

Dissertation  
submitted to the  
Combined Faculties of the Natural Sciences and  
Mathematics  
of the Ruperto-Carola-University of Heidelberg,  
Germany  
for the degree of  
Doctor of Natural Sciences

Put forward by  
Merve Şahinsoy  
born in Balıkesir

Oral examination on May 22, 2019



Identification of Higgs to  $b\bar{b}$  decays in  
boosted topologies and searches for  
resonances with boosted  $b$ -quark pairs  
in their final state with the ATLAS  
detector

Referees: Priv.-Doz. Dr. Oleg Brandt  
Prof. Dr. Stephanie Hansmann-Menzemer



## Abstract

With the discovery of the 125 GeV Higgs boson, both the measurements of Higgs properties and the searches for new physics with Higgs bosons moved into focus of the physics programme of high energy experiments. The most frequent decay channel of the Higgs boson, the  $h \rightarrow b\bar{b}$  process, is one of the natural choices to search for anomalous Higgs production. Focusing on this decay channel in boosted topologies, this thesis discusses the improved identification techniques for Higgs boson, and presents two searches using those techniques. The first analysis searches for pair-production of boosted Higgs bosons in the  $b\bar{b}b\bar{b}$  final state with  $36.1 \text{ fb}^{-1}$  of ATLAS  $pp$  data collected in 2015 and 2016 at  $\sqrt{s} = 13 \text{ TeV}$ . The analysis looks for resonances in the invariant mass region of Higgs pairs between 800 GeV and 3000 GeV. The second analysis searches for boosted  $b\bar{b}$  pairs produced in association with an additional initial state radiation jet using  $80.5 \text{ fb}^{-1}$  of ATLAS  $pp$  data collected in 2015, 2016 and 2017 at  $\sqrt{s} = 13 \text{ TeV}$ . Targeting the dijet invariant mass spectrum below 230 GeV, the analysis searches for new resonances predicted by simplified dark matter models as well as Standard Model boosted Higgs boson production. No indication of new physics observed in both analyses, and the results are used to constrain the phase space for beyond the Standard Model physics.

## Kurzzusammenfassung

Nach der Entdeckung des 125 GeV Higgs-Bosons konzentrierten sich die Hochenergieexperimente auf Messungen von Higgs-Eigenschaften sowie auf die Suche nach neuer Physik mit dem Higgs-Boson. Der häufigste Zerfallskanal des Higgs-Bosons, der  $h \rightarrow b\bar{b}$  Prozess, ist eine der natürlichen Möglichkeiten, um nach Higgs-Produktion zu suchen. Diese Arbeit legt den Fokus auf diesen Zerfallskanal in geboosteten Topologien und dokumentiert die Entwicklung neuer und verbesserter Identifikationstechniken für das Higgs-Boson, deren erfolgreiche Implementierung im Rahmen zweier Higgs-bezogener Suchen im ATLAS-Experiment demonstriert wird. Die erste Analyse sucht nach der Paarproduktion geboosterter Higgs-Bosonen im Endzustand  $b\bar{b}b\bar{b}$  mit  $36.1 \text{ fb}^{-1}$  ATLAS  $pp$  Daten, die 2015 und 2016 bei  $\sqrt{s} = 13 \text{ TeV}$  aufgezeichnet wurden. Die Analyse sucht nach Anzeichen für neue Physik im invarianten Massenbereich von Higgs-Paaren zwischen 800 GeV und 3000 GeV. Die zweite Analyse sucht nach geboosteten  $b\bar{b}$ -Paaren, die in Verbindung mit einem zusätzlichen im Initialzustand abgestrahltem Jet erzeugt wurden, wobei  $80.5 \text{ fb}^{-1}$  von ATLAS  $pp$  Daten verwendet wurden, die 2015, 2016 und 2017 bei  $\sqrt{s} = 13 \text{ TeV}$  aufgezeichnet wurden. Die Analyse konzentriert sich auf das niedrige Dijet Massenspektrum und sucht nach Resonanzen, die durch simplifizierte Modelle für dunkle Materie sowie durch die Higgs-Boson-Produktion im Standardmodell vorhergesagt werden. In keiner der Analysen werden Anzeichen neuer Physik beobachtet, und die Ergebnisse werden verwendet, um den Phasenraum für Physik jenseits des Standardmodells zu beschränken.



# Contents

<b>1</b>	<b>Overview</b>	<b>1</b>
<b>2</b>	<b>Theory</b>	<b>5</b>
2.1	Standard Model . . . . .	5
2.1.1	Fundamental particles and forces . . . . .	6
2.1.2	Quantum chromodynamics . . . . .	9
2.1.3	Electroweak theory and symmetry breaking . . . . .	10
2.1.4	The Higgs Boson . . . . .	11
2.2	Limitations of the Standard Model . . . . .	13
2.2.1	Gravity . . . . .	13
2.2.2	Dark Matter . . . . .	13
2.2.3	Hierarchy problem . . . . .	14
2.3	Beyond the Standard Model . . . . .	15
2.3.1	Extra dimension searches . . . . .	15
2.3.2	Dark matter searches . . . . .	16
2.3.3	Searches in the Higgs sector . . . . .	18
<b>3</b>	<b>The ATLAS Experiment</b>	<b>21</b>
3.1	The Large Hadron Collider . . . . .	21
3.2	The ATLAS detector . . . . .	24
3.2.1	Inner detector . . . . .	25
3.2.2	Calorimeters . . . . .	27
3.2.3	Muon spectrometer . . . . .	29

3.2.4	Trigger and data acquisition system . . . . .	30
<b>4</b>	<b>Data and Simulations</b>	<b>33</b>
4.1	Data . . . . .	33
4.2	Monte Carlo simulations . . . . .	34
4.2.1	Simulations of Standard Model processes . . . . .	34
4.2.2	Simulations of Beyond the Standard Model processes . . . . .	36
<b>5</b>	<b>Reconstruction</b>	<b>39</b>
5.1	Jets . . . . .	40
5.1.1	Jet algorithms . . . . .	41
5.1.2	Jet grooming . . . . .	44
5.1.3	Small- $R$ jets . . . . .	47
5.1.4	Large- $R$ jets . . . . .	48
5.1.5	Track jets . . . . .	50
5.1.6	$b$ -jets . . . . .	50
5.2	Muons . . . . .	52
<b>6</b>	<b>Boosted <math>h \rightarrow b\bar{b}</math> identification</b>	<b>55</b>
6.1	Boosted $h \rightarrow b\bar{b}$ tagger . . . . .	57
6.1.1	Large- $R$ jet selection . . . . .	57
6.1.2	Track jet selection . . . . .	58
6.1.3	Jet substructure and mass . . . . .	60
6.2	Higgs jet mass . . . . .	64
6.2.1	Combined mass for Higgs jets . . . . .	64
6.2.2	Muon correction with combined mass . . . . .	67
6.2.3	Higgs jet mass window . . . . .	70
6.3	Alternative Higgs jet reconstruction techniques . . . . .	76
6.4	Conclusion . . . . .	80

---

<b>7</b>	<b>Search for boosted di-Higgs production in the <math>b\bar{b}b\bar{b}</math> final state</b>	<b>81</b>
7.1	Overview . . . . .	81
7.1.1	Searching for new physics . . . . .	81
7.1.2	SM non-resonant Higgs boson pair production . . . . .	82
7.1.3	Analysis strategy . . . . .	83
7.2	Event Selection . . . . .	85
7.2.1	Trigger . . . . .	85
7.2.2	Preselection . . . . .	85
7.2.3	Final Event Categorisation . . . . .	86
7.2.4	Impact of jet mass choice . . . . .	88
7.3	Background modelling . . . . .	91
7.3.1	Data driven multijet background estimation . . . . .	92
7.3.2	Data driven normalisation of $t\bar{t}$ and multijet backgrounds . . . . .	102
7.4	Systematic Uncertainties . . . . .	102
7.5	Results . . . . .	103
7.5.1	Combined results with the resolved analysis . . . . .	106
7.6	Conclusion . . . . .	108
<b>8</b>	<b>Search for boosted di-<math>b</math>-jet resonances associated with a jet</b>	<b>109</b>
8.1	Overview . . . . .	109
8.1.1	Searching for boosted Higgs boson production . . . . .	109
8.1.2	Searching for a mediator to dark matter . . . . .	110
8.1.3	Analysis strategy . . . . .	112
8.2	Event selection . . . . .	112
8.2.1	Trigger . . . . .	112
8.2.2	Preselection . . . . .	113
8.2.3	Final event categorisation . . . . .	114
8.2.4	Impact of the track jet choice . . . . .	115
8.3	Background modelling . . . . .	117
8.3.1	Resonant backgrounds . . . . .	117

8.3.2	Non-resonant multijet background . . . . .	119
8.3.3	Data-driven multijet background: CR Reweighting . . . . .	120
8.3.4	Direct Fitting Method . . . . .	128
8.4	Systematic Uncertainties . . . . .	131
8.4.1	Luminosity . . . . .	132
8.4.2	Large- $R$ jet energy and mass uncertainties . . . . .	132
8.4.3	Flavour tagging uncertainties . . . . .	135
8.4.4	Background and signal modelling uncertainties . . . . .	137
8.4.5	Theoretical uncertainties . . . . .	139
8.5	Statistical Framework . . . . .	139
8.5.1	Search phase with BumpHunter . . . . .	139
8.5.2	Limit setting with BAT framework . . . . .	140
8.6	Results . . . . .	141
8.6.1	Validation of the background modelling with $V + \text{jets}$ . . . . .	141
8.6.2	Measuring the boosted Higgs production . . . . .	142
8.6.3	Searching the $Z'$ mediator . . . . .	143
8.7	Conclusions . . . . .	144
<b>9</b>	<b>Conclusions and outlook</b>	<b>145</b>
<b>A</b>	<b>Boosted <math>h \rightarrow b\bar{b}</math> identification</b>	<b>147</b>
A.1	High mass tail identification . . . . .	147
<b>B</b>	<b>Search for boosted di-Higgs production in the <math>b\bar{b}b\bar{b}</math> final state</b>	<b>149</b>
B.1	Performance of $m^{\text{calo}}$ and $m^{\text{comb}}$ . . . . .	149
B.2	Background prediction using MC simulations . . . . .	151
B.3	Performance of the $b$ -tagging algorithm in different kinematic regimes . . . . .	155
<b>C</b>	<b>Search for boosted di-<math>b</math>-jet resonances associated with a jet</b>	<b>159</b>
C.1	Fixed- $R$ vs variable- $R$ track jets . . . . .	159
C.2	CR reweighting . . . . .	161

