to physics analyses.

The level-1 (L1) trigger is hardware-based, and reduces the rate of events to 75-100 kHz using trigger chambers of the muon system, and calorimeter detectors with reduced granularity. This allows identifying events with high transverse energy electrons, photons, muons and jets,

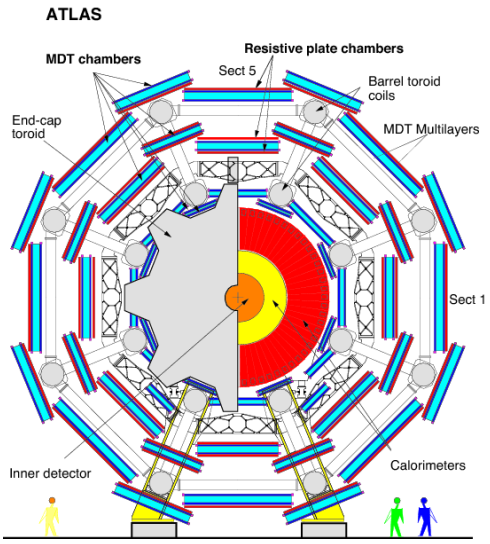


FIGURE 2.7: Schematic view of the muon spectrometer system.

or missing transverse energy. Simple selections addressed to trigger-level objects are designed in order to have a good acceptance for physics processes studied in ATLAS analyses and keep a low event-rate against unwanted events.

Events passing the L1-trigger requirements are transmitted to the software-based High-Level Trigger (HLT) system together with the η-ϕ information of relevant objects, in so-called Regions of Interest (RoIs). The HLT trigger exploits the full detector information within RoIs to further reduce the event rate to about 1 kHz. The remaining events are recorded to disks for further offline analyses.

2.3 Reconstruction techniques in the ATLAS experiment

In this section, an overview of the reconstruction methods used in ATLAS is given. The signals and energy deposits recorded in various locations of the detector are processed and matched to objects with a physics meaning, which can be interpreted in terms of particles and their decay products.

2.3.1 Tracks and vertices

The propagation of charged particles through the Inner Detector is expected to be recorded in the IBL, Pixel detector, SCT and TRT, in the form of signals referred to as $hits$. Tracks are reconstructed from such clusters of signals in the silicon pixel and microstrip sensors, and drift circles in the straw tube tracker. Hits are associated a precise spatial information, so that combinations of hits allow to fit the track direction with a high resolution. This reconstruction technique allows to compute the transverse momentum and trajectory of charged particles, exploiting the bending of tracks cause by the 2T axial magnetic field.

The spatial intersection of tracks is expected to occur at locations from which charged particles are emitted. Such intersection points are denoted as vertices. Identification techniques in ATLAS assume the primary vertex of the hard-scattering collision to present the largest sum of tracks transverse momentum squared.

In most events, vertices are connected by the tracks produced in the decay chain of hadrons. Yet displaced vertices are produced from the decay of long-lived particles. This results in a track offset with respect to the primary vertex, denoted as the impact parameter. This information is very useful in the context of b-Tagging, as described in Section 4.2.1

2.3.2 Photons and electrons

Photons and electrons are reconstructed from energy deposits in the electromagnetic calorimeter, with a fixed size cluster of cells. The granularity in the second sampling of electromagnetic calorimeter is 3x7 (5x5) in the barrel (endcap). A cluster with no matching track (two tracks compatible with a vertex or one track without hit in the innermost pixel layer) is compatible with an unconverted (converted) photon. Conversely, a cluster matched with a track is compatible with an electron. The most recent reconstruction techniques use dynamical size clusters (topoclusters) for photons and electrons reconstruction.

2.3.3 Muons

Muons are reconstructed from tracks in the ID and muon segments in the muon spectrometer. The reconstruction efficiency for muons with transverse momentum greater than 5 GeV is above 99% for most of the covered phase space. Also, some low quality muons can be reconstructed using the ID and energy deposit in the Tile calorimeter.

2.3.4 Jets

Energy deposits in the calorimeter cells are grouped into clusters. Jets are reconstructed from 3D topological clusters gathering cell energies greater than 4 (seed), 2 (neighbour) and 0 (outer crown) sigma of the (electronic + pileup) noise using the anti-kT algorithm [24] with radius 0.4 in most analyses. The energy of jets is calibrated in order to account for reconstruction biases [25]. Also, jets are associated a b-Tagging information, which denotes their probability to originate from the decay of a B-hadron, as described in Chapter 4 with more details.

2.3.5 Missing Transverse Energy

Momentum is conserved in the plane transverse to the beam axis. As a consequence, the vectorial sum of the transverse energy of all particles produced in an single event is expected to be close to null. This hypothesis relies on the assumption that partons collide with null transverse momentum[4]. It allows to build MET, defined for each event as the opposite of the momentum vector sum of objects reconstructed in the ATLAS calorimeters and muon-spectrometer. The intrinsic momentum of colliding partons is negligible in the transverse plane with respect to the resolution on MET, so we can take advantage of this property in individual events.

Potential biases on the MET accuracy are caused by pile-up and the resolution on objects energy measurements, especially for events with large jets multiplicity. MET is a quantity of prime interest for analyses expecting unreconstructed neutrinos in the final state, for which it will present potentially large values. Conversely, fully reconstructed events (with no neutrino expected in the final state) should allow for very little values of MET.

[4] beams are accelerated along the longitudinal direction only

Chapter 3

Simulation and event generators

The following sections present an overview of the principles and programs intended to event simulation to the purpose of most analyses performed by the ATLAS collaboration. This corresponds to Monte-Carlo event generators used for physics predictions in the analyses presented in this thesis. The GEANT4 toolkit and treatment of simulated events in ATLAS analyses is presented as well.

3.1 Motivation and overview

Simulation collectively refers to the set of techniques used in High Energy Physics aiming at predicting experimental observations from collider and detector experiments. Simulation models widely rely on the current theoretical knowledge of particle physics. Also, tunes of parameters used in the simulation are empirically adapted from the study of discrepancies between predictions and observations. Last but not least, Beyond the Standard Model predictions are built and allow experimentalists to exclude theory models or to constrain parameters of these models.

Simulation works as a step-by-step process starting from the generation of physics events. Events generation is implemented so that particle production, interactions and decays follow probability laws. The hadronization of quarks and gluons into observable objects is also subject to simulation programs, similarly to their interaction with detector devices. The output of any step of this process largely depends on tunings and models resulting from an equilibrium between our understanding of particles interactions and the accuracy of Monte-Carlo predictions with respect to experimental observations. Uncertainties in the generation process can largely impact physics analyses results. To this purpose, various Monte-Carlo programs and tunings are used in order to estimate systematic uncertainty in event generators.

3.2 Structure of an event

This section describes the principal items of events generation in the frame of a "truth" knowledge of particles nature, momentum, and interactions. Hence, the particles we deal with in this section should be seen from a theory perspective, and do not represent observed objects : particles detection, reconstruction and measurements are not part of this procedure. Figure 3.1 shows an overview of the main steps developing in Monte-Carlo events generation. Further details on the motivation and implementation of these processes are given in the following.

3.2.1 Hard process

In the context of experiments in hadron colliders, many particles interactions can occur from collisions, amongst which only a tiny fraction is likely to present some interest depending on what physics analysis is performed. An inclusive generation of events in numbers proportional to the respective cross-section of all physics processes occurring in collisions is obviously not

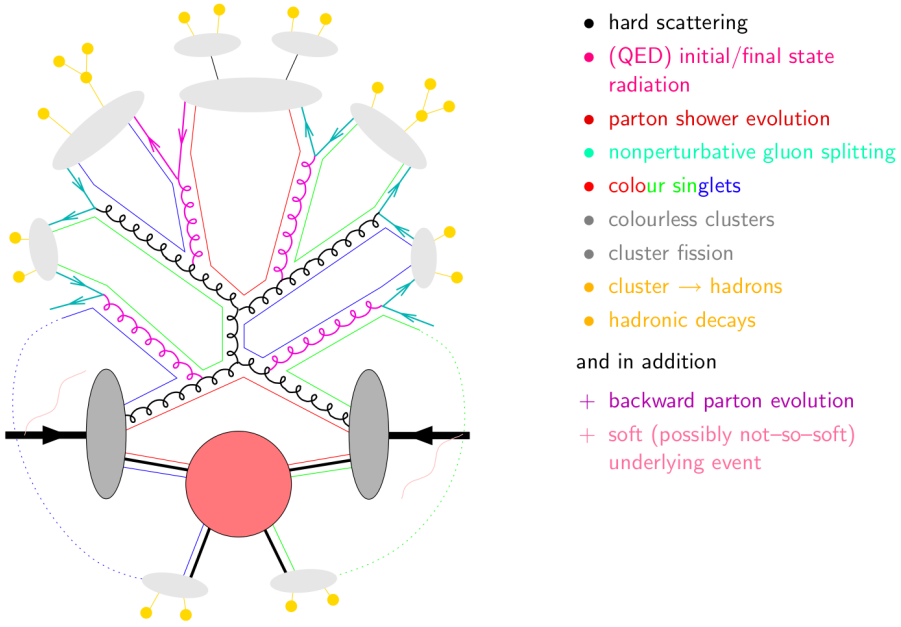


FIGURE 3.1: Schematic view of the main steps of physics processes adressed by Monte-Carlo events simulation.

the preferred approach, as it would lead to statistical issues for rare processes of most interest. Inversely, an individual generation of physics processes is performed by targeting a specific hard process, which represents the process with the highest momentum transfer in the event. This is performed by computing from perturbation theory the probability for a certain hard-scatter to occur. Collisions of composite objects such as protons require exploiting Parton Distribution Functions (PDFs) as probability models for the hard-scatter of elementary quarks within protons. Lowest order perturbation theory allows computing probability distributions for the outgoing partons. Hard process generation programs or libraries are more often referred to as Matrix Element (ME) generators, amongst which Madgraph [3.3.1] and the POWHEG library [3.3.2] are commonly used.

The Monte-Carlo simulation of hadron collisions assumes the QCD factorization theorem, separating the hard scatter from the rest of the event. This allows to compute the cross-section for perturbative processes which occur at the LHC. For instance, the process pp $\rightarrow$ X has inclusive cross-section :

$$\sigma_{\mathrm{X}} = \sum_{a,b} \int_0^1 dx_a dx_b f_a(x_a) f_b(x_b) \times \hat{\sigma}_{\mathrm{ab}\rightarrow \mathrm{X}}(x_a, x_b; \alpha_s, \alpha; \hat{s}, \hat{t}, \hat{u}) \tag{3.1}$$

Where :

- a and b denote the type and flavour of scattering partons carrying energy fractions x_a and x_b of the colliding protons respectively.
- $f_i(x)$ is the Parton Distribution Function, which encodes the probability for a parton i to carry a fraction x of the proton momentum.
- $\hat{\sigma_{ab}}$ is the differential cross-section for a given kinematics of the colliding partons a and b.
- $\hat{s}$, $\hat{t}$ and $\hat{u}$ are the Mandelstam variables which encode the energy, momentum and angles of scattering particles [26].

- α, α_S are the electromagnetic and strong coupling constants

Especially, Equation 3.1 enlights the dependency of cross-sections on protons PDFs. This is a key feature of Monte-Carlo predictions at the LHC, and requires dedicated uncertainties.

3.2.2 Parton shower

Parton Shower (PS) consists of a set of techniques used in order to approximate real emission corrections to the hard process by simulating the branching of a parton into two partons. This allows simulating QCD cascades of partons produced by colour charged quarks and gluons. The same technique is applied in the simulation of photon emissions[1]. Many difficulties in Parton Shower simulation arise in the choice of splitting functions and potential double counting interaction processes generated at different interaction orders[2] in the hard process, with respect to the subsequent Parton Shower. The technique for computing emission probabilities was first developed in the context of QED cascade in terms of Sudakov form factors[3] [27]. The splitting functions enhance the emission of additional partons in the soft and collinear limits. Especially, the simulation of consecutive emissions can rely on different approaches. For instance, angular ordering ensures the coherence of the colour structure in QCD cascades. In this approach, a quark-antiquark colour singlet can radiate gluons confined to the cones delimited by the two quarks. Conversely, p_T-ordering enforces the emission of radiations of decreasing energy. Other approaches were tested as well but are not discussed here.

Parton Showers provide a model for real emissions produced from incoming (outgoing) partons, namely Initial (Final) State Radiations, or shortly ISR (FSR). Especially, ISR simulation is performed after that of the hard process, which requires the impact of ISR to be propagated to the rest of the event going backward in time. Monte-Carlo generators exploit backward-evolution Sudakov algorithms and form factors to this purpose. Also, differences between simulation programs arise in the repartition of the recoil energy generated by the ISR amongst partons involved in the hard process.

A remarkable feature of QCD was pointed out by t'Hooft who showed that extending the description of QCD from $SU(3)_C$ to a $SU(N)$ gauge field theory allows to arrange the colour current of Feynman diagrams into planar flows in the $N \to \infty$ limit [28], while keeping a good description of QCD processes. In this description, the production of quarks which are both colour-connected and adjacent in space is enhanced, which drastically simplifies requirements from QCD colour coherence.

3.2.3 Conciliating fixed order and parton shower simulations

The simulation of a physics process from the appropriate Matrix Element at fixed order allows for reliable simulations of events with high multiplicity of outgoing observable[4] particles and jets. Moreover, this technique is very flexible to phase space cuts at generation level, which allows to optimize the acceptance of simulated events to the intended physics analyses. Nevertheless, the multiplicity of jets and accuracy of jet substructure is not optimal using Matrix Elements, and negative cross-sections can even be achieved in collinear regions. Oppositely, Parton Showers provide a generic method for the simulation of real corrections, are easy to match with hadronization tools, and are sensible to jets structure. A major drawback to the use of Parton Showers is the inefficiency of exclusive final states simulation.

[1] The emission of weak bosons can be simulated as well, but represents smaller corrections.

[2] the interaction order of events processes refers to the coupling order of the appropriate quantum theory in the Matrix Element, i.e. the number of vertices of the interaction

[3] To be accurate, Sudakov form factors encode the non-emission probability

[4] Arbitrary requirements are performed in this definition

Monte-Carlo generators take the best of the two methods. They rely on Matrix Elements expressions in the simulation of processes in the high energy regime of final state particles for exclusive final states, and on Parton Shower in the soft regime. The intermediate regime is left to an overlap of the two methods. Matching and merging are different strategies presented in the following, which allow combining fixed order simulation and Parton Shower consistently.

3.2.3.1 Matching

Matching consists of adjusting Parton Shower to Matrix Element predictions so that the corrections encoded in Sudakov form factors match to the parton-level calculation of the next-to-leading order prediction. In practice, this method is equivalent to exploiting a modified expression of the differential cross-section for events simulation. This procedure was first exploited in MC@NLO [29], and rearranged in the Powheg method [30] to overcome limitations from negative weight events. The KrkNLO approach [31] was developed more recently, and relies on the use of Parton Distribution Functions (PDFs) in a dedicated factorization scheme.

3.2.3.2 Merging

Merging is the strategy of combining events samples generated with Matrix Elements expressions computed at different orders of perturbation theory. This results in 'towers' of Matrix Elements with increasing number of additional particles and jets evolved with Parton Shower. The phase space in between two Matrix Elements expressions is divided using the merging scale. The hard regime is populated by events generated with pure Matrix Element expressions. Events undergo Parton Shower in the soft and collinear regimes. The difficulty of merging is to find the most performing merging variable.

Merging was first introduced with the CKKW method at LO [32]. The FXFX [33] and MEPS [34] merging schemes are now built-in the Madgraph 5 aMC@NLO and Sherpa generators respectively (Sherpa is presented in section 3.3.5) for merging at NLO. Alternatively, the computation of cross-sections in the UMEPS approach [35] focuses on the unitary nature of parton shower. Finally, the MINLO prescription [36] does not require a merging scale but relies on a fine use of renormalization scales.

3.2.4 Hadronization

QCD partons involved in the hard process and Parton Shower do not represent objects that can be observed in the final state. Colour confinement requires quarks and gluons to bind within hadrons observable in detectors. Lund strings and cluster model are the most widely used hadronization models in simulation programs, and are presented in the following.

3.2.4.1 The Lund String Model

The Lund string model [37] takes the quark-antiquark potential $V_{q\bar{q}}$ to be rising linearly with the distance $r_{q\bar{q}}$ between the quark and the anti-quark, as measured in the spectrum of quarkonium spectra and computed in lattice QCD models : $V_{q\bar{q}}(r) = \kappa r$, where $\kappa \simeq 1$ GeV/fm. In this picture, the system tend to accumulate energy as quarks are moving apart, which may eventually lead the system to split and create new quarks to restore low energy quarks systems. In the Lund string model, the presence of gluons in between quarks tends to enhance the production of hadrons in the quarks interspace. This effect is measurable, and its impact was tested with colourflow in the context of the VH(bb) analysis 6.4.

3.2.4.2 The Cluster Model

The cluster hadronization model [38] [39] is based on the colour preconfinement property of QCD [40], which can be inferred assuming that colour-connected partons originating from the QCD parton shower must be near-by in the phase space topology. Adjacent color-lines allow to form primary clusters which are iteratively fragmented toward low mass clusters which are transformed into the final hadrons. This hadronization model also provides a possible discrimination in term of colourflow variables.

3.2.5 Particle decays

The decay of unstable particles and hadrons is implemented in most general purpose Monte-Carlo generators. In the ATLAS experiment, processes in which the parton shower and hadronization is simulated using the Pythia 8 and HERWIG programs presented in sections 3.3.3 and 3.3.4 respectively, have their description of B-hadrons decay improved using the EvtGen tool [41], while tau-leptons decay is currently implemented in TAUOLA [42]. Processes simulated with Sherpa exploit the HADRONS++ package. These programs rely on libraries of decay channels of which many have their kinematic modeled.

3.2.6 Multi-Parton Interactions

A special feature of (anti-) protons collisions arises from that they consist of coloured quarks and gluons bound states. As colour confinement requires quarks to form colour neutral final states, Monte-Carlo generators must consider how protons remnants from a collision hard-scatter should evolve, re-hadronize and potentially interact again with each other. Multi-Parton Interactions (MPIs) are closely related to Minimum Bias events observed in detector experiments, which describe typical 'boring' (anti-) protons collisions, from which a few hadrons are expected to be produced.

3.3 Review of Monte-Carlo generation programs

3.3.1 Madgraph

The MadGraph5_aMC@NLO program [43] brings together the Madgraph5 [44] and aMC@NLO [29] [45] programs, and allows the simulation of hard-scatter events for SM and BSM physics in a user-friendly implementation. It includes the computation of cross sections, and provides a variety of tools relevant to event manipulation and preliminary analyses. Processes can be generated at the LO accuracy for any Lagrangian, and at NLO for QCD corrections to SM and many BSM models.

3.3.2 POWHEG

The POWHEG generator [46] addresses the generation of hard-scatter event at NLO. It relies on the POWHEG method [30] for handling the matching of NLO processes and Parton Shower. A noticeable feature of POWHEG is that it overcomes the generation of negative weight events which can cause statistical issues in samples exploited by physics analyses. If a fraction f of events is simulated with negative weights, the equivalent number of positive weight events which allows to achieve the same uncertainty level is $(1-2f)^2 \times N$. Especially, negative weight events are a major drawback of the MC@NLO method used in the MadGraph5_aMC@NLO generator.

3.3.3 PYTHIA

PYTHIA [47] [48] is a general purpose events generator which is widely used for LHC physics analyses. Many physics models, especially for perturbative and non perturbative QCD, were implemented in PYTHIA, hence extensive Parton Shower studies performed with this tool. PYTHIA comes with capabilities for hard process generation of several processes, but it is most commonly used to exploit subprocess events generated with Madgraph or POWHEG. QCD showers in PYTHIA are ordered in transverse momentum for MPI, ISR and FSR. One of principal features of PYTHIA resides in its hadronization model, the Lund string fragmentation which is described in Section 3.2.4.1.

3.3.4 HERWIG 7

The HERWIG 7 generator [49] is based on the HERWIG (Hadron Emission Reactions With Interfering Gluons) simulation software which was developed in the era of LEP. HERWIG provides several mechanisms for Matrix Element generation, although it is most commonly used for Parton Shower simulation with events generated with POWHEG. Especially, it comes with matching capabilities for many hard processes at NLO using the POWHEG method. Parton Shower in HERWIG is based on angular ordered emissions, and the hadronization procedure relies on the cluster model, as presented in Section 3.2.4.2.

3.3.5 Sherpa

Sherpa is a multipurpose event generator generally used independently of other generation programs presented here. Matrix Element generation is provided in Sherpa through the AMEGIC++ module [50] for Standard Model processes, although higher parton-multiplicity processes can be generated using the better suited COMIX generator [51]. The combined implementation of these two generators allows a generic implementation of merging events generated at different orders in QCD. The default parton shower algorithm relies on the Catani–Seymour dipole factorization formalism [52]. Last, hadronization in Sherpa is performed with the AHADIC++ module, which is based on a cluster model which enforces flavour-dependent transition scale between clusters and hadrons [53].

3.4 Detector response simulation

Analyses performed by the ATLAS collaboration exploit the previous Monte-Carlo programs in association to the GEANT 4 toolkit [54], which provides a simulation of the ATLAS detector response to particles crossing its sensitive material. A model for the detector was designed in order to faithfully account for the geometry, magnetic fields, passive and sensitive materials of the detector. Electromagnetic, hadronic and optical processes can involve a variety of particles over a wide energy range. These features make GEANT 4 a general purpose tool for various experiments, with a great adaptability for designing new detector and study analyses feasibility.

Tracks and energy deposits in sensitive materials of the detector are simulated as HITs in the GEANT 4 framework. On the other hand, the energy loss of particles in passive materials is taken into account in the evolution of particles. The digitization process transforms HITs into DIGITs which mimic the output signals of FrontEnd boards of the ATLAS detector. Pile-Up is taken into account by merging HITs from minimum bias events with that produced in hard-scatter processes.

3.5 ATLAS software

The ATLAS software is implemented in the ATHENA framework [55], which derives from the GAUDI architecture [56] used in several experiments. It provides all features for Monte-Carlo events generation and simulation of the detector response to simulated events. The reconstruction and identification of objects is treated equally for data and simulated events, while special corrections applied to Monte-Carlo events are applied later in analysis frameworks. Also, systematic uncertainties on simulation models and tunes make it necessary to produce alternative versions of the same processes in order to measure the impact of each uncertainty source on analysis results in terms of physics observables. Various analysis frameworks were developed, all based on the ROOT toolkit [57] which allows handy implementations of analysis selections and histogramming capabilities [58], with multivariate and fit methods in the TMVA [59] and RooFit [60] tools respectively.

Chapter 4

Continuous calibration of *b*-Tagging

4.1 Introduction

b-Tagging collectively refers to the activities which aim at identifying jets produced subsequently to the hadronization of a beauty quark, and the decay of the corresponding B-hadron. Such jets are mostly composed of the decay products of the B-hadron. Especially, the first present measurable properties which help to discriminate b-jets amongst all jets reconstructed in the hadronic calorimeter of ATLAS. This identification, namely tagging, is a crucial part in the event selection of many physics analyses. This chapter describes the conception of b-Tagging discriminants and their use in the ATLAS experiment [61].

4.1.1 Motivation

Collisions at the LHC produce final states with high density of hadrons in the volume of the ATLAS detector. The emission of gluons and production of light flavour quarks (up, down, strange) are largely favored with respect to heavy flavour quarks. Therefore, most jets reconstructed in the ATLAS detector originate from gluons or light flavour quarks, and do not present any interest for analyses seeking for b-jets in the final state. The tagging information on jets can be exploited to reject these background sources by comparing the number of b-tagged jets in reconstructed events to the number of expected b-quarks in signal processes.

4.1.2 Performance estimators

4.1.2.1 Label

Within Monte-Carlo simulations, jets are associated a flavour label which denotes their truth particle origin. Just as b-labelled jets (b-jets) originate from B-hadrons, a c-jet refers to a jet produced subsequently to the decay of a C-hadron, excluding any B-hadron origin. Such associations are performed by matching hadrons to jets depending on their angular separations. Also, jets originating from a τ-lepton decay can be identified with dedicated tools, and are denoted as τ-jets. Others jets (neither b- nor c- or τ) are collectively referred to as light-jets. As top quarks do not hadronize, there is no such thing as a top-jet. Most often, the top quark will decay to a beauty quark in a weak interaction process, which is likely to result in the observation of a b-jet in the final state.

4.1.2.2 *b*-Tagging discriminant

Any jet reconstructed from collisions is associated a tagging information, which simply indicates either it is likely to originate from a b-quark or not. This information is based on detector observables processed into b-Tagging algorithms which provide discriminant variables such as that described in section 4.2. Monte-Carlo jets also have this information after simulating their output in the detector.

4.1.2.3 Working Points

The tagging information ranges over several confidence levels, delimited by Working Points (WPs) which denote the expected acceptance for jets originating from b-quarks. In most physics analyses interested in b-jets, the tagging discriminant is used by picking the optimal Working Point in order to maximize the b-jets efficiency with respect to light and charm-jets rejection (inefficiency), depending on the background flavour composition of the analysis. The purpose of pseudo-continuous tagging techniques explained in Section 4.4 consists of exploiting simultaneously more than one of these confidence levels delimited by b-Tagging Working Points.

4.1.2.4 Scale Factors

Estimating b-Tagging algorithms performance on Monte-Carlo datasets allows to compare the efficiency of different algorithms and measure improvements from using new techniques. But given some cut on a b-tagging algorithm output, the tagging efficiency on b-jets is different in data with respect to Monte-Carlo events. For each single jet, this discrepancy is accounted for with a Scale Factor (SF) which depends on the jet kinematic (p_T, $|\eta|$) and multiplies the event weight. Light and charm-flavour jets suffer from identical effects, and the corresponding mistag rates need to be calibrated as well. Scale factors are calibrated in Monte-Carlo to data comparisons of the tagging efficiency in samples with high-purity of jets with the appropriate flavour being calibrated.

4.1.3 Key ingredients

The discrimination of true b-jets with respect to other flavours of jets takes advantage from some specific properties of B-hadrons, which result in observable quantities in the detector :

- Long lifetime, of the order of 1.5 ps ($c\tau \simeq 450$ μm), which increases the flight path length and makes the decay vertex of the B-hadron displaced with respect to the primary vertex.
- Hard fragmentation : a B-hadron retains 70% of the initial b-quark's momentum, which enhances to the property above.
- Large mass : results in a large transverse momentum, hence wide opening angle of the B-hadron decays with respect to the B-hadron axis.

As a consequence of the previous properties, B-hadron (and possibly C-hadron) decay products are produced from a displaced secondary (tertiary) vertex with a large impact parameter with respect to the primary vertex, as shown in Fig. 4.1.

4.2 b-Tagging algorithms

4.2.1 Impact parameter based algorithms

Tracks identified in the Inner Detector are matched to the reconstructed jets, based on their angular distance. Hence, the properties of tracks can be exploited in order to compute the jet tagging information. Especially, the tracks transverse and longitudinal impact parameters are critical for jet-tagging. The transverse impact parameter d_0 measures the distance of closest approach of the track to the primary vertex. The longitudinal impact parameter z_0 characterizes the track offset with respect to the primary vertex, along the beam axis (z coordinate). Impact

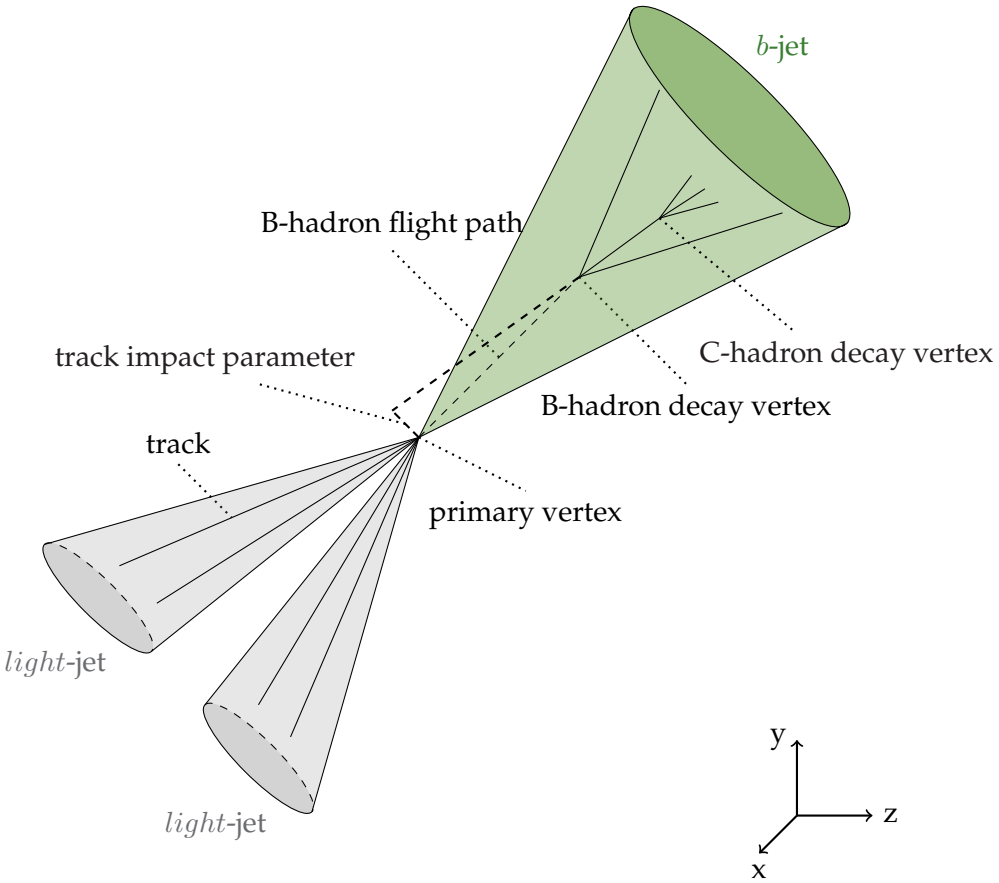


FIGURE 4.1: b-Tagging vertex topology

parameter variables are signed, which allows to further discriminate tracks which originate from a secondary vertex in the case of B-hadron decays. If the track extrapolation crosses the jet axis in front of the primary vertex, the sign is positive : this topology is characteristic of secondary tracks. Otherwise, the sign is negative, meaning that the track's production vertex is likely to be located behind the primary vertex with respect to the track direction. Tracks matched to b- and c jets will tend to have positive impact parameter values, while $light$-flavour jets will present symmetric distributions of their tracks impact parameter variables.

The IP3D tagging algorithm is an impact parameter based algorithm which takes advantage of the tracks properties described above. It combines the ratio of the probability density functions for the signed transverse and longitudinal impact parameter significance of these tracks for the b- and light-flavour jet hypotheses, into a single log likelihood ratio discriminant. A similar algorithm, IP2D, is defined from the transverse impact parameter only.

4.2.2 Secondary vertex finding algorithms

Tracks can also be exploited to find a secondary vertex corresponding to a potential B-hadron decay, as proposed by the SV0 and SV1 algorithms. The first step consists of building two-tracks vertices from candidate tracks, i.e. with large impact parameters. Vertices likely to originate from the decay of a (non-B) long-lived particle (K_S, Λ), photon conversion or hadronic interaction with the detector materials are removed. The surviving vertices are then combined using an iterative procedure, which removes tracks which don't fit to the inclusive vertex.

SV0 only relies on the decay length signed significance of the reconstructed secondary vertex. SV1 combines this latter information with other vertex properties :

- the mass of all tracks associated to the vertex
- the ratio of the sum of tracks energy associated to the vertex and the jet
- the number of two-tracks vertices

4.2.3 Decay chain reconstruction algorithm with JetFitter

The JetFitter algorithm aims at reconstructing the full decay chain through a multivertex fit, that is, from the event chronology :

$$\text{PrimaryVertex} \rightarrow \text{B-hadron} \rightarrow \text{C-hadron}$$

This method assumes that the primary vertex, B-hadron (secondary) and C-hadron (tertiary) decay verticies lie on the B-hadron flight path. Hence, any charged particle track originating from these decays should intersect the same line. This latter hypothesis, presents several advantages for this method :

- The multivertex fit performances are improved from constraining the secondary and tertiary vertices to lie on the B-hadron flight axis.
- Fewer tracks are needed in order to retrieve both vertices. This means that the performance of b-tagging is less degraded for incomplete topologies, and is less dependent on tracks reconstruction performances.

4.2.4 Multivariate algorithms

MV2 is a set of b-tagging discriminants built from Boosted Decision Tree techniques, combining the three basic algorithms previously introduced. The jets kinematic informaion (pT, $|\eta|$) are included in the training to allow the multivariate algorithm to exploit their correlations together with the tagging information. The full list of inputs variables for the MV2, SV and JetFitter algorithms are summarized in Table 4.1. The training dataset takes b-jets as signal sample, and a mixture of c- and light-jets as background sample, in proportions optimized in order to maximize the sensitivity to b-jets. MV2c10 is the most widely used version of MV2, and is trained against 93% of light-jets, and 7% of charm-jets.

4.2.5 Algorithms performances

Given some b-jet efficiency, the rejection of non-b jets is a good estimate of the performance of b-Tagging algorithms. It measures the fraction of tagged non b-labelled jets with respect to the fraction of tagged b-labelled jets.

The performances of the algorithms introduced are shown in Fig. 4.2. In these plots, MV1 is multivariate tagger, which is based on a Neural Network approach and was widely used at the run 1 of the LHC. The MV2 tagger is a major revision of the MV1 tagger. Fig. 4.3 shows the gain in rejection using MV2 with respect to MV1.

4.3 Integrated b-Tagging calibrations

4.3.1 b-jets calibration

In the results presented in the following, the calibration of b-jets [62] relies on a maximum likelihood fit approach performed on a dataset of jets in dilepton $t\bar{t}$ events. The integrated calibration consists of a simplified version of the procedure described in 4.4.1.2 where only two bins

Input	Variable	Description
Kinematic	$p_T(jet)$	Jet transverse momentum
	$\|\eta\|(jet)$	Jet pseudo-rapidity
IP2D, IP3D	$\log(P_b/P_{light})$	Likelihood ratio between the b- and light jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio between the b- and c-jet hypotheses
	$\log(P_c/P_{light})$	Likelihood ratio between the c- and light jet hypotheses
SV	m(SV)	Invariant mass of tracks at the secondary vertex assuming pion masses
	f_E(SV)	Fraction of the charged jet energy in the secondary vertex
	$N_{TrkAtVtx}$(SV)	Number of tracks used in the secondary vertex
	$N_{2TrkVtx}$(SV)	Number of two track vertex candidates
	L_{xy}(SV)	Transverse distance between the primary and secondary vertices
	L_{xyz}(SV)	Distance between the primary and secondary vertices
	S_{xyz}(SV)	Distance between the primary and secondary vertices divided by its uncertainty
	ΔR(jet,SV)	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex
JetFitter	$N_{2TrkVtx}$(JF)	Number of two track vertex candidates (prior to decay chain fit)
	m(JF)	Invariant mass of tracks from displaced vertices assuming pion masses
	S_{xyz}(SV)	Significance of the average distance between the primary and displaced vertices
	f_E(SV)	Fraction of the charged jet energy in the secondary vertices
	$N_{1-trkvertices}$(JF)	Number of displaced vertices with one track
	$N_{\geq 2-trkvertices}$(JF)	Number of displaced vertices with more than one track
	$N_{TrkAtVtx}$(SV)	Number of tracks from displaced vertices with at least two tracks
	$\Delta R(\overrightarrow{p_{jet}}, \overrightarrow{p_{vtx}})$	ΔR between the jet axis and the vectorial sum of the momenta of all tracks attached to displaced vertices

TABLE 4.1: The 24 input variables used by the MV2 b-tagging algorithms.

of the tagging discriminant distribution are considered. The calibration of b-tagging for b-jets results from the combination of individual fits performed in four channels : opposite flavour leptons ($e\mu$) and same flavour leptons ($ee/\mu\mu$) both subdivided into two and three jets regions. As no difference in the calibration results is expected from the various channels, a combination is performed using a Kalman Filter approach, which allows for a proper normalization of the SFs with respect to the corresponding efficiencies in the overall tagweight distribution.

Uncertainties on the Matrix Element modelling, Parton Shower, and hadronisation effects in $t\bar{t}$ events show the largest impact on the b-jets SFs. Also, energy calibration leads to important uncertainties, especially for jets with low transverse momentum value due to migrations amongst different kinematic bins.

4.3.2 c-jets calibration

The integrated calibration of b-Tagging SFs for c-labelled jets [63] exploits samples of jets produced in $t\bar{t}$ single lepton events. The selection aims at events in which a single c-quark is produced in the hadronic decay of a W-boson. The other W-boson is required to decay leptonically. The selection exploits the Kinematic Likelihood Fitter [64] which aims at reconstructing $t\bar{t}$ decays from proton collisions. This reconstruction technique allows to retrieve the most probable combination for matching jets to the expected $t\bar{t}$ decay products.

The flavour composition and modelling of the $t\bar{t}$ process is a major source of uncertainty to the b-Tagging SFs of c-jets. Uncertainties on the SFs associated to b- and light-jets also contribute to the systematic error on c-jets SFs. Last, statistical uncertainties caused by the limited number

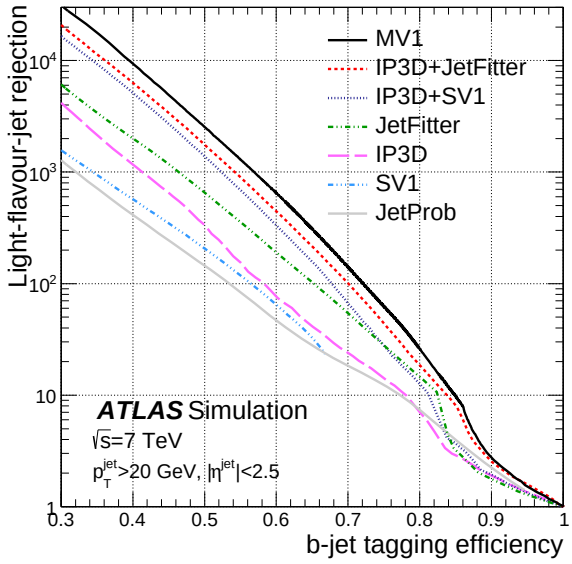


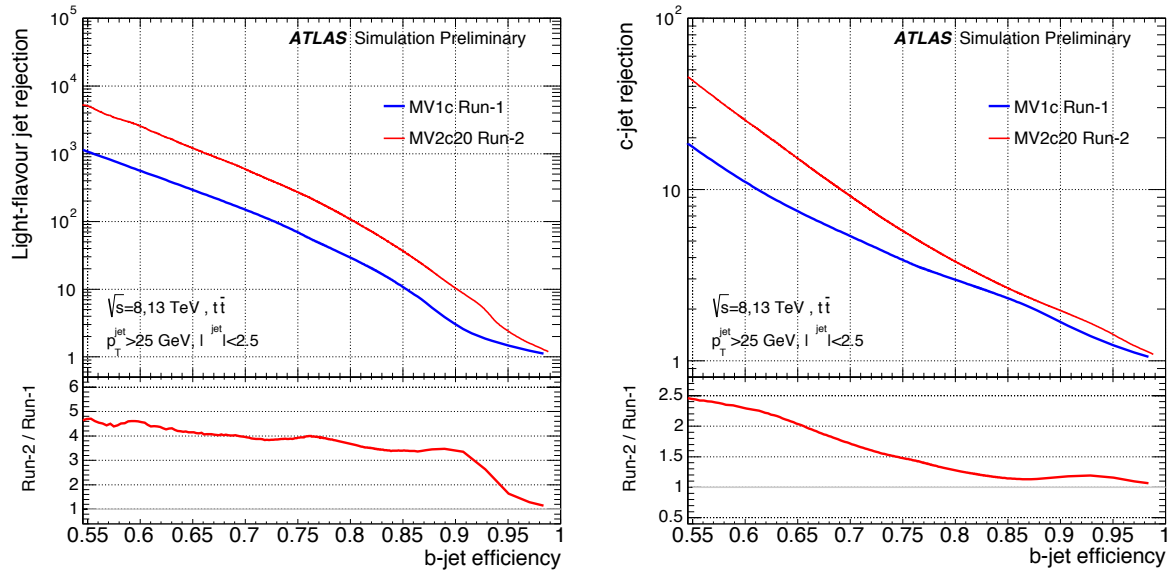
FIGURE 4.2: Light-flavour-jet rejection versus b-jet tagging efficiency, for various tagging algorithms. JetProb is an Impact Parameter based tagger mostly used at LEP and at the Tevatron.

of data events entering the c-jets calibration sample also contribute to increase the uncertainty on c-jets SFs.

This method can be extended by accounting for the shape of the b-Tagging multivariate discriminant in the fit procedure. This allows to retrieve the pseudo-continuous calibration of c-jets b-Tagging SFs.

4.3.3 Light-jets calibration

The efficiency for tagging light-flavour jets using a certain b-Tagging algorithm is referred to as the mistag rate of a tagging discriminant. The rejection of light jets is an important concern of most physics analyses because of the large cross-sections for processes with light-jets in the final state. In general, light-flavour jets are able to trick b-tagging algorithms due to the uncertainty on tracks direction caused by the finite resolution of the Inner Detector. In average, the signed decay length and lifetime-signed impact parameter distributions of prompt tracks associated to light-jets are symmetric around zero. Conversely, b-labelled jets are expected to be associated tracks pointing to the displaced secondary vertex, i.e. in front of the primary vertex with respect to the track direction. Hence, the mistag rate is approximated by the negative tag rate, which consists of the efficiency for tagging jets with reverted tracks signed decay length and lifetime-signed impact parameter. This procedure is straightforward for the basic tagging algorithms, while discriminant based on multivariate methods have an equivalent procedure defined. Distributions of the IP3D+JetFitter discriminant are displayed in Fig. 4.4 and show the impact of the negative tag procedure. In overall, the negative tag rate is expected to provide a good approximation for the mistag rate caused by detector resolution effects. Scale factors for light-jets are estimated on inclusive samples of jets using the negative tag method. Correction factors are applied in order to account for the contribution of long-lived particles (b and c-jets, K_S, Λ, etc.) which have asymmetric distributions of the signed decay length and lifetime-signed impact parameter.



(a) Light jets rejection

(b) Charm jets rejection

FIGURE 4.3: The light (a) and c-jet rejection (b) versus b-jet efficiency for the MV1c b-tagging algorithm using the Run-1 detector and reconstruction software (blue) compared to the MV2c20 b-tagging algorithm using the Run-2 setup (red).

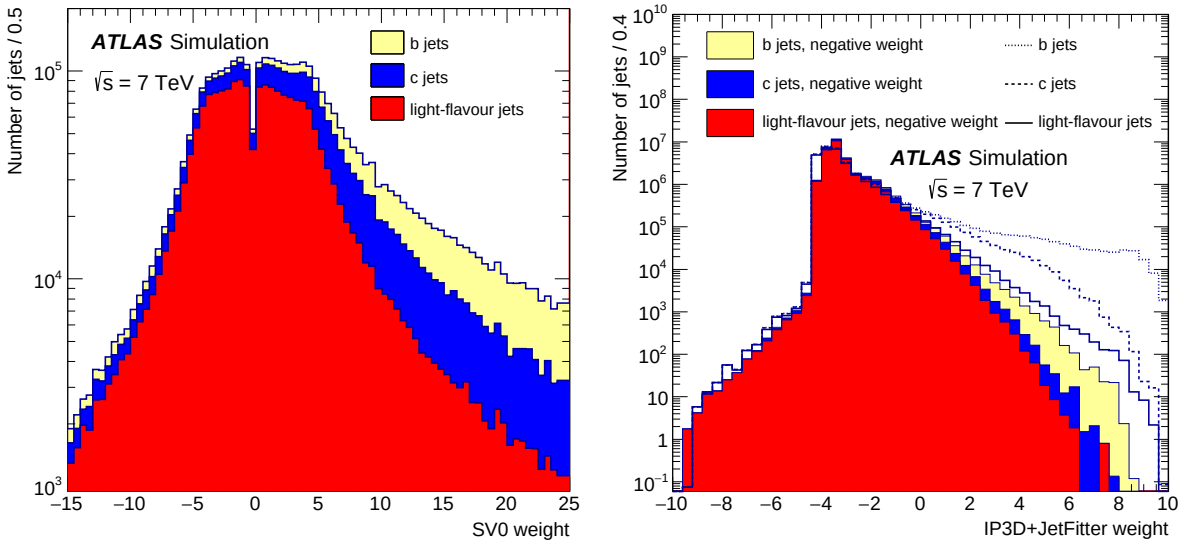
4.4 Continuous calibration of *b*-Tagging

Continuous tagging allows to exploit the exact value of b-Tagging discriminants in physics analyses, instead of simply denoting jets as tagged or not-tagged with respect to a fixed Working Point [65]. This feature provides much more tagging information on jets, as tagging algorithms depend on many input variables which are combined in a single discriminant. Continuous tagging can enhance the sensitivity of many physics analyses, but is generally found to complicate the analyses as well.

In the ATLAS experiment as in the next sections of this document, continuous tagging stands as a shortcut for pseudo-continuous tagging, as the calibration is performed on a binned distribution of tagging discriminants. From a pictorial representation, integrated calibrations can be seen as the calibration of b-Tagging discriminants in 2-bins histograms where explicit bin labels would be "not-tagged" and "tagged". From that perspective, (pseudo-)continuous tagging is an extension to $n-$bins histograms. Methods and results for continuous tagging calibration and use are presented in the next sections.

4.4.1 Methods

Several methods were developed in order to calibrate binned distributions of b-tagging discriminants. The number of bins which can be calibrated should result from an equilibrium between the gain in information provided by additional bins, and the statistical uncertainty which increases at the same time. The current binning results from historical choices, and simply consists of five bins delimited by the four working points used for integrated calibrations. Hence, the bin boundaries for continuous calibration distributions correspond to b-jet tagging efficiencies [100%, 85%, 77%, 70%, 60%, 0%].



(a) (b)

FIGURE 4.4: The SV0 tag weight distributions (a) and the IP3D+JetFitter standard and negative tag weight distributions (b) for b, c and light-flavour jets in a simulated inclusive jet sample. The cut value on the IP3D+JetFitter weight corresponding to 70% b-jet efficiency is 0.35. [61]

4.4.1.1 Derivation from cumulative calibrations

The continuous calibration of b-tagging discriminants can be derived from integrated tagging efficiencies respectively in data and Monte-Carlo samples :

$$SF_i^{continuous} = \frac{\epsilon_i^{data} - \epsilon_{i+1}^{data}}{\epsilon_i^{MC} - \epsilon_{i+1}^{MC}} = \frac{\epsilon_i^{MC} SF_i^{integrated} - \epsilon_{i+1}^{MC} SF_{i+1}^{integrated}}{\epsilon_i^{MC} - \epsilon_{i+1}^{MC}} \tag{4.1}$$

Where :

- i and $i+1$ are indices for consecutive integrated Working Points with decreasing b-jet tagging efficiency
- ϵ_i^{MC} and ϵ_i^{data} are respectively Monte-Carlo and data tagging efficiencies
- $SF_i^{continuous}$ is the integrated calibration scale factor for the Working Point with b-tagging MC efficiency ϵ_i^{MC}
- $SF_i^{continuous}$ is the continuous-tagging calibration scale factor for b-tagging MC efficiencies in the range $[\epsilon_i^{MC}; \epsilon_{i+1}^{MC}]$

From a pictorial perspective, the overall computation of this procedure simply consists of the ratio of the binned b-tagging discriminant normalised distributions in data and Monte-Carlo. As the tagging efficiency for jets depends on their kinematic, continuous scale factors are generally derived in different bins of transverse momentum and pseudo-rapidity of the jets, separately from the calculus on tagweight variables. The computation of uncertainties is constraint by the normalization of the tagweight efficiencies which must sum to unity. This

introduces correlations between tagweight bins which translates in the covariance matrix reflecting statistical uncertainties on the tag weight fractions. A gaussian error propagation is applied to the normalized distribution of tagging fraction p_i in each bin :

$$V_{ii} = ((1-2p_i)V_{ii}^N + p_i^2 \sum_k V_{kk}^N)/N_{tot}^2 \tag{4.2}$$

$$V_{ij} = (-p_j V_{ii}^N - p_i V_{jj}^N + p_i p_j \sum_k V_{kk}^N)/N_{tot}^2 \tag{4.3}$$

Where :

- N_i is the number of tagged jets in bin i
- $N_{tot} = \sum_i N_i$ is the total number of jets
- p_i the fraction of jets in tagweight bin i, $p_i = N_i/N_{tot} = \epsilon_i^{data} - \epsilon_{i+1}^{data}$
- V_{ii}^N are the coefficients of the diagonal covariance matrix $V_{ij}^N = \delta_{ij}(\Delta N_i)^2$

The estimate of systematic uncertainties on continuous SFs is derived from Equation 4.1 using integrated MC tagging efficiencies computed in systematic scenarios. The difference with respect to nominal continuous SFs is then taken as uncertainty.

The simplicity of this approach allows to apply this procedure to any integrated calibration. It is intended to the calibration of $light$.

4.4.1.2 Likelihood calibration

The continuous calibration for b- and c-jets[1] is performed using a direct fit of the binned tagging discriminant template to data. This maximum likelihood fit approach relies on the high-purity of b- and c-jets samples obtained from dilepton and single-lepton $t\bar{t}$ events respectively.

The likelihood function used to derive the b-jet calibration result in the two jets region is :

$$\begin{aligned} L(pT_1, pT_2, w_1, w_2) = \quad & [f_{bb}\ PDF_{bb}(pT_1, pT_2)\ PDF_b(w_1|pT_1)\ PDF_b(w_2|pT_2) \\ & + f_{bl}\ PDF_{bl}(pT_1, pT_2)\ PDF_b(w_1|pT_1)\ PDF_l(w_2|pT_2) \\ & + f_{ll}\ PDF_{ll}(pT_1, pT_2)\ PDF_l(w_1|pT_1)\ PDF_l(w_2|pT_2) \\ & + 1 \leftrightarrow 2]/2 \end{aligned} \tag{4.4}$$

Where :

- f_{bb}, f_{bl} and f_{ll} are the two jets flavour fractions (b stands for b-labelled jets and l for other jets, including c-jets)
- $PDF_{ff}(pT_1, pT_2)$ is the two jets kinematic PDF for jets with flavour label f
- $PDF_f(w|pT)$ is the PDF template for the b-tagging discriminant of a jet with flavour f and transverse momentum pT

A similar likelihood function can be written for three jets events, with additional terms in the combinatoric. PDFs are implemented as one or two-dimensional binned histograms. To preserve the normalization condition, the number of free parameters is set to the number of bins minus one. The binning for transverse momentum PDFs was chosen to be [20, 30, 60, 90,

[1] the calibration of τ-jets is extrapolated from that of c-jets

140, 200, 300] GeV. Depending on the analysis framework, jets with higer momentum values generally have tagging SF equal to that of the last kinematic bin, with artificially increased uncertainty. The $PDF_b(w|pT)$ template is fitted to data, while other PDFs and coefficients are extracted from Monte-Carlo expectations. The fit results from the four different channels are combined throug a χ^2 minimization procedure based on a Kalman filter approach, which allows taking into account both statistical and systematic uncertainties. While statistical uncertainties are uncorrelated, systematic uncertainties can be reduced thanks to correlations across the different channels.

The calibration of c-jets is derived from a single category in three bins of transverse momentum with bin edges [20, 40, 65, 140] GeV. This binning ensures similar statistical uncertainties in each kinematic bin. The dedicated likelihood function is written in order to individually account for the different flavour fractions and kinematics of b-, charm- and light-jets in the charm-jets calibration sample.

4.4.2 Systematic uncertainties

The calibration of b-jets suffers from the low statistics in bins of the b-tagging discriminant with the lowest purity in b-jets. This also applies more generally in kinematic regions with very low and very high transverse momentum of jets. The calibrations of light- and charm-flavour jets mostly suffer from systematic uncertainties in the flavour composition of the calibration samples. The energy calibration of jets is also an important source of uncertainty, as it induces migrations between the different kinematic categories.

4.4.3 Performances

A continuous calibration of the MV2c10 tagging algorithm was performed with data collected in 2015 and 2016, with total integrated luminosity 36.1 fb^{-1}. The results presented below were used in analyses with publications purpose, to measure potential improvements and allow software updates for the more complex implementation of continuous tagging. Especially, one can still notice how the calibration suffers from statistical uncertainties in very low and very high transverse momentum regions. Also, low (high) tagweight bins where very small b- (charm-) jets fractions are expected, show dominant statistical uncertainties as well. Especially, the calibration of b-jets should benefit from the increased integrated luminosity of data collected along the full run 2 of the LHC.

4.4.3.1 light-jets calibration

The pseudo-continuous calibration of b-tagging for light-flavour jets relies on integrated calibration methods described in Section 4.3.3 and the procedure explained in Section 4.4.1.1. Figure 4.5 presents the calibration results, where each distribution corresponds to a distinct region of jets kinematic. Each bin conforms to a single bin of the MV2c10 b-Tagging discriminant, delimited by consecutive Working Points of the appropriate integrated calibration. Very large scale factor values are measured, with important associated systematic uncertainties. In the new version of the ATLAS software, the update of this calibration presents scale factor values closer to one.

4.4.3.2 charm-jets calibration

The pseudo-continuous calibration of b-tagging for charm-jets relies on integrated calibration methods described in Section 4.3.2 and the procedure explained in Section 4.4.1.2. Figure 4.7 presents the calibration results, where each distribution corresponds to a distinct region of jets kinematic.

4.4.3.3 b-jets calibration

The pseudo-continuous calibration of b-tagging for b-jets relies on the procedure explained in Section 4.4.1.2. Figure 4.8 presents the calibration results, where each distribution corresponds to a distinct region of jets kinematic.

4.4.4 Use in physics analyses

Several analyses have an interest in using pseudo-continuous tagging. This concerns especially searches with b-quarks in the final state of signal processes, and a background contamination from light- and charm-jets. Pseudo-continuous tagging allows to loosen the b-tagging selection in order to increase the acceptance for signal events. This comes with a larger background, which generally decreases the inclusive ratio of yields of signal to background events. The pseudo-continuous calibration of b-tagging discriminants allows to exploit the information from tight b-tagging requirements at the same time. At the run 1 of the LHC, the VH(bb) analysis [66] exploited pseudo-continuous tagging in the form of categories, especially in order to better estimate the contamination from background processes with c-jets in intermediate regions of the b-tagging discriminant. The use of pseudo-continuous tagging was estimated to increase the sensitivity of the analyis by approximately 15%.

4.4.5 Prospects

Future versions of the pseudo-continuous calibration of b-Tagging might include a smoothing procedure. This method consists of fitting the binned distribution of b-Tagging SFs using a polynomial kernel estimator in order to retrieve a continuous shape with respect to kinematic variables (p_T, $|\eta|$). Preliminary versions of a smoothed pseudo-continuous calibration were performed but not used for publication purpose.

Updates of the continuous calibration for b-jets are already available with more data and improved taggers relying on deep learning multivariate techniques (DL1). The latest calibration results have support for physics analysis with publication purpose. Yet, the procedure performed to compute the pseudo-continuous calibration is unchanged.

Also, the pseudo-continuous calibration of b-Tagging for jets reconstructed from tracks (namely track-jets) could be performed with the same techniques as described previously. The principal limitations for performing this calibration were the novelty of methods and results with large uncertainties, so that no analysis would have been interested in exploiting such preliminary results.

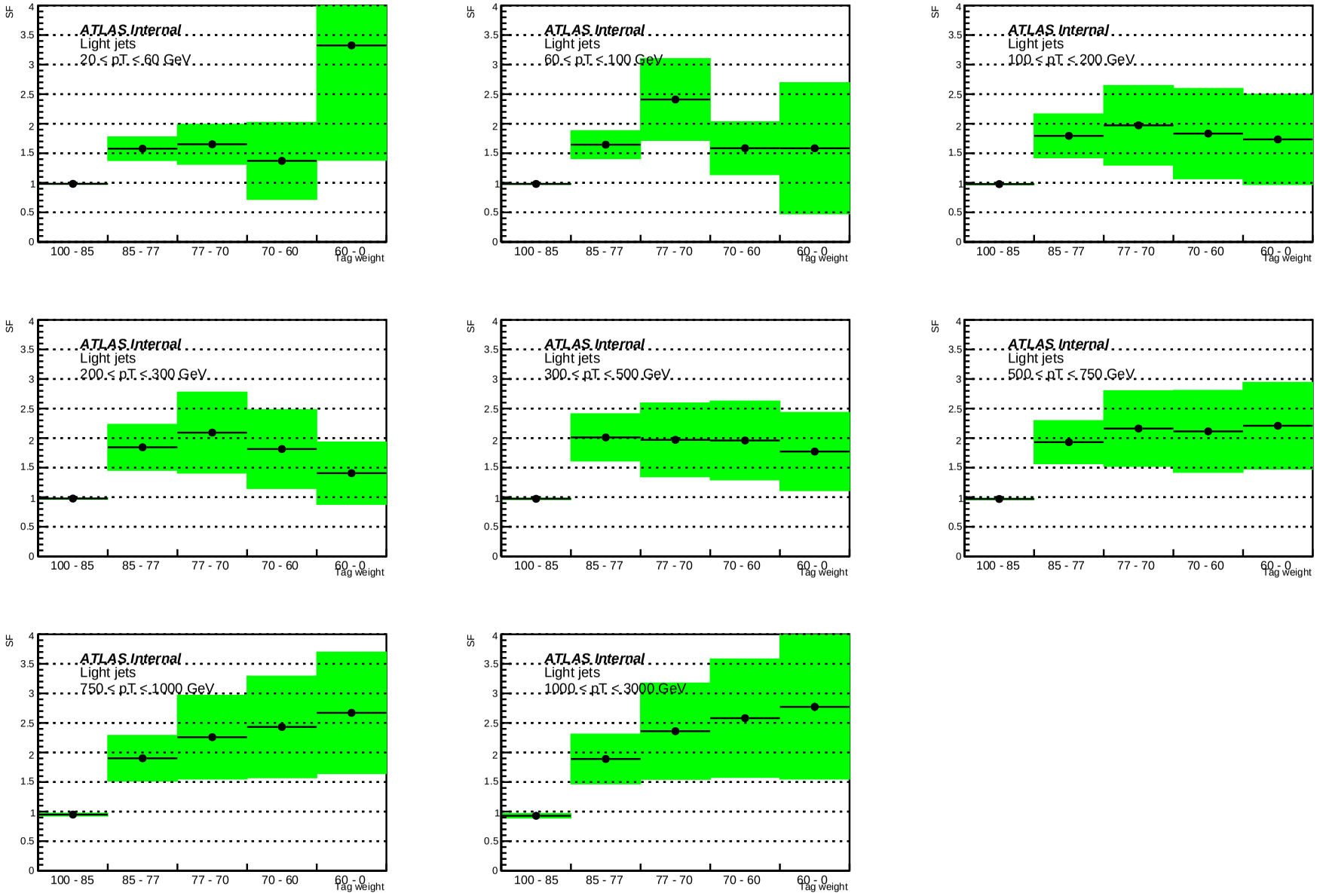


FIGURE 4.5: Light-jets preliminary calibrations results for the MV2c10 tagger using 36.1 fb^{-1} of data collected in 2015 and 2016. The black error bars show the statistical uncertainty, while green bands measure the total error (statistical and systematic). These results deal with jets with pseudo-rapidity $|\eta| < 1.0$. Within each different figure, which represent different transverse momentum regions, bins show scale factors and uncertainties for different tagweight bins with decreasing b-jet tagging efficiency.

