NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

PHYSICS DEPARTMENT OF SCHOOL OF APPLIED MATHEMATICAL AND PHYSICAL SCIENCES

Feasibility study for inclusive measurements of Higgs branching ratios to hadronic final states with the ATLAS detector

DIPLOMA THESIS

by

Athanasia - Konstantina Angelopoulou

Supervisors: Evangelos Gazis
Professor NTUA
Marco Battaglia
Professor UCSC

Athens, May 2020

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΦΥΣΙΚΗΣ ΣΧΟΛΗΣ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

Feasibility study for inclusive measurements of Higgs branching ratios to hadronic final states with the ATLAS detector

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Αθανασίας - Κωνσταντίνας Αγγελοπούλου

Επιβλέποντες: Ευάγγελος Γαζής
Καθηγητής Ε.Μ.Π.
Marco Battaglia
Καθηγητής UCSC

Αθήνα, Μάιος 2020

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΦΥΣΙΚΗΣ ΣΧΟΛΗΣ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

Feasibility study for inclusive measurements of Higgs branching ratios to hadronic final states with the ATLAS detector

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Αθανασίας - Κωνσταντίνας Αγγελοπούλου

Επιβλέποντες: Ευάγγελος Γαζής
Καθηγητής Ε.Μ.Π.
Marco Battaglia
Καθηγητής UCSC

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 29η Μαΐου 2020.

.....	.....	.....
Ευάγγελος Γαζής Καθηγητής Ε.Μ.Π.	Ηλίας Γλύτσης Καθηγητής Ε.Μ.Π.	Ιωάννης Ξανθάκης Καθηγητής Ε.Μ.Π.

.....

Αθανασία - Κωνσταντίνα Αγγελοπούλου
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Abstract

The investigation of the decay properties of the Higgs boson is an important part of the LHC physics programme. An accurate determination of the decay branching ratios to the hadronic states $b\bar{b}, c\bar{c}, gg$ is necessary for studying the Higgs couplings and thus clarifying the understanding of the particle mass generation mechanism and validating the predictions of the Standard Model. At the same time, possible deviations from these values could indicate the existence of new physics. Identifying hadronic states is based on flavour tagging jets and as current methods are able to mostly identify well b-jets, the branching ratio of the Higgs boson to charm quarks has not been measured.

This work studies the feasibility of doing an inclusive measurement of the Higgs hadronic branching ratios. The attempt at determining all three of them simultaneously is based on di-jet discriminants created using the response of a flavour tagging algorithm. For this to be possible, it's necessary to first extract the contribution from background processes attributed to the strong interaction. Thus, the flavour composition of the background is studied and a methodology for estimating it in the signal region is developed. This thesis shows that with the present flavour tagging accuracy, the branching ratios cannot be determined. Therefore, modifications to the output of the flavour tagging algorithm are examined. The study demonstrates the required level of efficiency and mistagging rate of flavour tagging for enabling the analysis with the goal to motivate further enhancements in the discrimination between b-, c- and light-flavour jets.

Keywords: Higgs boson, Flavour tagging, Hadronic decays, Branching ratios, Inclusive measurement, QCD background, MV2c10 algorithm

Περίληψη

Η διερεύνηση των ιδιοτήτων διάσπασης του μποζονίου Higgs αποτελεί σημαντικό κομμάτι του προγράμματος φυσικής του LHC. Ένας ακριβής προσδιορισμός των λόγων διακλάδωσης στις αδρονικές καταστάσεις $b\bar{b}, c\bar{c}, gg$ είναι απαραίτητος για την μελέτη των σταθερών σύζευξης του πεδίου Higgs και την διευκρίνιση του μηχανισμού απόκτησης μάζας των σωματιδίων αλλά και για την επιβεβαίωση των προβλέψεων του Καθιερωμένου Προτύπου. Ταυτόχρονα, πιθανές αποκλίσεις από αυτές τις τιμές θα μπορούσαν να υποδεικνύουν την ύπαρξη νέας φυσικής. Η αναγνώριση των αδρονικών καταστάσεων βασίζεται στο flavour tagging και καθώς οι τρέχουσες μέθοδοι μπορούν να αναγνωρίσουν καλά κυρίως τα b-jets, ο λόγος διακλάδωσης σε c κουάρκ δεν έχει μετρηθεί.

Η εργασία αυτή μελετά τη δυνατότητα επίτευξης μίας συνολικής μέτρησης των αδρονικών λόγων διακλάδωσης. Η απόπειρα για τον ταυτόχρονο προσδιορισμό τους βασίζεται σε di-jet discriminants σχηματισμένων από την απόκριση ενός αλγορίθμου flavour tagging. Για να είναι εφικτό αυτό θα πρέπει πρώτα να εξαχθεί η συνεισφορά διαδικασιών background που οφείλονται κυρίως στην ισχυρή αλληλεπίδραση. Συνεπώς μελετάται η σύνθεση του background και αναπτύσσεται μία μέθοδος για την εκτίμηση της συνεισφοράς του στην περιοχή του σήματος. Η διπλωματική εργασία δείχνει ότι με την τρέχουσα ακρίβεια του flavour tagging, οι λόγοι διακλάδωσης δε μπορούν να προσδιοριστούν. Στη συνέχεια, εξετάζονται τροποποιήσεις στην απόκριση του αλγορίθμου flavour tagging ώστε να προσδιοριστεί το απαραίτητο επίπεδο απόδοσης για να είναι εφικτή μία τέτοια ανάλυση. Σκοπός είναι να κινητοποιηθούν περαιτέρω βελτιώσεις στην διάκριση μεταξύ των jets διαφορετικών γεύσεων.

Λέξεις κλειδιά: Μποζόνιο Higgs, Flavour tagging, Αδρονικές διασπάσεις, Λόγοι διακλάδωσης, QCD background, Αλγόριθμος MV2c10

Acknowledgements

I would like to thank my professor and supervisor Evangelos Gazis for his advice, support and encouragement during the last couple of years. I am grateful for his belief in me and the opportunity he gave me to complete my diploma thesis at CERN.

I would also like to express my gratitude towards Prof. Marco Battaglia for guiding me through the completion of the thesis and sharing his knowledge and experience with me. His help and willingness to answer my questions were invaluable and through our collaboration I learnt important lessons regarding not only physics but generally the scientific approach.

Είμαι ευγνώμων που μοιράστηκα όλες τις εμπειρίες των τελευταίων χρόνων με την αγαπημένη μου φίλη Άννα, η οποία είναι πάντα δίπλα μου ακόμα και όταν μας χωρίζει φυσική απόσταση. Η προθυμία της να με ακούσει, η κατανόηση που πάντα επιδεικνύει και οι συμβουλές της με έχουν βοηθήσει ουσιωδώς στο να αντιμετωπίσω πολλές δύσκολες καταστάσεις.

Θα ήθελα επίσης να ευχαριστήσω τον Παναγιώτη που όλα αυτά τα χρόνια με έχει στηρίξει σε κάθε μου προσπάθεια και έχει αποτελέσει καθοριστικό παράγοντα στην διαμόρφωση της επιστημονικής μου σκέψης αλλά και εν γένει της προσωπικότητας μου. Η απεριόριστη υπομονή του και η μοναδική του ικανότητα να προσφέρει πρακτικές λύσεις σε κάθε είδους πρόβλημα δεν έχει σταματήσει να με εκπλήσσει.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την συνεχή αγάπη και φροντίδα καθώς και για την ηθική και υλική υποστήριξη που μου παρέχει σε κάθε στάδιο της ζωής μου.

Contents

List of Figures

List of Tables

Εκτεταμένη ελληνική περίληψη

Η παρούσα διπλωματική εργασία εξετάσει το κατά πόσο είναι εφικτό να γίνει μια συνολική μέτρηση όλων των λόγων διακλάδωσης των διασπάσεων του μποζονίου Higgs σε αδρονικές καταστάσεις. Η προτεινόμενη μέθοδος χρησιμοποιεί το αποτέλεσμα ενός αλγορίθμου flavour tagging χωρίς να θέτει κάποιο κατώφλι στην τιμή του εκ των προτέρων. Συνεπώς λαμβάνει υπόψιν όλες τις πιθανές γεύσεις των jets. Στη συνέχεια χρησιμοποιεί di-jet discriminants για να μετρηθεί η συμβολή κάθε τρόπου διάσπασης δίνοντας έτσι μία εκτίμηση για τους λόγους διακλάδωσης. Η δυσκολία της μεθόδου έγκειται στο μεγάλο background που οφείλεται στην παρουσία ζευγών κουάρκ-αντικουάρκ στον ανιχνευτή και παράγονται από διαδικασίες που εμπλέκουν την ισχυρή δύναμη (QCD background) και είναι δύσκολο να διαχωριστούν από τις διασπάσεις του Higgs. Γι΄ αυτό το λόγο πρέπει να μοντελοποιηθούν ξεχωριστά.

Θεωρητικό υπόβαθρο

Η σταθερά σύζευξης του πεδίου Higgs με ένα φερμιόνιο δίνεται από την Εξίσωση 2.72 η οποία δείχνει ότι η ισχύς της σύζευξης είναι ανάλογη της μάζας. Αυτή είναι μία σημαντική πρόβλεψη του Καθιερωμένου Προτύπου (Κ.Π) που είναι χρήσιμο να επαληθευτεί πειραματικά. Τυχούσες αποκλίσεις από τις προβλεπόμενες τιμές των σταθερών σύζευξης θα μπορούσαν να θεωρηθούν ενδείξεις για την στήριξη θεωριών εκτός του Κ.Π όπως υπερσυμμετρικών μοντέλων. Μία πειραματική μέτρηση μπορεί να γίνει εμμέσως μετρώντας τους λόγους διακλάδωσης οι οποίοι ορίζονται σύμφωνα με την Εξίσωση 3.2 όπου Γ_i είναι ο ρυθμός διάσπασης και Γ_{total} το συνολικό πλάτος διάσπασης. Ο ρυθμός διάσπασης εκφράζει την πιθανότητα ανά μονάδα χρόνου ότι ένα σωματίδιο θα διασπαστεί με ένα συγκεκριμένο τρόπο. Το συνολικό πλάτος διάσπασης ορίζεται ως το άθροισμα των ρυθμών διάσπασης για όλους τους πιθανούς τρόπους διάσπασης και είναι αντιστρόφως ανάλογος με την διάρκεια ζωής ενός σωματιδίου. Συγκεκριμένα για το μποζόνιο Higgs, η διάσπαση σε ένα ζεύγος φερμιονίου-αντιφερμιονίου έχει μερικό πλάτος που δίνεται από την Εξίσωση 3.3, σχετίζεται δηλαδή άμεσα με την μάζα του φερμιονίου. Αυτό ουσιαστικά αντανακλά πόσο ισχυρά αλληλεπιδρά το φερμιόνιο με το πεδίο Higgs. Οι λόγοι διακλάδωσης δίνονται στον Πίνακα 3.1. Οι διασπάσεις σε αδρονικές καταστάσεις αναφέρονται στις περιπτώσεις όπου το μποζόνιο Higgs διασπάται σε b-κουάρκ, c-κουάρκ και γκλουόνια. Το μποζόνιο Z έχει επίσης αδρονικές διασπάσεις άρα η μέθοδος υπολογισμού των λόγων διακλάδωσης μπορεί να εφαρμοστεί και στην περίπτωση του. Σε διάφορα υπερσυμμετρικά μοντέλα οι λόγοι διακλάδωσης λαμβάνουν διαφορετικές τιμές από αυτές του Κ.Π όπως φαίνεται στο Σχήμα 3.8α. Τα τελευταία πειραματικά αποτελέσματα για τους λόγους διακλάδωσης έχουν ακρίβεια που ακόμα επιτρέπει σενάρια εκτός του Κ.Π (Σχήμα 3.8β).

Πειραματικό υπόβαθρο

Το κύριο φυσικό αντικείμενο της εργασίας αυτής είναι τα jets. Ο όρος αυτός αναφέρεται σε ευθυγραμμισμένα σωματίδια που παράγονται κατά τον σχηματισμό αδρονίων από κουάρκ και γκλουόνια σε πειράματα επιταχυντών. Η ύπαρξη των jets οφείλεται στο χαρακτηριστικό της ισχυρής αλληλεπίδρασης να μην επιτρέπει την ύπαρξη καταστάσεων με μη μηδενικό φορτίο χρώματος. Τα αδρόνια που παράγονται έχουν παρόμοια ενέργεια και ορμή με αυτή των αρχικών κουάρκ, άρα μπορούν να λειτουργήσουν σαν 'αντιπρόσωποι' τους. Καθώς οι αδρονικές διασπάσεις του μποζονίου Higgs οδηγούν στην δημιουργία jets, είναι σημαντικό να μπορούν να ανιχνευτούν. Ένα ζεύγος κουάρκ-αντικουάρκ ανακατασκευάζεται ως ένα μεγάλης ακτίνας jet το οποίο αποτελείται από δύο jets μεταβλητής ακτίνας. Τα jets μεγάλης ακτίνας ανακατασκευάζονται στο καλορίμετρο του ανιχνευτή ενώ τα μεταβλητής ακτίνας ανακατασκευάζονται σύμφωνα με τις τροχιές που καταγράφονται στον εσωτερικό ανιχνευτή.

Η αναγνώριση της γεύσης ενός κουάρκ γίνεται εφαρμόζοντας την τεχνική του flavour tagging στα μεταβλητής ακτίνας jets. Αφορά κυρίως την διάκριση jets που περιέχουν ένα b-κουάρκ καθώς αυτό έχει ορισμένα χαρακτηριστικά που το διαφοροποιούν από τα υπόλοιπα. Η διάρκεια ζωής του είναι τέτοια που ταξιδεύει κάποια απόσταση προτού διασπαστεί χωρίς να είναι αρκετά μεγάλη ώστε να διαφύγει από τον ανιχνευτή. Αυτό οδηγεί σε μία συγκεκριμένη τοπολογία λόγω της ύπαρξης μιας δευτερεύουσας κορυφής που αντιπροσωπεύει το σημείο που διασπάστηκε το b-κουάρκ. Επίσης, λόγω της μεγάλης του μάζας, τα jets που έχουν προέλθει από ένα b-κουάρκ αποτελούνται από περισσότερα σωματίδια και έχουν μεγαλύτερο άνοιγμα. Αυτά τα χαρακτηριστικά εκμεταλλεύονται από τον αλγόριθμο MV2c10 ο οποίος παράγει μία διαφορετική κατανομή ανάλογα με την γεύση του jet (Σχήμα 4.6α). Η απόδοση του b-tagging ορίζεται σύμφωνα με την Εξίσωση 4.1 ενώ ο συντελεστής απόρριψης ισούται με το αντίστροφο της απόδοσης. Τα μεγέθη αυτά ανταναμκλούν την ικανότητα του αλγορίθμου να αναγνωρίσει τα σωστά jets ως έχοντα μια συγκεκριμένη γεύση και να απορρίψει τα υπόλοιπα.

Τεχνική φιτ

Η μέθοδος που αναπτύχθηκε για την διπλωματική βασίζεται στην δημιουργία templates από την έξοδο του αλγορίθμου MV2c10 για κάθε πιθανό τρόπο διάσπασης (gg, $c\bar{c}$, $b\bar{b}$). Χρησιμοποιούνται για το φιτ μίας συνολική κατανομής που περιλαμβάνει όλα τα πιθανά γεγονότα ανεξαρτήτως γεύσης από το οποίο ουσιαστικά προκύπτει ένα κλάσμα που αντιστοιχεί στον λόγο διακλάδωσης. Το φιτ βασίζεται στην ελαχιστοποίηση μίας συνάρτησης ελαχίστων τετραγώνων που ορίζεται από την Εξίσωση 5.1 και περιέχει ως ελεύθερες μεταβλητές τα κλάσματα των δύο από τα τρία templates και έναν παράγοντα κανονικοποίησης. Το τρίτο κλάσμα μπορεί να βρεθεί μέσω της Εξίσωσης 5.2 καθώς οι αδρονικοί λόγοι διακλάδωσης θα πρέπει να έχουν άθροισμα ένα.

Καθώς το φιτ ελαχίστων τετραγώνων χρησιμοποιείται σε όλη την εργασία, πρέπει να επιβεβαιωθεί η ορθότητα του μέσω διαφόρων στατιστικών τρόπων. Για να γίνει αυτό, η συνολική κατανομή παράγεται ως γραμμικός συνδυασμός των τριών templates ώστε να είναι γνωστή εξ αρχής η σύνθεση της. Αρχικά, εξετάζεται η γραμμικότητα του φιτ καθώς θα πρέπει να μπορεί να εξάγει ακριβώς τα κλάσματα που εισάγονται. Η γραμμικότητα πράγματι επιβεβαιώνεται στις γραφικές παραστάσεις του Σχήματος 5.3. Στη συνέχεια, θα πρέπει να ελεγχθεί ότι η κατανομή του pull που ορίζεται από την Εξίσωση 5.7 είναι κανονική. Πράγματι και για τα τρία κλάσματα είναι γκαουσιανή με μέση τιμή μηδέν και τυπική απόκλιση ένα (Σχήμα 5.4). Τέλος, το στατιστικό λάθος των αποτελεσμάτων του φιτ θα πρέπει να είναι ανάλογο με

το πλήθος των μετρήσεων. Σχεδιάζοντας το κλασματικό λάθος ως συνάρτηση του πλήθους των ιστογραμμάτων, προκύπτει ότι ακολουθείται μία σχέση ρίζας όπως και αναμένεται (Σχήμα 5.5).

Background

Για να κατανοηθεί καλύτερα το background των αδρονικών διασπάσεων, είναι χρήσιμο να βρεθεί η σύνθεση του. Αυτό μπορεί να γίνει μελετώντας γεγονότα Monte Carlo και εξετάζοντας jets μεγάλης ακτίνας. Επιλέγονται jets με μάζα $40 - 600$ GeV και εγκάρσια ορμή $250 - 1000$ GeV. Στη συνέχεια αναλύεται η υποδομή τους ταξινομώντας τα jets μεταβλητής ακτίνας από τα οποία αποτελούνται σύμφωνα με την εγκάρσια ορμή και απαιτείται να είναι μεγαλύτερη των 10 GeV. Το υπό εξέταση μεγάλης ακτίνας jet επιλέγεται αν αποτελείται από τουλάχιστον δύο jets μεταβλητής ακτίνας που ικανοποιούν την προαναφερθείσα προϋπόθεση. Έπειτα εξετάζεται η γεύση των δύο μεταβλητής ακτίνας jets και για κάθε πιθανό συνδυασμό η τιμή της μάζας και της εγκάρσιας ορμής του jet μεγάλης ακτίνας προστίθενται στα αντίστοιχα ιστογράμματα. Η κατηγοριοποίηση γίνεται ανάλογα αν το ζευγάρι jets προέρχεται από b, c ή ελαφριά κουάρκ (u, d, s) που συμβολίζονται με q. Μπορεί έτσι να βρεθεί το κλάσμα για κάθε γεύση κουάρκ-αντικουάρκ που αποτελεί το background. Οι δισδιάστατες γραφικές παραστάσεις του Σχήματος 6.1 απεικονίζουν πως μεταβάλλεται το κλάσμα των διαφόρων συνδυασμών γεύσεων ανάλογα με την μάζα και την εγκάρσια ορμή. Τα ελαφριά κουάρκ υπερισχύουν σε όλο το φάσμα ακόμα και για μεγάλες τιμές της ορμής. Στην περίπτωση των b-κουάρκ, το κλάσμα τους μειώνεται καθώς η εγκάρσια ορμή αυξάνεται κάτι το οποίο διαισθητικά δεν αναμενόταν. Η ομαλή εξάρτηση που παρατηρείται και στους δύο άξονες είναι μία σημαντική παρατήρηση καθώς δικαιολογεί την παρεμβολή ανάμεσα σε γνωστές περιοχές για την εύρεση της σύνθεσης σε κάποια άγνωστη περιοχή. Εξετάζεται επίσης πως μεταβάλλεται η σύνθεση του background ανάλογα με το αν τα μεταβλητής ακτίνας jets έχουν αναγνωριστεί ως b-jets (b-tagged) ή όχι. Τα διαγράμματα του Σχήματος 6.3 υποδεικνύουν μία διαφοροποίηση στην απόδοση του b-tagging ανάλογα με την εγκάρσια ορμή.

Η μέθοδος φιτ μπορεί να εφαρμοστεί στο background για να εξεταστεί πόσο καλά μπορεί να βρεθεί η σύνθεση του. Αναλύονται οι περιοχές μάζας στις οποίες επικρατεί σε σχέση με άλλες φυσικές διαδικασίες (40 - 70 GeV, 200 - 300 GeV και 300 - 400 GeV). Η δημιουργία των templates και των συνολικών κατανομών γίνεται με την ίδια διαδικασία που περιγράφηκε παραπάνω. Η διαφορά έγκειται στο ότι αφού έχουν επιλεχθεί τα δύο jets μεταβλητής ακτίνας οι τιμές του αλγορίθμου MV2c10 για κάθε ένα συνδυάζονται σε μία συνολική πιθανότητα σύμφωνα με την Εξίσωση 6.1. Εξετάζοντας έπειτα την γεύση των δύο κουάρκ μπορούν να κατασκευαστούν templates για κάθε συνδυασμό κουάρκ-αντικουάρκ. Καθώς ο συνδυασμός cb έχει την μικρότερη συμβολή, φτιάχνονται μόνο πέντε templates για τα qq, cc, bb, qb, qc αντίστοιχα. Δεν παρουσιάζουν σημαντική διαφοροποίηση στις διαφορετικές περιοχές μάζας συνεπώς μπορεί να χρησιμοποιηθεί ο μέσος όρος τους (Σχήματα 6.5α-6.5ε). Ο βέλτιστος αριθμός templates για το φιτ μπορεί να προσδιοριστεί λαμβάνοντας υπόψιν τόσο την ικανοποιητική περιγραφή του background όσο και τον περιορισμό των ελεύθερων μεταβλητών. Τρία templates δεν είναι αρκετά καθώς υπάρχει σημαντική συνεισφορά από τους ανάμεικτους όρους qc και qb. Ωστόσο πέντε templates δυσχεραίνουν σημαντικά το φιτ με αποτέλεσμα κάποιο κλάσμα να έχει πάντα μηδενική τιμή. Συνεπώς, τα δύο templates των qc και qb συνδυάζονται σε ένα υπολογίζοντας την μέση τιμή τους (Σχήμα 6.5φ). Τα αποτελέσματα αυτά συνοψίζονται στους Πίνακες 6.2 - 6.9. Καθώς υπάρχουν περιπτώσεις που το φιτ δεν ανταναλά σωστά τα προβλεπόμενα αποτελέσματα είναι απαραίτητο οι ελεύθερες μεταβλητές να περιοριστούν περαι-

τέρω. Το Σχήμα 6.6 αναπαριστά τις προβλεπόμενες τιμές και τα αποτελέσματα του φιτ για τα τέσσερα κλάσματα στις τρεις περιοχές μάζας. Η σύγκριση αυτή υποδηλώνει μία αντισυσχέτιση μεταξύ των τιμών για τα qq και $qcqb$ που επιτρέπει την αντικατάσταση τους στην συνάρτηση ελαχιστοποίησης του φιτ από μία συνολική παράμετρο που αντανακλά το άθροισμα τους. Για να υλοποιηθεί σωστά αυτή η αλλαγή εισάγεται ο σχετικός τους λόγος που ορίζεται σύμφωνα με την Εξίσωση 6.3 και μένει σταθερός στην προβλεπόμενη τιμή.

Μεθοδολογία

Σε αυτό το σημείο μπορεί να αναπτυχθεί πλήρως η μεθοδολογία για την εξαγωγή των λόγων διακλάδωσης. Βασίζεται στην τεχνική αφαίρεσης του background. Αρχικά πρέπει να υπολογιστεί μία καλή εκτίμηση του background στην περιοχή τους σήματος. Έπειτα καθορίζεται η σύνθεση του αθροίσματος των διαδικασιών background και σήματος στην περιοχή του σήματος καθώς σε αυτή την περιοχή υπάρχουν συμβολές και από τις δύο που δε μπορούν να αναλυθούν ξεχωριστά. Εφαρμόζοντας στη συνέχεια την Εξίσωση 7.1 μπορούν να υπολογιστούν οι αδρονικοί λόγοι διακλάδωσης. Οι λόγοι R_{bkg} και R_{sig} χρησιμοποιούνται για να αναπροσαρμοστούν τα κλάσματα κάθε γεύσης ανάλογα με το σχετικό ποσοστό γεγονότων σήματος και background στην περιοχή του σήματος και ορίζονται σύμφωνα με την Εξίσωση 7.3. Είναι επίσης σημαντικό να ακολουθηθεί σωστά η διάδοση σφαλμάτων στα διάφορα μεγέθη καθώς είναι αυτά που καθορίζουν την ακρίβεια της μεθόδου. Συνδυάζοντας τις Εξισώσεις 7.5 και 7.7 με τα σφάλματα που προκύπτουν από το φιτ διαδίδονται ορθώς τα σφάλματα στα τελικά κλάσματα.

Η εκτίμηση του background στην περιοχή του σήματος μπορεί να προσεγγιστεί με διάφορους τρόπους. Ο απλούστερος είναι απλά να μετρηθεί το background σε μία περιοχή που υπερισχύει (side-band subtraction). Για να είναι έγκυρος αυτός ο τρόπος θα πρέπει να μην αλλάζει η σύνθεση του σε σχέση με την μάζα, δηλαδή να επιδεικνύει τα ίδια χαρακτηριστικά και στην περιοχή του σήματος. Ωστόσο, όπως φάνηκε στην προηγούμενη ενότητα, αυτό δεν ισχύει για το QCD background. Καλύτερα λοιπόν θα ήταν να χρησιμοποιηθούν δύο περιοχές μάζας όπου υπάρχει μόνο background για την παρεμβολή στην περιοχή του σήματος. Η παρεμβολή μπορεί να υλοποιηθεί είτε με πολυώνυμο πρώτου βαθμού είτε με δευτέρου. Για την πρώτη περίπτωση χρησιμοποιούνται οι Εξισώσεις 7.8 και 7.9 ενώ για την δεύτερη οι Εξισώσεις 7.13 και 7.14. Τα αποτελέσματα του Πίνακα 7.1 δείχνουν ότι η περιοχή μάζας 40 - 70 GeV δε μπορεί να γίνει τόσο καλά φιτ όσο η 200 - 300 GeV καθώς έχει μεγαλύτερη συνεισφορά cc. Αυτό δυσχεραίνει το φιτ αφού το αντίστοιχο template έχει σημαντική επικάλυψη με τα υπόλοιπα. Συνεπώς, η πρώτη μέθοδος εκτίμησης του background δε μπορεί να εφαρμοστεί στην πράξη. Σύμφωνα με τον Πίνακα 7.2 η παρεμβολή με πολυώνυμο δευτέρου βαθμού διαδίδει μεγαλύτερα σφάλματα άρα προτιμότερη είναι η γραμμική παρεμβολή.

Στους Πίνακες 7.3 - 7.6 συνοψίζονται τα αποτελέσματα της εφαρμογής της μεθόδου σε απλοποιημένες περιπτώσεις έτσι ώστε να δοθεί μία εικόνα της ακρίβειας της. Προκύπτει ότι οι λόγοι διακλάδωσης του σήματος μπορούν να βρεθούν αρκετά καλά ενώ κάποιες μικρές αποκλίσεις εξηγούνται από την ύπαρξη του τέταρτου template για τον ανάμεικτο όρο. Ωστόσο, τα σφάλματα της μεθόδου είναι αρκετά μεγάλα και σχετίζονται με την αφαίρεση που χρησιμοποιείται για την εξαγωγή των λόγων διακλάδωσης. Καθώς ο αριθμός των γεγονότων background είναι πολύ μεγαλύτερος από αυτόν του σήματος, είναι φυσικό η μέθοδος να διαδίδει σημαντικά σφάλματα στα τελικά αποτελέσματα.

Η μεθοδολογία υπολογισμού των λόγων διακλάδωσης εφαρμόζεται πρώτα στην περιοχή σήματος των ZW. Οι λόγοι διακλάδωσης αφορούν και τα δύο μποζόνια καθώς η ανάλυση

γίνεται στην περιοχή μάζας 70 - 105 GeV και υπολογίζονται μέσω της Εξίσωσης 7.15. Επιλέγεται γραμμική παρεμβολή για την εκτίμηση του background ώστε να διαδοθούν μικρότερα σφάλματα. Γίνεται εφαρμογή της μεθόδου τόσο σε δείγματα MC όσο και σε δεδομένα χρησιμοποιώντας τους σχετικούς λόγους που δίνονται από την Εξίσωση 7.16. Τα αποτελέσματα για τις δύο περιπτώσεις παρουσιάζονται στους Πίνακες 7.7 και 7.8. Καθώς τα κλάσματα που προκύπτουν δεν έχουν τιμές ανάμεσα στο μηδέν και το ένα, ακολουθείται μία διαδικασία εξαγωγής φυσικών τιμών. Τα αποτελέσματα για τις περιοχές του background και του σήματος χρησιμοποιούνται για την δημιουργία κατανομών που στη συνέχεια δειγματοληπτούνται. Η αφαίρεση επαναλαμβάνεται με αυτές τις τιμές και κρατούνται μόνο οι τιμές που έχουν φυσικό νόημα. Βρίσκοντας την μέση τιμή και την τυπική απόκλιση τους προκύπτει μία καλύτερη εκτίμηση για τους λόγους διακλάδωσης που παρουσιάζεται στην τελευταία στήλη των προαναφερθέντων πινάκων. Ωστόσο τα αποτελέσματα αυτά δεν είναι αξιόπιστα καθώς μόνο ένα μικρό εύρος των κατανομών ανήκει στο $(0, 1)$. Αντίστοιχα γίνεται η εφαρμογή της μεθόδου στην περιοχή σήματος του μποζονίου Higgs. Οι δύο περιοχές του background παρεμβάλλονται γραμμικά ώστε να βρεθεί μία εκτίμηση της συμβολής του στην περιοχή μάζας 105 - 145 GeV. Χρησιμοποιώντας τους λόγους της Εξίσωσης 7.17 προκύπτουν τα αποτελέσματα του Πίνακα 7.10. Παρόλο που έχουν φυσικές τιμές δεν είναι σε συμφωνία με τις προβλεπόμενες από το MC τιμές.

Με βάση όλα τα παραπάνω είναι πλέον εμφανές ότι μία συνολική μέτρηση των αδρονικών λόγων διακλάδωσης δεν είναι εφικτή με την τρέχουσα ακρίβεια του flavour tagging. Η διακριτική ικανότητα μεταξύ των διαφόρων γεύσεων δεν είναι αρκετά μεγάλη και επιδεινώνεται από την εισαγωγή του τέταρτου template το οποίο παρουσιάζει μεγάλη επικάλυψη με τα υπόλοιπα. Τέλος, τα μεγάλα σφάλματα της μεθόδου οφείλονται στο μικρό ποσοστό γεγονότων σήματος το οποίο θα μπορούσε να αυξηθεί χρησιμοποιώντας την συσχετιζόμενη παραγωγή του μποζονίου Higgs με ένα μποζόνιο Z ή W.

Τροποποιήσεις

Καθώς η μέθοδος που αναπτύχθηκε δεν είναι εφικτή με την τρέχουσα ακρίβεια του αλγορίθμου MV2c10 είναι ενδιαφέρον να εξεταστούν τροποποιήσεις στην διακριτική ικανότητα ανάμεσα στις διαφορετικές γεύσεις. Η έξοδος του αλγορίθμου μπορεί να μεταβληθεί ανάλογα με την πραγματική γεύση κάθε jet. Ιδανικά η έξοδος θα ήταν μία κατανομή με μία δεδομένη κεντρική τιμή και εύρος που θα μειωνόταν καθώς η διακριτική ικανότητα του αλγορίθμου αυξάνεται. Πρακτικά αυτό ήδη ισχύει για τα b και τα ελαφριά jets. Οι τρεις κεντρικές τιμές μπορούν να είναι τα 0, 0.5, 1. Μία τροποποιημένη πιθανότητα μπορεί λοιπόν να οριστεί σύμφωνα με τις Εξισώσεις 8.1 και 8.2 όπου το a παίρνει τιμές από το 0 μέχρι το 1 και αντανακλά τον βαθμό τροποποίησης. Για $a = 1$ η πιθανότητα μένει αμετάβλητη. Εφαρμόζοντας αυτές τις σχέσεις για την τιμή του αλγορίθμου για όλα τα μεταβλητής ακτίνας jets, μπορούν να τροποποιηθούν τόσο τα templates όσο και οι συνολικές κατανομές. Το Σχήμα 8.1 παρουσιάζει τις γραφικές παραστάσεις της απόδοσης του b-tagging σε σχέση με τους συντελεστές απόρριψης για τα c και τα ελαφριά jets για διάφορες τιμές του a. Παρατηρείται ότι οι συντελεστές απόρριψης αυξάνονται σημαντικά μέχρι το σημείο που υπάρχει κορεσμός καθώς η απόδοση μειώνεται. Καθώς το a μικραίνει ο κορεσμός παρατηρείται για ολοένα και μεγαλύτερες τιμές της απόδοσης για τα b-jets. Αντίστοιχα, οι γραφικές παραστάσεις του Σχήματος 8.2 δείχνουν πως η επικάλυψη μεταξύ των templates μειώνεται σημαντικά για μικρότερες τιμές του a. Για $a = 0.7$ μπορούν να θεωρηθούν πλήρως διαχωρισμένα, αφού λαμβάνουν την μέγιστη τιμή τους σε διαφορετικά σημεία.