---

C.2.1	40 GeV cut on 3D reweighting $p_{T,J}, \Delta R_{bb}, \mathcal{R}_{p_T}$ . . . . .	161
C.2.2	3D reweighting with $p_{t,1}, \Delta R_{bb}, \mathcal{R}_{p_T}$ . . . . .	161
C.2.3	3D reweighting with $p_{t,2}, \Delta R_{bb}, \mathcal{R}_{p_T}$ . . . . .	163
C.2.4	2D reweighting with $p_{T,J}, \mathcal{R}_{p_T}$ . . . . .	164
C.2.5	2D reweighting with $\Delta R_{bb}, \mathcal{R}_{p_T}$ . . . . .	165
C.3	Signal contamination tests . . . . .	165
C.4	Systematic uncertainties . . . . .	167
C.4.1	Scale uncertainties . . . . .	167
C.4.2	Flavour tagging uncertainties . . . . .	171
<b>D</b>	<b>Release and GRL</b>	<b>183</b>
	<b>Acknowledgements</b>	<b>199</b>



# Chapter 1

## Overview

We are living in a remarkable era of particle physics, with a privilege of having results from the Large Hadron Collider (LHC) at unprecedented centre-of-mass energies of  $\sqrt{s} = 13$  TeV. With the discovery of what is often referred to as the last missing piece of the Standard Model (SM), the *Higgs boson*, we witness that all the theories and attempts to describe the SM and the Higgs mechanism that is responsible for the generation of particle masses in the past can finally be confronted with the experimental truth of today. At the same time, one of the greatest milestones in the future's particle physics history has been taken.

Today, the SM of elementary particle physics is an elegant and experimentally verified theory that matches perfectly the experimental results and holds the theoretical expectations in many senses. Yet, there are many significant questions to answer not only to understand the universe of the very small but also to explain phenomena seen at large scales. The mysterious dark matter and the missing graviton are two examples of these highly important questions. While ongoing searches aim to reveal the new physics beyond the standard model (BSM) to answer these questions at various experiments around worldwide, LHC continues to provide great opportunities to perform some of these searches with its unique dataset, the most energetic proton-proton collisions ever. Not only direct searches for new physics, but also searching the properties of the existent SM theory through precision measurements can give clues about new physics, both are pursued at LHC.

It is common to use the new discoveries as tools that lead to the subsequent ones in scientific journeys. Similarly, the Higgs boson became a new tool for particle physicists to use in new searches after its discovery. Naturally, there has been a dramatic increase in the theoretical and experimental studies involving the Higgs boson. Therefore, it is crucial to study the Higgs boson and develop advanced techniques for its identification in such an era.

In addition to probing unknown kinematic regions, the LHC provides the chance to observe heavy SM particles while they are travelling at significant fractions of the speed of light. The term *boosted object* refers to those particles, particularly the  $W$ ,  $Z$  bosons, the top quark and the Higgs boson. Due to the boost in a given process, decay products of these objects can overlap, failing the standard reconstruction algorithms. That is why boosted topologies require dedicated reconstruction techniques.

Focusing on the most frequent decay channel of the Higgs boson, the  $h \rightarrow b\bar{b}$  process, in the boosted regime, this thesis discusses the improved identification techniques for Higgs boson and presents two searches using those techniques. The identification methods

for boosted Higgs boson decays to a  $b\bar{b}$  pair are discussed extensively in Chapter 6.

The presented data analyses are performed using ATLAS data collected at the centre-of-mass energy of 13 TeV. Both of the searches are looking for boosted  $b\bar{b}$  pairs in their final state, with different motivations but mainly to find new physics. The first data analysis presented in Chapter 7 is performed with  $36.1 \text{ fb}^{-1}$  of ATLAS  $pp$  data and searches for the pair production of boosted Higgs bosons where both Higgs boson decay to a  $b\bar{b}$  pair, each leading 4  $b$ -quarks in the final state. This search is presented briefly focusing on the impact of the newly developed techniques for the boosted  $h \rightarrow b\bar{b}$  identification shown in Chapter 6, and the discussions about the most crucial backgrounds of boosted  $b\bar{b}$  pairs.

The second data analysis presented in Chapter 8 is performed with  $80.5 \text{ fb}^{-1}$  ATLAS  $pp$  data and searches for a boosted  $b\bar{b}$  pair with an additional jet from initial state radiation (ISR) in its final state, with two physics motivations: first searching for highly boosted Higgs bosons and second searching for a  $Z'$  dark matter mediator. This channel is studied for the first time in ATLAS, therefore the presented studies are discussed in detail.

In conclusion, both of the data analyses are optimised to identify boosted  $h \rightarrow b\bar{b}$  processes and they have very similar final states and backgrounds. Therefore in the following chapters, it is convenient to present a common foundation to follow both analyses. Chapter 2 and 3 briefly explain the theoretical and experimental foundations. Chapter 4 introduces the data and simulations used in this thesis. Chapter 5 explains the ATLAS reconstruction techniques for the relevant objects, and finally in Chapter 9 the conclusions and outlook of the thesis are presented.

**Author's contribution** The scientific results presented in this thesis were obtained within the ATLAS Collaboration. Since the modern high energy particle physics experiments require tremendous efforts, any scientific result produced within the ATLAS Collaboration is ultimately a result of a combined effort of its around 3000 members from all across the world. Significant amount of work is needed for the construction, operation, and maintenance of the detector. Therefore, the direct contributions from the author are explicitly listed and briefly described below.

The author actively contributed to the trigger operations, ensuring smooth data taking, as an expert on High Level Trigger software, and  $b$ -jet triggers.

The author contributed substantially to the development, optimisation and implementation of the identification techniques for boosted Higgs bosons decaying to  $b\bar{b}$  pair presented in Chapter 6. Being one of the two main analysers, she worked on the development of the ATLAS Run 2 Higgs jet tagger software and recommendations. Particularly focusing on the Higgs jet mass studies, she adapted and validated the newly introduced jet mass observable, the *combined jet mass* for boosted  $h \rightarrow b\bar{b}$  identification, and provided the first implementation of the muon correction for the combined mass to the ATLAS collaboration. She also derived the jet mass resolutions for systematic uncertainty estimations for Higgs samples and the presented backgrounds in Chapter 6. The analysis on the identification of initial state radiation contributions within the Higgs jet and their impact on its mass was performed by the author. She provided the idea of developing a  $p_T$  dependent Higgs mass window requirement and worked it out. Finally, the author implemented alternative jet reconstruction algorithms for boosted  $h \rightarrow b\bar{b}$  identification and studied their impact.

The author implemented the combined jet mass and muon correction to the search presented in Chapter 7, leading to notable improvements in the search sensitivity. She worked on the data-driven multijet background estimation method. In particular, she

investigated the modelling of  $b$ -tagged track jets, as well as validated the data-driven background estimation approach using MC simulations. She also performed an analysis using stable truth particles to study the performance of the  $b$ -tagging algorithm and the composition of the multijet background in the analysis.

For the search for boosted  $b\bar{b}$  pairs presented in Chapter 8, the author contributed to different aspects of the analysis from the start to the end. She contributed to the optimisation and validation of the analysis framework. She studied the impact of the alternative track jet reconstruction methods as well as the modelling of the  $b$ -tagged track jets and the multijet background. The author developed and tested the *Control region reweighting* method for multijet background estimation using MC simulations. Finally, she contributed to the systematic uncertainty estimations particularly by implementing and evaluating the  $b$ -tagging uncertainties, and studying the impact of a possible NNLO reweighting for  $t\bar{t}$  background.



# Chapter 2

## Theory

This chapter aims to provide a brief theoretical foundation and motivation for the readers who are not familiar with the topics presented in this thesis so that it is possible to follow the discussions throughout the thesis and understand the reasons why it is important to perform the presented searches. Chapter starts with the summary of the Standard Model briefly explaining the fundamental particles and their interactions. Then some of the open questions which can not be addressed by the Standard Model are discussed so that the reasons for the new physics searches can be put into perspective. Finally, the possible scenarios beyond the Standard Model are presented, particularly the ones which are used as benchmark models for the searches presented in this thesis. Although this chapter is aimed to give a broad overview, all the relevant references elaborating on each topic are provided for the readers who would like to steep themselves in the theoretical aspects.

### 2.1 Standard Model

The content of the Universe and particularly the structure of the matter was one of the biggest mysteries about our world since the beginning of the human history. Using the existing knowledge of their time, many attempts have been made to reveal the smallest building blocks of the ordinary matter around us. With the accumulation of the experimental results, the theories were either improved or refuted, providing more information for the new physics models.