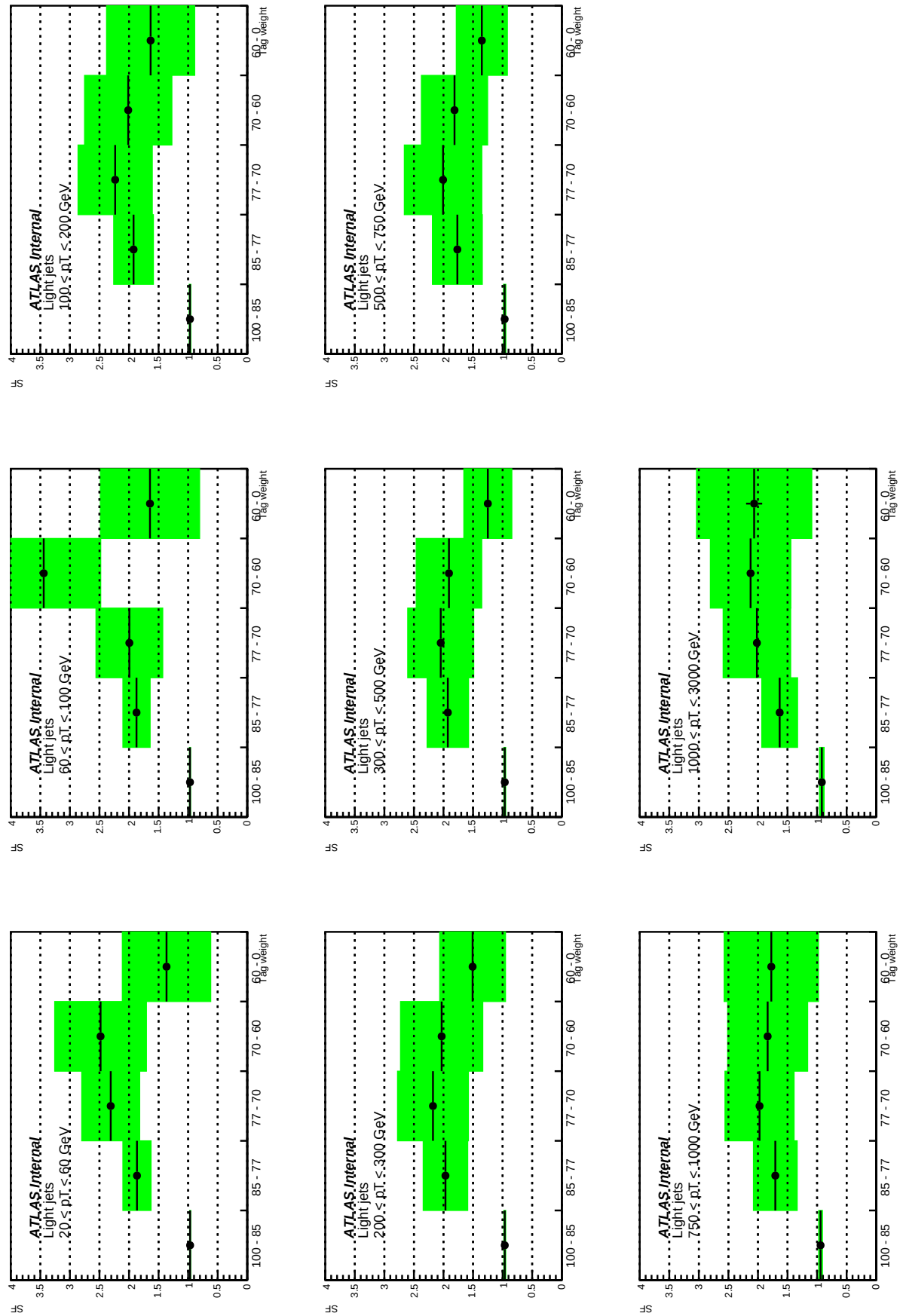


FIGURE 4.6: light-jets preliminary calibrations results for the MV2c10 tagger using 36.1 fb^{-1} of data collected in 2015 and 2016. The black error bars show the statistical uncertainty, while green bands measure the total error (statistical and systematic). These results deal with jets with pseudo-rapidity $1.0 < |\eta| < 2.5$. Within each different figure, which represent different transverse momentum regions, bins show scale factors and uncertainties for different tagweight bins with decreasing b-jet tagging efficiency.

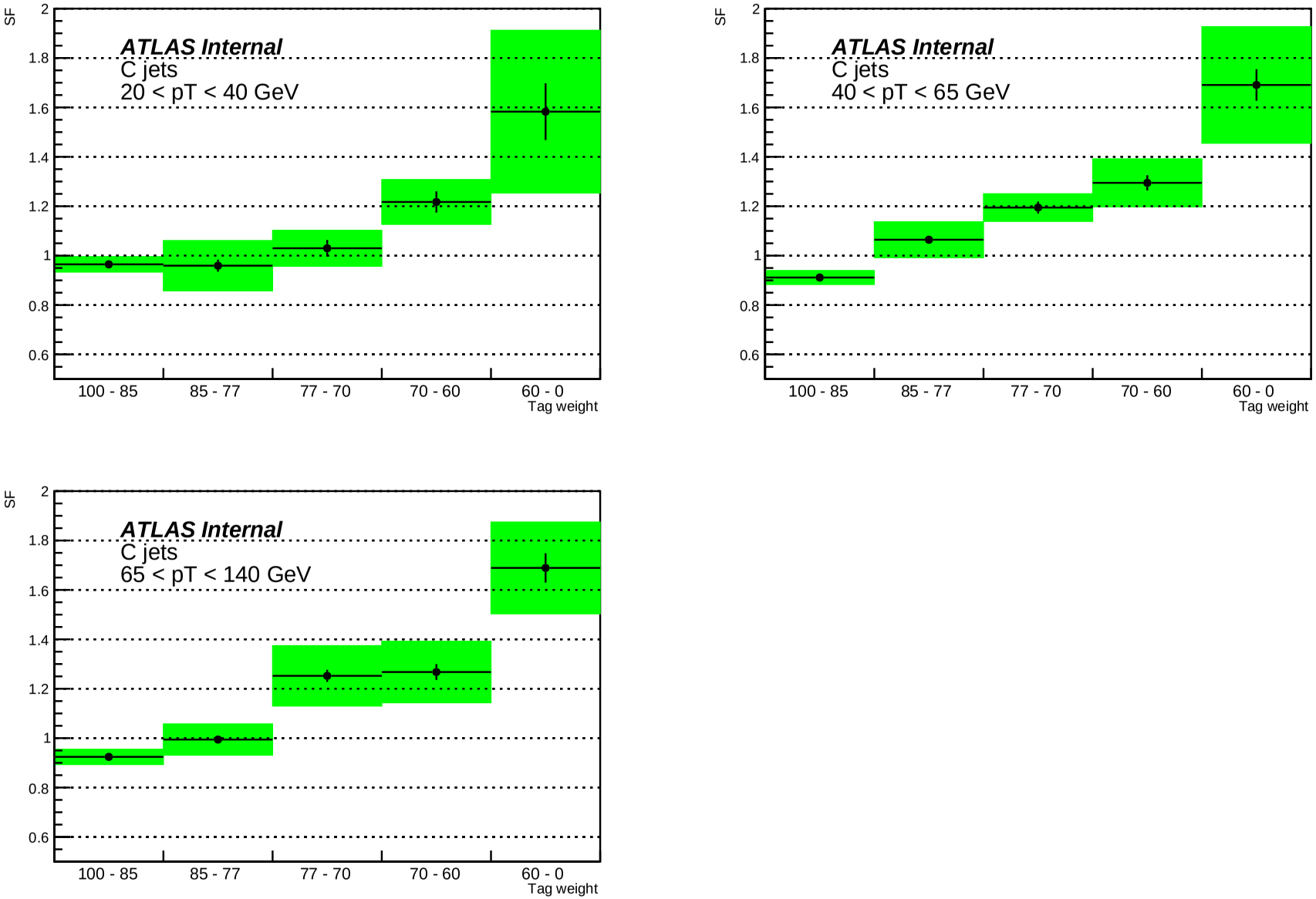


FIGURE 4.7: c-jets preliminary calibrations results for the MV2c10 tagger using 36.1 fb^{-1} of data collected in 2015 and 2016. The black error bars show the statistical uncertainty, while green bands measure the total error (statistical and systematic). Within each different figure, which represent different transverse momentum regions, bins show scale factors and uncertainties for different tagweight bins with decreasing b-jet tagging efficiency.

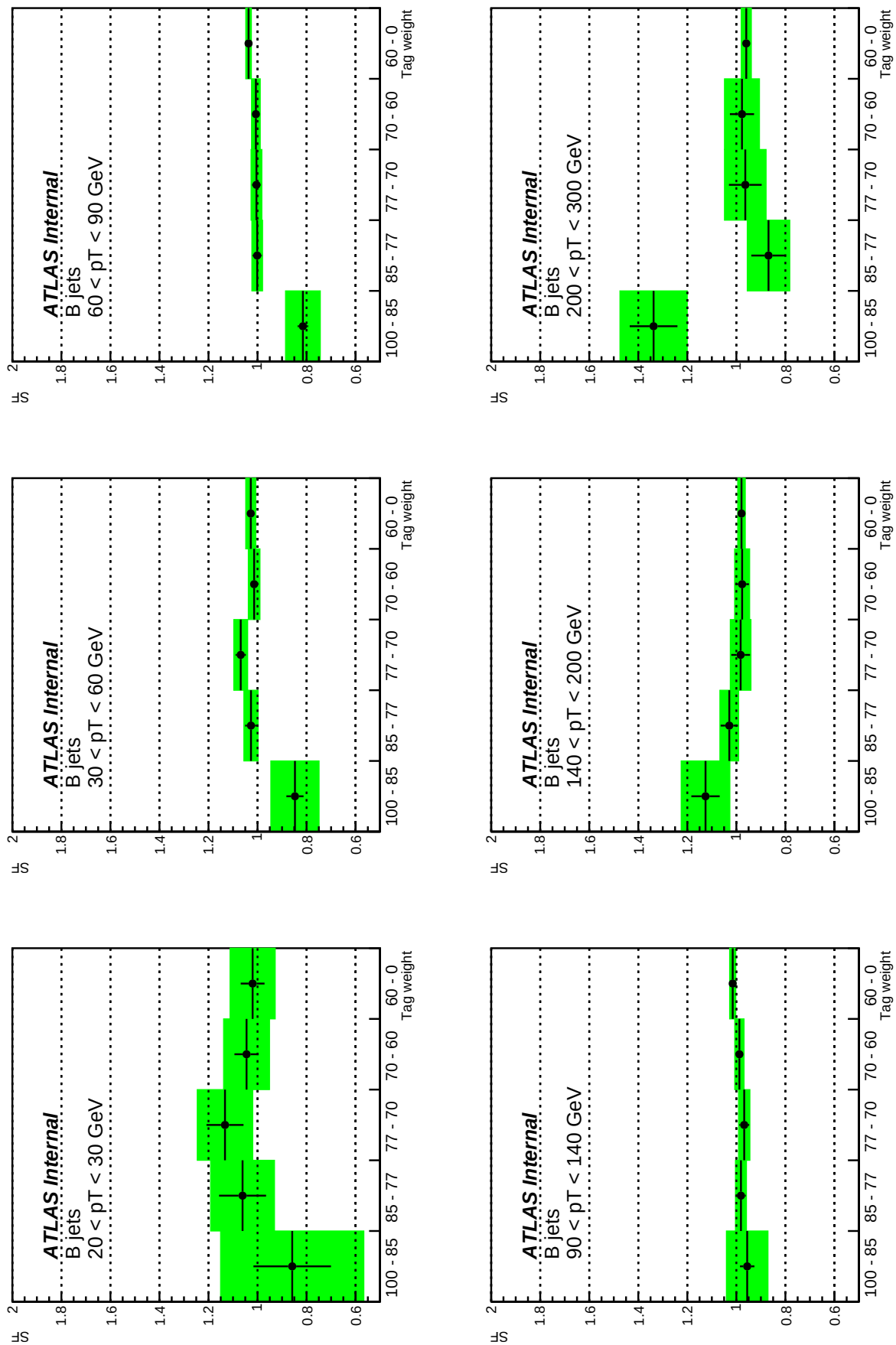


FIGURE 4.8: b-jets preliminary calibrations results for the MV2c10 tagger using 36.1 fb^{-1} of data collected in 2015 and 2016. The black error bars show the statistical uncertainty, while green bands measure the total error (statistical and systematic). Within each different figure, which represent different transverse momentum regions, bins show scale factors and uncertainties for different tagweight bins with decreasing b-jet tagging efficiency.

Chapter 5

Search for the SM VH(bb) process

5.1 Introduction

5.1.1 Context

In July 2012, the ATLAS [67] and CMS [68] collaborations reported the discovery of a new particle compatible with the Higgs boson predicted by the Standard Model at a mass about 125 GeV. This discovery was performed looking for decays of the Higgs boson to vector bosons (γ, Z, W), although the main decay mode for a m_H=125 GeV SM Higgs boson is $H \to b\bar{b}$ with branching ratio 58%. This specific channel is particularly challenging in hadron colliders such as the LHC because of important and complicated backgrounds.

Yet, the CDF and D0 collaborations observed an evidence in the combination of searches for the SM VH(bb) process [69]. Using 9.7 fb^{-1} of data collected at the Fermilab Tevatron collider with $\sqrt{s} = 1.96$ TeV revealed an excess of events in the mass range between 120 GeV and 135 GeV with global statistical significance of 3.1 standard deviations at 125 GeV.

In this section, we present the search for the decay of a Standard Model Higgs boson with m_H=125 into a $b\bar{b}$ pair when produced in association with a W or Z boson. The analysis relies on data collected with the ATLAS detector in LHC protons collisions at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 36.1 fb^{-1}. At the Run 1 of the LHC, the ATLAS [66] and CMS [70] experiments reported a statistical significance of respectively 1.4 and 2.1 standard deviations for the same search.

5.1.2 Signal, background processes and signatures

The SM VH(bb) analysis aims at characterizing events in which a Higgs boson is produced in association to a vector boson, namely $pp \to VH$ ($V = W, Z$), followed by the decay $H \to b\bar{b}$. Candidate events are categorised amongst three channels according to the number of charged leptons resulting from the decay of the vector boson, either 0, 1 or 2 lepton(s), aiming respectively at the $Z \to \nu_\ell \bar{\nu}_\ell$, $W \to \ell\nu_\ell$ and $Z \to \ell^\pm \ell^\mp$ decay processes ($\ell^\pm = e, \mu$).

Background processes lead to final states which can mimic the signature expected from signal events in the detector, but present no interest to the VH(bb) search. Hence, the selection for VH(bb) candidate events relies on objects expected to be reconstructed in the ATLAS detector in the final state of signal processes :

- Two jets resulting from the hadronisation of the b-quarks.
- Potentially one or two charged leptons.
- Potential missing transverse energy from unreconstructed neutrinos.

Motivations and techniques for using Missing Transverse Energy (MET) are presented in Section 2.3.5. In addition, specific kinematic signatures allow to further discriminate signal

from background events. For instance, momentum conservation in the transverse plane enforces back-to-back emission of the Higgs and the vector boson from the primary vertex. Hence, a specific selection is designed in order to sort signal from background events by exploiting individual or combined properties of the reconstructed objects.

A review of the signal and background processes considered in the analysis is presented in the following. Example Feynman diagrams are provided, and the detail of Monte-Carlo generators and cross-section orders used are presented in Table 5.1.

5.1.2.1 Signal processes

The Feynman diagrams of the VH(bb) signal processes are presented in Figure 5.1 at Leading Order. The analysis is split into three channels, according to the number of charged leptons in the final state, each lepton channel aiming at a different vector boson decay.

The SM prediction for VH production estimates the $q\bar{q} \to ZH$ and $gg \to ZH$ cross-sections respectively to $\sigma_{qqZH} = 0.7612\ ^{+0.5}_{-0.6}$ %(QCD scale) $\pm$ 1.7% (PDF) $\pm$ 0.9% (α_S) pb and $\sigma_{ggZH} = 0.1227\ ^{+25.1}_{-18.9}$ % (QCD scale) $\pm$ 1.8% (PDF) $\pm$ 1.6% (α_S) pb [12] at a center-of-mass energy $\sqrt{s}$ = 13 TeV. Especially, the $q\bar{q} \to WH$ production mode has inequal contributions from W^+H and W^-H because of charge asymetry at LHC collisions, with respectively $\sigma_{qqW^+H} = 0.8400$ pb and $\sigma_{qqW^-H} = 0.5328$ pb and relative uncertainties $^{+0.5}_{-0.7}$%(QCD scale) $\pm$ 1.7% (PDF) $\pm$ 0.9% (α_S).

At NLO (QCD), the $q\bar{q} \to ZH$ and $q\bar{q} \to WH$ signals are expected to produce one additional jet in the final state. Contributions from processes with a gluon and a quark in the initial state are also accounted in this production mode. On the other hand, the $gg \to ZH$ signal process has LO contributions from triangle and box Feynman diagrams, which are higher order QCD processes with respect to the $q\bar{q} \to ZH$ and $q\bar{q} \to WH$ signals.

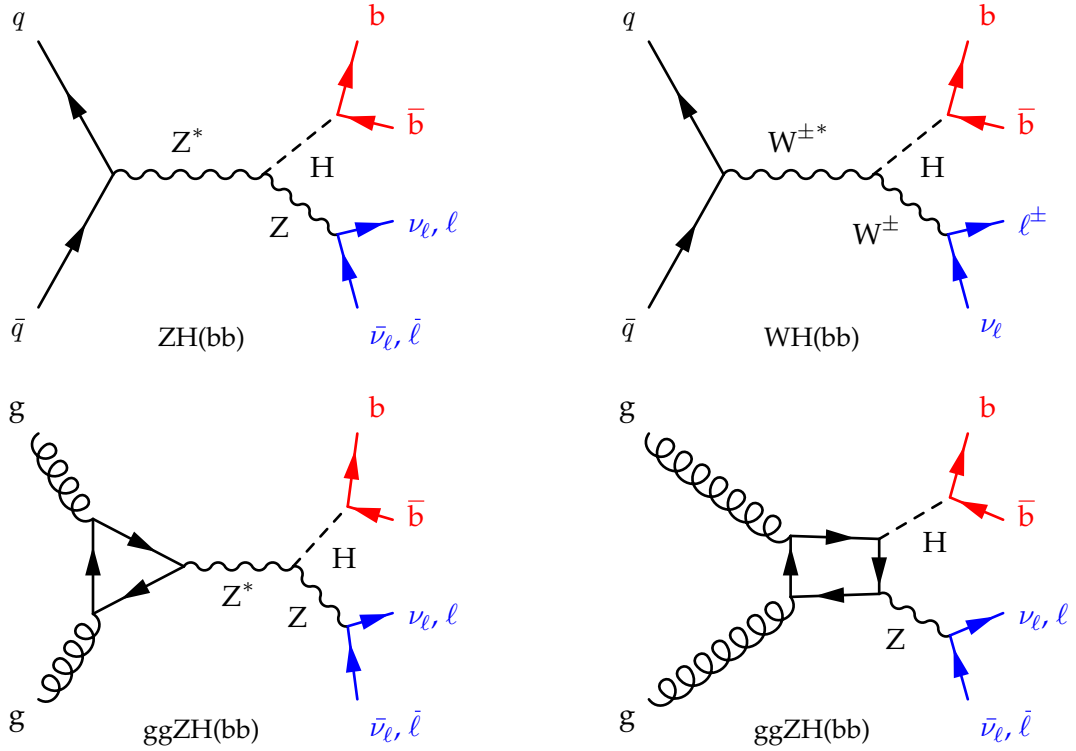


FIGURE 5.1: Example Feynman diagrams for the VH(bb) signal processes.

5.1.2.2 Background processes

Several processes are considered in the background modelling of the VH(bb) analysis. Their respective contributions largely depend on the lepton channel and typical event topology. The most relevant background processes are presented below with example Feynman diagrams, although many more diagrams actually contribute to the total background. Production cross-sections are given for proton-proton collisions with center-of-mass energy $\sqrt{s} = 13$ TeV [12].

- the $t\bar{t}$ (pair production of top quarks) background has inclusive NNLO production cross-section $\sigma_{t\bar{t}} = 831.76$ pb for $m_{top} = 172.5$ GeV, with Feynman diagrams displayed in Fig. 5.2. The $t\bar{t}$ background has important contributions in all 0-, 1- and 2-lepton(s) channels because of the multiple final state possibilities W bosons decays can generate in $t\bar{t} \rightarrow bW^+\bar{b}W^-$. The fully hadronic decay channel is neglected because of the absence of a 4-jets category in the 0-lepton channel, where it would present a dominant contribution. Detailed studies on the typical topologies of $t\bar{t}$ events contributing to the 0-lepton channel background are presented in Section 6.3.

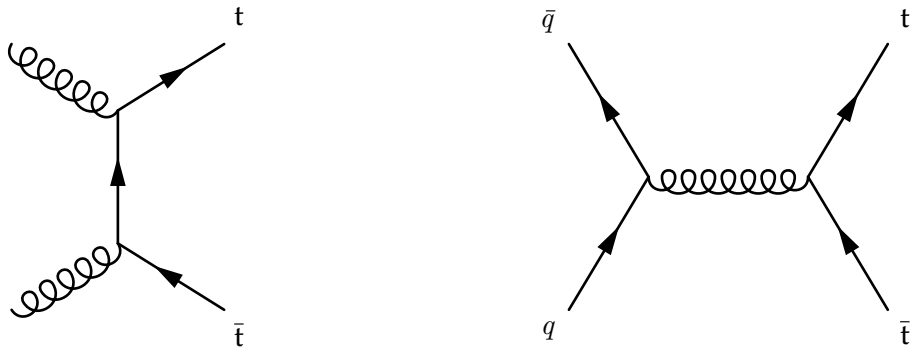


FIGURE 5.2: Example Feynman diagrams for the $t\bar{t}$ background.

- the V+jets background is split into two classes, namely Z+jets and W+jets, with major contributions respectively to the 0- and 2-, and 1-lepton Signal Regions. Possible Feynman diagrams are presented in Fig. 5.3. The inclusive production cross-section for these processes are $\sigma_{Z+jets} = 17512.9$ pb and $\sigma_{W+jets} = 59625.2$ pb. These large values have to be put in perspective to the jets flavour, which are very often light-flavour jets against which b-Tagging offers a very good rejection. Yet, V+jets events with b-jets represent a major background of the 0- and 2-leptons channels.

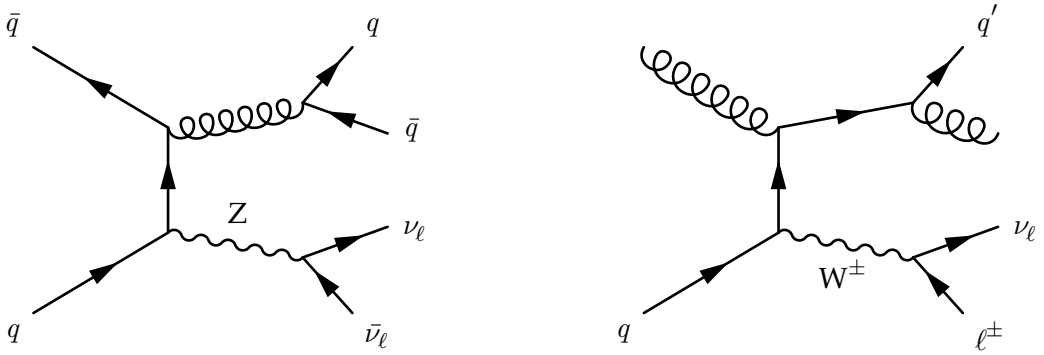


FIGURE 5.3: Example Feynman diagrams for the V+jets background.

- the semi-leptonic diboson background has production cross-section 71.366 pb, inclusively in the flavour of leptons and quarks from vector boson decays. Given the lepton

channels and jets multiplicities considered in the analysis, fully leptonic or hadronic decay modes are neglected. Example Feynman diagrams are displayed in Fig. 5.4.

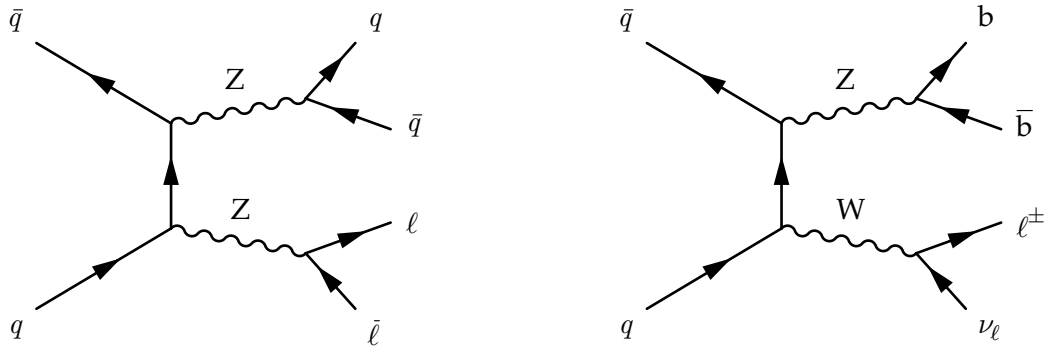


FIGURE 5.4: Example Feynman diagrams for the VV (diboson) background

- the single-top background has three production modes. The single-top t channel cross sections amounts to $\sigma_{single-topt} = 216.99$ pb. The single-top Wt channel (Fig. 5.5c) cross sections is $\sigma_{single-topWt} = 71.7$ pb. The single-top s channel (Fig. 5.5a) cross sections is $\sigma_{single-topWt} = 10.32$ pb. A very low contamination in sensitive regions of the analysis is caused by the s channel. At the end of the selection, the Wt and t channels have similar contributions to the total background yield.

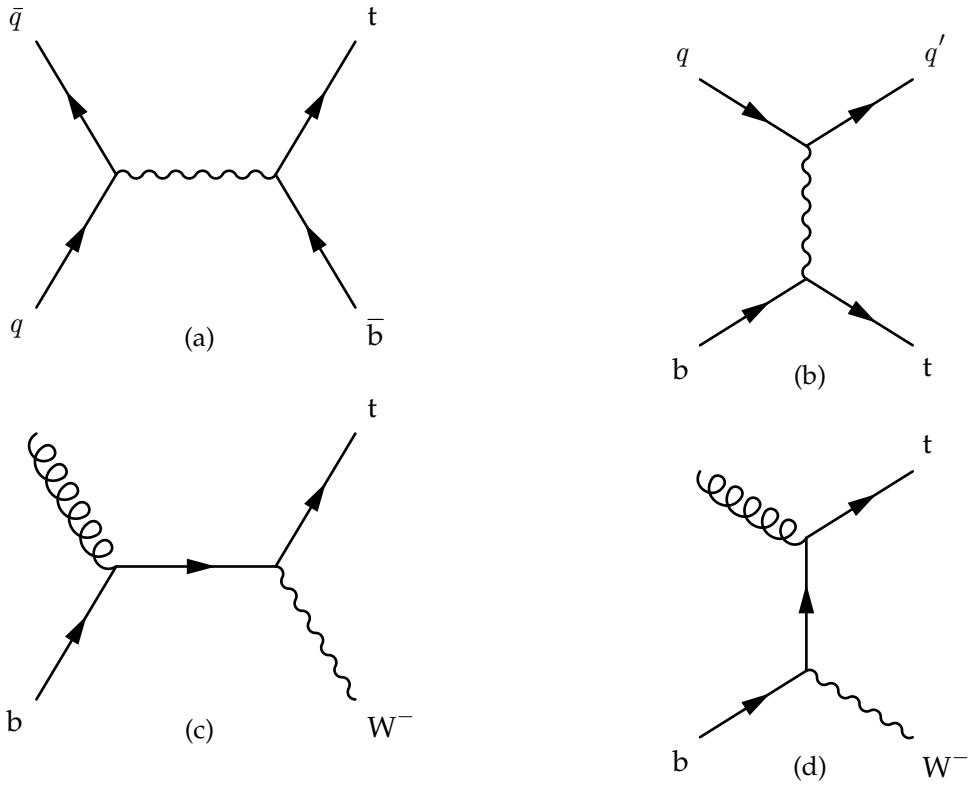


FIGURE 5.5: Example Feynman diagrams for the single-top background in the s-channel (A), t-channel (B) and W-associated channel (C,D) production modes.

- the multijet (QCD) background has inclusive production cross-section $\sigma_{QCD} \simeq 8 \times 10^{10}$ pb. Fig. 5.6 presents an example of Feynmnan diagram for this background. The contamination from multijet events is neglected in the 0- and 2-leptons channels where kinematic cuts allow an excellent rejection. In the 1-lepton channel, the QCD background is estimated using a data-driven method because of limitations in producing multijet samples with sufficient statistics.

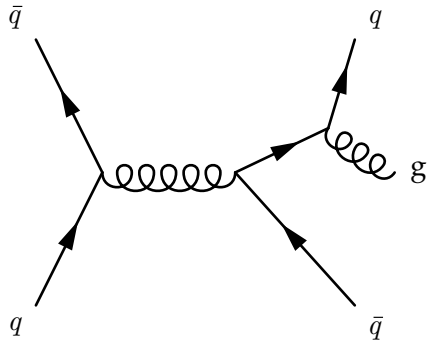


FIGURE 5.6: Example Feynman diagram for the QCD multijet background

5.1.3 Overview of the analysis strategy

The SM VH(bb) search consists of a multivariate analysis which aims at measuring the signal strength $\mu = \sigma_{observed}/\sigma_{SMexpected}$ for the VH(bb) process. It relies on the comparison of predictions based on Monte-Carlo simulations to data collected in the ATLAS detector, as presented in Section 5.2. The analysis depends on specific objects selected amongst all electrons, muons and jets reconstructed in the ATLAS detector, as described in Section 5.3. A selection is performed in order to retain events potentially relevant to the search, as presented in Section 5.4. Multivariate discrimination techniques allow to improve the separation of signal and background events using Boosted Decision Tree methods described in Section 5.5. In order to account for uncertainties in theory predictions, Monte-Carlo modelling and detector measurements, alternative Monte-Carlo simulations are considered as presented in Section 5.6. The Monte-Carlo template of the Boosted Decision Tree discriminant is fitted to data, as described in Section 5.7. This procedure allows to measure the signal strength, accounting for statistical and systematic sources of uncertainty at the same time. A cut-based analysis is performed as cross-check for better understanding the results of the multivariate method. Last, a multivariate analysis targeting diboson events is used as validation search. All results are presented in Section 5.8. Finally, details of further possible improvements and prospects were explored for future versions of the SM VH(bb) analysis, and described in Section 6.1.

5.2 Datasets

5.2.1 Experimental data

In 2015, the LHC started a new period of data taking, namely Run 2. The paper "Evidence for the H $\rightarrow$ b$\bar{\text{b}}$ decay with the ATLAS detector" published in 2017 relies on collision data collected in 2015 and 2016 at a center-of-mass energy of 13 TeV [71]. The integrated luminosity of data exploited by the SM VH(bb) search amounts to 3.2 $\pm$ 0.1 fb^{-1} (2015) + 32.9 $\pm$ 1.1 fb^{-1} (2016). These values account for data periods with no detector defects which would impact data quality and analyses reliability.

Process	ME Generator + Parton Shower	Cross-section order
Signal		
$qq \to WH$	PowhegBox v2 +	NNLO(QCD) + NLO(EW)
$\to l\nu bb$	GoSam + MiNLO + Pythia 8	
$qq \to ZH$	PowhegBox v2 +	NNLO(QCD) + NLO(EW)
$\to llbb/\nu\nu bb$	GoSam + MiNLO + Pythia 8	
$gg \to ZH$	PowhegBox v2 +	NLO(QCD) + NLL(QCD)
$\to llbb/\nu\nu bb$	Pythia 8	
Vector boson + jets		
$Z \to \nu\nu$	SHERPA 2.2.1	NNLO
$W \to \ell\nu$	SHERPA 2.2.1	NNLO
$Z/\gamma^* \to \ell\ell$	SHERPA 2.2.1	NNLO
Top-quark		
$t\bar{t}$	POWHEG + PYTHIA 8	NNLO + NNLL
s - channel	POWHEG + PYTHIA 6	NLO
t - channel	POWHEG + PYTHIA 6	NLO
Wt - channel	POWHEG + PYTHIA 6	NLO
Diboson		
WW	SHERPA 2.1.1	NLO
WZ	SHERPA 2.2.1	NLO
ZZ	SHERPA 2.2.1	NLO

TABLE 5.1: Monte-Carlo samples and cross section order used to normalise signal and background processes at $\sqrt{s}$ = 13 TeV.

5.2.2 Monte-Carlo samples

Monte-Carlo simulated events are used as prediction for the Standard Model backgrounds and VH(bb) signal samples. Events yields are normalised to their most accurate theoretical cross-section prediction. Detector and reconstruction effects for simulated events are performed using a GEANT 4 simulation and the standard ATLAS reconstruction software respectively. Pile-up effects are taken into account by overlaying minimum bias events simulated using soft QCD processes of Pythia 8 with the A2 tune and MSTW2008LO parton distribution functions. For all simulation samples but that generated using Sherpa, the decay of bottom and charm hadrons is described using EvtGen 1.2. The list of generators and models used to generate samples of simulated events is summarized in Table 5.1. Multijet events yields and shape distributions are estimated from data driven analyses described in section 5.4.5.

5.3 Objects selection

5.3.1 Trigger

Data events occuring in the ATLAS detector are recorded only if they activate at least one of several possible triggers. Trigger requirements differ for each lepton channel and depend on the data taking period.

- 0-lepton : a missing transverse energy trigger is used, requiring MET > 70 GeV in 2015 data, MET > 90 GeV and then MET > 110 GeV in 2016 data.

- 1-lepton : single electron events are required to satisfy one of the possible lowest un-prescaled electron triggers. Muons are not considered in the computation of MET for triggering, as they are not reconstructed with an optimal efficiency at trigger level. Hence, single muon events must satisfy the same MET trigger requirement than that in the 0-lepton channel. This corresponds to triggering on the transverse momentum of the W-boson in signal events in the 1-lepton channel. Studies showed that this choice allows to have $\simeq$ 98% trigger acceptance instead of 79% if using single muon triggers with respect to the offline selection.

- 2-leptons : at least one lepton is required to satisfy the un-prescaled single electron or single muon triggers.

Due to slight differences in triggers efficiency in Monte-Carlo with respect to data events, trigger scale factors are applied to account for a potential small bias, as described in Section 5.3.7.2.

5.3.2 Leptons

Leptons are categorised with increasing purity. *Loose* leptons characterize objects reconstructed as leptons within loose requirements. The *loose* leptons category presents a moderate purity in true-leptons, and is used in order to veto events with large multiplicity of leptons. Conversely, *signal* leptons allow to select events with high purity of true leptons, and are used for determining the lepton channel of events. Further details on the reconstruction of each generation of leptons are given below.

- Electrons : requirements for the reconstruction of electrons are described in Table 5.2.

Electron Selection	p_T	η	ID	d_0^{sig} w.r.t. BL	$\lvert\Delta z_0 sin\theta\rvert$	Isolation
VH-Loose	>7 GeV	$\lvert\eta\rvert < 2.47$	LH Loose	< 5	< 0.5 mm	Loose, using tracks
ZH-Signal	>27 GeV			Same as VH-Loose		
WH-Signal	Same as VH-Signal		LH Tight	Same as VH-Signal		Tight, using calorimeter

TABLE 5.2: Selection requirement for reconstructing (anti-)electron. The LH requirement comes from an electron identification Likelihood. d_0^{sig} defines the transverse impact parameter significance. z_0 represents the longitudinal impact parameter, and θ is the polar angle of the electron track.

- Muons : requirements for the reconstruction of muons are described in Table 5.3.

Muon Selection	p_T	η	ID	d_0^{sig} w.r.t. BL	$\lvert\Delta z_0 sin\theta\rvert$	Isolation
VH-Loose	>7 GeV	$\lvert\eta\rvert < 2.7$	Loose	< 3	< 0.5 mm	Loose, using tracks
ZH-Signal	>27 GeV	$\lvert\eta\rvert < 2.5$		Same as VH-Loose		
WH-Signal	>25 GeV	$\lvert\eta\rvert < 2.5$	Medium	< 3	< 0.5 mm	FixedCutHighPtTrackOnly

TABLE 5.3: Selection requirement for reconstructing (anti-)muons. d_0^{sig} defines the transverse impact parameter significance. z_0 represents the longitudinal impact parameter, and θ is the polar angle of the muon track.

- Tau leptons have a short lifetime, and decay either leptonically or hadronically. The first scenario leads to the production of an electron or a muon. The latter case is likely

to result in the production of a jet originating from the tau lepton decay, namely tau-jet. Such jets are removed from the list of Higgs-decay candidate jets. Also, the jets multiplicity of candidate events does not account for τ-jet. Studies were performed in order to improve the rejection of the ttbar background by applying a τ-veto to 0-lepton events. No significant gain was found from such a requirement, because of the decreased acceptance for the WH process which can lead to the production of a τ-lepton in the W-boson decay.

5.3.3 Jets

Following H $\rightarrow$ b$\bar{\text{b}}$ decays, beauty quarks subsequently hadronize and decay, producing sprays of particles in the form of cones which can be reconstructed as jets. Especially, b-tagged jets constitute the best candidates for recovering the energy and momentum of Higgs boson decay products. Jets reconstructed with the Anti-k_t algorithm with radius R=0.4 and meeting the requirements described in Table 5.4 are used in order to retrieve Higgs boson candidates.

The resolution on the Higgs boson mass is enhanced by improving the accuracy of the energy measurement of jets. Especially, b-tagged jets undergo corrections in addition to the JES calibration :

- Muon-in-jet correction : muon production is enhanced in B-hadron decays due to the significant branching ratio BR(B $\rightarrow$ $\mu^{\pm}$X) and BR(B $\rightarrow$ DX$\rightarrow$ $\mu^{\pm}$Y) in B-hadron and potential D-hadron decays. This mode requires special care because muons are long-lived particles with little energy loss going through the ATLAS calorimeters, hence minor energy contribution to b-jets. To compensate for this effect, the four-momentum of calorimeter topoclusters matched to muon tracks are replaced by that of muons as measured in the Inner Detector and Muon spectrometer.

- The $PtReco$ correction is a simple scaling factor intented to further improve jets response by compensating residual differences between the reconstructed and truth energy of jets. Especially, the value of the correction depends on the presence of charged leptons (electron, muon) within the jet radius. Also, the $PtReco$ correction aims at compensating the energy of unreconstructed neutrinos produced in the decay of B-hadrons and potential D-hadrons in signal events for instance. This method is developed with Monte-Carlo samples for which both the truth event and simulated detector response are known.

Jet Category	Selection Requirements		
Forward Jets	jet cleaning $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ $2.5 \leq \lvert\eta\rvert < 4.5$		
Signal Jets	jet cleaning $p_{\mathrm{T}} \geq 60\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.5$	or	$20\,\mathrm{GeV} \leq p_{\mathrm{T}} < 60\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.4$ JVT > 0.59

TABLE 5.4: Selection requirements applied to jets. Jet cleaning refers to the quality criteria which remove jets in regions with hot noisy calorimeter cells [72]. The Jet-Vertex-Taggeg (JVT) discriminant aims at removing pile-up jets.

5.3.4 *b*-Tagging

b-Tagging techniques are described in Chapter 4. In the context of the SM VH(bb) search, the selection of *b*-Tagged jets relies on the MV2c10 algorithm with 70% tagging efficiency for *b*-labelled jets in $t\bar{t}$ simulation samples. The choice of this algorithm followed the latest recommandations for ATLAS analyses targeting *b*-labelled jets. The 70% *b*-jet integrated efficiency working point was shown to present optimal expected performances in SM VH(bb) search. It allows both to keep a reasonable signal efficiency and reject large fractions of light- and charm-jets backgrounds. The *c*- and light-jets rejections using this *b*-Tagging Working Point are respectively estimated to 12.2 and 381.

Continuous *b*-Tagging allows to exploit more *b*-Tagging information in order to assess the probability for a jet to truly originate from a b-quark, as described in Chapter 4. In the context of the SM VH(bb) search, this technique allows at the same time to loosen *b*-Tagging requirements, which increases signal acceptance, and to keep sufficient *b*-Tagging information in order to discriminate light- against *b*-jets in the tagging regime with high *b*-jet purity. Also, dedicated categories (Control Regions) based on the *b*-Tagging score of jets allow to better estimate the background originating from *c*-jets. Tests using pseudo-continuous *b*-Tagging were performed. Improvements of the order of 15% were achieved in statistical significance. Extending this estimate to systematics with correct treatment of correlations presented important complications due to difficulties brought by additional categories and systematics in the fit, as experienced in the Run 1 VH(bb) publication [66].

5.3.5 Missing Transverse Energy

The analysis exploits Missing Transverse Energy, which is computed as presented in Section 2.3.5. MET allows to target events with unreconstructed neutrinos which can be produced in the vector boson decay.

Also, tracks unassociated to the reconstructed objects allow to define the MET soft term. This variable is a measure of the soft hadronic or electroweak activity in the detector, and was tested to improve the rejection of the $t\bar{t}$ background in the SM VH(bb) analysis as described in Section 6.3.5.

5.3.6 Missing Transverse Momentum

A track-based MissingET is computed from charged particles reconstructed in the Inner Detector and Muon Spectrometer, namely Missing Transverse Momentum (MissingPt, or MPT). Conversely to MET, Missing Transverse Momentum is not very sensible to pile-up as it exploits tracks associated to the primary vertex, but it does not account for neutral particles. Thus, events with unbalanced momentum of neutral particles in the azimutal plan will have enhanced MPT values, even in the absence of neutrinos.

5.3.7 Corrections to Monte-Carlo predictions

Monte-Carlo predictions rely on models based on events generators and detector response simulations. The developement of these models over time tend to improve the agreement of observables in Monte-Carlo to data comparisons. The most significant discrepancies are corrected in order not to bias physics analyses.

5.3.7.1 Pile-up reweighting

Simulations do not allow to faithfully predict the environment of hadrons collisions at the LHC at the time of Monte-Carlo events generation. MC samples are generated before data taking,

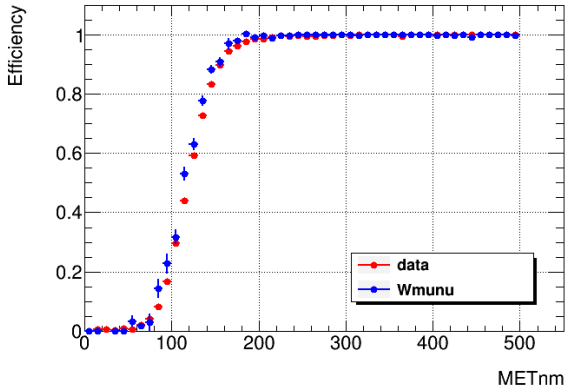


FIGURE 5.7: Measured MET trigger efficiency in data and simulation as function of the offline MET value for the HLT_xe70 trigger menu in W($\mu\nu$)+jets events.

relying on hypotheses for the profile of pile-up which characterizes the average number of interactions per bunch crossing. The profile of pile-up needs to be corrected in Monte-Carlo samples by reweighting simulated events in order to achieve the same average number of inelastic proton-proton interactions per bunch crossing in corrected Monte-Carlo and data samples.

5.3.7.2 Trigger efficiency Scale Factors

Events are preselected at trigger level, which restricts the events topology to a phase space relevant to the SM VH(bb) analysis. As trigger acceptance efficiencies was measured to be different in data with respect to simulation predictions, Scale Factors are applied to reweight Monte-Carlo events. The correction values are extracted from efficiency ratios in datasets with a simpler selection which allows a good comparison of data and Monte-Carlo trigger acceptance, as presented in Fig. 5.7. The trigger scale factor value is very close to unit in most events.

5.3.7.3 *b*-Tagging Scale Factors

Monte-Carlo events weights are rescaled in order to account for the difference in jets tagging efficiency and inefficiency between Monte-Carlo and data. This technique is described with more details in Chapter 4. Within any Monte-Carlo event, each jet is assigned a b-Tagging Scale Factor which multiplies the total event weight. The value of this Scale Factor depends on the label, transverse momentum, pseudo-rapidity, and result of the b-Tagging test of the jet.

5.4 Event selection

This section describes the selection and categories designed in the context of the SM VH(bb) search. Studies on the signal acceptance and contamination from other Higgs processes are presented as well. The methods used in order to estimate the multijet background are explained in the last part.

5.4.1 Selection

The event selection is optimised within each lepton channel in addition to a few common requirements. Selection requirements are described in Table 5.5 for the multivariate analysis. Exactly two b-tagged jets are required, targetting the b-quarks produced in the decay of the Higgs boson. Also, several cuts are exploited in the 0-lepton channel to reject most of the multijet background, as described in Section 5.4.5.1. The diboson analysis used as validation for

the SM VH(bb) search has identical selection, with the multivariate discriminant trained to separate diboson events from other processes. The cut-based analysis aiming at the VH(bb) process exploits additional requirements :

- In the region p_T^V < 200 GeV (p_T^V > 200 GeV), $\Delta R(b1, b2) \leq 1.8$ (1.2) is required.
- In the 1-lepton channel, the criterion m_T^W < 120 GeV improves the rejection of $t\bar{t}$ events, where m_T^W is the W-candidate transverse mass.
- In the 2-leptons channel, MET/$\sqrt{\text{HT}}$ < 3.5 $\sqrt{\text{GeV}}$ is required, where HT is defined as the scalar sum of the transverse momenta of all jets and charged leptons. This allows to further reduce the contamination from the $t\bar{t}$ background.

5.4.2 Categories

Several categories are exploited in the search, which allows to improve the analysis sensitivity. Signal Regions (SR) are defined as regions sensitive to signal processes with largest acceptance for VH(bb) events, and allow to fit the signal strength μ. Conversely, Control Regions (CR) allow to better estimate the dominant sources of uncertainty. Control Regions are designed as analysis categories with very high purity for targeted background processes and low signal acceptance, and are orthogonal to the Signal Regions. Monte-Carlo to data comparisons in these Control Regions allow to better understand Monte-Carlo predictions. The information gained from such comparisons is extrapolated from Control Regions towards Signal Regions, with appropriate extrapolation uncertainties. A summary of the categories used in the SM VH(bb) search is presented in Table 5.6, with additional details given below.

5.4.2.1 Signal Regions

Events are categorised into Signal Regions relying on the events jets multiplicity and measured value of the vector boson transverse momentum p_T^V separately. In the 0- and 1-lepton channels, two categories are made with events with two and three jets respectively, both in the high-p_T^V regime (p_T^V > 150 GeV). In the 2-leptons channel, events with respectively two, or three and more jets, are filled into separate categories. In addition to the high-p_T^V regime, a medium p_T^V regime (75 GeV < p_T^V < 150 GeV) is defined to gain further sensitivity to the VH(bb) process. The signal strength is extracted from a fit to data of the multivariate discriminant Monte-Carlo template (BDT) in all Signal Regions simultaneously.

The selection and categorization strategy relies on the transverse momentum of the vector boson. The p_T^V spectrum extends to larger values in signal events with respect to most background processes, which makes the signal purity to increase when moving to high-p_T^V regions. Hence, the sensitivity of the analysis is strongly related to the compromise between rejecting background events in low pTV regions, and signal acceptance.

In the cut-based analysis, the medium-p_T^V (150 GeV < p_T^V < 200 GeV) and high-p_T^V (p_T^V > 200 GeV) categories are defined in all channels. This splitting allows to further exploit the higher purity of signal events in the high-p_T^V regions. An additional p_T^V-region (75 GeV < p_T^V < 150 GeV) is defined in the 2-leptons channel, similarly to the multivariate analysis.

5.4.2.2 Control Regions

The SM VH(bb) relies on Control Regions to better estimate uncertainties from the W+HF (Heavy Flavor jets : bb + bc + cc+ bl labelled jets) and $t\bar{t}$ backgrounds in the 1- and 2-lepton(s)

Selection	0-lepton	1-lepton		2-lepton
		e sub-channel	μ sub-channel	
Trigger	$E_{\mathrm{T}}^{\mathrm{miss}}$	Single lepton	$E_{\mathrm{T}}^{\mathrm{miss}}$	Single lepton
Leptons	0 loose leptons with $p_{\mathrm{T}} > 7$ GeV	1 tight electron $p_{\mathrm{T}} > 27$ GeV	1 medium muon $p_{\mathrm{T}} > 25$ GeV	2 loose leptons with $p_{\mathrm{T}} > 7$ GeV ≥ 1 lepton with $p_{\mathrm{T}} > 27$ GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 150 GeV	> 30 GeV	–	–
$m_{\ell\ell}$	–	–		$81\text{ GeV} < m_{\ell\ell} < 101$ GeV
Jets	Exactly 2 or 3 jets			Exactly 2 or ≥ 3 jets
Jet p_{T}	> 20 GeV			
b-jets	Exactly 2 b-tagged jets			
Leading b-tagged jet p_{T}	> 45 GeV			
H_{T}	> 120 (2 jets), >150 GeV (3 jets)	–		–
$\min[\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}, \vec{\text{jets}})]$	$> 20°$ (2 jets), $> 30°$ (3 jets)	–		–
$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}, \vec{b}\vec{b})$	$> 120°$	–		–
$\Delta\phi(\vec{b}_1, \vec{b}_2)$	$< 140°$	–		–
$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}, \vec{E}_{\mathrm{T,trk}}^{\mathrm{miss}})$	$< 90°$	–		–
p_{T}^{V} regions	> 150 GeV			(75, 150] GeV, > 150 GeV
Signal regions	✓	$m_{bb} \geq 75$ GeV or $m_{\mathrm{top}} \leq 225$ GeV		Same-flavour leptons Opposite-sign charge ($\mu\mu$ sub-channel)
Control regions	–	$m_{bb} < 75$ GeV and $m_{\mathrm{top}} > 225$ GeV		Different-flavour leptons

TABLE 5.5: Summary of the event selection in the 0-, 1- and 2-lepton channels for the multivariate analysis.

Channel	SR/CR	Categories			
		2 b-Tagged jets			
		75 GeV < p_T^V < 150 GeV		p_T^V > 150 GeV	
		2 jets	3 jets	2 jets	3 jets
0-lepton	SR	-	-	BDT	BDT
1-lepton	SR	-	-	BDT	BDT
2-leptons	SR	BDT	BDT	BDT	BDT
1-lepton	W+HF CR	-	-	Yield	Yield
2-leptons	top eμ CR	m_{bb}	m_{bb}	Yield	m_{bb}

TABLE 5.6: Categories and variables used in the likelihood fit of the SM VH(bb) multivariate analysis.

channels respectively. In these channels, each Signal Region is associated a corresponding Control Region targetting a specific phase space in jets multiplicity and p_T^V.

- The W+HF Control Region aims at better constraining the normalization of the W+HF background normalization in 0- and 1-lepton events. It is defined from events which satisfy the 1-lepton selection described in Table 5.5 with cuts on m_{bb} and m_{top} reversed. This allows for a high purity in W+HF events in this Control Region. The yield of W+HF events expected from Monte-Carlo predictions is compared to that of data entering this category, and helps to improve the normalization of the W+HF background in the fit.

- The top $e\mu$ Control Region aims at better constraining the normalization and modelling of the $t\bar{t}$ background in 2-leptons events. The $t\bar{t}$ background is naturally reduced by requiring 81 GeV < m_{ll} < 101 GeV in the 2-leptons selection. Yet, its contribution is not negligible and can be better estimated with respect to Monte-Carlo predictions. Events with different flavour of leptons are filled into the top $e\mu$ Control Region. This requirement allows for a very high purity of top background, with more than 99% of events originating from the $t\bar{t}$ and Wt processes in Monte-Carlo estimates. The yield of $t\bar{t}$ events expected from Monte-Carlo predictions is compared to that of data entering this category, and helps to improve the normalization of the $t\bar{t}$ background in the fit. Besides, fits to data of the m_{bb} template distribution allow to further constrain systematic uncertainties on the dijet mass modelling for this specific background.

5.4.3 Signal acceptance

Due to the limited detector acceptance, reconstruction efficiency, and the selections applied, only a small fraction of signal events targeted by the VH(bb) search are selected in the sensitive region of the analysis, as illustrated in Table 5.7. In the 0- (1-) lepton channel, the acceptance for $ZH \to \nu_\ell\bar{\nu}_\ell b\bar{b}$ ($WH \to \ell\nu_\ell b\bar{b}$) is estimated to about 3% (1%) in Monte-Carlo simulations. The better acceptance in the two-leptons channel (about 10%[1]) relates to the additional medium-p_T^V categories (75 GeV < pTV < 150 GeV) and to the good efficiency of the ATLAS detector at reconstructing charged leptons. Especially, this allows to exploit events with lower vector boson transverse momentum values in the region. However, the two-leptons channel is statistically limited due to the lower cross-section times branching ratio.

[1]Only the $Z \to ee(\mu\mu)$ decays are considered in this estimate. The 2-leptons channel has very low acceptance for $Z \to \tau\tau$ events with leptonic decays of the τ-leptons because of the requirement 81 GeV < m_{ll} < 101 GeV. On the other hand, the 1-lepton channel aims at all leptonic decays of the W-boson.

Process	Cross-Section x Br [fb]	Acceptance [%]		
		0-lepton	1-lepton	2-leptons
$q\bar{q} \to (Z \to \ell\bar{\ell})(H \to b\bar{b})$	29.9	< 0.1	< 0.1	7.0
$gg \to (Z \to \ell\bar{\ell})(H \to b\bar{b})$	4.8	< 0.1	< 0.1	15.7
$q\bar{q} \to (W \to \ell\nu)(H \to b\bar{b})$	269.0	0.2	1.0	-
$q\bar{q} \to (Z \to \nu\bar{\nu})(H \to b\bar{b})$	89.1	1.9	-	-
$gg \to (Z \to \nu\bar{\nu})(H \to b\bar{b})$	14.3	3.5	-	-

TABLE 5.7: Acceptance for signal processes contributing to sensitive regions of the analysis.

In the 0-lepton channel, the $W \to \tau\nu_\tau$ decay is enhanced with respect to $W \to \mu\nu_\mu$ and $W \to e\nu_e$ because of the possible hadronic decay of the τ-lepton, $\tau \to W(q\bar{q})\nu_\tau$. Such events are expected to produce possibly one or two additional jets, but no charged lepton in the final state. Also, the acceptance for the $gg \to ZH$ is larger than that of the $q\bar{q} \to ZH$ process due to the larger transverse momentum of the vector boson in such events. This relates to the Parton Density Function of the proton, which favours the production of high-energy gluons with respect to quarks and anti-quarks especially.

Furthermore, as the acceptance for leptons reconstruction is limited by the detector coverage and efficiency, possibly one or two lepton(s) can be lost. This effect induces migrations from high- to low- lepton multiplicity channels, and is accounted for in Monte-Carlo estimates. On the contrary, fake leptons can induce events migrations in the opposite way by increasing the number of reconstructed leptons with respect to the nominal event. This effect is negligible for signal processes.

5.4.4 Contamination from other Higgs processes

The VH(bb) search exclusively seeks for the associated production of a Higgs boson, followed by the $H \to b\bar{b}$ decay. Then, any other Higgs production mode or decay channel is considered as a background. Besides, these processes represent a resonant background with dijet mass distribution possibly similar to that of the expected signal. An estimate of the contribution from these processes to the VH(bb) sensitive region was performed in each individual lepton channel. The contamination from these processes was finally considered as negligible, as presented in the following.

5.4.4.1 Higgs production modes

Associated production represents a small fraction (about $\simeq 5\%$) of the total Higgs boson production cross-section at m_H = 125 GeV. To ensure that the analysis is not sensitive to other Higgs boson production modes, the selection acceptance for these latter processes is estimated from Monte-Carlo samples with results shown in Table 5.8. In the 1- and 2- lepton(s) channels, the contamination from the ttH production mode in the 3-jets category is not negligible, but occurs in the low BDT region. The ttH(bb) process is expected to produce four b-jets from the top quarks and Higgs boson decays. Most likely, the selected pair of b-Tagged jets rarely correspond to that originating from the Higgs boson decay. In that case, the dijet mass is expected to be different from 125 GeV. Such events will most-likely not have any impact on the signal strength measurement.

	0-lepton channel		1-lepton channel		2-leptons channel	
	2 jets	3 jets	2 jets	3 jets	2 jets	3 jets
Gluon fusion	0	0	0	0	0	0
ttH	< 0.1	< 1	< 1	4.765	< 1	9.4
VBF	0	0	0	0	0	0

TABLE 5.8: Yields for other Higgs production modes than the VH(bb) Signal.

5.4.4.2 Higgs decay channels

The contamination from non-bb Higgs boson decay channels can be neglected, mostly due to the dominant branching ratio for the $H \to b\bar{b}$ process. The Higgs coupling to light quarks is expected to be very low, and will lead to the production of light-jets which are not likely to satisfy b-Tagging requirements. A similar reasonning stands for the $H \to c\bar{c}$ decay channel, yet to a more moderate extent. Assuming identical kinematic acceptance for $H \to c\bar{c}$ and $H \to b\bar{b}$ events, the number of VH(cc) events entering the VH(bb) sensitive region can be roughly estimated from the VH(bb) signal yields and multiplicative factors :

- $k_{c/b}^{b-Tagging}$ = 0.014 from the difference in b-Tagging efficiency between c-jets to b-jets.
- $k_{Hcc/Hbb}^{BR}$ = 0.049 from the difference in Branching Ratios for m_H = 125 GeV.

In this hypothesis, the number of events produced in the VH(cc) process is expected to be lower than 0.001% of the VH(bb) signal yield.

5.4.5 Multijet estimate

Multijet events represents a peculiar background for the VH(bb) search, and undergo different methods of rejection or estimation in each lepton channel.

5.4.5.1 0-lepton channel

The topology of 0-lepton signal events allows to design a set of kinematic cuts, called anti-QCD cuts, very efficient at rejecting multijet events :

- min$\Delta\phi(MET, jet) > 20^\circ$ (2jet) , $> 30^\circ$(3jet) : this cut allows to remove multijet events in which MET is largely enhanced by the energy mismeasurement of a jet. This results in MET being aligned to the jet direction, and provides a reliable cut for QCD rejection.
- $|\Delta\phi(MET, h)| > 120^\circ$: this requirement enforces the expectation for topologies in which the Higgs boson and vector boson are produced back-to-back in the transverse plane.
- $|\Delta\phi(jet1, jet2)| < 140^\circ$ ensures the small angular separation of jets in the transverse plane. This is expected for boosted Higgs boson decay products, especially in high-MET regimes.
- $|\Delta\phi(MET, MPT)| < 90^\circ$ is based on the reliability of MPT with respect to pile-up. Signal events entering the 0-lepton phase space are expected to miss transverse momentum (energy) in the Inner Detector (calorimeter) from unreconstructed neutrinos. Therefore, MET and MPT are expected to be aligned up to the neutral hadronic component of events, as encoded in this cut.

Cut	VH	Multijet	Multijet/VH	Multijet/VH 80 GeV < m(bb) < 160 GeV
20°	56.40	65.44	1.16	0.44
25°	55.26	33.23	0.60	0.23
30°	54.12	16.87	0.31	0.12
35°	52.94	8.57	0.16	0.06
40°	51.89	4.35	0.08	0.03

TABLE 5.9: Comparison of the multi-jet and VH yields for the 0 lepton 3jet category determined from the fit to the min$\Delta\phi(MET, jet)$ distribution. The second and third columns show the yields for the given selection, the fourth column gives the ratio, and the last column shows the ratio after scaling each yield by the fraction determined from the m(bb) distributions in the range 80 to 160 GeV. A statistical uncertainty on the multijet normalization of 5% is determined from the fit described in the text, while a 15% systematic uncertainty is estimated by varying the fit range at low min$\Delta\phi(MET, jet)$.

Estimating the sensitivity of the VH(bb) analysis to the multijet background is complicated because of the very low acceptance for simulated multijet Monte-Carlo events in the Signal Regions of the analysis. This latter issue is especially limiting due to the very high production cross-section for this process. As a consequence, having no event passing the selection, or having a few events with very large weights in the sensitive region does not allow to give a reliable estimate of the impact of multijet events on the VH(bb) search.