Η μέθοδος μπορεί να επαναληφθεί με τις τροποποιημένες τιμές του MV2c10. Οι Πίνακες 8.3 και 8.4 δείχνουν τα αποτελέσματα για το φιτ του background για διαφορετικές τιμές του a. Για $a = 0.9$ δεν παρατηρείται βελτίωση αλλά για $a = 0.7$ το κλάσμα του cc δεν είναι πλέον μηδενικό στην περιοχή χαμηλής μάζας. Για τις άλλες δύο τιμές η πραγματική σύνθεση ανταναικλάται με μεγάλη ακρίβεια. Έτσι, εφαρμόζοντας γραμμική παρεμβολή, δίνεται μία πολύ καλύτερη εκτίμηση του background στην περιοχή του σήματος (Πίνακας 8.5). Τα αποτελέσματα του Πίνακα 8.7 για τους λόγους διακλάδωσης των ZW συμφωνούν σε μεγάλο βαθμό με τις προβλεπόμενες τιμές. Παρόμοια βελτιώνονται τα αποτελέσματα της εφαρμογής της μεθόδου στο μποζόνιο Higgs. Ωστόσο, τα αποτελέσματα του Πίνακα 8.11 για την περιοχή σήματος δείχνουν ότι απαιτείται σημαντική αύξηση στην διακριτική ικανότητα του αλγορίθμου για να εξαχθεί μη αμελητέα τιμή του κλάσματος του cc. Επίσης, σε όλες τις περιπτώσεις του a, οι τελικοί λόγοι διακλάδωσης δεν έχουν φυσικές τιμές, κάτι που οφείλεται στον μικρό αριθμό γεγονότων σήματος. Εφαρμόζοντας την διαδικασία που περιγράφηκε στην προηγούμενη ενότητα για εξαγωγή φυσικών τιμών, λαμβάνονται πολύ καλές εκτιμήσεις των λόγων διακλάδωσης ιδίως για $a = 0.3$. Είναι λοιπόν φανερό ότι η αποτελεσματικότητα της μεθόδου σχετίζεται άμεσα με την ακρίβεια του flavour tagging. Μία συνολική μέτρηση των λόγων διακλάδωσης είναι εφικτή αν η διακριτική ικανότητα μεταξύ των c-jets και των υπολοίπων αυξηθεί σημαντικά.

Chapter 1

Introduction

Throughout the 20^{th} century, the Standard Model (SM) of the elementary particles was developed, to explain the properties observed at the subatomic scale. Its basis is formed by a unified description of the electromagnetic force and the weak force and the associated force-carrying particles, the photon, and the W and Z bosons. However, although the latter have mass, this property is not intrinsic to the SM and its introduction would violate a fundamental invariance of the theory. This problem was solved by the Brout-Englert-Higgs mechanism which predicts the existence of a neutral particle known as the Higgs boson. The search for it motivated the construction of particle accelerators, the LHC being the latest, but for many years there was no experimental observation. Finally, its discovery in 2012 by the ATLAS and CMS experiments marked a key moment in the history of particle physics and paved the way for a detailed exploration of its properties. The last few years there have been multiple findings about its spin, parity and decay channels. Nevertheless, there are still unanswered questions related to the Higgs boson nature. It is not determined if the observed particle is the one predicted by the SM as there are other types of Higgs bosons predicted by theories beyond the Standard Model (BSM). The experimental uncertainties leave room for alternative explanations.

One important property of the Higgs boson is its coupling to the mass of the particles it interacts with. The SM predicts that the strength of the interaction is proportional to the mass. The way to test this experimentally is through measuring the partial decay widths, which can be expressed through the branching ratios. Therefore, the measurement of the Higgs boson branching ratios in all possible decay modes constitutes an important part of the Higgs sector and its interest is twofold. First, it is essential for giving a precision test of the SM predictions. At the same time, a possible deviance from these predictions could be interpreted in terms of BSM physics and provide supporting evidence for theories such as the Minimally Supersymmetric Standard Model (MSSM). Non-standard Higgs couplings are allowed within the present uncertainties in various Higgs coupling measurements. Hence, it is essential to be able to measure all the branching ratios. However, currently at the LHC, it is difficult to determine the total and partial decay widths. More specifically, the $H \to c\bar{c}$ decay which has a non-negligible branching ratio, has not been observed experimentally yet as the c-quark cannot be tagged efficiently. For this reason, its important to test if the branching ratios and specifically the hadronic ones, could be measured inclusively.

The aim of the thesis is to propose a method for measuring the Higgs boson branching ratios to $b\bar{b}, c\bar{c}, gg$ using the response of a flavour tagging algorithm and study if it's possible

with the present accuracy. While usually the output of this algorithm is used by imposing a cut on its value which determines the flavour of a given jet, this study uses it inclusively. The selected events for the analysis consist of all the quark flavours. Therefore, it has the advantage of increased statistics. The goal is to then extract the contribution of each component using di-jet discriminants. These are formed by combining the output of the MV2c10 algorithm while also knowing the true flavour of each jet and represent a pair of quarks. A technique is developed to fit the discriminants to the sample and extract its composition. However, the difficulty of the method lies on the overwhelming number of hadronic jets present in the detector. There are quark-antiquark pairs produced by other processes involving the strong interaction which must be modelled separately. Hence, the present thesis studies this background and its flavour composition. Other discriminants must be introduced except for those corresponding to the three main decay channels. The study tries to find the exact contribution of the background processes so as to remove it and be left with only the flavour composition of the decays of a given boson. This will essentially give the branching ratios. The method is tested for both the Higgs boson and the ZW bosons as they also decay hadronically. At the course of this study, it is shown that is not possible to inclusively measure the branching ratios with the current response of the flavour tagging algorithm. Therefore, the study considers modified outputs of the algorithm which progressively increase the discrimination between the different flavours. The method is repeated for these cases so as to determine the required rejection rates for it to be feasible.

The thesis has the following structure. Chapter 2 gives a theoretical overview of the Standard Model and the Higgs mechanism and presents some relevant aspects of the Minimally Supersymmetric Standard Model. Chapter 3 focuses on the phenomenology of the Higgs boson, explaining its production and decay modes, while also mentioning relevant experimental results. Chapter 4 discusses experimental methods and physics objects integral to the present analysis. In chapter 5, the developed fitting method is analysed and validated. Chapter 6 is dedicated to the study of the background of the hadronic processes. Its flavour composition is studied and the fit method is applied to extract its composition. Chapter 7 develops the complete methodology of finding the branching ratios and then implements it in various cases. Chapter 8 examines modifications to the output of the flavour tagging algorithm and repeats the method using them. Finally, the conclusions of the thesis are summed up in chapter 9.

Chapter 2

Theoretical Framework

This chapter provides the theoretical basis for explaining physics concepts relevant to the topic of this thesis. The main features of the Standard Model are briefly discussed in the first section while the following section presents the Higgs mechanism. The final section mentions some basic ideas of the Minimally Supersymmetric Standard Model concerning the Higgs boson.

2.1 The Standard Model

The Standard Model of elementary particles is the theory describing three of the four known fundamental forces: the electromagnetic, weak, and strong interactions. It also classifies the known elementary particles (Figure 2.1) based on their interaction properties and physical quantities like mass, charge and spin. It includes twelve elementary particles of spin-1/2 called fermions, which are further classified into quarks and leptons based on their interactions. Fermions carry both electric charge (except for neutrinos) and weak isospin, so they interact with other fermions via electromagnetism and the weak interaction. However, quarks also carry colour charge which permits them to interact via the strong interaction, as well. Each fermion has a corresponding antiparticle. The SM also includes gauge bosons of spin-1 which are the force carriers that mediate the fundamental interactions: the photon mediates the electromagnetic force between charged particles, the $W^{\pm}$ and Z bosons mediate the weak interactions and the gluons mediating the strong interactions. The final piece of the SM is the Higgs boson, which is a spin-0 particle and explains why the other elementary particles are massive.

The mathematical formulation of the SM is based on combining the principles of special relativity and quantum mechanics in a quantum field theory. Elementary particles are treated as excitations of quantum fields which can be described by a Lagrangian $\mathcal{L} = \mathcal{L}_{free} + \mathcal{L}_{int}$. The first term represents each participating field and determines the propagator, while the latter consists of interaction terms. The SM is a gauge theory which means that its Lagrangian is invariant under local transformations from certain Lie groups. The transformations between possible gauges form the symmetry group of the theory. Specifically, the local $SU(3) \times SU(2) \times U(1)$ gauge symmetry essentially defines the SM, as all fundamental interactions are generated by demanding that a global invariance holds locally.

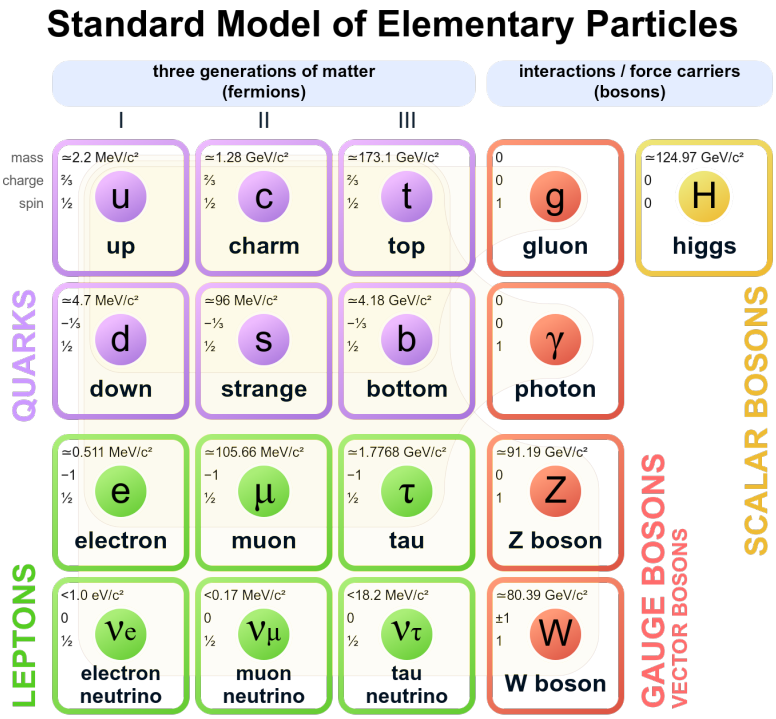


Figure 2.1: Elementary particles of the SM

2.1.1 Free particles

The dynamics of the quantum mechanical fields describing spin-0, spin-1/2 and spin-1 particles are determined by the appropriate Lagrangian [1]. The equations of motion can then be derived by the Euler-Lagrange equations:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi_i \right)} \right) - \frac{\partial \mathcal{L}}{\partial \phi_i} = 0 \tag{2.1}$$

A scalar (spin-0) particle is described as an excitation of a scalar field $\phi(x)$ with the Lagrangian:

$$\mathcal{L} = \frac{1}{2} \left(\partial_\mu \phi \right) \left(\partial^\mu \phi \right) - \frac{1}{2} m^2 \phi^2 \tag{2.2}$$

Applying the Euler-Lagrange equation gives:

$$\partial_\mu \partial^\mu \phi + m^2 \phi = 0 \tag{2.3}$$

which is the Klein-Gordon equation for a free scalar field $\phi(x)$.

A spin-1/2 particle corresponds to a field $\psi(x)$ for which the Lagrangian is:

$$\mathcal{L} = i \bar{\psi} \gamma^\mu \partial_\mu \psi - m \bar{\psi} \psi \tag{2.4}$$

The field $\psi(x)$ is a four-component complex spinor which satisfies the Dirac equation:

$$i \gamma^\mu \partial_\mu \psi - m \psi = 0 \tag{2.5}$$

Finally, a vector field A^μ has the Lagrangian:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}m^2A^\mu A_\mu \tag{2.6}$$

where:

$$F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu \tag{2.7}$$

and the Euler-Lagrange equation yields:

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = 0 \tag{2.8}$$

which is the Proca equation.

2.1.2 Local gauge invariance and electromagnetism

Interactions can be introduced by demanding a local gauge symmetry. While for example, the Dirac Lagrangian of 2.4 is invariant under a global phase transformation:

$$\psi \to \psi' = e^{i\theta}\psi \tag{2.9}$$

it is not invariant under a local one:

$$\psi(x) \to \psi'(x) = e^{iq\chi(x)}\psi(x) \tag{2.10}$$

For the Lagrangian to be invariant under local phase transformations, the derivative $\partial_m u$ is replaced with the covariant derivative D_μ:

$$\partial_\mu \to D_\mu = \partial_\mu + iqA_\mu \tag{2.11}$$

and the new field A_μ must transform as:

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu\chi \tag{2.12}$$

The new Lagrangian:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu A_\mu\psi \tag{2.13}$$

is invariant under local phase transformations. The field A_μ can be identified as the electromagnetic vector potential (photon) and thus satisfies the Proca Lagrangian 2.6 for $m = 0$. This term can be added to the Lagrangian, so the complete expression:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu A_\mu\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{2.14}$$

now describes one massive spinor field interacting with a massless vector field. Essentially, the principle of a $U(1)$ local gauge invariance has produced the field theory of electromagnetism known as quantum electrodynamics (QED).

2.1.3 Yang-Mills theory

The weak interaction is based on the $SU(2)$ symmetry group so the Lagrangian must be invariant under transformations of 2×2 unitary matrices. Only the left chiral components of the spinors carry weak isospin charge. Taking a two-component column vector such as:

$$\psi = \begin{pmatrix} \nu_e \\ e \end{pmatrix} \tag{2.15}$$

with each component being a Dirac spinor, the total Lagrangian will also be a Dirac Lagrangian:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi \tag{2.16}$$

The transformation that will be applied is:

$$\psi \rightarrow \psi' = (I + ig\boldsymbol{\alpha}(\boldsymbol{x})\cdot\mathbf{T})\psi \tag{2.17}$$

where $\mathbf{T}$ are the generators of $SU(2)$ and can be expressed in terms of the Pauli matrices as $\mathbf{T} = \sigma/2$. The covariant derivative is:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig\mathbf{W}^\mu \cdot \mathbf{T} \tag{2.18}$$

where $\mathbf{W} = \{W_1, W_2, W_3\}$ are the three new gauge fields. After finding the transformation rule for these fields (which is not trivial) and also including their free Lagrangians, Equation 2.16 can be expressed as:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi - \frac{1}{4}\mathbf{F}^{\mu\nu}\mathbf{F}_{\mu\nu} - g\bar{\psi}\gamma_\mu \mathbf{W}^\mu \cdot \mathbf{T}\psi \tag{2.19}$$

where

$$\mathbf{F}^{\mu\nu} \equiv \partial^\mu \mathbf{W}^\nu - \partial^\nu \mathbf{W}^\mu - g\mathbf{W}^\mu \times \mathbf{W}^\nu \tag{2.20}$$

However, although the resulting Lagrangian is invariant under local $SU(2)$ transformation, it describes two Dirac fields interacting with three massless vector gauge fields. This does not correspond to the weak interaction as its vector bosons are massive. As will be explained in 2.2.3, the Higgs mechanism can resolve this problem and the electromagnetic and weak interaction can be described in a unified description of $SU(2) \times U(1)$ symmetry.

2.1.4 Strong interaction

The SM also explains the strong interaction. The corresponding field theory, Quantum Chromodynamics (QCD) can be similarly obtained by extending the local gauge principle so that the Lagrangian is invariant under non-abelian $SU(3)$ local phase transformations. The transformations correspond to 3×3 unitary matrices. The fermion fields that carry colour charge (quarks) are represented as:

$$\psi = \begin{pmatrix} \psi_r \\ \phi_g \\ \psi_b \end{pmatrix} \tag{2.21}$$

with each component corresponding to a particular colour. The derivative is replaced with:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + i\alpha_s \boldsymbol{\lambda}\cdot\mathbf{A}_\mu \tag{2.22}$$

where $\mathbf{A}_\mu$ introduces eight gluon fields and $\boldsymbol{\lambda}$ are the eight 3×3 Gell-Mann matrices. The Lagrangian can be derived as [1]:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi - \alpha_s \bar{\psi}\gamma^\mu \boldsymbol{\lambda}\cdot\mathbf{A}_\mu \psi - \frac{1}{4} G_{\mu\nu} G^{\mu\nu} \tag{2.23}$$

where $G_{\mu\nu}$ is the gluon field strength tensor and is related to the gluon fields. This tensor introduces several terms corresponding to self-interactions of the gluon field.

The strong coupling constant α_s depends on the transferred four-momentum Q^2 and at large Q^2, a first-order perturbative calculation can be used to yield:

$$\alpha_s(Q^2) = \frac{12\pi}{(22 - 2n_f) \cdot \ln(Q^2/\Lambda^2_{QCD)}} \tag{2.24}$$

where $n_f = 6$ is the number of quark flavours and $\Lambda^2_{QCD} = 218$ MeV is the characteristic scale of the theory. All of the above result in QCD exhibiting two main properties:

- *Colour confinement*: Color charged objects cannot be isolated. This stems from gluon-gluon self-interactions, which are possible because gluons carry colour charge. When two quarks are separated, the extraction of a new quark–antiquark pair from the vacuum is energetically preferable over a further increase in distance. It explains the hadronization process which leads to the presence of jets in a detector (Section 4.2). This characteristic is also reflected in α_s, which increases rapidly as Q^2 decreases.

- *Asymptotic freedom*: The coupling constant α_s decreases as the energy scale of the interactions increase, which practically means that the interaction between quarks and gluons becomes asymptotically weaker. At very large Q^2, quarks can be considered to be free.

2.2 The Higgs mechanism

The Higgs mechanism can solve the problem of introducing the massive gauge fields by combining local gauge invariance with spontaneous symmetry breaking, as shown by Brout, Englert and Higgs [2, 3, 4]. It is described in this section by first displaying it in two simplified cases and then outlining its introduction in the SM, following the methodology found in [5].

2.2.1 Symmetry breaking for a real scalar field

First, it is shown how mass terms can arise and be identified in a Lagrangian. Assuming a scalar field ϕ with potential:

$$V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4 \tag{2.25}$$

the corresponding Lagrangian is:

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\right)\left(\partial^\mu\phi\right) - \frac{1}{2}\mu^2\phi^2 - \frac{1}{4}\lambda\phi^4 \tag{2.26}$$

The first term represents the kinetic energy of the scalar particle, the second its mass and the third can be thought as self-interactions of the scalar field. The ground state (vacuum) corresponds to the minimum of the above potential. The parameter λ must be positive for a finite minimum to exist but there are two cases for μ^2. If it is also positive, then the field configuration of the minimum energy is the trivial one: $\phi = 0$ and μ is indeed the mass. However, if μ^2 is negative, it can no longer represent the mass and the minimum occurs at:

$$\phi = \pm\upsilon = \pm\sqrt{\frac{-\mu^2}{\lambda}} \tag{2.27}$$

In this case, the field is said to have a non-zero vacuum expectation value υ. The choice between $\phi = +\upsilon$ and $\phi = -\upsilon$ breaks the symmetry of the Lagrangian (spontaneous symmetry breaking). This selection of an asymmetrical ground state conceals the true symmetry of the system. If the minimum $\phi = +\upsilon$ is chosen, the Lagrangian can then be expressed as excitations about it by introducing a new variable $\phi(x) = \upsilon + \eta(x)$. Using Equation 2.27 the Lagrangian is:

$$\mathcal{L}(\eta) = \frac{1}{2}\left(\partial_\mu\eta\right)\left(\partial^\mu\eta\right) - \lambda\upsilon^2\eta^2 - \lambda\upsilon\eta^3 - \frac{1}{4}\lambda\eta^4 + \frac{1}{4}\lambda\upsilon^4 \tag{2.28}$$

The second term is proportional to η^2 and can thus be interpreted as a mass:

$$m_\eta = \sqrt{2\lambda\upsilon^2} = \sqrt{-2\mu^2} \tag{2.29}$$

The third and fourth terms represent self-couplings and because the last term is constant, the Lagrangian can finally be written as:

$$\mathcal{L}(\eta) = \frac{1}{2}\left(\partial_\mu\eta\right)\left(\partial^\mu\eta\right) - \frac{1}{2}m_\eta^2\eta^2 + V(\eta) \tag{2.30}$$

with

$$V(\eta) = \lambda\upsilon\eta^3 + \frac{1}{4}\lambda\eta^4 \tag{2.31}$$

The two Lagrangians (2.26 and 2.31) represent the same physical system but the second one is necessary for applying perturbation theory.

2.2.2 Higgs mechanism for a complex scalar field

The above can be applied to a complex scalar field:

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \tag{2.32}$$

with the potential:

$$V(\phi) = \mu^2(\phi^*\phi) + \lambda(\phi^*\phi)^2 \tag{2.33}$$

and the corresponding Lagrangian:

$$\mathcal{L} = (\partial_\mu\phi)^*(\partial^\mu\phi) - V(\phi) \tag{2.34}$$

The Lagrangian is invariant under global $U(1)$ phase transformations of the form:

$$\phi \to \phi' = e^{i\alpha}\phi \tag{2.35}$$

as $\phi'^*\phi' = \phi^*\phi$. For $\mu^2 > 0$ the potential has a zero vacuum expectation value, while for $\mu^2 < 0$ the minima are determined by:

$$\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} = \upsilon^2 \tag{2.36}$$

Therefore, there is an infinite number of equivalent minima lying on a circle of radius $\sqrt{-\mu^2/\lambda}$. Choosing one breaks the rotational symmetry spontaneously (Figure 2.2). A

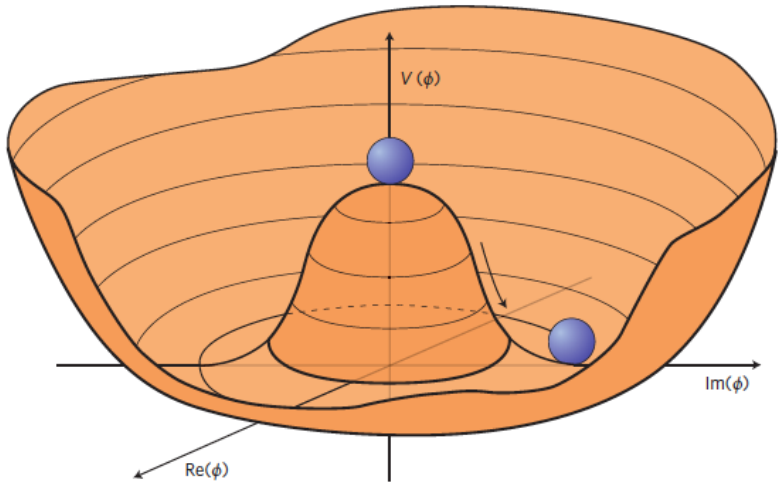


Figure 2.2: Higgs potential for $\mu^2 < 0$ which illustrates how the rotational symmetry is broken by choosing a non-zero vacuum expectation value [6].

specific ground state is chosen as:

$$\phi_{1_{min}} = \upsilon, \quad \phi_{2_{min}} = 0 \tag{2.37}$$

and the excitations of the vacuum can be considered by introducing new fields η and ξ with $\phi_1(x) = \eta(x) + \upsilon$ and $\phi_2(x) = \xi(x)$. The Lagrangian can be written in terms of η and ξ as:

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \eta\right)\left(\partial^\mu \eta\right) - \frac{1}{2}m_\eta^2\eta^2 + \frac{1}{2}\left(\partial_\mu \xi\right)\left(\partial^\mu \xi\right) - V_{int}(\eta, \xi) \tag{2.38}$$

where $m_\eta = \sqrt{2\lambda\upsilon^2}$ and V_{int} is:

$$V_{int}(\eta, \xi) = \lambda\upsilon\eta^3 + \frac{1}{4}\lambda\eta^4 + \frac{1}{4}\lambda\xi^4 + \lambda\upsilon\eta\xi^2 + \frac{1}{2}\lambda\eta^2\xi^2 \tag{2.39}$$

The terms with three or four powers of the fields correspond to interactions. The second term corresponds to a scalar massive particle, while the third to a massless one. The latter is known as a Goldstone boson and as it was proven in [7], spontaneous breaking of a continuous global symmetry is accompanied by the appearance of such a scalar massless particle. This is not compatible with the known elementary particles but can be remedied by applying spontaneous symmetry breaking to the case of local gauge invariance, which forms the Higgs mechanism.

To construct a theory that is invariant under local $U(1)$ transformations:

$$\phi(x) \to \phi'(x) = e^{ig\chi(x)}\phi(x) \tag{2.40}$$

the derivatives must be replaced with the covariant derivatives:

$$\partial_\mu \to D_\mu = \partial_\mu + igB_\mu \tag{2.41}$$

and the introduced gauge field B_μ should transform as:

$$B_\mu \to B'_\mu = B_\mu - \partial_\mu \chi(x) \tag{2.42}$$

In this case, the resulting Lagrangian:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + (D_\mu\phi)^*(D_\mu\phi) - \mu^2\phi^2 - \lambda\phi^4 \tag{2.43}$$

where $F^{\mu\nu} \equiv \partial^\mu B^\nu - \partial^\nu B^\mu$, exhibits local gauge invariance. As before, the potential is minimized and the Lagrangian is expressed as perturbations around the ground state:

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\eta\right)\left(\partial^\mu\eta\right) - \lambda\upsilon^2\eta^2 + \frac{1}{2}\left(\partial_\mu\xi\right)\left(\partial^\mu\xi\right) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2\upsilon^2 B_\mu B^\mu - V_{int} + g\upsilon B_\mu(\partial^\mu\xi) \tag{2.44}$$

The first three terms describe a massive scalar particle and a massless Goldstone boson. The fourth term is the $U(1)$ gauge field and while B was previously massless, it has now acquired mass, as shown by the fifth term. V_{int} contains interaction terms. Although the original problem of the mass has been resolved, the massive gauge boson has introduced an extra degree of freedom due to the additional longitudinal polarization state. There is also a direct coupling of the Goldstone field ξ to the gauge field B. The Goldstone field can be eliminated by making the transformation:

$$B_\mu \to B'_\mu = B_\mu + \frac{1}{g\upsilon}\partial_\mu\xi(x) \tag{2.45}$$

and choosing $\chi(x) = -\xi(x)/g\upsilon$, which in turn makes the transformation of $\phi(x)$ as:

$$\phi(x) \to \phi'(x) = e^{ig\frac{\xi(x)}{g\upsilon}}\phi(x) = e^{i\xi(x)/\upsilon}\phi(x) = \frac{\upsilon + \eta(x)}{\sqrt{2}} \tag{2.46}$$

The complex scalar field $\phi(x)$ has become entirely real and the Goldstone field $\xi(x)$ has been eliminated (unitary gauge). The Lagrangian can finally be written as:

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \lambda\upsilon^2\eta^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2\upsilon^2 B_\mu B^\mu \\ &+ g^2\upsilon B_\mu B^\mu\eta + \frac{1}{2}g^2 B_\mu B^\mu\eta^2 - \lambda\upsilon\eta^3 - \frac{1}{4}\lambda\eta^4\end{aligned} \tag{2.47}$$

It now describes a massive scalar field η (first two terms) and a massive gauge boson B (third and fourth term) which has absorbed the Goldstone boson. It also contains Higgs and gauge boson interaction terms (fifth and sixth) and Higgs self-interactions (seventh and eighth term). The mass of the gauge boson is identified as $m_B = g\upsilon$, while the Higgs mass is:

$$m_H = \sqrt{2\lambda}\upsilon \tag{2.48}$$

This means that the masses of the two bosons are directly related to the vacuum expectation value υ.

2.2.3 Generalization to the electroweak case

The previous methodology can be generalized to the electroweak case and create massive vector bosons for the electroweak interaction, as was proposed by Glashow, Salam and Weinberg [8, 9, 10]. In this case, the Higgs mechanism is needed to generate mass for a $SU(2) \times U(1)$ symmetry. To account for the W^+, W^- and Z bosons, three longitudinal degrees of freedom are needed, which are provided by three Goldstone bosons. There will also be a massive scalar particle corresponding to the excitations of the vacuum state. Therefore, the simplest Higgs model consists of two complex scalar fields:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.49}$$

where ϕ_0 and ϕ^+ are used to provide mass for the neutral and charged vector bosons, respectively. The corresponding Higgs potential is:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{2.50}$$

and the Lagrangian:

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) \tag{2.51}$$

The minima satisfy the equation:

$$\phi^\dagger \phi = \frac{v^2}{2} = -\frac{\mu^2}{2\lambda} \tag{2.52}$$

To break both the $SU(2)$ and $U(1)$ symmetries but preserve invariance under $U(1)$ symmetry so that the photon remains massless, the minimum is chosen as: $\phi^+_{min} = 0$ and $\phi^0_{min} = \frac{v}{\sqrt{2}}$. The field is written in the unitary gauge to eliminate the Goldstone bosons:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.53}$$

The masses of the gauge bosons can be identified by using the covariant derivative:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig\mathbf{T} \cdot \mathbf{W}_\mu + ig' \frac{Y}{2} B_\mu \tag{2.54}$$

where Y is the weak hypercharge, which is related to the third component of the weak isospin T_3 and the electric charge Q by:

$$Q = T_3 + \frac{1}{2} Y_W \tag{2.55}$$

This derivative mixes the fields W_1, W_2, W_3, B, producing the physically observable fields $W^\pm, Z^0, A$ as:

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \tag{2.56}$$

$$A = \cos\theta_W B \sin\theta_W W_3 \tag{2.57}$$

$$Z^0 = -\sin\theta_W B cos\theta_W W_3 \tag{2.58}$$

where θ_W is related to the ratio of the the couplings of $U(1)$ and $SU(2)$ as:

$$\tan\theta_W = \frac{g'}{g} \tag{2.59}$$

The mass terms that appear yield:

$$m_W = \frac{1}{2}gv, \quad m_A = 0, \quad m_Z = \frac{1}{2}v\sqrt{g^2+g'^2} \tag{2.60}$$

and the masses of the W and Z bosons satisfy:

$$\frac{m_W}{m_Z} = \frac{g}{\sqrt{g^2+g'^2}} = \cos\theta_W \tag{2.61}$$

These two couplings along with μ and λ are the four parameters of the Salam-Weinberg model. From the measured values of m_W, g, the vacuum expectation value of the Higgs field is found to be $v = 246$ GeV.