Finally, at the end of 19th century, electron was discovered with no internal structure as a first elementary particle [1]. Following this discovery, based on the electromagnetic interactions, Rutherford's atomic physics model came into the stage by revealing the existence of the subatomic particles [2]. Since then, thanks to many other discoveries, a well grounded and experimentally verified theory which predicts the existence of all the fundamental particles that we know and explains how they interact with each other, the *Standard Model* (SM), have been formulated.

The SM is a gauge quantum field theory which stays invariant under the local transformations of the unitary product group  $SU(3)_C \times SU(2)_L \times U(1)_Y$ . In the following sections, the fundamental particles and forces in the SM are introduced (Section 2.1.1) and then a glimpse of the theories behind the interactions is given (Section 2.1.2 and 2.1.3). Finally in Section 2.1.4, the long-sought solution for the generation of particle masses in the SM, the Higgs boson and Higgs mechanism is briefly discussed.

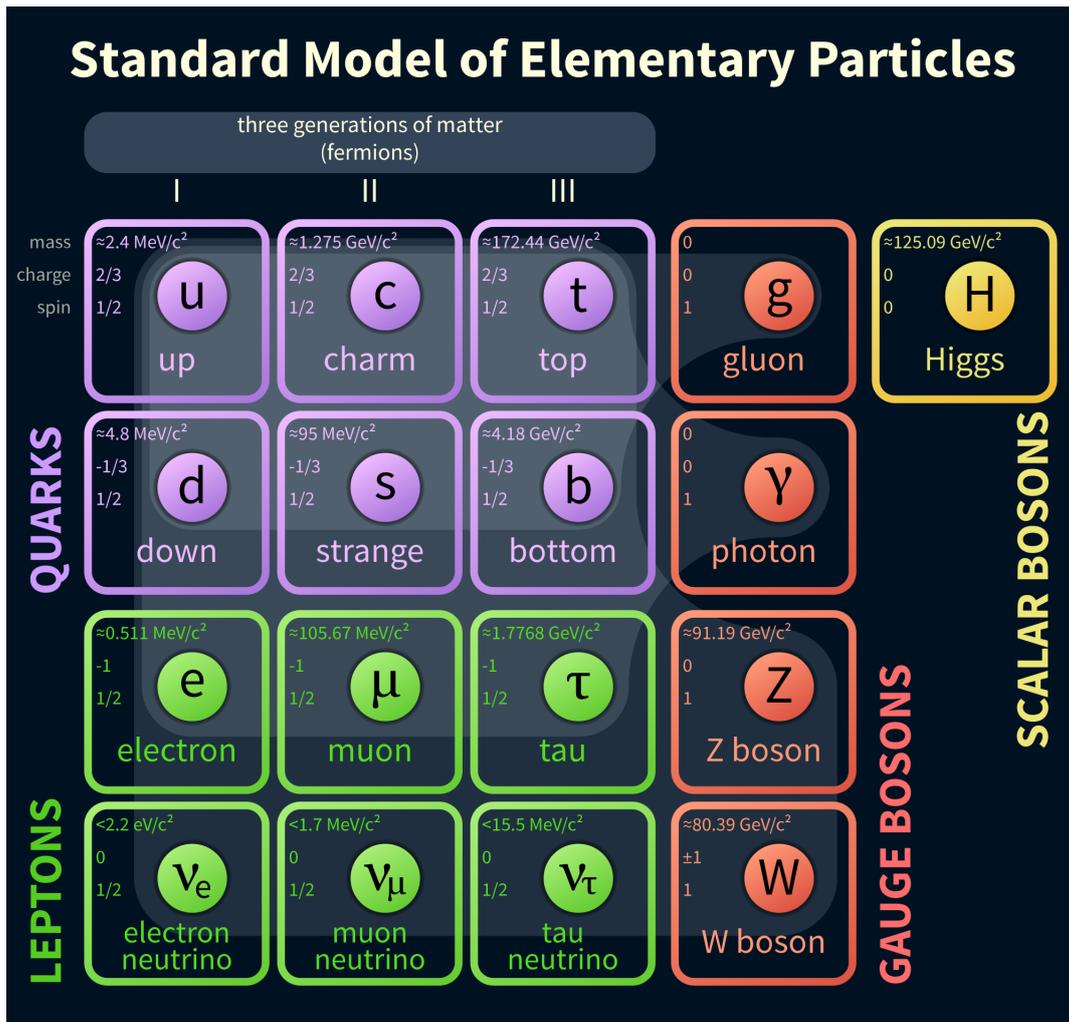


Figure 2.1: Fundamental particles of the Standard Model are shown together with the mass, charge and spin information [4]. The shaded colours represent the gauge boson interactions.

### 2.1.1 Fundamental particles and forces

In the present SM theory, there are 61 elementary particles in total which makes up all ordinary matter in the universe. This number 61 comes up by considering each quark and lepton with its anti-particles and different colour charges of quarks and gluons: 12 leptons, 36 quarks, 12 mediators and finally the Higgs boson [3].

Quarks and leptons are fermions with the half integer spin,  $1/2$ , following Fermi-Dirac statistics and obeying the Pauli exclusion principle. All fermions come in 6 different flavours, organised into 3 generations as shown in Figure 2.1. Quark also carries a colour charge out of the 3 possible colours, leading to 18 quarks in total. Both quarks and leptons have also their antiparticles that carry the opposite electric charge of the original particles.

The interactions between these particles are mediated by the gauge bosons with integer spins following Bose-Einstein statistics. Eight coloured gluons generated by  $SU(3)_C$  group mediate the strong force.  $W^\pm$  and  $Z$  bosons are the three mediators of the weak force that are generated initially massless by the  $SU(2)_L$  group and then are given mass through spontaneous symmetry breaking (SSB) that is explained in Section 2.1.3. Finally, photons are the mediators for the electromagnetic interaction generated by the

$U(1)_Y$  group after SSB. Both gluons and photons are massless. All gauge bosons have spin 1, while the Higgs boson is a scalar boson with spin 0. More detailed discussion about the Higgs boson is given later in Section 2.1.4. All these fundamental particles are presented with their classification, mass, charge, and spin information in Figure 2.1.

Four fundamental forces are identified in our Universe so far: strong, electromagnetic, weak and gravitational force. Properties of these forces are quite different in terms of their interaction strength and range as shown in Table 2.1. Each of the forces is carried by its mediator particle as mentioned above. Photons, gluons,  $W$  and  $Z$  bosons are the mediator particles that are observed experimentally for the three fundamental interactions. On the other hand, the mediator for the gravitational force, the graviton, is still not experimentally observed, and the mystery of extremely weak gravitational interaction is still to be solved (see Section 2.2.3).

Table 2.1: Fundamental Forces [3, 5].

Force	Strength	Range	Particles
Strong	1	$10^{-18}$	Gluon
Electromagnetic	$10^{-3}$	$\infty$	Photon
Weak	$10^{-8}$	$10^{-18}$	$W$ and $Z$ bosons
Gravitational	$10^{-37}$	$\infty$	Graviton

The three observed interactions are governed by the specific theories, namely the quantum chromodynamics (QCD), the theory of strong interaction, the quantum electrodynamics (QED) that describes the electromagnetic interaction, and finally the weak theory or quantum flavour dynamics (QFD) governs the weak interaction.

Each interaction is represented by a three-point vertex of the gauge boson with an incoming and outgoing fermion, as shown in Figure 2.2. The strengths of these interactions are determined by the coupling constant (or coupling strength), denoted as  $g$ . Often the coupling strengths are expressed with their dimensionless forms,  $\alpha \propto g^2$  to ease the calculations.

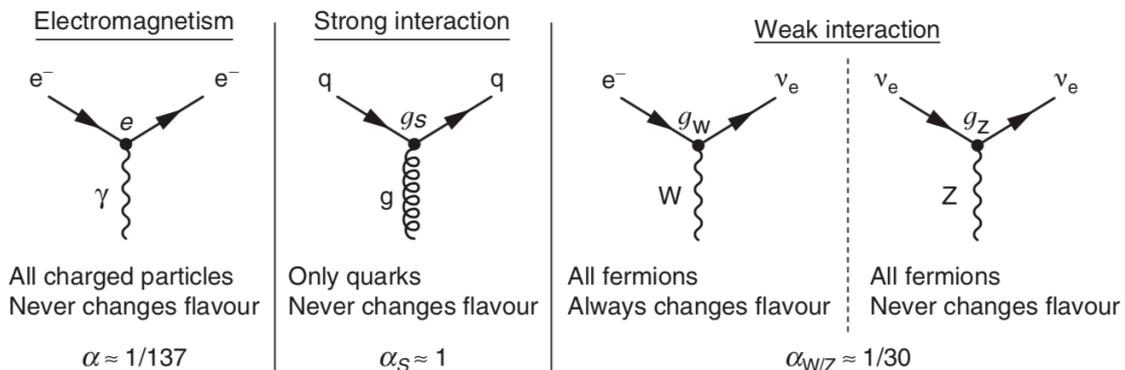


Figure 2.2: Examples of the Standard Model vertices for three fundamental interactions are shown. For electromagnetism, coupling strength is the electron charge,  $e$  [5].

In the following sections, three of the four fundamental interactions are briefly introduced.

### Electromagnetic interaction

Quantum electrodynamics describes the interaction between the light and ordinary matter. All charged particles interact electromagnetically. The interaction is mediated by a massless photon ( $\gamma$ ), thus the range of the electromagnetic force is infinite. As shown in Figure 2.2, the fundamental vertex of the interaction is either photon absorption or emission. The coupling constant of the interaction is known as *fine structure constant* which is approximately  $1/137$ . In all electromagnetic interactions, the fermion flavour and the electrical charge (Q) must be conserved [3].