The technique used to get around this limitation is to loosen the selection by removing one or several anti-QCD cuts of the selection, in order to look at the Monte-Carlo to data agreement in the region where multijet events are expected. The efficiency of anti-QCD cuts against multijet events is illustrated in Fig. 5.8 and 5.9 for events with two and three jets respectively. Yet, this also shows that a few events may pass the selection, which reduces the confidence in these stringent cuts.

Hence, a complementary data-driven approach was developed. This method consists of relaxing the cut on min$\Delta\phi(MET, jet)$ as displayed in Figures 5.10 and 5.11 in order to fit the shape of multijet events on data, after subtraction of the Monte-Carlo prediction for the Electroweak backgrounds. The shape of min$\Delta\phi(MET, jet)$ in multijet Monte-Carlo samples is found to be decreasing slightly more rapidly than an exponential function of the form $e^{-\alpha \cdot min\Delta\phi(MET,jet)}$ accounting for multijet events in the fit to data. This procedure gives a good agreement to data, so it is expected to provide a reliable estimate of the yield of the multijet background in Signal Regions. The residual multijet contamination is estimated by integrating the fitted function above the selected cut value. The historical cut showed a sufficient efficiency in the 2 jets region, while an optimisation for the 3 jets category is presented in Table 5.9. The selection now requires min$\Delta\phi(MET, jet)$ to be greater than 30° instead of 20° in the previous versions of the analysis, which allows for a fairly small contamination of the multijet background.

5.4.5.2 1-lepton channel

Multijet events in the 1-lepton channel generally consist of events with real muons or electrons produced in the semi-leptonic decay of heavy flavour hadrons. An additional background source in the electron sub-channel consists of events with electrons produced in the conversion of photons stemming from the decay of neutral pions in jets, or directly from π^0 Dalitz decays. Such leptons are not expected to be isolated, but do satisfy isolation requirements in a small fraction of events. The contamination from multijet events in Signal Regions is estimated from

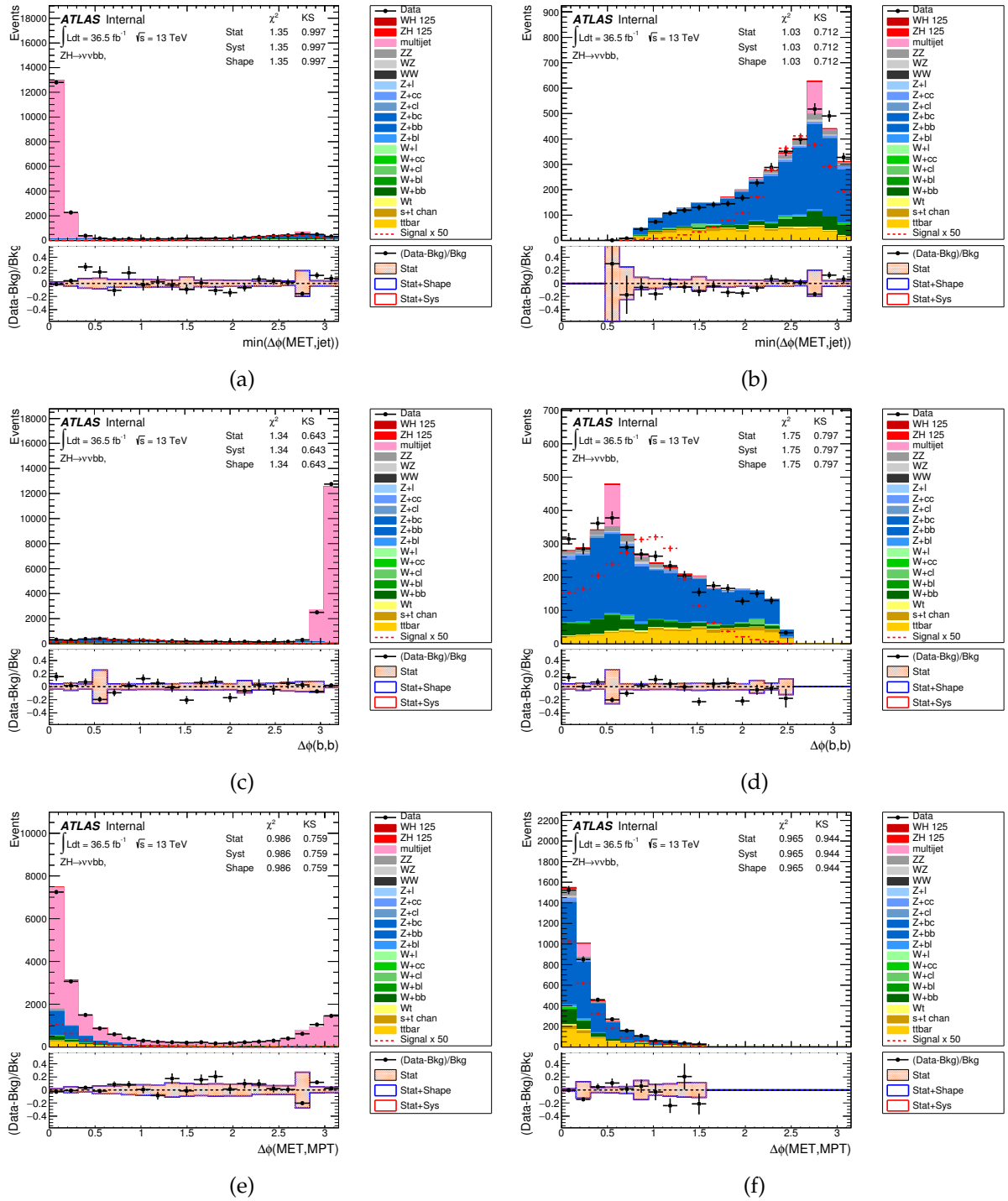
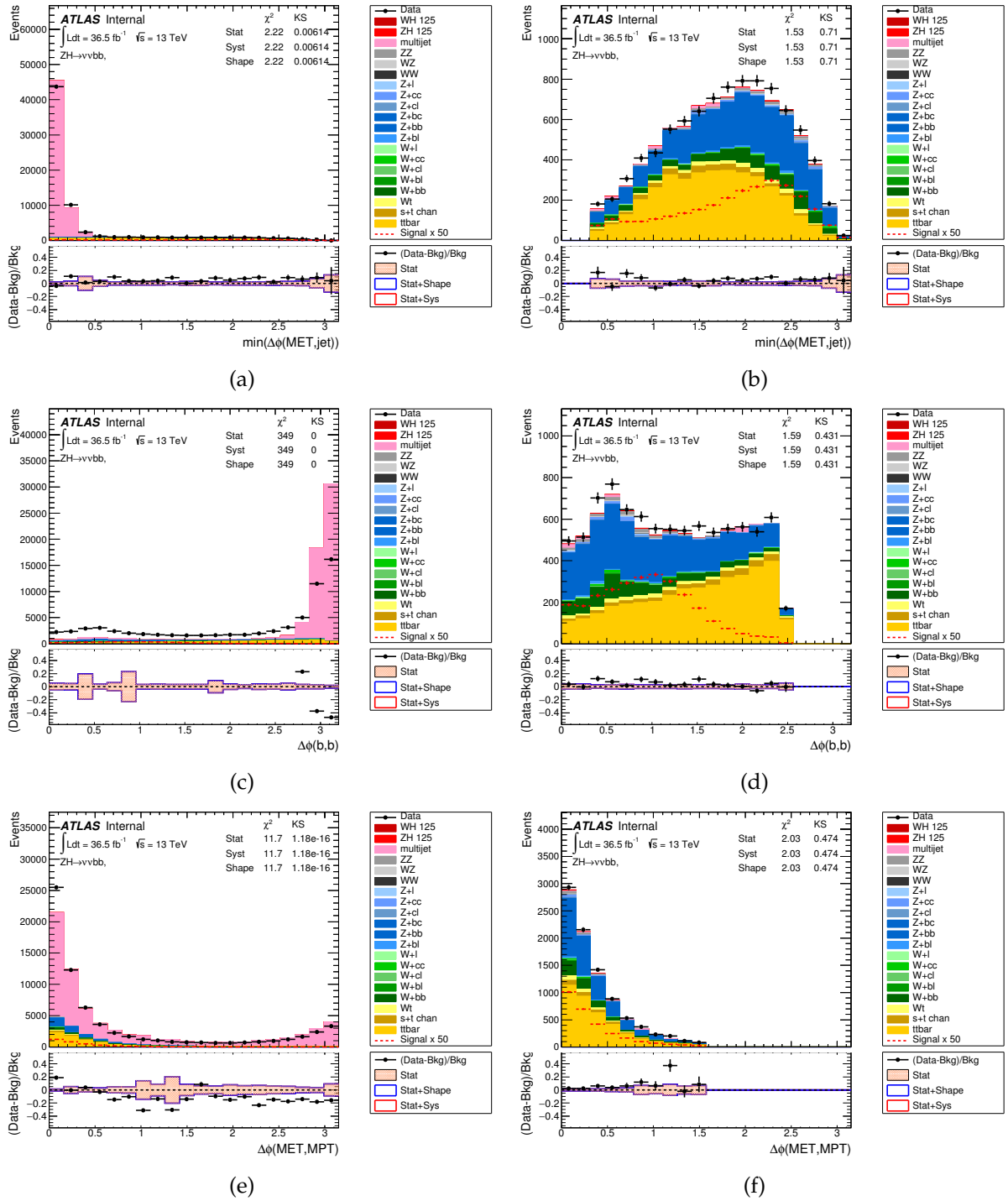


FIGURE 5.8: Distribution of $\min(\Delta\phi(MET, jet))$ in Figures (a) and (b), $|\Delta\phi(jet1, jet2)|$ in Figures (c) and (d), and $|\Delta\phi(MET, MPT)|$ in Figures (e) and (f) in the 2-tag 2-jets region. Anti-QCD cuts are removed in Figures (a), (c), (e) and applied in Figures (b), (d) and (f).



(a) (b) (c) (d) (e) (f)

FIGURE 5.9: Distribution of $\min(\Delta\phi(MET, jet))$ in Figures (a) and (b), $|\Delta\phi(jet1, jet2)|$ in Figures (c) and (d), and $|\Delta\phi(MET, MPT)|$ in Figures (e) and (f) in the 2-tag 3-jets region. Anti-QCD cuts are removed in Figures (a), (c), (e) and applied in Figures (b), (d) and (f).

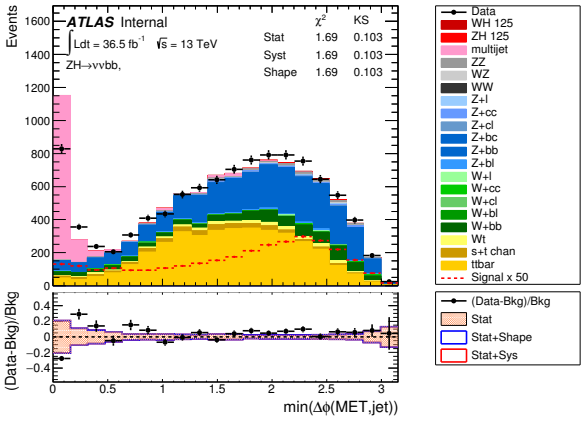


FIGURE 5.10: Distribution of min$\Delta\phi(MET, jet)$ in the 0-lepton channel for events with three jets. The requirement on min$\Delta\phi(MET, jet)$ was removed, hence larger QCD contamination. The distribution of data was fitted with a decreasing exponential shape accounting for the multijet background in order to estimate the contamination in the Signal Region.

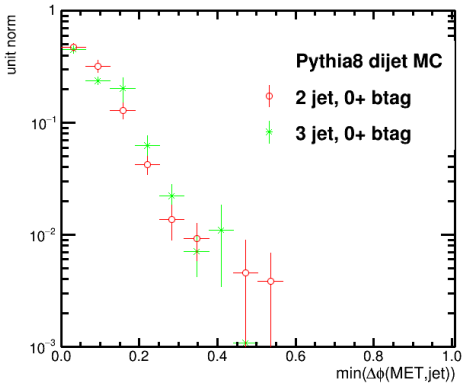


FIGURE 5.11: Distribution of min$\Delta\phi(MET, jet)$ in the 0-lepton channel of the QCD background prediction. The requirement on min$\Delta\phi(MET, jet)$ was removed, hence larger QCD contamination. The shape is decreasing more rapidly than the selected exponential function, hence conservative estimates of the multijet contamination.

dedicated Control Regions in the 2 and 3 jets categories, and in the electron and muon sub-channels individually.

The number of expected multijet events is estimated from Control Regions defined using the standard selection with inverted lepton isolation requirements. To account for the dependency of isolation cuts efficiency on the lepton kinematics, the extrapolation to non overlapping Signal Regions uses event-per-event correction Scale Factors relying on the lepton p_T and η, and on MET in the electron sub-channel. In order to improve the statistical precision of this procedure, the b-Tagging selection is loosened by requiring either one or two b-Tagged jets, instead of two.

In each individual Control Region, a fit to data of the W-candidate transverse mass distribution m_T^W allows to estimate the yield of multijet events. The variable m_T^W was selected for its good discrimination against EW backgrounds. The contribution from EW backgrounds is taken from MC predictions, with free normalizations for the W+jets and top-quark backgrounds. The subtraction of EW backgrounds MC templates also allows to retrieve the shape for discriminant variable distributions in multijet events, used in Signal Regions.

The multi-jet contribution in the 2-jets region is measured to be 4.5% (3%) of the total background contribution in the electron (muon) sub-channel, while in the 3-jets region it is found to be 0.6% (0.4%).

5.4.5.3 2-leptons channel

Multijet events are strongly suppressed in the 2-leptons channel by requiring two isolated leptons with dilepton invariant mass compatible with that of a Z boson. An estimate of the residual contamination is performed by fitting the dilepton mass distribution in same-sign lepton events, including expected Electroweak contributions from Monte-Carlo events and an exponential model for the multijet background. This estimate is propagated to the opposite-sign lepton region based on the assumption that the acceptance for opposite and same sign leptons are identical. The fraction of multijet events within 81 GeV < m_{ll} < 101 GeV and 100 GeV < m_{bb} < 140 GeV is about 8% of the expected signal with no resonant shape. The BDT distribution of such events is similar to that of the remaining backgrounds, and is likely to be absorbed by normalization factors. Thus, its contribution would have a negligible impact for the VH(bb) search and is not included in the fit model.

5.5 Multivariate method

Instead of extracting the VH(bb) signal process from the dijet mass distribution, which is expected to present a resonant excess in data at the Higgs boson mass value, a multivariate technique is performed in order to enhance the sensitivity of the analysis.

5.5.1 Boosted Decisions Trees

Boosted Decision Trees (BDT) are a classification technique which allows to exploit fine correlations between several input variables. A BDT consists of a collection of Decision Trees (DT), each gathering series of cuts on discriminating variables. This multivariate technique allows to classify events either as signal-like or background-like.

The training of a BDT consists of the consecutive training of several Decision Trees. This procedure aims at determining the set of optimal cuts maximizing the separation of Monte-Carlo true-signal against true-background events. Within each Decision Tree, events are divided into signal-like and background-like sub-samples. The $Boosted$ qualifier of a BDT comes from boosting techniques used in order to increase the separation power of a Decision Tree with respect to previously trained Decision Trees. This improvement is achieved by reweighting misclassified events. For instance, a true-signal event classified as background-like in output

of a Decision Tree is given a larger weight in the training of subsequent Decision Trees. Boosting allows the BDT to learn separation patterns of increasing complexity. Finally, each Decision Tree is assigned a weight determined from the purity of classification categories.

The optimisation of cuts in Decision Trees training for the SM VH(bb) search relies on the $Gini$ index :

$$g(c_i) = p(c_i)(1 - p(c_i)) \tag{5.1}$$

Where c_i represents the cut threshold with bin index i on a binned training variable x. The $Gini$ index is based on the purity $p(c_i) = S(c_i)/(S(c_i)+B(c_i))$ for a sample of events filled with $S(c_i)$ true-signal and $B(c_i)$ true-background events.

Each step of the training procedure of a Decision Tree consists of maximizing the separation gain $\Gamma(c_i)$ computed from the difference in $Gini$ index of a parent sample of events, and the $Gini$ indices of the signal-like and background-like sub-samples split with the selection criterion on c_i :

$$\Gamma(c_i) = g_{\text{parent sample}} - f_{\text{signal-like}}(c_i) \cdot g_{\text{signal-like}}(c_i) - f_{\text{background-like}}(c_i) \cdot g_{\text{background-like}}(c_i) \tag{5.2}$$

Where $f_{\text{signal-like}}$ and $f_{\text{background-like}}$ are the fractions of events filled into the signal-like and background-like sub-samples respectively. The variable which maximizes $\Gamma(c_i)$ is selected and the sample of training events is split at c_i^{max} into a signal-like and a background-like sub-samples, as illustrated in Fig. 5.12a. This procedure is repeated recursively for each sub-sample in the limit of parameters aiming at limiting overtraining, which is the specialisation of cuts against statistical fluctuations in training datasets. Each additional cut can be based on a different variable, which allows BDTs to exploit correlations and perform better than cut-based analyses.

The misclassification rate of a signal-like sub-sample corresponds to the ratio of the number of $true$-background events and the total number of events filled into the signal-like sub-sample, and conversely for background-like sub-samples. By construction, this error rate is lower than 0.5 since it is evaluated using the same events than that used to find the optimal cut :

$$err_i = \min(p(c_i), 1 - p(c_i)) \tag{5.3}$$

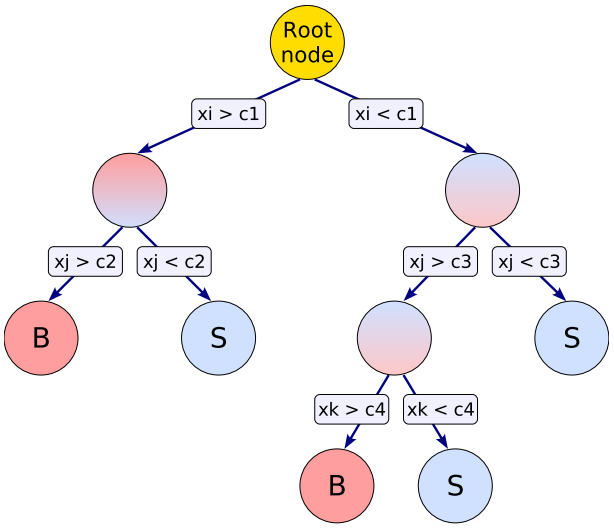
The boost weight α_i assigned to a specific Decision Tree i is computed from the misclassification rate err_i of training events after cut optimisation. The boosting procedure consists of scaling the weight of misclassified events by α_i before training the next Decision Tree :

$$\alpha_i = \frac{1 - err_i}{err_i} \tag{5.4}$$

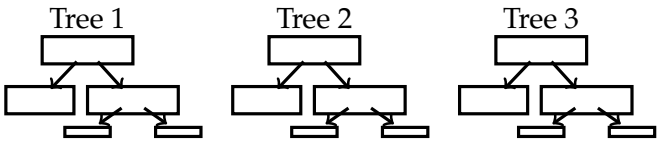
The evaluation of a BDT consists of assigning a BDT score to each event by applying the cuts optimised in the BDT training. Each Decision Tree independently classifies events either as signal-like or background-like, with respective output values $h_i(x)$ = 1 and -1. Decision Trees scores are combined according to their respective weight in order to determine the BDT score $y(x)$ of the event.

$$y(x) = \frac{1}{N_{DT}} \sum_{i=1}^{N_{DT}} ln(\alpha_i) h_i(x) \tag{5.5}$$

In the context of the SM VH(bb) search, dedicated BDTs are trained separately in each analysis category. Each BDT is first trained on Monte-Carlo simulated samples, and then evaluated on data. Each BDT is also evaluated on a different dataset of Monte-Carlo events than that used



(a) Sketch of a Decision Tree



(b) Sketch of a Boosted Decision Tree.

FIGURE 5.12: Illustration of how Boosted Decision Trees work.

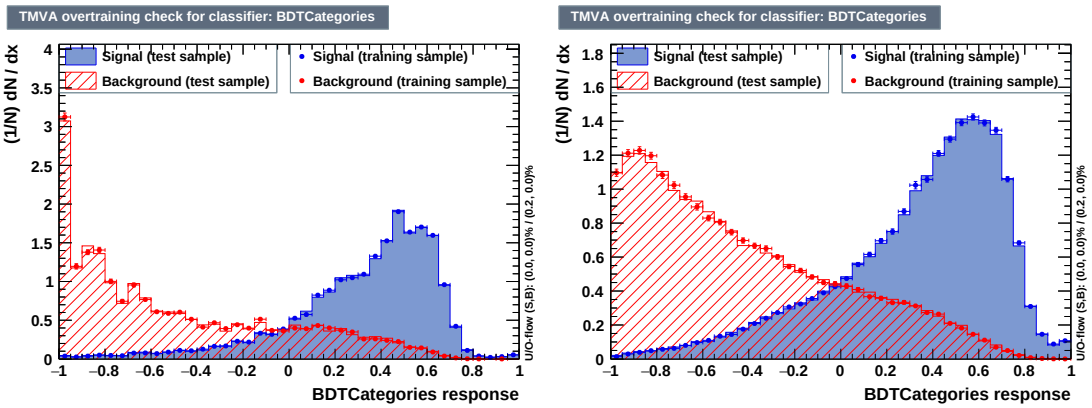


FIGURE 5.13: Validation of BDT overtraining in the 0-lepton channel in the two jets (left) and three jets (right) categories.

for training, following the k-Folding procedure described in Section 5.5.3.1. Signal events are expected to cluster at large MVA values, while background events are mostly rejected to low values, as shown in Figure 5.13.

5.5.2 Parameters

Table 5.10 describes the settings of the BDT used for the SM VH(bb) analysis.

The $AdaBoost$ boosting procedure consists of the reweighting and evaluation strategy described in Section 5.5.1. It comes with the $AdaBoostBeta$ learning rate parameter, denoted β which changes the weight α_i assigned to a Decision Tree i to α_i^β. Pruning methods allow to reduce the number of splitting nodes in a Decision Tree after it was trained to its maximum number of nodes allowed, instead of interrupting the node splitting at an earlier stage. The $Gini$ Index defined in Equation 5.1 is the quantity used to compare the performance of all possible cut values tested in the training procedure. Pruning is based on the observation that apparently insignificant cuts can nevertheless lead to relevant splits further down the tree. Yet, no pruning method is used in the SM VH(bb) search. The maximum tree depth parameter

TMVA Setting	Value	Definition
BoostType	AdaBoost	Boost procedure
AdaBoostBeta	0.15	Learning rate
SeparationType	GiniIndex	Node separation gain
PruneMethod	NoPruning	Pruning method
NTrees	200	Number of trees
MaxDepth	4	Maximum tree depth
nCuts	100	Number of equally spaced cuts tested per variable per node
nEventsMin	5%	Minimum number of events in a node

TABLE 5.10: BDT configuration parameters.

limits the number of subsample layers when splitting events into signal-like and background-like subsamples in the Decision Tree training. The $nCuts$ parameter defines the granularity of training variables binned distributions which are used in order to optimise cuts. These variables are tested in the form of histograms with number of bins $nCuts$. Large $nCuts$ value may largely increase the training time, while low values can lead to information loss. Finally, the $nEventsMin$ parameter define the minimal fraction of events a subsample should contain after applying cuts of a Decision Tree. All of these parameters were optimised in the context of the SM VH(bb) analysis performed at the Run 1 of the LHC using an iterative approach parsing the phase space of parameter values, and showed correct performances at Run 2.

5.5.3 Training strategy

5.5.3.1 k-Folding

k-Folding is a procedure used in order to control the impact of the limited Monte-Carlo statics on BDT performances. It consists of splitting the Monte-Carlo sample in several datasets (named folds) which are reciprocally used for training and testing several BDTs. For example, in the simple case of a two-folds multivariate analysis, Monte-Carlo events are distributed within two folds, f_1 and f_2 :

- BDT$_1$ is trained on events from fold f_2 and evaluated on events from fold f_1.
- BDT$_2$ is trained on events from fold f_1 and evaluated on events from fold f_2.

This technique allows to spot possible overtraining issues. Overtraining plots, such as Fig. 5.13, allow to control that the shapes of the BDT outputs are compatible between the training and testing datasets. In the SM VH(bb) search, events are split into two folds based on the parity of their Event Number value. Monte-Carlo event are associated a unique Event Number and Run Number with no physical meaning. This provides a good variable for training independent BDTs aiming at the same signal process.

5.5.3.2 Truth tagging

Truth tagging is a technique which aims at increasing the Monte-Carlo statistics available for BDT training, by removing the b-Tagging requirement of the selection. Instead, events are reweighted by the appropriate probability to pass the b-Tagging cut, which depends on jets flavour and kinematics. This procedure relies on the assumption that the b-Tagging score of jets does not depend on the topology of the events, so that no bias is introduced with respect to the nominal dataset. Truth tagging is especially useful for samples with mostly light-flavour jets which rarely satisfy the b-Tagging requirements, hence have limited Monte-Carlo statistics.

Category	2 jets	3 jets
VH	0.98	0.99
ttbar	0.91	0.95
Zbb	0.88	0.92
Wbb	0.78	0.85
All bkg	0.90	0.96

TABLE 5.11: Ratios of expected yields when using the direct or truth tagging strategy.

5.5.3.3 The truth/direct tagging discrepancy

The use of truth tagging was found to cause a 5% loss in statistical significance with respect to using direct tagging. This discrepancy was explained by an increase in the expected number of background events from Monte-Carlo predictions, as presented in Table 5.11. Although weights provided by the truth tagging procedure are supposed to conserve the expected number of events, this procedure requires similar kinematic templates in the datasets used for the computation and evaluation of these Scale Factors, which depend on the transverse momentum and pseudo-rapidity of jets. Discrepancies in the kinematic phase space were found to cause this bias. Events passing the VH(bb) selection have jets with lower average transverse momentum value than that used for computing the tagging efficiency maps from which are computed truth tagging Scale Factors. The average b-Tagging efficiency of simulated events in the SM VH(bb) analysis is therefore lower than that computed in b-Tagging efficiency maps.

Yet, no correction was provided to account for this effect. The bias was found to increase the yields of ttbar and V+jets backgrounds. These processes are naturally enhanced in the fitting procedure where their normalization is left free. Hence, this bias should allow to partially account for the underestimated yields of these processes.

5.5.4 Training variables

The nominal set of variables used to train the BDT is presented in Fig. 5.12. Additional variables were tested, as described in Section 5.5.4.1, aiming at a better discrimination of signal against background processes. To the same purpose, pre-processing transformations of these variables were considered, and are presented in Section 5.5.4.2. Further ideas of enhancements are described in Section 5.5.4.3 in order to use the post-fit normalization of the samples coming from the final fit. Last, propects for the use of Pseudo-Continuous tagging in the VH(bb) MVA analysis are shown in Section 5.5.4.4.

5.5.4.1 Testing new variables in the 0-lepton channel

The procedure to optimise the choice of variables for the 0-lepton BDT consists of an iterative process. The training dataset and other settings are identical to those used in the nominal VH(bb) search. The iterative procedure relies on a few observations and assumptions :

- The dijet mass mBB is the most relevant training variable for BDT performance. This can be seen from the frequency for cutting on mBB in the BDT evaluation in Table 5.13, which explains the strong correlation between mBB and the BDT output as illustrated in Fig. 5.14.
- Increasing the number of training variables can only lead to improved BDT performance, within statistical uncertainties.

Variable	Name	0-lep	1-lep	2-lep	Description
p_T^V	pTV		✓	✓	Vector boson pT
MET	MET	✓	✓	✓	Transverse Missing Energy
p_T^{jet1}	pTB1	✓	✓	✓	Leading jet pT
p_T^{jet2}	pTB2	✓	✓	✓	Sub-leading jet pT
m_{jj}	mBB	✓	✓	✓	Dijet mass
$\Delta R(jet_1, jet_2)$	dRBB	✓	✓	✓	Dijet angular separation
$\lvert\Delta\eta(jet_1, jet_2)\rvert$	dEtaBB	✓			Dijet pseudo-rapidity separation
$\Delta\phi(V, H)$	dPhiVBB	✓	✓	✓	Azimutal separation of Vector boson to Higgs
$\Delta\eta(V, H)$	dEtaVBB			✓	Rapidity separation of Vector boson to Higgs
$M_{eff}(M_{eff3})$	HT	✓			Scalar sum of MET and jets pT
$\min(\Delta\phi(\ell, jet))$	dPhiLBmin		✓		Azimutal separation of lepton to closest jet
mTW	mTW		✓		Transverse mass of W boson
m_{ll}	mLL			✓	Dilepton mass
$\Delta Y(W, H)$	dYWH		✓		Rapidity separation of Vector boson to Higgs
m_{top}	mTop		✓		Reconstructed top quark mass
Only in 3 Jet Events					
p_T^{jet3}	pTJ3	✓	✓	✓	Leading non-tagged jet pT
m_{jjj}	mBBJ	✓	✓	✓	Trijet mass

TABLE 5.12: Variables used to train the multivariate discriminant.

Under the assumptions listed above, an iterative procedure was used in order to seek for the optimal set of training variables :

1. Form a set of best training variables $\mathcal{V}_{best-train}$, containing only mBB to begin with.
2. Train as many BDT as there are variables v_{test} to be tested : BDT$\{\mathcal{V}_{best-train}, \text{v}_{test}\}$.
3. Move the tested variable which brings the largest improvement to $\mathcal{V}_{best-train}$. The improvement is measured with respect to the evolution of the statistical significance of the multivariate discriminant distribution following the rebinning treatment presented in Section 5.7.1.
4. Iterate from step 2 until no variable is left to be tested.

Variable	Frequency
mBB	3.59e-01
dRBB	2.16e-01
dEtaBB	1.50e-01
pTB1	7.34e-02
pTB2	7.23e-02
dPhiMETdijet	5.97e-02
HT	4.28e-02
MET	2.71e-02

Variable	Frequency
mBB	3.46e-01
dRBB	1.49e-01
pTB2	8.58e-02
mBBJ	8.34e-02
HT	8.34e-02
MET	7.65e-02
pTJ3	6.92e-02
dPhiMETdijet	4.77e-02
pTB1	3.08e-02
dEtaBB	2.82e-02

TABLE 5.13: Frequency for cutting on training variables in the multivariate discriminant in 2 jets (left) and 3 jets (right).

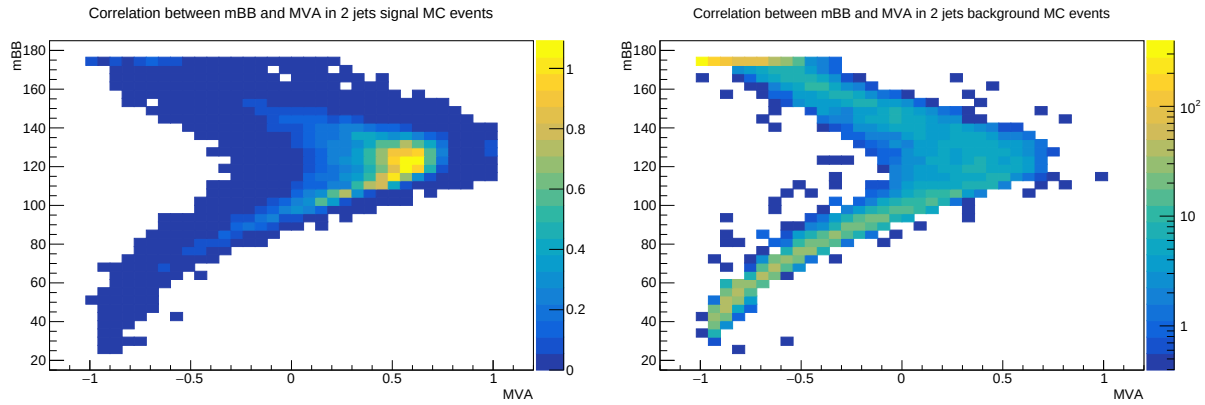


FIGURE 5.14: Correlation of the dijet mass and BDT discriminant in 2 jets simulated signal and background events.

Variable	Name	Description
$\|y_{bb}\|$	yBB	Dijet rapidity
$\|cos(\Theta^*)\|$	abscostheta	Decay angle
$\frac{pT_{B1}-pT_{B2}}{pT_{B1}+pT_{B2}}$	pTBA	Tagged jets pT asymetry
pT_H	pTH	Dijet pT
$\frac{pT_H-MET}{pT_H+MET}$	pTHMETA	MET to Higgs candidate pT asymetry
$MET' = 2 * \frac{m_{BB}}{dRBB}$	METprime	Recomputed MET
Only in 3 Jet Events		
$\Delta R(Bx, J3)$, (x = 1,2)	dRBxJ3	Angular separations in three jets events
$max\Delta R(B, J3), min\Delta R(B, J3)$	maxdRBJ3, mindRBJ3	Extreme angular separations in three jets events
$\|\eta_{J3}\|$	etaJ3	Non-tagged jet pseudo-rapidity

TABLE 5.14: Variables used to train the multivariate discriminant.

In addition to the variables inherited from the run 1 configuration, new variables listed in Table 5.14 were tested. The $|cos(\theta^*)|$ quantity relies on the specific decay topology of scalar particles in their center-of-mass frame. Conversely, $|y_{bb}|$ aims at the production of Higgs bosons in the VH process, which show greater rapidity than most backgrounds. The pTBA variable measures the unbalance between jets transverse momentum. Depending on the Higgs boson rapidity and decay kinematics, jet-candidates for Higgs boson decay products can present similar momentum, contrarily to backgrounds where jets are not produced from a single interaction. The dijet transverse momentum pTH translates the expectation for a boosted Higgs boson candidate in signal events, contrarily to backgrounds where jets are not produced close-by for instance. The unbalance between pTH and MET is measured with pTHMETA. This variable enforces the expectation for opposite vector momentum of the Higgs and vector boson. Hence, similar values of MET and dijet transverse momentum are expected in signal events, within reconstruction and identification effects. The recomputed MET is another definition of MET relying on the measured dijet transverse momentum and angular separation, potentially more resilient to pile-up. In events with three jets, the angular separations of the non b-Tagged jet with respect to b-Tagged jets were tested. The additional jet in signal events is likely to originate from a radiation emitted from b-quarks, hence low angular separation to jet-candidates for Higgs boson decay products.

The results of this iterative procedure are presented for the 2- and 3-jets categories in Fig.

5.15 and Fig. 5.16 respectively. The y-axis represents the tested statistical-only significance in the appropriate n-jets category. The x-axis gives the set of variables used in the BDT training : each x-bin represents a single BDT training, labelled according to the tested variable. Trainings of the same iteration are grouped in a single block. Within each block, tested variables are ranked by increasing BDT performance. The last variable from a bloc shows the best improvement and is added to the baseline set of variables $\mathcal{V}_{best-train}$. This ranking is less relevant in the last iterations, as correlations induce interplay between variables.

Noteworthy observations can be drawn from these results. First, the trainings significance reaches a plateau after very few iterations, which shows that not many training variables are needed in order to achieve optimal BDT performance, within statistical errors. This is especially true in the 2-jets category, where the most performant BDT trained with four variables {mBB, dRBB, pTB2, pTBA} reaches 98% of the overall best BDT significance. In the 3-jets category, eight variables seem to be needed in order to present good performance. Due to the limited size of training and testing datasets, statistical uncertainties are likely to explain the degradation in performance while moving from the most performant BDT of an iteration to the less performant training of the next series.

These results show that none of the new variables brings any gain to the BDT performance. Besides, it motivates arguments for reducing the number of training variables :

- Mismodelled variables may lower the discrimination power of the multivariate discriminant. If the BDT training relies on variables which do not correctly describe data, the set of selected cuts will show to be suboptimal.
- systematic uncertainties can induce important changes in variables value with respect to nominal Monte-Carlo events used for BDTs training. This can lead to underperforming BDTs if some systematics cause large variations to nominal templates in the fit to data.

In order to measure the potential gain in reducing the number of training variables, alternative BDTs were trained with reduced set of training variables. Their respective sensitivity were estimated in the 0-lepton phase space only, and compared to that of the nominal BDT configuration following the standard fit procedure with all systematic uncertainties considered. No gain in sensitivity appeared to be possible from reducing the number of training variables in this rough approach. Yet, more sophisticated procedures such as adversarial multivariate technique may allow to train BDTs so that the systematics information is exploited in the training procedure and its impact minimized. This approach was not tested as it is not included in the TMVA framework, and would require a different environment to be evaluated.

5.5.4.2 Transformation of variables

Various preprocessing treatments of the BDT input variables were explored as a way of optimising the BDT training efficiency.

- Overflow bin : the idea is to avoid very long tails in some of the training variables distributions. Considering a single variable, a few outstanding values would weaken the cuts granularity of the BDT training controled by the parameter nCuts presented in Table 5.10. With this transformation, values larger to a certain threshold are explicitely set to this threshold value.
- Rarity provides a nice alternative to the overflow bin transformation : this method consists of a redefinition of the binning of training variables, so that each bin contains the same fraction of either signal, background, or signal + background (with equal normalizations) events. Especially, if signal distributions present sharp peaks, rarity allows to

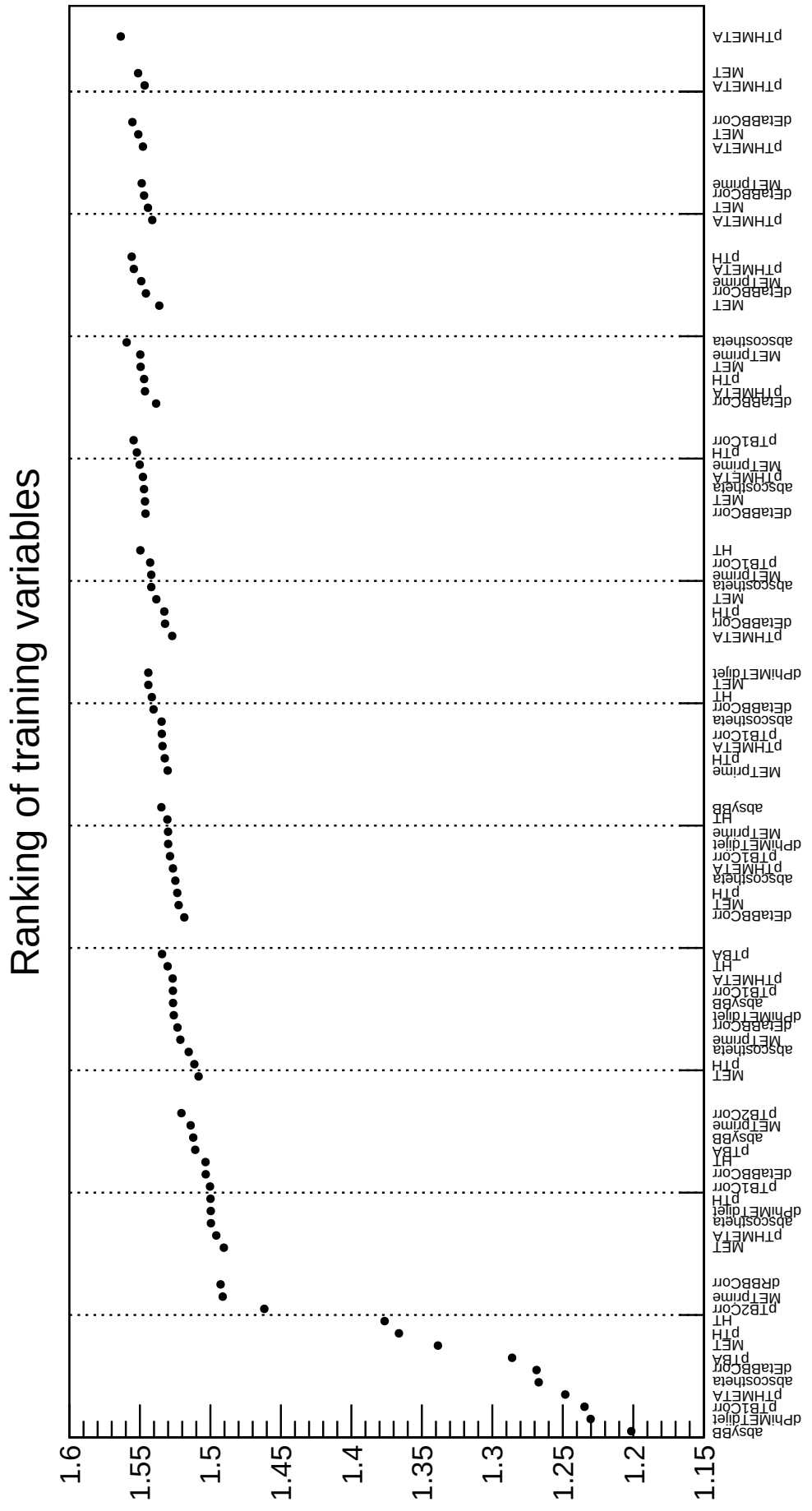


FIGURE 5.15: Variables optimisation in the 0-lepton channel 2 jets category. The order in which variables are picked is : $\mathcal{V}_{best-train}$ = {dRBB, pTB2, pTBA, yBB, dPhiMETdijet, HT, pTB1, abscostheta, pTH, METprime, dEtaBB, MET, pTHMETA}. The corresponding shapes are illustrated in Fig. 5.17, 5.18, 5.19.

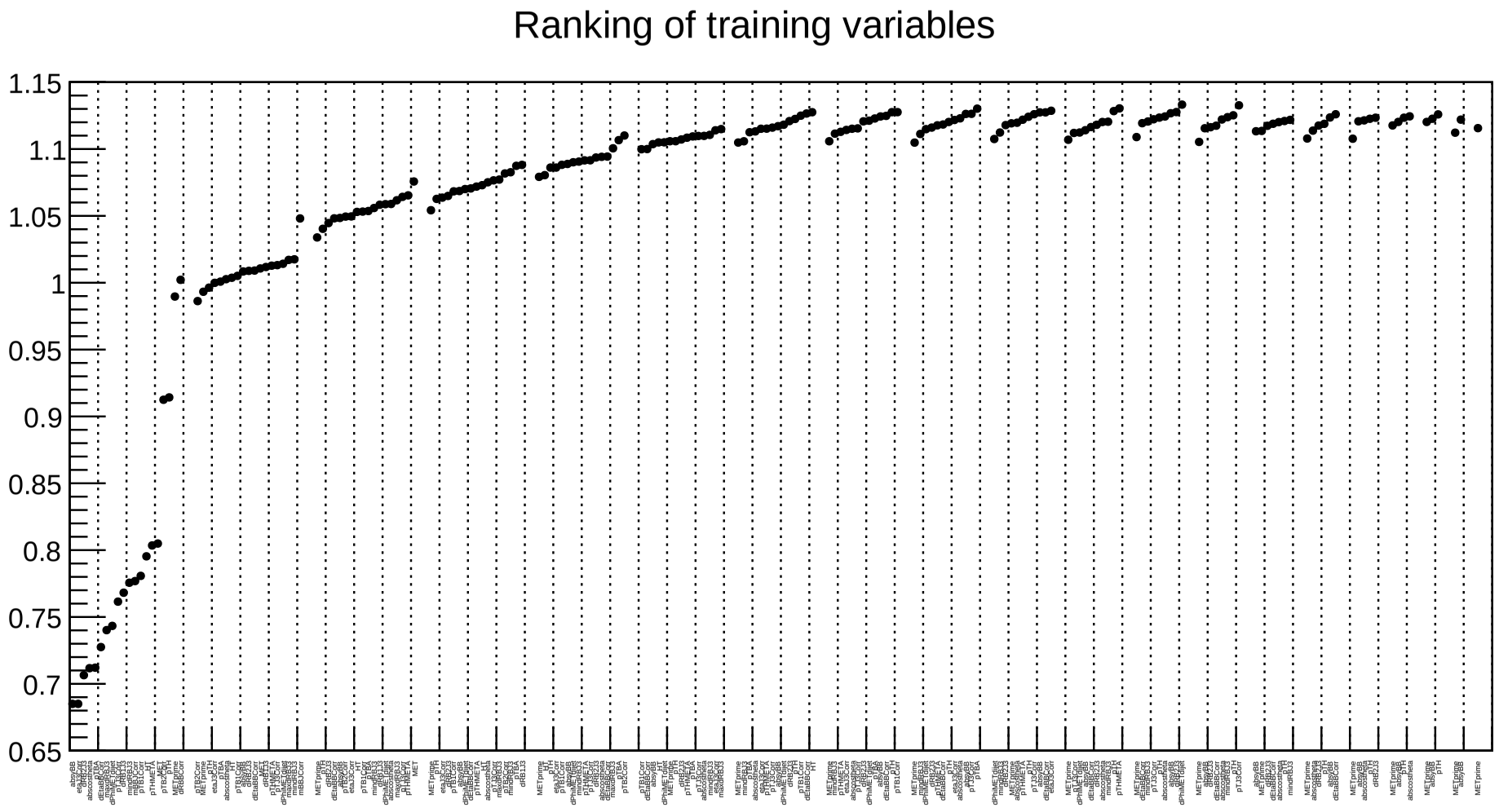


FIGURE 5.16: Variables optimisation in the 0-lepton channel 3 jets category. The order in which variables are picked is : $\mathcal{V}_{best-train}$ = {dRBB, mBBJ, MET, dRB1J3, pTB2, maxdRBJ3, HT, pTB1, pTBA, etaJ3, pTHMETA, dPhiMETdijet, mindRBJ3, dEtaBB, dRB2J3, costtheta, pTH, yBB, MET'}. The corresponding shapes are illustrated in Fig. 5.20, 5.21, 5.22, 5.23.

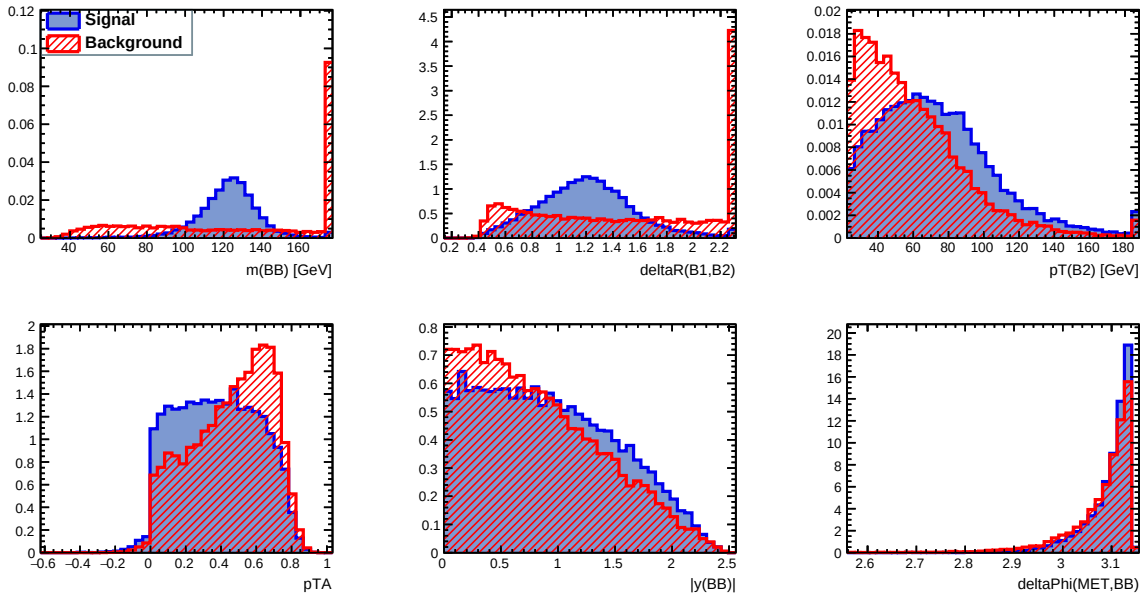


FIGURE 5.17: Shape of the m_{bb}, $\Delta R(b,b)$, pT(b2), pTA, |y(bb)| and $\Delta\phi$(MET,bb) variables in the 0-lepton channel 2 jets category.

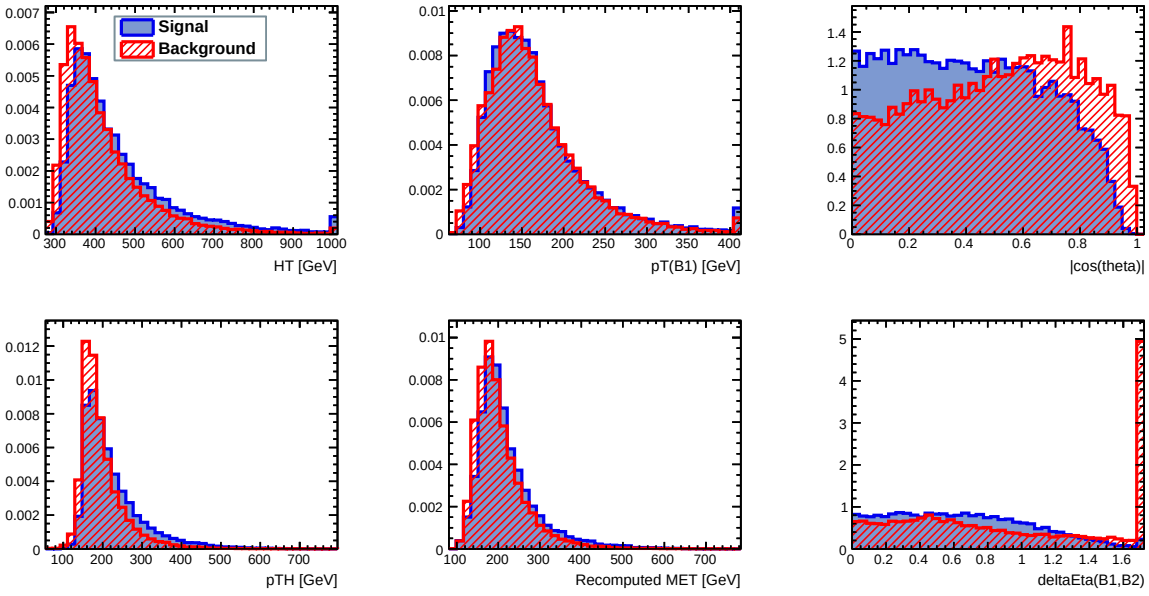


FIGURE 5.18: Shape of the HT, pT(b1), $|\cos(\Theta^*)|$, pTH, recomputed MET and $\Delta\eta$(b1,b2) variables in the 0-lepton channel 2 jets category.

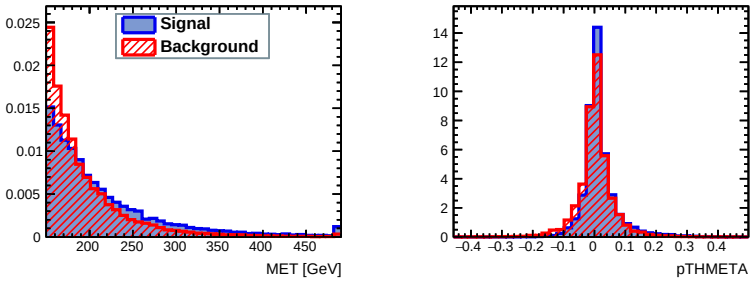


FIGURE 5.19: Shape of the MET and pTHMETA variables in the 0-lepton channel 2 jets category.

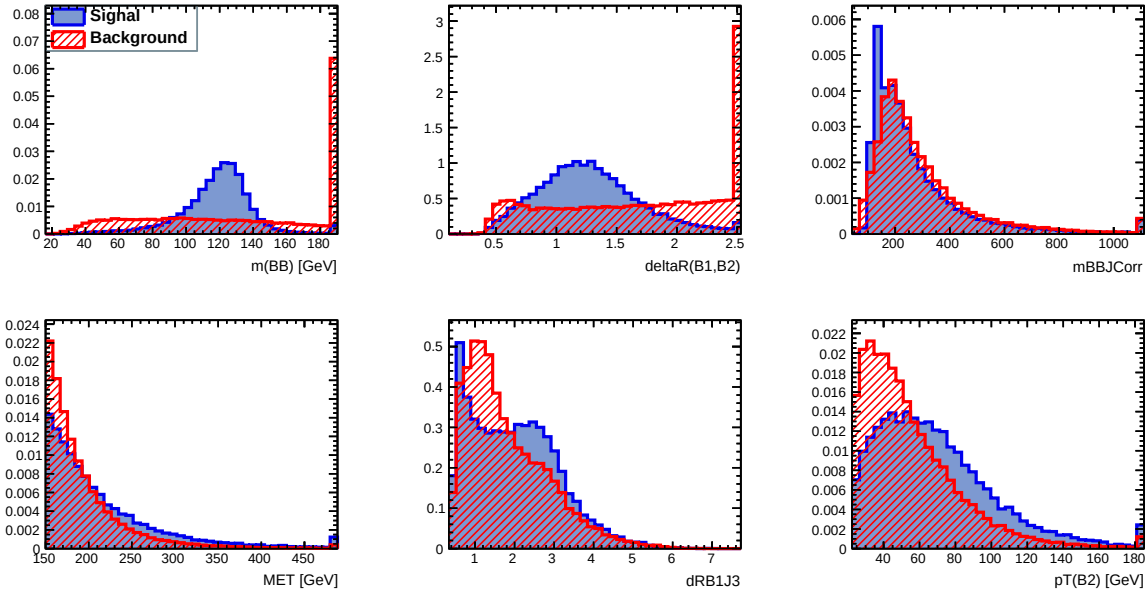


FIGURE 5.20: Shape of the m_{bb}, $\Delta R(b,b)$, m_{bbj}, MET, $\Delta R(b1,j3)$ and pT(b2) variables in the 0-lepton channel 3 jets category.

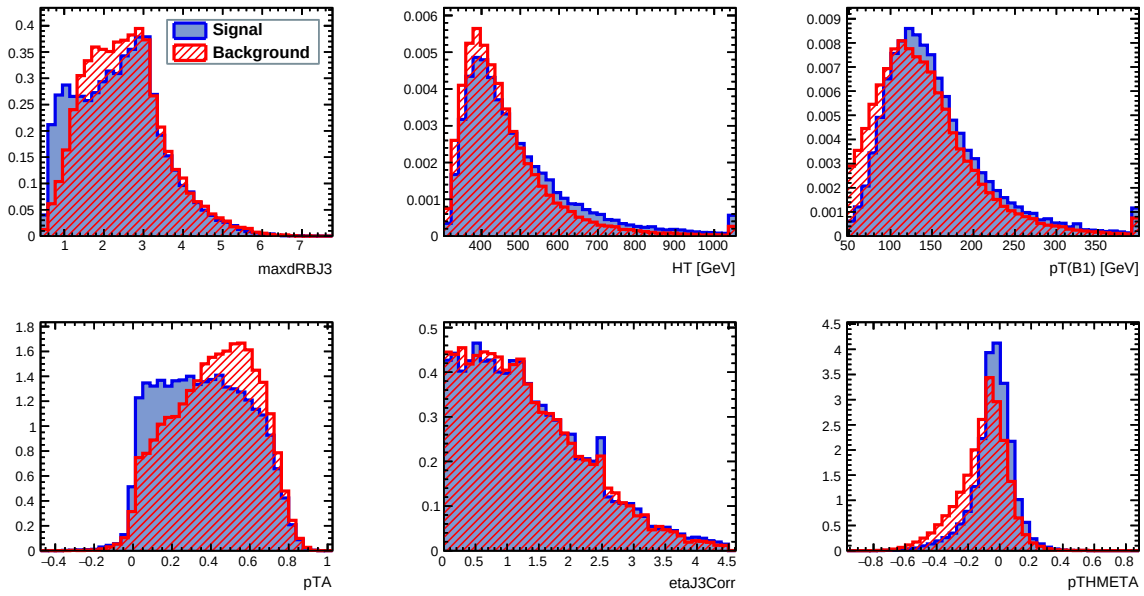


FIGURE 5.21: Shape of the max($\Delta R(b, j3)$), HT, pT(b1), pTA, $|\eta_{j3}|$ and pTHMETA variables in the 0-lepton channel 3 jets category.

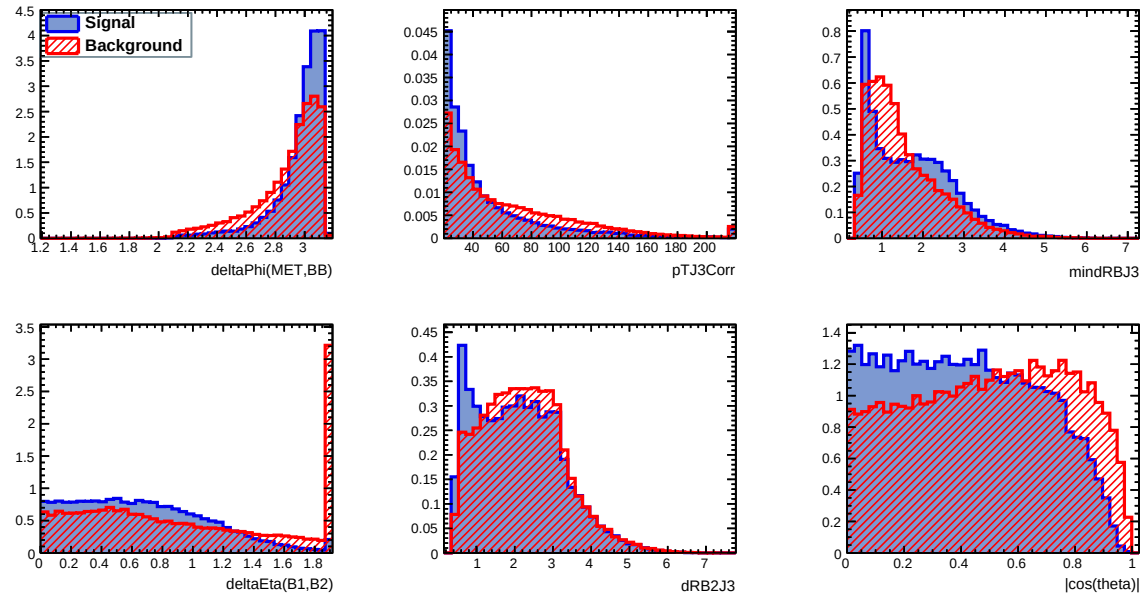


FIGURE 5.22: Shape of the $\Delta\phi$(MET,bb), pT(j3), min($\Delta R(b, j3)$), $\Delta\eta(b1, b2)$, $\Delta R(b2, j3)$ and $|\cos(\Theta^*)|$ variables in the 0-lepton channel 3 jets category.

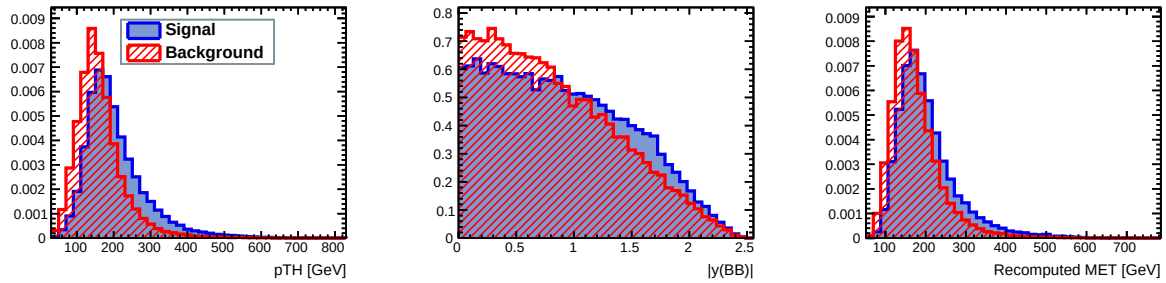


FIGURE 5.23: Shape of the pTH, $|y(bb)|$ and recomputed MET variables in the 0-lepton channel 3 jets category.

exploit fine structures and correlations. Similarly, it allows to avoid long tails especially in background samples.

- Decorrelation : this transformation was tested although no large improvement was expected in the context of the SM VH(bb) analysis, since BDTs naturally exploit correlations between variables. Decorrelation consists of re-defining variables values after diagonalising their correlation matrix, which removes linear correlations.