2.2.4 Fermion masses

The final step is to consider the masses of the fermions, as they can also be generated by the Higgs mechanism. The fermion mass term in the Dirac Lagrangian:

$$-m\bar{\psi}\psi = -m(\bar{\psi_L}\psi_R + \bar{\psi_R}\psi_L) \tag{2.62}$$

is not invariant under $SU(2)\times U(1)$ transformations, so it must be replaced with another one. As the weak interaction acts only on left-handed particles and right-handed antiparticles, a chiral representation is used. Left-handed states are placed in weak isospin $SU(2)$ doublets with $T_3 = 1/2$ for the upper component and $T_3 = -1/2$ for the lower, while right-handed states are placed in singlets with $T = T_3 = 0$. The doublets and singlets are the following:

$$L_L = \begin{pmatrix}\nu_L^e\\ e_L\end{pmatrix}, \begin{pmatrix}\nu_L^\mu\\ \mu_L\end{pmatrix}, \begin{pmatrix}\nu_L^\tau\\ \tau_L\end{pmatrix}, \quad L_R = e_R, \mu_R, \tau_R \tag{2.63}$$

$$Q_L = \begin{pmatrix}u_L\\ d_L\end{pmatrix}, \begin{pmatrix}c_L\\ s_L\end{pmatrix}, \begin{pmatrix}t_L\\ b_L\end{pmatrix}, \quad U_R = u_R, c_R, t_R, \quad D_R = d_R, s_R, b_R \tag{2.64}$$

The two complex scalar fields of the Higgs mechanism are also placed in an $SU(2)$ doublet, so by introducing the Higgs field as:

$$\mathcal{L}_f = -g_f(\bar{\psi_L}\phi\psi_R + \bar{\psi_R}\phi^\dagger\psi_L) \tag{2.65}$$

the Lagrangian is invariant due to the presence of two $SU(2)$ doublets. The constant g_f is known as the Yukawa coupling. For the electron doublet, using Equation 2.53, the term becomes:

$$\mathcal{L}_e = -\frac{g_e}{\sqrt{2}}\left[\begin{pmatrix}\bar{\nu_e} & \bar{e}\end{pmatrix}_L \begin{pmatrix}0\\ v+h(x)\end{pmatrix} l_R + \bar{e_R}\begin{pmatrix}0 & v+h(x)\end{pmatrix}\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L\right] \tag{2.66}$$

$$= -\frac{g_e}{\sqrt{2}}v(\bar{e_L}e_R + \bar{e_R}e_L) - \frac{g_e}{\sqrt{2}}h(\bar{e_L}e_R + \bar{e_R}e_L) \tag{2.67}$$

The Yukawa coupling is not predicted by the Higgs mechanism, but in order to be consistent with the observed mass, can be chosen as:

$$g_e = \sqrt{2}\frac{m_e}{v} \tag{2.68}$$

so 2.66 becomes:

$$\mathcal{L}_e = -m_e\bar{e}e - \frac{m_e}{v}\bar{e}eh \tag{2.69}$$

The first term gives the mass of the lepton and represents the coupling to the Higgs field, while the second shows a coupling between the lepton and the Higgs boson.

Equation 2.65 can be used only for the lower components of the doublets due to the zero vacuum expectation value of the Higgs field in the upper component. Therefore, another mechanism is needed for up-type quarks, which can be achieved by using the conjugate doublet ϕ_c instead of ϕ:

$$\phi_c = -i\sigma_2\phi* = \begin{pmatrix} -\phi^{0*} \\ \phi^- \end{pmatrix} \tag{2.70}$$

with the term in the Lagrangian being:

$$\mathcal{L}_f = -g_f(\bar{\psi_L}\phi_c\psi_R + \bar{\psi_R}\phi_c^\dagger\psi_L) \tag{2.71}$$

This way, all fermions can acquire mass and the Yukawa couplings to the Higgs field are given by:

$$g_f = \sqrt{2}\frac{m_f}{v} \tag{2.72}$$

2.3 Minimal Supersymmetric Standard Model

Although the SM has been proven successful by having many of its predictions experimentally confirmed, it is not a complete theory of the fundamental interactions and cannot explain various phenomena. The main motivations which call for BSM theories are the naturalness and hierarchy problem, the gauge coupling unification problem and a candidate particle for dark matter. All of these can be resolved by assuming that there is a symmetry relating fermions and bosons, called a supersymmetry (SUSY) [11]. Supersymmetry predicts the existence of a partner to every known particle which differs in spin by 1/2 and has various phenomenological advantages. A low energy model of SUSY realization is the Minimal Supersymmetric Standard Model (MSSM), which is characterised as minimal because it considers only the minimum number of new particle states and new interactions consistent with phenomenology [12].

Supesymmetric models involve chiral multiplets whose complex conjugates belong to multiplets of the opposite chirality, thus they cannot couple together in the Lagrangian. This, along with the fact that scalar fields belong in the same chiral multiplets together with spin-1/2 fields, lead to the need for an even number of Higgs multiplets consisting of pairs of Higgs doublets with opposite hypercharge. Therefore, the MSSM contains two scalar Higgs doublets, H_U and H_D, with hypercharges +1 and -1, respectively. The neutral components of H_U and H_D have vacuum expectations values v_D and v_U, which are normalised so that:

$$v^2 = v_U^2 + v_D^2 \simeq (246\text{ GeV})^2 \tag{2.73}$$

Their ratio is defined as:

$$\tan\beta \equiv \frac{v_U}{v_D} \tag{2.74}$$

The two scalar Higgs doublets correspond to eight degrees of freedom, three of which give mass to the vector bosons, while the remaining five create physical scalar states: the CP-even neutral Higgs bosons h (light) and H (heavy), one of which must be the SM higgs boson, the CP-odd neutral Higgs A and two charged Higgs states $H^{\pm}$. The Higgs sector at tree level is fully described by the Z boson mass m_Z and two free parameters, $\tan\beta$ and one Higgs boson mass, conventionally chosen to be the CP-odd Higgs boson mass, m_A. The other tree-level Higgs boson masses are then given in terms of these parameters. The couplings of the supersymmetric Higgs bosons are different for up-type and down-type fermions and can be expressed relatively to the SM couplings. Their normalised values are shown in Table 2.1. The angle α is the mixing angle that diagonalises the CP-even Higgs squared-mass matrix.

Higgs field	Up-type fermions	Down-type fermion	Gauge bosons
h	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$	$\sin(\beta-\alpha)$
H	$\sin\alpha/\sin\beta$	$-\cos\alpha/\cos\beta$	$\cos(\beta-\alpha)$
A	$1/\tan\beta$	$\tan\beta$	0

Table 2.1: MSSM Higgs couplings to fermions and gauge bosons normalised to the SM values [12].

The MSSM has multiples regimes based on values of the parameters $\tan\beta$ and m_A, with very different Higgs bosons couplings properties [13]. However, most can be ruled out by phenomenological constraints and the main two that are of interest are the decoupling regime and the SUSY regime. In the decoupling regime, the h boson has SM-like properties, while the heavier Higgs states decouple from the gauge bosons. This happens for $m_A \gg m_Z$ (several hundred GeV) or at lower values of m_A for large $\tan\beta$. In the SUSY regime, the impact of light SUSY particles is big. In particular, one-loop corrections from strongly-interacting SUSY particles result in a Δb correction in the $b\bar{b}$ coupling [14]. The resulting coupling can then be expressed as:

$$g_{hbb} \approx g_{Abb} \approx \tan\beta(1-\Delta b) \tag{2.75}$$

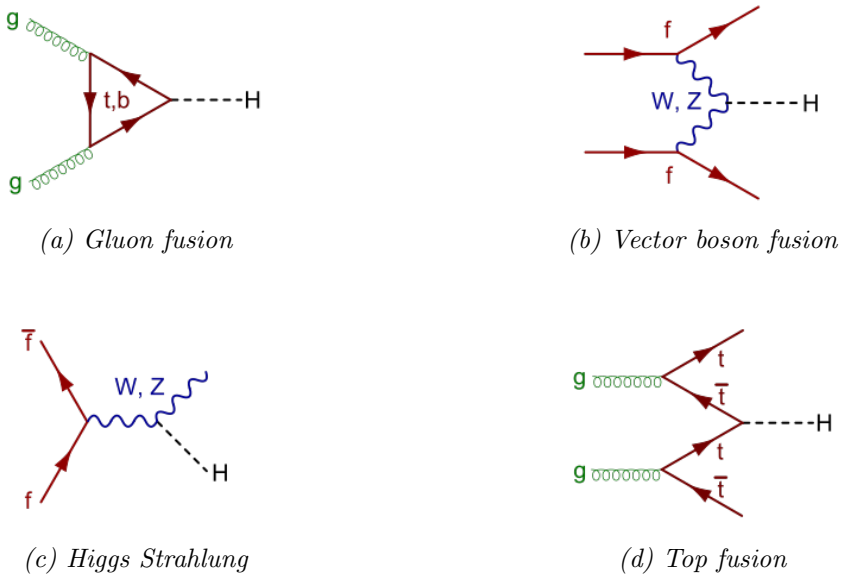
Chapter 3

Phenomenological Aspects

This chapter introduces some aspects of the phenomenology of the Higgs boson, which are useful for understanding the goals of the present study. It demonstrates the experimental profile of the Higgs boson, mentioning the most important developments and briefly presenting some results related to BSM physics.

3.1 Production modes

The Standard Model predicts that Higgs boson could be produced in a number of ways, as it couples to all massive particles, the most common of which are described in the following part. The corresponding Feynman diagrams are depicted in Figure 3.1, while their cross section is shown in Figure 3.2.



(a) Gluon fusion *(b) Vector boson fusion*

(c) Higgs Strahlung *(d) Top fusion*

Figure 3.1: Main Leading Order Feynman diagrams contributing to the Higgs production

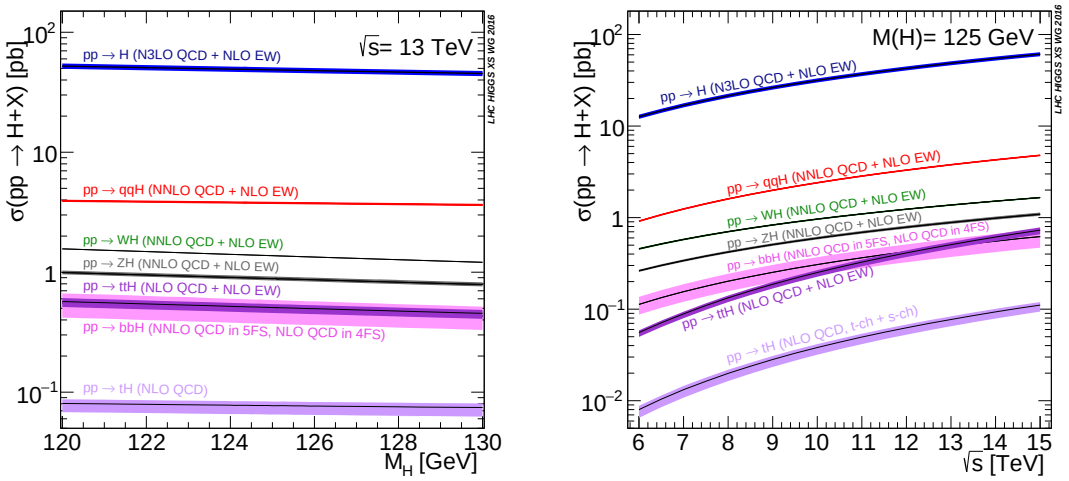
Gluon fusion: This is the dominant process at the LHC, being about ten times more likely than any of the other processes, as it involves the collision of two gluons binding a

hadron together, such as the proton. The two gluons combine to form a loop of virtual quarks, because the Higgs boson does not couple to the massless gluons. Since the coupling to the Higgs boson is proportional to the mass, this process is more likely for bottom and top quarks. When the Higgs boson is produced though this mechanism, there are no additional particles expected in the event, except for the Higgs boson decay products, so the identification of its final states is made more difficult due to QCD radiation from the colour field.

Vector boson fusion: The VBF production mode is the second leading one and occurs when two fermions (not necessarily of the same type) exchange a virtual W or Z boson, which in turn couples to the Higgs boson. This process is also important, because it results in easily identifiable final states, consisting of just the decay products of the Higgs boson and two forward jets from the break-up of the colliding protons.

Higgs Strahlung: During the collision of a fermion with an antifermion, they can merge to form a virtual W or Z boson, which if it carries enough energy, can radiate a Higgs boson. Since a proton is made of quarks and gluons, the LHC has a much lower probability of creating a Higgs boson through this channel.

Top fusion: A small contribution to the total Higgs production cross-section comes from two colliding gluons, each of which decay into a heavy quark-antiquark pair. A quark and antiquark from each pair can then combine to form a Higgs particle. This mode is a unique way to understand the direct coupling to the top quark, but it is challenging to tackle experimentally due to the high $t\bar{t}$ QCD production at the LHC.



(a) *(b)*

Figure 3.2: (a) Higgs boson production cross sections at $E_{cm} = 13$ TeV as a function of Higgs boson mass (b) Higgs boson production cross sections as a function of the centre of mass energies [15].

For the present study, gluon fusion and Higgs Strahlung are of interest, since the method for determining the Higgs branching ratios could be applied to both production modes in order to draw a comparison between the two. On the one side, the hadronic Higgs production has the advantage of increased statistics, but at the same time there is an overwhelming source of background from the QCD direct production of jets (QCD background). On the other side, the production in association with a Z or W boson is the

most sensitive production mode, as it provides a clean experimental signature due to the presence of leptons from the decay of the vector boson.

3.2 Decay modes

The Higgs boson can theoretically decay to most Standard Model particles, however the likelihood of each decay depends on the mass of the particle due to the proportionality of the coupling to the mass. A Higgs boson with a mass of 125 GeV/c^2 has a lifetime of approximately $1.6 \cdot 10^{-22}$ s, so the products of its decay are crucial in order to infer its presence in a particle physics experiment. The main quantity used for studying the decays of a particle is the decay rate Γ, which is essentially the probability per unit time that a particle will decay. If multiple decay modes are available, a decay rate Γ_i can be associated with each mode and the total rate is the sum of the individual ones:

$$\Gamma_{total} = \sum_{i=1}^{n} \Gamma_i \tag{3.1}$$

When the mass of an elementary particle is measured, the total rate shows up as the irreducible width of the shape of the distribution, which is why the decay rate is also referred to as decay width. The particle's lifetime is related to Γ_{total} as: $\tau = \frac{1}{\Gamma_{total}}$. The probability of the decay by a specific mode is expressed with the branching ratio. It is defined as:

$$BR_i = \frac{\Gamma_i}{\Gamma_{total}} \tag{3.2}$$

For the Higgs boson, the values of the branching ratios for the different decay modes are shown in Figure 3.3 and in Table 3.1. The branching ratios to hadronic products ($b\bar{b}$, $c\bar{c}$, gg) studied in this thesis, represent almost the 70% of the total Higgs decays.

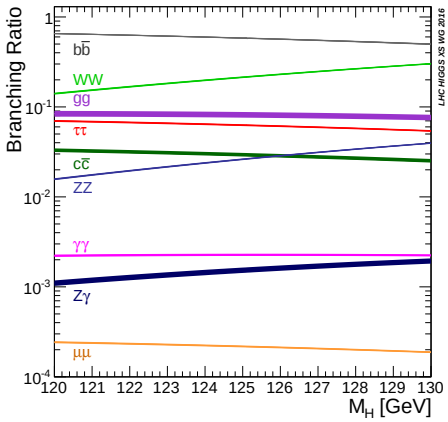


Figure 3.3: Branching ratios for various decay modes of a SM Higgs boson as a function of the Higgs mass [15].

Applying the Feynman rules for the decay of a Higgs boson to a pair of fermion-antifermion, the partial decay width can be computed as:

$$\Gamma\left(H \to q\bar{q}\right) = N_c \frac{m_f^2 m_H}{8\pi v^2} \tag{3.3}$$

Decay Mode	Branching ratio
$H \to b\bar{b}$	57.9%
$H \to WW^*$	21.7%
$H \to gg$	8.2%
$H \to \tau^+\tau^-$	6.2%
$H \to c\bar{c}$	2.9%
$H \to ZZ^*$	2.7%
$H \to \gamma\gamma$	0.2%
$H \to Z\gamma$	0.2%

Table 3.1: Branching ratios of the Higgs boson for $m_H = 125$ GeV [15].

where the factor N_c accounts for the number of colours; it's value is 1 for leptons and 3 for quarks [5]. From the above formula, it is evident that the decay rate is proportional to the fermion mass, that is why the Higgs boson couples strongly to b-quarks. The term $\frac{m_f}{v}$, as can be seen from Equation 2.72, is related to the coupling to the Higgs field and is the fundamental physical quantity to measure. Therefore, a measurement of the branching ratios is directly related to the determination of the fermion couplings.

The method for determining the branching ratios to hadronic final states presented in this study, is also applied to the Z signal region, so its decay is mentioned here for completeness. Z bosons decay into a fermion and an antifermion: approximately 20% for neutrinos, 10% for leptons and 70% for hadrons. The hadronic branching ratios are shown in the following table.

Decay Mode	Hadronic branching ratio (69.91%)
$(u\bar{u} + c\bar{c})/2$	11.6%
$(d\bar{d} + s\bar{s} + b\bar{b})/3$	15.6%
$c\bar{c}$	12.03%
$b\bar{b}$	15.12%

Table 3.2: Hadronic branching ratios of Z boson [16].

3.3 Higgs experimental profile

Following the observation of the Higgs boson [17, 18], a detailed exploration of its properties has taken place at the LHC at the different runs at $\sqrt{s} = 7$, 8 and 13 TeV. The Run 2 dataset at $\sqrt{s} = 13$ TeV corresponds to an integrated luminosity of approximately 156 fb^{-1}. A thorough description of the status of the Higgs boson physics can be found in [19], while this section presents only the results relevant to the topic of the study.

The most important decay channels at the LHC are: $H \to \gamma\gamma$, ZZ^*, WW^*, $\tau\tau$, $b\bar{b}$. For a given m_H, there are multiple factors that influence the sensitivity of a channel such as the production cross section, the decay branching ratio, the reconstructed mass resolution, the selection efficiency and the background in the final state. The $H \to ZZ^* \to 4l$ and $H \to \gamma\gamma$ channels have the best reconstructed mass resolution. Although the former

decay channel has a small branching ratio (Table 3.1), the final state topology has a clean experimental signature and the mass peak can be precisely reconstructed. The $H \to WW^* \to l\nu_l l'\bar{\nu}_{l'}$ has a significant branching ratio but the presence of neutrinos, which are not reconstructed in the final state, results in a poor mass resolution. Finally, the $H \to b\bar{b}$ and the $H \to \tau\tau$ channels suffer from large backgrounds, which makes the reconstruction difficult.

The Higgs mass is a free parameter in the SM so its measurement is important. During Run 1, the ATLAS and CMS experiments independently measured the Higgs boson mass, using samples of proton-proton collision data collected in 2011 and 2012, which correspond to 5 fb^{-1} of integrated luminosity at $\sqrt{s} = 7$ TeV and 20 fb^{-1} at $\sqrt{s} = 8$ TeV for each experiment. Based on the combined samples of the two, the mass was measured to be 125.09 ± 0.24 GeV [20]. ATLAS and CMS have been improving the precision of the mass measurements during the last years, with the latest results being 124.97 ± 0.24 GeV [21] and 125.38 ± 0.14 GeV [22], respectively. ATLAS measurements in the $\gamma\gamma$ and the $4l$ channels with 36.1 fb^{-1} of collision data and their combination are shown in Figure 3.4. The CMS result has the best precision so far and was achieved by similarly measuring in the same two channels and combining with Run 1 data.

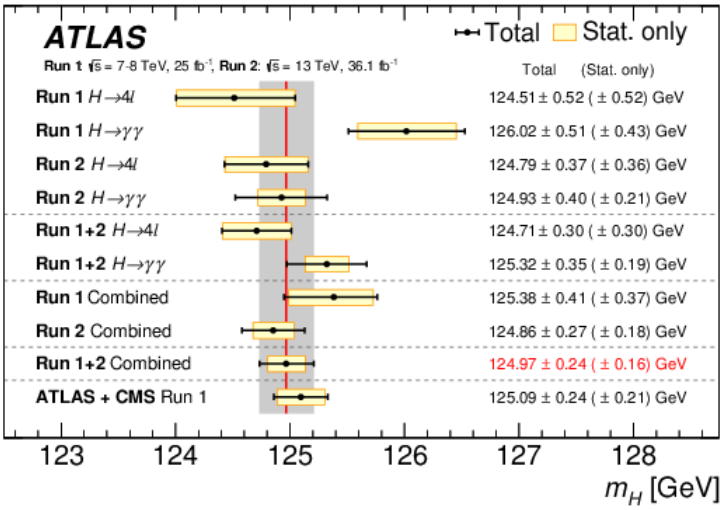
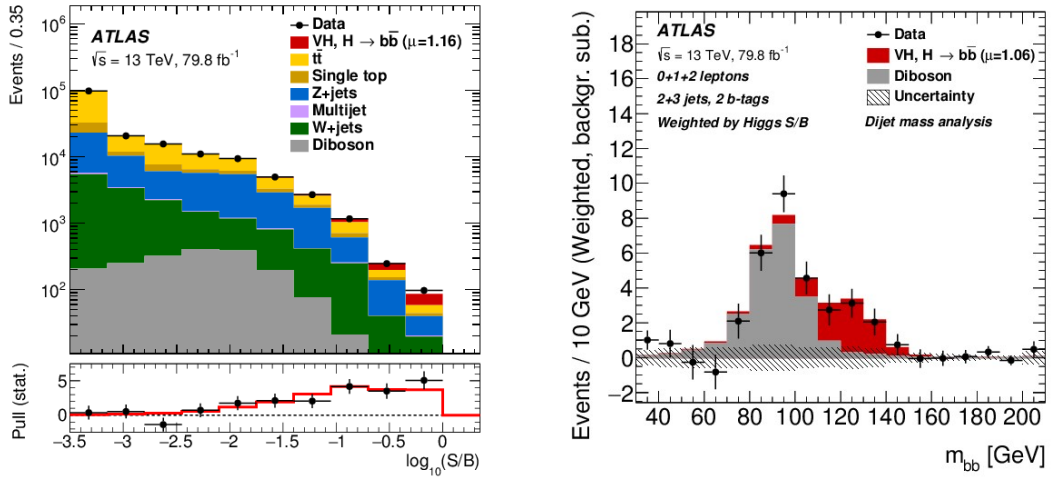


Figure 3.4: Summary of the ATLAS Higgs boson mass measurements from individual and combined analyses. The red vertical line and grey shaded column indicate the central value and the total uncertainty of the combined ATLAS Run 1 + 2 measurement, respectively [21].

A significant development is the observation of the $H \to b\bar{b}$ decay, which was only recently achieved [23]. Although it is the favoured decay of the Higgs boson with the corresponding branching ratio being almost 58%, it hadn't been previously observed. A search for it in the most copious production process for the Higgs boson, the gluon fusion, was very challenging due to the overwhelming QCD background. That is why less dominant production processes, such as the associated production with a vector boson, had to be considered, as they have features not present in the QCD background. The leptonic decays of the vector bosons make possible the reduction of the multi-jet background. In the analysis, events are selected in 0-, 1- and 2-lepton channels, based on the

number of charged leptons l (electrons or muons), to explore the $ZH \to \nu\nu b\bar{b}$, $WH \to l\nu b\bar{b}$ and $ZH \to llb\bar{b}$ signatures, respectively. The dominant background processes after the event selection are V + jets, $t\bar{t}$, single top quark production and diboson process. The Higgs boson candidate is formed by requiring events to have exactly two b-tagged jets in all channels and the main discriminant variables are the invariant mass m_{bb}, the reconstructed transverse momentum of the vector boson and the jet cone size which are combined in a Boosted Decision Tree. The results are shown in the following figures.



(a) *(b)*

Figure 3.5: (a) Event yields as a function of $\log(S/B)$ *for data, background and a Higgs boson signal with* $m_H = 125$ *GeV, with S being the fitted signal and B the fitted background yields. The lower panel shows the pull of the data relative to the background. (b) The distribution of* m_{bb} *in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes [23].*

Finally, it is also relevant to present the latest results for the Higgs boson production cross sections and branching ratios, which can be extracted from combined measurements [24]. The combination is based on analyses investigating the decay modes $H \to \gamma\gamma$, ZZ^*, WW^*, $\tau\tau$, $b\bar{b}$, $\mu\mu$ and uses 79.8 fb^{-1} of collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector. The parameter of interest is the production cross section times branching ratio $(\sigma \cdot BR)_{if}$ for a specific mode i and a decay with final state f. This is because the two quantities cannot be treated independently, as each observed process involves at least two Higgs boson coupling strengths. This parameter is then normalised to the SM expectation values:

$$\mu_{if} = \frac{\sigma_i}{\sigma_i^{SM}} \cdot \frac{BR_f}{BR_f^{SM}} \tag{3.4}$$

and represents a specific signal yield. The results are presented in Figure 3.6. The combination of the various μ_{if} gives a global signal strength of $\mu = 1.11^{+0.09}_{-0.08}$. Modifications of the Higgs boson couplings related to BSM physics can be studied using the κ framework. It describes the Higgs coupling properties in terms of coupling strength modifiers κ_i which are defined as:

$$\kappa_i^2 = \frac{\sigma_i}{\sigma_i^{SM}} \quad \text{or} \quad \kappa_f^2 = \frac{\Gamma_f}{\Gamma_f^{SM}} \tag{3.5}$$

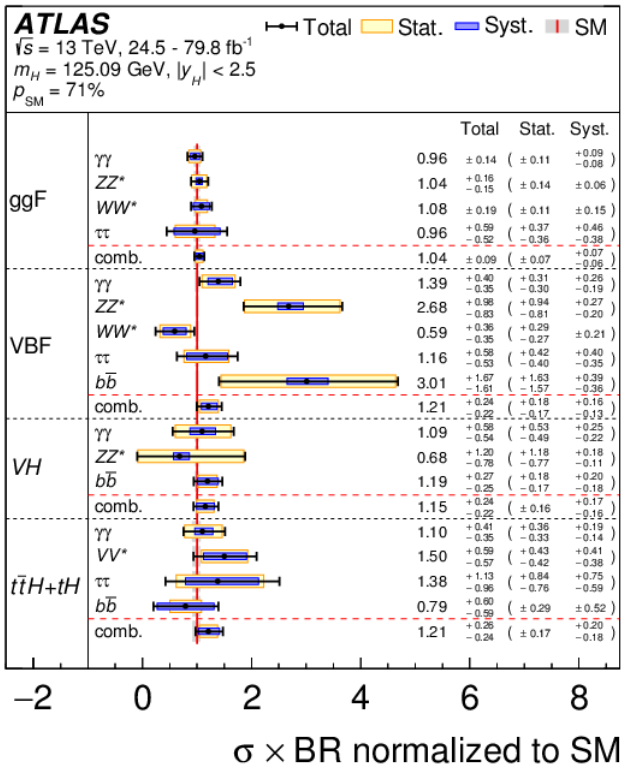


Figure 3.6: Cross sections times branching ratio for the main production processes in each relevant decay mode, normalised to their SM predictions [24].

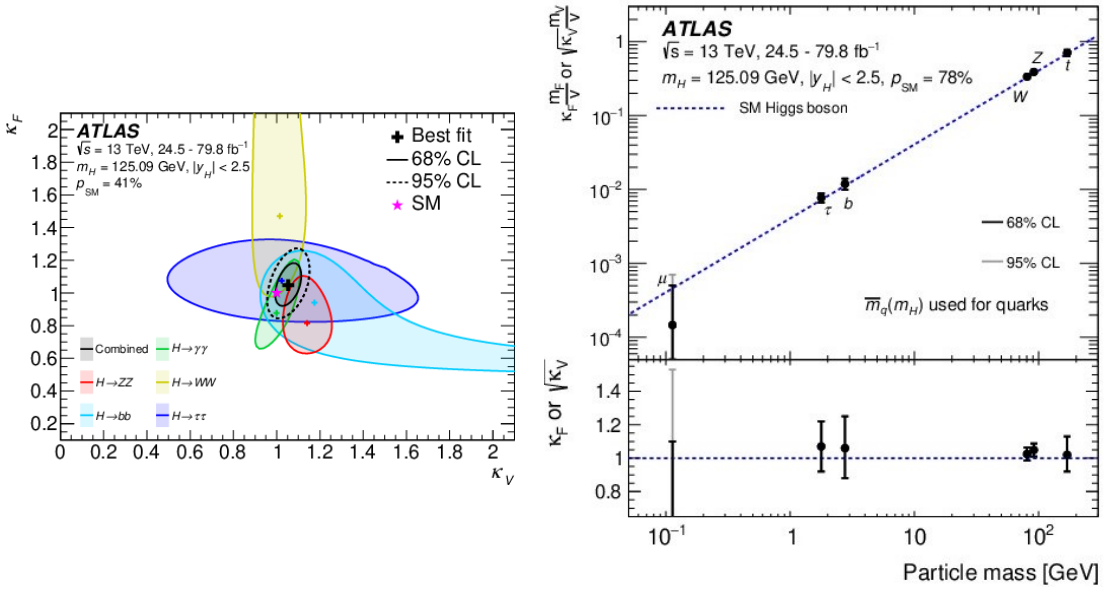
They are related to the product of the cross section and the branching ratio by the equation:

$$(\sigma \cdot BR)_{if} = \frac{\sigma_i^{SM}\kappa_i^2 \cdot \Gamma_f^{SM}\kappa_f^2}{\Gamma_H^{SM}\kappa_H^2} \Rightarrow \mu_{if} = \frac{\kappa_i^2 \cdot \kappa_f^2}{\kappa_H^2} \tag{3.6}$$

where κ_H^2 adjusts the SM Higgs width to take into account the modifications induced by the deformed Higgs boson couplings. With all the modifiers set to 1, the SM is reproduced. A model with universal coupling strength scale factors $\kappa_V = \kappa_W = \kappa_Z$ and $\kappa_F = \kappa_t = \kappa_b = \kappa_\tau = \kappa_\mu$ is probed. A fit can be performed either globally for these two parameters or separately for each of the five major decay channels, yielding ten parameters: κ_V^f and κ_F^f with f indicating the decay mode. The best-fit values and uncertainties from a combined fit are $\kappa_V^f = 1.05 \pm 0.04$ and $\kappa_F^f = 1.05 \pm 0.09$. The results of the fit are shown in Figure 3.7a. In a different model, the scale factors are treated independently and are used for defining the reduced coupling strength scale factors for the weak bosons and the fermions as:

$$y_V = \sqrt{\kappa_V}\frac{m_V}{\upsilon} \quad \text{or} \quad y_F = \kappa_F\frac{m_F}{\upsilon} \tag{3.7}$$

In Figure 3.7b they are plotted against the particle mass. Taking into account all the aforementioned results, it can be concluded that no significant deviations from Standard Model predictions are observed.



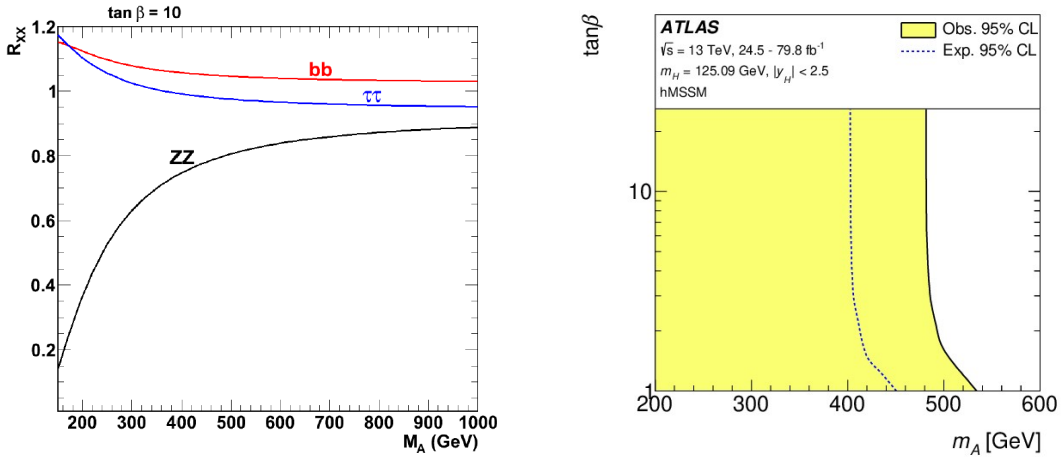
(a) *(b)*

Figure 3.7: (a) Negative log-likelihood contours in the (κ_V^f, κ_F^f) plane for the individual decay modes and their combination. (b) Reduced coupling strength modifiers for fermions and for weak gauge bosons as a function of their masses. The lower inset shows the ratios of the values to their SM predictions [24].

3.4 Branching ratios and BSM physics

As explained in Section 2.3, the MSSM involves different couplings of the Higgs boson to the elementary particles and particularly important are those concerning the $b\bar{b}$ decay (Table 2.1, Equation 2.75). On the one hand, the decay modes $H, A \to b\bar{b}, \tau^+, \tau^-$ dominate when $\tan\beta$ is large. On the contrary, the correction Δb can obtain significant values in some scenarios leading to a significant reduction of the $h_{b\bar{b}}$ coupling [25]. The ratios of the h branching ratios to their SM values $R_{XX} \equiv BR(h \to XX)/BR(H_{SM} \to XX)$ as a function of m_A are shown in Figure 3.8a. The modified couplings could evidently manifest in a measurement of the Higgs branching ratios. Taking into consideration Equations 3.1 and 3.2, the large decay width of $H \to b\bar{b}$ could lead to a propagation of its coupling deviation to all other branching ratios by affecting the total decay width. Therefore, measuring different branching ratios than the theoretically predicted ones could be attributed to non-SM couplings and unveil SUSY scenarios.