### Strong interaction

Similar to the electrical charge in electrodynamics, colour charge is playing a key role in the strong interaction. Only quarks and gluons are the subjects of this force with their non zero colour charges. As mentioned in Section 2.1.1, there are 8 different physical gluon states which carry the different combinations of colour and anti-colour charges. On the other hand, quarks and antiquarks can only exist in one of the three orthogonal colour anti-colour states, respectively.

Three fundamental vertex can be counted: a quark may emit (or absorb) a gluon, a gluon may emit (or absorb) a gluon and gluons can interact with each other. What a gluon does in a quark-gluon vertex is simply carrying the difference of the colour charges between the quarks. In other words, it renders possible that the quarks change their colour states from one to another by carrying some of the charges themselves.

In strong interaction, a dimensionless quantity called *isospin* must be conserved [3].

### Weak interaction

In flavour dynamics there are two types of weak interaction, known as *charged-current interaction* and *neutral-current interaction* with respect to the charge of the boson mediating the interaction. While leptons can absorb a  $W^\pm$  boson to be converted into their corresponding neutrinos, quarks can emit and absorb the  $W^\pm$  bosons changing their flavour via the charged current interaction. On the other hand, in the neutral-current interaction, both the quarks and the leptons can absorb and emit the  $Z$  boson without change their flavour.

When quarks are changing their flavours via charged-current interaction, they follow the probabilities given in Cabibbo-Kobayashi-Maskawa (CKM) matrix [6, 7] to be converted into one of the three opposite type quarks, as given below in 2.1. The transition probabilities are proportional to the square of the given matrix elements. As the elements of CKM matrix indicate, the transitions mostly occur between the same generations of quarks. More massive quarks decay to produce lighter quarks via weak interaction, as it is true also for leptons.

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97428 & 0.2253 & 0.00347 \\ 0.2252 & 0.97345 & 0.0410 \\ 0.00862 & 0.0403 & 0.99915 \end{bmatrix} \quad (2.1)$$

A quantum number referred to as *weak isospin* ( $T$ ) plays an important role for the weak interactions. The third component of the weak isospin ( $T_3$ ) must be conserved in weak interactions. While all weak bosons can couple to particles with a third component

of the weak isospin, only  $Z$  bosons can couple to the particles which carry an electric charge.

### 2.1.2 Quantum chromodynamics

Quantum chromodynamics (QCD) has distinctive features which are necessary to know in order to understand the observed phenomena in particle physics. In this section, the colour confinement and the asymptotic freedom phenomena are very briefly described, and the Lagrangian of the theory is given.

Up to this point, the information given for strong interactions was very similar to the electromagnetic interactions except the two differences. First, eight gluons mediate the force instead a photon, and second, gluons are self-interacting particles. There is a major difference between the electrical charges and the colour charges. The colour charged particles have never been observed as free particles. They are observed in bound states with other colour charged particles forming colourless *hadrons*, which are composite particles comprised of two or more quarks. This phenomena in QCD is known as *colour confinement* [5].

Another major feature of QCD is the strength of its coupling constant,  $\alpha_s$ . At low-energy scales,  $\alpha_s$  is large ( $\sim O(1)$ ) meaning that each additional vertex (*higher order diagrams*) in a given process, significant contributions are expected unlike the QED processes where the simplest Feynman diagrams have the major contributions to the final sum of the amplitudes (matrix elements). This issue creates additional challenges on top of the usual divergence problems faced in the amplitude calculations. In QED, divergent integrals can be solved by the mathematical approaches, known as regularisation and normalisation, ignoring the regions where  $\alpha$  blows up at the energies of about  $10^{280}$  [3]. But for the  $\alpha_s$  the situation is different, and that was a major showstopper for the particle physicists for years. Finally in 1973, David Gross and Frank Wilczek discovered a phenomena, so-called *asymptotic freedom* [8]. The discovery allows to perform calculations without worrying about the large values of  $\alpha_s$ , since  $\alpha_s$  is found to be asymptotically decreasing with the increasing energies. In other words, at high energies, quarks interact weakly allowing perturbative calculations. At low energies, the interaction becomes strong resulting to the confinement of quarks and gluons within composite hadrons [3].

Lagrangian of QCD ( $\mathcal{L}_{QCD}$ ) describes the dynamics of the quarks and gluons, consisting of the free propagators of the gluon and quark fields, the self interactions of gluons, and finally the quark gluon interaction:

$$\mathcal{L}_{QCD} = -\frac{1}{4} \sum_{a=1}^8 G_{\mu\nu}^a G_a^{\mu\nu} + \sum_{i,j=1}^{N_f} \bar{q}_i (i\gamma^\mu D_\mu - m_q \delta_{ij}) q_j. \quad (2.2)$$

The first term of the Lagrangian represents the 8 gluon fields  $G_\mu^a$ . The second term is summed over 6 quark flavours ( $N_f$ ) where  $q_i$  denotes the quark spinors and  $D_\mu$  represents the gauge covariant derivative. For more details and mathematical discussion, the given reference can be seen [5].

### 2.1.3 Electroweak theory and symmetry breaking

Both the electromagnetic and the weak interactions are well understood and unified in a single theory, so-called the *electroweak theory*. The unified electroweak theory is described by the product of the groups  $SU(2)_L \times U(1)_Y$ . The  $SU(2)_L$  group generates three gauge boson fields  $W_\mu^i$ ,  $i = 1, \dots, 3$ , and the  $U(1)_Y$  group generates a single gauge boson field  $B$  with the conserved quantum number  $Y$ , the *weak hypercharge*:

$$Y = 2(Q - T_3), \quad (2.3)$$

where  $Q$  is the electric charge and  $T_3$  is the third component of weak isospin as introduced in Section 2.1.1. While the weak isospin describes how a particle transforms under  $SU(2)$ , the hypercharge describes how it transforms under  $U(1)$ .

The components of the Lagrangian for the electroweak theory can be written as:

$$\mathcal{L}_{SU(2) \times U(1)} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_\phi + \mathcal{L}_f + \mathcal{L}_{\text{Yukawa}}, \quad (2.4)$$

where the  $\mathcal{L}_{\text{gauge}}$  term describes the interaction between the  $W$  and  $B$  vector bosons, the term  $\mathcal{L}_\phi$  represents a scalar potential which interacts with itself and the other gauge bosons, the  $\mathcal{L}_f$  is the kinematic term for the fermions, and finally the  $\mathcal{L}_{\text{Yukawa}}$  term describes an interaction between the scalar field  $\phi$  and the fermions.

$\mathcal{L}_{SU(2) \times U(1)}$  has no mass term, preserving the local gauge invariance. On the contrary to this theoretical fact,  $W$  and  $Z$  bosons are massive particles i.e. masses of around 80 GeV [9]. This contradiction was a disturbing problem in particle physics for some time, until the mechanism of *spontaneous symmetry breaking* (SSB) was introduced to explain the phenomena:

$$SU(3)_C \times SU(2)_L \times U(1)_Y \xrightarrow{\text{SSB}} SU(3)_C \times U(1)_{EM}. \quad (2.5)$$

Through the SSB, the unified electroweak symmetry is broken to the electromagnetic subgroup  $U(1)_{EM}$ . As a result of this process the  $W$ , and  $Z$  bosons gain their masses, while photons remain massless. With this approach, the problem evolved to another quest which is to find a mechanism responsible from the symmetry breaking. Finally around 1960s, why and how this symmetry is broken was explained by a mechanism, so-called the *Brout-Englert-Higgs* (BEH) or *Higgs* mechanism [10, 11]. The main components of the electroweak Lagrangian after SSB can be written as:

$$\mathcal{L}_{U(1)_{EM}} = \mathcal{L}_K + \mathcal{L}_N + \mathcal{L}_C + \mathcal{L}_H + \mathcal{L}_{HV} + \mathcal{L}_{WWV} + \mathcal{L}_{WWVV} + \mathcal{L}_Y, \quad (2.6)$$

where  $\mathcal{L}_K$  is the kinetic term including the mass terms,  $\mathcal{L}_N$  and  $\mathcal{L}_C$  are neutral and charged current interactions,  $\mathcal{L}_H$  and  $\mathcal{L}_{HV}$  are the Higgs interactions with itself and the gauge bosons,  $\mathcal{L}_{WWV}$  and  $\mathcal{L}_{WWVV}$  are gauge three and four point self interactions, and finally  $\mathcal{L}_Y$  is the Yukawa interactions between the Higgs field and the fermions.