Negligible improvements were achieved from such transformations. Hence, no motivation for supporting complex transformations was experienced. Simple overflow bins were defined for kinematic variables with long tails.

5.5.4.3 Process normalization scaling

Rescaling the respective yield of Monte-Carlo background samples prior to the BDT training is physically motivated by the result of fits to data with fixed signal-strength μ=1, which shows that the normalization of V + heavy flavour jet (Zbb, Wbb) and $t\bar{t}$ processes need to be scaled in order for the Monte-Carlo prediction to fit to data. Because of the different topologies of these processes, the optimal set of cuts of a BDT may undergo large variations while changing their relative normalization. Hence, training the BDT using appropriate yields for each background process is a possible way to improve the fitted sensitivity of the analysis. This procedure was tested and showed a 1.8% loss in significance, within statistical and systematic uncertainties. A possible conclusion is that the nominal BDT training already produces cuts resilient to changes in the normalization of the main backgrounds to some extent. This can be due to constraints limiting the BDT overtraining, which also prevent the training from a specialisation of the BDT on discrimination patterns exclusive to specific background processes. Yet, this small decrease in performances is not significant given the statistical and systematic uncertainties.

5.5.4.4 BDT enhancement from Continuous *b*-Tagging

A possible improvement to the BDT consists of exploiting the binned information of jets b-Tagging scores as input variables to BDT. Improvements of the order of 15% on the statistical significance were observed using preliminary continuous-tagging calibrations. Yet, this improvement is expected to be mitigated once using systematics in the fitting procedure.

5.6 Systematic uncertainties

Systematic uncertainties refer to experimental and modelling sources of error for the central value of the measured signal-strength. The impact of these uncertainties is evaluated as variations on the signal strength. Systematic uncertainties considered for the final result are treated as uncorrelated Nuisance Parameters (NP) in the fit.

5.6.1 Theory uncertainties

Theory systematics are considered for signal production due to cross-section uncertainties from PDF and scale variations, and Branching ratio uncertainties. Additional Nuisance Parameters encode shape variations of the vector boson transverse momentum and dijet mass distributions. The impact of the dominant systematic on signal acceptance from Parton Shower models was investigated in Section 6.1.2. The uncertainty from acceptance effects is also estimated from scale and PDF+α_S variations. The impact of Electroweak corrections on the modelling of the vector boson transverse momentum is considered. Also, uncertainties in the signal cross-section are computed to estimate the impact of unaccounted higher-order EW corrections. In overall, theory systematics have a small impact on the significance.

5.6.2 Experimental systematics

Experimental uncertainties entering the SM VH(bb) search are :

- Luminosity : the normalization of Monte-Carlo samples relies on measures of the integrated luminosity collected by the ATLAS detector, which is L = 36.1 $\pm$ 1.2 fb^{-1}.
- Trigger : uncertainties on the trigger efficiency.
- Objects reconstruction, identification and energy measurement.
- b-Tagging : b-Tagging scale factors are estimated by comparing tagging efficiencies in Monte-Carlo and data samples, and are impacted by modelling and experimental uncertainties. An appropriate treatment would allow to correlate them with analysis uncertainties of similar nature (e.g. correlate b-Tagging modelling systematics with SM VH(bb) search modelling systematics, etc.), but this approach is not available in the current framework.

5.6.3 Modelling systematics

Modelling systematics collectively refer to variations in exploiting different simulation models and tunes. The principal sources of modelling systematics come from uncertainties in Matrix Element generation, and Parton Shower simulation. The comparison of Matrix Element generators and Parton shower algorithms is facilitated by the unified Matrix Element events format output, used as input to Parton Shower simulations. This allows to uncorrelate the impact of Parton Shower from Matrix Element models, and reciprocally. For the main backgrounds of the analysis, the shape uncertainties on m_{bb} and p_T^V are defined from data to Monte-Carlo comparisons in dedicated regions. Otherwise, comparisons between different generators are used.

Because of practical considerations, different generators are used to generate nominal samples of different processes. Yet, this should not impact physics results since alternative generators and their tunes are picked in order to estimate the enveloppe uncertainty for modelling.

5.7 Statistical treatment

5.7.1 Statistical rebinning

A procedure is applied to prevent physics results from depending on the binning of distributions. This treatment was designed because of the limited Monte-Carlo statistics. Especially, multivariate discriminants are trained in order to achieve an optimal signal to background separation, as described in Section 5.5. Hence, a rebinning transformation is applied to sensitive multivariate discriminant distributions [73]. The procedure intends to preserve a good signal purity in sensitive bins, in compromise to reasonnable statistical uncertainties. The rebinned distribution is more resilient to statistical fluctuations in Monte-Carlo and data distributions, and to the evaluation of systematic uncertainties. The transformation relies on the rebinning function which depends on variables :

$$Z(I[k,l]) = Z(z_s, n_s(I[k,l]), N_s, z_b, n_b(I[k,l]), N_b) \tag{5.6}$$

With :

- I[k, l] : bins interval from bin k to l.
- n_s(I[k, l]) : number of signal events in bins I[k, l].
- n_b(I[k, l]) : number of background events in bins I[k, l] (all backgrounds merged).
- N_s : total number of signal events.
- N_b : total number of background events (all backgrounds merged).
- z_s : Rebinning parameter for signal events.
- z_b : Rebinning parameter for background events.

The definition of the rebinning threshold function is :

$$Z = z_s n_s/N_s + z_b n_b/N_b \tag{5.7}$$

The rebinning alorithm consists of an iterative procedure starting from the most sensitive bins, with highest BDT scores :

1. Consider I[k, l] as new bin candidate :
   a) At the very first iteration, k = l = N_{bins}.
   b) If a new bin was just defined, consider k = l as the last unmerged bin.
   c) Otherwise, increase the size of the interval undergoing the optimisation by merging to it the bin of the multivariate discriminant just to the left of I[k, l] : k = k-1. One should have k $\neq$ l and l = last unmerged bin.
2. Evaluate the rebinning function on I[k, l] :
   a) If Z > 1, bins from indices k to l are merged within a new single bin which is ignored in the next rebinning iterations. Then, restart at step 1. b).
   b) If Z < 1, restart at step 1. c).

5.7.2 Pruning

Due to the important number of processes, categories and systematic sources, a very large number of histograms should be processed in the fit. Therefore, systematic uncertainties which lead to very small variations with respect to the nominal distributions of the multivariate discriminant are pruned away in order to accelerate the fit. For each individual sample, the pruning procedure relies on overall and bin-per-bin comparisons of varied and nominal histograms in each region :

- The conditions for pruning a normalization systematic are either :
  - the variation is lower than 5%.
  - the $\pm\ 1\sigma$ variations lead to same sign changes in the normalization.
- The conditions for pruning a shape systematic are either :
  - all bins of the $\pm\ 1\sigma$ varied distributions present less than 0.5% yield difference with respect to the nominal distribution.
  - only one of the $\pm\ 1\sigma$ variations is non-zero.

Also, both normalization and shape systematic uncertainties of a Monte-Carlo sample are pruned away in a given region if the contribution for the corresponding process is less than 2% of the total background, within additional requirements regarding signal Monte-Carlo templates.

5.7.3 Smoothing

Systematic uncertainties which may induce changes in the measurements of selected events, such as JES, can introduce additional statistical fluctuations affecting the template of systematic variations. Hence, a smoothing procedure is applied to the shape of varied distributions in each region, which allows to reduce the impact of such fluctuations.

5.7.4 Fit Model

The parameter of interest measured in the SM VH(bb) search consists of the signal strength μ for observing VH(bb) events with respect to the Standard Model expectation :

$$\mu = \frac{\sigma_{observed} \times Br(\mathrm{H} \to \mathrm{b\bar{b}})}{\sigma_{SMexpected} \times Br(\mathrm{H} \to \mathrm{b\bar{b}})} \tag{5.8}$$

The signal strength is measured by fitting a binned likelihood function to multivariate discriminant distributions of data events collected with the ATLAS detector. This section describes the construction of the likelihood function, and how it is used in order to assess the confidence of the analysis for observing the VH(bb) process.

5.7.4.1 Systematics-free likelihood function

The likelihood model is built from the assumption that within each bin i of multivariate distributions, the probability for observing n_i data events follows a Poisson-law with expectation Monte-Carlo prediction within signal strength hypothesis μ : $\mu\ \mathrm{s}_i$ + b_i. In a systematic-free model, the likelihood function reads :

$$\mathcal{L}(\mu) = \prod_{i \in bins} Pois(n_i | \mu s_i + b_i) = \prod_{i \in bins} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} \tag{5.9}$$

The Maximum-Likelihood Estimate (MLE) of $\hat{\mu}$ is the value $\mathcal{L}(\hat{\mu})$ for which the likelihood function has its global maximum with respect to μ. It allows to retrieve the best estimator for the signal-strength value, denoted as $\hat{\mu}$. This approach relies on the fact that appropriate probability-density function and parameters of the likelihood function will lead to larger likelihood-estimates than for wrong parameter values.

5.7.4.2 Systematics Nuisance Parameters

Particle physics experiments generally include many systematic uncertainties, for instance from differences in theoretical models or potential biases in experimental measurements. These are accounted for in the likelihood model by Nuisance Parameters (NPs) θ and resolution σ_θ initially estimated through auxiliary measurements. The likelihood-function penalty term for shifting a NP value from its auxiliary measurement is defined usually as a gaussian term which multiplies the systematic-free likelihood function 5.9 :

$$\mathcal{L}_{Aux}(\theta) = \frac{1}{\sigma_\theta\sqrt{2\pi}} e^{-\frac{(\theta-\theta_{Aux})^2}{2\sigma_\theta^2}} \tag{5.10}$$

Nuisance Parameters are built so that their impact on fitted distributions affects either the shape or the normalization of simulation inputs. The Maximum-Likelihood estimator for shape Nuisance Parameters is computed from interpolations and extrapolations of the one sigma up and down variations of the systematic uncertainty with respect to the nominal distribution. Normalization Nuisance Parameters are defined in order to balance the yields between different analysis categories (lepton channel, jet multiplicity, Control to Signal Region) or amongst specific background processes.

Also, free floating normalization factors are defined for the main background processes. These later parameters do not penalize the likelihood function, and are completely determined from the fit to data.

5.7.4.3 Monte-Carlo statistics Nuisance Parameters

Monte-Carlo simulated samples have an intrinsic statistical error due to the finite number of events generated. Hence, the normalization of the overall background template is scaled by per-bin γ-factors with Poisson probability terms for the available statistics in each bin. In this approach, the systematic-free likelihood function 5.9 transforms to :

$$\mathcal{L}(\mu, \gamma_i) = \prod_{i\in bins} Pois(n_i|\mu s_i + \gamma_i b_i) \prod_{i\in bins} Pois(m_i|\gamma_i\tau_i) \tag{5.11}$$

In the latter equation, m_i and τ_i express the equivalent number of unweighted Monte-Carlo events for background processes. They are computed as $(b_i/\delta b_i)^2$ where δb_i is the statistical uncertainty on the weighted bin content b_i. The product of Poisson terms transcribes the probability law for a given statistical repartition of unweighted events amongst bins of the discriminant distribution. Because of practical limitations, no γ-stat parameters are defined for signal samples, although it would be needed in a complete model. Yet, this approximation is estimated to have negligible impact because of the low statistical uncertainty for the available signal samples.

An even more complete model shall account for the intrinsic statistical uncertainty of individual Monte-Carlo datasets, which are each generated to cover well defined physics-processes. This method would require much more computing ressources, hence the inclusive γ-stat approach. The effect of this approximation is supposedly absorbed by other Nuisance Parameters and normalization factors.

5.7.5 Test statistic

In order to test a certain hypothesis for the signal strength value μ, the $test\ statistic$ is defined from the difference of log-likelihood values :

$$q_\mu = -2\ \ln(\frac{\mathcal{L}(\mu, \hat{\hat{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}) \tag{5.12}$$

Where $\hat{\mu}$ and $\hat{\theta}$ are the estimators value that maximize the likelihood function. Conversely, $\hat{\hat{\theta}}_\mu$ defines the best Nuisance Parameters estimator value for a fixed signal strength hypothesis μ. Especially, in the context of a discovery where the background-only hypothesis $\mu = 0$ is tested, the previous formula of the test statistic is refined :

- if $\hat{\mu} > 0 : q_0 = -2\ \ln(\mathcal{L}(0, \hat{\hat{\theta}}_0)/\mathcal{L}(\hat{\mu}, \hat{\theta}))$
- if $\hat{\mu} < 0 : q_0 = 0$

The result of this profiling method provides a test statistic value which allows different interpretations of the fit result :

- $q_\mu \gg 1$ characterizes measurements with important disagreement of the fixed μ hypothesis with respect to the data measurement $\hat{\mu}$.
- $q_\mu \gtrapprox 0$ characterizes measurements with very good agreement of the fixed μ hypothesis with respect to the data measurement $\hat{\mu}$.

The test statistic allows to estimate the level of disagreement for a specific parameter of interest hypothesis μ with respect to its best measured value $\hat{\mu}$, quantified by the p-value as :

$$p_\mu = \int_{q_\mu}^{\infty} f(q'_\mu|\mu)\, \mathrm{d}q'_\mu \tag{5.13}$$

Where q_μ is fixed to the measured test-statistic value for the tested hypothesis μ, and $f(q'_\mu|\mu)$ defines the probability density function for all possible test-statistic values q'_μ. The shape of $f(q'_\mu|\mu)$ is not extracted from the distribution of q_μ, which depends on experimental data. In the limit of large datasets, $f(q'_\mu|\mu)$ behaves as a non-central chi-square function [74] [75], assuming that $\hat{\mu}$ is gaussian distributed around μ'. The non-centrality term cancels in the special case $\mu' = \mu$, which further simplifies the estimate of $f(q_0|0)$. The validation of this statistical model is performed by fitting the Monte-Carlo predictions for all samples superimposed, denoted as Asimov dataset. The respective normalizations of signal and backgrounds processes are known, which allows to ensure the reliability of the procedure.

Given $\hat{\mu}$, an ensemble of toy experiments consists of repeating multiple times this measurement with alternative data : p_μ quantifies the fraction of alternative experiments which would observe a worse data to Monte-Carlo agreement, with respect to the likelihood fit value. Small values of p_μ indicate low probability for rejecting the μ hypothesis if it is true. In the frame of physics searches aiming at the confirmation for a certain signal, the p_0 value is particularly relevant for rejecting the background-only ($\mu = 0$) hypothesis.

For discoveries, the p-value is most often expressed in standard deviations with respect to the normal standard distribution, namely significance :

$$Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0} \tag{5.14}$$

Where Φ^{-1} is the inverse cumulative Gaussian distribution centered at zero with unit standard deviation. Conventionally, Z > 5σ is required for claiming a discovery (p_0 = 2.87 10^{-7}), whereas Z > 3σ allows to present an evidence for a signal.

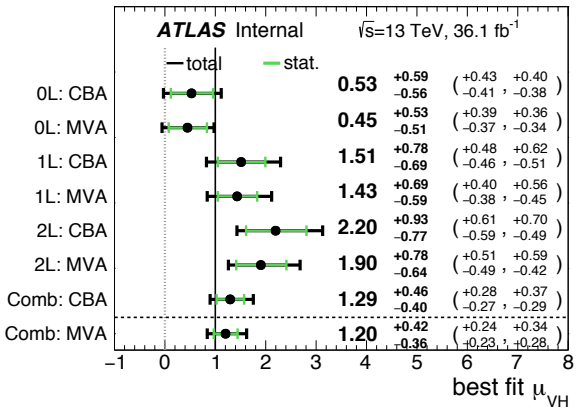


FIGURE 5.24: The μ-value measured in the cut-based and MVA analyses combined fit to data is compared to the values obtained when running the combined fit with the signal strength floated separately for each of the leptonic channels.

5.8 Results

This section presents the results published with data collected in 2015 and 2016 of the Run 2 of the LHC, and their combination with the SM VH(bb) search performed with Run 1 data.

5.8.1 Run 2

At the Run 2 of the LHC, the search for the SM VH(bb) process was performed using a multivariate discriminant [71]. As described in the analysis overview 5.1.3, the multivariate analysis procedure was first validated through a diboson search. Also, a cut-based analysis aiming at VH events was used as cross-check for the SM VH(bb) multivariate analysis.

5.8.1.1 Diboson search

For the diboson search aiming at VZ(bb) events, the signal strength is measured to be :

$$\hat{\mu}_{VZ} = 1.11^{+0.12}_{-0.11}(stat.)^{+0.22}_{-0.19}(syst.) \tag{5.15}$$

This result corresponds to an observed significance of 5.8 standard deviations for the VZ(bb) process. The expected significance for this analysis amounts to 5.3 standard deviations. The measured signal strength is compatible with the SM prediction ($\mu_{VZ} = 1.0$).

5.8.1.2 VH(bb) cut-based analysis

Aiming at VH(bb) events as signal process, the cut-based analysis relies on a fit to the dijet-mass distribution. The fit combined across channels gives a signal-strength of :

$$\hat{\mu}_{VH} = 1.29^{+0.28}_{-0.27}(stat.)^{+0.37}_{-0.29}(syst) \tag{5.16}$$

The corresponding observed significance is 3.5 standard deviations, in comparison to an expected significance of 2.8 standard deviations. Figure 5.25 presents the distribution of m_{bb} in data after subtraction of all backgrounds except for the diboson processes, as obtained with the dijet-mass analysis.

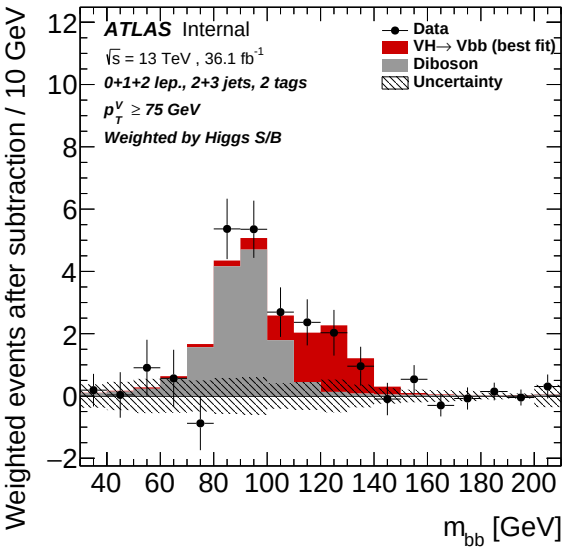


FIGURE 5.25: Distribution of m_{bb} in data after subtraction of all backgrounds except for the diboson processes, as obtained with the dijet-mass analysis. The contributions from all lepton channels, pTV intervals and number-of-jets categories are summed weighted by their respective value of the ratio of fitted Higgs boson signal and background. The expected contribution of the associated WH and ZH production of a SM Higgs boson with m_H = 125 GeV is shown scaled by the measured combined signal strength ($\hat{\mu}$ = 1.29). The size of the combined statistical and systematic uncertainty for the fitted background is indicated by the hatched band [71].

5.8.1.3 VH(bb) multivariate analysis

The multivariate analysis targeting at VH(bb) events in the SM shows a p_0 value of 0.019% for observing in a background-only hypothesis a result at least as signal-like for m_H = 125 GeV. This corresponds to an observed significance of 3.5 standard deviations, to be compared to a 3.0 standard deviations expected significance. The fitted signal strength is :

$$\hat{\mu}_{VH} = 1.20^{+0.24}_{-0.23}(stat.)^{+0.34}_{-0.28}(syst) \tag{5.17}$$

Figure 5.24 presents a comparison of the fitted signal strength in combined and individual lepton channels fits. This result presents a good compatibility with respect to the measurement of the cut-based analysis.

A ranking of the Nuisance Parameters with largest impact on the signal strength uncertainty is presented in Fig. 5.26. Systematic uncertainty sources are grouped into categories in Table 5.15. This breakdown emphasizes the impact from modelling and b-Tagging systematics. Last, the scale factors applied to the normalization of the main backgrounds are written in Fig. 5.16. The latter scale factor values present a very good agreement with that fitted in the diboson analysis.

Post-fit distributions of the multivariate discriminant in the 0-, 1- and 2-lepton(s) channels are presented in Fig. 5.27, 5.28 and 5.29 respectively. These account for the fitted signal strength and values of normalization and shape Nuisance Parameters. The post-fit distributions of variables fitted in the 1-lepton and 2-leptons Control Regions are shown in Fig. 5.28 and 5.30 respectively. These distributions correspond to the fitted yield values presented in Figure 5.17.

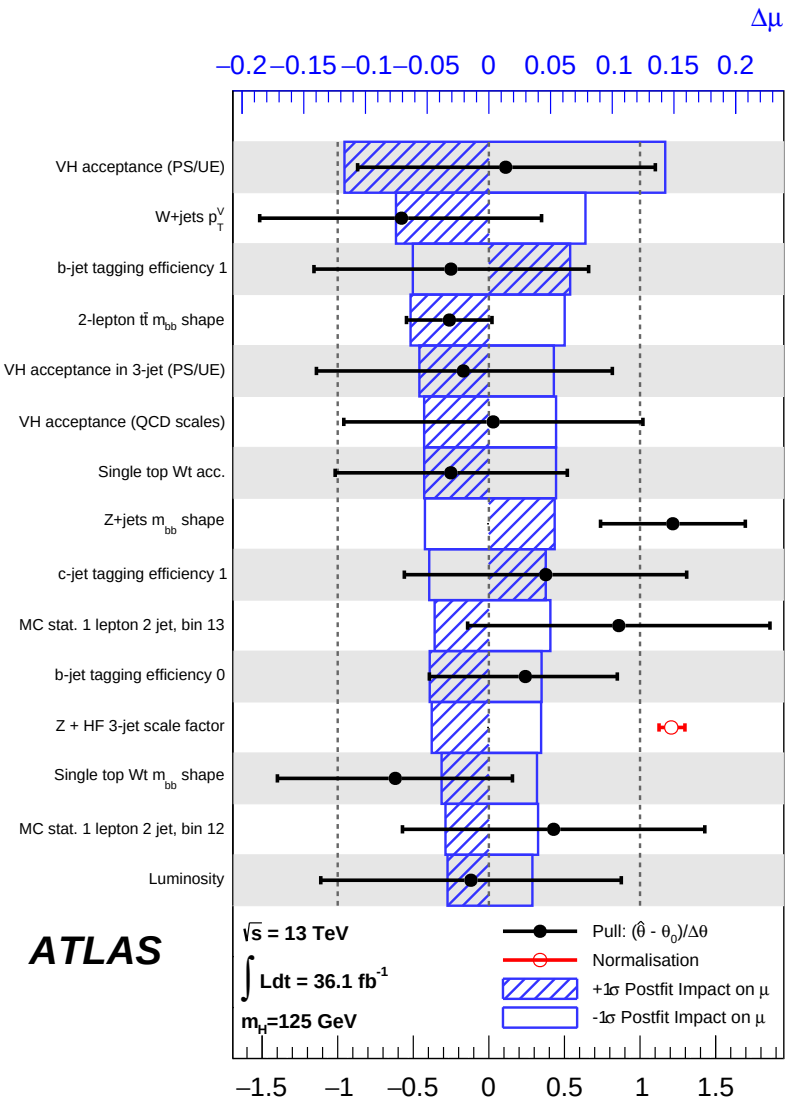


FIGURE 5.26: Impact of systematic uncertainties on the fitted Higgs boson signal-strength parameter μ for the nominal MVA analysis applied to the 13 TeV data. The systematic uncertainties are listed in decreasing order of their impact on μ. The boxes show the variations of μ, refering to the top x-axis, when fixing the corresponding individual Nuisance Parameter θ to its fitted value θ modified upwards or downwards by its fitted uncertainty, and performing the fit again, with all the other parameters allowed to vary, so as to take correlations between systematic uncertainties properly into account. The hatched and open areas correspond to the upwards and downwards variations, respectively. The filled circles, referring to the bottom x-axis, show the deviations of the fitted Nuisance Parameters θ from their nominal values θ_0, expressed in terms of standard deviations with respect to their nominal uncertainties $\Delta\theta$. The associated error bars show the fitted uncertainties of the Nuisance Parameters, relative to their nominal uncertainties. The open circle with its error bar, also referring to the bottom x-axis, shows the fitted value and uncertainty of the Z + HF normalisation parameter in the 3-jet categories that is freely floating in the fit. This normalisation parameter has a nominal value of one. As explained in section 7.1, the b-tagging uncertainties are decomposed into uncorrelated components; the labels 0 and 1 refer to the leading and second-leading components [71].

Source of uncertainty		σ_μ
Total		0.39
Statistical		0.24
Systematic		0.31
Experimental uncertainties		
Jets		0.03
$E_{\mathrm{T}}^{\mathrm{miss}}$		0.03
Leptons		0.01
b-tagging	b-jets	0.09
	c-jets	0.04
	light jets	0.04
	extrapolation	0.01
Pile-up		0.01
Luminosity		0.04
Theoretical and modelling uncertainties		
Signal		0.17
Floating normalisations		0.07
Z + jets		0.07
W + jets		0.07
$t\bar{t}$		0.07
Single top quark		0.08
Diboson		0.02
Multijet		0.02
MC statistical		0.13

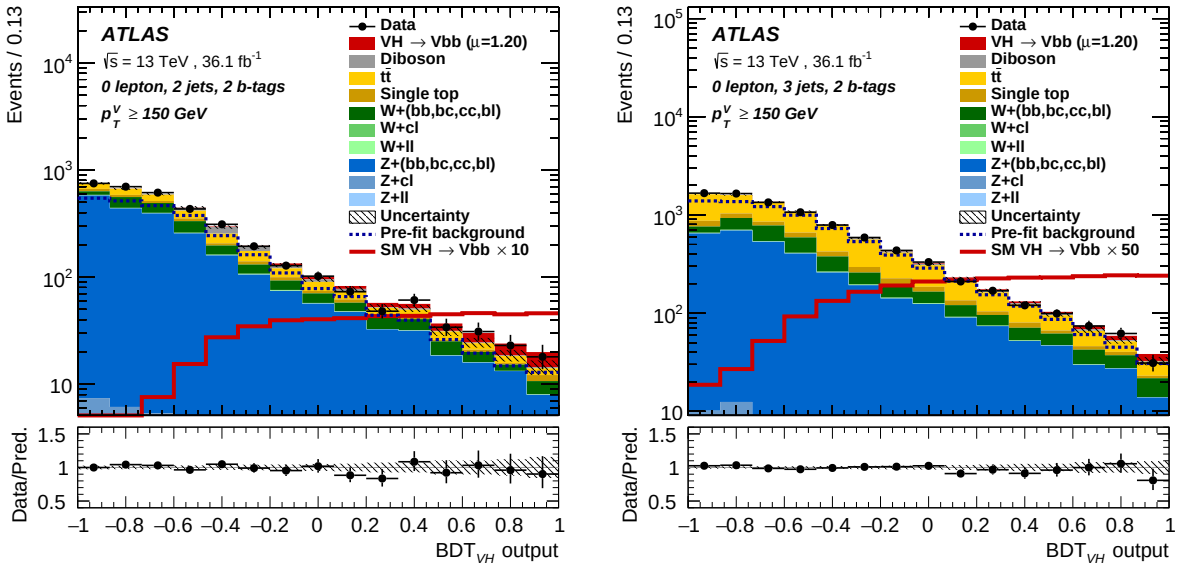
TABLE 5.15: Breakdown of the contributions to the uncertainties in μ. The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations. The b-tagging extrapolation uncertainty refers to the extrapolation of the b-Tagging calibration for jets with transverse momentum larger than 300 GeV.

Process and category	Nomalisation factor
$t\bar{t}$ 0- and 1-lepton	0.90 ± 0.08
$t\bar{t}$ 2-leptons 2 jets	0.97 ± 0.09
$t\bar{t}$ 2-leptons 3 jets	1.04 ± 0.06
W+HF 2 jets	1.22 ± 0.14
W+HF 3 jets	1.27 ± 0.14
Z+HF 2 jets	1.30 ± 0.10
Z+HF 3 jets	1.22 ± 0.09

TABLE 5.16: Factors applied to the nominal normalizations of the $t\bar{t}$ (Powheg + Pythia 8), W + HF and Z + HF (Sherpa) backgrounds, as obtained from the global fit to the 13 TeV data for the nominal multivariate analysis, used to extract the Higgs boson signal. Errors include statistical and systematic uncertainties.

Signal regions	0-lepton		1-lepton		2-lepton			
	$p_{\mathrm{T}}^{V} > 150$ GeV, 2-b-tag		$p_{\mathrm{T}}^{V} > 150$ GeV, 2-b-tag		75 GeV $< p_{\mathrm{T}}^{V} < 150$ GeV, 2-b-tag		$p_{\mathrm{T}}^{V} > 150$ GeV, 2-b-tag	
Sample	2-jet	3-jet	2-jet	3-jet	2-jet	≥3-jet	2-jet	≥3-jet
$Z + ll$	9.0 ± 5.1	15.5 ± 8.1	< 1	–	9.2 ± 5.4	35 ± 19	1.9 ± 1.1	16.4 ± 9.3
$Z + cl$	21.4 ± 7.7	42 ± 14	2.2 ± 0.1	4.2 ± 0.1	25.3 ± 9.5	105 ± 39	5.3 ± 1.9	46 ± 17
Z + HF	2198 ± 84	3270 ± 170	86.5 ± 6.1	186 ± 13	3449 ± 79	8270 ± 150	651 ± 20	3052 ± 66
$W + ll$	9.8 ± 5.6	17.9 ± 9.9	22 ± 10	47 ± 22	< 1	< 1	< 1	< 1
$W + cl$	19.9 ± 8.8	41 ± 18	70 ± 27	138 ± 53	< 1	< 1	< 1	< 1
W + HF	460 ± 51	1120 ± 120	1280 ± 160	3140 ± 420	3.0 ± 0.4	5.9 ± 0.7	< 1	2.2 ± 0.2
Single top quark	145 ± 22	536 ± 98	830 ± 120	3700 ± 670	53 ± 16	134 ± 46	5.9 ± 1.9	30 ± 10
$t\bar{t}$	463 ± 42	3390 ± 200	2650 ± 170	20640 ± 680	1453 ± 46	4904 ± 91	49.6 ± 2.9	430 ± 22
Diboson	116 ± 26	119 ± 36	79 ± 23	135 ± 47	73 ± 19	149 ± 32	24.4 ± 6.2	87 ± 19
Multi-jet e sub-ch.	–	–	102 ± 66	27 ± 68	–	–	–	–
Multi-jet μ sub-ch.	–	–	133 ± 99	90 ± 130	–	–	–	–
Total bkg.	3443 ± 57	8560 ± 91	5255 ± 80	28110 ± 170	5065 ± 66	13600 ± 110	738 ± 19	3664 ± 56
Signal (fit)	58 ± 17	60 ± 19	63 ± 19	65 ± 21	25.6 ± 7.8	46 ± 15	13.6 ± 4.1	35 ± 11
Data	3520	8634	5307	28168	5113	13640	724	3708

TABLE 5.17: The fitted Higgs boson signal and background yields for each signal region category in each channel after the full selection of the multivariate analysis. The yields are normalised by the results of the global likelihood fit. All systematic uncertainties are included in the indicated uncertainties. An entry of '-' indicates that a specific background component is missing in a certain region, or that no simulated events are left after the analysis selection [71].



(a) (b)

FIGURE 5.27: BDT postfit distributions obtained in the 0-lepton channel Signal Regions for the combined unconditional VH fit on 2015 and 2016 data in the 2-tag 2-jets region (a) and 2-tag 3-jets region (b). The best-fit signal value is used.

5.8.2 Comparison with Run 1

The SM VH(bb) search was performed with data collected at the Run 1 of the LHC. This dataset corresponds to 4.7 and 20.3 fb^{-1} integrated luminosities from pp collisions at $\sqrt{s}$ = 7 TeV and 8 TeV respectively. The observed (expected) deviation from the background-only hypothesis corresponds to a significance of 1.4 (2.6) standard deviations, which corresponds to a signal strength $\hat{\mu} = 0.52 \pm 0.32$ (stat.) $\pm$ 0.24 (syst.) for a Higgs boson mass of 125.36 GeV (the result is independent of the value m_H within about 1 GeV). Especially, individual signal strengths for each lepton channel fitted simultaneously showed a low 0-lepton $\hat{\mu}$ value, as presented in Fig. 5.31. This is observed in the Run 2 analysis as well, and will be subject to special attention in future versions of the analysis.

This result was combined with the Run 2 search, with proper treatment of correlations between the two analyses described in section 5.8.3.

5.8.3 Run 1 - Run 2 combination

A combination of analyses with data collected at $\sqrt{s}$ = 7 TeV, 8 TeV and 13 TeV (2015+2016) was performed [71]. No change was implemented in the Run 1 analysis [66], but systematic uncertainties were correlated for the dominant uncertainties.

Important improvements in the experimental apparatus (inclusion of the IBL) and b-Tagging algorithm performances (improved multivariate techniques) and use (pseudo-continuous tagging was used in Run 1) lead to the choice of decorrelating b-Tagging systematics between Run 1 and Run 2 analyses. No significant impact was observed from correlating jet energy scale systematic uncertainties between 7 TeV, 8 TeV and 13 TeV. Yet, a weak correlation scheme was adopted, where only the b-jet specific jet energy scale uncertainty is correlated.

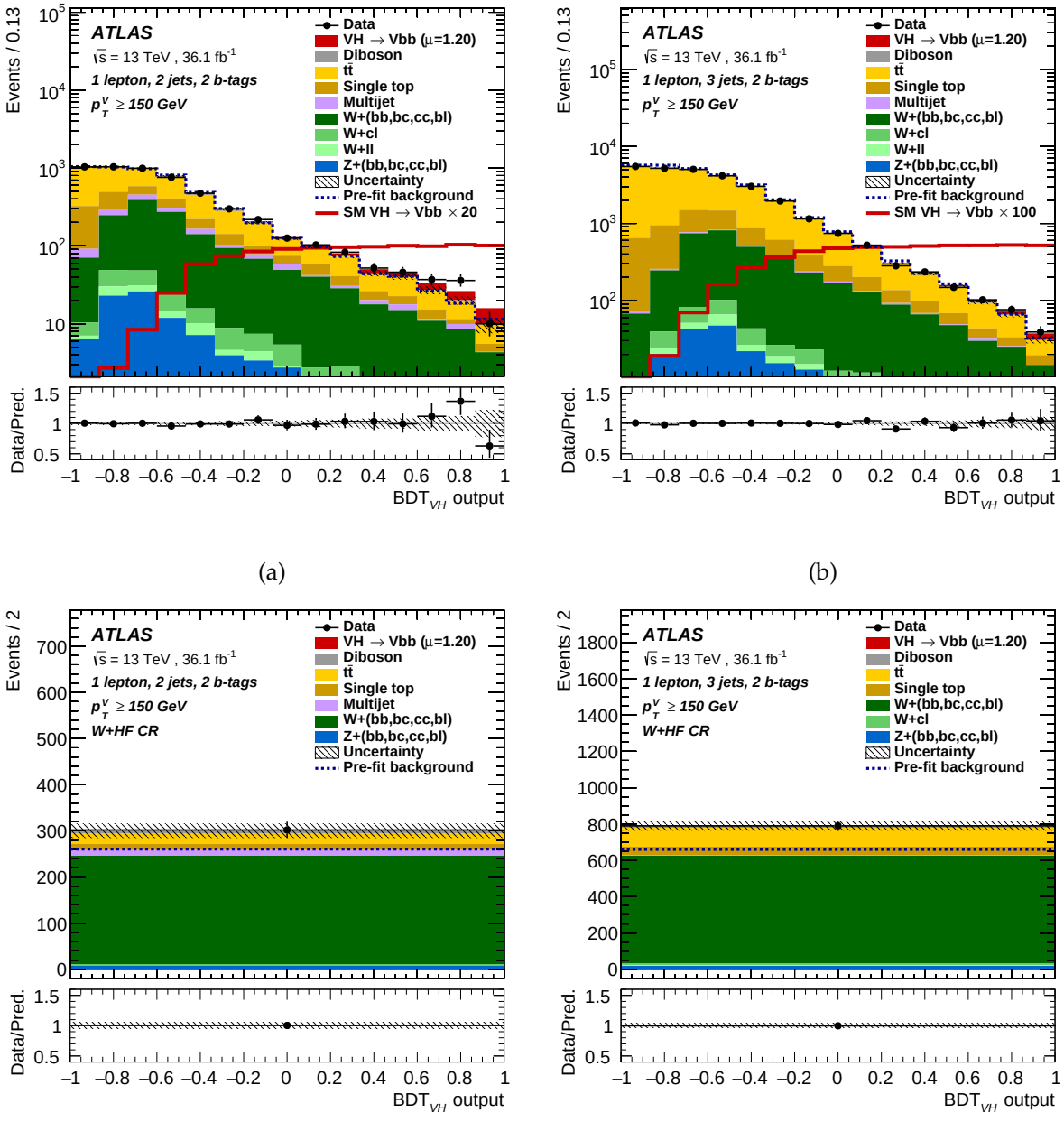


FIGURE 5.28: BDT postfit distributions obtained in the 1-lepton channel for the combined unconditional VH fit on 2015 and 2016 data in the 2-tag 2-jets SR (a), CR (c), and 2-tag 3-jets SR (b) and CR (d). The best-fit signal value is used.

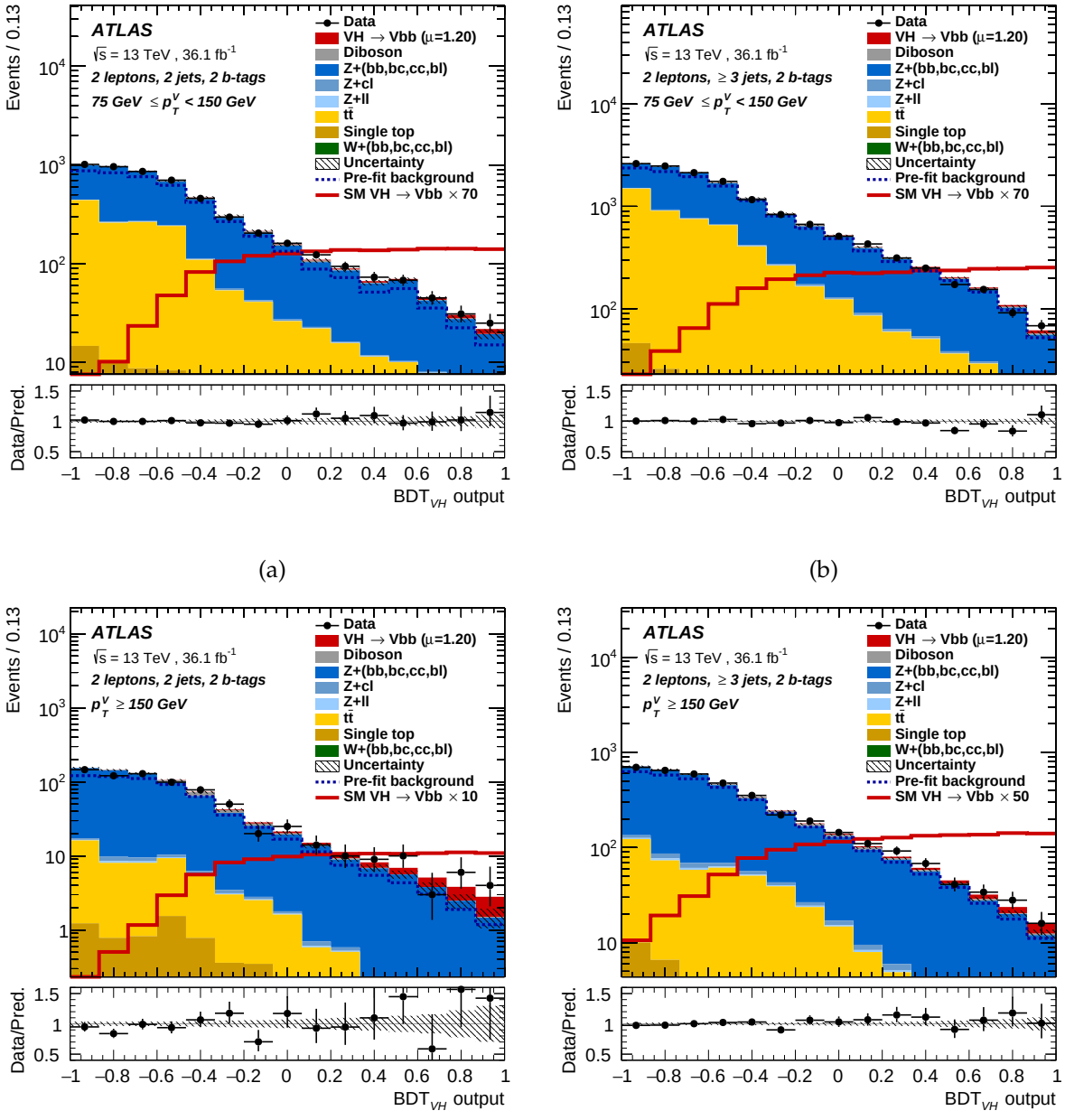


FIGURE 5.29: BDT postfit distributions obtained in the 2-leptons channel Signal Regions for the combined unconditional VH fit on 2015 and 2016 data. The medium p_T^V region is shown in the 2-tag 2-jets (a) and 2-tag 3-jets (b) categories (the 3-jets category is filled with events with three or more jets in the 2-leptons channel). The high p_T^V region is displayed in the 2-tag 2-jets (c) and 2-tag 3-jets (d) categories. The best-fit signal value is used.

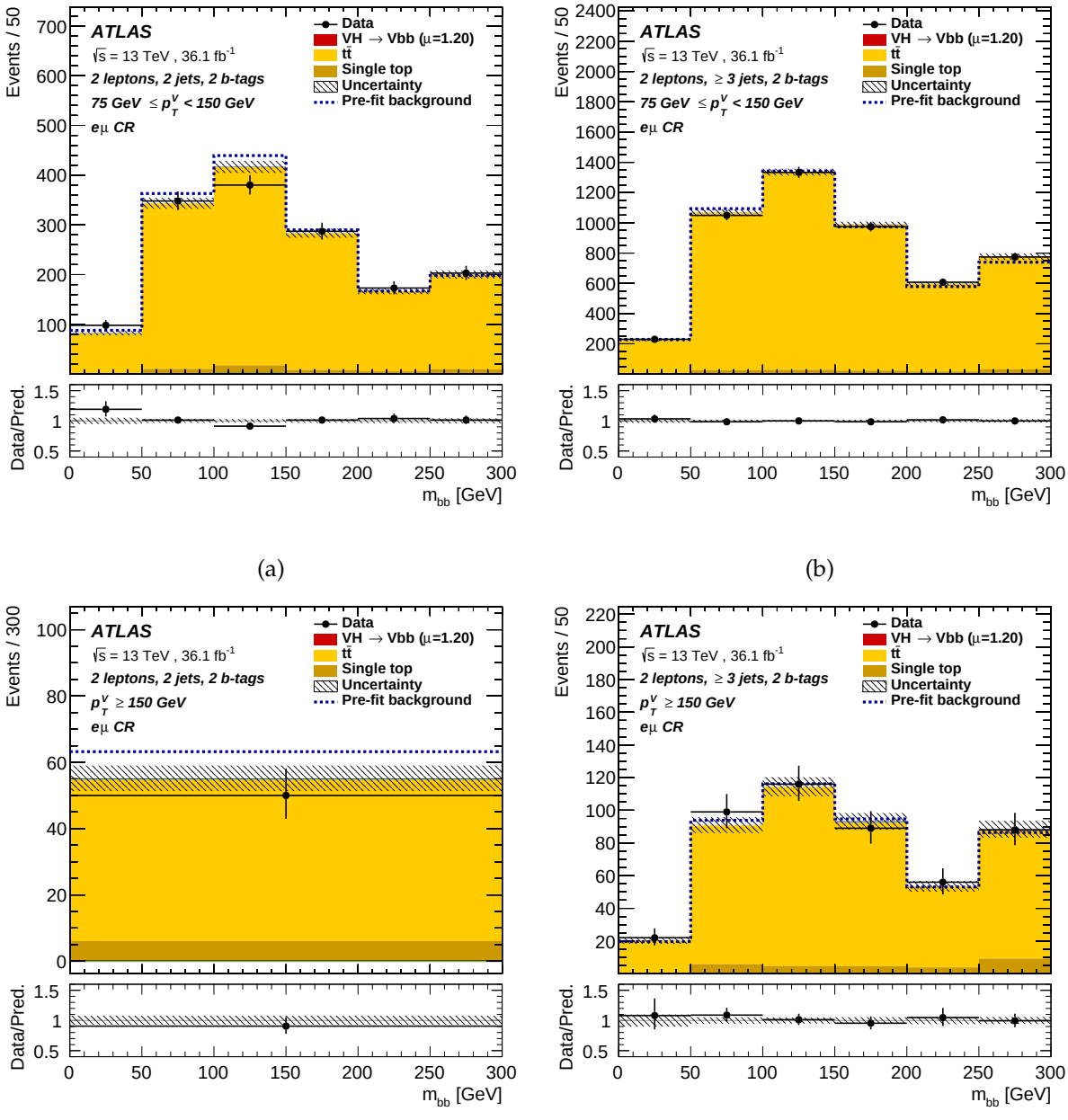


FIGURE 5.30: m_{bb} postfit distributions obtained in the 2-leptons channel $e\mu$ Control Regions for the combined unconditional VH fit on 2015 and 2016 data. The best-fit signal value is used.

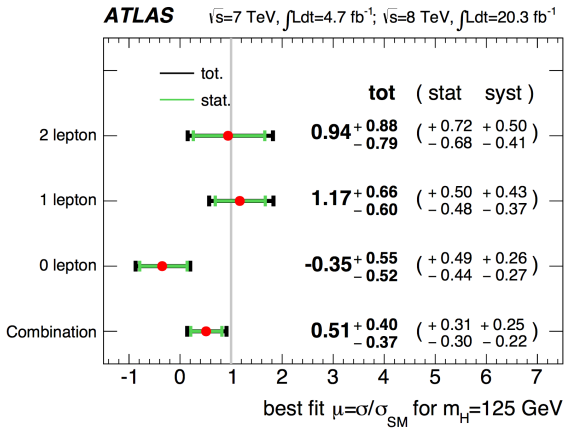


FIGURE 5.31: The fitted values of the Higgs boson signal-strength parameter μ for mH = 125 GeV for the 0-, 1- and 2-lepton channels and the combination of the three channels, with the 7 and 8 TeV datasets combined.The individual $\hat{\mu}$ values for the lepton channels are obtained from a simultaneous fit with the signal strength for each of the lepton channels floating independently [71].

The possibility for correlating modelling systematics is limited by changes or updates in the centre-of-mass energy, Monte-Carlo generators, object and event selection, and in the software tools used for simulation, reconstruction and analysis. Different correlation configurations were tested, from uncorrelated to fully correlated for the $t\bar{t}$, Z + HF, and W + HF normalizations and systematic shape variations across the two datasets. The impact on the combined signal strength was found to be smaller than 1% for all cases considered, while the effect on the signal strength uncertainty was found to be smaller than 4%.

Theory uncertainties for the overall cross-section of the Higgs boson signal, on the $H \to b\bar{b}$ branching ratio and on the p_T^V dependent NLO EW corrections are correlated across datasets.

For the tested Higgs boson mass of 125 GeV, the signal strength fitted in all channels combined is $\hat{\mu} = 0.90 \pm 0.18$ (stat.) $^{+0.21}_{-0.19}$ (syst.). This corresponds to a observed significance of 3.6 standard deviations, which can be compared to the expectation of 4.0 standard deviations.

5.8.4 Comparison to CMS

The CMS collaboration performed a search for the VH(bb) process similarly to ATLAS. The 2017 publication [76] for this analysis relies on data collected at the LHC in proton-proton collisions at $\sqrt{s} = 13$ TeV, with an integrated luminosity of 35.9 fb^{-1}. An excess of data was observed with respect to the background-only hypothesis, which corresponds to a signal strength of $\mu = 1.2 \pm 0.4$ for the VH(bb) targeted process. The observed (expected) significance for this search amounts to 3.3 (2.8) standard deviations. The combination with run 1 allows to achieve an observed (expected) significance of 3.8 (3.8) standard deviations, and to measure a signal strength $\hat{\mu} = 1.06^{+0.31}_{-0.29}$.

5.8.5 Observation of the $H \to b\bar{b}$ decay in ATLAS

The ATLAS collaboration published the observation of the $H \to b\bar{b}$ decay in 2018 [77]. This result is the combination of the searches performed at the run 1 and run 2 of the LHC for the decay of the Standard Model Higgs boson into a pair of beauty quarks produced in association to a vector boson, a pair of top quarks and in the vector boson fusion production mode respectively. The observed (expected) significance is 5.4 (5.5) standard deviations in this combination.

The search for the VH(bb) process is the most sensitive channel, and benefits from an increased dataset with respect to the 2017 publication. This channel measures an observed (expected) significance of 4.9 (4.3) standard deviations using 79.8 fb^{-1} of data collected at the run 2 of the LHC.

5.8.6 Observation of the $H \to b\bar{b}$ decay in CMS

On another side, the CMS collaboration reported the observation of the Higgs boson decay to bottom quarks as well [78]. This results from the combination of searches for the $H \to b\bar{b}$ process in several production modes of the Higgs boson. The observed (expected) significance is 5.6 (5.5) standard deviations, which corresponds to a signal strength of 1.04 ± 0.20.

Chapter 6

Additional studies on the SM VH(bb) analysis

6.1 Studies dedicated to better understanding the published result

Following the 2016 and 2017 publications [79] [71], investigations were performed aiming at a better understanding of unexpected Nuisance Parameters pulls and estimated pre-fit errors.

6.1.1 The Z+Jets normalization issue

MC samples are normalized to their best known cross-section estimate. For most variables, the superimposition of MC samples distributions with appropriate normalization is expected to fit to data within systematic and statistical uncertainties. Yet, MC to data comparisons sometimes present a disagreement in the expected and observed number of events. To account for global scaling uncertainties, dominant background processes are associated individual floating normalization Nuisance Parameters, which allow for an unconstrained scaling of the corresponding event yields. In the SM VH(bb) search analysis, ttbar, Z+HF and W+HF processes are concerned by this procedure. The fitting procedure described in Section 5.7.4 allows to extract the scale factors for these floating normalization Nuisance Parameters, which are presented in Table 5.16 for the 2017 result.

The normalization of the Z+HF processes is significantly underestimated in the 0- and 2-leptons channels where it represents the main background. The floating normalization Nuisance Parameter for this background is particularly constrained in the 2-leptons categories due to the larger purity in Z+jets events. An additional constrained Nuisance Parameter is needed in the 0-lepton channel, which allows to further increase the yield of the Z+HF background in 0-lepton categories on top of the floating normalization Nuisance Parameter shared by the 0- and 2-lepton(s) channels. The uncertainty on the ratio of the yields in the 2- and 0-lepton(s) channels for this process was estimated to 7% from tunes and generators comparisons using alternative Madgraph Z+jets samples in the high-p_T^V region of the two channels. Nominal yields and variables shapes are estimated from Sherpa Z+Jets samples (Section 7.2.2). The constrained Nuisance Parameter dedicated to the 0-lepton channel is pulled by 1.3 standard deviations in the combined fit in order to achieve a good data to Monte-Carlo agreement in both 0- and 2-lepton(s) channels. The following sections describe several studies performed in order to better understand the source of this discrepancy between the 0- and 2-lepton(s) channels.

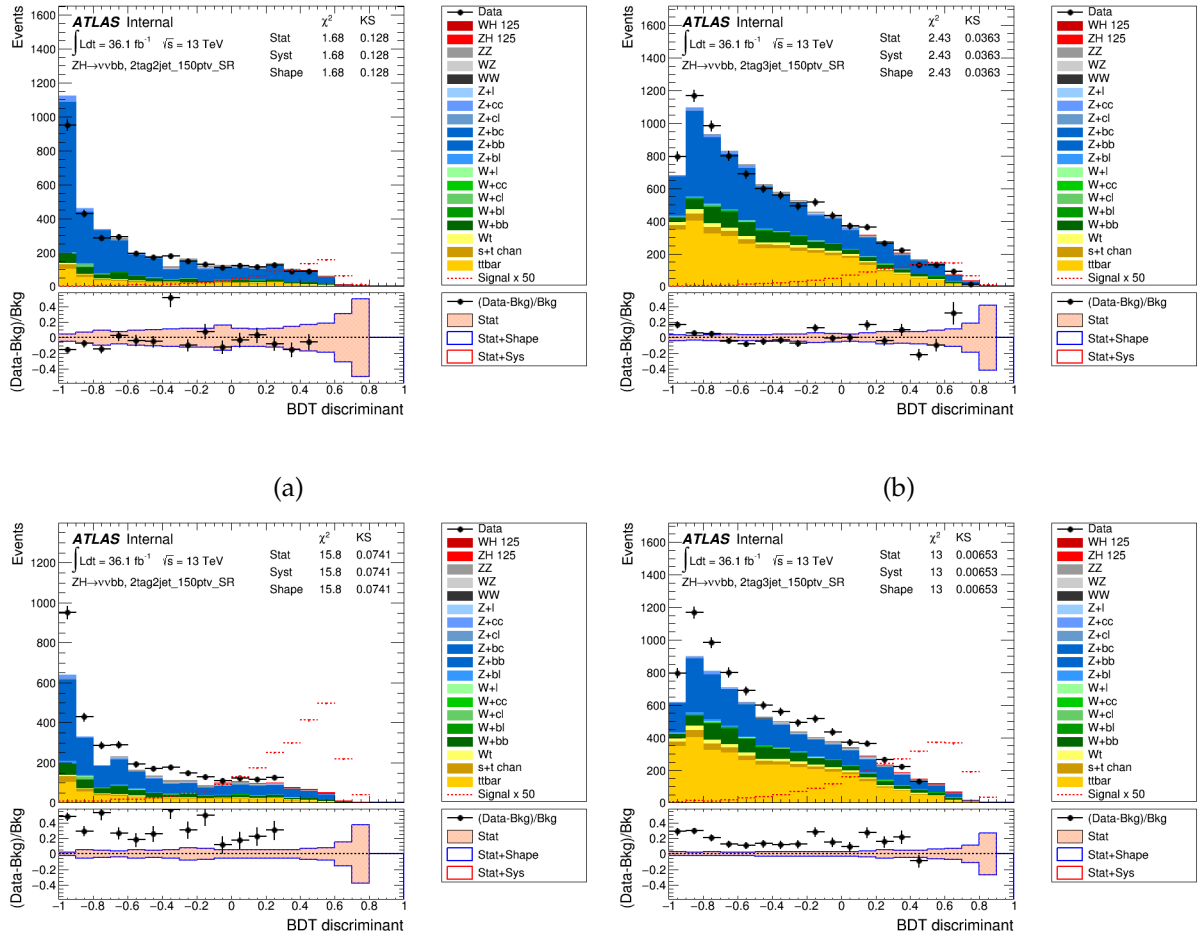
The first approach developed in order to explain why an additional nuisance parameter is required in the 0-lepton channel aimed at spotting potential bugs in the 0-lepton code, or any requirement in the 0-lepton selection wich would lower the acceptance of simulated Z+jets events with respect to the 2-leptons selection. The strategy adopted for this study consists of running the 0-lepton framework on 1- and 2-lepton(s) events. Events with 1- and 2-lepton(s) are transformed by adding the lepton(s) momentum vector(s) to the MET vector. This is equivalent to running the 0-lepton selection using p_T^V instead of MET. The analysis aims at similar

Category	Channel	Source (code processed)		
		0-lepton code	0/1/2-lepton(s) code	ICHEP publication
2-jets	0-lepton	1.34	1.34	1.34
	1-lepton	1.13	1.11	1.04
	2-lepton	1.24	1.22	1.22
3-jets	0-lepton	1.18	1.18	1.18
	1-lepton	1.00	0.99	0.97
	2-lepton	1.19	1.19	1.15

TABLE 6.1: Data to MC ratio measured in events with three jets, in the different lepton channels. The ratio was computed running either the 0-lepton code or the appropriate n-lepton code. The last column presents approved data to MC ratios. Especially unofficial estimates lack the multijet contribution in 1-lepton categories, and events with more than 3 jets in the 2-leptons channel.

regions of the p_T^V spectrum in all leptons channels, which allows to run efficiently the 0-lepton selection on 1- and 2-lepton(s) events. The data to Monte-Carlo yield ratios computed with this approach are presented in Table 6.1. No discrepancy seems to be produced by running the 0-lepton framework with respect to the appropriate framework for 1- and 2-lepton(s) events. This procedure has special interest in the 2-leptons channel which presents a background composition enriched in Z+jets events. Yet, both the 1- and 2-lepton(s) channels show consistent results in running different frameworks on the appropriate datasets, despite the very different background compositions.

In a second time, alternative Z+jets samples generated with Madgraph were tested and showed a better agreement in data to Monte-Carlo yields comparisons. In the 2- (3-) jets 0-lepton channel Signal Region, the data to MC background pre-fit yields ratio evolves from 1.36 to 0.94 (1.19 to 1.02) using Madgraph instead of Sherpa Z+jets samples. Corresponding distributions of the multivariate discriminant are presented in Fig. 6.1. Yet, these Madgraph samples have lower statistics and are generated at LO in QCD, which motivates the use of Sherpa samples for the final analysis.



(a) (b)

(c) (d)

FIGURE 6.1: Multivariate discriminant pre-fit distributions obtained in the 0-lepton channel Signal Regions. Figures (a) and (b) ((c) and (d)) exploit Madgraph (Sherpa) samples in the 2- and 3-jets categories respectively.

In order to track the origin of this discrepancy, acceptance and variables shape comparisons of Madgraph and Sherpa Z+jets samples were performed at truth-level. The requirements on truth-event variables are intented to mimic the selection at reconstructed level. Cuts are applied in the following order :

- 0: All events
- 1 : 0 lepton, at least 2 signal-jets (pT > 20 GeV and $|\eta| < 2.5$)
- 2 : 2 or 3 jets
- 3 : MET >= 150 GeV
- 4 : $\Delta\phi(MET, MPT) < \pi/2$
- 5 : min$\Delta\phi(MET, jet) < 20^{\circ}$
- 6 : SumPt > 120 (150) GeV in 2 (3) jets categories
- 7 : pT(leading b-jet) > 45 GeV
- 8 : $\Delta\phi(b-jet, b-jet) < 140^{\circ}$
- 9 : $\Delta\phi(MET, dijet) > 120^{\circ}$

The ratio of Sherpa to Madgraph cutflow distributions for the 0-lepton selection are presented separately for inclusive samples and for events with two b-labelled jets in Figure 6.2. The discrepancy in the first bin is caused by a slight difference in cross-sections which is not corrected here. The two first requirements on lepton and jets multiplicities slightly increase the Sherpa to Madgraph yield ratio, whose final value is expected to be about 0.8. The MET requirement (MET > 150 GeV) largely impacts this ratio, which does not evolve much with the remaining selection cuts. A comparisons of the shape of the MET and p_T^Z distributions between Sherpa and Madgraph samples is shown in Figure 6.3. As expected from the cutflow evolution, events generated with Madgraph tend to have larger MET values than those simulated with Sherpa.

Assuming a better modelling of p_T^Z in Madgraph than Sherpa samples, the feature that made this discrepancy invisible in modelling systematics is related to the fit model and Nuisance Parameters. The uncertainty on the modelling of the Z boson transverse momentum shape is encoded within a single Nuisance Parameter, which changes the slope of the vector boson transverse momentum if it is pulled. This modelling uncertainty is estimated in 2-leptons Control Regions very pure in Z+Jets events, and is shown in Figure 6.4. In overall, the enveloppe allows for a 10% scaling of events weights for events with p_T^V greater than 150 GeV. At this point, a new interpretation of the different fit configurations can be proposed :

- 0-lepton fit : MET is required to be greater than 150 GeV and is expected to take values close to p_T^Z. Hence, the impact of the p_T^Z shape systematic is degenerated with the floating normalization scale. Hence, the p_T^Z shape Nuisance Parameter is not pulled while the freely floating normalization scale factor absorbs the potential mismodelling of p_T^Z.
- 2-leptons fit : thanks to the medium- and high-p_T^V regions (75 < p_T^V < 150 GeV and p_T^V > 150 GeV), the impact of the p_T^Z shape Nuisance Parameter is no longer degenerated with the freely floating normalization factor. In this result, the p_T^Z shape Nuisance Parameter is pulled by -0.15 standard deviations.