The analyses of the Higgs decay modes can be used to set exclusion limits on the parameters of the MSSM. The latest results extracted from [24] show the values of m_A and $\tan\beta$ excluded by fits to the measured rates of Higgs boson production and decays (Figure 3.8b). The decoupling limit, in which all Higgs boson couplings tend to their SM value, corresponds to $m_A \to \infty$. The observed (expected) lower limit at 95% CL on the CP-odd Higgs boson mass is at least $m_A > 480$ GeV ($m_A > 400$ GeV) for $1 \leq \tan\beta \leq 25$ and increases to $m_A > 530$ GeV ($m_A > 450$ GeV) at $\tan\beta = 1$. The results of Section 3.3 have measured accuracies that still permit the deviations predicted for many BSM scenarios.



(a) *(b)*

Figure 3.8: (a) The h branching ratios normalised to their SM value as a function of m_A for $\tan\beta = 10$ [26].(b) Likelihood contours in the m_A-$\tan\beta$ plane for both the data and the expectation of the SM Higgs sector. The regions to the left of the solid contour are excluded [24].

Chapter 4

Experimental Framework

In this chapter, the main physics objects used in the study, jets and leptons, are presented. The chapter aims to review the characteristics of each and explain how they can be reconstructed. A brief overview of the ATLAS detector is also given, so as to underline the role of the various systems in the reconstruction process. Finally, the chapter explains the flavour tagging process which is central for the present thesis.

4.1 The ATLAS detector

The coordinate system used by ATLAS [27] needs to be specified as it is used in the following parts. The origin of the right-handed coordinate system is at the interaction point (IP) in the center of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, with ϕ being the azimuthal angle around the z-axis. Pseudorapidity η is used to describe the angle of a particle relative to the beam axis and is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$, while the angular separation ΔR is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. The transverse momentum p_T is the component of the momentum perpendicular to the beam axis: $p_T = p\cos\theta$.

ATLAS is a general-purpose particle detector covering nearly the entire solid angle around the collision point. The magnet configuration [28] consists of a thin superconducting solenoid surrounding the inner detector cavity, providing a 2T axial magnetic field and three large superconducting toroids (one barrel and two end-caps) around the calorimeters.

The Inner Detector (ID) [29] is the first part of ATLAS to see the decay products of the collisions, so it is very compact and highly sensitive. It is used to reconstruct the particle tracks in the region of acceptance $|\eta| < 2.5$ and consists of discrete, high-resolution semiconductor pixel and strip detectors in its inner part and straw-tube tracking detectors in its outer part. The latter also provide measurements of transition radiation that are used in electron identification. Several other parameters can be determined, such as the p_T which can be measured from the curvature of the tracks and the sign of the charge of the particle, which can be inferred from the direction of the curvature. Finally, it is possible to precisely determine the origin of the vertex from which the particles originate, by determining the transverse and longitudinal impact parameter of tracks. The transverse impact parameter d_0 is defined as the shortest distance between a track and the beam line in the transverse plane, while the longitudinal impact parameter z_0 is defined as the

distance in z between the primary vertex and the point on the track used to evaluate d_0.

A calorimeter system surrounds the inner tracking detector, covering the pseudorapidity range $|\eta| < 4.9$. Calorimeters are based on the principle of electromagnetic showers, which occur when a high-energy electron radiates a bremsstrahlung photon as it interacts in a medium. The photon in turn, produces an e^+e^- pair. This process produces a cascade of photons, electrons and positrons and continues until the average energy of the particles falls below a critical energy. The produced scintillation light can be collected and is proportional to the total energy of the original electron/photon. They are designed to absorb most of the particles (except muons and neutrinos) coming from a collision, by forcing them to deposit all their energy within the detector. They typically consist of layers of a "passive" high-density material like lead, interleaved with layers of an "active" medium, such as liquid argon. ATLAS uses two types of calorimeters: a high-granularity liquid argon (LAr) electromagnetic calorimeter [30] and a scintillator-tile hadronic calorimeter [31]. The latter is based on the strong interaction between both charged and neutral hadrons with atomic nuclei. Hadronic showers occupy a significant volume in the detector due to the relatively large distance between nuclear interactions.

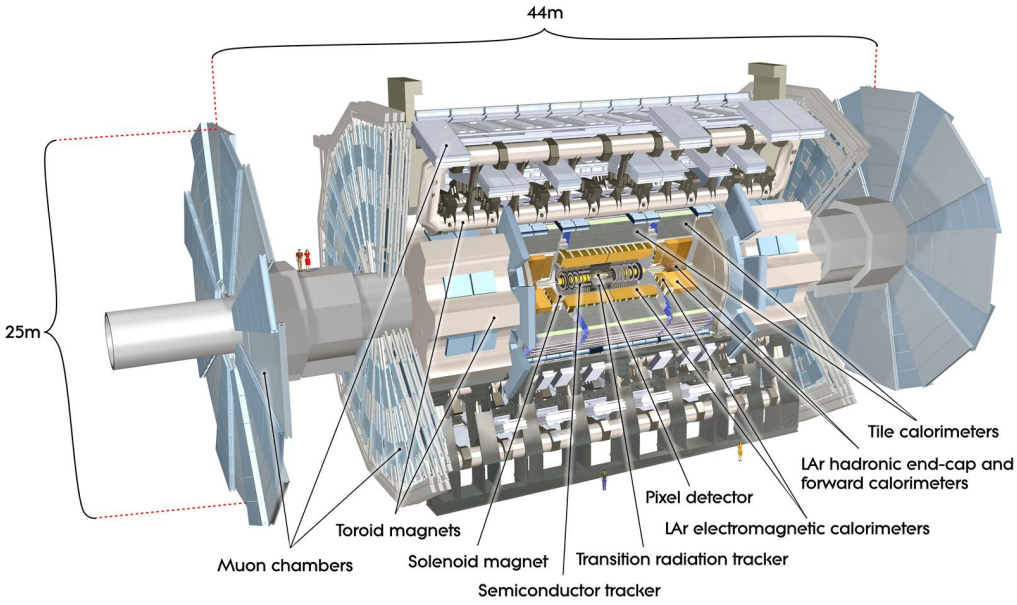


Figure 4.1: Cut-away view of the ATLAS detector showing the various sub-detectors [27].

The calorimeter is surrounded by the muon spectrometer (MS) [32]. Due to their mass and lack of strong interaction, muons usually pass through the ID and the calorimeters undetected, so it is necessary to have a dedicated system for measuring their energy and momentum. The muon spectrometer measures the deflection of muons with $|\eta| < 2.7$ using multiple layers of high-precision tracking chambers located in a toroidal field of approximately 0.5T or 1T in the central and end-cap regions of ATLAS, respectively. In the barrel region, the chambers are arranged in three cylindrical layers around the beam axis, while in the transition and end-cap regions, the chambers are installed in planes perpendicular to the beam, also in three layers. The muon spectrometer defines the overall dimensions of the ATLAS detector.

The sub-detectors can be seen in Figure 4.1 and the signatures of each particle type in Figure 4.2.

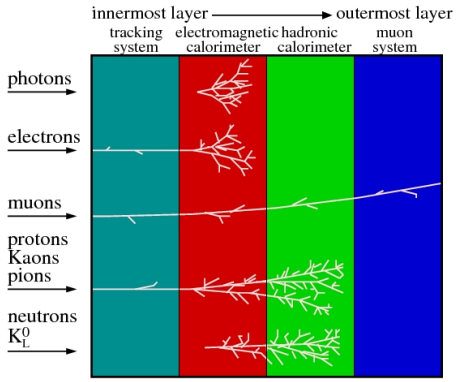


Figure 4.2: Signatures of the different particle types in the various detector systems [33].

4.2 Reconstruction

An essential part of the analysis of ATLAS data is the reconstruction of physics objects in the events, based on their experimental signature in the detector. There are multiple steps involved in the reconstruction of different objects, which are described in the following part.

Tracks: The first step in the reconstruction process concerns the tracks of the particles. Tracks can be reconstructed by measuring localised ionisation signals (hits) in the ID released by electrically charged particles [34]. By combining the individual hits on different layers, it is hence possible to identify tracks that correspond to the trajectories of the particles. The reconstructed tracks can be used to reconstruct interaction vertices, of which, the one with the highest sum of squared transverse momenta of associated tracks is selected as the primary vertex.

Jets: In the case of the hadronic Higgs production, the decay products are jets. Jets refer to collimated hadrons and other particles produced by the hadronization of a quark or gluon in collider experiments. Due to colour confinement, which allows only for colorless states, quarks or gluons produced by high-energy collisions cannot exist individually. They combine with quarks and antiquarks spontaneously created from the vacuum, to form hadrons. Most of the quark's initial energy is carried in the motion-energy of the new hadrons, so the total energy and direction of the jet is quite similar to the energy and direction of the initial quark. As a result, jets can function as proxies to the high energy quarks and gluons produced in a collision. Identifying and measuring them allows the reconstruction of the kinematics of the elementary QCD interactions. It is an essential step, since jets are involved in many different physics processes. Jets can be identified by reconstructing the full hadronic $q\bar{q}$ system as a single large-radius (large-R) jet and then flavour tagged, using variable radius track sub-jets (Section 4.3).

- **Large Radius Jets:** The hadronic decay products can be clustered within a single large-R calorimeter jet by reconstructing topological clusters of calorimeter cells [35], in an attempt to extract the significant signal from a background of electronic noise

and other sources of fluctuations, such as pile-up. These form the set of constituents from which large-R calorimeter jets are reconstructed, using the anti-k_t algorithm [36] with a radius parameter of R = 1.0. The algorithm involves forming a collection of jets through an iterative recombination of the input clusters. The characteristic shape of the output jets is a cone with angular width driven by the free parameter R.

- **Variable Radius Track Jets:** The distribution of the energy inside a jet, which is referred as jet substructure, contains crucial information for discriminating Higgs jets from multijet background jets. As the transverse momentum of jets increases, the decay products become more collimated and the reconstruction of jet substructure observables becomes more difficult due to the finite angular resolution of the ATLAS calorimeter. However, this can be overcome by using tracks reconstructed in the inner detector, which provides a good angular resolution for charged jet components. Jets based on tracks (referred to as track jets) are reconstructed using the variable radius jet algorithm [37], which lets the cone size of a jet vary as $\Delta R \propto \frac{1}{p_T}$. The variable ΔR corresponds to a fixed angular size of a jet cone, which correctly models the shower of hadrons coming from two partons.

Leptons: For the case of the associated production of a Higgs boson and a W or Z boson, the leptonic decay of the vector boson can be used to disentangle the signal from the background by reconstructing the leptons involved. Electrons are reconstructed from topological clusters of energy deposits in the calorimeter, since their characteristic experimental signature is the electromagnetic shower. They can then be matched to a track in the ID. Muons are feebly interacting in the calorimeters, while having quite a long lifetime. They appear as charged particle tracks in the ID and MS, with a finite curvature caused by the magnetic field of the solenoid and toroid systems. Their selection procedure begins with two independent track fits in the ID and MS, that are then combined using different algorithms depending on the signature in each sub-system. Finally, due to their high mass, the taus have a very short lifetime, which results in them typically decaying before reaching the ID. The tau is the only lepton that can decay into hadrons, so through the weak interaction it can decay to either purely leptonic states or hadronic (approximately 65% of the time). In the former case, electrons and muons are reconstructed as described previously. while in the latter, the reconstruction makes use of the hadronic calorimeter.

An event with only leptonic products recorded by the ATLAS detector, is shown in Figure 4.3. It corresponds to the Higgs boson production in the four leptons decay channel. The two leading electrons are represented by green tracks and green EM calorimeter deposits, while the subleading muons are indicated by two red tracks. Recoiling against the four lepton candidate in the left hemisphere, is a dimuon pair in the right hemisphere indicated by the red tracks.

4.3 Flavour tagging

Jets containing b-hadrons (b-jets) and to a much lesser extent c-quarks have some distinct features, allowing them to be distinguished from other jets. First, hadrons with b-quarks travel the ideal distance for detection by the LHC detectors. They have sufficient lifetime that they travel some distance (a few millimetres) before they decay, while at the

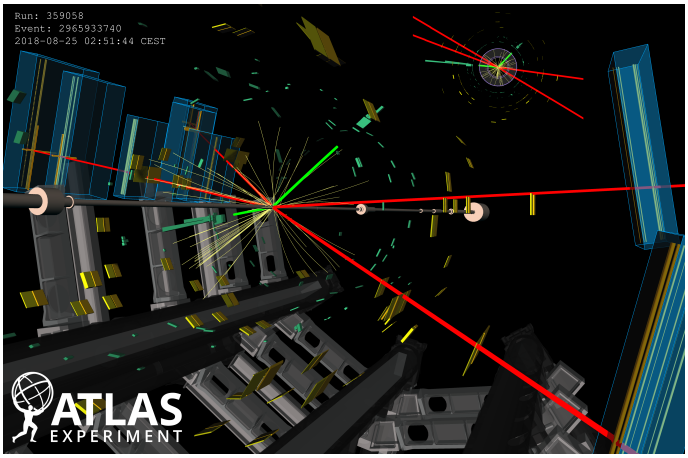


Figure 4.3: Event display of a 2e2μ ZH candidate [38].

same time their lifetimes are not as high as those of light quark hadrons, which escape the detector. This results in a significant number of hadrons in a b-jet coming from a different point than the collision point, which is called a secondary vertex and is the location where the bottom hadron decayed (Figure 4.4). Furthermore, due to the large mass of the b-quark, its decay products have higher transverse momentum, so b-jets are wider and have more constituents particles (multiplicities). Finally, b-jets can contain low-energy leptons due to direct and indirect semileptonic decays.

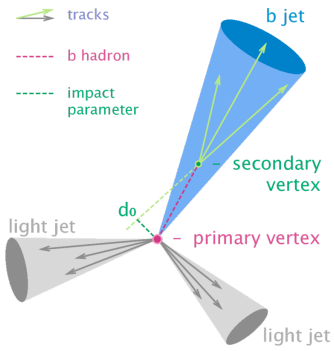
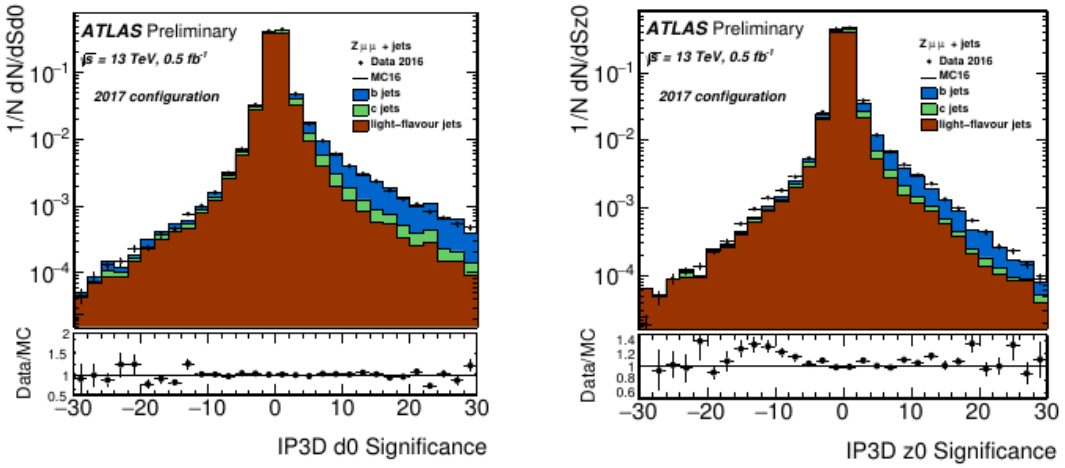


Figure 4.4: Comparison between the characteristic properties of light-flavour jets and b-jets. The trajectory of the b-hadron, the secondary vertex and the impact parameter d_0 of a track from its decay are highlighted.

The identification of b-jets is fundamental for studying a large number of processes, including the decays of the Higgs boson. In order to measure the Higgs branching ratios to hadronic final states, it is necessary to distinguish b-jets from c-jets and light-flavour jets. This can be achieved through flavour tagging, which is based on three distinctive features used by different b-tagging algorithms: impact-parameter based algorithms, an inclusive secondary vertex reconstruction algorithm and a decay chain multi-vertex reconstruction algorithm. The outputs of these algorithms are combined in multivariate discriminants (MV2) which provide the best separation among the different flavours hypotheses. Tracks used for b-tagging are reconstructed using the ID and associated to jets.

The IP2D/IP3D algorithms involve the use of the impact parameters of the tracks in a jet, as the main discriminating variables. The IP2D tagger uses the transverse impact parameter significance d_0/σ_{d_0}, whereas IP3D uses both the transverse and the longitudinal impact parameter significance $z_0 \sin\theta/\sigma_{z_0 \sin\theta}$ (Figure 4.5). In general, a larger impact parameter significance indicates a clear secondary vertex signal, which in turns represents the presence of a heavy jet. The probability distributions of the significances are employed to define ratios of b-jet and light-flavour jet hypotheses, which are then combined into a single discriminant. The secondary vertex finding algorithm reconstructs a single displaced secondary vertex in a jet. It uses as input all pairs of tracks associated to each jet, forming two-track vertices that are then discarded based on their compatibility with the decay of long-lived particles (K_s or Λ), photon conversions or hadronic interactions with the detector material. An inclusive secondary vertex is then reconstructed from the accepted tracks. Finally, the decay chain multi-vertex reconstruction algorithm, JetFitter, exploits the topological structure of weak b- and c-hadron decays inside the jet and tries to reconstruct all decay vertices, assuming they lie on the same flight path [39].



(a) *(b)*

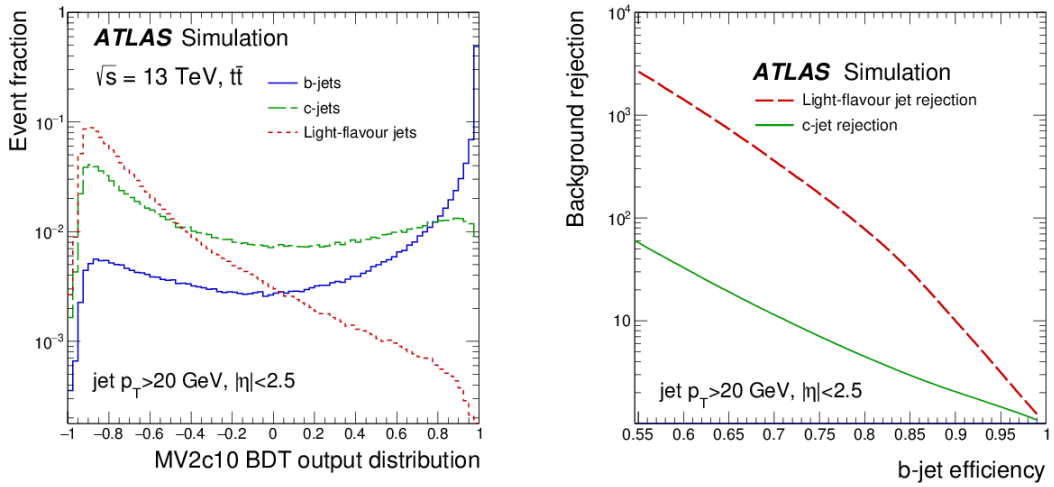
Figure 4.5: Data-MC comparisons of the transverse (a) and longitudinal (b) impact parameter significance values for IP3D selected tracks in the leading jet of the Z decay to muons plus jets dominated sample [39].

The three baseline algorithms mentioned above, contribute to b-jet discrimination by providing a number of weakly correlated variables. Their output is combined using a Boosted Decision Tree (BDT) and the algorithm is known as MV2c. It is trained using $t\bar{t}$ events while controlling the c- and light-flavour jet background. The analysis in this thesis employ the MV2c10 tagger in a configuration where the training sample contains 7% c-jets and 93% light-flavour jets. The efficiency is defined as:

$$\epsilon_j = \frac{Number\ of\ jets\ of\ flavour\ j\ passing\ cut}{Number\ of\ jets\ of\ flavour\ j} \tag{4.1}$$

while the rejection is defined as $1/\epsilon_j$. Regarding c-jets, they can be distinguished by taking advantage of the fact that charm hadron decays have a single secondary vertex with intermediate charged decay multiplicity and a decay distance comparable to that of

b-hadrons. The charm decay products can be separated from the b-hadron ones because of their smaller decay multiplicity, which makes the average energy per c-hadron larger than that of the b-quark decay products. The output of the MV2c10 algorithm and the light-flavour and c-jet rejection rates as a function of the b-jet efficiency are shown in Figure 4.6.



(a) *(b)*

Figure 4.6: (a) The MV2c10 output in simulated $t\bar{t}$ events. (b) Rejection factors as a function of the b-jet tagging efficiency of the MV2c10 algorithm [40].

An example of an event including jets recorded by the ATLAS detector is shown in Figure 4.7. The two central high p_T b-tagged jets are shown as green and yellow bars, corresponding to the energy deposition in the electromagnetic and hadronic calorimeters, respectively. The missing energy E_T^{Mss} is shown as a white dashed line. The excerpt shows three vertices and their associated tracks: the primary vertex (yellow) and the two secondary vertices (blue) compatible with the presence of a decay of a b-hadron.

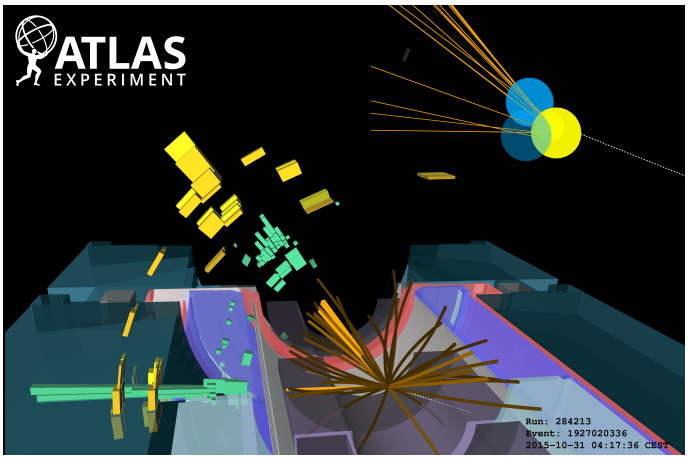


Figure 4.7: Candidate $H \to b\bar{b}$ decay event with two b-tagged jets and no lepton recorded in 2015 [41].

Chapter 5

Fitting Technique

This chapter presents the developed fitting technique for an inclusive measurement of the branching ratios. First, the technique is analysed in detail and then it is validated with various statistical tests in order to examine its accuracy.

5.1 Definition

The fit method used in this thesis is based on templates created using the output of the multivariate discriminant MV2 (Section 4.3). For each population of interest (gg, $c\bar{c}$ $b\bar{b}$), there is a corresponding template. These are then used to fit a sample which contains the MV2 outputs for every hadronic event and simultaneously extract the contribution of each population. This is essentially the branching ratio. The advantage of this method lies on the fact that it uses a pure sample without discriminating and therefore, has increased statistics. This wouldn't be possible if a cut on the flavour was imposed a priori, as is the case when taking only b-tagged events. The gg and $c\bar{c}$ contributions would be left out because they are difficult to tag while at the same time correspond to a non-negligible branching ratio.

The method is based on a least square fit. Specifically, TMinuit [42] is used to find the minimum value of a multi-parameter function (the "FCN"). The value of the FCN generally depends on one or more variable parameters, whose meaning is user defined but whose trial values are determined by MINUIT. By providing the initial parameter values and uncertainties and a set of instructions, MINUIT finds the values of the parameters for which the minimisation is achieved, along with their uncertainties. The FCN written for this analysis calculates the χ^2 between the templates of the three different populations and the data. It is built as:

$$\chi^2 = \left(\frac{h_d - N(fr_1 h_1 + fr_2 h_2 + fr_3 h_3)}{\sigma} \right)^2 \tag{5.1}$$

where h_d, h_1, h_2, h_3 are the values of the data and the three templates, respectively, fr_1, fr_2, fr_3 are the fitted fractions, N is a normalisation factor and σ is the data error. The variable parameters of the fit are fr_1, fr_2 and N. The normalisation factor is a free parameter, while the fractions are constraint between 0 and 1 as they correspond to branching ratios. The third fraction is given by the equation:

$$fr_1 + fr_2 + fr_3 = 1 \Rightarrow fr_3 = 1 - fr_1 - fr_2 \tag{5.2}$$

given that the hadronic branching ratios should add up to one. The fit results can be thought as representing the mean of a Gaussian distribution with standard deviation equal to the error. Given a functional relationship $Q = f(x, y)$ between several measured variables (x, y), the uncertainty in Q using error propagation is given as:

$$\sigma_Q^2 = \sigma_x^2 \left(\frac{\partial Q}{\partial x}\right)^2 + \sigma_y^2 \left(\frac{\partial Q}{\partial y}\right)^2 + 2\sigma_{xy} \frac{\partial Q}{\partial x} \frac{\partial Q}{\partial y} \tag{5.3}$$

where σ_x, σ_y are the standard deviations of x and y and σ_{xy} is their covariance. Applying the above relation to Equation 5.2 yields:

$$\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2 + 2\sigma_{12}} \tag{5.4}$$

where σ_1, σ_2 are the errors of the fitted fractions and σ_{12} is their covariance, which is also given from TMinuit and is usually different from zero.

5.2 Validation

Before applying the method to the cases of interest, it is necessary to check the goodness of the fit. The fit method is initially tested in a simplified process. First, three linear functions which can take only positive values are defined on the interval $[0, 1]$ and are used to randomly fill three histograms (Figure 5.1). These histograms are the templates used for the fitting and are normalised so as to have a total area equal to one. Then, a histogram representing the data to be fitted is created by taking the sum of the templates weighted by a fraction for each component. It is scaled with a constant corresponding to the total number of events (Figure 5.2). The bin content (bc) can be considered as counts following a Poisson distribution. Thus, the bin error is set to $\sqrt{bc}$ as this is a good estimate for the standard deviation of the distribution. The fit method is used to extract the fractions of the first and second template from the sum. The input fractions and the results of the fit for two cases are shown in Table 5.1. The third fraction and its error can be computed using Equations 5.2 and 5.4. The results show that the input values are fitted with great accuracy. In order to ensure that the fit is linear and the fitted values correspond correctly to the injected ones, five fits are performed and the results of the fit are plotted against the input. Indeed, as shown in Figure 5.3 the relationship is linear for all three fractions.

	case 1		case 2	
	Input	Fit results	Input	Fit results
N	-	10000 ± 100	-	10000 ± 100
fr_1	0.5	0.50000 ± 0.02477	0.12	0.120000 ± 0.02610
fr_2	0.2	0.20000 ± 0.07917	0.04	0.039998 ± 0.08177
fr_3	0.3	0.30000 ± 0.05882	0.84	0.840002 ± 0.05943
σ_{12}	-	-0.00171	-	-0.00192

Table 5.1: Input and extracted values for two cases

A useful method for testing parameter estimation techniques is checking the pull distribution. For this reason, a set of pseudo-data is constructed. This can be achieved by

"smearing" the weighted sum in order to replicate the statistical fluctuations (measurement errors) of the data and performing the fit multiple times. The pull distribution must be smoothed, that is why the content of every bin of the histogram is modified by adding a random value from a Gaussian distribution with parameters:

$$\mu = 0, \quad \sigma = \sqrt{bc} \tag{5.5}$$

The bin error is also ensured to be the square root of the bin content. The pull variable for a parameter p is defined as:

$$\frac{p_{estm} - p_{true}}{\sigma_p} \tag{5.6}$$

where p_{estm} and p_{true} are the estimated and the true value of the parameter, respectively and σ_p is the estimated error on the parameter. If the fit is correct, the distribution of the pull values is a standard normal distribution, i.e. a Gaussian distribution centred around zero with a standard deviation of one. In this case, for every fraction, the pull value is:

$$\frac{fr_i^f - fr_i^i}{\sigma_i} \tag{5.7}$$

with the superscripts f and i denoting the fitted value and the input. The fit is repeated 1000 times for input fractions 0.5, 0.2, 0.3. In every fit, the pull value is calculated and filled into a histogram. These histograms are then fitted with a Gaussian function, so as to extract the relevant parameters and check if they correspond to a standard normal distribution. The pull histograms along with the fitted functions are shown in Figure 5.4. As it is shown in Table 5.2, for all three fractions the parameters ensure that the statistical error of the fit method is correct.

	Pull fr_1	Pull fr_2	Pull fr_3
Constant	226.799 ± 8.950	226.897 ± 8.789	224.798 ± 8.651
Mean	0.1423 ± 0.0337	-0.0407 ± 0.0341	-0.0016 ± 0.0339
Sigma	1.0231 ± 0.0235	1.0411 ± 0.0233	1.0496 ± 0.0231

Table 5.2: Gaussian parameters corresponding to the fit performed on the pull distributions

Finally, it must be ensured that the error of the fitted fractions scales appropriately with the number of entries of the data histogram. Specifically, the fractional error, which is defined as:

$$\frac{\sigma_i}{fr_i} \tag{5.8}$$

should drop according to the square root of the number of entries. In order to check this, the fit is repeated seven times, for 1000, 5000, 10000, 20000, 30000, 40000 and 50000 entries. The fractional error is plotted against the number of entries in Figure 5.5 and it indeed decreases according to the expected way.