The linear combination of the  $W_\mu^1$  and  $W_\mu^2$  fields produces  $W^\pm$  bosons, the  $B$  and  $W_\mu^3$

fields mix with the *Weinberg angle* ( $\theta_W$ ) to form photons and  $Z$  bosons as given below [9]:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2), \\ Z_\mu^0 &= \cos \theta_W W_\mu^3 - \sin \theta_W B_\mu, \\ A_\mu &= \sin \theta_W W_\mu^3 + \cos \theta_W B_\mu, \\ \cos \theta_W &= \frac{g}{\sqrt{g^2 + g'^2}}. \end{aligned} \quad (2.7)$$

### 2.1.4 The Higgs Boson

This section briefly introduces the Higgs mechanism, and then describes the Standard Model Higgs boson which is produced as a result of the Higgs mechanism. As introduced in previous section, the Higgs mechanism is responsible for the mass generation for the  $W$  and  $Z$  weak gauge bosons through electroweak symmetry breaking. In order to obtain the mass terms in electroweak Lagrangian, the mechanism introduces a generic complex scalar doublet:

$$\phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix}, \quad (2.8)$$

where  $\phi^+(x)$  is a charged and  $\phi^0(x)$  is a neutral complex scalar field. The corresponding Higgs potential and Lagrangian are written as:

$$V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2, \quad (2.9)$$

$$\mathcal{L}_\phi = (D_\mu\phi)^\dagger(D^\mu\phi) + V(\phi^\dagger\phi). \quad (2.10)$$

Two parameters of this potential must satisfy the relations of  $\mu^2 < 0$  and  $\lambda > 0$  so that the symmetry breaking can be explained by the resulting shape of the potential, known as *Mexican hat potential* as shown in Figure 2.3. As can be seen from the figure, it is a symmetric potential and has local minima around the origin. Introducing a new field

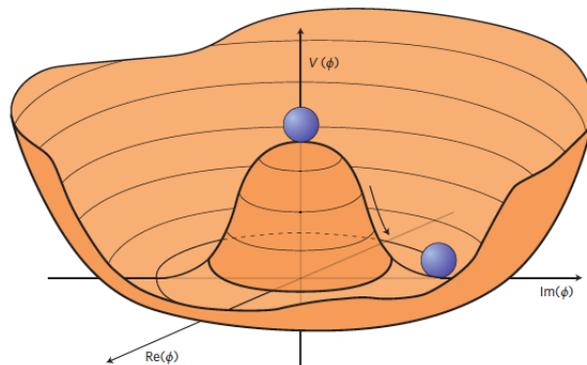


Figure 2.3: A Mexican hat potential that leads to spontaneous symmetry breaking [12].

variable  $v$ , with  $v = \sqrt{-\frac{\mu^2}{\lambda}}$ , the Higgs potential can be rewritten creating its minimum at  $\phi^\dagger\phi = \frac{v^2}{2}$ . This value is called the *vacuum expectation value* (VEV) and it corresponds

to the vacuum state of the Higgs field. Choosing a specific VEV to excite the complex scalar Higgs doublet:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \eta(x) \end{pmatrix}, \quad (2.11)$$

the symmetry of the system is broken creating the *Higgs boson*, the real scalar written as  $\eta$  with a mass of  $m_\eta = \sqrt{2}\lambda v$  [13]. The initial Lagrangian in Equation 2.10 takes a new form:

$$\mathcal{L} = \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{mass}} + \mathcal{L}_{\text{int}} \quad (2.12)$$

where  $\mathcal{L}_{\text{mass}}$  is the mass term of the gauge fields, the interaction term is  $\mathcal{L}_{\text{int}}$  and the Higgs field term is  $\mathcal{L}_{\text{Higgs}}$  with its kinetic, mass and self-interaction terms:

$$\mathcal{L}_{\text{Higgs}} = \frac{1}{2}(\partial_\mu h)^2 - \frac{1}{2}m_h^2 h^2 - \frac{m_h^2}{2v}h^3 - \frac{m_h^2}{8v^2}h^3. \quad (2.13)$$

As a result of the process, the  $W$  and  $Z$  bosons acquire their masses with the relations below:

$$\begin{aligned} m_Z &= \frac{\sqrt{g^2 + g'^2}}{2}v \\ m_W &= \frac{gv}{2} = m_Z \cos \theta_W. \end{aligned} \quad (2.14)$$

Although the Higgs mechanism was predicted in 60s, the Higgs boson could only be observed in 2012 at LHC with a mass of 125 GeV [14, 15]. The four production mechanism of Higgs boson are shown in Figure 2.4. The gluon fusion (ggf) process has the largest cross section at LHC where no additional particle is expected but only the Higgs boson. The second largest Higgs boson production is the vector boson fusion (VBF), and the third one is the associated production mode with an electroweak vector boson  $W$  or  $Z$ , so-called the *Higgsstrahlung* mode. Finally, the Higgs production with  $t\bar{t}H$ .

The Higgs boson couples to all fermions with a coupling strength proportional to the fermion mass ( $m_f$ ). The interaction vertex for  $H \rightarrow ff$  can be written as [5]:

$$-i\frac{m_f}{v} = -i\frac{m_f}{2m_W}g_W. \quad (2.15)$$

Since the fermion mass plays a role in the interaction vertex, the Higgs boson couples preferentially to the most massive particles. For the 125 GeV Higgs boson, the dominant decay channel is the  $h \rightarrow b\bar{b}$  channel.

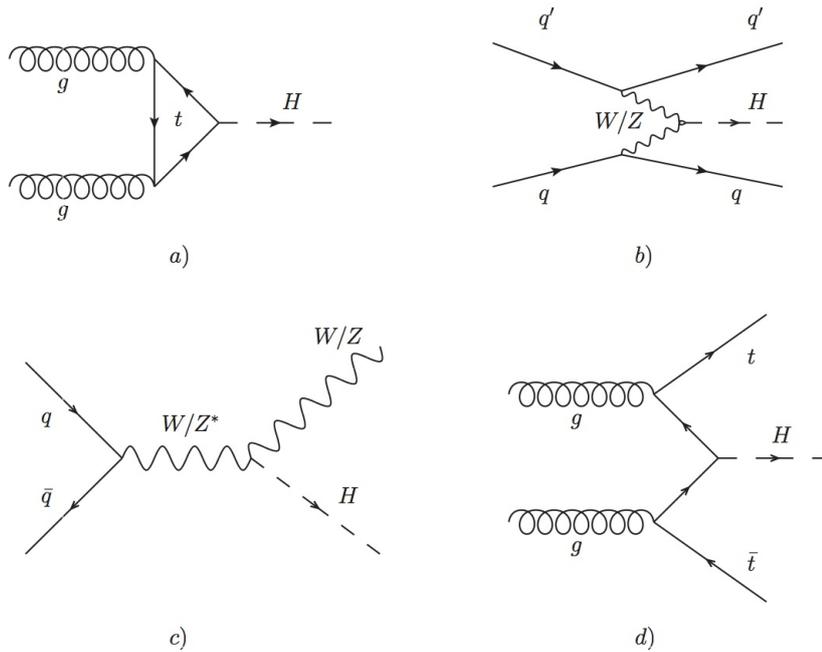


Figure 2.4: Different Higgs boson production modes are shown: a) Gluon fusion (ggF) b) Vector boson fusion (VBF) c) Higgsstrahlung and d)  $t\bar{t}H$  channel [16].

## 2.2 Limitations of the Standard Model

While the SM can be considered as a complete and experimentally verified theory especially after the discovery of its last missing piece, the Higgs boson, it has still shortages and limitations which can be recognised by its inability to answer certain questions. Although it is still controversial in the community, it is believed that SM is an effective theory of a more fundamental one. Therefore, some of these questions are also referred to as not only questions but the problems of the SM such as nonzero mass of neutrino, hierarchy problem, matter antimatter symmetry etc [17]. Whether or not the unexplained issues are the problems of the SM, this section describes only the shortages of the current SM that are relevant to the context of this thesis. More detailed information on the subject can be found in the literature.

### 2.2.1 Gravity

As mentioned above, the SM does not contain the gravity, one of the four fundamental forces observed in nature. This is considered as an important deficiency of the SM, implying that the new theories are needed. The simple addition of *graviton* to the SM as the force carrier, does not work yet looking at the experimental results. Therefore, new theories are needed to explain this phenomena.

### 2.2.2 Dark Matter

There are several evidences from the astronomical observations, predicting the existence of a different type of matter than the known, ordinary (baryonic) matter in the universe. While the ordinary matter is composed of the SM particles, the particle content of this unknown matter, so-called the *dark matter*, is still a mystery. One of the most convincing evidence for the presence of dark matter comes from the rotation curves of galaxies [18].

The measurements of the galaxy rotation curves shows that the rotational circular velocity ( $v_c$ ) is almost constant at large distances although it is expected to decrease with the relation  $v_c \propto \sqrt{r}$  considering the mass of the galaxies. In order to explain this observation, there must be additional invisible mass. Another well known proof is the observation of the *Bullet Cluster*, which consists of two large colliding clusters of galaxies as shown in Figure 2.5. Using gravitational lensing, the total mass were reconstructed and overlaid the background. The obtained results indicate that there must be additional matter other than the baryonic matter to explain the observation [19].

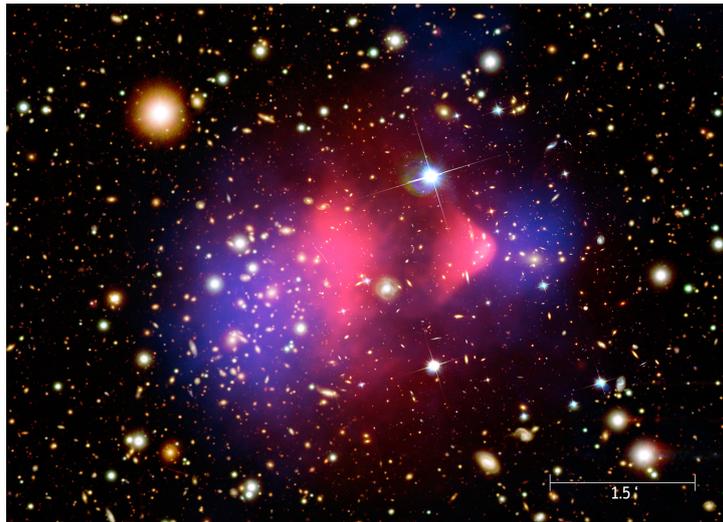


Figure 2.5: The bullet cluster (1E0657-56). The background images show the location of galaxies, with most of the larger yellow galaxies associated with one of the clusters. The overlaid pink features show x-ray emission from hot, intra-cluster gas. The overlaid blue features show a reconstruction of the total mass from measurements of gravitational lensing. This appears coincident with the locations of the galaxies, implying it has a similarly small interaction cross-section. However, there is far more mass than that present in the stars within those galaxies, providing strong evidence for the existence of an additional reserve of dark matter [19].