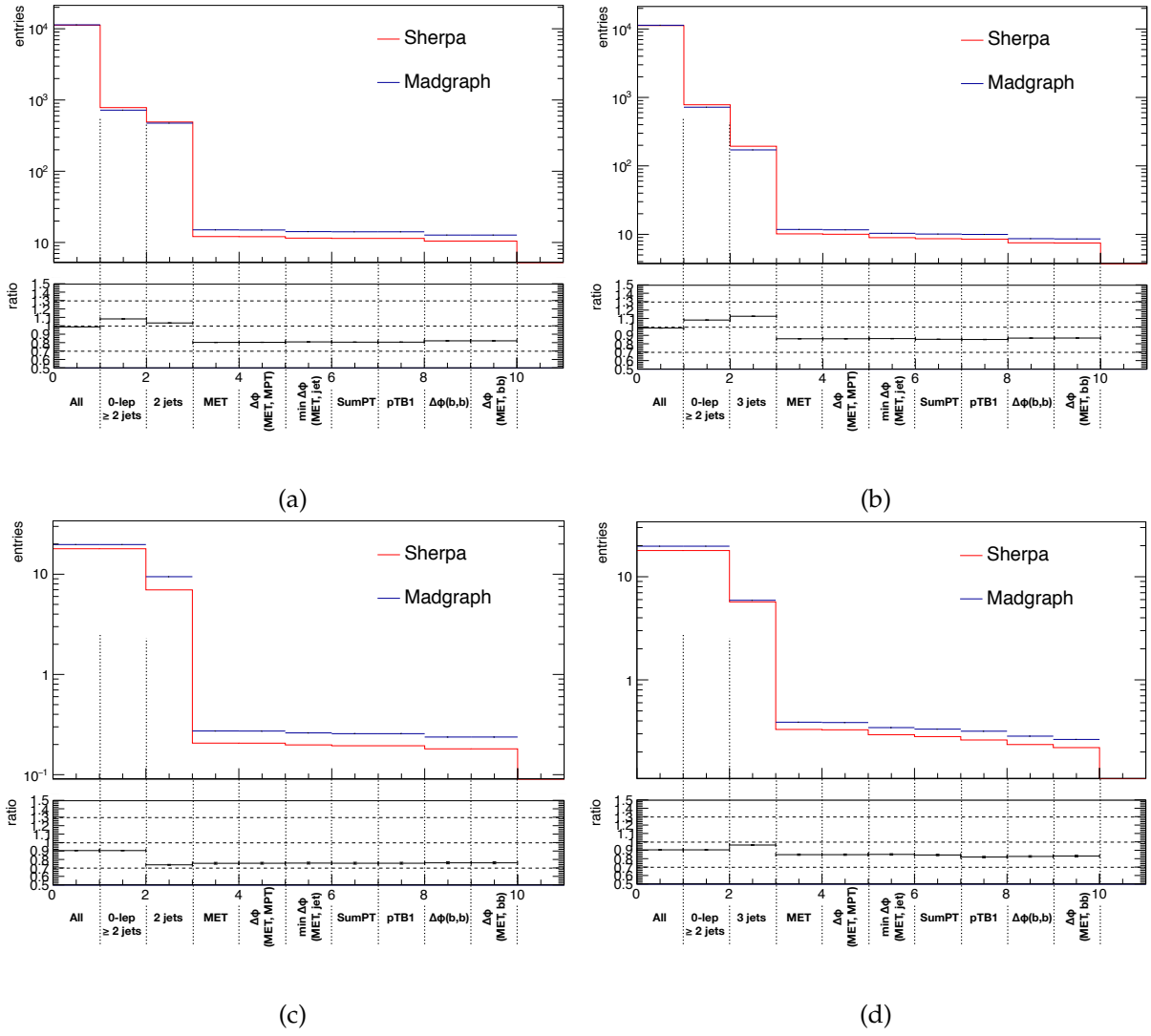


FIGURE 6.2: Comparison of Madgraph and Sherpa acceptance to consecutive analysis requirements (cutflow). Figures (a) and (b) do not have any requirement on the jets flavour, while Figures (c) and (d) focus on events with two b-labelled jets. Figures (a) and (c) ((b) and (d)) show the present the cuts efficiency ratio for events with 2- (3-) jets from the second bin of the cutflow.

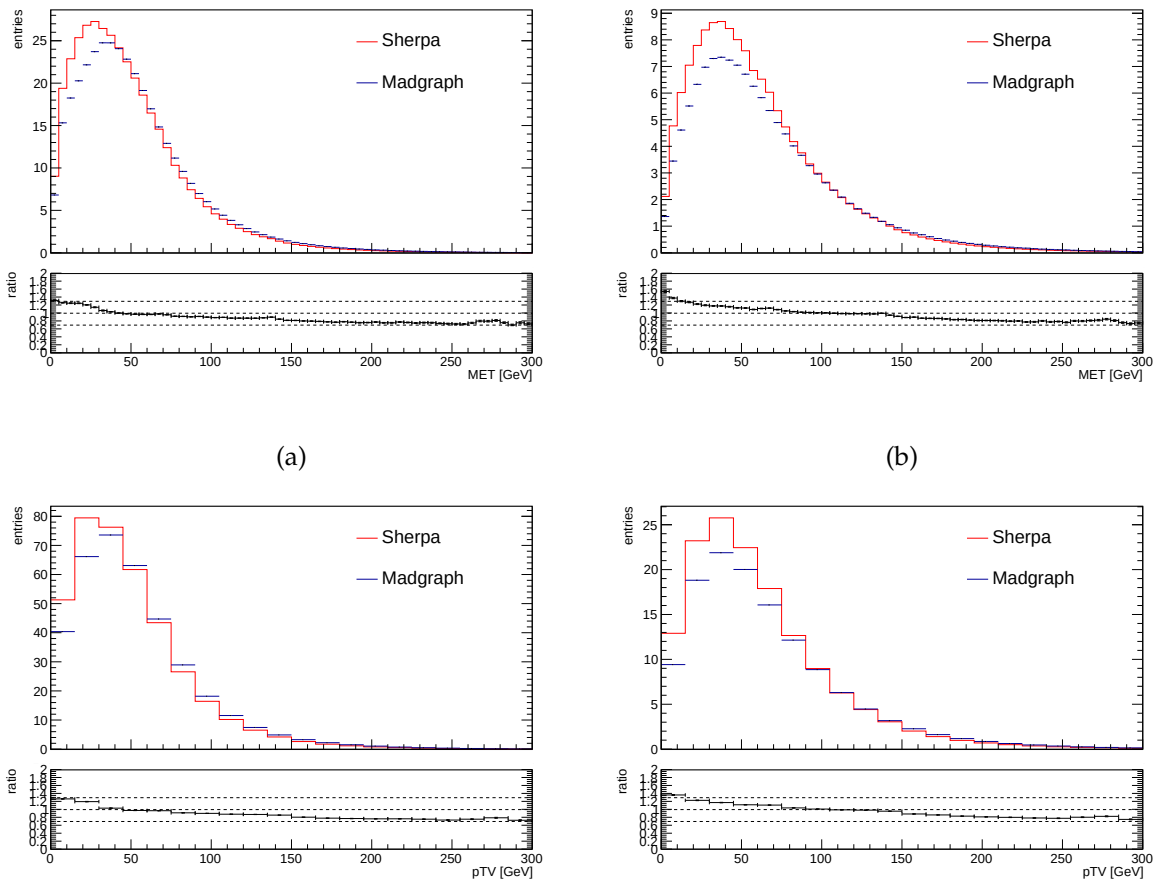
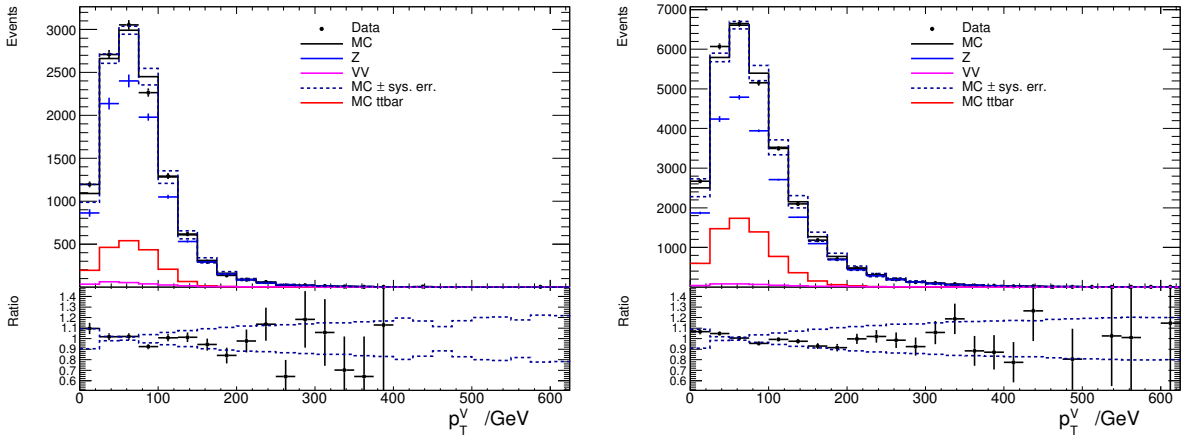


FIGURE 6.3: Comparison of Madgraph and Sherpa TruthMET ((a) and (b)) and pTV ((c) and (d)) obtained in the 0-lepton 2- ((a) and (c)) and 3-jets ((b) and (d)) 0-ptag categories. No filter on jets flavour is applied in these plots.



(a) (b)

FIGURE 6.4: Vector boson transverse momentum distribution in the 2-leptons CR where the region 110 GeV < m_{bb} < 140 GeV is excluded. Figure (a) (b) displays events with two (three) jets.

- Combined fit : the shape and normalization Nuisance Parameters are mostly constrained in the 2-leptons categories. The p_T^Z shape Nuisance Parameter is pulled by -0.12 standard deviations. Therefore, the persistent yield discrepancy in the 0-lepton channel is accounted for with the constrained scaling Nuisance Parameter. Other effects not understood may be compensating for the normalization of Z+HF processes in the 2-leptons channel, which is not expected to be different from that in the 0-lepton channel.

These results enlight tensions in the fit strategy. The 0-lepton fit does not allow for a consistent use of the modelling systematic on the vector boson transverse momentum shape and the normalization Nuisance Parameter which account for the same effect. In the combined fit, this reasoning does not explain the persistent discrepancy in 0-lepton categories while a good agreement is observed in the 2-leptons channel. Further investigations will be required to better understand this discrepancy.

6.1.2 Parton Shower systematic on signal samples

The uncertainty on the signal strength due to generator tuning and modelling of Underlying Events (UE) and Parton Shower (PS) in MC signal samples is encoded within a single Nuisance Parameter, named UEPS uncertainty. Fit rankings of the impact of Nuisance Parameters on the signal strength estimated from Asimov Datasets and fit to data showed a dominant impact from the UEPS Nuisance Parameter, as shown in Fig. 5.26. As it measures a signal acceptance modelling uncertainty, this systematic does not impact the significance, but it affects the signal strength uncertainty. The pre-fit error from this Nuisance Parameter is estimated for each individual lepton channel from $q\bar{q} \to VH$ samples at truth-level. The uncertainty from Underlying Events modelling is extracted from comparisons of signal acceptance for Madgraph samples with varied PDF tunes [80]. The uncertainty from Parton Shower models is estimated by comparing the acceptance for $q\bar{q} \to VH$ events generated with Powheg whith Parton Shower performed with Pythia 8 and Herwing 7 alternatively. The overall UEPS uncertainty consists of the maximal discrepancy on signal acceptance between UE and PS in each individual sensitive region of the analysis. Although they supposedly encode different effects, the two methods are expected to significantly overlap and lead to similar uncertainty estimates. In this final

estimates, the PS uncertainy was often significantly larger than the UE error. Yet, cutflow comparisons on the acceptance for MC samples used for the PS estimate showed much smaller differences in acceptance than that used in the fit. Further investigations revealed a normalization issue in the program used for acceptance systematic estimate computations. Updated values are presented in Table 6.2.

	0L: ZH $\to \nu\bar{\nu}b\bar{b}$		1L: WH $\to \ell\bar{\nu}b\bar{b}$		2L: ZH $\to \ell\bar{\ell}b\bar{b}$	
Category	2j	3j	2j	3j	2j	$\geq$3j
PS tune	7.37 %	5.50%	7.46%	6.69%	6.36%	7.67%
Initial Herwig 7	10.01%	10.42%	12.05%	16.47%	13.87%	14.38%
Updated Herwig 7	3.31%	3.69%	11.92%	16.34%	6.89%	7.37%
Initial UEPS	10.01%	10.42%	12.05%	16.47%	13.87%	14.38%
Updated UEPS	7.37%	5.50%	11.92%	16.34%	6.89%	7.67%

TABLE 6.2: Acceptance uncertainty for A14 tune variations and Parton Shower model in each analysis region. Shown in the bottom row is the acceptance uncertainty from comparing the Powheg+Herwig7 samples to the nominal.

This result shows an overestimate of the uncertainty on the measured signal strength, and will help to understand future estimates of signal acceptance systematics.

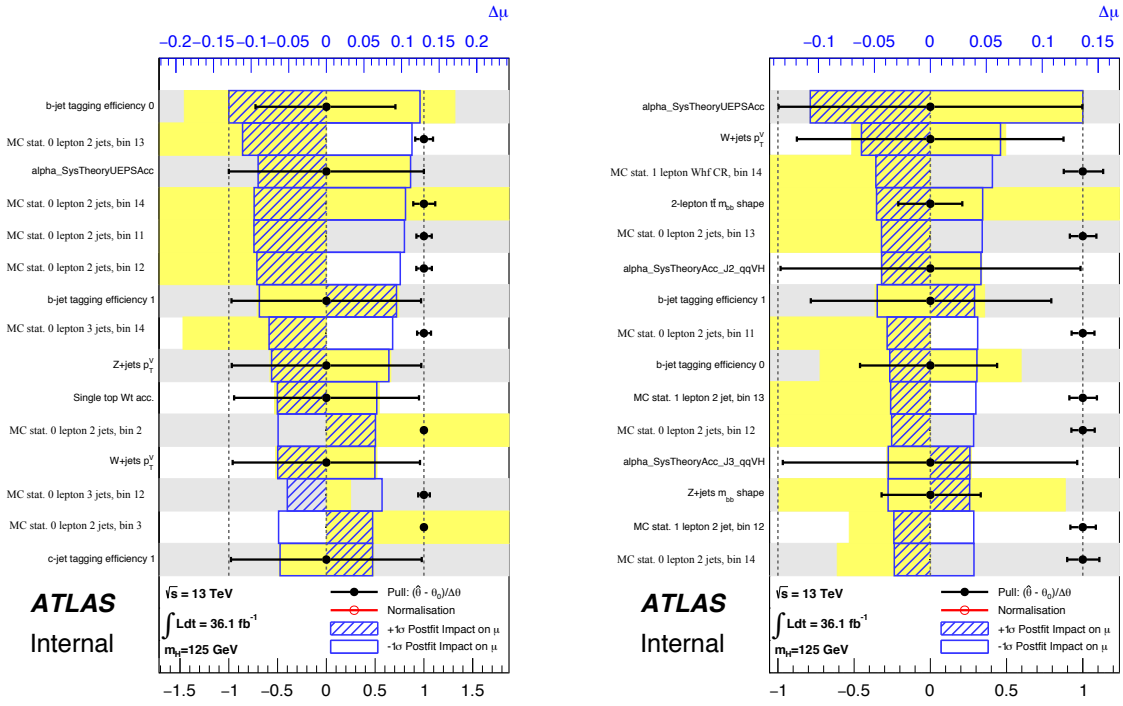
6.2 Optimisation of Monte-Carlo filters

Ranking the impact of systematics on the fitted signal strength uncertainty allowed to enlight a large impact of the limited Monte-Carlo statistics, as shown in Fig. 6.5. Especially, the γ-stat Nuisance Parameters attached to the most sensitive bins of the multivariate discriminant largely increase the uncertainty on $\hat{\mu}_{VH}$. The main limitation to this issue relates to the generation efficiency of Monte-Carlo samples. Simulations target with difficulty the most sensitive phase space of the SM VH(bb) analysis, which is very restricted. This situation leads to respectively low (high) statistical uncertainty in regions with low- (high-) interest for the analysis. The respective contributions of each MC sample to the statistical uncertainty in the most sensitive bins of the 0-lepton multivariate discriminant distribution are presented in Table 6.3, and show a dominant impact of the Z+Jets and $t\bar{t}$ processes. The following studies present how the simulation efficiency was improved for these samples.

6.2.1 Z+jets

Z+Jets MC samples used for background estimates at the 2017 publication were generated using a max(HT, p_T^Z)) filter, where HT measures the scalar sum of the transverse momentum of jets in the event. In the phase space of the 0-lepton channel of the VH(bb) search, the Z+Jets process with Z decaying to neutrinos is a major background contribution. For this particular process and decay mode, the value of p_T^Z is very close to the truth-level MET, assuming potentially low transverse momentum neutrinos from B- and D-hadron decays. As the sensitivity of the VH(bb) search is closely correlated to p_T^V, MC generation should particularly target at high-p_T^V events in Z+Jets samples. The choice of the max(HT, p_T^V) filter for Z+Jets MC samples is not natural for 0-lepton background estimates, which targets high-MET events with low jets multiplicity. This filter naturally increases the generation efficiency for high-jets multiplicity events, which subsequently decreases the acceptance for high-MET events.

Comparisons of the statistical uncertainties for 0-lepton Z+Jets samples were performed between p_T^V filtered Sherpa 2.2 MC samples with respect to Sherpa 2.2.1 MC samples generated



(a) 0-lepton fit (b) Combined fit

FIGURE 6.5: Impact of systematic uncertainties for the fitted Higgs boson signal-strength parameter μ for the nominal MVA analysis applied to the 13 TeV data. The luminosity was scaled to prospect value L = 150 fb^{-1}. This ranking highlights the impact of MC statistics on the signal-strength uncertainty.

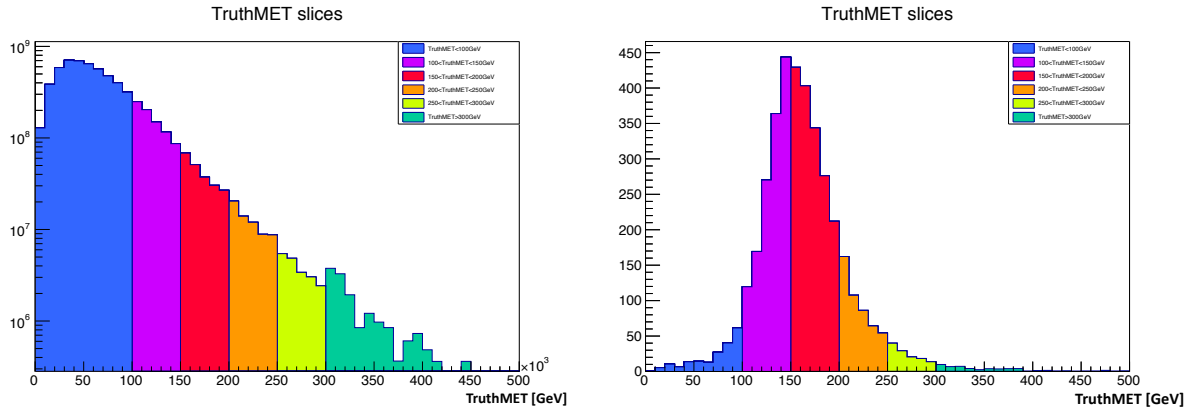
Bin index	Zbb	Wbb	stopWt	ttbar	s/sqrt(b)
4	0,36	0,10	0,09	0,30	0,69
3	0,55	0,09	0,02	0,24	0,88
2	0,28	0,15	0,06	0,42	1,01
1	0,47	0,07	0,05	0,34	1,17
0	0,11	0,17	0,24	0,45	1,27
4	0,08	0,04	0,13	0,69	0,44
3	0,18	0,11	0,08	0,54	0,54
2	0,07	0,06	0,06	0,78	0,65
1	0,16	0,10	0,37	0,31	0,76
0	0,09	0,10	0,08	0,53	0,93

2 jets
3 jets
Increasing sensitivity

Fraction of stat. error square : $Error_{sample}^2$ / $Error_{all\_bkg}^2$

Dominant impact
Non negligible impact

TABLE 6.3: Dominant contributions of MC samples to the statistical error, in the form of fraction to the total squared statistic uncertainty. The most sensitive BDT bin after statistical rebinning has index 0. This shows dominant uncertainties from $t\bar{t}$, Zbb, single-top Wt and Wbb.



(a) (b)

FIGURE 6.6: Truth distribution of p_T^Z in Z+jets simulated events. Figure (a) presents the output of MC generation with no selection applied, while Figure (b) has 0-lepton requirements applied at reconstructed level. Most events are generated in the slice with filter TruthMET < 100, although a very small fraction of these events enter the 0-lepton phase space.

Slice [GeV], x = p_T^V, max(HT, p_T^V)	$\Delta^2_{stat}(max(HT, p_T^V))/\Delta^2_{stat}(p_T^V)$
x ∈ [70, 140]	1.66
x ∈ [140, 280]	1.94
x ∈ [280, 500]	4.75
Combined	1.90

TABLE 6.4: Ratio of max(HT, p_T^V) to p_T^V filtered MC samples squared statistical uncertainties.

with a max(HT, p_T^V) filter, and presented in Table 6.4. Simulated events were further filtered by requiring a B-hadron with transverse momentum larger than 5 GeV to get closer to the analysis irreducible background. We showed that adopting a p_T^V filter allows to decrease the statistical uncertainty by a factor of two for the same number of generated events. The p_T^V variable is now used as default filter quantity for the new 0-lepton Z+Jets MC samples.

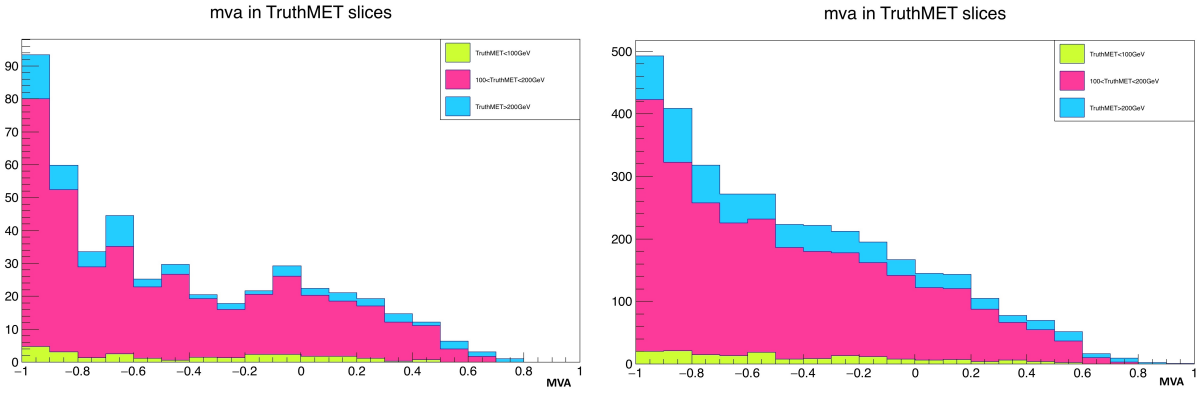
Further studies were performed aiming at improving the edge of p_T^V slices to better optimize the efficieny of MC events. Distributions of p_T^Z in Fig. 6.6 for events passing the 0-lepton selection show very low acceptance acceptance for events with p_T^Z < 100 GeV, generated in the slice $p_T^Z \in$ [70; 140] GeV. The definition of the slice is quite inefficient to the 0-lepton VH(bb) phases space, as the differential cross-section for events with 100 GeV < p_T^Z < 140 GeV represents only 38% of the inclusive cross-section for events with 70 GeV < p_T^Z < 140 GeV. A proposal for slices with edges [0; 70; 100; 140; 280; 500] GeV in p_T^Z was submitted to the ATLAS modelling group, but not accepted so far.

6.2.2 ttbar

Studies similar to those presented in Section 6.2.1 were performed to optimize the generation efficiency of $t\bar{t}$ events. MC samples generated with Powheg + Pythia 8 using a truth-level MET (TruthMET) filter were requested for SUSY analyses with slices and number of events described in Table 6.5.

Slice	Number of generated events
300 > TruthMET > 200 GeV	15 M events
400 > TruthMET > 300 GeV	3 M events
TruthMET > 400 GeV	2 M events

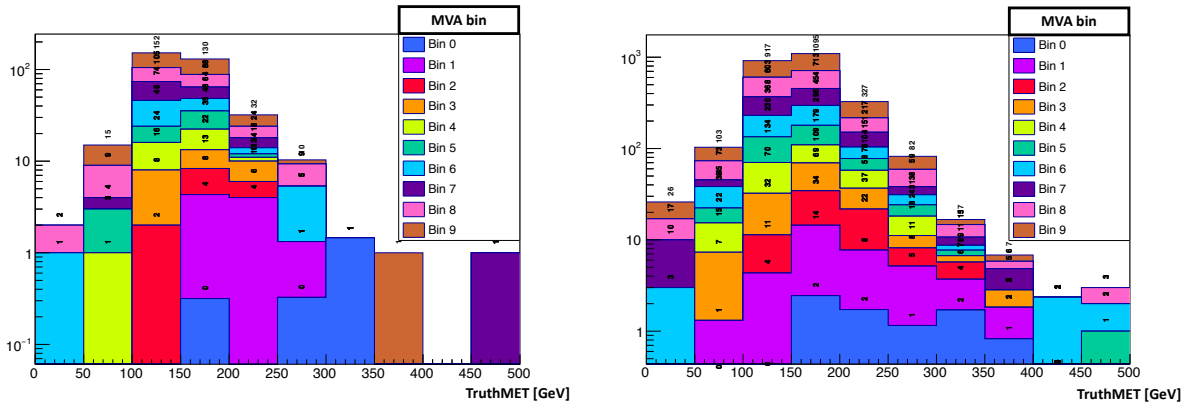
TABLE 6.5: Number of events in newly available $t\bar{t}$ samples generated with TruthMET filter



(a) (b)

FIGURE 6.7: BDT discriminant distribution of $t\bar{t}$ events in the 2-jets (a) and 3-jets (b) categories, split in TruthMET slices.

These samples can easily be exploited for VH(bb) background estimates. Basic computations show a 1.9% gain in statistical significance from integrating these events to the inclusive $t\bar{t}$ sample in the 0-lepton channel. Yet, the region 100 GeV < TruthMET < 200 GeV is responsible for an important fraction of $t\bar{t}$ events in sensitive BDT bins, as presented in Figures 6.7 and 6.8. Studies for an addition slice were performed, and showed a potential overall gain of 4.5% on the 0-lepton statistical significance by generating 20M events with 100 GeV < TruthMET < 200 GeV, for which a request was performed. Looking for further optimisations within this latter slice showed very limited gain in further splitting this region in more slices.



(a) (b)

FIGURE 6.8: Truth MET distribution in $t\bar{t}$ events in the 2-jets (a) and 3-jets (b) categories. Events are further categorized based on their index in the multivariate discriminant distribution after statistical rebinning. The most sensitive bin has index 0.

6.3 Studies of the ttbar background

6.3.1 Acceptance for ttbar events

The 0-lepton SR phase space is restricted to events with two or three jets, MET greater than 150 GeV, and no electron or muon identified in the final state. Hence, the topology of $t\bar{t}$ events should lead to a very little acceptance in the 0-lepton SR. Yet, its large cross-section yields to a major contribution to the total background. This motivated studies in order to better understand the topology of these events. The contamination from $t\bar{t}$ events is categorized with respect to top-quarks decays in the next sections.

6.3.1.1 Fully hadronic ttbar decays

Fully hadronic $t\bar{t}$ decays are expected to produce about six jets, from the b-quark and hadronic W-decay for each top (anti-)quark. Besides, no MET is expected, except from mismeasurements of jets transverse momentum and unreconstructed neutrinos produced in B- and D-hadrons decay. Running the analysis on fully hadronic $t\bar{t}$ MC samples confirmed the negligible acceptance for this decay mode in the 0-lepton SR.

6.3.1.2 Semi-leptonic ttbar decays

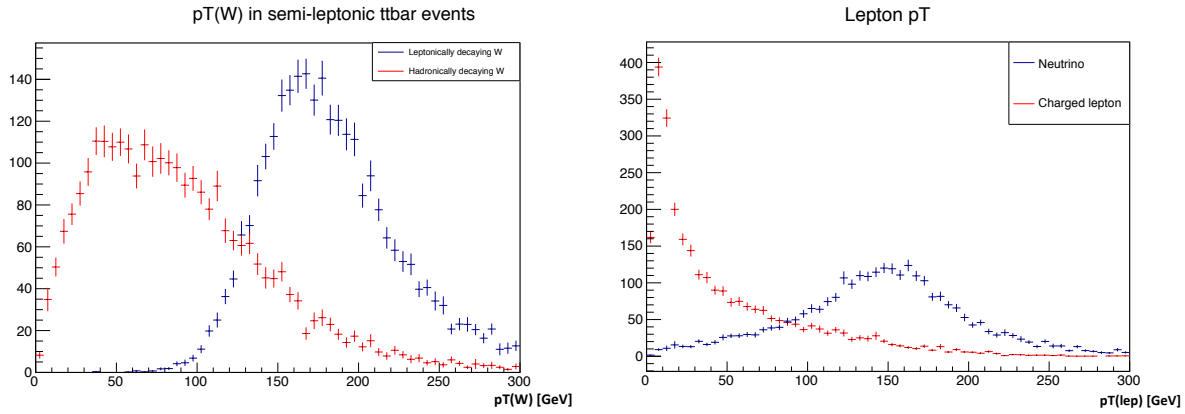
Semi-leptonic events $t\bar{t} \rightarrow bW^{\pm}(q\bar{q})\bar{b}W^{\mp}(\ell\nu)$ represent an important fraction of the $t\bar{t}$ background, especially in categories with high-jets multiplicity. Table 6.6 summarizes the fractions of semi- and dileptonic decays for different jets multiplicities, based on the truth of MC samples for events passing the 0-lepton selection.

Semi-leptonic $t\bar{t}$ decays are expected to produce about four jets at leading order. The lepton decay generates MET from the unreconstructed neutrino. This enhances the efficiency with respect to the selection requirement MET > 150 GeV. The selection does not veto events with an identified tau lepton. This enhances the contribution from events with a tau-lepton produced in the W-boson decay, which represent 75% of this background[1]. Conversely, semi-leptonic events with a non-identified electron (muon) constitute 11% (14%) of the contamination, where the

[1] the τ-lepton can decay leptonically, but a lower acceptance is expected with respect to hadronic τ-lepton decays.

	$t\bar{t}$ decay	
Number of jets	Semi-leptonic	Dileptonic
2 jets	41%	59%
3 jets	76.5%	23.5%
$\geq$ 4 jets	91%	9%

TABLE 6.6: Fractions of semi- and dileptonic decays of $t\bar{t}$ events depending on the jets multiplicity in the 0-lepton phase space.



(a) (b)

FIGURE 6.9: Transverse momentum distributions of the W-bosons (a) and leptons (b) produced in top-quarks decay. In Fig. 6.9a, the blue (red) distribution characterizes events with leptonically (hadronically) decaying vector boson. In Fig. 6.9b, the blue (red) distribution characterizes the neutral (charged) lepton. The charged lepton is picked in the truth of MC simulation, prior to showering.

charged lepton has an average transverse momentum about 12 GeV in truth simulation, prior to showering. This low value enhances mis-reconstruction and mis-identification probabilities. Besides, further studies reveal that events which enter the 0-lepton SR tend to combine specific topologies in the decay of top-quarks and W-bosons :

- The hadronically decaying W-boson is emitted with low transverse momentum in the top-quark decay, as shown in Fig. 6.9a. Hence, jets are often produced below the transverse momentum threshold in the objects selection. Conversely, the corresponding b-quark is emitted with high transverse momentum.

- The leptonically decaying W-boson is emitted with high transverse momentum in the top-quark decay (Fig. 6.9a). The W-boson decay tend to produce high-pT neutrinos collinear to the W-boson propagation axis. This enhances the efficiency for producing large MET in the final state. Conversely, the corresponding charged lepton is emitted with low-pT value, hence low reconstruction and identification efficiencies.

This specific topology of $t\bar{t}$ events leads to final states with two b-quarks, light-flavour quarks with low energy, a soft charged lepton and a high transverse momentum neutrino, as represented schematically in Fig. 6.10.

Yet, about half of these events do not satisfy MET > 150 GeV from the unreconstructed neutrino in the $t\bar{t}$ decay, as shown in Fig. 6.9b. This characterizes the importance of reconstruction biases in the measurement of MET, especially from jets energy mis-measurement.

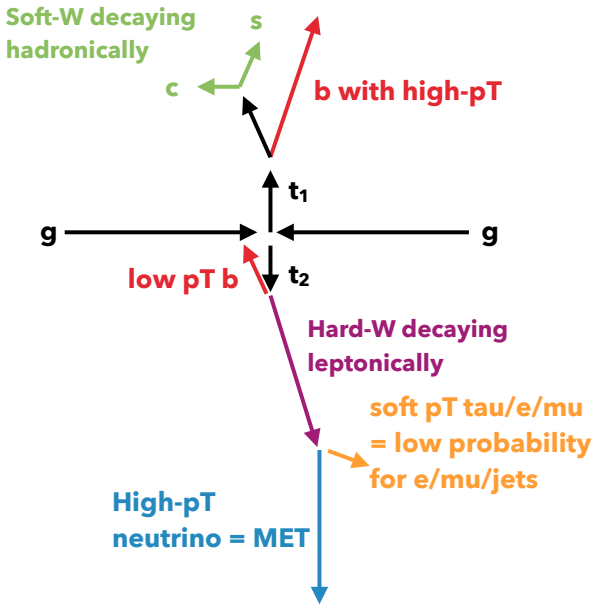


FIGURE 6.10: Schematic topology of $t\bar{t}$ events entering the 0-lepton phase space.

6.3.1.3 Dileptonic ttbar decays

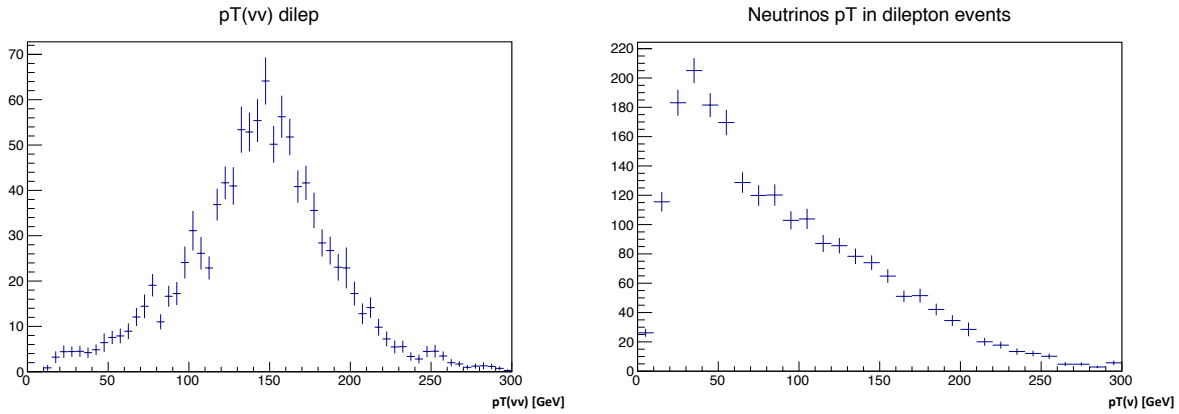
The previous study was extended to $t\bar{t}$ events where both W-bosons decay leptonically : $t\bar{t} \rightarrow bW^{+}(\ell^{+}\nu)\bar{b}W^{-}(\ell^{-}\bar{\nu})$. The flavour of leptons in this process for events satisfying the 0-lepton selection is presented in Table 6.7. About half events have both W-bosons decaying to a τ-lepton, while 40% have a single τ emitted. Again, acceptance, reconstruction and identification efficiency allow for the contamination of events with electrons and muons. The acceptance for events with τ-leptons is increased in the 0-lepton channel, as they can produce a jet in hadronic decays which represent 65% of τ decays.

Lepton flavour	e^{-}	μ^{-}	τ^{-}
e^{+}	3%	3%	12%
μ^{+}	2%	2%	8%
τ^{+}	12%	10%	47%

TABLE 6.7: Fractions of leptons flavour in dileptonic decays of $t\bar{t}$ events entering the 0-lepton phase space.

The distribution of MET generated from neutrinos presents a shape centered at MET $\simeq$ 150 GeV similar to that of semi-leptonic decays, as shown in Fig. 6.11a. The transverse momentum distribution of individual neutrinos in Fig. 6.11b covers a wider range of possible values. Correlations show that often one neutrino has transverse momentum larger than 100 GeV while the other is much softer (pT $\simeq$ 40 GeV). Yet, some events present medium pT neutrinos (50 GeV < pT(ν) < 100 GeV), which means that the neutrinos must be more aligned in order to enhance MET value. Conversely, back-to-back neutrinos have cancelling momentum, with suppressed acceptance to the 0-lepton selection. Yet, neutrinos alignment is not the only cause for increased MET values. Events with both neutrinos satisfying pT(ν) < 100 GeV have an average difference between truth and reconstructed MET of about 40 GeV. On the other hand, events with at least one neutrino satisfying pT(ν) > 100 GeV present an average difference between truth and

reconstructed MET lower than 10 GeV. Mismeasurements of jets energy are expected to cause such discrepancies, and can increase the acceptance for events with increased MET.



(a) (b)

FIGURE 6.11: Transverse momentum distributions of the di-neutrino system (a) and individual neutrinos (b) produced in top-quarks decay.

This specific topology of $t\bar{t}$ events leads to final states with two b-quarks, soft charged leptons and neutrinos which must be quite aligned depending on their transverse momentum balance, as represented schematically in Fig. 6.12.

6.3.2 $t\bar{t}$ decays in sensitive multivariate discriminant bins

Table 6.8 presents the fractions of semi- and dileptonic decays in the most sensitive region of the 0-lepton multivariant discriminant distribution. Semi-leptonic events have the largest contribution in both 2- and 3-jets categories, contrarily to the inclusive numbers shown in Table 6.6. This could motivate the use of semi-leptonic $t\bar{t}$ samples with filters relying on top-quark decays for future improvements of the SM VH(bb) search with respect to MC statistical uncertainties.

6.3.3 Reconstructed τ-jets in $t\bar{t}$ events

As presented in the previous sections, most of the $t\bar{t}$ background originates from events with a τ-lepton produced in the decay of W-bosons. The identification of τ-jets in the SM VH(bb) search relies on the medium efficiency Working Point of the dedicated τ-selection based on a BDT discriminant with electron Overlap Removal. The fractions of reconstructed τ-jets multiplicity are presented in Table 6.9 for events entering the 0-lepton 2- and 3-jets categories.

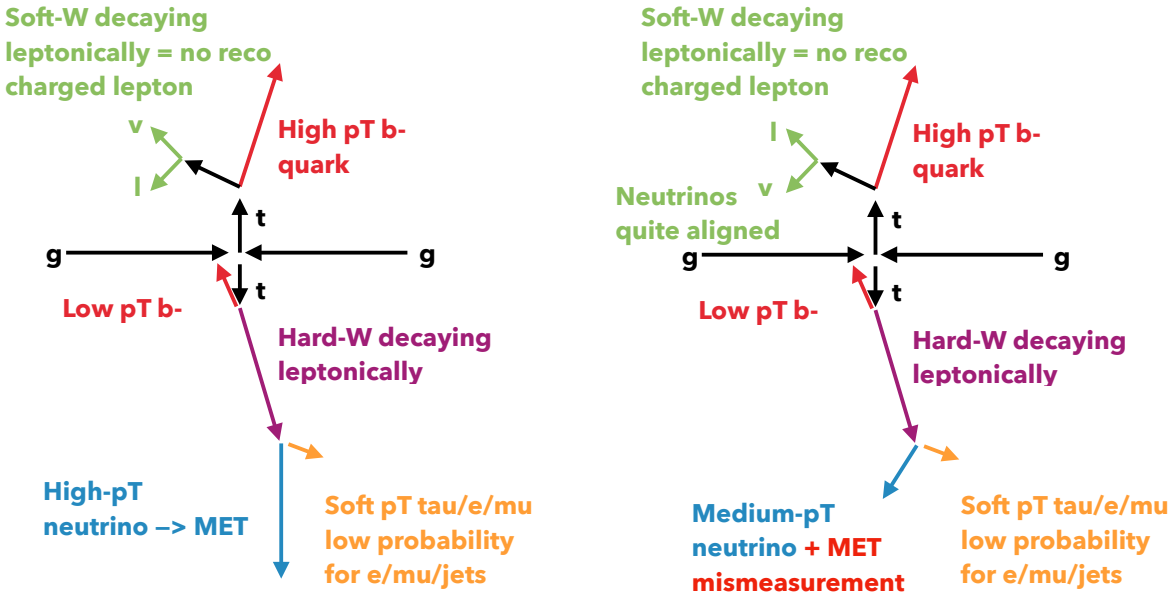


FIGURE 6.12: Schematic topology of $t\bar{t}$ events with dilepton decay entering the 0-lepton phase space. Figure (a) corresponds to the topology where a single neutrino is responsible for most of the generated MET. Figure (b) draws the toplogy for events with soft neutrinos, with MET value enhanced by mis-reconstruction effects.

Table 6.10 presents the fractions of reconstructed τ-jets matched to a truth τ-lepton in MC simulations for $t\bar{t}$ and WH(bb) events. Reconstructed τ-jets very often originate from hadronic τ-leptons for both processes. Tests for removing events with an identified τ-jet were performed in the analysis selection, and did not show a significant improvement.

6.3.4 Flavour composition

The $t\bar{t}$ background is caracterized by a large multiplicity of jets caused by the potentially large number of quarks following the top quark and W bosons decays in the 0-lepton channel. The contamination from fully hadronic $t\bar{t}$ decays is negligible in the 2 and 3 jets regions. Semi-leptonic events are expected to produce about four jets. Acceptance and reconstruction effects can increase the background contamination in the Signal Region, particularly in the 3-jets category. Especially, the loss of a b-quark favours the selection of events in which a c-jet is mistagged as a b-jet. This is also true for light-jets to a lower extent, due to the better rejection of light-jets with b-Tagging. Jets flavour composition is shown in Table 6.11 using the label of jets after full selection. This result allows to measure the potential gain in rejecting events with c- and light-labelled jets by improving b-Tagging algorithms.

6.3.5 Discrimination of $t\bar{t}$ events using track-jets and MET soft term

The high density particles in $t\bar{t}$ events is expected to increase the hadronic activity measured in calorimeters and the charged component identified using the Inner Detector. Especially, the track-jets multiplicity is expected to be enhanced with respect to signal processes and the other backgrounds. The MET soft term was also tested to improve the discrimination against the $t\bar{t}$ background. The variables described in the following were added to the nominal set of BDT training variables of the 0-lepton channel :

	Fraction of decays			
Bin	sing. lepton	dilepton	nEntries	s/sqrt(b)
4	58 %	42 %	19	0,69
3	87 %	13 %	15	0,88
2	63 %	38 %	8	1,01
1	89 %	11 %	9	1,17
0	100 %	0 %	4	1,27
4	76 %	24 %	114	0,44
3	79 %	21 %	82	0,54
2	83 %	17 %	46	0,65
1	80 %	20 %	30	0,76
0	100 %	0 %	25	0,93

TABLE 6.8: Fractions of semi- and dileptonic decays in the most sensitive region of the 0-lepton multivariant discriminant distribution. The most sensitive BDT bin after statistical rebinning has index 0. The fractions are computed with respect to the total number of unweighted events in the corresponding bin, denoted as nEntries.

	τ-jets multiplicity		
Number of jets	0	1	2
2 jets	73%	25%	2%
3 jets	82%	17%	1%

TABLE 6.9: Fractions of τ-jets multiplicities in $t\bar{t}$ events depending on the jets multiplicity in the 0-lepton phase space.

- Inclusive number of track jets with transverse momentum greater than 7 GeV.
- Number of track jets within/outside the ellipse defined with respect to the b-Tagged dijets axis. This variable was inspired from the CMS analysis, as a simplistic definition of colourflow. A proper description and test of colourflow is presented in Section 6.4.
- MET soft term, defined as the scalar sum of tracks unassociated to the reconstructed objects.
- Scalar/Vector sum of track-jets transverse momentum. These variables provide alternative definitions of HT and MET using track-jets only.
- The number of fat-jets reconstructed in the event was tested with no strong motivation. Fat jets are reconstructed using the Anti-kT algorithm with radius parameter R=1.0, and are expected to match systems of near-by jets of radius 0.4. The decay of boosted top-quarks should lead to such topologies, but concerns a minor fractions of the overall $t\bar{t}$ background.

Process	Reconstructed τ-jet matching test			
	Hadronic truth-τ within dR < 0,2	Truth e/μ within dR < 0,2	Parton within dR < 0.4	Unmatched
$t\bar{t}$	90%	0%	5%	5%
WH(bb)	95%	1%	2%	2%

TABLE 6.10: Fractions for matching reconstructed τ-jets to hadronically decaying τ-leptons, electrons or muons or partons (quarks) in MC events with one reconstructed τ-jet and entering the 0-lepton phase space. The matching procedure relies on angular separation tests, with priority given to hadronic truth-τ, then truth e/μ, and partons (quarks matched to a jet)

2 jets	bb	bc	bl	btau
150 < MET < 200 GeV	85.6 %	11.2 %	2.5 %	0.6 %
MET > 200 GeV	66.4 %	25.4 %	5.9 %	1.6 %
3 jets	bb	bc	bl	btau
150 < MET < 200 GeV	85.8 %	11.1 %	2.3 %	0.7 %
MET > 200 GeV	78.8 %	16.9 %	3.8 %	0.6 %

TABLE 6.11: Flavour composition of $t\bar{t}$ events entering the 0-lepton Signal Regions, with further MET slicing

The statistical significance increases by 1.7% with the re-trained BDT, which is not sufficient to motivate a change in the analysis. An alternative strategy consists of training dedicated BDTs against $t\bar{t}$ events and other backgrounds separately. No easy gain was found from this approach, which would require optimised BDT parameters and cuts for each individual multivariate method. The nominal BDT trained inclusively against all EW backgrounds already allows for a very efficient rejections of most $t\bar{t}$ events.

As a conclusion, a proper inclusion of these variables would probably require additional modelling systematics, which would dilute the small gain in significance. A similar study was performed, in which several multivariate discriminants were trained each against one of the dominant background processes ($t\bar{t}$, Z+jets and W+jets). An optimisation of cuts based on statistical significances showed no gain by separating the background sources in the BDT training procedure.

6.4 Colourflow

A possible improvement to the current SM VH(bb) search consists of exploiting the colour flow between QCD partons. Colourflow derives from the conservation of colour charges in strong interactions as described by QCD. Quarks pairs produced in the decay of a colour singlet such as the Higgs boson produce a different pattern of hadrons than quarks originating from colour charged particles such as gluons, which are connected to the underlying event. This allows to compute event superstructure variables, to potentially discriminate signal from background events. In the SM VH(bb) search, signal events have colour connected b-quarks at Matrix-Element level, contrary to $t\bar{t}$ or single-top events. In ATLAS, this effect can be measured from calorimeter clusters and from tracks matched to jets [81].

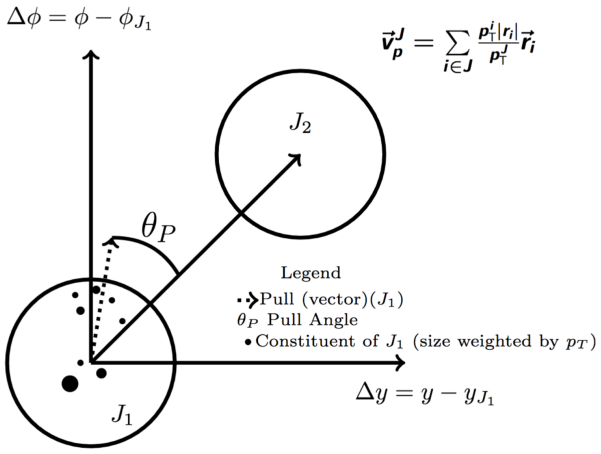


FIGURE 6.13: Schematic view of a jet's pull vector computation

6.4.1 Colourflow observables

The base object of colourflow studies is the jet pull-vector [82], defined from the weighted vector sum of vectors $\overrightarrow{r}_i$ binding the jet (J) center to individual jet constituents (i), as illustrated in Fig. 6.13.

$$\overrightarrow{v}_P^J = \sum_{i \in J} \frac{p_T^i |r_i|}{p_T^J} \overrightarrow{r}_i \tag{6.1}$$

The pull angle $\Delta\Theta$ of a jet is defined as the angular separation between the jet pull vector and the line connecting its center to that of a second jet. Given a pair of jets, two distinct pull angles can be computed, which allows to exploit their correlation for further discrimination.

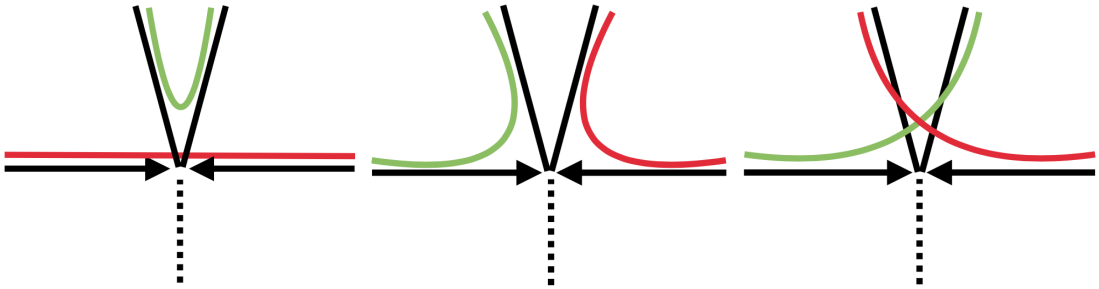
In addition, the norm of the pull-vector of a jet can be exploited to enhance signal to background separation. However, this quantity is expected to provide little improvement and is not presented in the following.

6.4.2 TruthAnalysis with a single-event

6.4.2.1 Analysis setup

A preliminary estimate of the discrimination power of colourflow variables is presented in the following. A simple analysis was performed at truth-level in order to mimic the SM VH(bb) selection without reconstruction and systematics effects. A single Madgraph ZH $\rightarrow \nu\bar{\nu}b\bar{b}$ event was picked at Matrix Element level, and its colourflow configuration modified. Three colourflow topologies were tested in order to measure the maximal impact of the hadronisation process on colourflow observables :

- Colour-singlet : this configuration leads to a pair of colour-connected b-quarks in the final state, carrying opposite colour charges (blue-$\overline{\text{blue}}$, red-$\overline{\text{red}}$, green-$\overline{\text{green}}$)
- Colour-charged : this topology leads to a pair of b-quarks in the final state, each carrying different colour charges (e.g. blue-$\overline{\text{red}}$, green-$\overline{\text{blue}}$, ...). The colour flow connects b-quarks to the beam remnants.
- Reverted colour-charged : this configuration is similar to the previous "Colour-charged" topology with the colour connections of b-quarks to beam remnants inverted.



(a) Colour singlet (b) Colour charged (c) Inverted colour charged

FIGURE 6.14: Schematic view of tested colour-flow configurations.

Variable	Value
m_{bb}	125 GeV
$\Delta R(b,b)$	1.37
pT(Z)	159 GeV
$\Delta\Phi(V,bb)$	2.7
p_T(B1)	89 GeV
p_T(B2)	96 GeV

FIGURE 6.15: Observables of the event used.

These colour configurations are illustrated in Fig. 6.14. Especially, the colour-charged and reverted colour-charged topologies are identical but reverting the momentum of quarks and corresponding anti-quarks produced in the boson decay.

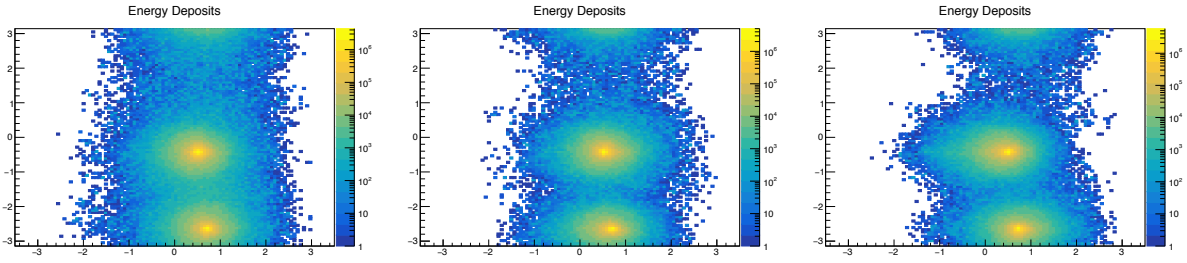
The single-event was picked for its signal-like topology with respect to the 0-lepton channel SM VH(bb) requirements, as shown in Table 6.15. A simple selection is performed at truth-level in order to mimic the analysis without reconstruction effects and systematics :

- Jets are reconstructed from interacting truth-particles using the Anti-kT algorithm with distance parameter R=0.4. Jets with p_T > 20 GeV, $|\eta|$ < 2.5 are preselected.
- Two b-labelled jets are required. A jet is b-labelled if a B-hadron is standing in the jet area within a cone of radius parameter 0.4 from the jet center.
- The dijet mass of b-labelled jets is required to stand in the region 100 GeV < m_{bb} < 150 GeV.
- MET > 130 GeV (allows enhanced stat. with respect to the nominal cut at 150 GeV, as the vector boson only has 159 GeV transverse momentum).
- $\Delta\Phi(b,b) < 140°$
- $\Delta\Phi(MET,bb) > 120°$
- $\min(\Delta\Phi(MET,b)) > 20°$

The set of cuts described above defines the tight selection of the analysis, while a looser selection made simply requiring two b-labelled jets in order to gain in statistics.

6.4.2.2 Results

All three colourflow topologies were hadronized 100,000 times each with Pythia 8 and Herwig 7. Figure 6.16 represents the location of final state stable particles in the η-ϕ plane, superimposed for all hadronisations performed with Pythia 8, for each colour topology respectively.



(a) Colour singlet (b) Colour charged (c) Inverted colour charged

FIGURE 6.16: Illustration of the jet constituents repartition for each tested colourflow configurations. The x- (y-) axis correspond to the η- (ϕ-) coordinates.

The corresponding distributions of the pull angle for leading and sub-leading jets are presented in Fig. 6.17 for the Pythia 8 and Herwig 7 Parton Shower models. Different computations of colourflow are tested, exploiting either stable particles entering jets constituents, i.e. candidate to interactions with the hadronic calorimeter (neutral colourflow) or charged stable particles which would be reconstructed as tracks in the Inner Detector (charged colourflow). Colour singlet processes are expected to show pull-angles peaking around zero, which reflects the alignment of the pull vector of one of the jets with the dijet axis. Colour-charged processes do not present such a feature, and have pull-vectors collinear to the beam axis, hence a misalignment with respect to the dijet axis. Also, reversing the colour connection shifts the peak by a value of $\pm\pi$. Finally, it is noticeable that charged colourflow worsens the discrimination between the colour singlet configuration and the colour charged topologies.

6.4.3 TruthAnalysis with ATLAS MC samples

The previous selection is now performed on Monte-Carlo samples with multiple events. Distributions of the neutral and charged colourflow pull-angle for the leading jet are presented in Fig. 6.18 and 6.19 respectively. They show $ZH(\nu\bar{\nu}b\bar{b})$ signal events and a Z+Jets dataset generated with a b-jet filter and requiring 140 GeV < p_T^Z < 280 GeV. This sample is the largest background in the most sensitive category of the 0-lepton channel in the SM VH(bb) search. The background process shows a flat shape with respect to that of the signal. This is expected due to the different kinematic of events entering the analysis phase space (conversely to the single-event study).

An intriguing feature is the peaking shape of the pull angle distribution of Z+jets events for the leading jet in the Loose selection. This feature partially cancels when moving to the tight selection, and it was shown that the requirement 100 GeV < m_{BB} < 150 GeV is responsible for this change. This probably reflects the correlation of MET and the dijet mass with the jets angular separation. It directly relates to the overlap between jets in the $\eta - \phi$ plane and the number of constituents that can be interchanged.

6.4.4 Analysis at reconstructed level

In order to confirm the discrimination showed at Truth level, the analysis should be extended at reconstructed level. However, calorimeter clusters composing jets are early removed from files in the successive steps of derivations intended to physics-analysis purpose. A compromise is to use charged-tracks associated to the jets with Ghost-matching in order to compute a "charged" colourflow. This approach is also more resilient to pileup, due to the association of charged tracks to the primary vertex. The results showed in the following were obtained using this latter approach.

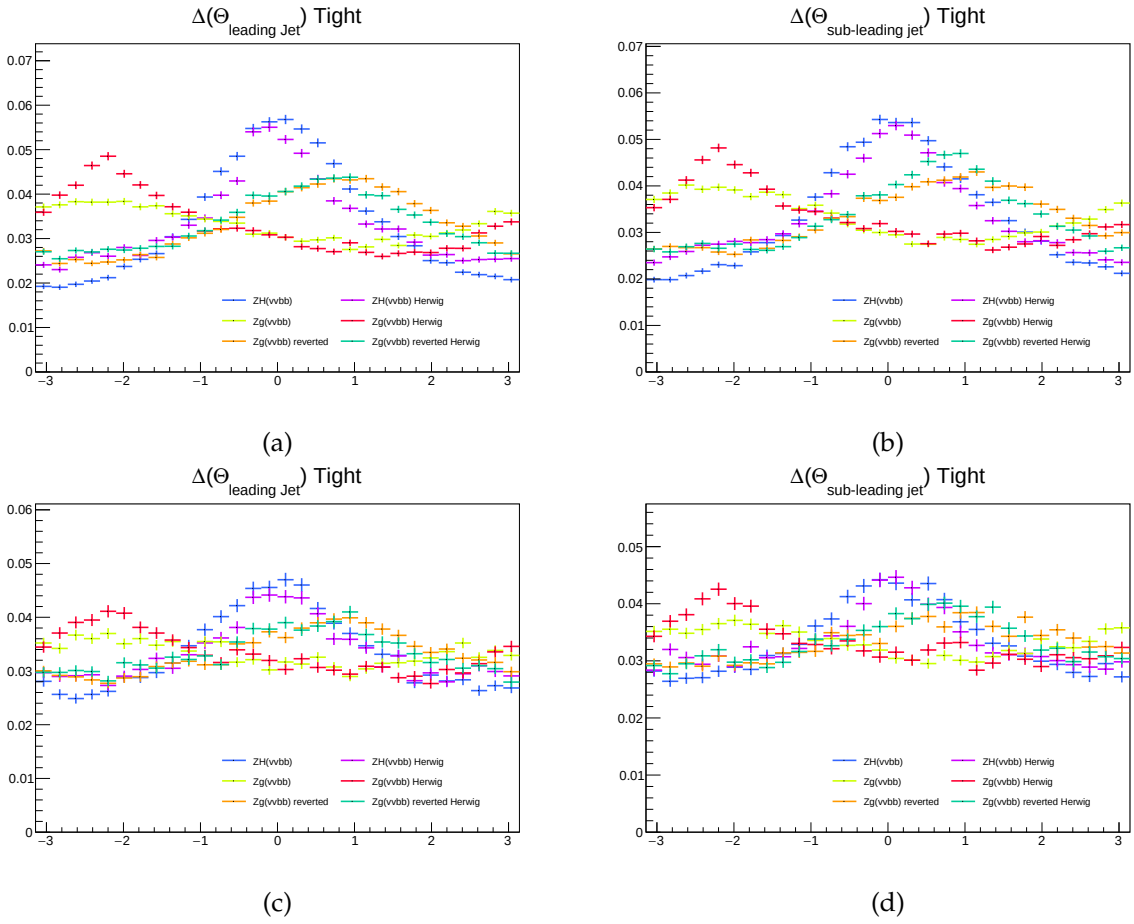


FIGURE 6.17: Pull angle distribution for colourflow exploiting stable particles entering jets constituents (Fig. A,B) or charged particles only (Fig. C,D). Figures show for the leading (Fig. A,C) and sub-leading (Fig. B,D) jets the distribution for each tested colour-flow configurations and Pythia 8 or Herwig Parton Shower programs.