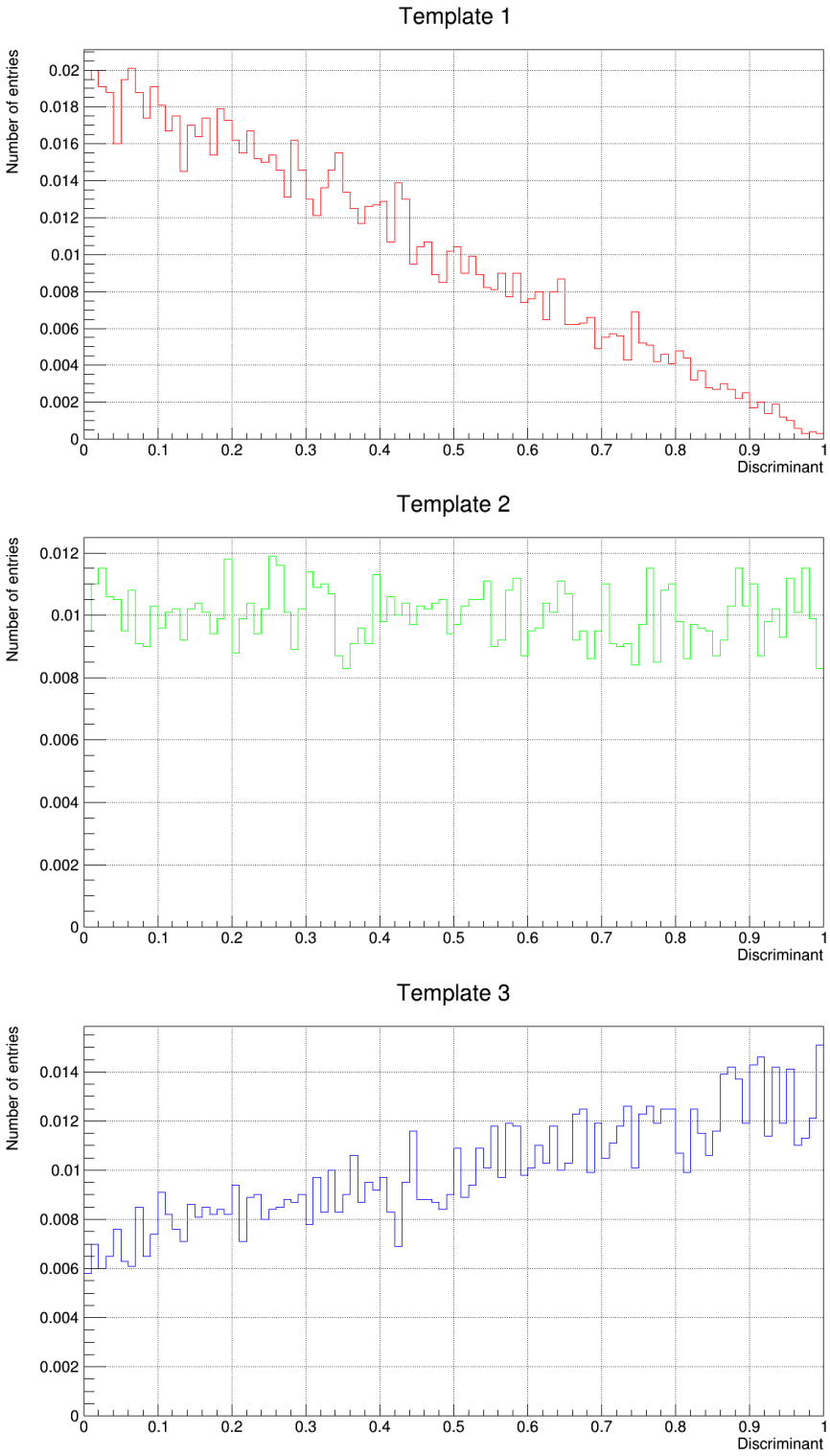
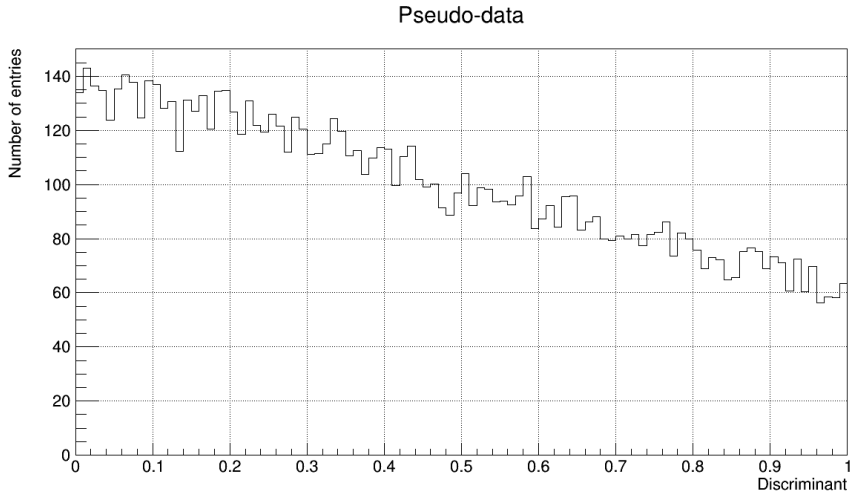
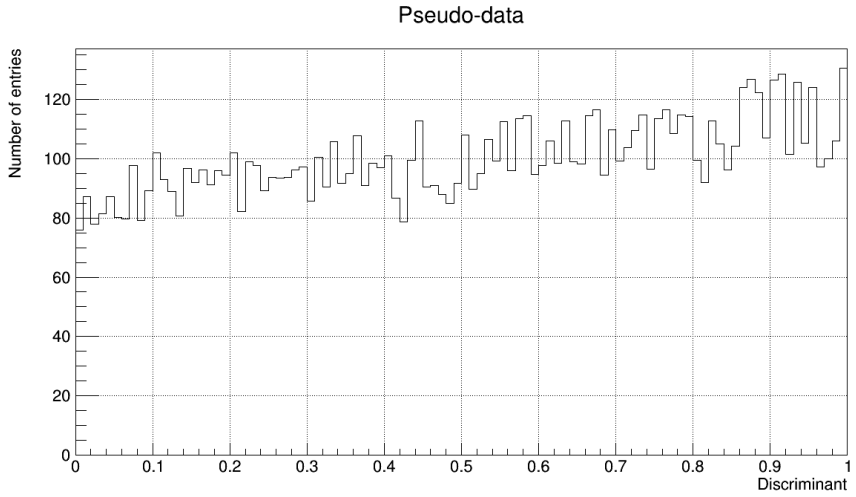


Figure 5.1: Templates based on linear functions



(a)



(b)

Figure 5.2: Pseudo-data created by taking the weighted sum of the three templates for fractions (a) 0.5, 0.2, 0.3 (b) 0.12, 0.04, 0.84

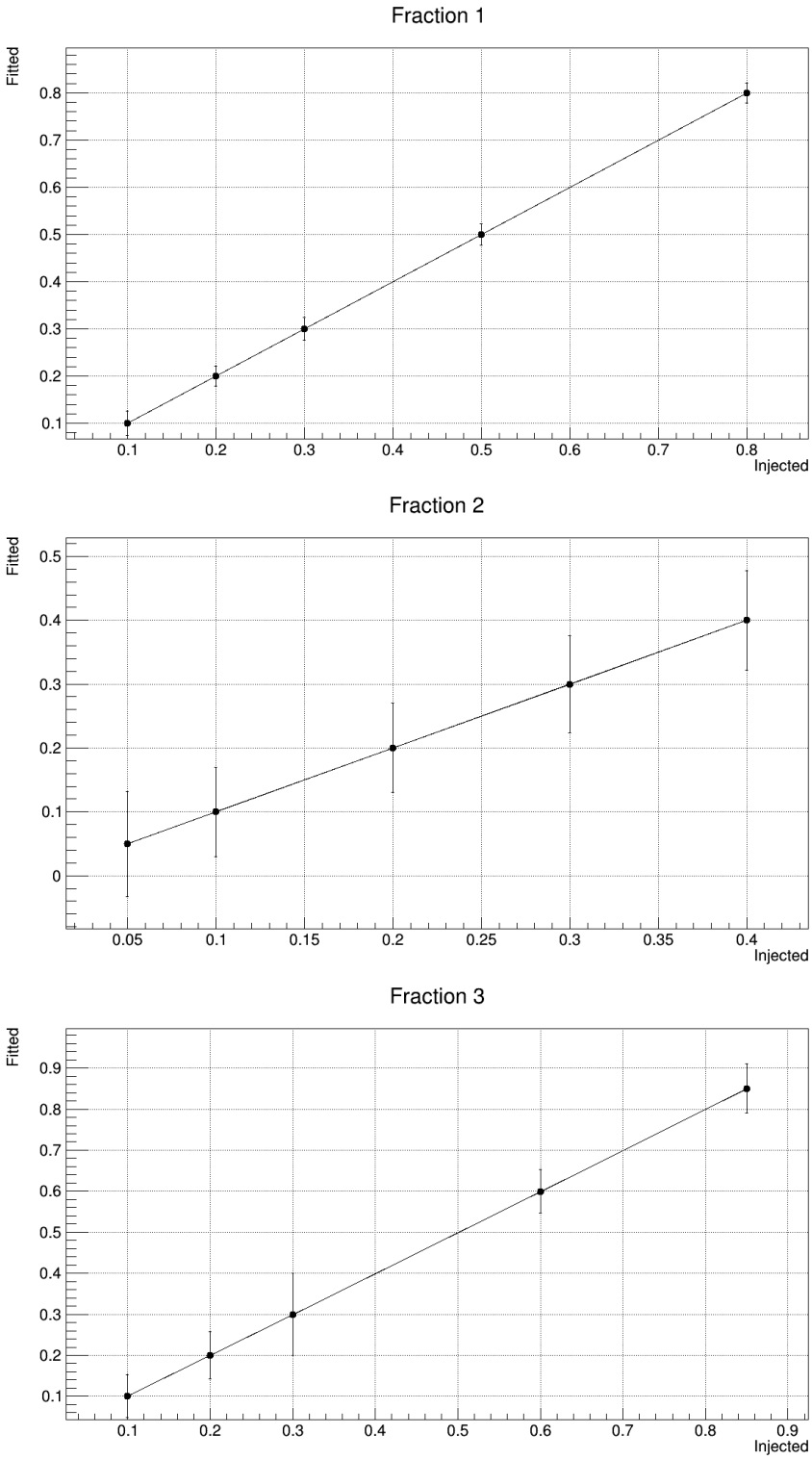


Figure 5.3: Fitted vs injected fractions for five cases

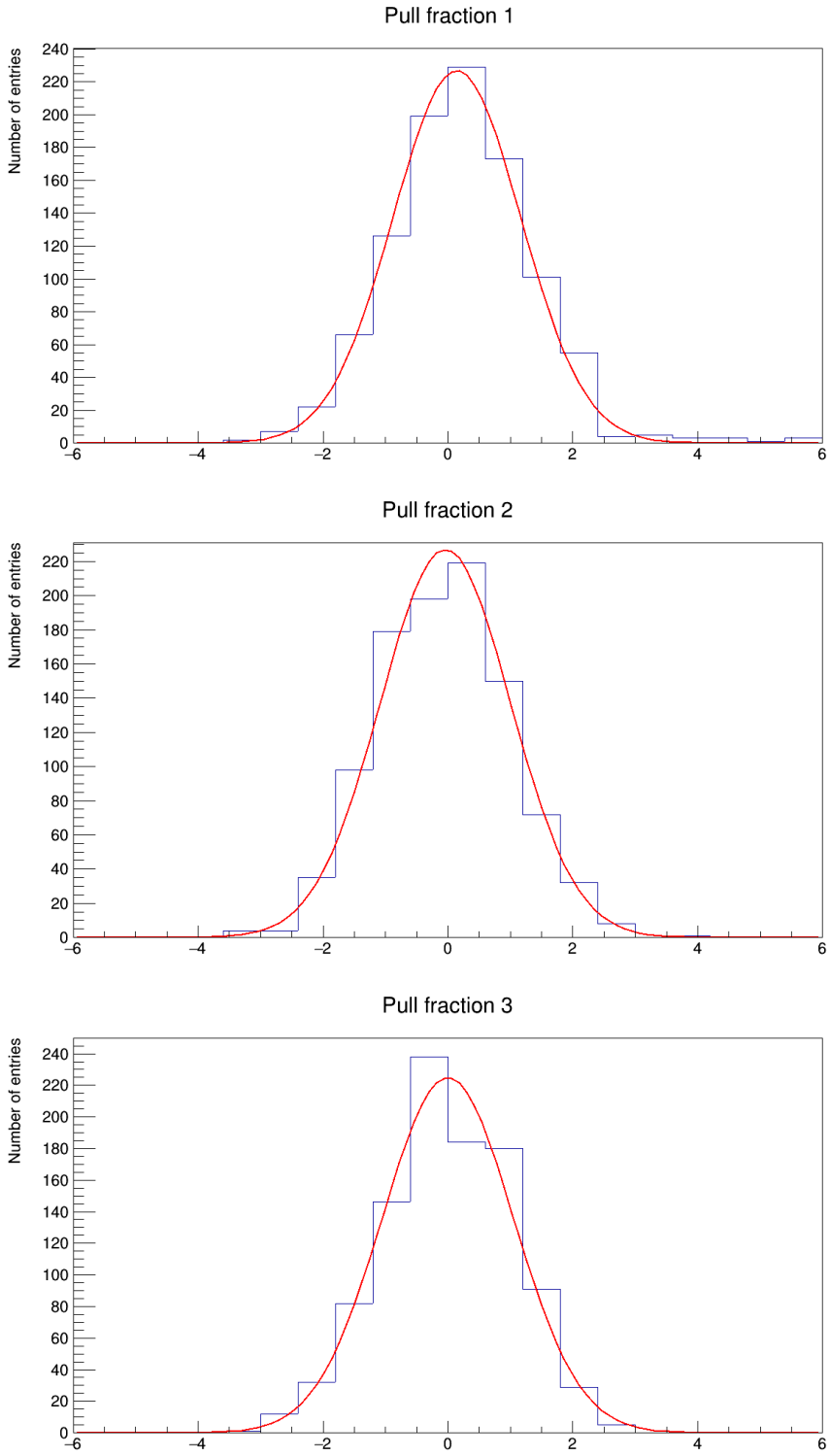


Figure 5.4: Pull distributions and fitted functions for the three fractions

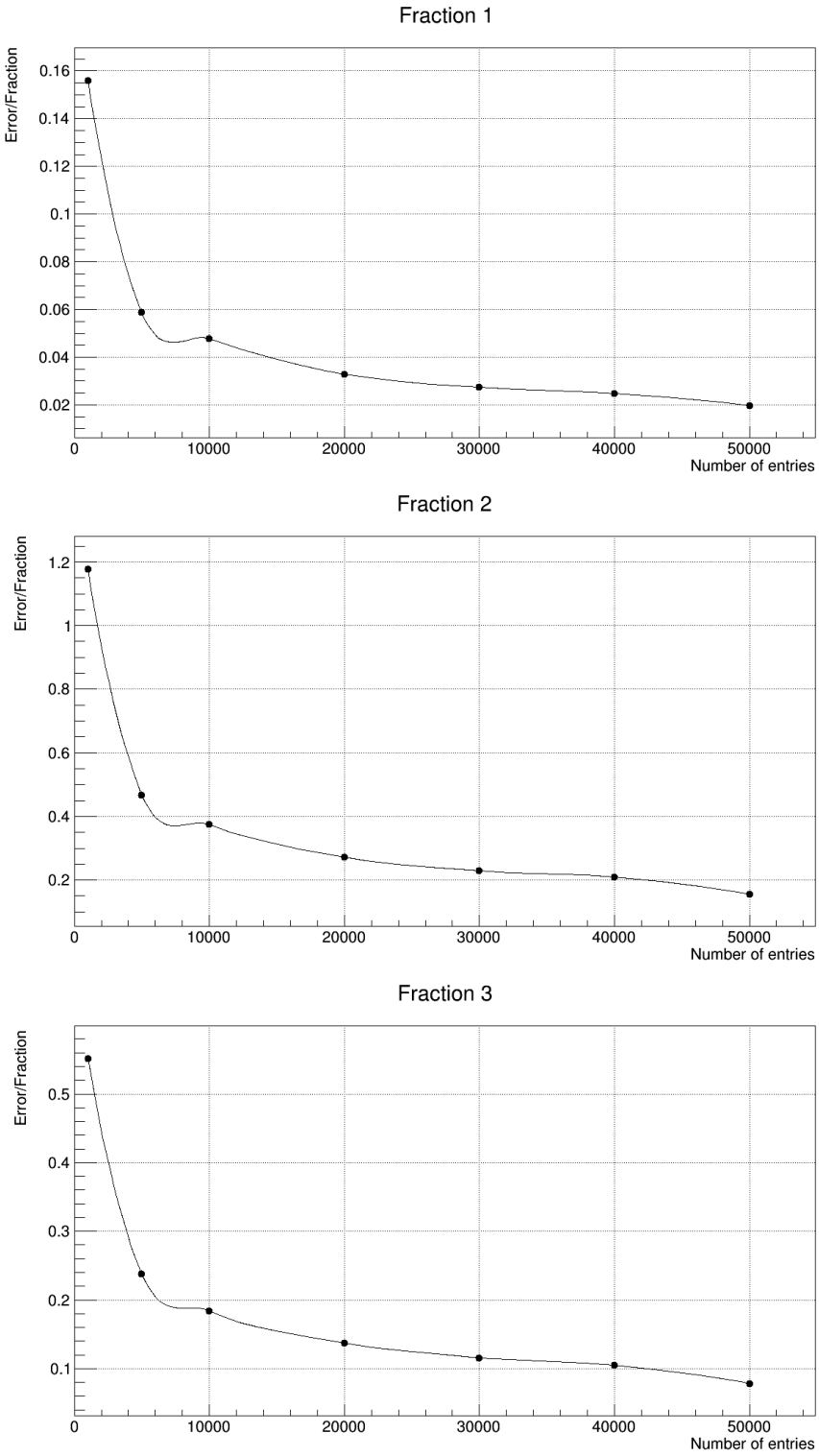


Figure 5.5: Fractional errors as a function of the number of entries of the data histogram

Chapter 6

Background

This chapter focuses on the main background of the present analysis, the QCD background. First, its flavour composition is determined and then the developed fit method is applied to it. This is done for both Monte Carlo (MC) and data, coming from official ATLAS samples.

6.1 Flavour composition of the QCD background

As it was already explained in Sections 3.2 and 4.2, the main background for the Higgs hadronic decays in the gluon fusion mode, as well as the Z decays is attributed to hadronic jets created by QCD processes. It is therefore extremely important to have a better understanding of it, that is why its flavour composition is studied. In practice, the QCD background is present in the detector in the form of quark-antiquark pairs forming large-R jets. All the combinations of quarks (except for the top, as it does not form hadrons) are possible, but it is sufficient to focus on b, c and light quarks (u, d, s) denoted by q.

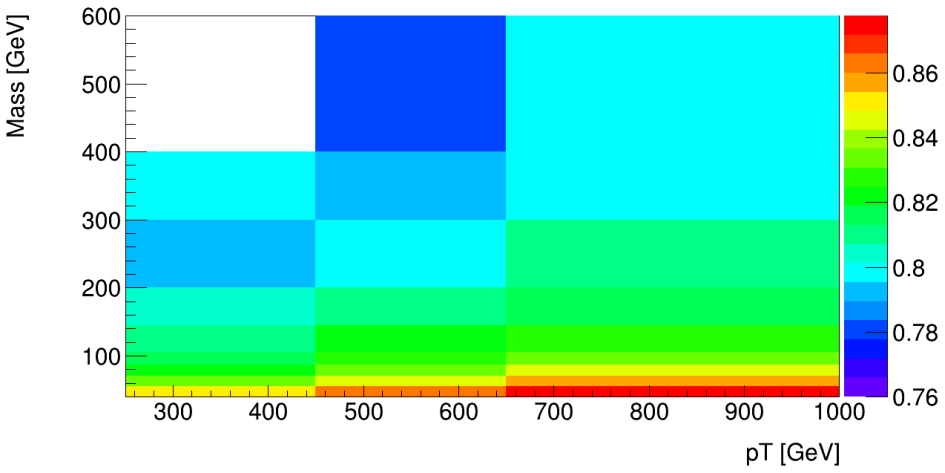
The flavour composition of the QCD background is determined by analysing MC generated events (by Pythia 8) and examining large-R ($R = 1.0$) jets. The analysis of the background covers the whole mass spectrum corresponding to a large-R jet mass of $40-600$ GeV. Jet quality cuts are applied to remove spurious jets. Only large-R jets with transverse momentum between 250 and 1000 GeV are selected. Their substructure is determined by analysing the variable radius track jets. These sub-jets are sorted according to their p_T and are required to have $p_T > 10$ GeV as jets with a smaller p_T value are unlikely to come from hadrons. The considered large-R jet must contain at least two sub-jets. If it does, the two leading sub-jets in p_T are selected as they correspond to di-jet events. If all these conditions are met, a histogram is filled with the large-R jet mass and p_T. Then, the flavour of the two sub-jets is found by examining the value of the variable jet flavour. For every combination of the flavour of the two sub-jets, a corresponding histogram is filled with the mass and p_T values and is then divided by the overall histogram. That way, a fraction for each component of the background can be computed and thus, the flavour composition of the background is fully specified. The flavour composition can be depicted in two dimensional histograms of the large-R jet mass and transverse momentum so as to see how the composition changes with respect to these two quantities (Figure 6.1). The binning for the mass was chosen in such a way so as to indicate the dominant physical pro-

cess and discriminate regions where either the background or the signal is prominent. In the latter case, it is possible to further separate different signal regions. This information is summed up in Table 6.1. The momentum is separated in three regions: $250 - 450$ GeV, $450 - 650$ GeV and $650 - 1000$ GeV. The light quarks dominate in the whole spectrum, even in the high p_T regions. It is interesting to note that in the case of b-quarks (Figure 6.1f), the result is counter-intuitive since its fraction in the background decreases as the p_T increases in the low mass regions. Its peak coincides with the Z peak, as it is expected. Moreover, all components exhibit a smooth dependence on the mass and p_T, which can justify an interpolation between different points. This conclusion is particularly useful for finding the QCD background composition in the signal region (Section 7.1).

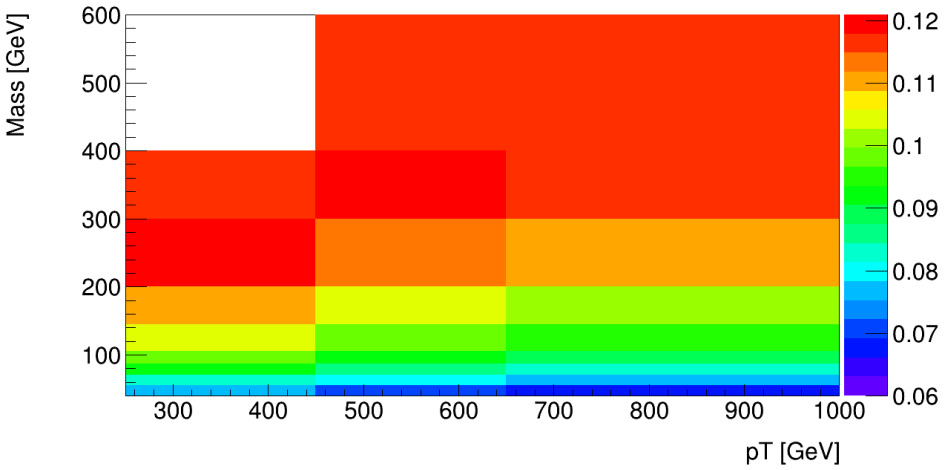
The flavour composition is also studied separately in the three p_T regions in the $70-210$ GeV mass range, while discriminating the cases where the jets are b-tagged or not. This is achieved by examining the value of the b-tagging cut for each of the two leading p_T variable radius track jets. The b-tagging in the simulation has an efficiency of 77%. If the value is one for both jets, the event is considered b-tagged, while if it is zero for both of them, it is classified as anti-b-tagged. The flavour composition of the QCD background in the three p_T regions as a function of the mass of the large-R jet is shown in Figure 6.2. The b-tagged case is shown in Figure 6.3 and the anti-b-tagged case in Figure 6.4. The three plots of Figure 6.3 show a difference in the performance of the b-tagging in the three p_T regions. The fraction of the true b-jets which are b-tagged is smaller in the $650 - 1000$ GeV region.

Mass (GeV)	40-55	55-70	70-87	87-105	105-145
Dominant physical process	QCD	QCD	W	Z	H
Mass (GeV)	145-200	200-300	300-400	400-600	
Dominant physical process	t	QCD	QCD	QCD	

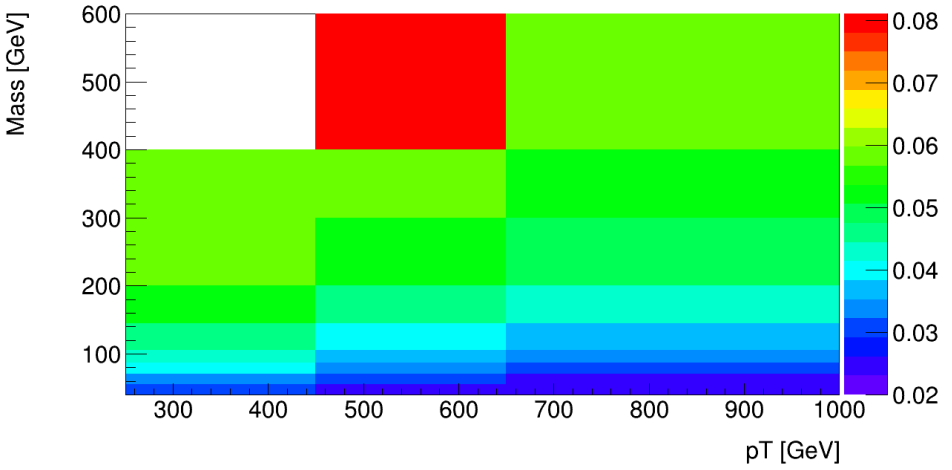
Table 6.1: Mass binning



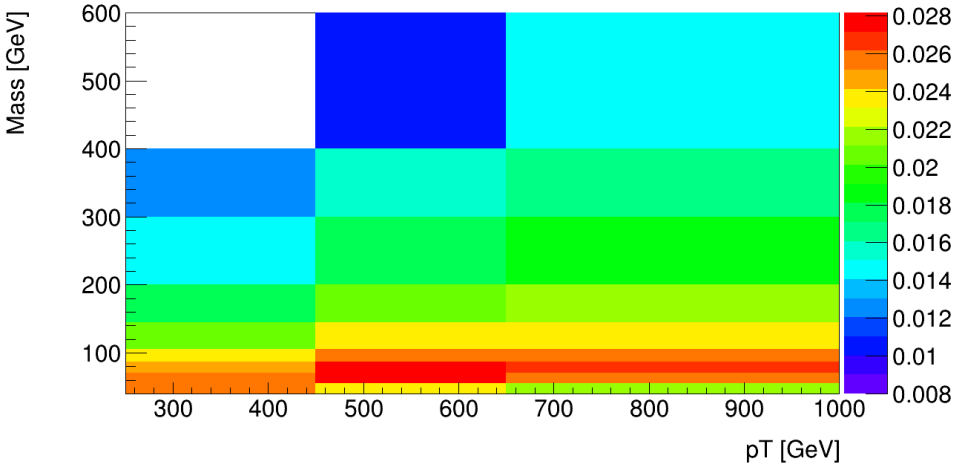
(a) qq


Mass [GeV]
600
500
400
300
200
100
300
400
500
600
700
800
900
1000
pT [GeV]
0.12
0.11
0.1
0.09
0.08
0.07
0.06

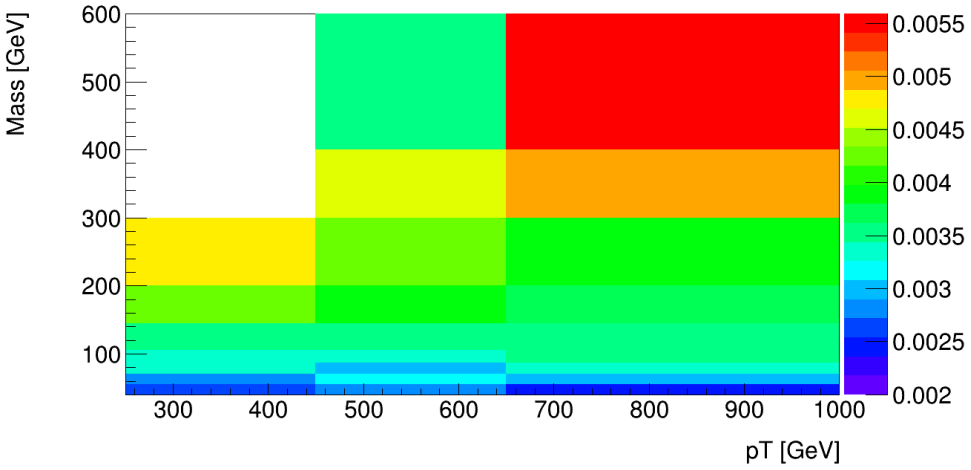
(b) qc


Mass [GeV]
600
500
400
300
200
100
300
400
500
600
700
800
900
1000
pT [GeV]
0.08
0.07
0.06
0.05
0.04
0.03
0.02

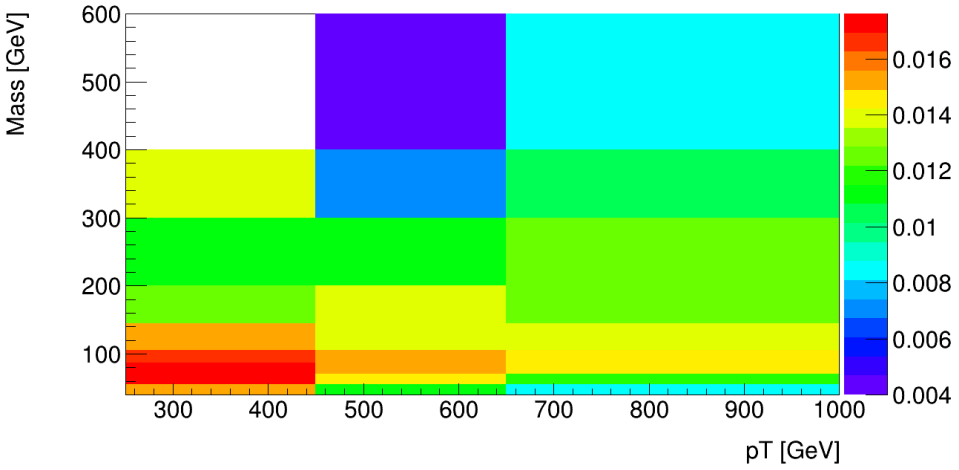
(c) qb


Mass [GeV]
600
500
400
300
200
100
300
400
500
600
700
800
900
1000
pT [GeV]
0.028
0.026
0.024
0.022
0.02
0.018
0.016
0.014
0.012
0.01
0.008

(d) cc



(e) cb



(f) bb

Figure 6.1: Fractions of the different flavour components of the QCD background as a function of the large-R jet mass and p_T.

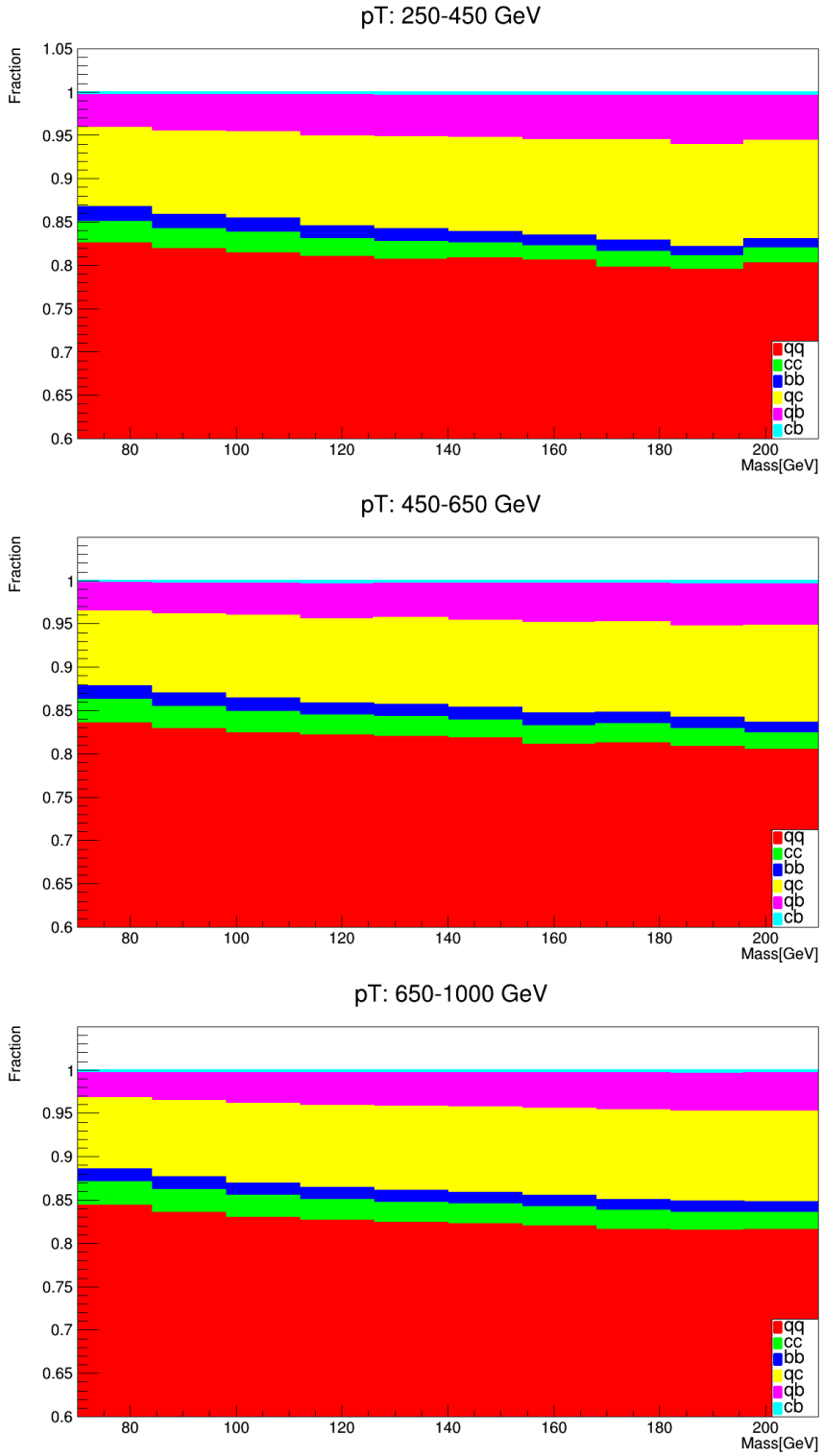


Figure 6.2: Fractions of the different flavour components of the QCD background in the $70-210$ *GeV large-R jet mass range in three separate* p_T *regions. The vertical axis starts at 0.6.*