Although the content of the DM is still unknown, there is an estimation about its amount from the temperature fluctuations of the cosmic microwave background, which is 26.6% of the known Universe [20]. Considering that the total amount of ordinary matter in the Universe is only about 4%, not knowing the nature of this sizeable matter is quite pressing. Therefore, this problem can be considered as one of the biggest questions in today's particle physics. Although the question is not particularly towards to SM and its layout, clearly there is no answer given by the current SM to explain the situation, which requires some new physics beyond the standard model.

### 2.2.3 Hierarchy problem

Although the discovery of the Higgs boson solves the mystery of the electroweak symmetry breaking and completes the SM, the Higgs boson mass at 125 GeV triggers fundamental questions about the scales of the SM. While the electroweak scale is  $O(100 \text{ GeV})$ , the Planck scale is at  $O(10^{19} \text{ GeV})$ . This huge difference between the weak scale and the gravitational scale is an open question and considered as an unnatural phenomena, and it is referred to as *naturalness* or hierarchy problem. This situation is a result of the fact that the Higgs boson mass is quadratically sensitive to the mass scale of physics that it

couples to, which requires severe *fine tuning* of the Higgs boson mass to be able to protect the Higgs mass against large (divergent) quantum corrections expected to be in the TeV regimes [9]. This situation can be avoided with the presence of new physics beyond the SM.

## 2.3 Beyond the Standard Model

As presented in Section 2.2, there are various unanswered questions which need additional theories beyond the standard model (BSM) to answer them. These new theories are candidates to be *new physics* and their signatures are searched heavily in the high energy physics experiments. Although there are plenty of crucial BSM theories to mention that are actively searched today like supersymmetry, leptoquarks, vector-like quarks etc, only the models that are relevant with this thesis are explained in this section. In each of the subsections, firstly the general motivation behind the search category is introduced by relating the searches to the open questions that they address. Then, the focus moves to the specific models which are subjects of the presented searches in this thesis.

### 2.3.1 Extra dimension searches

As the name implies, there are models proposing additional space or time dimensions beyond the observed four dimensions. Although there is no consensus on why those extra dimensions are not observed yet, different hypotheses address this question. For instance, one approach is that we are trapped on some type of hypersurfaces embedded in the bigger space of higher dimension and can't realise or exist in other dimensions. Another one suggests that the other dimensions are too small to be observed and looped back on themselves [21]. Among various approaches, the *braneworld* scenarios, that define the higher dimensional space-time as *bulk* and the hypersurface of the visible three dimensional world as *brane*, are significant attractions for the collider searches predicting massive particles coming from the extra dimensions [9, 22, 23].

While the predictions for the extra number of dimensions vary from model to model, one of their common benefit is to be able to solve the hierarchy problem (see Section 2.2.3) [22–24]. The classical approach to solve the hierarchy problem is to unify the theories of the electromagnetism and gravitation. Such a unification is possible by introducing extra dimension(s). In this way, the weakness of the gravitational force can be explained by its propagation through these extra dimensions. The idea is that the graviton resides in another dimension and propagates to us from there. In this respect, extra dimensions can address two of the open questions of the SM (see Section 2.2.3 and 2.2.1).

One of the possible propagations of the graviton is through the compact extra dimensions within the bulk Randall-Sundrum (RS) model [23]. The Kaluza Klein (KK) excitation of the graviton ( $G_{KK}$ ) predicted in bulk RS model is used as benchmark model for the pair produced Higgs boson search presented in Chapter 7. Therefore, brief information on the model is given here, while other models are left to the reader to follow from the given references.

RS model predicts one extra spatial dimension and uses a warped geometry allowing not only graviton but also all the SM gauge fields and fermions to propagate through the space-time dimensions [23]. Through the special curvature of the extra dimension, gravity is suppressed exponentially by a *warp factor*, given as  $\lambda_\pi \sim M_{\text{Pl}} e^{-kR\pi}$ , and it

solves the hierarchy problem by the exponential of a small number  $kR$  where  $R$  is the radius,  $k$  is the curvature scale of the extra dimensions, and  $M_{\text{Pl}} = 2.4 \times 10^{18}$  GeV is the Planck scale [25].

The model predict signatures of the Kaluza Klein graviton  $G_{KK}^*$  at the TeV scale, an accessible range for LHC, and leaves the mass of the graviton as a free parameter. Another free parameter is defined as  $c = k/\bar{M}_{\text{Pl}}$  where  $\bar{M}_{\text{Pl}}$  is reduced Planck scale. If the fermions are localised in the SM brane, the dominant graviton production will be gluon fusion, and the produced gravitons decay into two Higgs bosons with branching fraction ranging from 6.43% for gravitons with a mass of 500 GeV to 7.66% with a mass of 3 TeV [26].

### 2.3.2 Dark matter searches

Numerous evidences (see Section 2.2.2) indicate the existence of dark matter (DM). Consequently, there are many theoretical attempts to explain the DM and lots of experiments are designed to observe its nature. As the theorists need constraints from the experiments to test their hypotheses, the experimentalists need theoretical assumptions in order to design their experiments. One of the most common theoretical assumptions in the community is that the DM candidate is interacting with baryonic matter but the interaction strength is at an energy scale around weak interactions or weaker than that. Several hypotheses are using this assumption suggesting different DM candidates. However, the most popular candidates are the *weakly interacting massive particles* (WIMPs), as they naturally explain the current abundance of DM in the Universe [9].

The experimental detection methods for DM can be categorised according to its interaction with the ordinary matter. Denoting the DM candidate as  $\chi$ , and the ordinary matter as  $q$ , these methods are given below together with the few examples of the experiments using them.

- Direct detection: the process of scattering,  $\chi q \rightarrow \chi q$  e.g. CRESST II [27], DAMA [28, 29], LUX [30], XENON [31].
- Indirect detection: the annihilations of DM,  $\chi\chi \rightarrow qq$  e.g. ANTARES [32], ICEDUBE [33], AMS [34].
- Colliders: mainly the production of DM,  $qq \rightarrow \chi\chi$  e.g. LHC [35].

The focus in this thesis is on the LHC searches, therefore it is useful to give a glimpse of the theoretical approaches to determine the search strategies for the collider physics. As already mentioned, many BSM theories address different SM problems. Some of the BSM theories like supersymmetry, extra dimensions, little Higgs, predict new particles in their frameworks, which can also be WIMP candidates. There are searches to find the signatures of such models at LHC focusing to constrain the parameter spaces defined by these models.

On the other hand, in order to search for the DM strategically and constrain the parameter space particularly for the DM interactions, dedicated search strategies are needed in addition to the other BSM searches. Therefore, many theoretical approaches have been developed using the effective field theories (EFT) and the simplified models to search for the DM candidates at the colliders. While earlier the main focus was on the effective vertex paradigm with EFT, the current strategy is mostly to use the generic approach called *simplified models*, which resolves the effective vertex as illustrated in

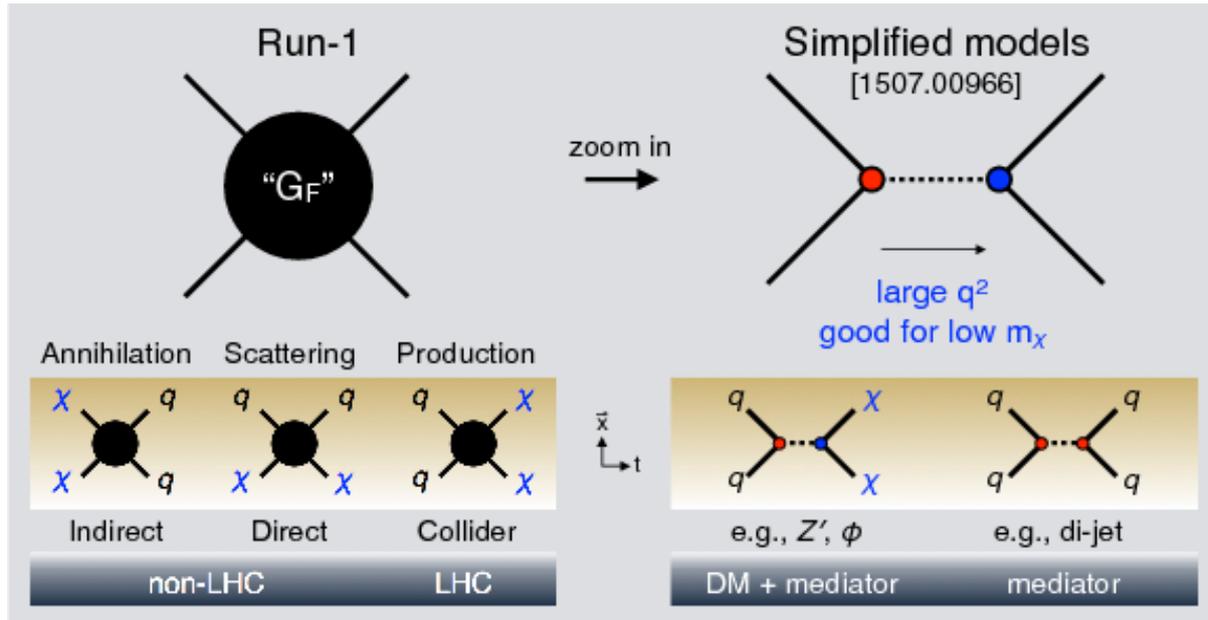


Figure 2.6: DM candidates are denoted as  $\chi$  and the ordinary matter as  $q$ , on the left representation of the EFT and on the right simplified models [36].