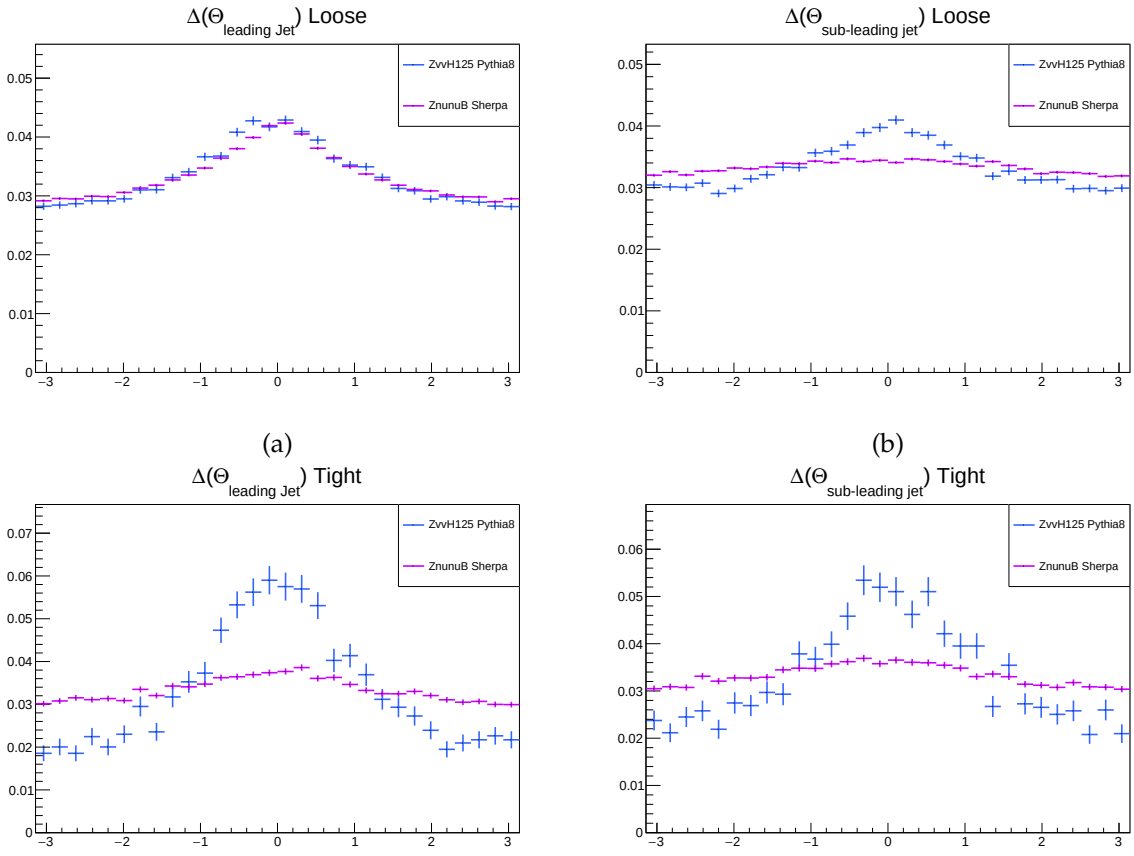


FIGURE 6.18: Pull-angle distribution of the leading jet in ZH $\rightarrow \nu\bar{\nu}b\bar{b}$ (blue) and Z+jets events (purple) in a computation of colourflow with all interacting stable particles.

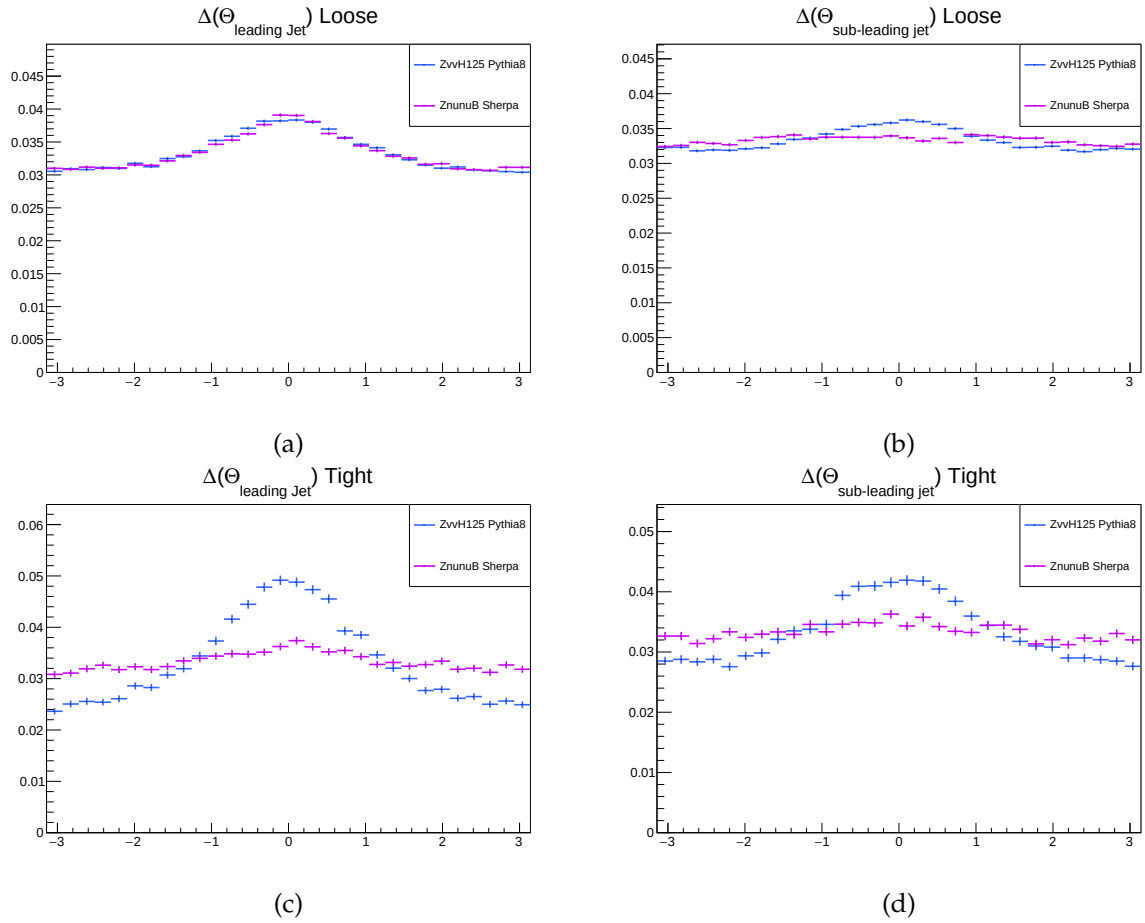
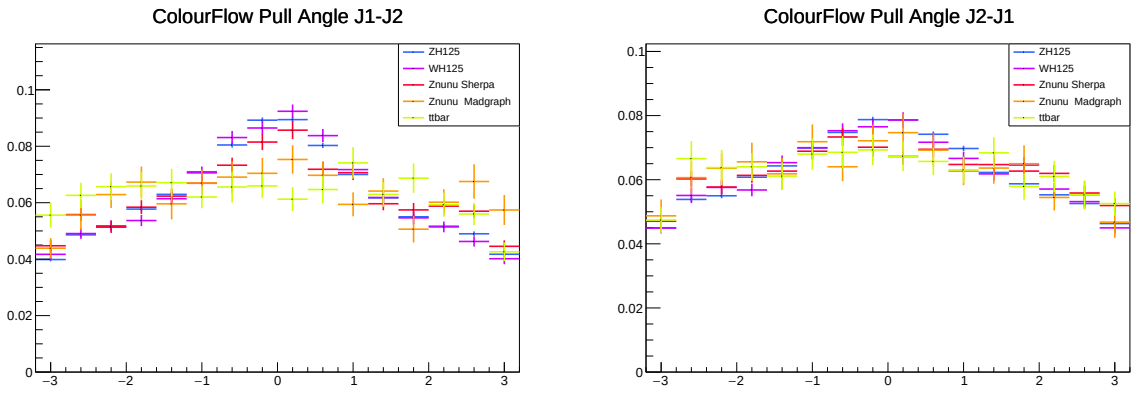


FIGURE 6.19: Pull-angle distribution of the leading jet in ZH $\rightarrow \nu\bar{\nu}b\bar{b}$ (blue) and Z+jets events (purple) in a computation of colourflow with charged particles only.



(a) Leading jet (b) Subleading jet

FIGURE 6.20: Charged colourflow pull-angle distribution at reconstructed level in events with two jets. The Jx-Jy notation advises the pull angle of jet x is computed with respect to the x-y dijet-axis.

The analysis is now performed at reconstructed level using tracks and the standard MVA selection described in Sec. 5.4. The ZH $\rightarrow \nu\bar{\nu}b\bar{b}$, WH $\rightarrow \ell\bar{\nu}b\bar{b}$, $t\bar{t}$ and Z+jets processes were considered. Pull angle distributions are presented for events with two and three jets, respectively in Fig. 6.20 and 6.21. Events with three jets allow for an important number of combinations for different jets pairs. This combinatoric approach aims at discriminating events with mis-identified jets with b-Tagging. A minor discrimination is visible from the shape of the pull angle distribution in ttbar events, but training the SM VH(bb) 0-lepton multivariate discriminant with colourflow information did not show any improvement. Furthermore, systematics effects are expected to decrease the discrimination provided by colourflow techniques. Therefore, the investigations to use colourflow in the VH(bb) analysis were stopped.

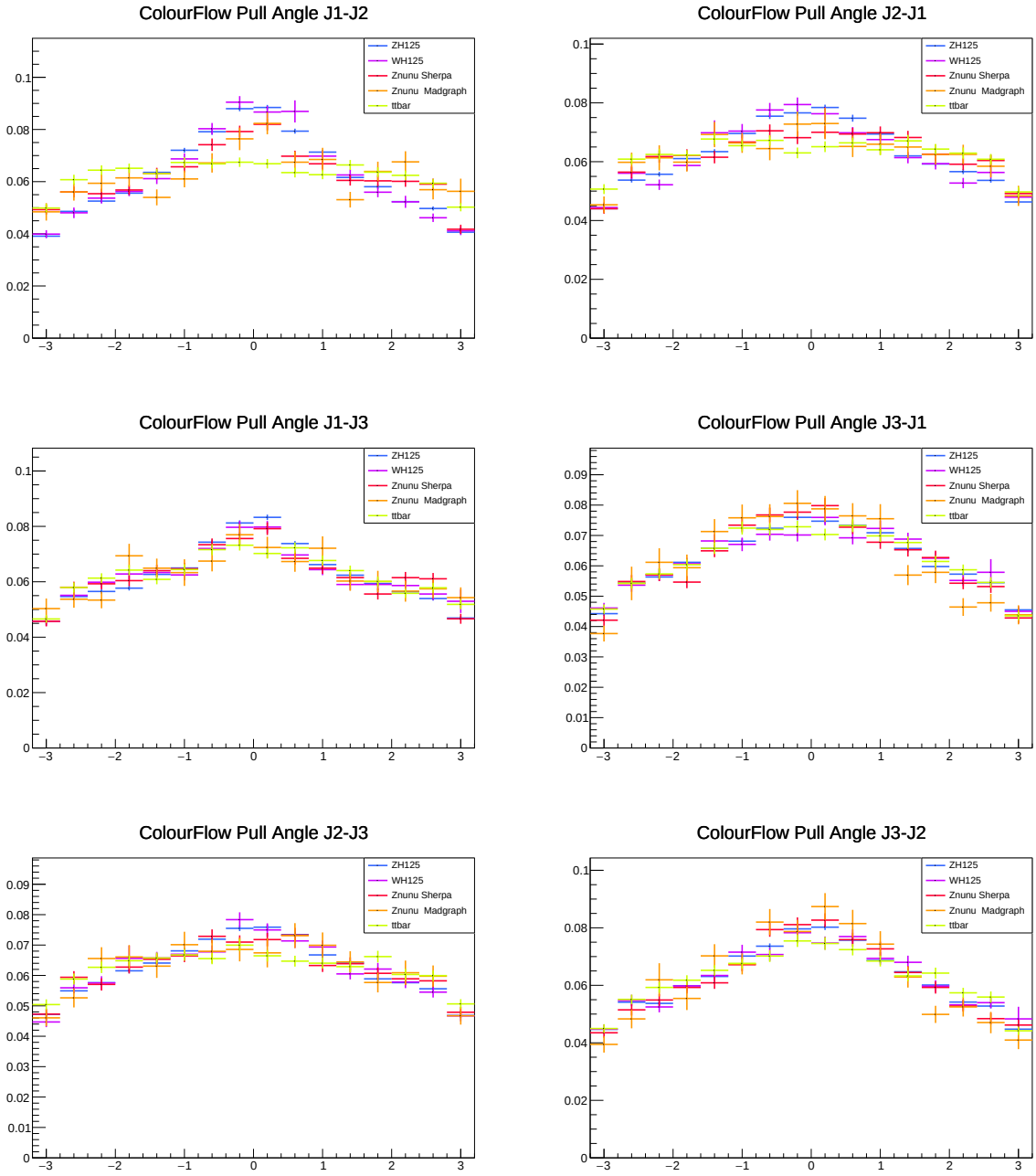


FIGURE 6.21: Charged colourflow pull-angle distribution at reconstructed level in events with three jets. The Jx-Jy notation advises the pull angle of jet x is computed with respect to the x-y dijet-axis.

Chapter 7

Search for top quark decays $t \to qH$ with $H \to \gamma\gamma$

A search for Flavour Changing Neutral Currents in the decay of a top quark into an up-type (q = u, c) quark and a Higgs boson decaying into two photons was performed at 13 TeV in the ATLAS experiment and published in 2017 [83]. This chapter presents optimisation studies performed after this publication, aiming at improvements on the upper limit of the tcH and tuH couplings.

7.1 Introduction

7.1.1 Motivation

The discovery of the Higgs boson in 2012 by the ATLAS and CMS collaborations allows for the search of new interaction processes not predicted in the Standard Model. Especially, the abundant production of top-quarks at the LHC motivates the search for exotic top-quark decays involving the tqH coupling. At tree level, the Standard Model only allows the decay of the top-quark to down-type quarks in association to a W-boson, dominated by the $t \to bW$ process. Higher order Electroweak processes involve loops and have their cross-section strongly suppressed as predicted by the Glashow-Iliopoulos-Maiani mechanism. Diagrams with a down-type quark in the loop have further phase-space restrictions due to the mass difference with respect to the W-boson. The amplitude for the Matrix Element of such processes incorporates a factor $(m_b/m_W)^4$. In contrast, processes like $b \to s\gamma$ allow for top-quarks in the loop hence a larger cross-section proportional to $(m_t/m_W)^4$.

Flavour Changing Neutral Currents enable top-quark to up-type quarks decays, such as $t \to qZ$, $t \to q\gamma$ and $t \to qH$ with $q = u, c$ as illustrated in Fig. 7.1.

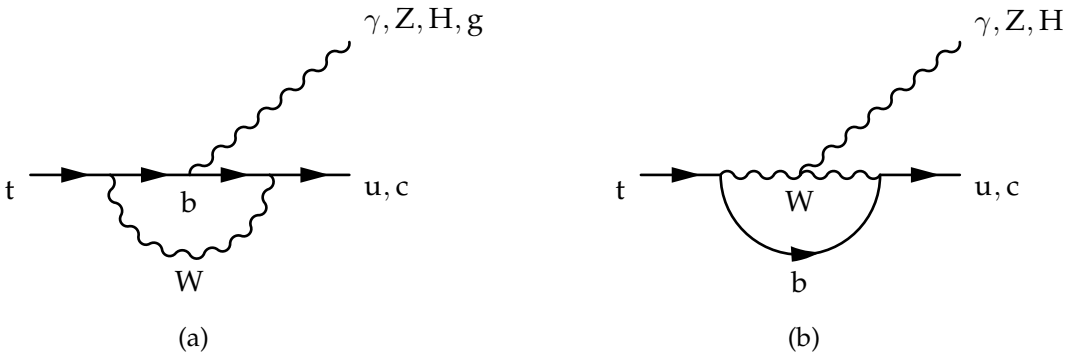


FIGURE 7.1: Examples of Standard Model 1-loop diagrams contributing to FCNC top decays. In Diagram A, the photon, Z or Higgs boson can be attached to any side of the fermion line. Especially, the Higgs boson coupling to top quarks is favoured.

Process	Model							
	SM	QS	FV-2HDM	FC-2HDM	MSSM	RPV SUSY	TC2	RS
$t \to u\gamma$	$3.7 \cdot 10^{-16}$	$7.5 \cdot 10^{-9}$	x	x	$2 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	x	$\sim 10^{-11}$
$t \to uZ$	$8 \cdot 10^{-17}$	$1.1 \cdot 10^{-4}$	x	x	$2 \cdot 10^{-6}$	$3 \cdot 10^{-5}$	x	$\sim 10^{-9}$
$t \to ug$	$3.7 \cdot 10^{-14}$	$1.5 \cdot 10^{-7}$	x	x	$8 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	x	$\sim 10^{-11}$
$t \to uH$	$2 \cdot 10^{-17}$	$4.1 \cdot 10^{-5}$	$5.5 \cdot 10^{-6}$	x	10^{-5}	$\sim 10^{-6}$	x	x
$t \to c\gamma$	$4.6 \cdot 10^{-14}$	$7.5 \cdot 10^{-9}$	$\sim 10^{-6}$	$\sim 10^{-9}$	$2 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$\sim 10^{-6}$	$\sim 10^{-9}$
$t \to cZ$	$1 \cdot 10^{-14}$	$1.1 \cdot 10^{-4}$	$\sim 10^{-7}$	$\sim 10^{-10}$	$2 \cdot 10^{-6}$	$3 \cdot 10^{-5}$	$\sim 10^{-4}$	$\sim 10^{-5}$
$t \to cg$	$4.6 \cdot 10^{-12}$	$1.5 \cdot 10^{-7}$	$\sim 10^{-4}$	$\sim 10^{-8}$	$8 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$\sim 10^{-4}$	$\sim 10^{-9}$
$t \to cH$	$3 \cdot 10^{-15}$	$4.1 \cdot 10^{-5}$	$1.5 \cdot 10^{-3}$	$\sim 10^{-5}$	10^{-5}	$\sim 10^{-6}$	x	$\sim 10^{-4}$

TABLE 7.1: Theoretical values for the branching fractions of FCNC top quark decays predicted by the SM and exotic extensions [103].

Models Beyond the SM predict the existence of new bosons interacting with SM particles. Loop-diagrams mediated by such new bosons would enhance the cross-section for FCNC interactions, and indirectly attest for the existence of new physics. Such models are presented in the following, and the theoretical values for the branching fractions of FCNC top quark decays shown in Table 7.1.

- Quark-singlet model (QS) [84] [85] [86]
- Two-Higgs doublet model [87] [88] [89] [90]
  - Two-Higgs doublet model with flavour-conservation (FC-2HDM)
  - Two-Higgs doublet model without flavour-conservation (FV-2HDM)
- Minimal supersymmetric model (MSSM) [91] [92] [93] [94] [95]
- SUSY with R-parity violation [96] [97]
- Topcolour-assisted Technicolour model (TC2) [98]
- Warped extra dimensions (RS) [99] [100]
- Little Higgs model with T-parity conservation [101]
- Composite Higgs models [102]

7.1.2 Signal, background processes and signatures

The search for Flavour Changing Neutral Currents presented in the following aims at characterizing events in which a top quark transforms into an up-type (q = u, c) quark or reciprocally, with the emission of a Higgs boson decaying into two photons. The main channels for top-quark production in proton-proton collisions either lead to the production of a single top or pair of top quark(s). The SM decaying top leads to the production of jets and leptons resulting from the hadronisation of a b-quark and W-boson decay. Both leptonic and hadronic W-boson decay channels are considered in the analysis. In the pair production mode, the Higgs boson is emitted in the exotic decay of the second top-quark $t \to qH$. This production channel adds complexity to the final state with an additional jet.

7.1.2.1 Signal processes

Signal processes for the search of FCNC in the decay of a top quark into an up-type (q = u, c) quark and a Higgs boson decaying into two photons arise from two distinct processes :

- SM pair-production of top quarks, followed by separate decays t $\to$ qH and $\bar{\text{t}}$ $\to$ $\bar{\text{b}}$W (or charge conjugate process), followed by the decay H $\to$ $\gamma\gamma$. The Feynman diagram for this process is presented in Fig. 7.2 at leading-order. The inclusive NNLO+NNLL pair-production cross-section is $\sigma_{t\bar{t}} = 831.76$ pb for $m_{top} = 172.5$ GeV.

  The probability for two top quarks decaying in t$\bar{\text{t}}$ $\to$ qH$\bar{q}$H is negligible, given the currrent limits $B(\text{t} \to \text{cH}) < 2.2 \times 10^{-3}$ and $B(\text{t} \to \text{uH}) < 2.4 \times 10^{-3}$ measured in the Run 2 analysis summarised in Section 7.1.3.

- Single-top production in the interaction qg $\to$ tH, followed by the decays H $\to$ $\gamma\gamma$ and t $\to$ bW. Fig. 7.3 illustrates the Feynman diagram for this process at leading-order.

  The different cross-sections and kinematics allow to disentangle the tcH and tuH couplings in this production mode. In single-top events, the parton q involved in the tqH interaction arises from protons PDF, which increases the probability for tuH with respect to tcH. However, valence up-quarks enhance the production of the Higgs boson with large rapidity for which the acceptance is lower. The NLO cross-section for these processes are respectively $\sigma_{\text{ug}\to\text{tH,tH}\to\text{bW}\gamma\gamma} \simeq 1.8$ fb and $\sigma_{\text{cg}\to\text{tH,tH}\to\text{bW}\gamma\gamma} \simeq 0.3$ fb assuming B(t $\to$ qH = 10^{-3}).

  For each of these processes, a single chirality operator must be considered out of two. In the ug $\to$ tH process for instance, they are encoded in the effective Lagrangian as :

  1. $\mathcal{L} \propto (\phi^\dagger\phi - v^2/2)\bar{t_L}\tilde{\phi}u_R$
  2. $\mathcal{L} \propto (\phi^\dagger\phi - v^2/2)\bar{u_L}\tilde{\phi}t_R$

  The chirality of the operator does not change the cross-section of the process, and negligible differences were observed in the resulting kinematics.

Signal yields estimated from MC simulations are normalised to the appropriate top-quark production mode cross-section which multiply several decay term probabilites. The branching ratio for a Higgs boson decaying to photons is predicted to $B(\text{H} \to \gamma\gamma) = (2.27 \pm 0.07) \times 10^{-3}$ for $m_H = 125.09$ GeV in SM predictions, which is in agreement with experimental measurements. The probability for the t $\to$ qH decay is subtracted from the inclusive branching ratio for the SM decay t $\to$ bW. Hence, the inclusive cross-section used in the normalisation of signal samples reads :

$$\sigma_{\text{t}\to q\text{H}(\gamma\gamma)} = \sigma_{\text{top-production}} \times B(\text{H} \to \gamma\gamma) \times 2 \times B(\text{t} \to q\text{H}) \times (1 - B(\text{t} \to q\text{H})) \tag{7.1}$$

The analysis is split into two channels, according to the presence of a lepton originating from the decay of the W-boson in the reconstructed final state. W-boson decay branching fractions $B(\text{W} \to q\bar{q}) = 67.6\%$ and $B(\text{W} \to \text{e}^-\nu_\text{e}, \mu^-\nu_\mu) = 21.64\%$ lead to different expected signal yields. Furthermore, the leptonic and hadronic channels of the analysis present very different purity of signal with respect to background yields due to the diphoton+jets background which has major contribution in the hadronic channel. Further categories can be designed, based on the number of reconstructed jets, in order to discriminate pair-production of top quarks against single-top production. This latter separation allows to disentangle tuH from tcH interactions, exploiting different expected yields and events kinematic, as described in Section 7.4.1.

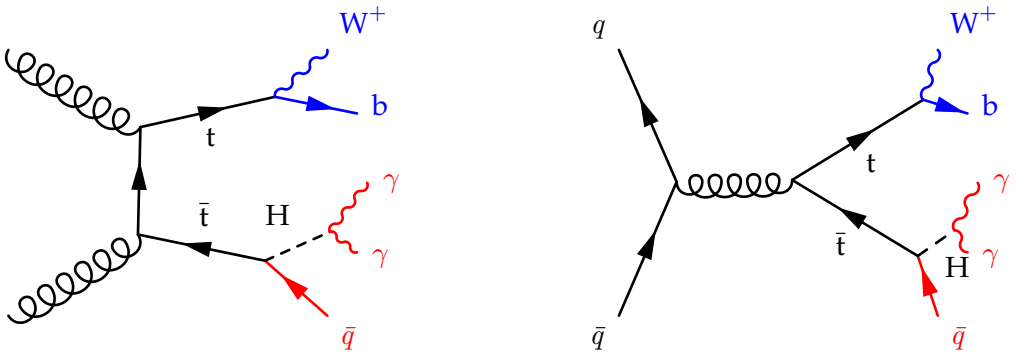


FIGURE 7.2: Possible Feynman diagrams for the top-quark pair production FCNC signal process.

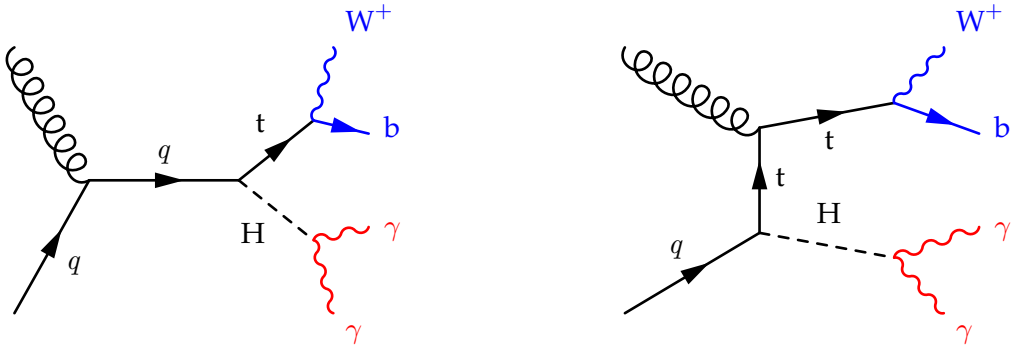


FIGURE 7.3: Possible Feynman diagrams for the single-Top FCNC signal process.

7.1.2.2 Background processes

Several SM processes contribute to the analysis background. Their respective impact on the analysis sensitivity mainly results from their cross-section, cross-section uncertainty, and diphoton mass shape. In the hadronic channel, processes leading to the production of two photons and jets have the largest cross-section amongst all considered backgrounds. In the leptonic channel, the production of a photon in association to a pair of top-quarks, where an electron originating from the decay of a top-quark can be misidentified as a photon for instance, yields to the most signal-like topologies amongst non-resonant backgrounds.

Both hadronic and leptonic channels are affected by SM Higgs production processes. The resonant shape of the diphoton mass distribution for these backgrounds is degenerated with that of the FCNC signal. Hence, a special care should be provided to the yield estimate of these processes. Besides, uncertainties on the production cross-section for these processes will impact the sensitivity of the FCNC analysis.

- SM Higgs production processes can present final states similar to that of the FCNC signal. Especially, in the ttH production channel, objects acceptance and reconstruction effects can lead to the multiplicity of jets and leptons expected in signal events. The efficiency for this specific channel is enhanced by the production of top-quarks, similarly to the FCNC signal.

- Diphoton events with additional partons constitute the majoritary background of the analysis, both in the hadronic and leptonic channels. Possible Feynman diagrams for such processes are illustrated in Fig. 7.4. The selection allows for little acceptance to this non-resonant background, but its large cross-section enhances its impact on the analysis

sensitivity. Last, $W(Z)\gamma\gamma$, $t\bar{t}\gamma$ and $t\bar{t}\gamma\gamma$ processes complete the picture of non resonant backgrounds. The monotonous shape of the diphoton mass distribution for these backgrounds allows yields in the sensitive region of the analysis to be estimated from fits in the sidebands of the the $m_{\gamma\gamma}$ distribution, as described in the analysis overview in Section 7.1.2.4

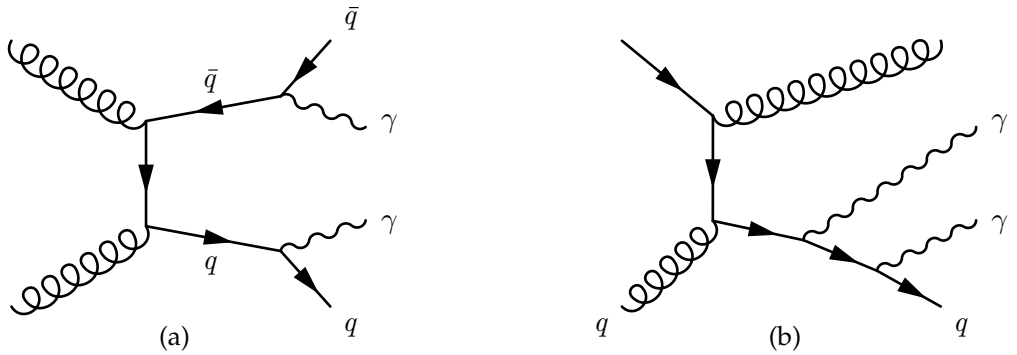


FIGURE 7.4: Possible Feynman diagrams for the $\gamma\gamma$ + jets background process. Diagram (A) can yield to the production of top-quarks which constitues the $t\bar{t}$ + $\gamma\gamma$ background.

7.1.2.3 Expected signatures

The separation of signal and background events exploits reconstructed objects which are expected in the final state of targetted signal processes :

- A pair of photons with diphoton mass $m_{\gamma\gamma} \simeq m_H$.
- A jet resulting from the hadronisation of a b-quark.
- In hadronic W-boson decays, two additional jets are expected.
- In leptonic W-boson decays, a single charged lepton and MET resulting from the unreconstructed neutrino.
- In the top-quark pair-production mode, an additional jet is expected from the hadronisation of the quark from the exotic top decay.

Additional signatures and kinematic requirements on systems of jets and photons allow to target top-quark candidates and reject combinatoric backgrounds. The analysis also exploits individual kinematic measurements and flavour-tagging properties of jets. Hence, the analysis results strongly depend on the performance and reliability of jets reconstruction and flavour-tagging techniques.

7.1.2.4 Overview of the analysis strategy

The search for Flavour Changing Neutral Currents in the decay of a top quark into an up-type (q = u, c) quark and a Higgs boson decaying into two photons is improved with respect to the 2017 publication summarised in Section 7.1.3. Limits on the Branching ratio for the $t \to qH$ decay are measured from yields estimates for signal-candidate events. Optimisation studies are performed using Monte-Carlo predictions, while the results are extracted with data collected in the ATLAS detector, as described in Section 7.2. Objects reconstructed in the ATLAS detector

are pre-selected using requirements presented in Section 7.3. Selected objects are exploited in the selection and categorisation of candidate events, as described in Section 7.4. The analysis benefits from the specific topology of signal processes by using multivariate techniques which allow to further discriminate signal-candidate events. The yield of signal-candidate events in data is measured in the mass range centered about $m_H = 125$ GeV within a few GeV from a fit to the diphoton mass distribution in the range 100 GeV < $m_{\gamma\gamma}$ < 160 GeV to determine and subtract the background.

7.1.3 The Run 2 publication

In 2017, a search for the t $\to$ qH with H $\to \gamma\gamma$ was performed in the ATLAS experiment [83] [104]. This analysis relies on data collected in 2015 and 2016 at $\sqrt{s}$ = 13 TeV with integrated luminosity 36.1 fb^{-1}. The subdominant single-top production mode is not considered with the other FCNC signals. The next sections describe the events selection, fit procedure and results of the analysis.

7.1.3.1 Selection and categories

The selection is designed in order to target different final states of the Standard Model decaying top quark t $\to$ bW. The leptonic channel aims at the process $t\bar{t} \to q\mathrm{H}(\gamma\gamma)\mathrm{bW}^{\pm}(\ell^{\pm}\nu)$, while the hadronic channel allows to target the process $t\bar{t} \to q\mathrm{H}(\gamma\gamma)\mathrm{bW}^{\pm}(q\bar{q})$ in pair production of top-quarks. The set of requirements applied to candidate events is summarised in Table 7.2.

Events are pre-selected by requiring at least one combination of jets which satisfies 152 GeV < m($\gamma\gamma$ jet) < 190 GeV. This cut allows to target jets which are good candidates for the quark produced in the exotic decay of the top-quark, t $\to$ qH. At least one selected jet is asked to satisfy b-tagging requirement, which corresponds to a 77% efficiency to b-jets in $t\bar{t}$ events. Further categories are designed within each analysis channel, based on the mass of the jet(s) and leptons expected in the final state of the SM decaying top-quark t $\to \mathrm{bW}^{\pm}(q\bar{q}, \ell^{\pm}\nu)$:

- In the leptonic channel, events which fullfill the requirement 130 GeV < m($\ell^{\pm}\nu$ jet) < 210 GeV are filled into the Lepton-Tight category of the analysis (Lepton 1). Such events present reliable final states targetted by the purest category of the analysis. Other events are filled into the Lepton-Loose category of the analysis (Lepton 2).
- In the hadronic channel, events which fullfill the requirement 120 GeV < m(jet jet jet) < 220 GeV are filled into the Hadron-Tight category of the analysis (Hadron 1). Such events present fully reconstructed final states, but strongly rely on the accuracy of jets reconstruction within a phase space with large jets multiplicity. Other events are filled into the Hadron-Loose category of the analysis (Hadron 2).

7.1.3.2 Fitting procedure

The result of the analysis is extracted from a likelihood fit. In the hadronic channel, the diphoton invariant mass spectrum is fitted using a resonant signal function centered at the Higgs boson mass. The same function normalized to Monte-Carlo estimates is used as model for the resonant SM Higgs background. The non-resonant background yield in the Signal Region is estimated from a background function constrained using data in sidebands of the $m_{\gamma\gamma}$ spectrum. The unbinned likelihood fit relates to the product over all events of the expected $m_{\gamma\gamma}$ distribution function. Constrained Gaussian factors allow to encode the uncertainty from Nuisance Parameters.

In the leptonic channel, the low number of events does not allow to fit the diphoton mass distribution. A single bin is defined in the Signal Region (SR) in the range 122 GeV < $m_{\gamma\gamma}$ < 129

Selection	Leptonic channel		Hadronic channel
	Electron sub-channel	Muon sub-channel	
Diphoton trigger	HLT selection "$g35 - loose - g25 - loose$" : leading cluster with pT > 35 GeV, subleading cluster with pT > 25 GeV		
Diphoton selection	Two photons with tight identification criteria, looose calorimeter and track isolation $p_T(\gamma_1)$ > 40 GeV, $p_T(\gamma_2)$ > 30 GeV, 100 GeV < $m_{\gamma\gamma}$ < 160 GeV		
Lepton selection	Exactly one electron, no muon pT > 15 GeV	Exactly one muon no electron pT > 10 GeV	No identified lepton
mT	mT > 30 GeV		
Jets multiplicity	2 or more jets		4 or more jets
Jets used in the selection	Only the 2 jets with highest pT		Only the 4 jets with highest pT
b-Jets	At least one b-Tagged jet		
Mass cut targeting the t $\rightarrow qH$ decaying Top	At least one combination of jets satisfies the mass cut 152 GeV < m($\gamma\gamma$ jet) < 190 GeV		
Mass cut targeting the SM decaying Top (Tight category only)	At least one combination of jet(s) not used in the previous computation of m($\gamma\gamma$ jet) satisfies 130 GeV < m($\ell^{\pm}\nu$ jet) < 210 GeV		120 GeV < m(jet jet jet) < 220 GeV

TABLE 7.2: Summary of the selection applied to candidate events in the 2017 search.

Selection	Hadronic		Leptonic	
Category	1	2	1	2
Signal t $\to$ cH	2.4	3.7	0.82	0.23
SM Higgs boson resonant background	1.1	3.1	0.24	0.22
Other background	16	63	0.14	0.29
Total background	17	66	0.38	0.51
Data	14	69	2	1

TABLE 7.3: Numbers of FCNC signal events in the SR (122 GeV < $m_{\gamma\gamma}$ < 129 GeV) for the fitted t $\to$ cH branching ratio B = 6.9 x 10^{-4}. The numbers of events for the SM Higgs boson production and the fitted non-resonant background are also shown, together with the number of observed events in data, in the four categories.

GeV to measure the yield of candidate signal events in data. This is encoded in the likelihood function as a Poisson term. Similarly, the background yield is controlled from the number of events observed in sidebands. The ratio α of the numbers of background events expected in the SR and CR is fitted and allows to subtract the fitted number of background events in the SR. The uncertainty on α is estimated from variations in the background shape of $m_{\gamma\gamma}$.

7.1.3.3 Results

Table 7.3 presents the fitted number of signal and background events, and yield of data events observed in the SR. The associated upper limit on the Branching ratio for the t $\to$ cH process is B < 2.2 x 10^{-3} at the 95% CL. This value can be compared to the expected limit in the absence of signal B < 1.6 x 10^{-3}. The corresponding observed (expected) limit on the Branching ratio for the t $\to$ uH process is B < 2.4 x 10^{-3} (B < 1.7 x 10^{-3}) at the 95% CL.

In the following, new techniques are introduced in the hadronic channel of the analysis such as the use of c-tagging and multivariate methods. The improvements expected from these techniques bring additional complications to the analysis, for which several solutions are still being explored. Hence, the procedures presented will not meet the final version of the analysis, and the preliminary results should be seen as prospects for the forthcoming complete analysis.

7.2 Datasets

7.2.1 Experimental data

Data collected with the ATLAS detector in 2015 and 2016 at the run 2 of the LHC were exploited in the publication. Data recorded in 2017 increase the integrated luminosity from 36.1 fb^{-1} to 80.1 fb^{-1}. The analysis is expected to benefit from this larger dataset both on the data statistical side and fit performances, as background yields in the sensitive region of the analysis are constrained from a fit to data in sidebands of the $m_{\gamma\gamma}$ spectrum. Yet, the next publication for this search will most likely profit from the full run 2 dataset using data collected from 2015 to 2018, which could have an integrated luminosity of $\sim$ 140 fb^{-1}. Statistical studies are performed on data collected in 2015, 2016 and 2017, blinded in the sensitive region 120 GeV < $m_{\gamma\gamma}$ < 130 GeV. Correspondingly, optimisation studies presented in the following rely on Monte-Carlo samples intended to meet the same period.

Sample	Process	ME Generator	Parton Shower	σxBr(fb)
Pair-production with t $\to$ cH decay	$t\bar{t} \to W(\ell\nu)cH(\gamma\gamma)$	Powheg	Pythia 8	1.224
	$t\bar{t} \to W(q\bar{q})cH(\gamma\gamma)$			2.548
Pair-production with t $\to$ uH decay	$t\bar{t} \to W(\ell\nu)uH(\gamma\gamma)$	Powheg	Pythia 8	1.224
	$t\bar{t} \to W(q\bar{q})uH(\gamma\gamma)$			2.548
Single-top production with tuH coupling	$qg \to W(\ell\nu)uH(\gamma\gamma)$	MG5_aMC@NLO	Pythia8	0.599
	$qg \to W(q\bar{q})uH(\gamma\gamma)$			1.227
Single-top production with tcH coupling	$qg \to W(\ell\nu)cH(\gamma\gamma)$	MG5_aMC@NLO	Pythia8	0.0865
	$qg \to W(q\bar{q})cH(\gamma\gamma)$			0.177

TABLE 7.4: Description of Monte-Carlo simulations for pair production and single-top signal samples. The cross-section accounts for the decays specified in the process assuming Br = $1.e^{-3}$ for a single flavour and 0 for the other flavour.

Process	ME Generator	Parton Shower	Cross-section order	σxBr(fb)
ggH	Powheg	Pythia 8	N3LO (QCD)	110.3
VBFH	Powheg	Pythia 8	NNLO (QCD) + NLO (EW)	8.585
WH	Powheg	Pythia 8	NNLO (QCD) + NLO (EW)	3.117
ZH	Powheg	Pythia 8	NNLO (QCD) + NLO (EW)	2.006
$t\bar{t}H$	Powheg	Pythia 8	NLO (QCD) + NLO (EW)	1.151
bbH	Powheg	Pythia 8	NNLO (QCD) ($gg \to bbH$ @NLO)	1.104
tHjb	MadGraph5 LO	Pythia 8	NLO (QCD)	0.168
tWH	MG5_aMC@NLO	Herwig ++	NLO (QCD)	0.034

TABLE 7.5: Description of Monte-Carlo simulations for the resonant SM Higgs backgrounds considered in the FCNC search.

7.2.2 Monte-Carlo samples

Optimisation studies are performed using Monte-Carlo simulations of the targetted signal processes and backgrounds. Events yields are normalized to the most accurate theoretical cross-section prediction for each individual process. The detector response for generated stable particles is simulated using a model of the ATLAS detector within the GEANT4 framework [54]. Minimum bias events generated with Pythia 8 are overlaid onto hard scattering events during the digitization, which provides a model for pile-up. Table 7.4 lists signal samples generated for this analysis. The cross-section for the single-top FCNC processes does not depend on the choice of chirality operator. The resonant SM Higgs background samples are presented in Table 7.5. The non-resonant diphoton + jets background is simulated with Sherpa using a merging strategy on the number of additional partons. Events with 0 or 1 additional partons are generated at NLO with QCD Parton Shower, while events with 2 and 3 additional partons are generated at LO with dedicated Matrix Elements. Only events with 90 GeV < $m_{\gamma\gamma}$ < 175 GeV at generation level are exploited. No simulation of the single photon + jets background is exploited in the analysis, but the corresponding yield estimate is about 15% of that for the diphoton + jets background. The $t\bar{t}\gamma$ background is generated using Madgraph+Pythia8 in fast simulation with at least one top quark decaying leptonically. Last, V+$\gamma\gamma$ processes are simulated with Sherpa requiring $m_{\gamma\gamma}$ > 80 GeV and at least one photon with p_T^{γ} > 17 GeV. Events with 0 addtionnal partons are generated at NLO, while events with 1 and 2 additional partons are generated at LO.

7.3 Objects selection

7.3.1 Trigger

Events pre-selection relies on a diphoton trigger applied to data and emulated to MC samples, with efficiency corrections applied to simulated events. Two calorimeter electromagnetic clusters with measured transverse energy greater than 35 GeV and 25 GeV are required in the trigger selection. Candidate calorimeter clusters must satisfy shower shape criteria designed to match electromagnetic showers initiated by photons.

7.3.2 Photons

The analysis exploits photons reconstructed in the electromagnetic calorimeter from clusters with transverse energy deposit greater than 2.5 GeV. The combination of tracking and calorimeter informations allows to discriminate converted and unconverted photons candidates with respect to isolated electrons. Two photons meeting tight identification criteria, loose calorimeter and track isolation are required in the events final state [105]. The selected photons are also required to meet the trigger condition.

7.3.3 Jets

The analysis relies on jets reconstructed from topological calorimeter clusters using the anti-k_t algorithm with radius parameter R=0.4. Jets are required to have transverse momentum larger than 30 GeV and to stand within detector acceptance $|\eta| < 4.4$ in pseudo-rapidity. In the central region $|\eta| < 2.4$ within Inner Detector acceptance, the contamination from pile-up jets is mostly suppressed by cutting on the JVT discriminant (JVT > 0.59). Forward regions of the detector benefit from the fJVT discriminant, which relies on jet shapes and topological jet correlations in pile-up interactions.

b- and c-Tagging techniques allow to attest how likely does a jet originate from the hadronization of a b- or c-quark respectively. The 77% b-jet tagging efficiency Working Point is used in the analysis, targeting at b-jets originating from SM decaying top quarks. The corresponding efficiencies for tagging c- and light-labelled jets are 16.1% and 0.74% respectively.

c-Tagging is a technique in development which relies on c-quark hadronization and decay properties, targeting at c-jets produced in the t $\to$ cH decay. Yet, c-Tagging offers a smaller discrimination and severely suffers from the contamination of mistagged b-jets. The c-Tagging discriminants are MV2c100 and MV2cl100, which rely on the same techniques as the MV2 algorithms used for b-Tagging as described in Section 4.2.4. The MV2c100 discriminant consists of a BDT aiming at the separation of b- and c-jets, while MV2cl100 aims at the separation of light- and c-jets. Dedicated requirements are performed using MV2c100 and MV2cl100 respectively. The loose c-Tagging Working Point (MV2c100 < 0.58 and MV2cl100 > 0.19) used in the analysis offers a compromise between c-jet tagging efficiency and b- (light-)jets rejection. The loose c-Tagging Working Point provides 41% tagging efficiency for c-labelled jets in MC simulations. The corresponding acceptance for b- and light-labelled jets are 25% and 5%.

Jets are associated with individual Scale-Factors depending on their kinematic and the output of b- or c-Tagging tests, which multiply the events weight. These Scale-Factors are calibrated separately for b- and c-Tagging. Thus, testing jets for c- and b-Tagging simultaneously does not allow for a consistent use of the corresponding calibrations. A combined calibration accounting for the correlations between the two techniques will be needed to enable this feature. Hence, the analysis is designed in order to avoid this issue and test jets either for b-Tagging or c-Tagging, depending on their kinematics as described in Section 7.4.2.2.

7.3.4 Electrons

Electrons are reconstructed from energy clusters in the electromagnetic calorimeter within $|\eta_{cl}| < 2.47$ associated to tracks in the Inner Detector. Electrons candidates with transverse momentum larger than 15 GeV must fullfill medium quality requirements, loose calorimeter and track isolation criteria within a cone of $\Delta R < 0.2$ to be considered in the analysis.

7.3.5 Muons

Medium quality requirements are applied to muons candidates standing in the pseudo-rapidity range $|\eta| < 2.7$ with transverse momentum larger than 10 GeV. Muons should also pass track isolation requirements within a cone of radius $\Delta R < 0.3$ with a (p_T, η) independent efficiency of 99%.

7.3.6 Overlap removal

The following Overlap-Removal procedure is applied to candidate events :

1. Identified photons are preserved.
2. Electrons nearby photons within $\Delta R(\mathrm{e}, \gamma) < 0.4$ are removed.
3. Jets nearby photons or electrons within $\Delta R(\mathrm{jet}, \mathrm{e}) < 0.2$ or $\Delta R(\mathrm{jet}, \gamma) < 0.4$ are removed.
4. Muons nearby photons or jets within $\Delta R(\mu, \gamma/\mathrm{jet}) < 0.4$ are removed.
5. Electrons nearby the remaining jets within $\Delta R(\mathrm{e}, \mathrm{jet}) < 0.4$ are removed.

7.3.7 Missing Transverse Energy

Missing Transverse Energy is computed from the negative vector sum of reconstructed objects energy in the transverse plane. Jets with transverse momentum lower 20 GeV or which do not satisfy JVT requirements are substituted by their associated tracks matched to the primary vertex, which enter the dedicated soft term of Missing Transverse Energy. Unassociated isolated tracks with transverse momentum grater than 5 GeV also enter the definition of the MET soft term.

7.4 Events selection

7.4.1 Selection

The event selection is updated with respect to the 2017 publication 7.2. Two non-overlapping selections are defined, targeting respectively at $\mathrm{t\bar{t}}$ (pair production) and single-top production modes of top quarks. The single-top selection allows to disentangle the tcH and tuH processes due to different production cross-sections and events kinematics. A common pre-selection is presented in Table 7.6. The pair production and single-top selections are described in the following, and summarized in Table 7.7 and 7.8 respectively.

- In the pair-production selection, the combinatoric approach is extended, with respect to the published analysis, to the five jets with highest transverse momentum. This allows to increase the efficiency of the mass cut requirements on top-candidates. This change especially enhances the acceptance for signal events in the hadronic channel, as jets emitted in the W-boson decay are often soft and can be supplanted by jets produced

Common Selection	Leptonic channel		Hadronic channel
	Electron sub-channel	Muon sub-channel	
Diphoton trigger	HLT selection "$g35 - loose - g25 - loose$" : leading cluster with pT > 35 GeV, subleading cluster with pT > 25 GeV		
Diphoton selection	Two photons with tight identification criteria, looose calorimeter and track isolation $p_T(\gamma_1)$ > 40 GeV, $p_T(\gamma_2)$ > 30 GeV, 100 GeV < $m_{\gamma\gamma}$ < 160 GeV		
Lepton selection	Exactly one electron, no muon pT > 15 GeV	Exactly one muon, no electron pT > 10 GeV	No identified lepton
mT	mT > 30 GeV		

TABLE 7.6: Common requirements applied to the pair production ($t\bar{t}$) and single-top selections.

Pair production selection ($t\bar{t}$)	Leptonic channel	Hadronic channel
Pre-selection	Fulfill the common selection (Table 7.6)	
Jets multiplicity	2 or more jets	4 or more jets
Jets used in the selection	Only the 5 jets with highest pT	
b-Jets	At least one b-Tagged jet	
Mass cut targeting the t $\to qH$ decaying Top	At least one jet satisfies the mass cut 152 GeV < m($\gamma\gamma$ jet) < 190 GeV	
Mass cut targeting the SM decaying Top (Tight category only)	At least one combination of jet(s) not used in the previous computation of m($\gamma\gamma$ jet), amongst which at least one b-Tagged jet, satisfy	
	130 GeV < m($\ell^{\pm}\nu$ b-jet) < 210 GeV	120 GeV < m(b-jet jet jet) < 220 GeV

TABLE 7.7: Summary of the event requirements in the leptonic and hadronic channels of the pair production ($t\bar{t}$) selection.

from hard radiations or pile-up. To limit the background contamination, at least one b-Tagged jet is required in the combination of jets targeting the SM decaying top. Tagging properties of the c-quark jet candidate are exploited in the further categorisation of events, as described in Section 7.4.2.

- The single-top selection is applied to events failing the $t\bar{t}$ requirements. In the leptonic channel, at least one-jet is required, targeting the b-quark. In the hadronic channel, at least three jets are required, aiming at the SM decaying top. All events are treated using a combinatoric approach over the five jets with highest transverse momentum, targeting at the reconstruction of the SM decaying top. Candidate combinations are also required to present at least one b-Tagged jet.

Single-top selection	Leptonic channel	Hadronic channel
Pre-selection	Fulfill the common selection (Table 7.6) and fail the pair production selection (Table 7.7)	
Jets multiplicity	1 or more jets	3 or more jets
Jets used in the selection	Only the 5 jets with highest pT	
b-Jets	At least one b-Tagged jet	
Mass cut targeting the SM decaying Top	At least one combination of jets, amongst which at least one b-Tagged jet, satisfy	
	130 GeV < m($\ell^{\pm}\nu$ b-jet) < 210 GeV	120 GeV < m(b-jet jet jet) < 220 GeV

TABLE 7.8: Summary of the event requirements in the leptonic and hadronic channels of the single-top selection.

7.4.2 Categories

Events passing the selection are separated into non-overlapping categories. These categories are designed as a mean to keep a good signal acceptance, and exploit regions with increased signal to background yield ratio. Table 7.10 summarizes the expected yields for individual processes using the categorization described in the next sections. The single-top FCNC signals are separated in this table for the different helicity operators. In the rest of the analysis, samples accounting for the different helicity operators are normalized to half of their expected cross-section and mixed in order to increase the effective statistics. The yield for the total background expectation seems larger than that of data (which are blinded in the range 120 GeV < $m_{\gamma\gamma}$ < 130 GeV) because of the approximate cross-section estimate for the $\gamma\gamma$+jets background.

7.4.2.1 Kinematic categories

Kinematic categories are defined, based on the result of mass cut requirements defined in the selection. The correspondance with respect to the expected targetted final states is described in the following and summarized in Table 7.9.

- In the $t\bar{t}$ selection, all events are required to present a good jets combination to the $t \to qH$ decay, as encoded in the cut on m($\gamma\gamma$ jet). If at least one combination allows for the reconstruction of a good SM decaying top candidate (m(b-jet jet jet) or m($\ell^{\pm}\nu$ b-jet) mass cut requirements), the event is filled into the tight category of the selection (Categories Had1 and Lep1). Other events are placed into the loose category (Categories Had2 and Lep2).
- Events which fail both the loose and the tight above requirements but present at least one combination with a good SM decaying top candidate are filled into the category aiming at single-top events (Categories HadtH and LeptH).

Kinematic category	Had1/Lep1	Had2/Lep2	HadtH/LeptH
Good jets candidate to $t \to qH$	x	x	-
Good jets candidate to $t \to bW^{\pm}$	x	-	x

TABLE 7.9: Summary of the categorisation strategy, based on top-quark candidate systems which are found from reconstructed objects.

7.4.2.2 c-Tagging categorisation

In $t\bar{t}$ categories, c-Tagging is expected to differentiate between $t \to cH$ and $t \to uH$ decays. The c-Tagging requirement consists of two cuts, aiming at the discrimination of c-jets against light- and b-jets respectively.

Kinematic categories are each split into tagging categories :

- If the jet used in the selected ($\gamma\gamma$ jet) combination is c-Tagged, the event is placed in tagging category A.
- Otherwise, the event enters tagging categories B.

Overall, eight $t\bar{t}$ categories are defined, from the hadronic and leptonic channels (Had, Lep), kinematic requirements (1, 2) and tagging category (A, B). Single-top categories (tH) only consist of the hadronic and leptonic channels with no further categorisation.

The calibration of scale factors does not allow for a consistent use of b- and c-tagging, as correlations cannot be taken into account with respect to systematic uncertainties. Hence, a criterion is defined in order to select if a jet should be tested with respect to b- or c-tagging requirements. In the pair-production channel, the exotic top quark decay is expected to produce two photons and a jet matching to the c-quark. Hence, jets which satisfy 152 GeV < m($\gamma\gamma$ jet) < 190 GeV are tested for c-tagging only, as they appear as good candidates to match the c-quark. Conversely, jets which fail the previous requirement are tested against b-tagging requirements only.

7.4.3 Improvement of the selection in the hadronic channel

In the 2017 publication, only the $m_{\gamma\gamma}$ distribution is exploited in the likelihood fit procedure. In the hadronic channel where the surviving number of events is large enough, optimisations of the analysis are possible by exploiting other kinematic variables which offer a discrimination of signal processes against backgound events. Yet, the combinatoric approach used in the events selection requires special care in the computation of such variables, as several combinations of jets possibly satisfy the selection requirements for the same analysis category. Optimisation studies were performed in order to select the most-signal like combination within each candidate event. Jets combinations of a single event are ranked based on matching criteria relying on the truth of $t\bar{t}$ cH($\gamma\gamma$)bW$^{\pm}$($q\bar{q}$) MC simulations. The recycling procedure below describes the requirements of combinations ranking :

1. All reconstructed jets match to Matrix Element partons (Bin [All matched] in Fig. 7.5) :
   (a) The jet used in the computation of m($\gamma\gamma$ jet) satisfies ΔR(jet, c-quark) < 0.4
   (b) The b-Tagged jet used in the computation of m(b-jet jet jet) satisfies ΔR(b-jet, b-quark) < 0.4
   (c) Other jets of the combination are each matched to a different W-boson decay product within $\Delta R < 0.4$
2. c- and b-jets match to Matrix Element c and b quarks (Bin [b and c matched] in Fig. 7.5) :
   (a) The jet used in the computation of m($\gamma\gamma$ jet) satisfies ΔR(jet, c-quark) < 0.4
   (b) The b-Tagged jet used in the computation of m(jet jet b-jet) satisfies ΔR(b-jet, b-quark) < 0.4
   (c) Other jets are not matched to the appropriate partons.
3. The c-jet matches to Matrix Element c (Bin [c matched] in Fig. 7.5) :
   (a) The jet used in the computation of m($\gamma\gamma$ jet) satisfies ΔR(jet, c-quark) < 0.4
   (b) Other jets are not matched to the appropriate partons.
4. The b-jet matches to Matrix Element c (Bin [b matched] in Fig. 7.5) :
   (a) The b-Tagged jet used in the computation of m(jet jet b-jet) satisfies ΔR(b-jet, b-quark) < 0.4.
   (b) Other jets are not matched to the appropriate partons.
5. No jet can be matched to the approriate partons (Bin [No match] in Fig. 7.5).

Different approaches relying on reconstructed kinematic variables are tested, aiming at the most performing matching fractions for signal procceses in the previous ranking strategy. The most relevant strategies tested consist of selecting :

- The combination which uses highest pT jets in the computation of m($\gamma\gamma$ jet) first, and m(b-jet jet jet) then.