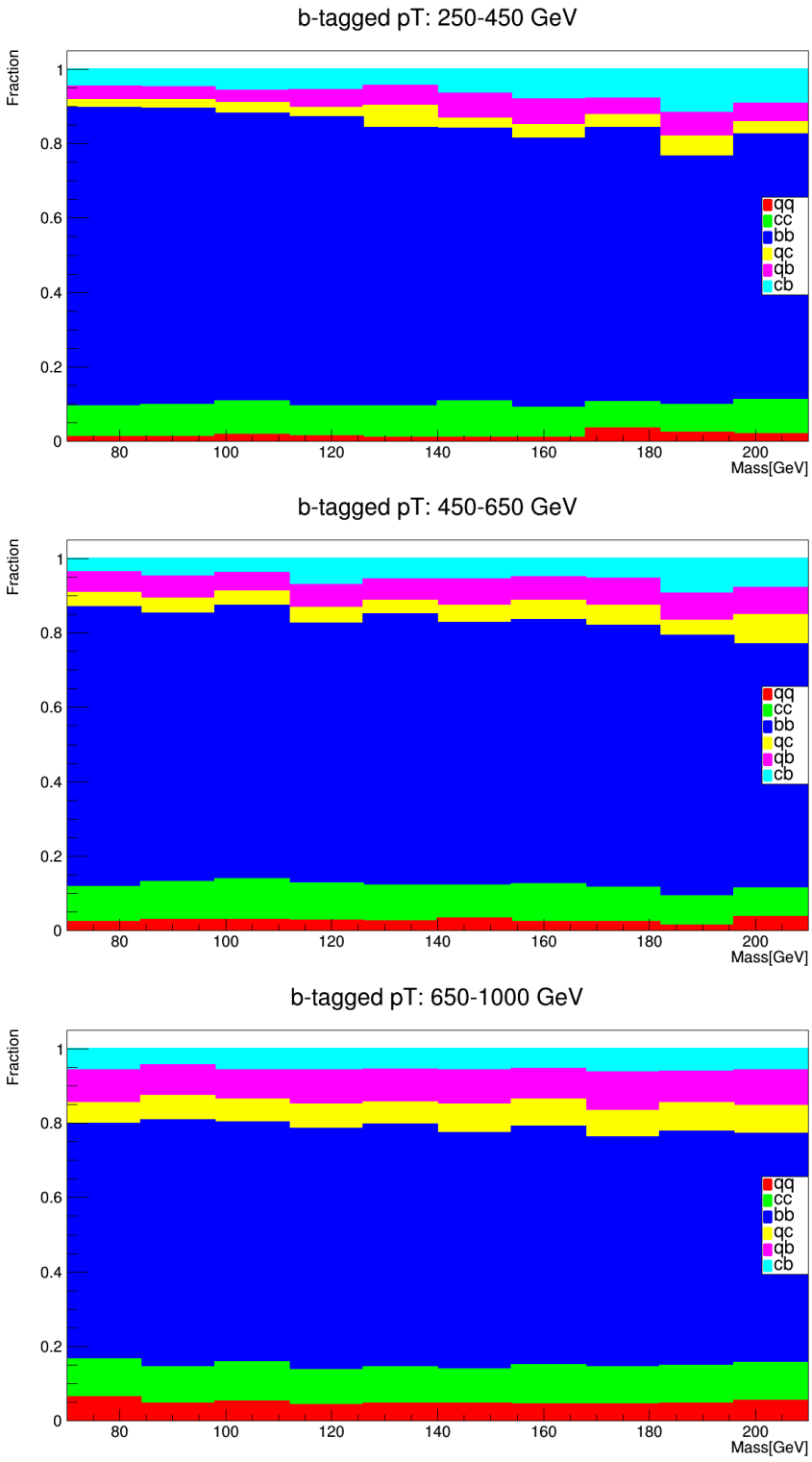


Figure 6.3: Fractions of the different flavour components of the QCD background in the $70 - 210$ *GeV large-R jet mass range in three separate* p_T *regions for b-tagged events.*

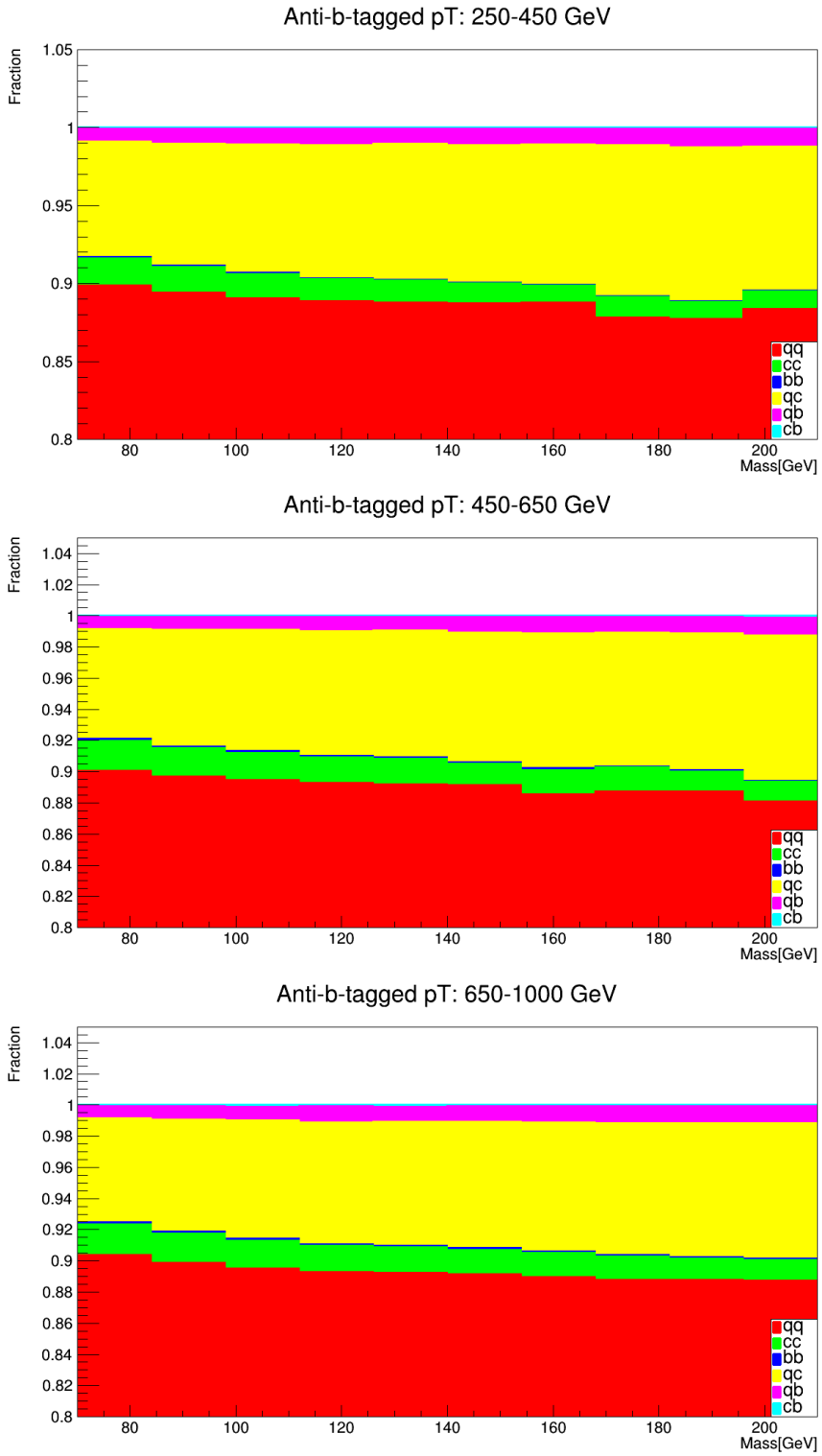


Figure 6.4: Fractions of the different flavour components of the QCD background in the $70-210$ *GeV large-R jet mass range in three separate* p_T *regions for anti-b-tagged events. The vertical axis starts at 0.8.*

6.2 Fitting the QCD background

6.2.1 Monte Carlo

The fit method must be applied to the background in order to examine how well it can be fitted and compare the results with the true flavour composition (Figure 6.1). This can be done in the three mass regions where it is dominant: 40 - 70 GeV, 200 - 300 GeV and 300 - 400 GeV. The templates and the samples to fit are created by selecting MC events in a similar process as in Section 6.1. The large-R jet should have the mass and momentum corresponding to each region of interest. There is no need to have p_T dependent templates but it is useful to create different distributions corresponding to the three p_T regions ($250-450$ GeV, $450-650$ GeV and $650-1000$ GeV) in order to fit them separately. The large-R jets should have at least two sub-jets with $p_T > 10$ GeV. Then, the MV2c10 value for each of the two leading variable radius track jets (denoted by pr_1, pr_2) is used, to create a combined probability for the large-R jet. The MV2c10 output (Figure 4.6) ranges from -1 to 1, so by using the formula:

$$pr_c = 0.25(pr_1 + 1)(pr_2 + 1)[1 - \log(0.25(pr_1 + 1)(pr_2 + 1))] \tag{6.1}$$

the two outputs are combined into a total probability pr_c that ranges from 0 to 1 so as to reflect its physical meaning. This value is filled in a histogram to create the "data" to be fitted. The templates are created by examining the flavour of the leading sub-jets and filling the different flavour histograms with the combined probability.

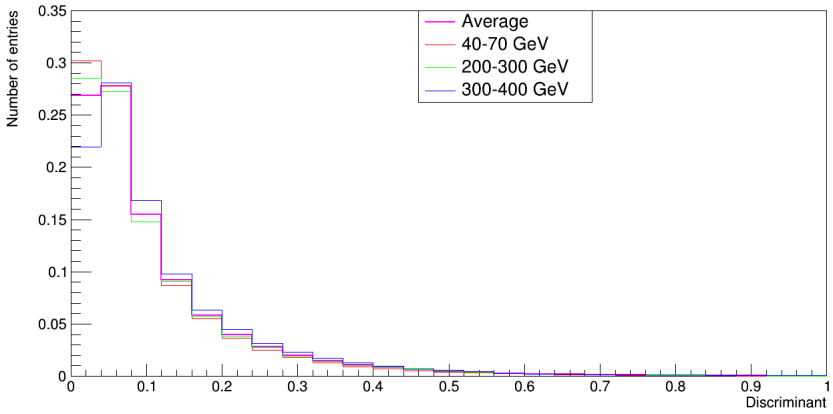
Five templates for each mass region are formed which correspond to di-jet discriminants for the qq, cc, bb, qb and qc quark pairs (Figures 6.5a-6.5e). As shown in Figure 6.1e, cb has the smallest contribution on the background so it is not taken into consideration. The templates are normalised so as to have an area of one. They are compatible enough that it is not necessary to have mass dependent discriminants. Therefore, their average value is computed to create one template for each flavour component, which can in turn be used for fitting all three mass regions.

The optimal number of templates for fitting the background is determined by finding a balance between sufficiently describing the background and limiting the number of free parameters of the fit. Therefore, the fit is performed in various mass and p_T regions for three different cases (Tables 6.2 - 6.9). Initially, the fit is tested using three templates (qq, cc, bb). Three templates are not sufficient to describe the background as there is significant contribution from the qc and qb terms. Therefore, the fit is then tested using all five templates. However, in this case, the contribution of either qc or qb is lost because the discriminants of cc, qc, qb are too similar, especially in the first bins. That is why qc and qb are combined into a single template, by computing the average value of the two histograms. Thus, for all subsequent fits, four templates are used: qq, cc, bb and $qcqb$ (Figure 6.5f) for representing the mixed terms.

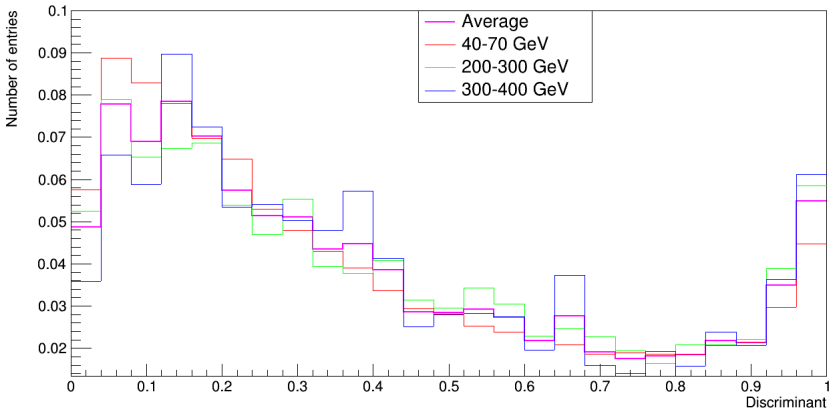
In the different fits, all the fractions are free parameters, except the one corresponding to bb, which can be found by generalising Equation 5.2 to include the extra fractions. The errors of the fit in the cases using four and five templates can be propagated to bb, by modifying Equation 5.4 as:

$$\sigma_{bb} = \sqrt{\sum_{i=1}^{n} \sigma_i + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n} \sigma_{ij}} \tag{6.2}$$

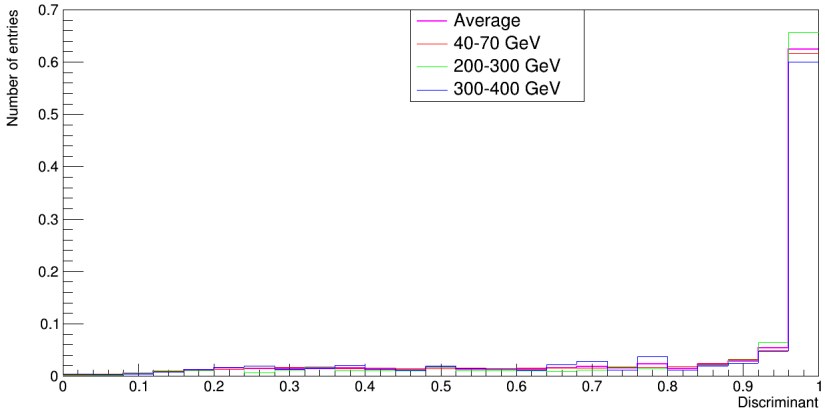
where n is the number of the free fractions.



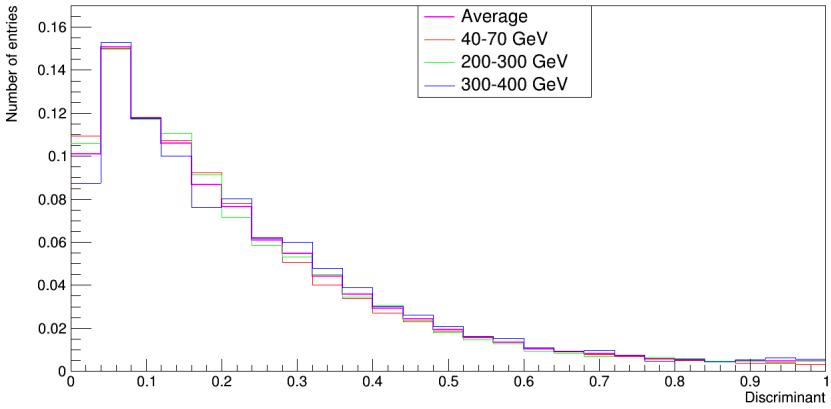
(a) qq



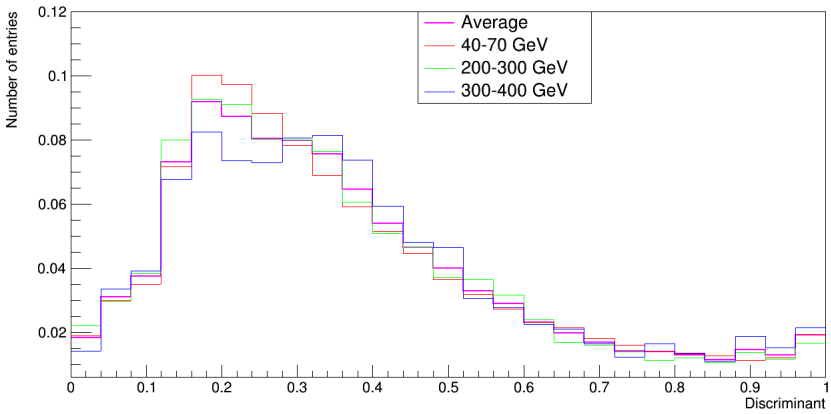
(b) cc



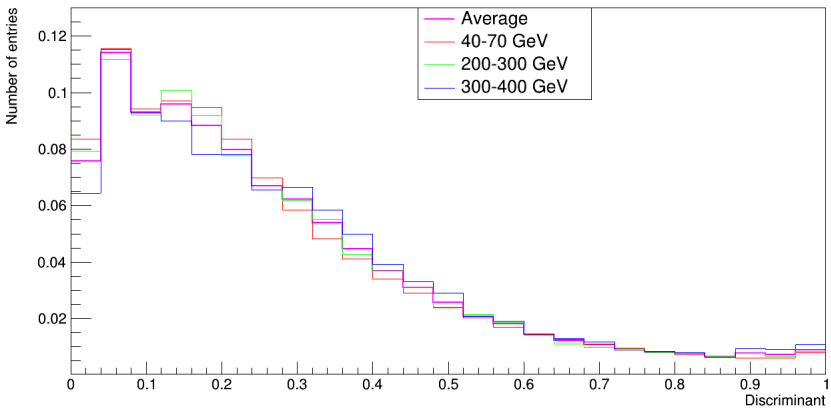
(c) bb



(d) qc



(e) qb



(f) qcqb

Figure 6.5: QCD templates based on the MV2c10 output for quark pairs of different flavour in the three mass regions and their average value

	True	Fit		
		3 templates	4 templates	5 templates
fr_{qq}	0.8428	0.9206 ± 0.0006	0.9100 ± 0.0010	0.9188 ± 0.0006
fr_{cc}	0.0256	0.0656 ± 0.0007	0.0495 ± 0.0013	0.0328 ± 0.0015
fr_{bb}	0.0160	0.0137 ± 0.0002	0.0151 ± 0.0002	0.0161 ± 0.0028
fr_{qc}	0.0803	-	-	$\approx 0 \pm 0.00007$
fr_{qb}	0.0326	-	-	0.0323 ± 0.0014
fr_{qcqb}	0.1129	-	0.0025 ± 0.0019	-

Table 6.2: Mass $= 40 - 70$ *GeV,* $p_T = 250 - 450$ *GeV*

	True	Fit		
		3 templates	4 templates	5 templates
fr_{qq}	0.7938	0.9097 ± 0.0046	0.8771 ± 0.0068	0.9004 ± 0.0047
fr_{cc}	0.0146	0.0810 ± 0.0052	0.0395 ± 0.0092	0.0105 ± 0.0113
fr_{bb}	0.0108	0.0093 ± 0.0015	0.0126 ± 0.0015	0.0141 ± 0.0024
fr_{qc}	0.1187	-	-	$\approx 0 \pm 0.0018$
fr_{qb}	0.0573	-	-	0.0749 ± 0.0105
fr_{qcqb}	0.1760	-	0.0708 ± 0.0150	-

Table 6.3: Mass $= 200 - 300$ *GeV,* $p_T = 250 - 450$ *GeV*

	True	Fit		
		3 templates	4 templates	5 templates
fr_{qq}	0.8557	0.8636 ± 0.0009	0.7646 ± 0.0013	0.6853 ± 0.0016
fr_{cc}	0.0257	0.1335 ± 0.0010	$\approx 0 \pm 0.0001$	$\approx 0 \pm 0.0001$
fr_{bb}	0.0127	0.0028 ± 0.0003	0.0128 ± 0.0002	0.0140 ± 0.0003
fr_{qc}	0.0754	-	-	0.3007 ± 0.0017
fr_{qb}	0.0274	-	-	$\approx 0 \pm 0.00003$
fr_{qcqb}	0.1028	-	0.2226 ± 0.0014	-

Table 6.4: Mass $= 40 - 70$ *GeV,* $p_T = 450 - 650$ *GeV*

	True	Fit		
		3 templates	4 templates	5 templates
fr_{qq}	0.7999	0.8761 ± 0.0021	0.8151 ± 0.0035	0.8550 ± 0.0099
fr_{cc}	0.0177	0.1185 ± 0.0023	0.0281 ± 0.0049	0.0146 ± 0.0036
fr_{bb}	0.0109	0.0054 ± 0.0006	0.0123 ± 0.0007	0.0125 ± 0.0010
fr_{qc}	0.1144	-	-	0.0262 ± 0.0180
fr_{qb}	0.0529	-	-	0.0917 ± 0.0106
fr_{qcqb}	0.1673	-	0.1445 ± 0.0068	-

Table 6.5: Mass $= 200 - 300$ *GeV,* $p_T = 450 - 650$ *GeV*

	True	Fit		
		3 templates	4 templates	5 templates
fr_{qq}	0.7919	0.8933 ± 0.0067	0.8403 ± 0.0123	0.8822 ± 0.0071
fr_{cc}	0.0158	0.1039 ± 0.0075	0.0371 ± 0.0148	0.0228 ± 0.0149
fr_{bb}	0.0072	0.0028 ± 0.0142	0.0075 ± 0.0014	0.0077 ± 0.0550
fr_{qc}	0.1210	-	-	$\approx 0 \pm 0.0550$
fr_{qb}	0.0596	-	-	0.0872 ± 0.0142
fr_{qcqb}	0.1806	-	0.1151 ± 0.0223	-

Table 6.6: Mass $= 300 - 400$ *GeV,* $p_T = 450 - 650$ *GeV*

	True	Fit		
		3 templates	4 templates	5 templates
fr_{qq}	0.8695	0.7639 ± 0.0006	0.5882 ± 0.0008	0.4545 ± 0.0010
fr_{cc}	0.0231	0.2455 ± 0.0007	$\approx 0 \pm 0.00002$	$\approx 0 \pm 0.00002$
fr_{bb}	0.0099	$\approx 0 \pm 0.0002$	0.0080 ± 0.0001	0.0102 ± 0.0001
fr_{qc}	0.0704	-	-	0.5353 ± 0.0010
fr_{qb}	0.0243	-	-	$\approx 0 \pm 0.000002$
fr_{qcqb}	0.0947	-	0.4038 ± 0.0009	-

Table 6.7: Mass $= 40 - 70$ *GeV,* $p_T = 650 - 1000$ *GeV*

	True	Fit		
		3 templates	4 templates	5 templates
fr_{qq}	0.8086	0.7971 ± 0.0014	0.6648 ± 0.0021	0.5616 ± 0.0026
fr_{cc}	0.0187	0.2050 ± 0.0016	$\approx 0 \pm 0.0003$	$\approx 0 \pm 0.0028$
fr_{bb}	0.0122	$\approx 0 \pm 0.0005$	0.0144 ± 0.0005	0.0165 ± 0.0028
fr_{qc}	0.1088	-	-	0.4218 ± 0.0027
fr_{qb}	0.0478	-	-	$\approx 0 \pm 0.0001$
fr_{qcqb}	0.1566	-	0.3208 ± 0.0022	-

Table 6.8: Mass $= 200 - 300$ *GeV,* $p_T = 650 - 1000$ *GeV*

	True	Fit		
		3 templates	4 templates	5 templates
fr_{qq}	0.8002	0.8157 ± 0.0026	0.6760 ± 0.0038	0.5716 ± 0.0051
fr_{cc}	0.0164	0.1870 ± 0.0029	$\approx 0 \pm 0.0009$	$\approx 0 \pm 0.0071$
fr_{bb}	0.0104	$\approx 0 \pm 0.0009$	0.0124 ± 0.0012	0.0145 ± 0.0071
fr_{qc}	0.1163	-	-	0.4139 ± 0.0052
fr_{qb}	0.0516	-	-	$\approx 0 \pm 0.0005$
fr_{qcqb}	0.1679	-	0.3116 ± 0.0039	-

Table 6.9: Mass $= 300 - 400$ *GeV,* $p_T = 650 - 1000$ *GeV*

Even with four templates, there are cases where some fitted fractions are equal to zero. The fit can generally extract the *bb* contribution with greater accuracy than all the others since the b-jet discrimination is better. Figure 6.6 shows the MC predicted flavour composition along with the fitted values in the three main mass regions for $p_T = 450 - 650$ GeV. The comparison with the true values in the three bins shows that the fit results for the *qq* and *qcqb* fractions are anti-correlated. Hence, the two individual fractions are discarded as free parameters in the fit and replaced with an overall parameter. In order to fit them together, a rate is introduced in the fit, defined as:

$$r = \frac{qq}{qq + qcqb} \tag{6.3}$$

which is kept constant at its MC predicted value. This change is useful for handling the overlap between the templates, which is significant and usually results in losing the contribution of the *cc* pair.

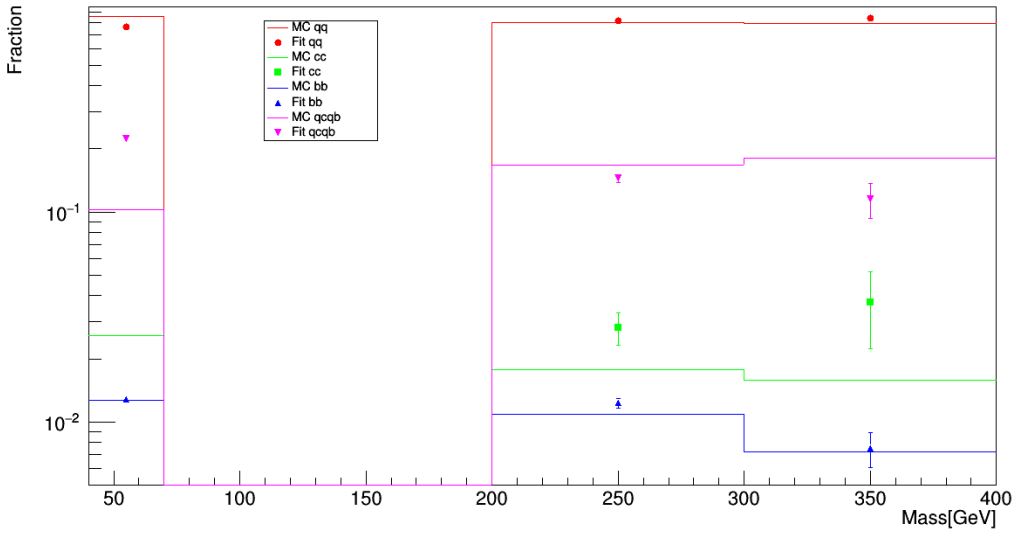
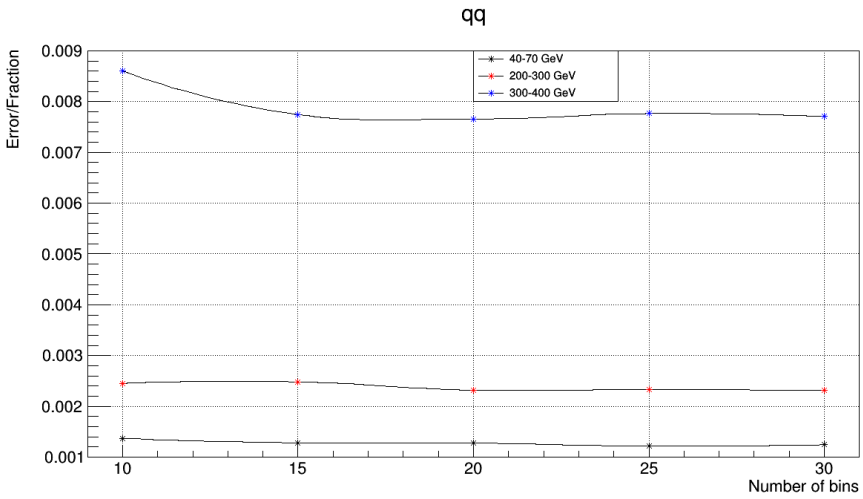
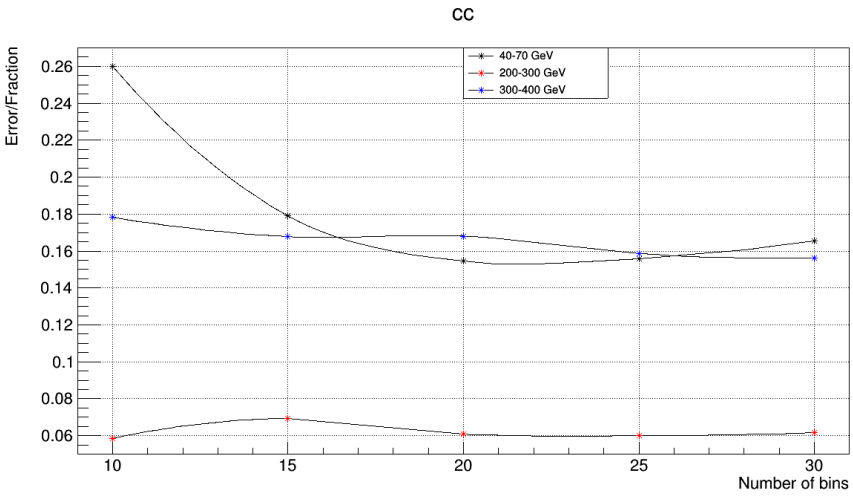
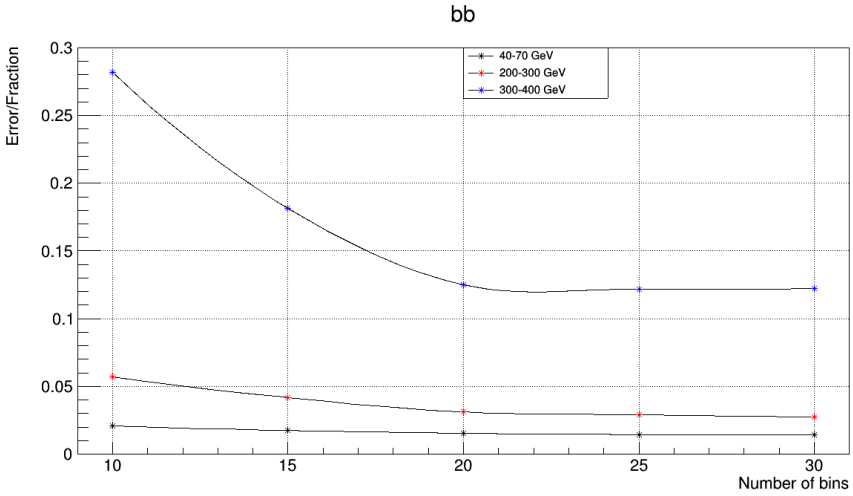


Figure 6.6: Comparison between MC predicted and fitted values for the different flavour fractions in the three QCD mass bins for $p_T = 450 - 650$ GeV

The number of bins of the histograms also affects the fit results, so the optimal selection should be made. A large number of bins gives a more accurate description of the distributions but also makes the fit more complex. For this reason, multiple fits are performed while the number of bins ranges from 10 to 30. Figure 6.7 shows how the fractional error of the fit parameters changes with respect to the number of bins in the three mass regions. This was tested in the $p_T = 450 - 650$ GeV region and using four templates for the fit. Based on these results the number of bins was chosen as 25.


qq
Error/Fraction
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
10
15
20
25
30
Number of bins
40-70 GeV
200-300 GeV
300-400 GeV


cc
Error/Fraction
0.26
0.24
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
10
15
20
25
30
Number of bins
40-70 GeV
200-300 GeV
300-400 GeV


bb
Error/Fraction
0.3
0.25
0.2
0.15
0.1
0.05
0
10
15
20
25
30
Number of bins
40-70 GeV
200-300 GeV
300-400 GeV

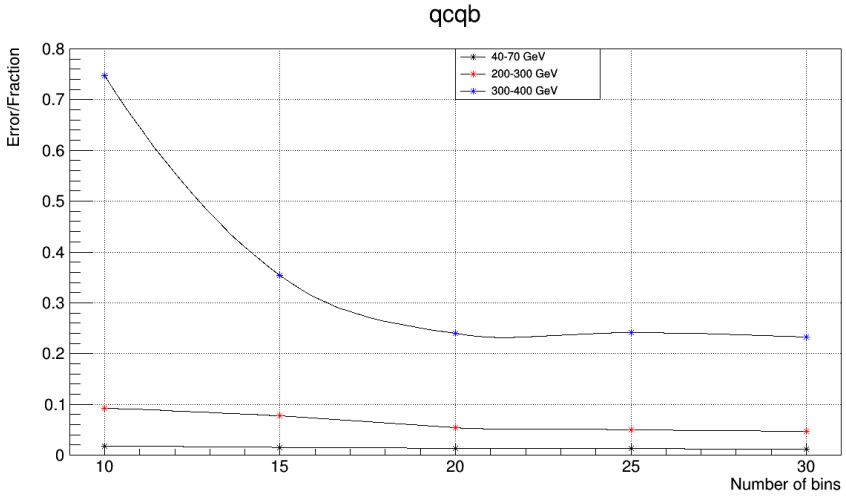


Figure 6.7: Fractional error as a function of the number of bins

6.2.2 Data

Using Equation 6.3 the fit method can be applied to data in the QCD dominant regions. The results are shown in the following table. Comparing with the MC predicted flavour composition, there are significant discrepancies. Although some contributions from t and ZW decays are expected, they are negligible. The high mass region reflects better the true composition but has larger parameter errors as it has less events than the low mass region.

	m = 40-70 GeV (r=0.8927)	m = 200-300 GeV (r=0.8270)
$fr_{qq+qcqb}$	0.8512 ± 0.0008	0.8877 ± 0.0016
fr_{cc}	0.1463 ± 0.0009	0.1003 ± 0.0018
fr_{bb}	0.0024 ± 0.0002	0.0121 ± 0.0005

Table 6.10: Data $p_T = 450 - 650$ GeV

Chapter 7

Methodology

In this chapter, a complete methodology for finding the hadronic branching ratios in a specified signal region corresponding to the decays of a boson (either Z, W or h), is developed. Then, the chapter analyses various methods for handling the background contribution and compares them. It includes simplified cases for testing and finally, the methodology is applied to various cases of interest.