Figure 2.6 [36]. To accommodate a realistic WIMP scenario and restrict the smallest possible set of benchmark models, simplified models are complete enough to give an accurate description of the physics at the scale probed by colliders [37]. They predict a single fermionic DM candidate and a BSM mediator to allow the interactions. The main parameters of the simplified models can be listed as couplings to the quarks ( $g_q$ ), leptons ( $g_l$ ), DM ( $g_{DM}$ ), and masses of the DM ( $m_{DM}$ ) and the mediator ( $m_{med}$ ).

In the analysis presented in Chapter 8, simplified axial-vector mediator model is used as a benchmark model to interpret the analysis results [37, 38]. This model predicts a spin-1 s-channel mediator, denoted with  $Z'$ . Depending on the chosen values of the model parameters, different scenarios can be foreseen. For instance in the *leptophobic* scenario, the axial-vector model can only couple to the quarks with the following values of parameters:  $g_l = 0, g_{DM} = 1.0$ . Alternatively, it can equally couple to both quarks and leptons such as a model with  $g_q = 0.1, g_l = 0.1, g_{DM} = 1.0$ .

The DM production at the LHC can only be recognised with an additional object in the final state of the process such as an initial state radiation jet, boson etc. that recoils against the DM candidates. These types of searches are referred to as Mono- $X$  searches where  $X$  can be replaced by the object involved in the process. In the signature of the process, only the additional object can be identified by the detector, while the DM candidates can be recognised as the imbalance in the total energy in the transverse plane. Although these searches provide plenty of final states and possibilities to look for, the parameter spaces that they address are pretty much overlapping with each other.

Considering the interactions given in the simplified models, one can see that not only the process of  $qq \rightarrow \chi\chi$  but also the process of  $qq \rightarrow qq$  can chase the DM signatures in colliders. It is possible to perform a direct search for the DM mediator by searching for the two quarks in the final state of the process. These searches are referred to as dijet (resonance) searches, and in the context of the simplified DM mediator model, they are quite strong tools to span for the parameter space of the model. In Figure 2.7, the exclusion limits of the previous Mono- $X$  and dijet searches are unified and shown for the

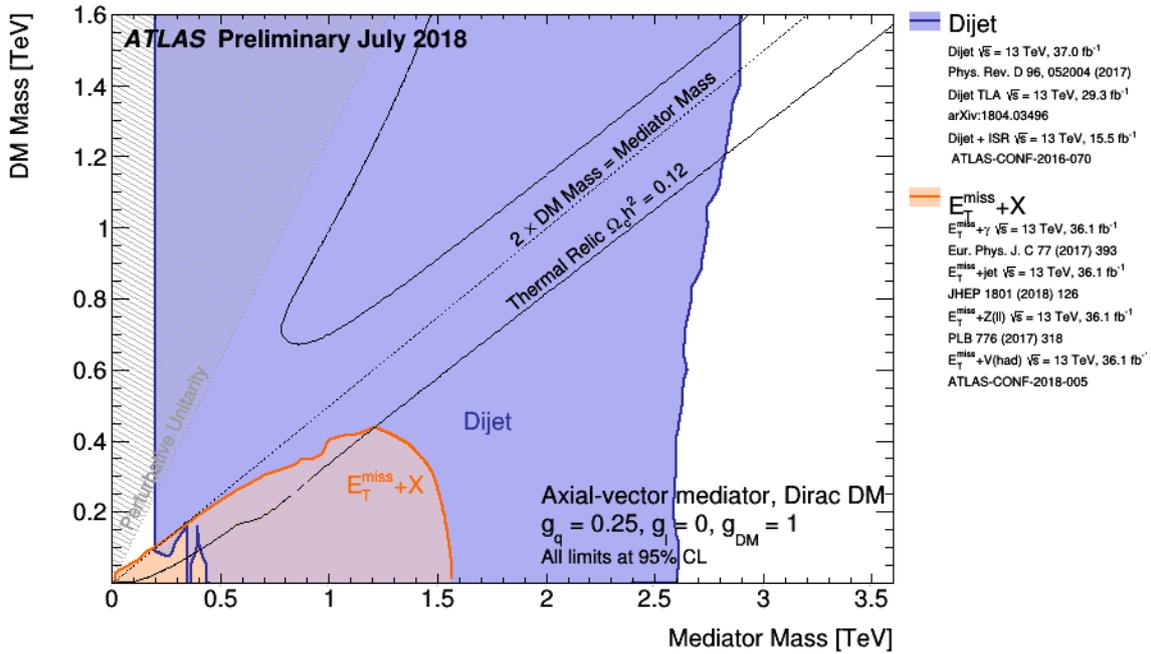


Figure 2.7: Regions in a dark matter mass-mediator mass plane excluded at 95% CL by a selection of ATLAS dark matter searches, for one possible interaction between the Standard Model and dark matter, the leptophobic axial-vector mediator as described in [38]. The exclusions are computed for a dark matter coupling  $g_{DM} = 1.0$ , a quark coupling  $g_q = 0.25$  universal to all flavors, and lepton coupling  $g_l$  set to zero [39].

leptophobic axial-vector mediator model.

### 2.3.3 Searches in the Higgs sector

Both direct and indirect searches in the Higgs sector can lead us to the signatures of new physics beyond the SM. Indirect searches aim to measure the properties of the Higgs boson such as its spin or couplings to the other particles. Any possible deviation from the SM expectations can indicate a new physics. On the other hand, many BSM scenarios suggest to extend the Higgs sector by adding a new Higgs like scalar particle, or predicting modified Higgs couplings and decays allowing to perform direct searches for new physics within the Higgs sector [40–42]. While the new physics models in Higgs sector quite often address to the hierarchy problem and the Higgs mass at 125 GeV, some of the models provide explanations for CP violation or predict dark matter candidates [9]. LHC provides unique opportunities to perform these searches as being the owner of the discovery for the Higgs boson at 125 GeV. The production cross sections of the Higgs boson at  $\sqrt{s} = 13$  TeV LHC are given in Figure 2.8 for different production modes.

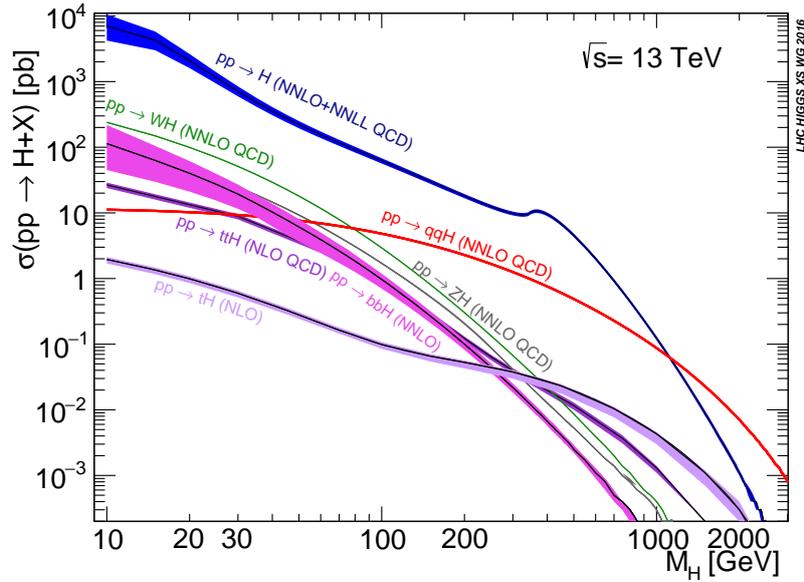


Figure 2.8: Higgs boson production cross sections at  $\sqrt{s} = 13$  TeV [43].

So far, the Higgs boson properties are measured in agreement with the SM predictions, as shown in Figure 2.9 for the measured couplings of Higgs boson. It is important to perform these measurements as precise as possible using information obtained from different SM processes.

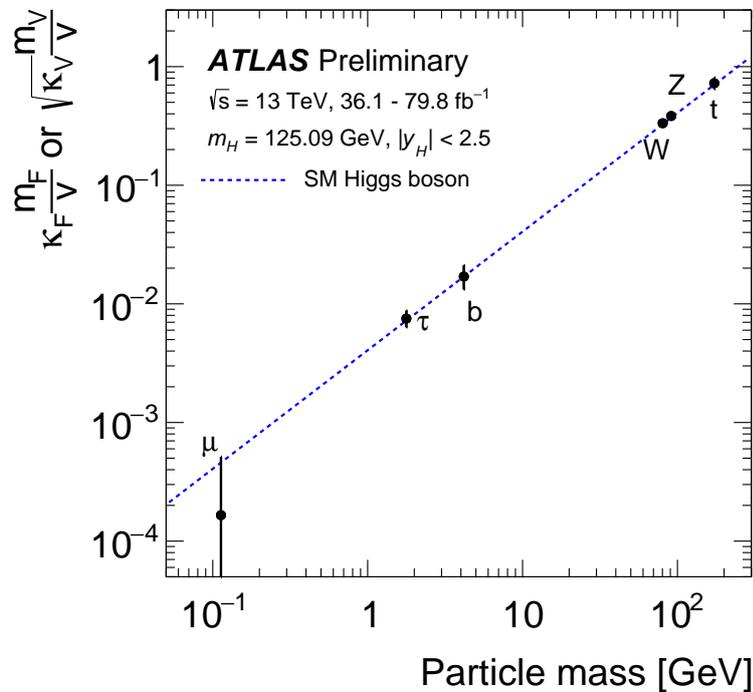


Figure 2.9: Higgs boson coupling strength to each particle as a function of particle mass compared with Standard Model prediction (blue dotted line) [44].