Sample	Process	Analysis category									
		Had1A	Had1B	Had2A	Had2B	HadtH	Lep1A	Lep1B	Lep2A	Lep2B	LeptH
FCNC signal Pair-production with t → cH/uH decay	$t\bar{t} \to W(\ell\nu)cH(\gamma\gamma)$	0.217	0.366	0.478	0.744	1.273	1.178	2.021	0.19	0.32	3.107
	$t\bar{t} \to W(q\bar{q})cH(\gamma\gamma)$	3.284	5.53	2.028	3.884	10.57	0	0	0	0	0.005
	$t\bar{t} \to W(\ell\nu)uH(\gamma\gamma)$	0.057	0.537	0.123	1.118	1.24	0.215	3.164	0.039	0.478	2.869
	$t\bar{t} \to W(q\bar{q})uH(\gamma\gamma)$	0.883	8.13	0.576	5.46	10.05	0.002	0.002	0	0	0.002
FCNC signal Single-top production with tuH coupling	$qg \to W(\ell\nu)uH(\gamma\gamma)$ $(\bar{t_L}\tilde{\phi}u_R)$	0.002	0.022	0.006	0.037	0.329	0.007	0.104	0.001	0.023	2.29
	$qg \to W(q\bar{q})uH(\gamma\gamma)$ $(\bar{t_L}\tilde{\phi}u_R)$	0.06	0.458	0.062	0.453	5.68	0	0	0	0	0.003
	$qg \to W(\ell\nu)uH(\gamma\gamma)$ $(\bar{u_L}\tilde{\phi}t_R)$	0.003	0.025	0.007	0.049	0.332	0.007	0.125	0.001	0.022	2.26
	$qg \to W(q\bar{q})uH(\gamma\gamma)$ $(\bar{u_L}\tilde{\phi}t_R)$	0.053	0.441	0.064	0.349	5.57	0	0	0	0	0.001
FCNC signal Single-top production with tcH coupling	$qg \to W(\ell\nu)cH(\gamma\gamma)$ $(\bar{c_L}\tilde{\phi}t_R)$	0.001	0.006	0.002	0.009	0.073	0.006	0.033	0	0.005	0.479
	$qg \to W(q\bar{q})cH(\gamma\gamma)$ $(\bar{c_L}\tilde{\phi}t_R)$	0.022	0.109	0.018	0.097	1.176	0	0	0	0	0
	$qg \to W(\ell\nu)cH(\gamma\gamma)$ $(\bar{t_L}\tilde{\phi}c_R)$	0.001	0.006	0.002	0.011	0.071	0.005	0.034	0	0.006	0.492
	$qg \to W(q\bar{q})cH(\gamma\gamma)$ $(\bar{t_L}\tilde{\phi}c_R)$	0.025	0.127	0.019	0.108	1.154	0	0	0	0	0.001
Total signal tuH		1,06	9,61	0,84	7,47	23,20	0.231	3.395	0.041	0.523	7.425
Total signal tcH		3,55	6,14	2,55	4,85	14,32	1.189	2.088	0.19	0.331	4.084
SM Higgs resonant background	Gluons fusion (ggH)	0.031	0.23	0.067	0.657	2.946	0	0	0	0	0.007
	Vector Boson Fusion (VBFH)	0.002	0.017	0.014	0.124	0.282	0	0	0	0	0
	W-boson associated production (WH)	0.007	0.063	0.015	0.117	0.545	0	0.006	0	0.006	0.07
	Z-boson associated production (ZH)	0.013	0.068	0.022	0.147	0.663	0	0	0	0.001	0.013
	top-quarks associated production ($t\bar{t}H$)	0.741	2.69	0.722	3.01	5.27	0.276	1.112	0.165	0.641	2.19
	b-quarks associated production (bbH)	0.001	0.01	0.006	0.058	0.119	0	0	0	0	0.002
	tHjb	0.016	0.106	0.045	0.223	0.703	0.004	0.03	0.002	0.013	0.273
	single-top associated production (tWH)	0.006	0.032	0.01	0.058	0.135	0.002	0.012	0.002	0.014	0.053
Total SM Higgs		0,817	3,216	0,901	4,394	10,663	0.282	1.160	0.169	0.675	2.608
Non-resonant background	Diphoton + jets	24.5	152.1	64.5	412	1149	0.121	0.32	0.009	0.074	1.312
	$t\bar{t} + \gamma$	4.21	18.24	5.75	26.7	58.8	1.794	11.18	1.436	7.66	27.3
	V+$\gamma\gamma$	0.118	0.197	0.279	1.261	2.533	0.018	0.333	0.037	0.362	2.951
Total background		29.6	173.8	71.4	444.4	1221.6	2.215	12.993	1.651	8.771	34.171
Data blinded		19	132	51	300	1039	0	5	1	6	39

TABLE 7.10: Expected yields for an integrated luminosity of 80.1 fb^{-1} based on Monte-Carlo simulations for FCNC signal, SM Higgs and non-resonant backgrounds considered in the search. Data are blinded in the range 120 GeV < $m_{\gamma\gamma}$ < 130 GeV.

- The combination which allows the value of m($\gamma\gamma$ jet) closest to 172.8 GeV, aiming at the best jet candidate for the t $\to$ qH decay.
- The combination which allows the value of m(b-jet jet jet) closest to 172.8 GeV, aiming at the best jets candidates for the t $\to$ bW($q\bar{q}$) decay.
- The combination which has the two jets closest to the b-tagged jet (if several jets are b-Tagged, the b-Tagged jet with highest transverse momentum is selected). This strategy also seeks the best jets candidates for the t $\to$ bW($q\bar{q}$) decay, based on the assumption that quarks are emmited near-by.
- The combination wich maximises the pT of the softest jet. This approach is expected to enhance the rejection of jets produced from pile-up and hard radiations.

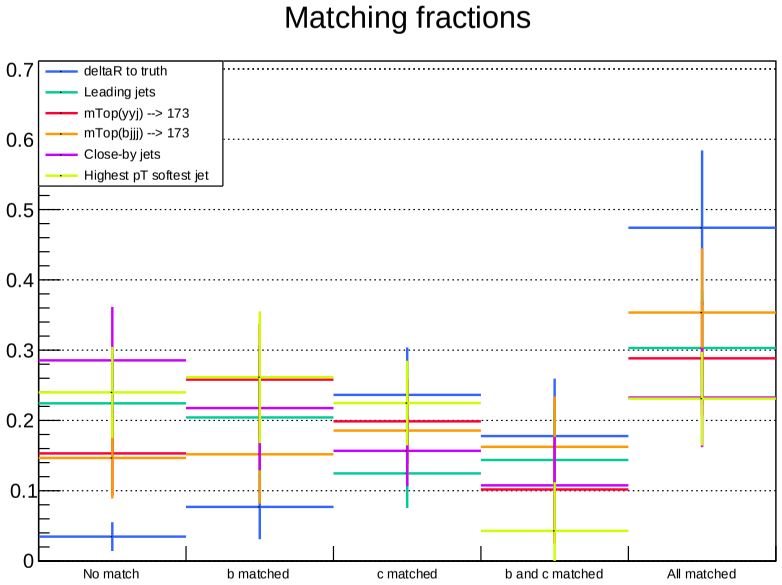
The strategies presented above exploit variables computed at reconstructed level. Their performances are compared to the approach based on the truth of MC simulations which consists of matching jets to quarks depending on their angular separation. This latter truth-based strategy allows to measure the maximal achievable matching fractions. Figure 7.5 presents the individual results in each category of the t$\bar{\text{t}}$ selection for events where combinations interchange the jet used in the computation of m($\gamma\gamma$ jet). It is noticeable that the strategy of selecting the combination with m($\gamma\gamma$ jet) closest to 172.8 GeV leads to the best matching performances of the c-jet to c-quark in most categories. Figure 7.6 presents the fractions for the number of jets used in the computation of m(b-jet jet jet) which are matched to truth quarks. Within tightest categories, strategies relying on the transverse momentum of jets show the best performances. The strategy based on m(b-jet jet jet) shows slightly lower matching fractions, but seems an intuitive choice in categories with m(b-jet jet jet) satisfying tight selection requirement. Within loose categories, the strategy wich picks the two jets closest to the b-Tagged jet presents optimal performances and is adopted for the rest of the analysis.

The final strategy adopted for picking the most signal-like combination is :

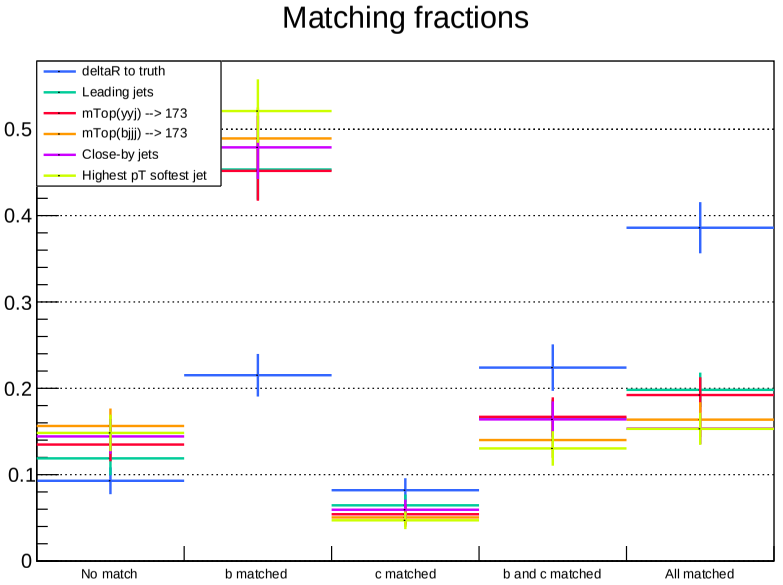
1. The categories are sorted with priority Had1A > Had1B > Had2A > Had2B. Combinations entering the best possible category are considered for the following ranking procedure, while other combinations are discarded.
2. If several jets can be used in the computation of m($\gamma\gamma$ jet), the jet which minimises $|$m($\gamma\gamma$ jet)-171$|$ is selected. The targetted value m_{top} = 171 GeV results from an optimisation, looking for the reconstructed mass of the jet and photons system matched to top-quark decay products.
3. In the tightest category, in which good jets candidates are found for the t $\to$ bW$^{\pm}$ decay, the combination which minimises $|$m(b-jet jet jet)-170$|$ is selected.
4. In the loose category, the two jets closest to the b-Tagged jet with highest transverse momentum value are selected.

7.5 Study of events kinematic

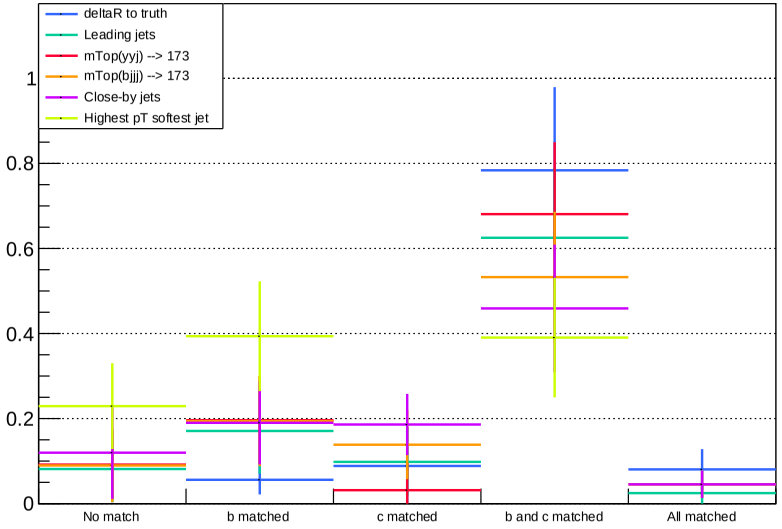
Kinematic variables are designed in the hadronic channel in order to allow for further signal against background discrimination. The large number of objects expected in the final state offers multiple possibilies, both on the individual and combined level. This Table 7.11 presents the variables which were tested and exploited into multivariate methods described in Section 7.6, separately for the pair-production and single-top selections.



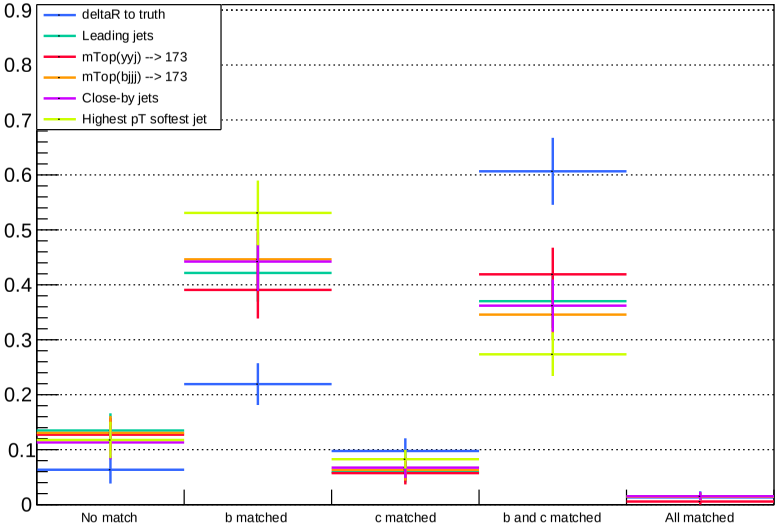
(a) Tight selection with a c-Tagged jet.



(b) Tight selection, no c-Tagged jet.

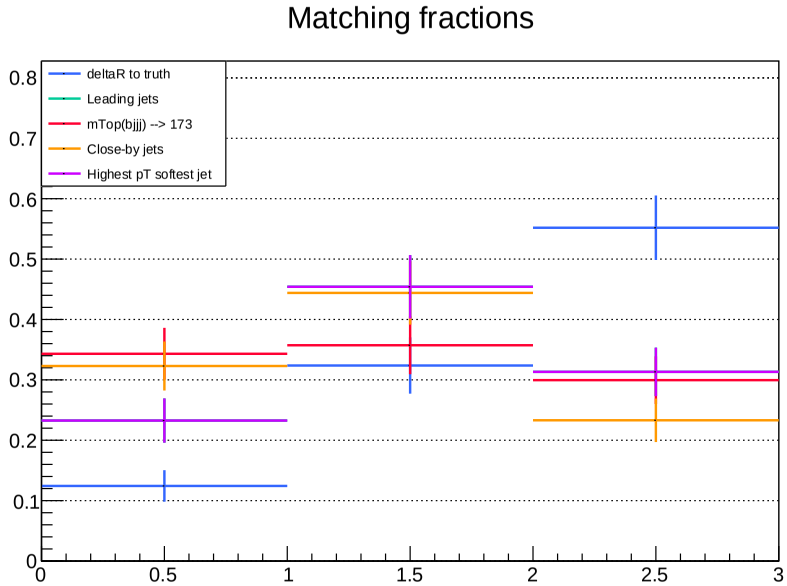


(c) Loose selection with a c-Tagged jet.

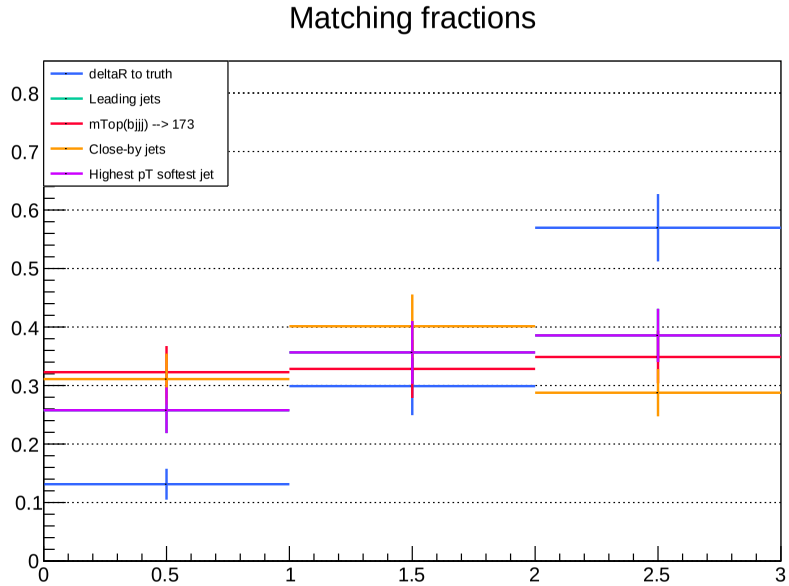


(d) Loose selection, no c-Tagged jet.

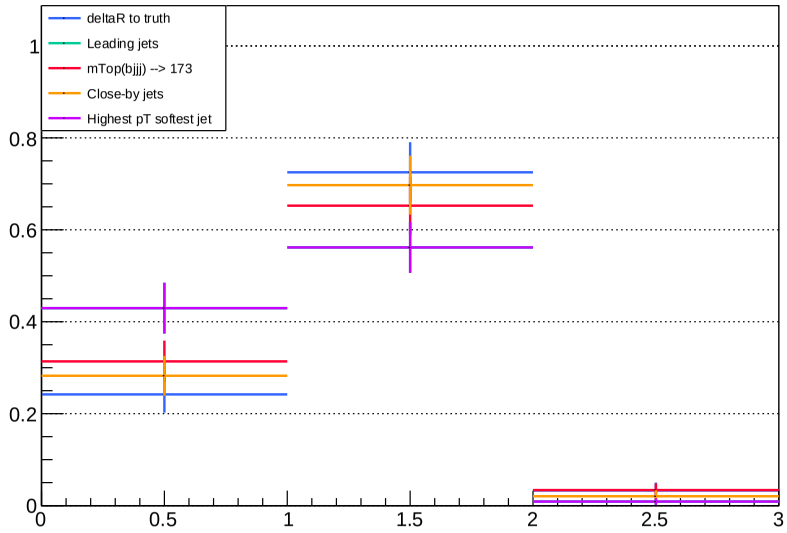
FIGURE 7.5: Matching score fractions in events interchanging the jet used in the computation of m($\gamma\gamma$ jet) for the different procedure tested in order to retrieve the most signal-like combinations. Matchings were performed between reconstructed and corresponding truth events in FCNC signal samples with hadronic decay of the vector boson. Distributions a and b (c and d) correspond to the tight (loose) selection of the $t\bar{t}$ analysis.



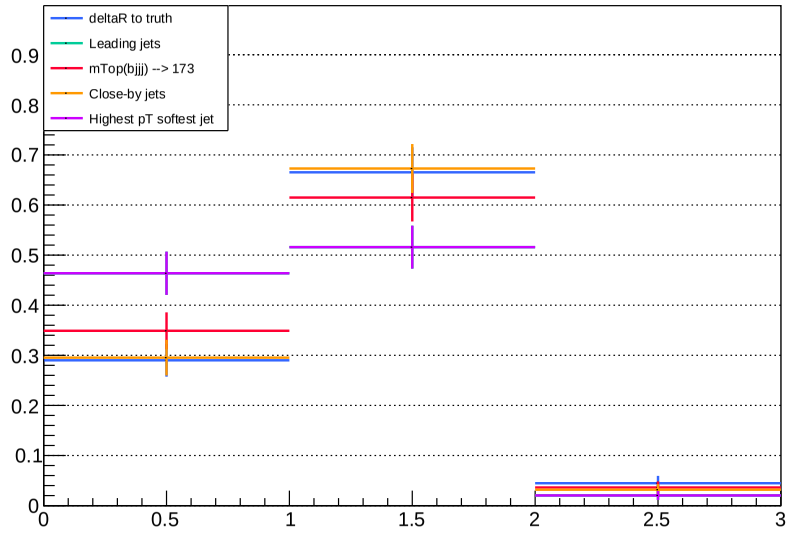
(a) Tight selection with a c-Tagged jet (Category Had1A).



(b) Tight selection, no c-Tagged jet (Category Had1B).



(c) Loose selection with a c-Tagged jet (Category Had2A).



(d) Loose selection, no c-Tagged jet (Category Had2B).

FIGURE 7.6: Number of jets matching fractions for the different procedure tested in order to retrieve the most signal-like combinations. Matchings were performed between reconstructed and corresponding truth events in FCNC signal samples with hadronic decay of the vector boson. Distributions a and b (c and d) correspond to the tight (loose) selection of the $t\bar{t}$ analysis. The strategy relying on the jets with highest transverse momentum leads to results identical to that based on the softest jet.

Variable level	Variable name	Description	Selection	
			Pair-production	Single-top
Full event	HT	Scalar sum of selected jets pT	x	x
	HTAllJets	Scalar sum of all jets pT	x	x
	$m(\mathrm{t\bar{t}})$	Mass of the $\mathrm{t\bar{t}}$ system	x	-
	$p_T^{\mathrm{t\bar{t}}}$	Transverse momentum of the $\mathrm{t\bar{t}}$ system	x	-
	$y(\mathrm{t\bar{t}})$	Rapidity of the $\mathrm{t\bar{t}}$ system	x	-
t $\to qH$ decaying top exploiting the jet selected in the computation of m(jet $\gamma\gamma$)	m(jet $\gamma\gamma$)	Top quark candidate system mass	x	-
	Meff(jet $\gamma\gamma$) $= p_T^{\text{jet}} + p_T^{\gamma_1} + p_T^{\gamma_2}$	Exotic top-quark candidate system effective mass	x	-
	Meff(b-jet jet jet) = $p_T^{b-\text{jet}} + p_T^{\text{jet}_1} + p_T^{\text{jet}_2}$	SM top-quark candidate system effective mass	x	-
	Meff($j\gamma\gamma bjj$) = Meff($j\gamma\gamma$) + Meff(bjj)	Top-quark pair candidate system effective mass	x	-
	p_T, η(jet)	Jet transverse momentum and pseudo-rapidity	x	x
	ΔR, $\Delta\phi$(jet,$\gamma\gamma$)	Angular separations of the jet and diphoton system	x	-
	$(max, min) \otimes (\Delta R, \Delta\phi)$ (jet,$\gamma_{1,2}$)	Extreme angular separations of the jet and a single-photon	x	-
	ΔR, $\Delta\phi(\gamma, \gamma)$	Diphoton system angular separations	x	x
	$p_T, \eta, y(\gamma, \gamma)$	Transverse momentum and (pseudo-) rapidity of the diphoton system	x	x
	$p_T, \eta(\gamma_{1,2})$	Individual photons transverse momentum and pseudo-rapidity	x	x
SM decaying top exploiting the jets selected in the computation of m(b-jet jet jet)	m(b-jet jet jet)	Top quark candidate system mass	x	x
	Meff(b-jet jet jet) = $p_T^{b\text{-jet}} + p_T^{\text{jet}_1} + p_T^{\text{jet}_2}$	Top quark candidate system effective mass	x	x
	p_T, η(b-jet)	b-Tagged jet transverse momentum and pseudo-rapidity	x	x
	$p_T, \eta(\text{jet}_{1,2})$	light-jets transverse momentum and pseudo-rapidity	x	x
	ΔR, $\Delta\phi$(b-jet,W)	Angular separations of the b-jet and vector boson candidate jets system	x	x
	m(W)	Mass of the vector boson candidate jets system	x	x
	$(max, min) \otimes (\Delta R, \Delta\phi)$ (b-jet,$\text{jet}_{1,2}$)	Extreme angular separations of the b- and light-jets	x	x

TABLE 7.11: Summary of the variables explored in the hadronic channel to improve our understanding of events topology and gain in signal against background.

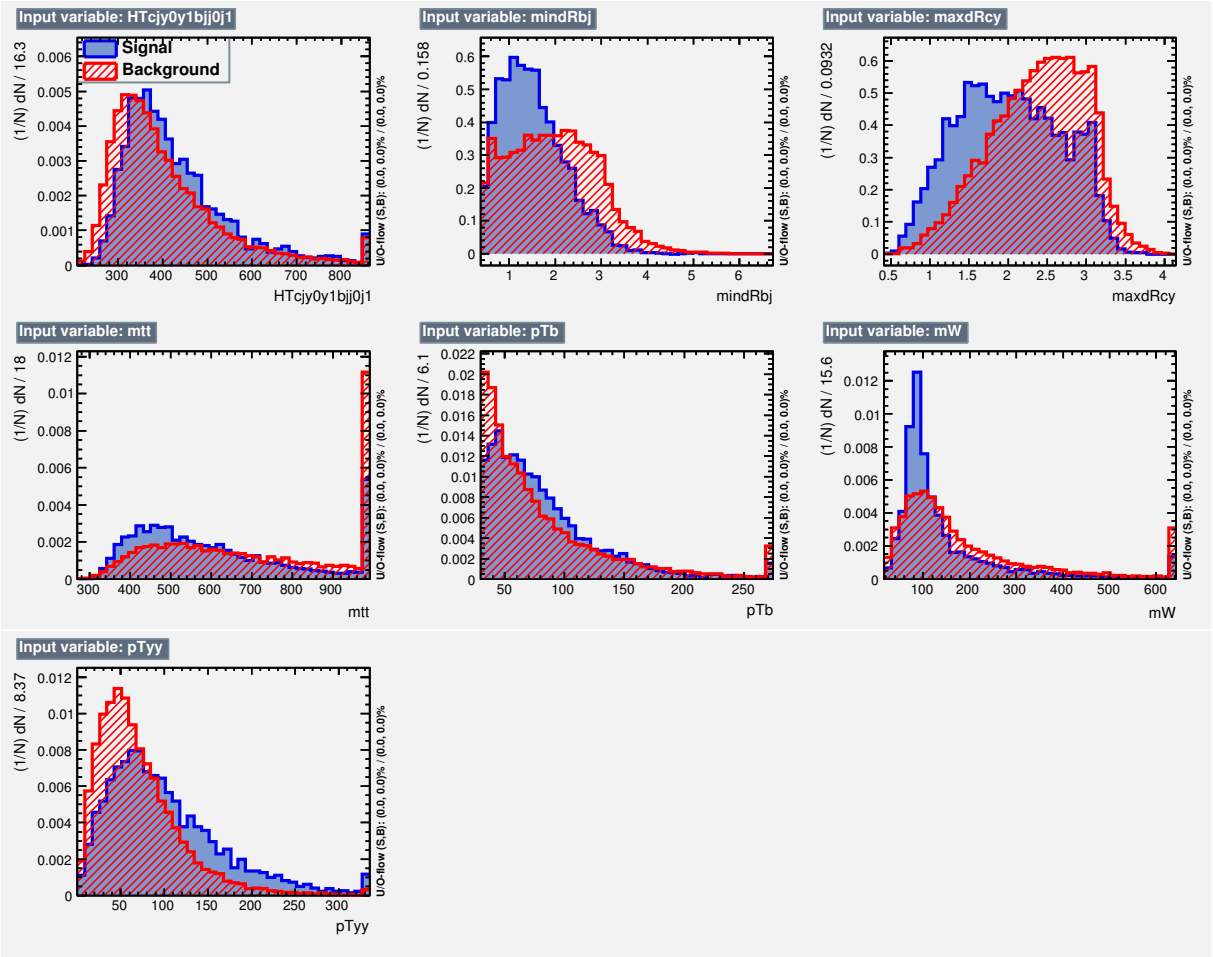


FIGURE 7.7: Shape of the most discriminating variables in the most sensitive category (1A) in the hadronic channel of the pair-production analysis. These shapes are very similar in other categories of the analysis. In the loose categories (2A and 2B), the variables m_W and $m(\mathrm{t\bar{t}})$ present a smaller discrimination due to the selection of jets which do not result from the hadronization of the W-boson decay products.

The selection of training variables is described in the next section. The shape of the most discriminating variables are presented in Figures 7.7 and 7.8 for the pair-production and single-top analyses respectively, where the shapes of the background originate from the $\gamma\gamma$+jets process.

7.6 Multivariate method

Multivariate techniques identical to that described in Chapter 5 are used in order to improve the separation of signal and background events. A ranking is performed in order to limit the number of training variables and provide only relevant informations to the BDT. The ranking procedure is similar to that described in Section 5.5.4.1, but going the opposite way in successive iterations. Starting from a set formed of all training variables, each iteration in the ranking procedure allows to remove the least discriminating variable. This approach allows the BDT training to exploit all correlations starting from the initial set of variables, and remove variables which provide the least information. The minimal set of relevant variables is that for which removing the least discriminating variable significantly decreases the evaluated significance.

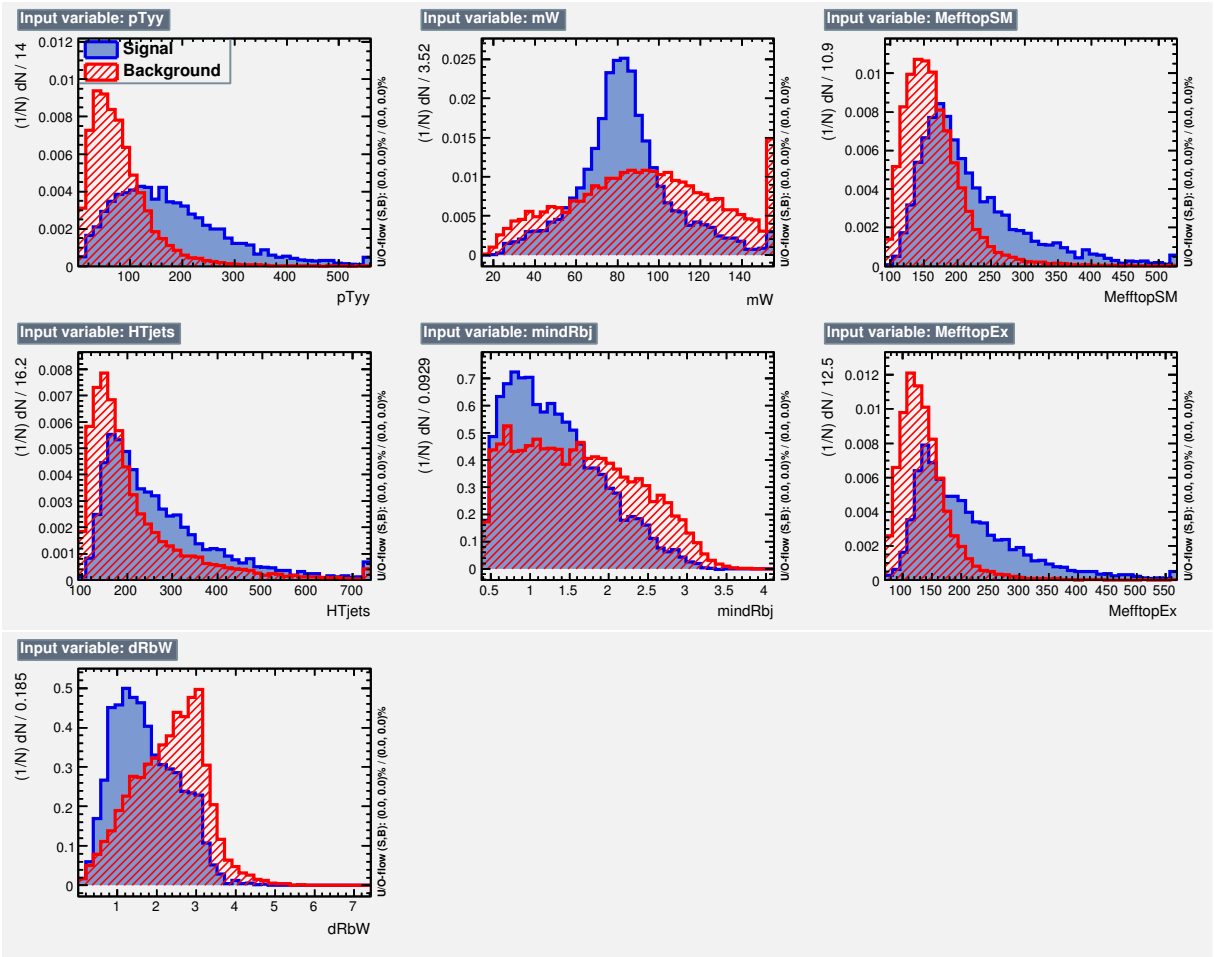


FIGURE 7.8: Shape of the most discriminating variables in the hadronic channel of the single-top analysis.

Variable name	Intermediate ranking	Final ranking	Kept into the final BDT
Meff(jet $\gamma\gamma$)	4		-
Meff(jet $\gamma\gamma$ b-jet jet jet)	1	1	x
Meff(b-jet jet jet)	3		-
HT	2		-
HTAllJets	5		-
$max\Delta R$(jet,$\gamma_{1,2}$)	1	3	x
$min\Delta R$(jet,$\gamma_{1,2}$)	4	10	-
$min\Delta R$(b-jet,jet)	2	2	x
$max\Delta\phi$(jet,$\gamma_{1,2}$)	6		-
$min\Delta\phi$(jet,$\gamma_{1,2}$)	8		-
ΔR(jet,$\gamma\gamma$)	10		-
$\Delta\eta$(jet,$\gamma\gamma$)	3	11	-
$\Delta\phi$(jet,$\gamma\gamma$)	12		-
ΔR(b-jet,W)	5		-
$\Delta\phi$(b-jet,W)	7		-
$\Delta\eta$(b-jet,W)	11		-
$min\Delta\phi$(b-jet,jet)	9		-
costthetastar	13	9	-
$m(\mathrm{t\bar{t}})$	3	4	x
$p_T(\mathrm{t\bar{t}})$	8		-
$y(\mathrm{t\bar{t}})$	13		-
$p_T(\gamma\gamma)$	1	7	x
$\eta(\gamma\gamma)$	6		-
p_T(b-jet)	5	5	x
p_T(jet$_0$)	14		-
p_T(jet$_1$)	4	8	-
η(b-jet)	7		-
η(jet$_0$)	9		-
η(jet$_1$)	11		-
m(W)	2	6	x
p_T(jet$_0$)/p_T(jet$_1$))	12		-
p_T(b-jet) / (p_T(jet$_0$)+p_T(jet$_1$))	10		-

TABLE 7.12: Ranking of the BDT training variables in the hadronic channel with pair-production selection. In the loose categories (2A and 2B), very few events have selected jets matching to the partons of the hard process, especially to decay products of the W-boson. Hence, the variables m_W and $m(\mathrm{t\bar{t}})$ are discarded in the training of BDTs dedicated to these categories.

Due to the large number of variables tested, rankings were split between base-kinematic, angular separation, and event-level sets of variables. Variables pre-selected amongst the latter sets are ranked in a final optimisation step. As negligible differences are observed in the kinematics of single-top FCNC processes from the choice of chirality operator, the corresponding samples generated with identical statistics are merged in order to double the effective Monte-Carlo statistics in the BDT training procedure. Only the non-resonant backgrounds are considered in the training procedure. A low discrimination power is expected against the SM Higgs background, but dedicated multivariant could be trained if discriminating variables are identified.

Tables 7.12 and 7.13 present the individual and recombined ranking results for events which satisfy the pair-production and single-top analysis requirements respectively. The ranking of variables is based on a statistical significance estimate[1]. The final choice of training variables results from the ranking plot presented in Figures 7.9 and 7.10. The shapes of the multivariate discriminant in each category are shown in Fig. 7.11.

[1] $\sigma = \sqrt{2((s+b)ln(1+s/b)-s)}$

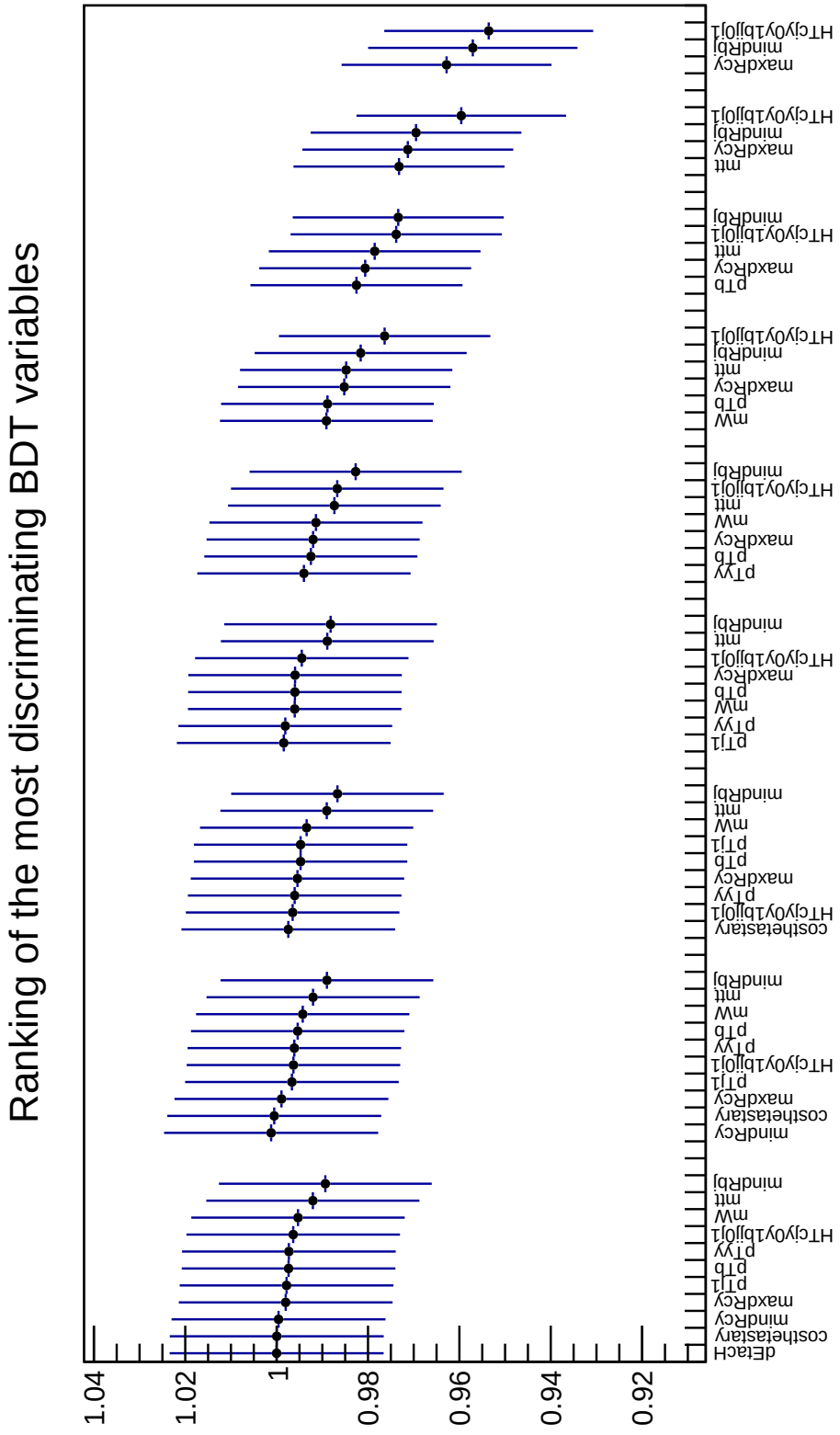


FIGURE 7.9: Ranking of the most discriminating variables in the BDT training for the pair-production categories. The performances are presented in terms of statistical significance for all categories combined. The significances are scaled so that the most performing BDT has significance value equal to one.

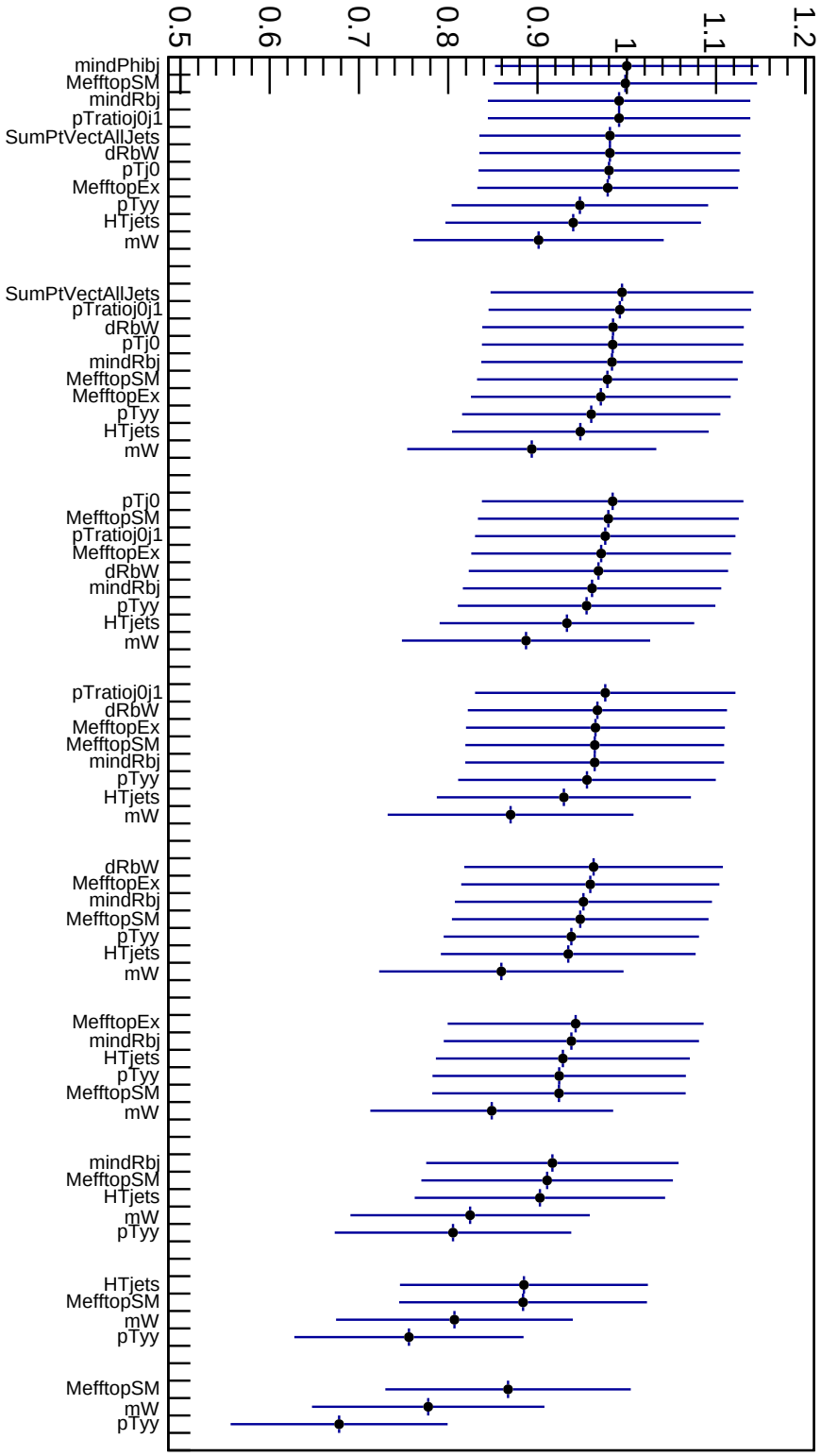
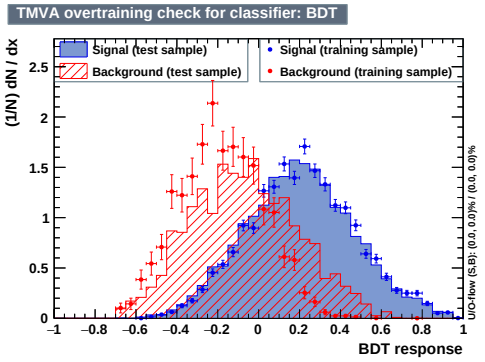
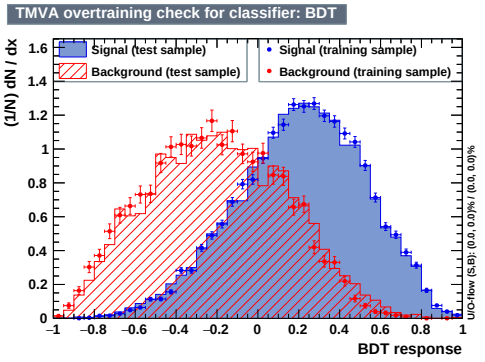


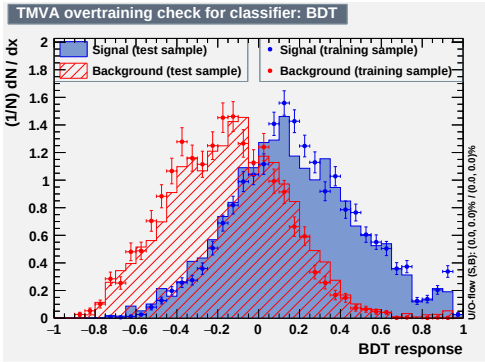
FIGURE 7.10: Ranking of the most discriminating variables in the BDT training for the single-top category. The performances are presented in terms of statistical significance for all categories combined. The significances are scaled so that the most performing BDT has significance value equal to one.



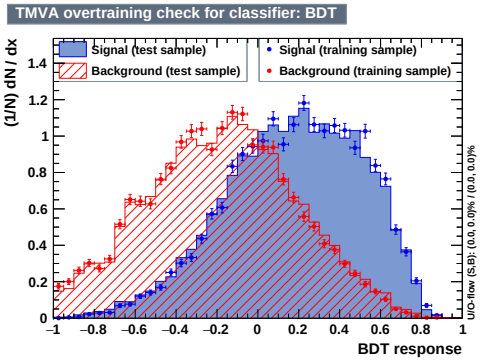
Category 1A



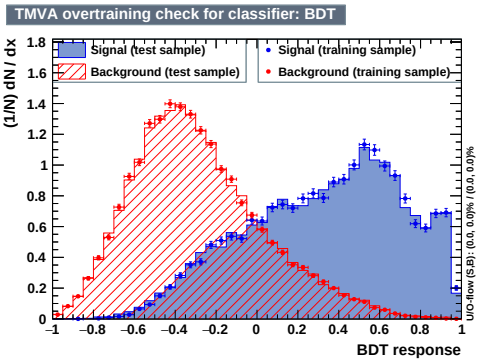
Category 1B



Category 2A



Category 2B



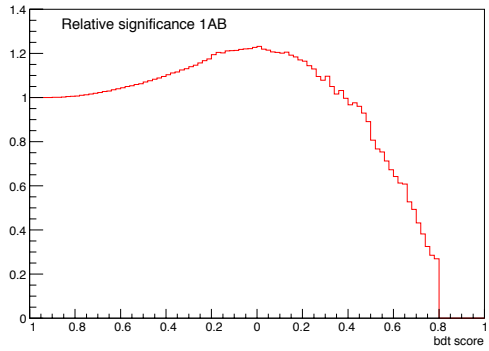
Single-top category

FIGURE 7.11: Shape of the multivariate discriminant in each category of the hadronic channel of the analysis.

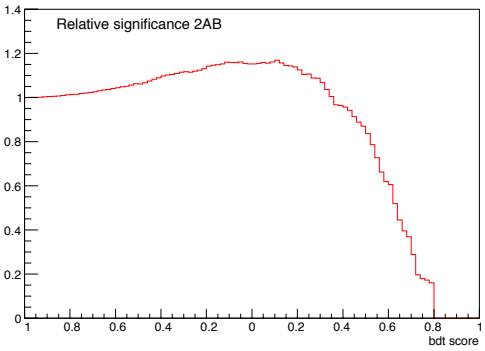
Variable name	Intermediate ranking	Final ranking	Kept into the final BDT
Meff($\gamma\gamma$)	1	6	x
Meff($\gamma\gamma$ b-jet jet jet)	7		-
Meff(b-jet jet jet)	2	3	x
HT	5		-
HTAllJets	3	4	x
Vector sum of selected jets pT	6		-
Vector sum of all jets pT	4	10	-
$min\Delta R$(b-jet,jet)	2	5	x
ΔR(b-jet,W)	1	7	x
$\Delta\phi$(b-jet,W)	6		-
$min\Delta\phi$(b-jet,jet)	3	11	-
$\Delta\eta$(b-jet,W)	4		-
costthetastar	5		-
$p_T(\gamma\gamma)$	1	1	x
$\eta(\gamma\gamma)$	8		-
p_T(b-jet)	7		-
p_T(jet$_0$)	3	9	-
p_T(jet$_1$)	10		-
η(b-jet)	6		-
η(jet$_0$)	5		-
η(jet$_1$)	9		-
m(W)	2	2	x
p_T(jet$_0$)/p_T(jet$_1$))	4	8	-
p_T(b-jet) / (p_T(jet$_0$)+p_T(jet$_1$))	11		-

TABLE 7.13: Ranking of the BDT training variables in the hadronic channel with single-top selection.

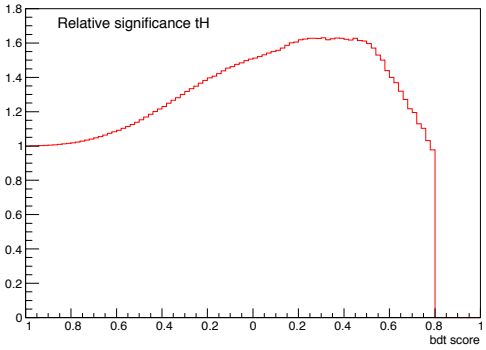
A selection criterion is designed in order to discard events with low multivariate discriminant value. The threshold value is optimised simultaneously over categories of the pair-production and single-top analyses by computing the profile of the statistical significance for different cut values, as presented in Fig. 7.12. The most performing criterion requires the multivariant discriminant to have value greater than 0.1 in the categories targeting the pair-production channel. A separate optimisation in the single-top category shows an optimised cut value of 0.5. This latter requirement has a rather low signal acceptance, but rejects most of the background at the same time. The expected yields after these cuts are presented in Table 7.14.



Pair-production selection, categories 1A/1B



Pair-production selection, categories 2A/2B



Single-top category

FIGURE 7.12: Relative significance as a function of the multivariate discriminant cut value in Signal Regions of the hadronic channel of the analysis. The cut value is optimised simulatenously in the categories 1A and 1B (2A and 2B) of the analysis, and separately in the single-top category, exploiting a statistical significance inclusive on the signal and background event yields.

Sample	Process	Analysis category				
		Had1A	Had1B	Had2A	Had2B	HadtH
FCNC signal Pair-production with t → cH/uH decay	$t\bar{t} \to W(\ell\nu)cH(\gamma\gamma)$	0.117	0.175	0.235	0.368	0.078
	$t\bar{t} \to W(q\bar{q})cH(\gamma\gamma)$	2.178	3.539	1.187	2.281	1.454
	$t\bar{t} \to W(\ell\nu)uH(\gamma\gamma)$	0.023	0.263	0.053	0.569	0.077
	$t\bar{t} \to W(q\bar{q})uH(\gamma\gamma)$	0.508	5.3	0.267	3.302	1.447
FCNC signal Single-top production with tuH coupling	$qg \to W(\ell\nu)uH(\gamma\gamma)$ $(\bar{t_L}\tilde{\phi}u_R)$	0.002	0.008	0.001	0.018	0.093
	$qg \to W(q\bar{q})uH(\gamma\gamma)$ $(\bar{t_L}\tilde{\phi}u_R)$	0.033	0.265	0.037	0.266	2.2
	$qg \to W(\ell\nu)uH(\gamma\gamma)$ $(\bar{u_L}\tilde{\phi}t_R)$	0	0.013	0.004	0.026	0.092
	$qg \to W(q\bar{q})uH(\gamma\gamma)$ $(\bar{u_L}\tilde{\phi}t_R)$	0.03	0.263	0.031	0.203	2.072
FCNC signal Single-top production with tcH coupling	$qg \to W(\ell\nu)cH(\gamma\gamma)$ $(\bar{c_L}\tilde{\phi}t_R)$	0	0.003	0.001	0.003	0.014
	$qg \to W(q\bar{q})cH(\gamma\gamma)$ $(\bar{c_L}\tilde{\phi}t_R)$	0.011	0.06	0.009	0.051	0.317
	$qg \to W(\ell\nu)cH(\gamma\gamma)$ $(\bar{t_L}\tilde{\phi}c_R)$	0	0.002	0	0.005	0.013
	$qg \to W(q\bar{q})cH(\gamma\gamma)$ $(\bar{t_L}\tilde{\phi}c_R)$	0.013	0.066	0.008	0.055	0.305
Total signal tuH		0,60	6,11	0,39	4,38	5,98
Total signal tcH		2,32	3,85	1,44	2,76	2,18
SM Higgs resonant background	Gluons fusion (ggH)	0.009	0.074	0.022	0.217	0.509
	Vector Boson Fusion (VBFH)	0	0.004	0.003	0.036	0.04
	W-boson associated production (WH)	0.001	0.03	0.007	0.046	0.074
	Z-boson associated production (ZH)	0.005	0.029	0.007	0.066	0.097
	top-quarks associated production ($t\bar{t}H$)	0.473	1.769	0.37	1.683	1.291
	b-quarks associated production (bbH)	0	0.001	0.002	0.026	0.003
	tHjb	0.008	0.043	0.018	0.093	0.143
	single-top associated production (tWH)	0.003	0.019	0.005	0.031	0.058
Total SM Higgs		0,499	1,969	0,434	2,198	2,215
Non-resonant background	Diphoton + jets	6.72	34.7	14.34	95.4	27.1
	$t\bar{t} + \gamma$	2.3	8.07	1.538	8.48	3.31
	V+$\gamma\gamma$	0.013	0.011	0.006	0.052	0.006
Total background		9.53	44.8	16.3	106.1	32.6
Data blinded		6	36	15	75	20

TABLE 7.14: Expected yields after applying the BDT cut for an integrated luminosity of 80.1 fb^{-1} based on Monte-Carlo simulations for FCNC signal, SM Higgs and non-resonant backgrounds considered in the search. Data are blinded in the range 120 GeV < $m_{\gamma\gamma}$ < 130 GeV.

7.7 Statistical analysis

The branching ratio of the t $\to qH$ decay is extracted from a fit to data using a distinct likelihood component for each category of the analysis. The likelihood function used in the combined fit consists of the product of the latter likelihoods. In the hadronic channel, an unbinned fit to the $m_{\gamma\gamma}$ distribution is performed, while the leptonic estimate is based on event counting in a signal bin (121 GeV < $m_{\gamma\gamma}$ < 129 GeV) and a control bin (105 GeV < $m_{\gamma\gamma}$ < 121 GeV and 129 GeV < $m_{\gamma\gamma}$ < 160 GeV). In the following, the fit procedure is detailed for the hadronic selection which requires a specific method to find the most appropriate parameterisation of the shape of the $m_{\gamma\gamma}$ spectrum in signal and background processes respectively. An additional study is dedicated to the uncertainty estimate related to this parameterisation.

7.7.1 Fit model

In each Signal Region of the analysis, a double-sided Crystal Ball function (a Gaussian function with power-law tails on both sides) is used to describe the shape of $m_{\gamma\gamma}$ for signal processes and the resonant SM Higgs backgrounds :

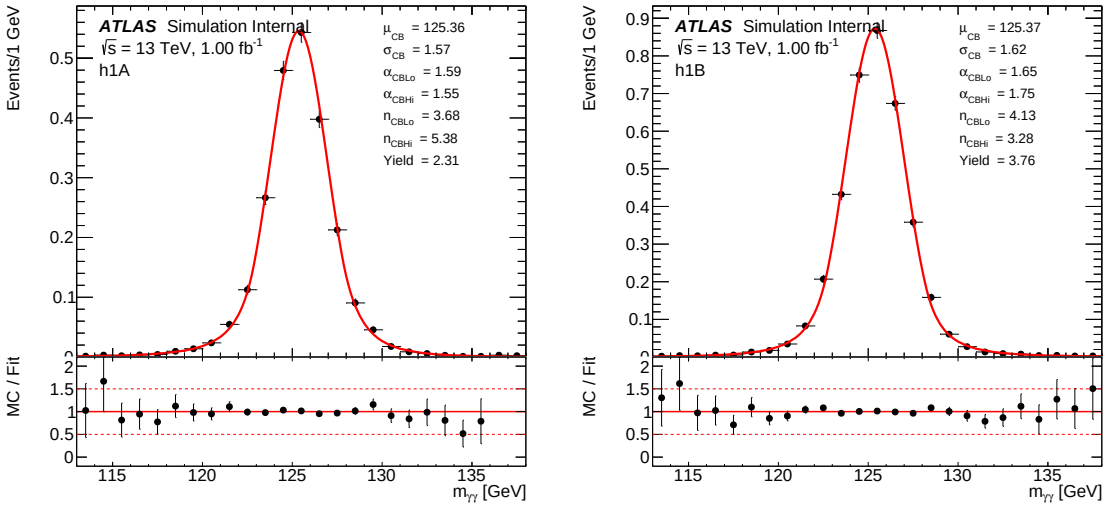
$$\mathcal{N} \cdot \begin{cases} e^{-t^2/2} & \text{if } -\alpha_- \leq t \leq \alpha_+ \\ e^{-\alpha_-^2/2} \left[\frac{\alpha_-}{n_-} \left(\frac{n_-}{\alpha_-} - \alpha_- - t \right) \right]^{-n_-} & \text{if } t < -\alpha_- \\ e^{-\alpha_+^2/2} \left[\frac{\alpha_+}{n_+} \left(\frac{n_+}{\alpha_+} - \alpha_+ + t \right) \right]^{-n_+} & \text{if } t > \alpha_+, \end{cases} \tag{7.2}$$

Where :

- $\mathcal{N}$ is a normalisation parameter.
- $t = \Delta m_X / \sigma_{CB}$
- $\Delta m_X = m_X - \mu_{CB}$
- μ_{CB} is the peak of the Gaussian distribution.
- σ_{CB} represents the width of the Gaussian part of the function.
- α_- (α_+) is the point where the Gaussian becomes a power law on the low (high) mass side ($\alpha_\pm$>0).
- n_- (n_+) is the exponent of this power law.

The function parameters are extracted from a fit to the mass spectrum of the FCNC signal in Monte-Carlo predictions after selection, as illustrated in Figure 7.13 for two out of the five categories. The same shape is used to model the resonant SM Higgs background. For this reason, the accuracy of yield estimates for these processes is crucial to the sensitivity of the analysis.

Different background models are tested to describe the shape of $m_{\gamma\gamma}$ for the non-resonant backgrounds : a decreasing exponential, second-, third- and fourth-order polynomial. The parameterisation choice can cause a bias on the number of signal events estimated from the fit to data. Hence, a systematic uncertainty is associated to the function model used to describe the $m_{\gamma\gamma}$ of background processes. We denote as *spurious signal* the number of fitted signal events which are induced by the inaccuracy of the background model used in the fit [106]. In the current fit model, this corresponds to the fitted signal yield in a MC dataset filled with background events only. A background model with too many free parameters will adjust every statistical fluctuation in data, so no signal event will be observed. On the other hand, a background model with too few parameters is likely to give a wrong description of the background



Category 1A Category 1B

FIGURE 7.13: Signal shape parameterisation of the $m_{\gamma\gamma}$ spectrum in the tightest Signal Regions of the pair-production hadronic channel of the analysis.

shape and might lead to a bias of the fitted signal yield. The background model for the $m_{\gamma\gamma}$ spectrum is required to induce an average spurious signal lower than 10% of the number of expected signal events at the limit, or smaller than 20% of the corresponding uncertainty. An optimisation study is performed, similar to that presented in Section 7.9. In every category, the shape of the $m_{\gamma\gamma}$ spectrum is taken as a distinct third-order polynomial function, whose coefficients are extracted from a fit to data.

In the preliminary results presented in Section 7.8, the number of spurious signal events are extrapolated from those estimated in the 2017 search, and shown in Table 7.15. These values are scaled to the available integrated luminosity of 80.1 fb^{-1} which meets the data taking period from 2015 to 2017. The spurious signal is also scaled in order to account for the lower background acceptance due to the use of a multivariate discriminant. The new categorisation with c-tagging categories is the major obstacle to this extrapolation. Hence, the number of spurious signal events are defined so that the sum of the spurious signal in tightest categories with and without a c-tagged jet (1A and 1B) meet that of the tightest category in the 2017 paper, and similarly in the loose categories of the analysis. The spurious signal in the single-top category is further extrapolated from that of the tightest category, which presents a similar phase space regarding the selection requirements. A preliminary version of the updated procedure for estimating the uncertainty related to the parameterisation of $m_{\gamma\gamma}$ in background events is presented in Section 7.9.

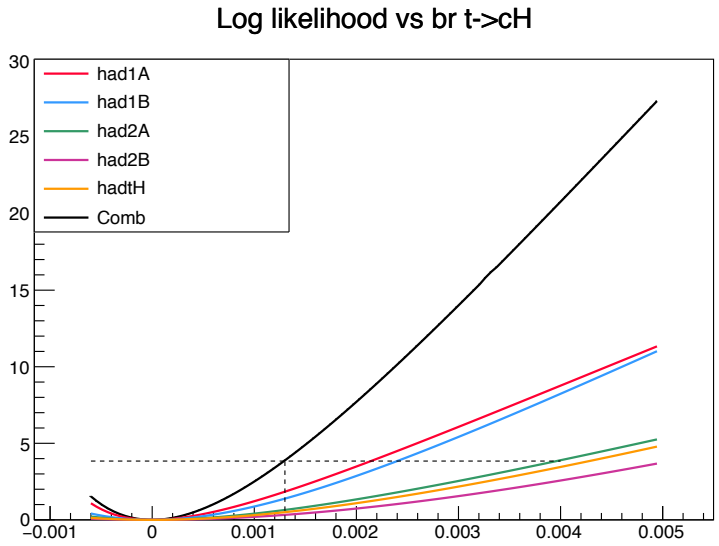
Similarly, experimental systematic uncertainties are extrapolated from the 2017 publication. The values for systematic uncertainties used in the tight categories of the pair-production selection and in the single-top selection are duplicated from that which were estimated in the tight category of the paper, and treated as correlated. The same procedure is performed in the loose-categories of the analysis respectively. A major drawback is the treatment of c-tagging uncertainties, which are not considered in the following.