7.1 Definition

The hadronic branching ratios can be extracted by applying a background subtraction method. This consists of three steps. First, a good estimate of the flavour fractions of the QCD background process in the signal region must be found. This can be approached in various ways as will be shown in Section 7.1.1. Then, the flavour composition of the combined background and signal processes can be determined by applying the fit method to the signal region. The signal region has a contribution from both the background processes and the relevant signal, so fitting it gives the overall fractions. The signal's fractions can then be found by using the following equation:

$$F^i = F^i_{bkg} R_{bkg} + F^i_{sig} R_{sig} \Longrightarrow \tag{7.1}$$

$$F^i_{sig} = \frac{F^i - F^i_{bkg} R_{bkg}}{R_{sig}} \tag{7.2}$$

where F, F_{bkg} and F_{sig} are the overall, background and signal fractions, respectively and i is used to represent the different decay modes. The rates R_{bkg} and R_{sig} are used to rescale the flavour fractions according to the relative amount of signal and background events in the signal region and are defined as:

$$R_{bkg} = \frac{n_{bkg}}{n_{bkg} + n_{sig}}, \quad R_{sig} = \frac{n_{sig}}{n_{bkg} + n_{sig}} \tag{7.3}$$

where n_{bkg} and n_{sig} are the number of expected background and signal events in the signal region and can be found from the integral of the MC histograms. It is important to correctly propagate the errors to the extracted fractions, as those reflect the accuracy of the method. They can be determined by considering the general expression:

$$x = \frac{a - bc}{d} \tag{7.4}$$

whose error σ_x can be expressed as a function of the individual fractional errors as:

$$\sigma_x^2 = \sigma_{a/d}^2 + \sigma_{bc/d}^2 = \left(\frac{a}{d}\right)^2 \cdot \left(\frac{\sigma_a^2}{a^2} + \frac{\sigma_d^2}{d^2}\right) + \left(\frac{bc}{d}\right)^2 \cdot \left(\frac{\sigma_b^2}{b^2} + \frac{\sigma_c^2}{c^2} + \frac{\sigma_d^2}{d^2}\right) \tag{7.5}$$

Equation 7.5 can then be applied to Equation 7.1 with the fractional errors for the fractions given by the fit, while those of the rates can be computed as:

$$\frac{\sigma_{R_{bkg}}^2}{R_{bkg}^2} = \frac{\sigma_{n_{bkg}}^2}{n_{bkg}^2} + \frac{\sigma_{n_{bkg}}^2 + \sigma_{n_{sig}}^2}{(n_{bkg} + n_{sig})^2}, \quad \frac{\sigma_{R_{sig}}^2}{R_{sig}^2} = \frac{\sigma_{n_{sig}}^2}{n_{sig}^2} + \frac{\sigma_{n_{bkg}}^2 + \sigma_{n_{sig}}^2}{(n_{bkg} + n_{sig})^2} \tag{7.6}$$

They can be further simplified by using the errors: $\sigma_{n_{bkg}} = \sqrt{n_{bkg}}$ and $\sigma_{n_{sig}} = \sqrt{n_{sig}}$. This yields:

$$\frac{\sigma_{R_{bkg}}^2}{R_{bkg}^2} = \frac{1}{n_{bkg}} + \frac{1}{n_{bkg} + n_{sig}}, \quad \frac{\sigma_{R_{sig}}^2}{R_{sig}^2} = \frac{1}{n_{sig}} + \frac{1}{n_{bkg} + n_{sig}} \tag{7.7}$$

7.1.1 Background estimation

Estimating the background contribution in the signal region can be handled in various ways. The simplest method that can first be attempted, is to measure the flavour fractions in a side-band where only the background process is dominant. For the side-band subtraction method to be valid, the background process should exhibit the same characteristics both in the background and in the signal region. As far as the QCD background is concerned, this practically means that the flavour composition should be constant with respect to the mass. However, as this is not the case for most of the flavour components, a better way would be to use two mass regions to interpolate the QCD fractions in the signal region. First, a simple linear interpolation is tried, since the flavour composition in individual p_T bins is approximately linear as a function of the mass (Figure 6.1) for most flavours. A linear interpolation at an interval (x_0, x_1) with known values (x_0, y_0) and (x_1, y_1), gives for the unknown value y at x:

$$y = y_0 + (x - x_0)\frac{y_1 - y_0}{x_1 - x_0} \tag{7.8}$$

The propagated error can then be found as:

$$\sigma_y^2 = \sigma_{y_0}^2 + \frac{(x - x_0)^2(y_1 - y_0)^2}{(x_1 - x_0)^2}\left(\frac{\sigma_x^2 + \sigma_{x_0}^2}{(x - x_0)^2} + \frac{\sigma_{y_1}^2 + \sigma_{y_0}^2}{(y_1 - y_0)^2} + \frac{\sigma_{x_1}^2 + \sigma_{x_0}^2}{(x_1 - x_0)^2}\right) \tag{7.9}$$

The above equations can be applied to interpolate between the low $(40 - 70$ GeV) and high $(200 - 300$ GeV) mass regions. In this case the x values and their errors are:

$$x_0 = 55 \text{ GeV}, \quad \sigma_{x_0} = 15 \text{ GeV}, \tag{7.10}$$

$$x_1 = 250 \text{ GeV}, \quad \sigma_{x_1} = 50 \text{ GeV} \tag{7.11}$$

while the y values correspond to the fitted fractions and their errors are given from the fits. For the case of the ZW bosons, the signal region is $70 - 105$ GeV so:

$$x = 87.5, \quad \sigma_x = 17.5 \text{ GeV} \tag{7.12}$$

However, some flavour components, especially the bb pair, deviate from the linear approximation as their peak is exhibited at the intermediate mass regions. So a higher order interpolation is needed. It's possible to use the functional dependence on the mass of the MC predicted flavour fractions and rescale it to fit the fitted fractions in the two mass regions, in order to obtain more accurate results. This is implemented by first fitting the three MC predicted fractions (low mass, high mass and signal region) with a second order polynomial:

$$y = p_2 x^2 + p_1 x + p_0 x \tag{7.13}$$

The MC predicted fractions have an error which should also be taken into account when computing the values and errors for the parameters p_0, p_1, p_2. Then, the two fitted fractions are fitted with a second order polynomial but p_1, p_2 are fixed at the MC predicted values, while only p_0 is a free parameter. The propagated error is:

$$\sigma_y^2 = x^4 \sigma_{p_2}^2 + x^2 \sigma_{p_1}^2 + \sigma_{p_0}^2 \tag{7.14}$$

To obtain the interpolated values, x is replaced with the center mass value of the signal region. Table 7.1 shows the MC predicted and the fitted values for the low and high mass regions. For each fit, the MC predicted rate r (Equation 6.3) was used. Table 7.2 shows the estimation of the background in the ZW signal region with the three aforementioned methods.

	m = 40-70 GeV		m = 200-300 GeV	
	True	Fit ($r = 0.8927$)	True	Fit ($r = 0.8270$)
$fr_{qq+qcqb}$	0.9585	0.9260 ± 0.0011	0.9672	0.9716 ± 0.0025
fr_{cc}	0.0257	0.0662 ± 0.0012	0.0177	0.0140 ± 0.0028
fr_{bb}	0.0127	0.0078 ± 0.0003	0.0109	0.0144 ± 0.0006

Table 7.1: QCD fits, $p_T = 450 - 650$ GeV

	True	Side-band	Linear	Quadratic
$fr_{qq+qcqb}$	0.955	0.9260 ± 0.0011	0.9336 ± 0.0059	0.9281 ± 0.0271
fr_{cc}	0.0263	0.0662 ± 0.0012	0.0575 ± 0.0067	0.0600 ± 0.0031
fr_{bb}	0.0155	0.0078 ± 0.0003	0.0089 ± 0.0009	0.0121 ± 0.0018

Table 7.2: QCD background estimation, Mass $70 - 105$ GeV, $p_T = 450 - 650$ GeV

The low mass region cannot be fitted as well as the high mass region due to its flavour composition, which has a bigger cc contribution. This makes the fit more difficult as the cc template has significant overlap with the others. Therefore, although the side-band (low mass region) subtraction would be valid for the $qq + qcqc$ and cc fractions, it cannot be done in practice. The errors of the high mass region are larger because it has fewer events. The quadratic interpolation was expected to give the best estimates, but this is not the case because the high mass region propagates a bigger error than in the linear interpolation case. Hence, the quadratic interpolation tends to give values closer to the low mass fractions.

7.2 Injection tests

The methodology of the previous section is first validated with some simple injection tests. The aim is to see if it is feasible to extract the flavour composition of a signal, when the background in the signal region is known exactly. By simplifying the problem this way, it's possible to determine the accuracy of the method and get a picture of its errors. In all of the following cases, the test signal is produced by multiplying the templates used for the fit with a constant which corresponds to the flavour fraction and scaling it with the total number of ZW events. The signal region consists of the sum of the signal and the background process. The latter is considered to be the same as in the background region. Therefore, applying Equation 7.1 should give exactly the flavour composition of the signal. The following tables show the fit results in the background and signal region, denoted by BR and SR, respectively and the subtracted values. The background in every case is the QCD process in the $70 - 105$ GeV mass region with $p_T = 450 - 650$ GeV.

	BR($r = 0.8703$)	SR ($r = 0.8703$)	Subtracted
$fr_{qq+qcqb}$	0.9461 ± 0.0013	0.9477 ± 0.0013	1.0800 ± 1.2589
fr_{cc}	0.0404 ± 0.0015	0.0389 ± 0.0015	-0.0865 ± 0.1842
fr_{bb}	0.0135 ± 0.0004	0.0134 ± 0.0004	0.0070 ± 0.0504

Table 7.3: Signal: $qq * n_z$

	BR($r = 0.8703$)	SR($r = 0.8703$)	Subtracted
$fr_{qq+qcqb}$	0.9461 ± 0.0013	0.9346 ± 0.0013	-0.0387 ± 1.2507
fr_{cc}	0.0404 ± 0.0015	0.0521 ± 0.0015	1.0429 ± 0.1900
fr_{bb}	0.0135 ± 0.0004	0.0133 ± 0.0004	-0.0042 ± 0.0513

Table 7.4: Signal: $cc * n_z$

	BR($r = 0.8703$)	SR($r = 0.8703$)	Subtracted
$fr_{qq+qcqb}$	0.9461 ± 0.0013	0.9342 ± 0.0013	-0.0671 ± 1.2501
fr_{cc}	0.0404 ± 0.0015	0.0409 ± 0.0015	0.00894 ± 0.1849
fr_{bb}	0.0135 ± 0.0004	0.0248 ± 0.0005	0.9777 ± 0.0588

Table 7.5: Signal: $bb * n_z$

The signal flavour contribution can be found relatively well and the small discrepancies are attributed to the existence of the template for the mixed term in the fit. However, the errors of the method are large and this is intrinsic to the subtraction. The number of background events is much larger than the signal events, making the relative rates of Equation 7.3 approximately 0.988 for the background and 0.012 for the signal. Hence, it is expected that the method will propagate significant errors to the final result.

	BR ($r = 0.8703$)	SR ($r = 0.8703$)	Subtracted
fr_{q+qcqb}	0.9461 ± 0.0013	0.9413 ± 0.0013	0.5306 ± 1.2548
fr_{cc}	0.0404 ± 0.0015	0.0419 ± 0.0015	0.1680 ± 0.1857
fr_{bb}	0.0135 ± 0.0004	0.0169 ± 0.0004	0.3014 ± 0.0518

Table 7.6: Signal: $(0.5qq + 0.2cc + 0.3bb) * n_{ZW}$

7.3 Applications

7.3.1 ZW signal region

The methodology of finding the branching ratios can be applied to the ZW signal region. The extracted fractions are not branching ratios, but can be thought of as overall fractions for the two vector bosons, denoted by F_V. Therefore, Equation 7.1 becomes:

$$F_V^i = \frac{F^i - F_{QCD} R_{QCD}}{R_V} \tag{7.15}$$

Taking into account both the values and the errors of Table 7.2, the best choice is to use the linear interpolation results, as the quadratic interpolation has a non-negligible error for the $qq + qcqb$ fraction. This gives the best estimate for the background in the signal region and is summed up in the second column of Table 7.7. The method is tested for MC and data samples in the $p_T = 450 - 650$ GeV region, using for both the MC rates:

$$R_{QCD} = 0.9883, \quad R_V = 0.0117 \tag{7.16}$$

For the MC case, adding the QCD and ZW histograms in the signal region forms the overall histogram to be fitted. The results are summed up in the following tables.

	BR	SR ($r = 0.8703$)	F_V	F_V (physical)
$fr_{qq+qcqb}$	0.9336 ± 0.0059	0.9451 ± 0.0013	1.9129 ± 1.3388	0.5068 ± 0.0878
fr_{cc}	0.0575 ± 0.0067	0.0412 ± 0.0015	-1.3270 ± 0.5827	0.0790 ± 0.0777
fr_{bb}	0.0089 ± 0.0009	0.0137 ± 0.0004	0.4141 ± 0.0836	0.4142 ± 0.0408

Table 7.7: Results of the method for the ZW signal region, $p_T = 450 - 650$ *GeV, MC*

	BR	SR ($r = 0.8703$)	F_V	F_V (physical)
$fr_{qq+qcqb}$	0.8573 ± 0.0047	0.8644 ± 0.0009	1.4580 ± 1.2040	0.4737 ± 0.1505
fr_{cc}	0.1387 ± 0.0059	0.1271 ± 0.0010	-0.8483 ± 0.5331	0.1359 ± 0.1258
fr_{bb}	0.0040 ± 0.0012	0.0086 ± 0.0003	0.3903 ± 0.1074	0.3904 ± 0.0827

Table 7.8: Results of the method for the ZW signal region, $p_T = 450 - 650$ *GeV, Data*

After the subtraction, a physical value between 0 and 1 is obtained only for the bb fraction. That is why, the method is repeated in a "toy" study. The results in the BR and SR correspond to distributions with mean equal to their value and width equal to the error. Sampling them 10000 times and each time repeating the subtraction, produces a

distribution for the subtracted fractions. These are shown in Figures 7.1 and 7.2. Then, it's possible to consider only the part of the distribution in the physical range. Taking the mean and standard deviation of these parts produces fractions in the desired range. As this process will give fractions that no longer sum up to one, it is done only for two of them (cc, bb) and the third one is found by subtracting from one. The physical values correspond to the fifth column of Tables 7.7 and 7.8. Nevertheless, the $qq + qcqb$ and cc fractions have too few entries in the (0, 1) range to consider the results reliable. Therefore, they do not represent good estimates for the F_V.

7.3.2 Higgs signal region

The application of the method to the Higgs signal region follows the same steps as the ZW region, with the only difference concerning the background estimation. The low and high mass regions are used to find the composition in the 105 - 145 GeV mass region. A linear interpolation is used. This is done for the $p_T = 250 - 450$ GeV region. The results are shown in Table 7.9. Given that there are few Higgs events in the MC samples used for the analysis, only the large radius jet with the highest p_T is selected from the QCD sample. This way, the number of background events taken into consideration is reduced. The MC rates are:

$$R_{QCD} = 0.9878, \quad R_H = 0.0122 \tag{7.17}$$

Using the developed methodology, the Higgs hadronic branching ratios, F_H can be estimated. The results are shown in Table 7.10. Some discrepancies are expected because the Higgs event selection is based on identifying the decay mode, while the method is based on the flavour tagging output for the variable radius jets. This mainly affects the decay to gluons, as the final state does not always contain only light-flavour jets. Although the branching ratios have physical values, they are not in agreement with the MC predicted ones. The results have large errors due to the small signal to background ratio.

	m = 40-70 GeV		m = 105-145 GeV	m = 200-300 GeV	
	True	Fit ($r = 0.8720$)	True	True	Fit ($r = 0.7999$)
$fr_{qq+qcqb}$	0.9514	0.9805 ± 0.0003	0.9578	0.9683	0.9865 ± 0.0013
fr_{cc}	0.0265	$\approx 0 \pm 0.00004$	0.0209	0.0145	$\approx 0 \pm 0.0003$
fr_{bb}	0.0185	0.0195 ± 0.0003	0.0168	0.0114	0.0135 ± 0.0013

Table 7.9: QCD fits, $p_T = 250 - 450$ GeV

	BR	SR ($r = 0.8289$)	F_H
$fr_{qq+qcqb}$	0.9826 ± 0.0011	0.9712 ± 0.0005	0.0405 ± 2.3645
fr_{cc}	$\approx 0 \pm 0.0001$	$\approx 0 \pm 0.00004$	$\approx 0 \pm 0.0108$
fr_{bb}	0.0174 ± 0.0011	0.0288 ± 0.0005	0.9595 ± 0.1154

Table 7.10: Results of the method for the Higgs signal region, $p_T = 250 - 450$ GeV, MC

It is clear that with the current flavour tagging response, it is not possible to have an inclusive measurement of the different flavour fractions. This is mostly attributed to the fact that the developed method introduces a fourth template, which has significant

overlap with the other three. Consequently, the fit no longer works sufficiently. Moreover, the method has large errors which could be reduced by an increase in statistics. That is why a selection based on a production mode with less background, such as the Z/W associated production would be preferable. The topology of the decay would make it easier to increase the signal to background ratio.

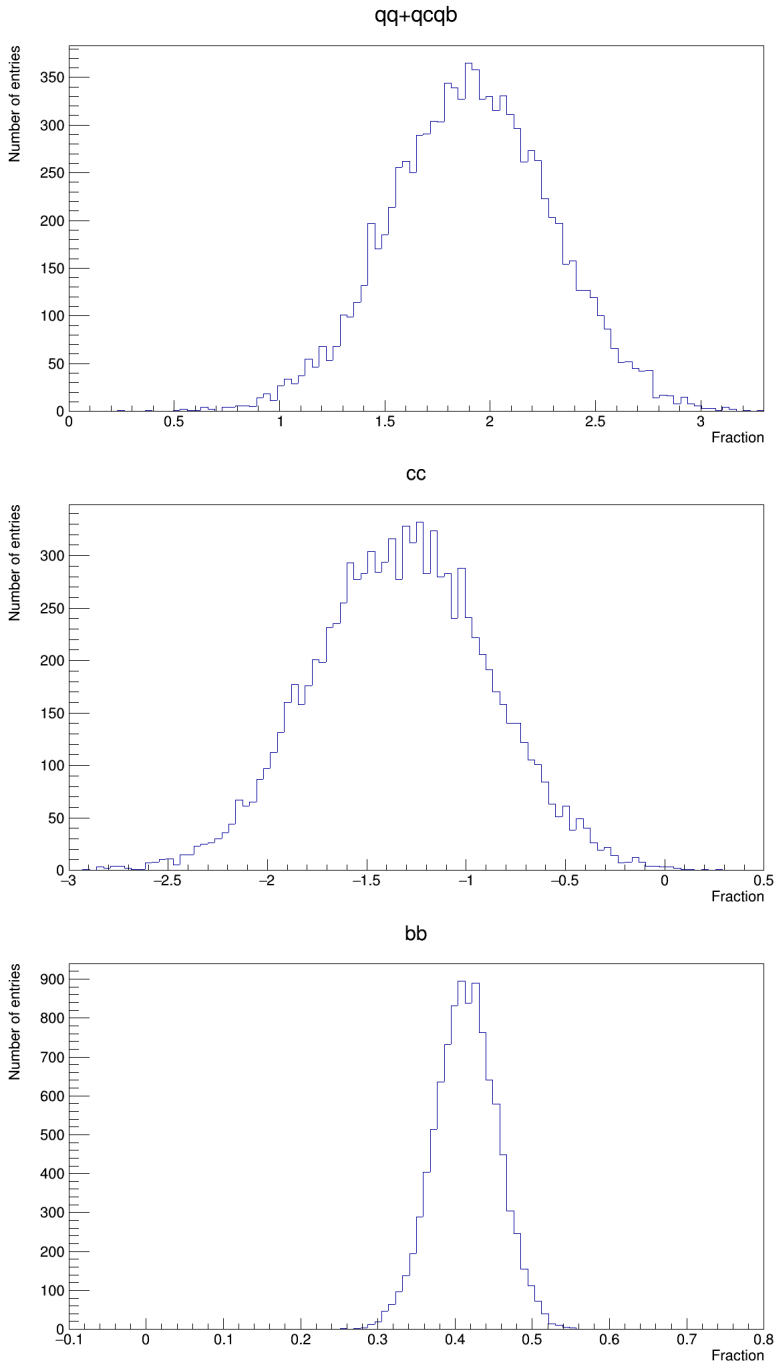


Figure 7.1: Distributions of F_V after subtraction, MC

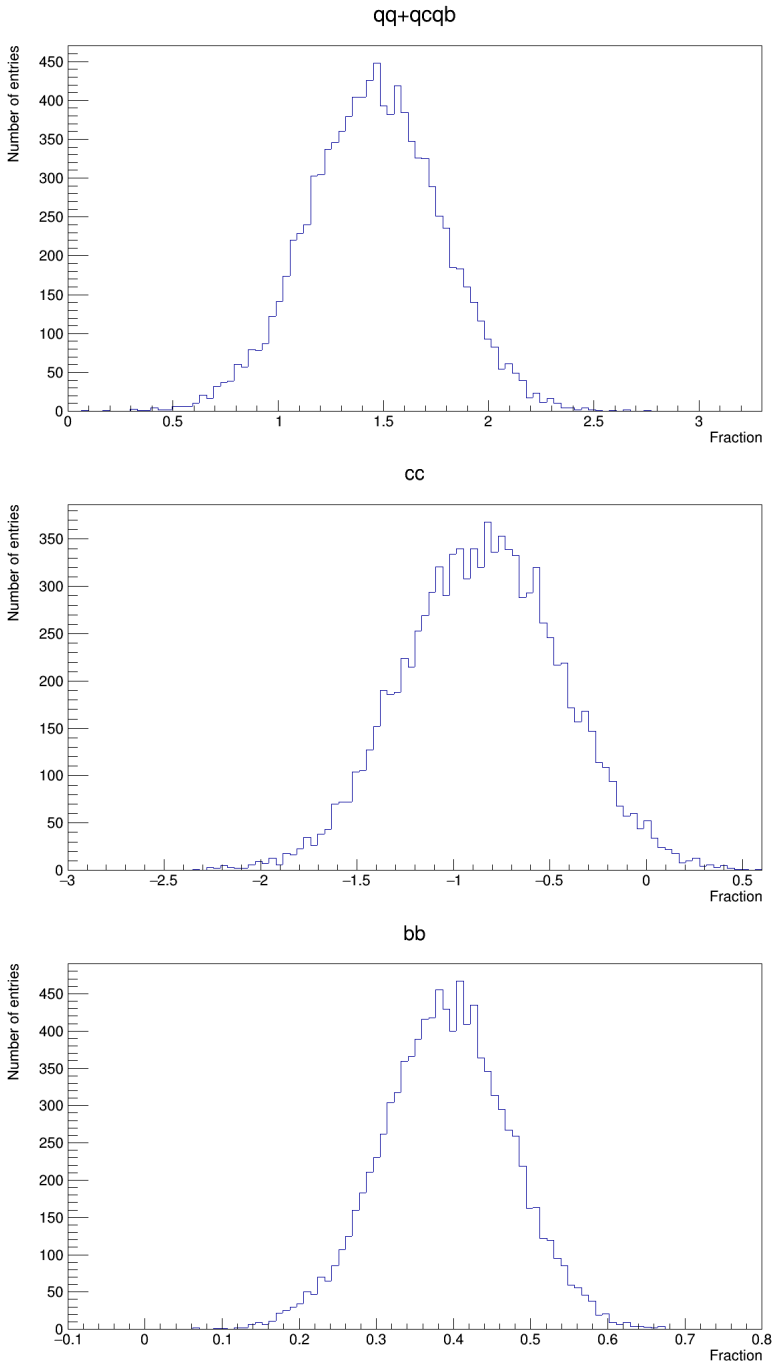


Figure 7.2: Distributions of F_V after subtraction, Data

Chapter 8

Modifications

This chapter examines possible modifications of the output of the flavour tagging algorithm and how they could affect the results of the previous chapters. The attempt at finding the branching ratios inclusively is repeated with modified templates.

8.1 Modified MV2c10 output

As it was concluded in the previous chapter, the developed method is not feasible with the present accuracy of the MV2c10 algorithm. It is now interesting to see how the results will change by modifying the MV2c10 output to increase the discrimination between c-, b- and light-flavour jets. A simple way to have a systematic modification of the output, would be to change the probability according to the flavour of the jet. Ideally the output for each flavour would be a distribution centred around a specific probability value, whose width would decrease as the discrimination improves. This is presently the case for the light-flavour jets and b-jets but not for the c-jets. Considering a shifted output of the algorithm so that the values are between 0 and 1, the three centres can be chosen as 0, 0.5 and 1, respectively. Based on this, the output is changed by modifying the probability as:

$$pr' = cent - dist * a \tag{8.1}$$

with

$$dist = cent - pr \tag{8.2}$$

where pr, pr' are the unmodified and modified probability, respectively, $cent$ is the centre value for each flavour and a is a number ranging from 0 to 1, which determines how much the probability is modified. For $a = 1$, the MV2c10 output remains the same. The di-jet discriminants can then be found using the same formula as before (Equation 6.1). This way, the flavour templates can be formed, while also taking all the correlations into account.

Figure 8.1 shows the light-flavour jet and c-jet rejection factors as a function of the b-jet tagging efficiency. The dashed lines correspond to the unmodified output, while the solid ones correspond to modified cases for different values of a. Examining the plot from left to right, a takes the values: 0.95, 0.9, 0.8, 0.7, 0.6, 0.5. The changes to the output cause the rejection rate for both the light-flavour jets and the c-jets to increase significantly and saturate as the b-jet efficiency decreases. As a decreases, the discrimination between

the flavours is improved, so the saturation continues for even larger b-jet efficiency values. For $a = 0.5$, the light-flavour jet rejection rate obtains its peak value regardless of the b-jet efficiency. This is an ideal but practically unachievable case.

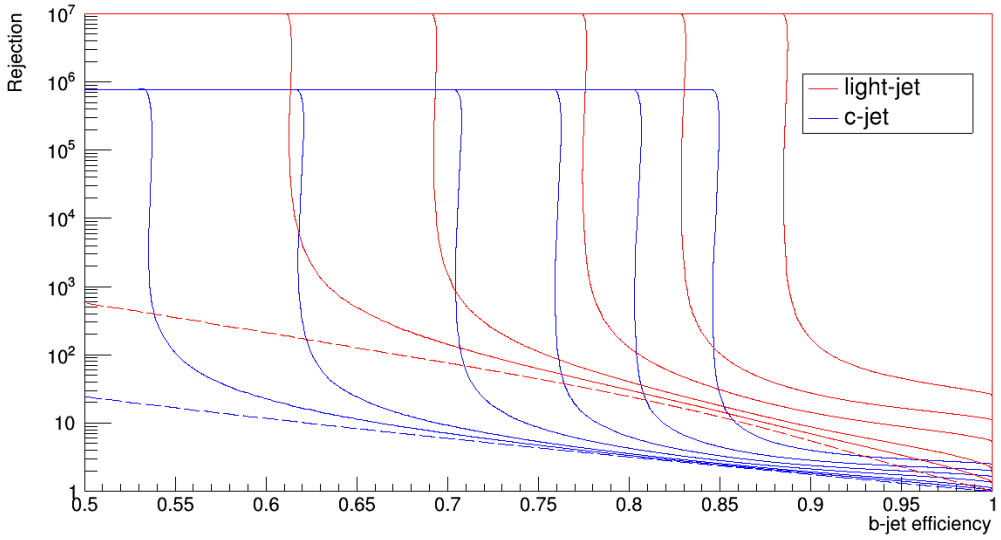
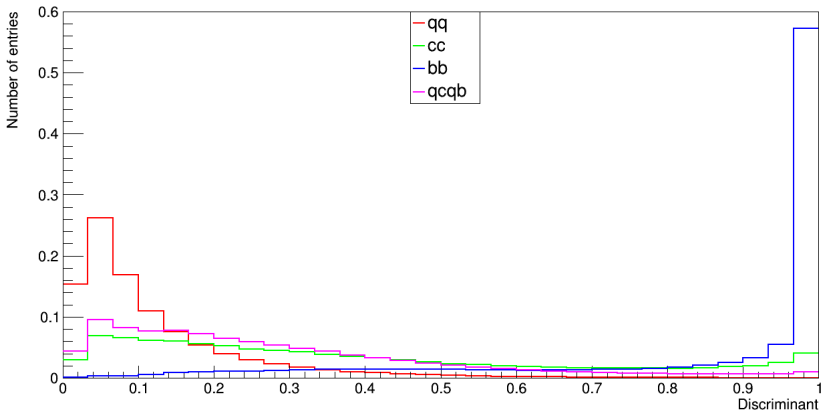
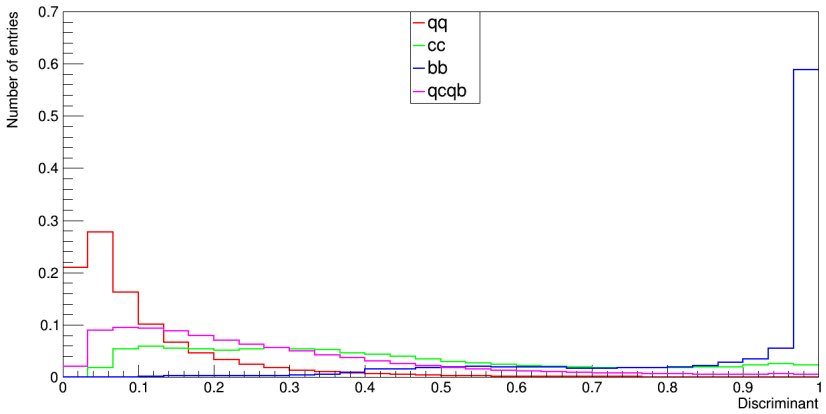


Figure 8.1: Rejection factors as a function of the b-jet tagging efficiency of the modified and unmodified MV2c10 output

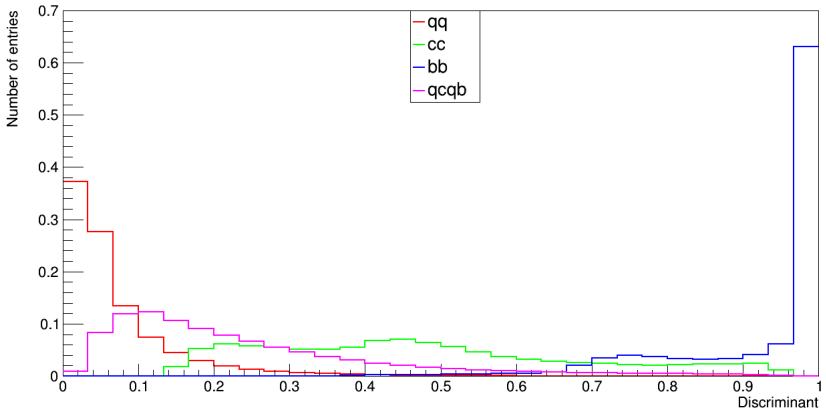
The four main templates used in the present study are shown in Figure 8.2 for different values of a. In the unmodified case ($a = 1$), there is significant overlap between the templates and that is why the fit method is not reliable. For $a = 0.7$, the discrimination becomes clear, as all the templates exhibit their peak value at different points.