Category	Pair-production				Single-top
	1A	1B	2A	2B	
SM Higgs background yield	0.92	3.58	0.77	3.77	3.42
Spurious signal	0.18	0.95	0.48	3.21	0.95

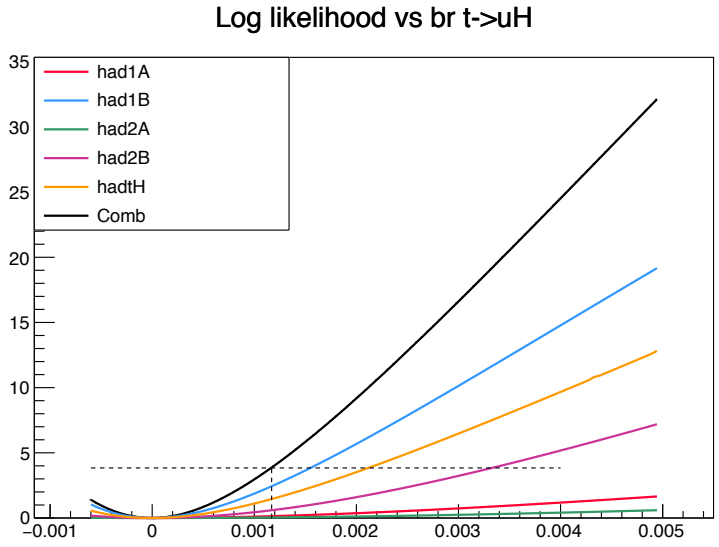
TABLE 7.15: Expected SM Higgs background yields based on Monte-Carlo simulations, and spurious signal estimates extrapolated from the 2017 publication. The relative uncertainty on the SM Higgs background yields is 15% (20%) in the categories 1A/1B/tH (2A/2B) respectively.

7.8 Prospects

The sensitivity of the analysis is measured with respect to the Branching ratios for the t $\rightarrow$ cH and t $\rightarrow$ uH decays in the range $0.5 \times 10^{-3} \leq$ Br $\leq 5 \times 10^{-3}$ (not both non-zero simultaneously), taking support from the limits observed in the 2017 search. This procedure exploits a test statistic based on the profile likelihood ratio. The fit is performed to a background Asimov distribution taken from a third-order polynomial fitted to the $m_{\gamma\gamma}$ shape of the non-resonant simulated backgrounds, normalized to an integrated luminosity of 80.1 fb^{-1}. The profile likelihood ratio is presented in Fig. 7.14 for the t $\rightarrow$ cH and t $\rightarrow$ uH decay processes separately. The dotted line presents the intersection of the combined fit likelihood ratio to the Branching ratio axis with a 95% confidence level, which occurs at Br(t $\rightarrow$ cH) = 0.113% and Br(t $\rightarrow$ uH) = 0.112% respectively. The tightest categories (1A and 1B) of the pair-production selection are the most sensitive to the t $\rightarrow$ cH process. On the other hand, the tight category with no c-tagged jet and the single-top category contribute the most to the sensitivity of the analysis to t $\rightarrow$ uH. Some caution should be taken in this prospect result due to the absence of c-tagging uncertainties in this study.



$t \to cH$



$t \to uH$

FIGURE 7.14: Profile likelihood ratio as a function of the Branching ratio for the $t \to qH$ decay exploiting the hadronic channel of the analysis. The figures show the result from fits performed in each category individually and in a combined fit respectively, for an integrated luminosity of 80.1 fb^{-1}. Br($t \to uH$) is fixed to zero in the profile likelihood ratio for the $t \to cH$ process.

7.9 Update of the background shape parameterisation

7.9.1 Preliminary estimate using the procedure from the 2017 publication

A proper estimate of the spurious signal is performed separately in each of the new categories. First, the $m_{\gamma\gamma}$ spectrum of the non-resonant simulated backgrounds is smoothed using the $RooKeysPdf$ implementation provided in the ROOT analysis framework [107]. Different $mirrors$ are tested in order to deal with the boundary conditions, and the smoothing parameter ρ is varied in order to achieve the best agreement with the $m_{\gamma\gamma}$ spectrum of Monte-Carlo predictions. The results of the smoothing procedure are illustrated in Fig. 7.15. Then, toy experiments are generated from the smoothed shape of $m_{\gamma\gamma}$ and simulate the $m_{\gamma\gamma}$ spectrum with respect to the statistics available in data. Within each toy experiment, the background model is fitted to the shape of the simulated $m_{\gamma\gamma}$ spectrum. Different functions are tested as background model in order to achieve a good compromise between a low spurious signal and a smooth shape fitting data. The difference measured in the Signal Region between the fitted and the true number of events is the bias denoted as spurious signal for the toy experiment. This procedure is repeated for a large number of toys, and the average value of the spurious signal is selected as final estimate. Alternatively, a fit model including the signal allows to measure the spurious signal as the number of fitted signal events. This latter method is used to verify the validity of the first method, and a good agreement is observed.

Figure 7.16 presents the distribution of spurious signal yields observed in toy experiments for each category individually. In the tightest category of the pair-production selection with a c-tagged jet (category 1A), this procedure fails because of the very low number of data events, hence issues in the fitting procedure. A solution under study consists of generating toys with a much larger number of events, and rescaling back the fitted spurious signal to the statistics measured in data, as described in Section 7.9.2.2. The decreasing exponential and second-order polynomial do not describe correctly the background shape, hence overestimates of the spurious signal. Conversely, the fourth-order polynomial gives very low spurious signal estimates in every category, but is likely to partly absorb the signal resonance in the $m_{\gamma\gamma}$ of data. The third-order polynomial appears as a reasonnable compromise, and is selected for the final spurious signal estimate and background model in the fit to data.

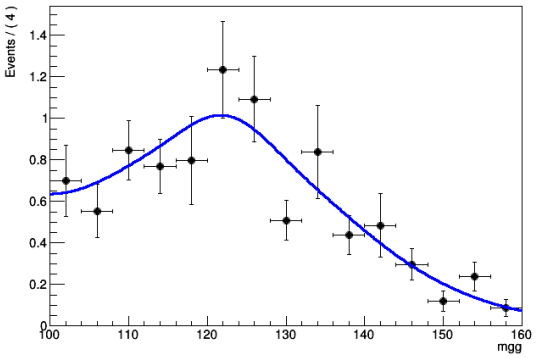
7.9.2 Update of the procedure for spurious signal estimates

7.9.2.1 Design of Control Regions

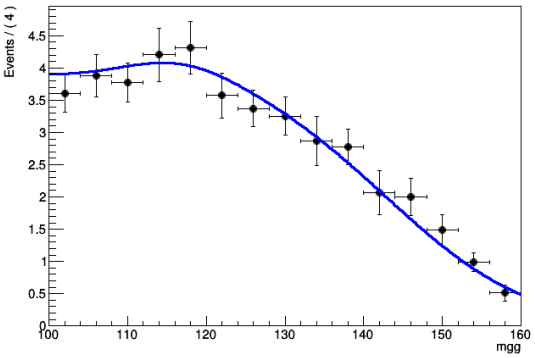
The low number of data events observed in the sensitive categories of the analysis causes uncertainties on the choice of the function used as background model. Especially, the estimate of the number of spurious signal events is driven by a procedure which exploits the shape of the $m_{\gamma\gamma}$ spectrum of simulation samples for the non-resonant backgrounds. In the final fit, the parameters of the background model are fully extracted from the fit to data. Yet, the spurious signal estimate is correct in the limit that the selected function must show enough flexibility to the shape of the $m_{\gamma\gamma}$ spectrum of data.

In order to provide a more robust approach, Control Regions are designed by reverting the isolation or identification requirements on the selected photons. The nominal selection requires the two photons with highest transverse momentum to satisfy loose isolation and tight identification requirements (tight isolated photons). Two orthogonal Control Regions are created, regarding the multiplicity of tight isolated photons, either zero or one. The event yields for the different tight isolated photons multiplicities are presented in Table 7.16. In the Signal Region, this is compared to the expected yield for the dominant non-resonant background.

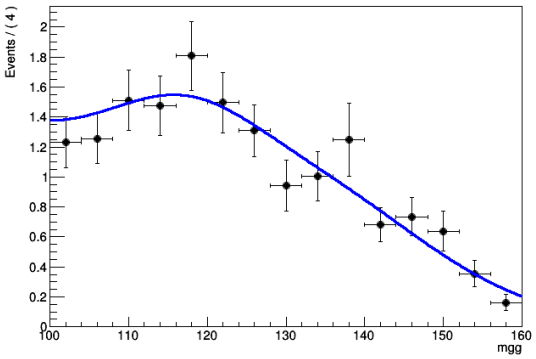
A comparison of the shape of the $m_{\gamma\gamma}$ spectrum is illustrated in Fig. 7.17 and 7.18 before and after applying the requirement on the multivariate discriminant. A reasonable agreemeent



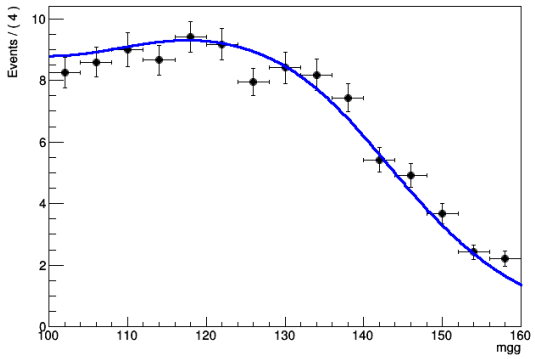
Category 1A



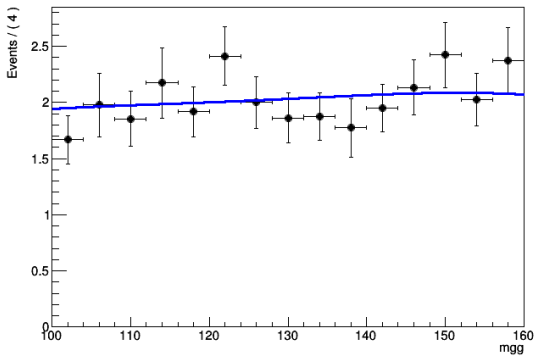
Category 1B



Category 2A

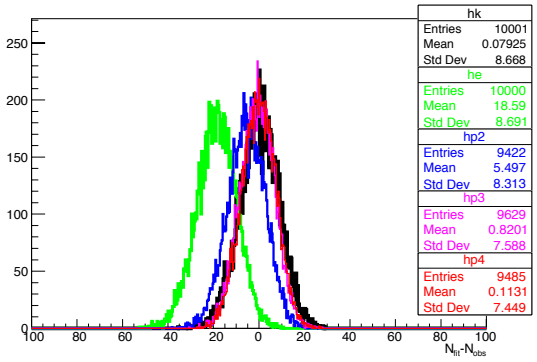


Category 2B

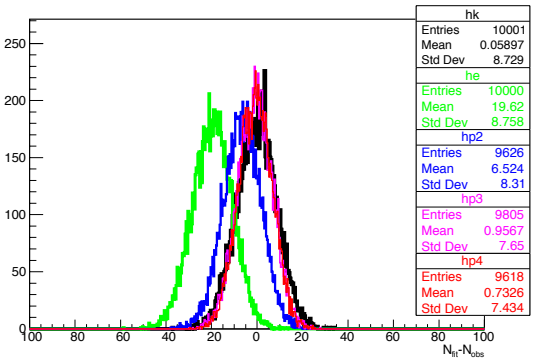


Category tH

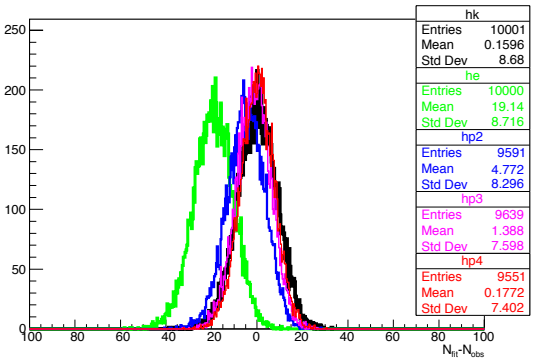
FIGURE 7.15: Shape of the $m_{\gamma\gamma}$ spectrum smoothed in each category individually. The slightly increasing shape in the single-top category is found to decrease at higer values of $m_{\gamma\gamma}$. This shape is likely to be caused by kinematic cuts and studies are ongoing to better understand it.



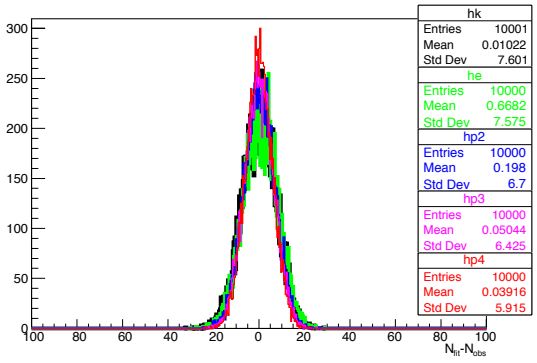
Category 1B



Category 2A



Category 2B



Category tH

FIGURE 7.16: Distributions of the spurious signal estimate in toy experiments generated in each category individually. In the category 1A, the procedure fails because of the very low number of data events, hence issues in the fitting procedure. A slighlty modified procedure is under study and should allow to avoid this issue. In each figure, he and $hp_{2,3,4}$ are the decreasing exponential and polynomial models respectively. In the current procedure, the mean of the distribution meets the final spurious signal estimate. The model denoted as hk consists of the smoothed shape used to generate the toys, and presents the smallest bias as expected.

	data			MC $\gamma\gamma$+jets
Requirement (isolated tight photons)	0 photon	1 photon	2 photons	2 photons
Had 1A	27	31	6	6,7
Had 1B	200	176	36	34,6
Had 2A	39	68	15	14,3
Had 2B	527	453	75	95,3
Had tH	107	178	20	27,1

TABLE 7.16: Observed number of data with reverted isolation cuts. In the selection requiring two isolated and tight photons in data, events are blinded in the range 120 GeV < $m_{\gamma\gamma}$ < 130 GeV.

is observed, which could motivate the use of these Control Regions in the procedure for estimating the spurious signal.

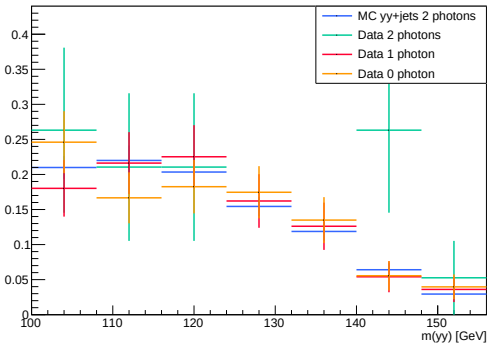
7.9.2.2 Statistical issues in toys

The procedure which allows to estimate the number of spurious signal events exploits toy experiments. Within each toy, a distribution of the $m_{\gamma\gamma}$ spectrum for the non-resonant backgrounds is simulated with the statistics available in data from the shape estimated on Monte-Carlo simulations. Due to the very low statistics in the categories of the analysis, this approach causes troubles in the fit procedure due to the low number of constraints. An alternative approach is being explored, which consists of generating the distribution of the $m_{\gamma\gamma}$ spectrum for a much higher integrated luminosity. The spurious signal estimate is then rescaled to the available integrated luminosity.

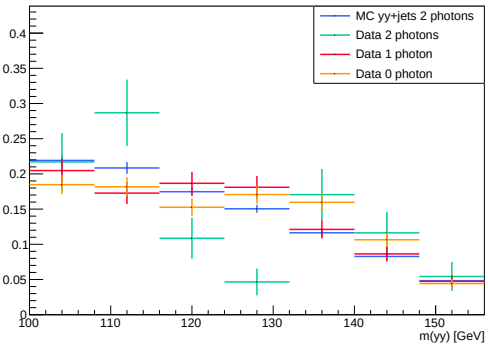
7.10 Conclusion

Preliminary studies are presented, aiming at improvements to the search for the decay of a top quark into an up-type (q = u, c) quark and a Higgs boson decaying into two photons in the ATLAS experiment. Most ideas are dedicated to the hadronic channel of the analysis. An additional selection aiming at the single-top production channel is introduced, which allows to disentangle the tcH and tuH couplings. Moreover, a new categorisation using c-tagging is designed in order to target separately the t $\to$ cH and t $\to$ uH decays. New techniques are introduced to improve the analysis such as the use of additional kinematic variables which are combined into a multivariate discriminant separating signal and background events. Especially, an optimisation of the combinatoric approach is presented, in order to select the combination of jets which looks the most to signal events.

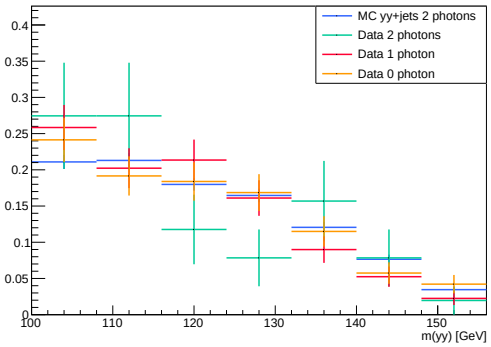
Yet, the ideas presented above bring new complications to the analysis. The use of c-tagging is limited since b- and c-tagging discriminants are calibrated separately for now. The new categories present a low statistics, which causes troubles to estimate the spurious signal and find the most appropriate shape to the $m_{\gamma\gamma}$ spectrum. The use of Control Regions with photon isolation and identification requirements reverted is being studied, and the full run 2 dataset will be used for the next publication of the search which will benefit from the increased integrated luminosity.



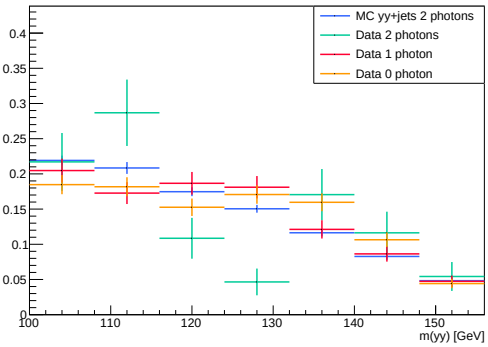
Category 1A



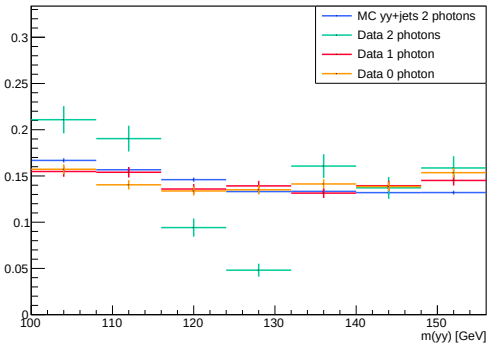
Category 1B



Category 2A

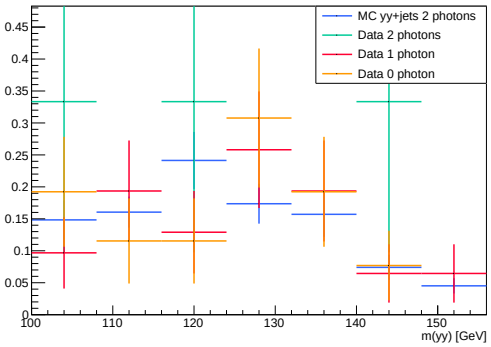


Category 2B

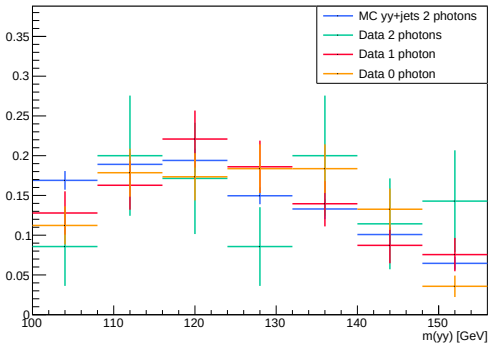


Single-top category

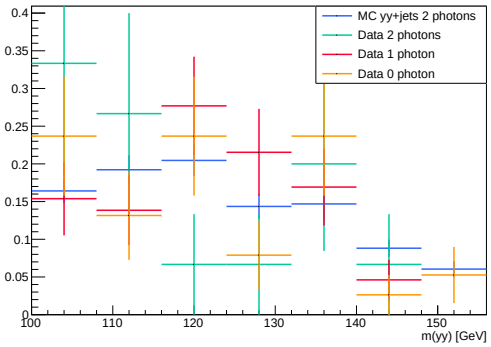
FIGURE 7.17: Shape of the $m_{\gamma\gamma}$ spectrum in data and $\gamma\gamma$+jets simulations depending on the multiplicity of tight isolated photons ($n\ photon(s)$ in the legend) before applying the BDT cut. The deficit in the distribution of data with two isolated photons is due to the blinding in the range 120 GeV < $m_{\gamma\gamma}$ < 130 GeV.



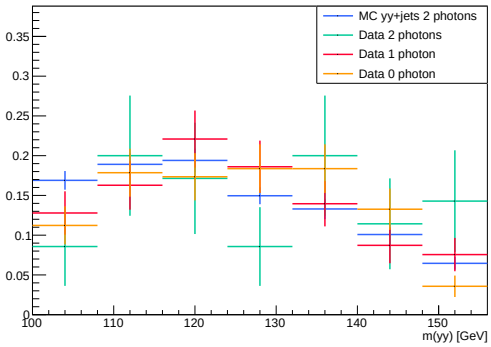
Category 1A



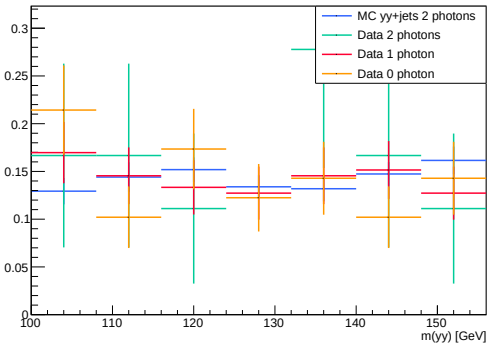
Category 1B



Category 2A



Category 2B



Single-top category

FIGURE 7.18: Shape of the $m_{\gamma\gamma}$ spectrum in data and $\gamma\gamma$+jets simulations depending on the multiplicity of tight isolated photons ($n\ photon(s)$ in the legend) after applying the BDT cut. The deficit in the distribution of data with two isolated photons is due to the blinding in the range 120 GeV < $m_{\gamma\gamma}$ < 130 GeV.

Conclusion

In this thesis, three aspects of my work within the ATLAS experiment are described. The first one is dedicated to the extraction of correction factors to be applied on the simulation to match the data. The second is the analysis of the Higgs boson decay to a pair of b-quarks. The last one is a search for BSM in the interaction between a top quark and the Higgs boson.

My implication in the ATLAS collaboration started with updates of the pseudo-continuous calibration of b-jets flavour identification. The pseudo-continuous calibration of b-Tagging in the ATLAS experiment was developped during the run 1 of the LHC. An update of the framework was performed, and calibration results have been produced. The uncertainties on the pseudo-continuous b-Tagging calibration are dominated by systematic sources.

The second topic of the thesis is the search for a Higgs boson decaying into beauty quarks, produced in association with a vector boson. I was especially involved in the channel looking for a Z boson decaying into neutrinos. The analysis was optimised with respect to the experience acquired during the run 1 of the LHC. Many ideas have been explored in order to improve the multivariate analysis. The evidence for the VH(bb) process was presented using data collected in 2015 and 2016. The signal strength is measured to $\hat{\mu}_{VH} = 1.20^{+0.24}_{-0.23}(stat.)^{+0.34}_{-0.28}(syst)$, which corresponds to an observed significance of 3.5 standard deviations. The uncertainty on this result is dominated by systematic sources. Hence, several studies have been performed in order to better understand specific aspects of this result. A special attention was given to the modelling of background processes to improve the generation efficiency of Monte-Carlo simulation. Also, the use of coulourflow in the VH(bb) search appears to provide a limited gain in the rejection of background processes with the current jet reconstruction performances. The analysis was updated adding data collected in 2017, and lead to the observation of the Higgs boson decay into beauty quarks.

At last, the final study presented in this work is a major revision of the search for a top quark decaying into an up-type (q = u, c) quark and a Higgs boson decaying into two photons. This process is forbiden at tree level in the SM and highly supressed at loop level, with an expected branching ratio of 10^{-15}, making it a very good probe for BSM physics. The analysis is now designed in order to disentangle the tuH and tcH couplings by targetting both the single- and pair-production modes of top quarks, and exploiting c-tagging in the categorisation. The use of kinematic variables was developped, and multivariate methods have been implemented in the analysis. Preliminary prospect results are presented. With the set of data accumulated by the end of the LHC run 2, the analysis will be sensitive to branching ratio below the permil level and could provide the first constraints on realistic BSM models.

Remerciements

Je remercie Achille Stocchi de m'avoir accueilli au Laboratoire de l'Accélérateur Linéaire ainsi que d'avoir accepté de faire partie de mon jury.

Je remercie également Giovanni Marchiori et Paolo Meridiani d'avoir d'avoir accepté d'être rapporteurs, ainsi que Matteo Cacciari et Frédéric Deliot en tant qu'examinateurs de ma thèse.

J'adresse mes remerciements à Jean-Baptiste de Vivie, pour ta patience, tes enseignements, et plus généralement pour le plaisir que j'ai eu à travailler avec toi.

Merci à Daniel Fournier pour tous nos échanges, tant à propos de l'analyse que d'autres sujets divers, j'ai beaucoup appris à ton contact.

J'ai pu bénéficier de l'aide de nombreuses personnes au cours de mon travail de thèse, et souhaite à ce titre remercier Nicolas Morange et Marumi Kado. Merci également au groupe Hbb et notamment Jean-François Grivaz, David Delgove ainsi que Camilla Maiani pour l'aide et les conseils dont ils m'ont fait profiter alors que je commençais ma thèse.

Merci à l'ensemble du groupe ATLAS du Laboratoire de l'Accélérateur Linéaire de m'avoir accueilli et permis de réaliser une thèse dans les meilleures conditions possibles. Je remercie également les étudiants du laboratoire pour la bonne ambiance dont j'ai pu largement profiter.

Le travail de thèse ne m'a laissé que trop peu de temps libre, mais je garderai d'excellents souvenirs des tentatives de mensonge de Mathieu au loup-garou, des destructions de décor de Florian dans Broforce, et de Timothée ... et sa bonne humeur ! Enfin, merci à Mitaines, Scorpi et Onze pour nos explorations, et les découvertes, rencontres et fêtes qu'elles ont amenées, et qui m'ont enrichi bien plus que je ne saurais le dire.

J'ai beaucoup de chance d'avoir ma famille pour me soutenir et m'encourager, et souhaite la remercier très sincèrement.

Bibliography

[1] William Rowan Hamilton. "XV. On a general method in dynamics; by which the study of the motions of all free systems of attracting or repelling points is reduced to the search and differentiation of one central relation, or characteristic function". en. In: *Philosophical Transactions of the Royal Society of London* 124 (Jan. 1834), pp. 247–308. ISSN: 0261-0523, DOI: 10.1098/rstl.1834.0017. URL: http://rstl.royalsocietypublishing.org/content/124/247.

[2] Emmy Noether. "Invariant variation problems". In: *Transport Theory and Statistical Physics* 1.3 (Jan. 1971), pp. 186–207. ISSN: 0041-1450. DOI: 10.1080/00411457108231446. URL: https://doi.org/10.1080/00411457108231446.

[3] Abdus Salam and J. C. Ward. "Weak and electromagnetic interactions". en. In: *Il Nuovo Cimento (1955-1965)* 11.4 (Feb. 1959), pp. 568–577. ISSN: 1827-6121. DOI: 10.1007/BF02726525. URL: https://link.springer.com/article/10.1007/BF02726525.

[4] Steven Weinberg. "A Model of Leptons". In: *Physical Review Letters* 19.21 (Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[5] Sheldon L. Glashow. "The renormalizability of vector meson interactions". In: *Nuclear Physics* 10 (Feb. 1959), pp. 107–117. ISSN: 0029-5582. DOI: 10.1016/0029-5582(59)90196-8. URL: http://www.sciencedirect.com/science/article/pii/0029558259901968.

[6] M. Gell-Mann. "Isotopic Spin and New Unstable Particles". In: *Physical Review* 92.3 (Nov. 1953), pp. 833–834. DOI: 10.1103/PhysRev.92.833. URL: https://link.aps.org/doi/10.1103/PhysRev.92.833.

[7] Kazuhiko Nishijima. "Some Remarks on the Even-odd Rule". en. In: *Progress of Theoretical Physics* 12.1 (July 1954), pp. 107–108. ISSN: 0033-068X. DOI: 10.1143/PTP.12.107. URL: https://academic.oup.com/ptp/article/12/1/107/1897250.

[8] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Physical Review Letters* 10.12 (June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[9] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". en. In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652. URL: https://academic.oup.com/ptp/article/49/2/652/1858101.

[10] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Physical Review Letters* 13.9 (Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[11] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physical Review Letters* 13.16 (Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[12] *CERN Yellow Report At 13 TeV, LHC Physics*. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV.

[13] *CERN Yellow Report BR, LHC Physics*. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR.

[14] Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. "Search for Neutrinos from the Sun". In: *Physical Review Letters* 20.21 (May 1968), pp. 1205–1209. DOI: 10.1103/PhysRevLett.20.1205. URL: https://link.aps.org/doi/10.1103/PhysRevLett.20.1205.

[15] The Super-Kamiokande Collaboration and Y. Fukuda et al. "Evidence for oscillation of atmospheric neutrinos". In: *Physical Review Letters* 81.8 (Aug. 1998). arXiv: hep-ex/9807003, pp. 1562–1567. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.81.1562. URL: http://arxiv.org/abs/hep-ex/9807003.

[16] B. Pontecorvo. "Neutrino Experiments and the Problem of Conservation of Leptonic Charge". In: *Sov.Phys.JETP* 26 (1968), pp. 984–988.

[17] F. Zwicky. "On the Masses of Nebulae and of Clusters of Nebulae". en. In: *The Astrophysical Journal* 86 (Oct. 1937), p. 217. ISSN: 0004-637X. DOI: 10.1086/143864. URL: http://adsabs.harvard.edu/abs/1937ApJ....86..217Z.

[18] Edwin Hubble. "A relation between distance and radial velocity among extra-galactic nebulae". en. In: *Proceedings of the National Academy of Sciences* 15.3 (Mar. 1929), pp. 168–173. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.15.3.168. URL: http://www.pnas.org/content/15/3/168.

[19] R. D. Peccei and Helen R. Quinn. "CP conservation in the presence of pseudoparticles". In: *Physical Review Letters* 38 (June 1977), pp. 1440–1443. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.38.1440. URL: http://adsabs.harvard.edu/abs/1977PhRvL..38.1440P.

[20] Michael Gronau, C. N. Leung, and Jonathan L. Rosner. "Extending limits on neutral heavy leptons". In: *Physical Review D* 29.11 (June 1984), pp. 2539–2558. DOI: 10.1103/PhysRevD.29.2539. URL: https://link.aps.org/doi/10.1103/PhysRevD.29.2539.

[21] *The ATLAS Experiment at the CERN Large Hadron Collider - IOPscience*. URL: http://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003/meta.

[22] Alessandro La Rosa. "The ATLAS Insertable B-Layer: from construction to operation". In: *arXiv:1610.01994 [physics]* (Oct. 2016). arXiv: 1610.01994. URL: http://arxiv.org/abs/1610.01994.

[23] Bartosz Mindur. *ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider*. fr. Mar. 2016. DOI: 10.1016/j.nima.2016.04.026, 10.1016/j.nima.2016.04.026. URL: https://cds.cern.ch/record/2139567.

[24] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008). arXiv: 0802.1189, pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://arxiv.org/abs/0802.1189.

[25] *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*. fr. Aug. 2015. URL: https://cds.cern.ch/record/2044941.

[26] S. Mandelstam. "Determination of the Pion-Nucleon Scattering Amplitude from Dispersion Relations and Unitarity. General Theory". In: *Phys. Rev.* 112 (4 1958), pp. 1344–1360. DOI: 10.1103/PhysRev.112.1344. URL: https://link.aps.org/doi/10.1103/PhysRev.112.1344.

[27] V.V. Sudakov. "Vertex Parts at Very High Energies in Quantum Electrodynamics". In: *JETP* Vol. 3, No. 1 (July 1956), p. 65. URL: http://www.jetp.ac.ru/cgi-bin/e/index/e/3/1/p65?a=list.

[28] G. 't Hooft. "A planar diagram theory for strong interactions". In: *Nuclear Physics B* 72.3 (Apr. 1974), pp. 461–473. ISSN: 0550-3213. DOI: 10.1016/0550-3213(74)90154-0. URL: http://www.sciencedirect.com/science/article/pii/0550321374901540.

[29] S. Frixione and B. R. Webber. "Matching NLO QCD computations and parton shower simulations". In: *Journal of High Energy Physics* 2002.06 (June 2002). arXiv: hep-ph/0204244, pp. 029–029. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2002/06/029. URL: http://arxiv.org/abs/hep-ph/0204244.

[30] Paolo Nason. "A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms". In: *Journal of High Energy Physics* 2004.11 (Nov. 2004). arXiv: hep-ph/0409146, pp. 040–040. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2004/11/040. URL: http://arxiv.org/abs/hep-ph/0409146.

[31] S. Jadach et al. "Matching NLO QCD with parton shower in Monte Carlo scheme - the KrkNLO method". en. In: (Mar. 2015). DOI: 10.1007/JHEP10(2015)052. URL: https://arxiv.org/abs/1503.06849.

[32] S. Catani et al. "QCD Matrix Elements + Parton Showers". en. In: *Journal of High Energy Physics* 2001.11 (Nov. 2001). arXiv: hep-ph/0109231, pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2001/11/063. URL: http://arxiv.org/abs/hep-ph/0109231.

[33] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". In: *Journal of High Energy Physics* 2012.12 (Dec. 2012). arXiv: 1209.6215. ISSN: 1029-8479. DOI: 10.1007/JHEP12(2012)061. URL: http://arxiv.org/abs/1209.6215.

[34] Stefan Hoeche et al. "QCD matrix elements + parton showers: The NLO case". In: *Journal of High Energy Physics* 2013.4 (Apr. 2013). arXiv: 1207.5030. ISSN: 1029-8479. DOI: 10.1007/JHEP04(2013)027. URL: http://arxiv.org/abs/1207.5030.

[35] Leif Lonnblad and Stefan Prestel. "Unitarising Matrix Element + Parton Shower merging". In: *Journal of High Energy Physics* 2013.2 (Feb. 2013). arXiv: 1211.4827. ISSN: 1029-8479. DOI: 10.1007/JHEP02(2013)094. URL: http://arxiv.org/abs/1211.4827.

[36] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. "MINLO: Multi-scale improved NLO". In: *Journal of High Energy Physics* 2012.10 (Oct. 2012). arXiv: 1206.3572. ISSN: 1029-8479. DOI: 10.1007/JHEP10(2012)155. URL: http://arxiv.org/abs/1206.3572.

[37] B. Andersson et al. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31 –145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7. URL: http://www.sciencedirect.com/science/article/pii/0370157383900807.

[38] Thomas D. Gottschalk. "A realistic model for e+e- annihilation including parton bremsstrahlung effects". In: *Nuclear Physics B* 214.2 (Apr. 1983), pp. 201–222. ISSN: 0550-3213. DOI: 10.1016/0550-3213(83)90658-2. URL: http://www.sciencedirect.com/science/article/pii/0550321383906582.

[39] Thomas D. Gottschalk. "An improved description of hadronization in the QCD cluster model for e+e- annihilation". In: *Nuclear Physics B* 239.2 (July 1984), pp. 349–381. ISSN: 0550-3213. DOI: 10.1016/0550-3213(84)90253-0. URL: http://www.sciencedirect.com/science/article/pii/0550321384902530.

[40] D. Amati and G. Veneziano. "Preconfinement as a property of perturbative QCD". In: *Physics Letters B* 83.1 (Apr. 1979), pp. 87–92. ISSN: 0370-2693. DOI: 10.1016/0370-2693(79)90896-7. URL: http://www.sciencedirect.com/science/article/pii/0370269379908967.

[41] David J. Lange. "The EvtGen particle decay simulation package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines 462.1 (Apr. 2001), pp. 152–155. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(01)00089-4. URL: http://www.sciencedirect.com/science/article/pii/S0168900201000894.

[42] Stanislaw Jadach, Johann H. Kuhn, and Zbigniew Was. "TAUOLA - a library of Monte Carlo programs to simulate decays of polarized leptons." In: *Computer Physics Communications* 64.2 (May 1991), pp. 275–299. ISSN: 0010-4655. DOI: 10.1016/0010-4655(91)90038-M. URL: http://www.sciencedirect.com/science/article/pii/001046559190038M.

[43] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (July 2014). arXiv: 1405.0301. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2014)079. URL: http://arxiv.org/abs/1405.0301.

[44] Johan Alwall et al. "MadGraph 5 : Going Beyond". In: *Journal of High Energy Physics* 2011.6 (June 2011). arXiv: 1106.0522. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2011)128. URL: http://arxiv.org/abs/1106.0522.

[45] S. Frixione, P. Nason, and B. R. Webber. "Matching NLO QCD and parton showers in heavy flavour production". In: *Journal of High Energy Physics* 2003.08 (Aug. 2003). arXiv: hep-ph/0305252, pp. 007–007. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2003/08/007. URL: http://arxiv.org/abs/hep-ph/0305252.

[46] Carlo Oleari. "The POWHEG-BOX". In: *Nuclear Physics B - Proceedings Supplements* 205-206 (Aug. 2010). arXiv: 1007.3893, pp. 36–41. ISSN: 09205632. DOI: 10.1016/j.nuclphysbps.2010.08.016. URL: http://arxiv.org/abs/1007.3893.

[47] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 Physics and Manual". In: *Journal of High Energy Physics* 2006.05 (May 2006). arXiv: hep-ph/0603175, pp. 026–026. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2006/05/026. URL: http://arxiv.org/abs/hep-ph/0603175.

[48] Torbjorn Sjostrand. "PYTHIA 8 Status Report". In: *arXiv:0809.0303 [hep-ph]* (Sept. 2008). arXiv: 0809.0303. URL: http://arxiv.org/abs/0809.0303.

[49] Johannes Bellm et al. "Herwig 7.0 / Herwig++ 3.0 Release Note". In: *The European Physical Journal C* 76.4 (Apr. 2016). arXiv: 1512.01178. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-016-4018-8. URL: http://arxiv.org/abs/1512.01178.

[50] Frank Krauss, Ralf Kuhn, and Gerhard Soff. "AMEGIC++ 1.0, A Matrix Element Generator In C++". en. In: *Journal of High Energy Physics* 2002.02 (2002), p. 044. ISSN: 1126-6708. DOI: 10.1088/1126-6708/2002/02/044. URL: http://stacks.iop.org/1126-6708/2002/i=02/a=044.

[51] Tanju Gleisberg and Stefan Hoeche. "Comix, a new matrix element generator". In: *Journal of High Energy Physics* 2008.12 (Dec. 2008). arXiv: 0808.3674, pp. 039–039. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/12/039. URL: http://arxiv.org/abs/0808.3674.

[52] Steffen Schumann and Frank Krauss. "A parton shower algorithm based on Catani-Seymour dipole factorisation". In: *Journal of High Energy Physics* 2008.03 (Mar. 2008). arXiv: 0709.1027, pp. 038–038. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/03/038. URL: http://arxiv.org/abs/0709.1027.

[53] Jan-Christopher Winter, Frank Krauss, and Gerhard Soff. "A modified cluster-hadronization model". In: *The European Physical Journal C* 36.3 (Aug. 2004). arXiv: hep-ph/0311085, pp. 381–395. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s2004-01960-8. URL: http://arxiv.org/abs/hep-ph/0311085.

[54] S. Agostinelli et al. "Geant4 simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (July 2003), pp. 250–303. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8. URL: http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[55] P. Calafiura et al. "The Athena control framework in production, new developments and lessons learned". In: (2005), pp. 456–458.

[56] G. Barrand et al. "GAUDI, A software architecture and framework for building HEP data processing applications". In: *Computer Physics Communications*. CHEP2000 140.1 (Oct. 2001), pp. 45–55. ISSN: 0010-4655. DOI: 10.1016/S0010-4655(01)00254-5. URL: http://www.sciencedirect.com/science/article/pii/S0010465501002545.

[57] Rene Brun and Fons Rademakers. "ROOT, An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. New Computing Techniques in Physics Research V 389.1 (Apr. 1997), pp. 81–86. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(97)00048-X. URL: http://www.sciencedirect.com/science/article/pii/S016890029700048X.

[58] Ilka Antcheva et al. "ROOT - A C++ Framework for Petabyte Data Storage, Statistical Analysis and Visualization". In: *Computer Physics Communications* 180.12 (Dec. 2009). arXiv: 1508.07749, pp. 2499–2512. ISSN: 00104655. DOI: 10.1016/j.cpc.2009.08.005. URL: http://arxiv.org/abs/1508.07749.

[59] A. Hoecker et al. "TMVA - Toolkit for Multivariate Data Analysis". In: *arXiv:physics/0703039* (Mar. 2007). arXiv: physics/0703039. URL: http://arxiv.org/abs/physics/0703039.

[60] Wouter Verkerke and David Kirkby. "The RooFit toolkit for data modeling". In: *arXiv:physics/0306116* (June 2003). arXiv: physics/0306116. URL: http://arxiv.org/abs/physics/0306116.

[61] ATLAS Collaboration. "Performance of b-Jet Identification in the ATLAS Experiment". In: *Journal of Instrumentation* 11.04 (Apr. 2016). arXiv: 1512.01094, P04008–P04008. ISSN: 1748-0221. DOI: 10.1088/1748-0221/11/04/P04008. URL: http://arxiv.org/abs/1512.01094.

[62] ATLAS Collaboration. "Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s}=13$ TeV". In: *Journal of High Energy Physics* 2018.8 (Aug. 2018). arXiv: 1805.01845. ISSN: 1029-8479. DOI: 10.1007/JHEP08(2018)089. URL: http://arxiv.org/abs/1805.01845.

[63] *Measurement of b-tagging Efficiency of c-jets in $t\bar{t}$ Events Using a Likelihood Approach with the ATLAS Detector*. fr. Mar. 2018. URL: https://cds.cern.ch/record/2306649.

[64] Johannes Erdmann et al. "A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 748 (June 2014), pp. 18–25. ISSN: 0168-9002. DOI: 10.1016/j.nima.2014.02.029. URL: http://www.sciencedirect.com/science/article/pii/S0168900214001855.

[65] E. Le Menedeu et al. *Continuous b-tagging for the ATLAS experiment*. fr. Jan. 2014. URL: https://cds.cern.ch/record/1644430.

[66] ATLAS Collaboration. "Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector". In: *Journal of High Energy Physics* 2015.1 (Jan. 2015). arXiv: 1409.6212. ISSN: 1029-8479. DOI: 10.1007/JHEP01(2015)069. URL: http://arxiv.org/abs/1409.6212.

[67] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012). arXiv: 1207.7214, pp. 1–29. ISSN: 03702693. DOI: 10.1016/j.physletb.2012.08.020. URL: http://arxiv.org/abs/1207.7214.

[68] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012). arXiv: 1207.7235, pp. 30–61. ISSN: 03702693. DOI: 10.1016/j.physletb.2012.08.021. URL: http://arxiv.org/abs/1207.7235.

[69] T. Aaltonen et al. "Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron". In: *Phys.Rev.Lett.* 109 (2012), p. 071804. DOI: 10.1103/PhysRevLett.109.071804.

[70] CMS Collaboration. "Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks". In: *Physical Review D* 89.1 (Jan. 2014). arXiv: 1310.3687. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.89.012003. URL: http://arxiv.org/abs/1310.3687.

[71] ATLAS Collaboration. "Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector". In: *Journal of High Energy Physics* 2017.12 (Dec. 2017). arXiv: 1708.03299. ISSN: 1029-8479. DOI: 10.1007/JHEP12(2017)024. URL: http://arxiv.org/abs/1708.03299.

[72] *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector*. fr. July 2015. URL: https://cds.cern.ch/record/2037702.

[73] Daniel Buscher. "Search for Higgs bosons with b-jets in the final state in proton-proton collisions with the ATLAS experiment". fr. PhD thesis. Nov. 2016. URL: https://cds.cern.ch/record/2232472.

[74] Abraham Wald. "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large". In: *Transactions of the American Mathematical Society* 54.3 (1943), pp. 426–482. ISSN: 0002-9947. DOI: 10.2307/1990256. URL: http://www.jstor.org/stable/1990256.

[75] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011). arXiv: 1007.1727, pp. 1–19. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: http://arxiv.org/abs/1007.1727.

[76] CMS Collaboration. "Evidence for the Higgs boson decay to a bottom quark-antiquark pair". In: *Physics Letters B* 780 (May 2018). arXiv: 1709.07497, pp. 501–532. ISSN: 03702693. DOI: 10.1016/j.physletb.2018.02.050. URL: http://arxiv.org/abs/1709.07497.

[77] ATLAS Collaboration. "Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector". In: *arXiv:1808.08238 [hep-ex]* (Aug. 2018). arXiv: 1808.08238. URL: http://arxiv.org/abs/1808.08238.

[78] CMS Collaboration. "Observation of Higgs boson decay to bottom quarks". In: *Physical Review Letters* 121.12 (Sept. 2018). arXiv: 1808.08242. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.121.121801. URL: http://arxiv.org/abs/1808.08242.

[79] "Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a $b\bar{b}$ pair in pp collisions at 13 TeV using the ATLAS detector". In: (2016).

[80] *ATLAS Run 1 Pythia8 tunes*. fr. Nov. 2014. URL: https://cds.cern.ch/record/1966419.

[81] ATLAS Collaboration. "Measurement of colour flow using jet-pull observables in $t\bar{t}$ events with the ATLAS experiment at $\sqrt{s} = 13$ TeV". In: *arXiv:1805.02935 [hep-ex]* (May 2018). arXiv: 1805.02935. URL: http://arxiv.org/abs/1805.02935.

[82] Jason Gallicchio and Matthew D. Schwartz. "Seeing in Color: Jet Superstructure". In: *Physical Review Letters* 105.2 (July 2010). arXiv: 1001.5027. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.105.022001. URL: http://arxiv.org/abs/1001.5027.

[83] ATLAS Collaboration. "Search for top quark decays $t \rightarrow qH$, with $H\to\gamma\gamma$ in $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector". In: *Journal of High Energy Physics* 2017.10 (Oct. 2017). arXiv: 1707.01404. ISSN: 1029-8479. DOI: 10.1007/JHEP10(2017)129. URL: http://arxiv.org/abs/1707.01404.

[84] F. del Aguila, J. A. Aguilar-Saavedra, and R. Miquel. "Constraints on top couplings in models with exotic quarks". In: *Physical Review Letters* 82.8 (Feb. 1999). arXiv: hep-ph/9808400, pp. 1628–1631. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.82.1628. URL: http://arxiv.org/abs/hep-ph/9808400.

[85] J. A. Aguilar-Saavedra and B. M. Nobre. "Rare top decays t -> c gamma, t -> c g and CKM unitarity". In: *Physics Letters B* 553.3-4 (Feb. 2003). arXiv: hep-ph/0210360, pp. 251–260. ISSN: 03702693. DOI: 10.1016/S0370-2693(02)03230-6. URL: http://arxiv.org/abs/hep-ph/0210360.

[86] J. A. Aguilar-Saavedra. "Effects of mixing with quark singlets". In: *Physical Review D* 69.9 (May 2004). arXiv: hep-ph/0210112. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.69.099901. URL: http://arxiv.org/abs/hep-ph/0210112.

[87] Santi Bejar, Jaume Guasch, and Joan Sola. "Loop Induced Flavor Changing Neutral Decays of the Top Quark in a General Two-Higgs-Doublet Model". In: *Nuclear Physics B* 600.1 (Apr. 2001). arXiv: hep-ph/0011091, pp. 21–38. ISSN: 05503213. DOI: 10.1016/S0550-3213(01)00044-X. URL: http://arxiv.org/abs/hep-ph/0011091.

[88] Itzhak Baum, Gad Eilam, and Shaouly Bar-Shalom. "Scalar FCNC and rare top decays in a two Higgs doublet model "for the top"". In: *Physical Review D* 77.11 (June 2008). arXiv: 0802.2622. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.77.113008. URL: http://arxiv.org/abs/0802.2622.

[89] Kai-Feng Chen et al. "When the Higgs meets the Top: Search for t –> ch^0 at the LHC". In: *Physics Letters B* 725.4-5 (Oct. 2013). arXiv: 1304.8037, pp. 378–381. ISSN: 03702693. DOI: 10.1016/j.physletb.2013.07.060. URL: http://arxiv.org/abs/1304.8037.

[90] David Atwood, Sudhir Kumar Gupta, and Amarjit Soni. "Constraining the flavor changing Higgs couplings to the top-quark at the LHC". In: *Journal of High Energy Physics* 2014.10 (Oct. 2014). arXiv: 1305.2427. ISSN: 1029-8479. DOI: 10.1007/JHEP10(2014)057. URL: http://arxiv.org/abs/1305.2427.

[91] Chong Sheng Li, R. J. Oakes, and Jin Min Yang. "Rare decays of the top quark in the minimal supersymmetric model". In: *Physical Review D* 49.1 (Jan. 1994), pp. 293–298. DOI: 10.1103/PhysRevD.49.293. URL: https://link.aps.org/doi/10.1103/PhysRevD.49.293.

[92] G. M. de Divitiis, R. Petronzio, and L. Silvestrini. "Flavour changing top decays in supersymmetric extensions of the standard model". In: *Nuclear Physics B* 504.1-2 (Oct. 1997). arXiv: hep-ph/9704244, pp. 45–60. ISSN: 05503213. DOI: 10.1016/S0550-3213(97)00476-8. URL: http://arxiv.org/abs/hep-ph/9704244.

[93] Jorge L. Lopez, D. V. Nanopoulos, and Raghavan Rangarajan. "New Supersymmetric Contributions to $t->cV$". In: *Physical Review D* 56.5 (Sept. 1997). arXiv: hep-ph/9702350, pp. 3100–3106. ISSN: 0556-2821, 1089-4918. DOI: 10.1103/PhysRevD.56.3100. URL: http://arxiv.org/abs/hep-ph/9702350.

[94] D. Delepine and S. Khalil. "Top flavour violating decays in general supersymmetric models". In: *Physics Letters B* 599.1-2 (Oct. 2004). arXiv: hep-ph/0406264, pp. 62–74. ISSN: 03702693. DOI: 10.1016/j.physletb.2004.08.025. URL: http://arxiv.org/abs/hep-ph/0406264.

[95] Jian Jun Liu et al. "$t\to cV$ via SUSY FCNC couplings in the unconstrained MSSM". In: *Physics Letters B* 599.1-2 (Oct. 2004). arXiv: hep-ph/0406155, pp. 92–101. ISSN: 03702693. DOI: 10.1016/j.physletb.2004.07.062. URL: http://arxiv.org/abs/hep-ph/0406155.

[96] J. J. Cao et al. "SUSY-induced FCNC top-quark processes at the Large Hadron Collider". In: *Physical Review D* 75.7 (Apr. 2007). arXiv: hep-ph/0702264. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.75.075021. URL: http://arxiv.org/abs/hep-ph/0702264.

[97] Jin Min Yang, Bing-Lin Young, and X. Zhang. "Flavor-changing Top Quark Decays In R-Parity Violating SUSY". In: *Physical Review D* 58.5 (July 1998). arXiv: hep-ph/9705341. ISSN: 0556-2821, 1089-4918. DOI: 10.1103/PhysRevD.58.055001. URL: http://arxiv.org/abs/hep-ph/9705341.

[98] Gongru Lu et al. "The rare top quark decays $t\to cV$ in the topcolor-assisted technicolor model". In: *Physical Review D* 68.1 (July 2003). arXiv: hep-ph/0303122. ISSN: 0556-2821, 1089-4918. DOI: 10.1103/PhysRevD.68.015002. URL: http://arxiv.org/abs/hep-ph/0303122.

[99] Kaustubh Agashe, Gilad Perez, and Amarjit Soni. "Flavor Structure of Warped Extra Dimension Models". In: *Physical Review D* 71.1 (Jan. 2005). arXiv: hep-ph/0408134. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.71.016002. URL: http://arxiv.org/abs/hep-ph/0408134.

[100] Kaustubh Agashe, Gilad Perez, and Amarjit Soni. "Collider Signals of Top Quark Flavor Violation from a Warped Extra Dimension". In: *Physical Review D* 75.1 (Jan. 2007). arXiv: hep-ph/0606293. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.75.015002. URL: http://arxiv.org/abs/hep-ph/0606293.

[101] Bingfang Yang, Ning Liu, and Jinzhong Han. "Top Quark FCNC Decay to 125GeV Higgs boson in the Littlest Higgs Model with T-parity". In: *Physical Review D* 89.3 (Feb. 2014). arXiv: 1308.4852. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.89.034020. URL: http://arxiv.org/abs/1308.4852.

[102] Kaustubh Agashe and Roberto Contino. "Composite Higgs-Mediated FCNC". In: *Physical Review D* 80.7 (Oct. 2009). arXiv: 0906.1542. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.80.075016. URL: http://arxiv.org/abs/0906.1542.

[103] J. A. Aguilar-Saavedra. "Top flavour-changing neutral interactions: theoretical expectations and experimental detection". In: *arXiv:hep-ph/0409342* (Sept. 2004). arXiv: hep-ph/0409342. URL: http://arxiv.org/abs/hep-ph/0409342.

[104] Daniel Fournier and Jean-Baptiste De Vivie De Regie. *Search for Flavour Changing Neutral Currents in $t\rightarrow qH$, with $H\to\gamma\gamma$, and limit on the tqH couplings*. fr. Nov. 2016. URL: https://cds.cern.ch/record/2233615.

[105] ATLAS Collaboration. "Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data". In: *The European Physical Journal C* 76.12 (Dec. 2016). arXiv: 1606.01813. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-016-4507-9. URL: http://arxiv.org/abs/1606.01813.

[106] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[107] Kyle S. Cranmer. "Kernel Estimation in High-Energy Physics". In: *Computer Physics Communications* 136.3 (May 2001). arXiv: hep-ex/0011057, pp. 198–207. ISSN: 00104655. DOI: 10.1016/S0010-4655(00)00243-5. URL: http://arxiv.org/abs/hep-ex/0011057.

université PARIS-SACLAY | ÉCOLE DOCTORALE PHENIICS

Titre : Étude du boson de Higgs dans son canal de désintégration en quarks beauté avec l'expérience ATLAS au run 2 du LHC

Mots clés : ATLAS, LHC, boson, Higgs, quark, beauté

Résumé : Dans ce manuscrit sont décrits trois aspects de mon travail de thèse auprès de l'expérience ATLAS. Le premier est consacré à l'extraction des corrections sur l'étiquetage de jets issus de quarks beauté, appliquées aux simulations afin de mieux décrire les données expérimentales. Le second est l'étude de la désintégration du boson de Higgs en paire de quarks beauté. Le dernier aspect est la recherche de nouvelle physique à travers l'interaction d'un quark top et d'un boson de Higgs.

Mon implication dans la collaboration ATLAS a commencé avec la mise à jour de la calibration pseudo-continue des techniques d'identification de jets issus de quarks beauté (b-jets). La calibration pseudo-continue de l'étiquetage des b-jets (b-Tagging) dans l'expérience ATLAS a été développée durant le run 1 du LHC. Une mise à jour du code a été effectuée. Les incertitudes sur la calibration pseudo-continue du b-Tagging sont dominées par les sources d'erreur systématique.

Le second sujet de la thèse est la recherche du boson de Higgs dans son canal de désintégration en quarks beauté, produit en association avec un boson vecteur (VH(bb)). J'ai notamment participé à l'étude du canal pp $\rightarrow$ ZH, dans lequel un boson Z se désintègre en neutrinos. L'analyse a été optimisée respectivement à l'expérience acquise au cours du run 1 du LHC. Plusieurs idées ont été explorées afin d'améliorer l'analyse multivariée. L'évidence du processus VH(bb) a été obtenue avec les données collectées en 2015 et 2016. La force du signal (rapport des sections efficaces mesurée et prédite) a été mesurée à $\hat{\mu}_{VH} = 1.20^{+0.24}_{-0.23}(stat.)^{+0.34}_{-0.28}(syst.)$, ce qui correspond à une significance observée de 3.5 significations statistiques. L'incertitude sur ce résultat est dominée par des sources d'erreur systématique. En conséquence, plusieurs études ont été menées afin de mieux comprendre certains aspects de ce résultat. Une attention particulière a été apportée à la simulation des processus de bruit de fond dans le but d'améliorer l'efficacité de génération des simulations Monte-Carlo. Aussi, l'utilisation de nouvelles variables liées au flux de couleur entre les deux quarks beauté (colourflow) dans la recherche du processus VH(bb) semble apporter un gain limité dans la rejection des processus de bruit de fond avec les performances actuelles de reconstruction des jets. L'analyse a été menée en ajoutant les données collectées en 2017, et a conduit à l'observation non ambigüe du signal de désintégration du boson de Higgs en quarks beauté.

Enfin, la dernière étude présentée est une révision de la recherche du signal d'un quark top dans sa désintégration en quark up ou charm et un boson de Higgs se désintégrant en paire de photons. Ce processus est interdit avec un vertex unique dans le Modèle Standard, et largement supprimé dans les canaux impliquant des boucles, avec un taux d'embranchement de 10^{-15}, ce qui en fait un canal pertinent dans la recherche de nouvelle physique. L'analyse est conçue afin de différencier les couplages et en ciblant à la fois les modes de production d'un unique quark top, ou d'une paire de quarks top, et en exploitant les techniques d'identification de c-jets dans la catégorisation. L'utilisation de variables cinématiques a été développée, et des méthodes multivariées mises en oeuvre dans l'analyse. Les premières perspectives de résultats attendus sont présentées. Les données collectées au cours du run 2 du LHC devraient permettre à l'analyse d'être sensible à un taux d'embranchement inférieur à 1 et pourrait apporter les premières contraintes sur certains modèles de physique au délà du modèle standard.

Title : Study of the Higgs boson decay into b-quarks with the ATLAS experiment run 2

Keywords : ATLAS, LHC, boson, Higgs, quark, beauty

Abstract : In this thesis, three aspects of my work within the ATLAS experiment are described. The first one is dedicated to the extraction of correction factors to be applied on the simulation to match the data. The second is the analysis of the Higgs boson decay to a pair of b-quarks. The last one is a search for BSM in the interaction between a top quark and the Higgs boson.

My implication in the ATLAS collaboration started with updates of the pseudo-continuous calibration of b-jets flavour identification. The pseudo-continuous calibration of b-Tagging in the ATLAS experiment was developped during the run 1 of the LHC. An update of the framework was performed, and calibration results have been produced. The uncertainties on the pseudo-continuous b-Tagging calibration are dominated by systematic sources.

The second topic of the thesis is the search for a Higgs boson decaying into beauty quarks, produced in association with a vector boson. I was especially involved in the channel looking for a Z boson decaying into neutrinos. The analysis was optimised with respect to the experience acquired during the run 1 of the LHC. Many ideas have been explored in order to improve the multivariate analysis. The evidence for the VH(bb) process was presented using data collected in 2015 and 2016. The signal strength is measured to $\hat{\mu}_{VH} = 1.20^{+0.24}_{-0.23}(stat.)^{+0.34}_{-0.28}(syst)$, which corresponds to an observed significance of 3.5 standard deviations. The uncertainty on this result is dominated by systematic sources. Hence, several studies have been performed in order to better understand specific aspects of this result. A special attention was given to the modelling of background processes to improve the generation efficiency of Monte-Carlo simulation. Also, the use of coulourflow in the VH(bb) search appears to provide a limited gain in the rejection of background processes with the current jet reconstruction performances. The analysis was updated adding data collected in 2017, and lead to the observation of the Higgs boson decay into beauty quarks.

At last, the final study presented in this work is a major revision of the search for a top quark decaying into an up-type (= ,) quark and a Higgs boson decaying into two photons. This process is forbiden at tree level in the SM and highly supressed at loop level, with an expected branching ratio of 10^{-15}, making it a very good probe for BSM physics. The analysis is now designed in order to disentangle the and couplings by targetting both the single- and pair-production modes of top quarks, and exploiting c-tagging in the categorisation. The use of kinematic variables was developped, and multivariate methods have been implemented in the analysis. Preliminary prospect results are presented. With the set of data accumulated by the end of the LHC run 2, the analysis will be sensitive to branching ratio below the permil level and could provide the first constraints on realistic BSM models.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France