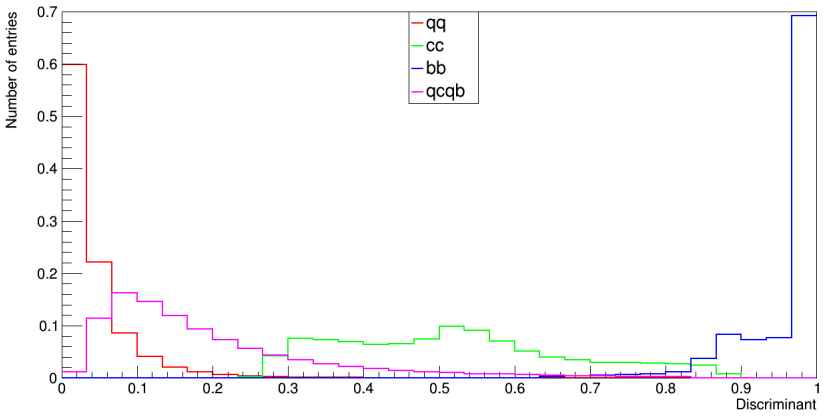
(a) $a = 1$



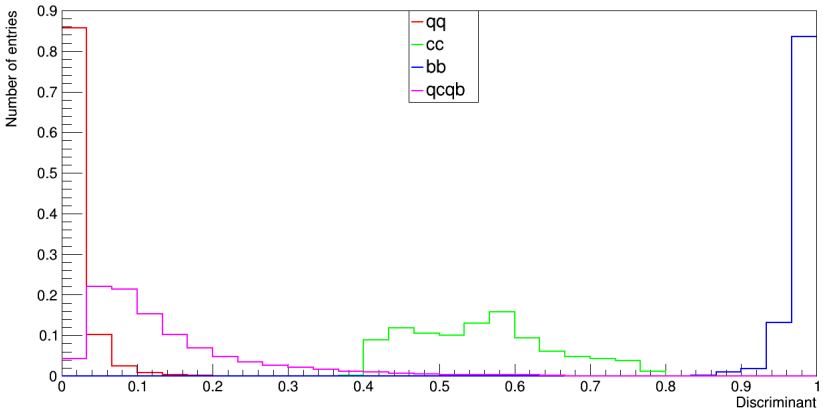
(b) $a = 0.9$



(c) $a = 0.7$



(d) $a = 0.5$



(e) $a = 0.3$

Figure 8.2: Modified flavour templates for different values of a

8.2 Applications

8.2.1 ZW signal region

The modified templates can first be used to test the method in the ZW signal region. This time the sample is in the $p_T = 250 - 450$ GeV region and the rates are:

$$R_{QCD} = 0.9358, \quad R_V = 0.0642 \tag{8.3}$$

The relative rate of signal to background is larger than in the case of the previous chapter and this will affect the overall errors of the subtraction. The modifications of Equation 8.1 are also applied to the background and signal distributions. First, the subtraction is attempted with the original histograms, without the modifications, so as to be able to compare the results. The true and fitted values for the fractions of the background are shown in Table 8.1. In this p_T region, the fit is unable to extract the cc contribution. The background composition is found using a linear interpolation between the two mass regions. The results of the subtraction are shown in Table 8.2. The errors are indeed smaller with this sample and now the fractions have physical values. This can be attributed to the better flavour tagging performance in the first p_T region compared to the others (Figure 6.3). Nevertheless, the cc fraction cannot be estimated.

	m = 40-70 GeV		m = 70-105 GeV	m = 200-300 GeV	
	True	Fit ($r = 0.8720$)	True	True	Fit ($r = 0.7999$)
$fr_{qq+qcqb}$	0.9514	0.9792 ± 0.0002	0.9529	0.9683	0.9868 ± 0.0010
fr_{cc}	0.0265	$\approx 0 \pm 0.00002$	0.0243	0.0145	$\approx 0 \pm 0.0002$
fr_{bb}	0.0185	0.0208 ± 0.0003	0.0187	0.0114	0.0132 ± 0.0010

Table 8.1: QCD fits with unmodified MV2c10, $p_T = 250 - 450$ GeV

	BR	SR ($r = 0.8478$)	F_V
$fr_{qq+qcqb}$	0.9805 ± 0.0010	0.9684 ± 0.0004	0.7932 ± 0.1231
fr_{cc}	$\approx 0 \pm 0.00004$	$\approx 0 \pm 0.00005$	$\approx 0 \pm 0.0007$
fr_{bb}	0.0195 ± 0.0010	0.0316 ± 0.0004	0.2068 ± 0.0159

Table 8.2: Results of the method for the ZW signal region, $p_T = 250 - 450$ GeV

The next step is to see how the QCD fits will be affected by the modified templates. Tables 8.3 and 8.4 show the results of the fit in the low and high mass regions for different values of a. For $a = 0.9$, there is no improvement, so it is not useful to further consider this case, but for $a = 0.7$ the cc fraction is no longer zero in the low mass region. The next two values of a reflect the true flavour composition much better and with greater accuracy. These results are linearly interpolated to give an estimate of the background in the signal region (Table 8.5). The fit results for the signal region are shown in Table 8.6. Finally, the overall fractions for the vector bosons F_V are presented in Table 8.7 along with the MC predicted values. For $a = 0.5$ and $a = 0.3$, they agree rather well with the predicted results. The two cases have almost the same accuracy. Compared with the $p_T = 450 - 650$ GeV region, the errors have decreased as there are more ZW events.

	Fit ($a = 0.9$)	Fit ($a = 0.7$)	Fit ($a = 0.5$)	Fit ($a = 0.3$)
$fr_{qq+qcqb}$	0.9786 ± 0.0002	0.9637 ± 0.0005	0.9547 ± 0.0003	0.9535 ± 0.0003
fr_{cc}	$\approx 0 \pm 0.00003$	0.0160 ± 0.0005	0.0257 ± 0.0003	0.0275 ± 0.0002
fr_{bb}	0.0214 ± 0.0002	0.0204 ± 0.0002	0.0196 ± 0.0002	0.0190 ± 0.0002

Table 8.3: QCD fits with modified MV2c10, Mass 40-70 GeV, $p_T = 250 - 450$ GeV

	Fit ($a = 0.9$)	Fit ($a = 0.7$)	Fit ($a = 0.5$)	Fit ($a = 0.3$)
$fr_{qq+qcqb}$	0.9861 ± 0.0009	0.9852 ± 0.0008	0.9748 ± 0.0014	0.9742 ± 0.0010
fr_{cc}	$\approx 0 \pm 0.0002$	$\approx 0 \pm 0.0011$	0.0116 ± 0.0013	0.0138 ± 0.0008
fr_{bb}	0.0139 ± 0.0009	0.0148 ± 0.0013	0.0135 ± 0.0007	0.0120 ± 0.0006

Table 8.4: QCD fits with modified MV2c10, Mass 200-300 GeV, $p_T = 250 - 450$ GeV

	BR $a = 0.7$	BR $a = 0.5$	BR $a = 0.3$
$fr_{qq+qcqb}$	0.9673 ± 0.0028	0.9580 ± 0.0026	0.9570 ± 0.0026
fr_{cc}	0.0133 ± 0.0021	0.0234 ± 0.0018	0.0252 ± 0.0018
fr_{bb}	0.0194 ± 0.0008	0.0186 ± 0.0008	0.0178 ± 0.0009

Table 8.5: QCD background estimation with modified MV2c10, Mass $70 - 105$ GeV, $p_T = 250 - 450$ GeV

	SR $a = 0.7$	SR $a = 0.5$	SR $a = 0.3$
$fr_{qq+qcqb}$	0.9531 ± 0.0005	0.9422 ± 0.0004	0.9410 ± 0.0003
fr_{cc}	0.0163 ± 0.0005	0.0292 ± 0.0004	0.0315 ± 0.0003
fr_{bb}	0.0306 ± 0.0003	0.0287 ± 0.0002	0.0274 ± 0.0002

Table 8.6: Signal region fits with modified MV2c10, Mass $70 - 105$ GeV, $p_T = 250 - 450$ GeV

	F_V True	F_V $a = 0.7$	F_V $a = 0.5$	F_V $a = 0.3$
$fr_{qq+qcqb}$	0.7251 ± 0.0062	0.7463 ± 0.1272	0.7111 ± 0.1251	0.7092 ± 0.1251
fr_{cc}	0.1292 ± 0.0021	0.0606 ± 0.0314	0.1137 ± 0.0272	0.1231 ± 0.0261
fr_{bb}	0.1427 ± 0.0022	0.1930 ± 0.0124	0.1752 ± 0.0126	0.1676 ± 0.0139

Table 8.7: F_V with modified MV2c10, $p_T = 250 - 450$ GeV

8.2.2 Higgs signal region

Similarly to the ZW case, the MV2c10 output for the variable radius jets of the Higgs selected events, can be modified according to Equation 8.1. The results of the QCD fits are shown in Tables 8.8 and 8.9 and differ slightly from Tables 8.3 and 8.4 because there are fewer background events taken into consideration. Linearly interpolating in the Higgs signal region, gives the background estimation of Table 8.11. The results agree more with the MC predicted composition of Table 7.9 as a decreases. There is no significant improvement between $a = 0.5$ and $a = 0.3$. The fit values in the signal region of Table 8.11 indicate that in order to get a non-negligible value for the cc fraction, a considerable increase in the discrimination between the different flavour jets is needed. This is also reflected in the values of the branching ratios after subtracting the background (Table 8.12). Contrary to the ZW signal region, the fractions do not have physical values even in the case of $a = 0.3$, due to the small signal to background ratio. Nevertheless, physical branching ratios are obtained by considering only the physical parts of the distributions (Figures 8.3 - 8.5) for two of the three fractions. The results are shown in Table 8.13. For $a = 0.3$, a good estimate of the branching ratios is achieved.

	Fit ($a = 0.7$)	Fit ($a = 0.5$)	Fit ($a = 0.3$)
$fr_{qq+qcqb}$	0.9650 ± 0.0006	0.9561 ± 0.0004	0.9545 ± 0.0004
fr_{cc}	0.0159 ± 0.0006	0.0251 ± 0.0004	0.0273 ± 0.0003
fr_{bb}	0.0191 ± 0.0003	0.0187 ± 0.0002	0.0182 ± 0.0002

Table 8.8: QCD fits with modified MV2c10, Mass 40-70 GeV, $p_T = 250 - 450$ GeV, Leading jet

	Fit ($a = 0.7$)	Fit ($a = 0.5$)	Fit ($a = 0.3$)
$fr_{qq+qcqb}$	0.9841 ± 0.0026	0.9729 ± 0.0018	0.9725 ± 0.0013
fr_{cc}	0.0009 ± 0.0025	0.0130 ± 0.0016	0.0149 ± 0.0011
fr_{bb}	0.0151 ± 0.0010	0.0140 ± 0.0009	0.0125 ± 0.0008

Table 8.9: QCD fits with modified MV2c10, Mass 200-300 GeV, $p_T = 250 - 450$ GeV, Leading jet

	BR $a = 0.7$	BR $a = 0.5$	BR $a = 0.3$
$fr_{qq+qcqb}$	0.9718 ± 0.0032	0.9622 ± 0.0028	0.9610 ± 0.0030
fr_{cc}	0.0105 ± 0.0026	0.0208 ± 0.0021	0.0228 ± 0.0020
fr_{bb}	0.0177 ± 0.0008	0.0171 ± 0.0009	0.0162 ± 0.0010

Table 8.10: QCD background estimation with modified MV2c10, Mass $105 - 145$ GeV, $p_T = 250 - 450$ GeV

Based on all the aforementioned observations, it can be concluded that as the overlap between the templates decreases, the method developed in this study produces more accurate results. Therefore, an inclusive measurement of the branching ratios is possible and directly related to the improvement of the MV2c10 algorithm and especially to the increase in the discrimination between the c-jets and the other two flavours.

	SR $a = 0.7$	SR $a = 0.5$	SR $a = 0.3$
$fr_{qq+qcqb}$	0.9633 ± 0.0008	0.9531 ± 0.0006	0.9525 ± 0.0005
fr_{cc}	0.0078 ± 0.0008	0.0198 ± 0.0005	0.0219 ± 0.0004
fr_{bb}	0.0289 ± 0.0004	0.0272 ± 0.0004	0.0256 ± 0.0004

Table 8.11: Signal region fits with modified MV2c10, Mass $105-145$ *GeV,* $p_T = 250-450$ *GeV*

	F_H True	F_H $a = 0.7$	F_H $a = 0.5$	F_H $a = 0.3$
$fr_{qq+qcqb}$	0.1442 ± 0.0084	0.2686 ± 2.3557	0.2133 ± 2.3275	0.2631 ± 2.3265
fr_{cc}	0.0499 ± 0.0047	-0.2147 ± 0.2261	-0.0627 ± 0.1802	-0.0518 ± 0.1775
fr_{bb}	0.8059 ± 0.0249	0.9462 ± 0.0938	0.8494 ± 0.0943	0.7887 ± 0.0999

Table 8.12: F_H *with modified MV2c10,* $p_T = 250 - 450$ *GeV*

	F_H True	F_H $a = 0.7$	F_H $a = 0.5$	F_H $a = 0.3$
$fr_{qq+qcqb}$	0.1442 ± 0.0084	$\approx 0 \pm 0.0599$	0.0725 ± 0.0726	0.1214 ± 0.0872
fr_{cc}	0.0499 ± 0.0047	0.0626 ± 0.0538	0.0780 ± 0.0627	0.0898 ± 0.0711
fr_{bb}	0.8059 ± 0.0249	0.9444 ± 0.0264	0.8495 ± 0.0365	0.7888 ± 0.0505

Table 8.13: Physical F_H *with modified MV2c10,* $p_T = 250 - 450$ *GeV*

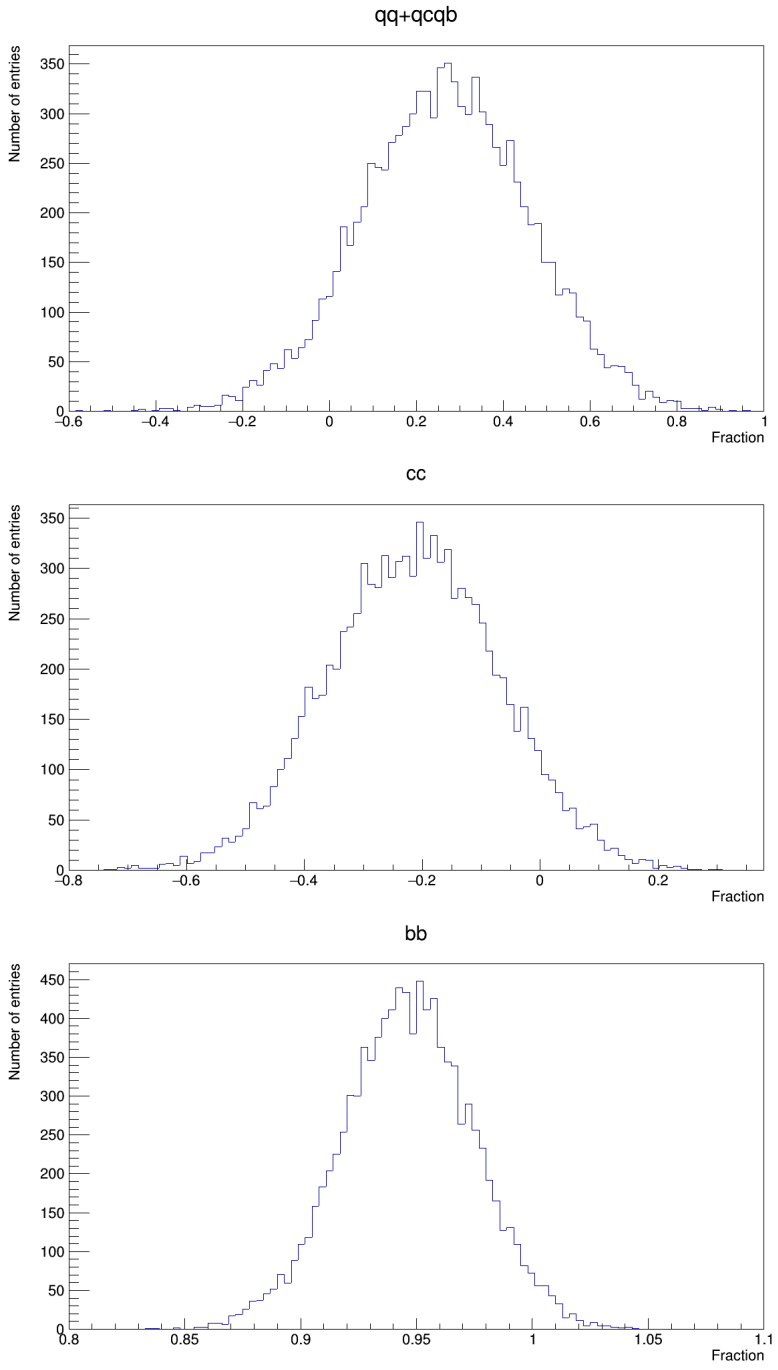


Figure 8.3: Distributions of F_H after subtraction, $a = 0.7$

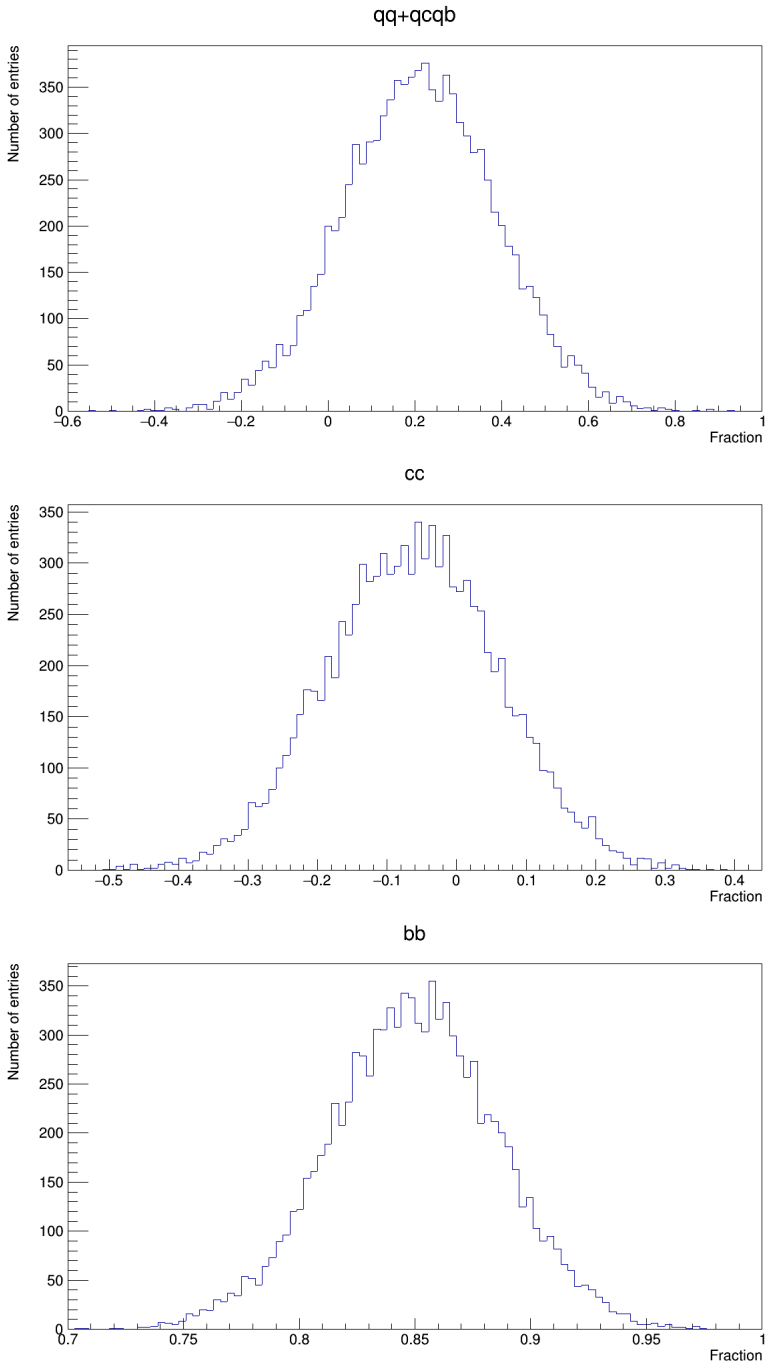


Figure 8.4: Distributions of F_H after subtraction, $a = 0.5$

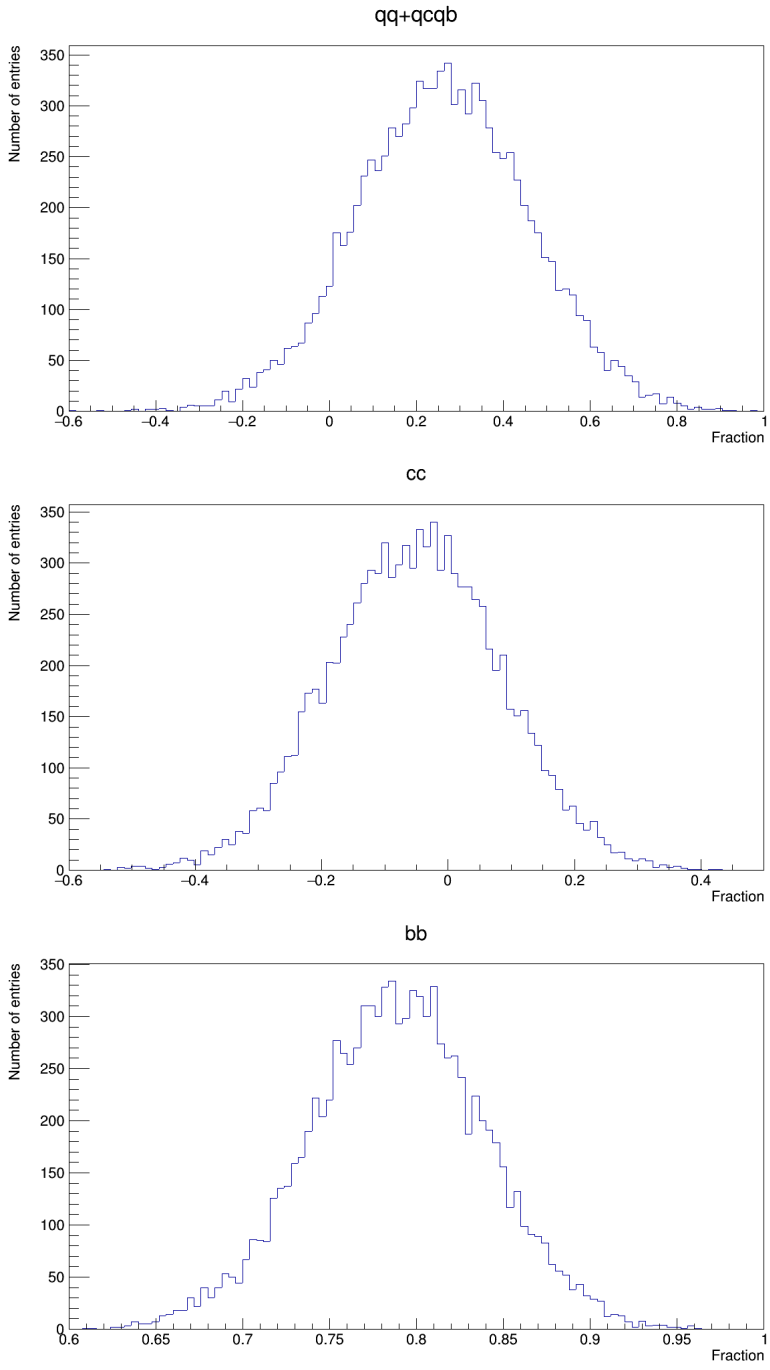


Figure 8.5: Distributions of F_H after subtraction, $a = 0.3$

Chapter 9

Conclusions

Considering all the results of the previous chapters, several conclusions can be drawn. First of all, Chapter 5 showed that it is possible to fit three templates to find the fraction of each in a given sample. The accuracy of the fit is directly related to the number of entries of the sample. In practice, this means that the proposed method would benefit from more statistics. Chapter 6 provided a better understanding of the QCD background by examining its flavour composition. It is important to know how exactly it changes with respect to the mass and the p_T in order to model it efficiently, as it is a sizeable background for many other processes of interest and new physics searches. Furthermore, it was shown that the QCD background has a significant percentage of pairs of light and heavy flavour quarks, so a fourth template representing them, is necessary for describing it effectively. Chapter 7 demonstrated that a simple side-band subtraction of the background is not viable and interpolating between two background regions is a better alternative. The present study determined that it is not possible to conduct an inclusive measurement of the hadronic branching ratios with the current flavour tagging output and that such a method would have limited accuracy. The inability to discriminate well the c-quark from the others was reflected in the results and was exacerbated by the introduction of the fourth template. Chapter 8 demonstrated that an inclusive measurement is possible and directly related to the relation between the rejection rates and the efficiency of the MV2c10 algorithm. The resulting fractions are in accordance with the MC predicted values for significantly improved tagging performance. The results of this thesis for the Higgs branching ratios could be expanded by considering the Z/W associated production.

It is worth mentioning the future prospects for measuring all the branching ratios and the impact this would have. In recent years, there have been direct searches for the decay to charm quarks which are a vital step in fully understanding the mass coupling [44]. The attempt at an inclusive measurement will also benefit from c-tagging techniques which seem feasible at the High-Luminosity LHC [43]. Finally, finding absolute branching ratios is an integral part of future linear collider programmes in order to determine the Higgs boson nature [45, 46].

Bibliography

[1] David J Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. New York, NY: Wiley, 2008. URL: https://cds.cern.ch/record/111880.

[2] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[3] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[4] Peter W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Physics Letters* 12.2 (1964), pp. 132–133. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)91136-9. URL: http://www.sciencedirect.com/science/article/pii/0031916364911369.

[5] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139525367.

[6] John Ellis. *Higgs Physics*. 2013. arXiv: 1312.5672 [hep-ph].

[7] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. "Broken Symmetries". In: *Phys. Rev.* 127 (3 Aug. 1962), pp. 965–970. DOI: 10.1103/PhysRev.127.965. URL: https://link.aps.org/doi/10.1103/PhysRev.127.965.

[8] Sheldon L. Glashow. "The renormalizability of vector meson interactions". In: *Nucl. Phys.* 10 (1959), pp. 107–117. DOI: 10.1016/0029-5582(59)90196-8.

[9] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc.* C680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[10] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[11] Stephen P. Martin. "A SUPERSYMMETRY PRIMER". In: *Advanced Series on Directions in High Energy Physics* (July 1998), pp. 1–98. ISSN: 1793-1339. DOI: 10.1142/9789812839657_0001. URL: http://dx.doi.org/10.1142/9789812839657_0001.

[12] A. Djouadi. "The anatomy of electroweak symmetry breaking Tome II: The Higgs bosons in the Minimal Supersymmetric Model". In: *Physics Reports* 459.1-6 (Apr. 2008), pp. 1–241. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2007.10.005. URL: http://dx.doi.org/10.1016/j.physrep.2007.10.005.

[13] Marco Battaglia et al. "The Higgs sector of the phenomenological MSSM in the light of the Higgs boson discovery". In: *Journal of High Energy Physics* 2012.9 (Sept. 2012). ISSN: 1029-8479. DOI: 10.1007/jhep09(2012)107. URL: http://dx.doi.org/10.1007/JHEP09(2012)107.

[14] P. Bechtle et al. "The light and heavy Higgs interpretation of the MSSM". In: *The European Physical Journal C* 77.2 (Feb. 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-4584-9. URL: http://dx.doi.org/10.1140/epjc/s10052-016-4584-9.

[15] D. de Florian et al. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. 869 pages, 295 figures, 248 tables and 1645 citations. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG. Oct. 2016. DOI: 10.23731/CYRM-2017-002. URL: https://cds.cern.ch/record/2227475.

[16] C. Amsler et al. URL: http://pdg.lbl.gov/2009/tables/rpp2009-sum-gauge-higgs-bosons.pdf.

[17] G. 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: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020. URL: http://dx.doi.org/10.1016/j.physletb.2012.08.020.

[18] S. Chatrchyan et al. "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), pp. 30–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021. URL: http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[19] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[20] G. Aad et al. "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and $\sqrt{s} = 8$ TeV with the ATLAS and CMS Experiments". In: *Physical Review Letters* 114.19 (May 2015). ISSN: 1079-7114. DOI: 10.1103/physrevlett.114.191803. URL: http://dx.doi.org/10.1103/PhysRevLett.114.191803.

[21] M. Aaboud et al. "Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector". In: *Phys. Lett. B* 784.arXiv:1806.00242 (June 2018), 345–366. 22 p. DOI: 10.1016/j.physletb.2018.07.050. URL: https://cds.cern.ch/record/2621113.

[22] CMS Collaboration. *A measurement of the Higgs boson mass in the diphoton decay channel*. 2020. arXiv: 2002.06398 [hep-ex].

[23] M. Aaboud et al. "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector". In: *Phys. Lett. B* 786.arXiv:1808.08238 (Aug. 2018), 59–86. 28 p. DOI: 10.1016/j.physletb.2018.09.013. URL: https://cds.cern.ch/record/2636066.

[24] G. Aad et al. "Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment". In: *Phys. Rev. D* 101.arXiv:1909.02845 (Sept. 2019), 012002. 48 p. DOI: 10.1103/PhysRevD.101.012002. URL: https://cds.cern.ch/record/2688596.

[25] Marco Battaglia et al. "An update of the constraints on the phenomenological MSSM from the new LHC Higgs results". In: *Physics Letters B* 720.1-3 (Mar. 2013), pp. 153–160. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2013.02.001. URL: http://dx.doi.org/10.1016/j.physletb.2013.02.001.

[26] Alexandre Arbey, Marco Battaglia, and Farvah Mahmoudi. "Supersymmetric heavy Higgs bosons at the LHC". In: *Physical Review D* 88.1 (July 2013). ISSN: 1550-2368. DOI: 10.1103/physrevd.88.015007. URL: http://dx.doi.org/10.1103/PhysRevD.88.015007.

[27] G. Aad et al. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08003. 437 p. DOI: 10.1088/1748-0221/3/08/S08003. URL: https://cds.cern.ch/record/1129811.

[28] *ATLAS magnet system: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/338080.

[29] G. Aad et al. "The ATLAS Inner Detector commissioning and calibration". In: *The European Physical Journal C* 70.3 (Aug. 2010), pp. 787–821. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-010-1366-7. URL: http://dx.doi.org/10.1140/epjc/s10052-010-1366-7.

[30] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331061.

[31] *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331062.

[32] *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331068.

[33] Christian Lippmann. "Particle identification". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 666 (Feb. 2012), pp. 148–172. ISSN: 0168-9002. DOI: 10.1016/j.nima.2011.03.009. URL: http://dx.doi.org/10.1016/j.nima.2011.03.009.

[34] T. Cornelissen et al. *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*. Tech. rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002. Geneva: CERN, Mar. 2007. URL: https://cds.cern.ch/record/1020106.

[35] G. Aad et al. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5004-5. URL: http://dx.doi.org/10.1140/epjc/s10052-017-5004-5.

[36] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-ktjet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[37] David Krohn, Jesse Thaler, and Lian-Tao Wang. "Jets with variable R". In: *Journal of High Energy Physics* 2009.06 (June 2009), pp. 059–059. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2009/06/059. URL: http://dx.doi.org/10.1088/1126-6708/2009/06/059.

[38] *Measurements of the Higgs boson inclusive, differential and production cross sections in the 4ℓ decay channel at $\sqrt{s}$ = 13 TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2019-025. Geneva: CERN, July 2019. URL: http://cds.cern.ch/record/2682107.

[39] *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. Tech. rep. ATL-PHYS-PUB-2017-013. Geneva: CERN, July 2017. URL: https://cds.cern.ch/record/2273281.

[40] M. Aaboud et al. "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). ISSN: 1029-8479. DOI: 10.1007/jhep08(2018)089. URL: http://dx.doi.org/10.1007/JHEP08(2018)089.

[41] M. Aaboud et al. "Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector". In: *Journal of High Energy Physics* 2017.12 (Dec. 2017). ISSN: 1029-8479. DOI: 10.1007/jhep12(2017)024. URL: http://dx.doi.org/10.1007/JHEP12(2017)024.

[42] F. James and M Winkler. *MINUIT User's Guide*. http://inspirehep.net/record/1258345/files/mnusersguide.pdf.

[43] *Prospects for $H \to c\bar{c}$ using Charm Tagging with the ATLAS Experiment at the HL-LHC*. Tech. rep. ATL-PHYS-PUB-2018-016. Geneva: CERN, Aug. 2018. URL: http://cds.cern.ch/record/2633635.

[44] M. Aaboud et al. "Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment". In: *Physical Review Letters* 120.21 (May 2018). ISSN: 1079-7114. DOI: 10.1103/physrevlett.120.211802. URL: http://dx.doi.org/10.1103/PhysRevLett.120.211802.

[45] Marco Battaglia. *Measuring Higgs Branching Ratios and telling the SM from a MSSM Higgs Boson at the e+e- Linear Collider*. 1999. arXiv: hep-ph/9910271 [hep-ph].

[46] Yambazi Banda, Tomáš Laštovička, and Andrei Nomerotski. "Higgs boson hadronic branching ratios at the ILC". In: *Physical Review D* 82.3 (Aug. 2010). ISSN: 1550-2368. DOI: 10.1103/physrevd.82.033013. URL: http://dx.doi.org/10.1103/PhysRevD.82.033013.