In this thesis, both of the presented searches measure the Higgs boson productions

decaying to  $b\bar{b}$  pairs at  $\sqrt{s} = 13$  TeV LHC in order to search for an anomalous production of the Higgs boson(s) as discussed in Section 7.1 and 8.1. While the first analysis presented in Chapter 7 searches for the pair produced Higgs bosons, the analysis presented in Chapter 8 searches for the boosted Higgs boson production. Moreover, a new scalar resonance decaying to Higgs boson pairs is considered as a candidate for new physics in the context of the search for pair produced Higgs bosons.

# Chapter 3

## The ATLAS Experiment

This chapter gives an overview of the Large Hadron Collider (LHC) in Section 3.1 and then introduces the ATLAS experiment which is one of the two multi-purpose experiments located on the LHC ring. In Section 3.2, the ATLAS detector is described together with the nomenclatures that are necessary for the reader to be able to follow the discussions throughout this thesis. Afterwards, each major component of the ATLAS detector: the tracking system, the calorimeters, the muon system, and the trigger system are briefly explained.

### 3.1 The Large Hadron Collider

The LHC is one of the most fabulous scientific and technological achievements of humanity with a higher centre-of-mass energy than has ever been explored at any previous collider. It is a two-ring superconducting hadron collider, located in 26.7 km long LEP tunnel that lies between 45m and 170m below the surface around the France-Switzerland border [35]. The European Organisation for Nuclear Research (CERN) is managing this project with the support of many countries and institutions since 1994 to explore both the SM and BSM at a centre-of-mass energy of up to 14 TeV. Although the accelerator was designed to produce mainly proton-proton ( $pp$ ) collisions, there are also dedicated periods for ion beam collisions.

In LHC, there are two high-energy particle beams travelling in opposite directions in separate beam pipes at close to the speed of light. Before coming to the LHC ring, those particles are pre-accelerated within the CERN accelerator complex (see Figure 3.1), which provides beams not only for the LHC but also for other experiments [45]. Initial injection starts with a linear accelerator (LINAC) and then the beams are sent to the three circular accelerators. Protons are injected into the Proton Synchrotron Booster (PSB), while lead ions are injected into the Low Energy Ion Ring as a second step. After that, both follow the same path through the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS). Finally after SPS, these beams are injected into the LHC. Inside the LHC, they are guided by strong magnetic fields produced by superconducting electromagnets. To maintain the superconducting state of the magnets, the temperature is kept as  $-271.3^{\circ}\text{C}$ , colder than the outer space, using liquid helium.

Each beam consists of several proton packages, called bunches, which filled approximately with  $10^{11}$  protons. Depending on the beam conditions and the purpose of runs, special filling schemes can be designed and there can be different numbers of colliding bunches in the beams. In other words, the total number of collisions per second per

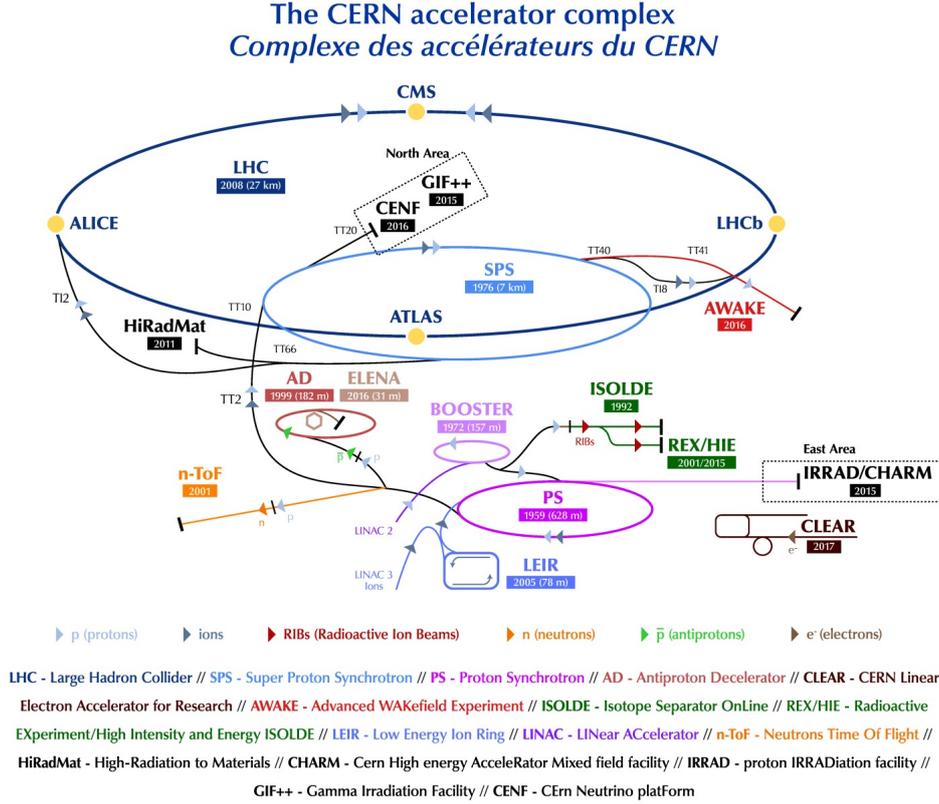


Figure 3.1: CERN Accelerator Complex [45].

area, which happen at the interaction points of the two colliding beams, changes depending on the filling scheme of LHC. Typically in colliders, this measure is expressed as instantaneous luminosity ( $\mathcal{L}$ ) and for LHC it is defined as:

$$\mathcal{L} = \frac{N_1 N_2 f}{A}, \quad (3.1)$$

where  $A$  is the cross sectional size of the beam,  $N_i$  the number of particles in each beam and  $f$ , so called the *revolution frequency*, representing the frequency of the bunch interaction in the LHC. This calculated value then can be related to the interaction rate for a desired physics process depending on the cross section ( $\sigma$ ) of the relevant process:

$$\mathcal{L}\sigma = dN/dt. \quad (3.2)$$

From this point of view, it is clear that the higher the luminosity is, the higher the event rate is. Therefore luminosity represents the scientific relevance of the collider by providing more collisions and observation opportunity.

The final commissioning of the LHC started by the end of the 2009. Since then until the end of 2011, it was operated at a centre-of-mass energy of 7 TeV (Table 3.1), mainly to gain enough confidence in the machine protection system and to explore the limits of the machine itself. In 2012, with a centre-of-mass energy of 8 TeV, the LHC successfully delivered  $23.1 \text{ fb}^{-1} pp$  data and enabled the discovery of Higgs boson [14, 15]. After three years of operation (Run 1), the machine was stopped for upgrade and

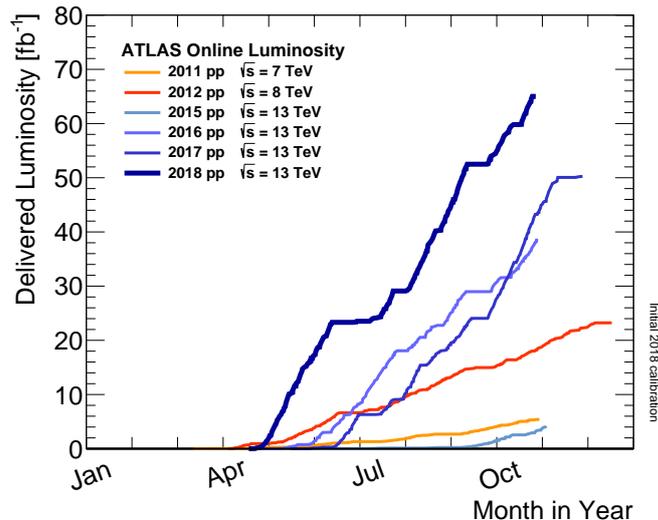


Figure 3.2: Cumulative luminosity as a function of time delivered to ATLAS during stable beams and for high energy  $pp$  collisions between 2011 and 2018 [46].

maintenance studies (LS1) to be able to perform at the centre-of-mass energy of 13 TeV during Run 2. Between 2015 and 2018, with an outstanding performance by exceeding the design luminosity value, LHC delivered more than  $150 \text{ fb}^{-1}$  of  $pp$  data (see Figure 3.2). According to the current schedule, in 2020 Run 3 is going to start with the aim of doubling the instantaneous luminosity. After 2025, with the major upgrades that will happen in Long Shutdown 3 (LS3), LHC will operate as High Luminosity LHC (HL-LHC) until 2035, delivering 5 to 7 times of the design instantaneous luminosity, and accumulating  $3000 \text{ fb}^{-1}$   $pp$  data.

Table 3.1: LHC operations.

Year	Purpose	COM energy	Integrated $\mathcal{L}$ [ $\text{fb}^{-1}$ ]
2010	Commissioning	7 TeV	0.04
2011	Run 1 - Exploring limits	7 TeV	6.1
2012	Run 1 - Production	8 TeV	23.1
2013	Long Shutdown 1	-	-
2015	Run 2 - Production	13 TeV	160
2019	Long Shutdown 2	-	-
2020	Run 3 - Production	14 TeV	-
2023	Long Shutdown 3	-	-
2025	HL-LHC	14 TeV	-

On the LHC ring, there are four main detectors to explore the high energy beam collisions. ATLAS and CMS are general purpose detectors for high luminosity  $pp$  data with the targeted luminosity value up to  $L = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . Different than these two biggest detectors which surrounds the entire collision point, there is LHCb detector focusing to the particles that are thrown forward direction of the collision. LHCb studies mainly B-physics with a lower luminosity target of  $L = 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ . And lastly, ALICE detector analyses heavy ion collisions, which are lead-lead (Pb-Pb), Pb-p or p-Pb collisions, up to luminosity of  $L = 10^{27} \text{ cm}^{-2}\text{s}^{-1}$ .

The two searches presented in this thesis use the ATLAS  $pp$  data collected during

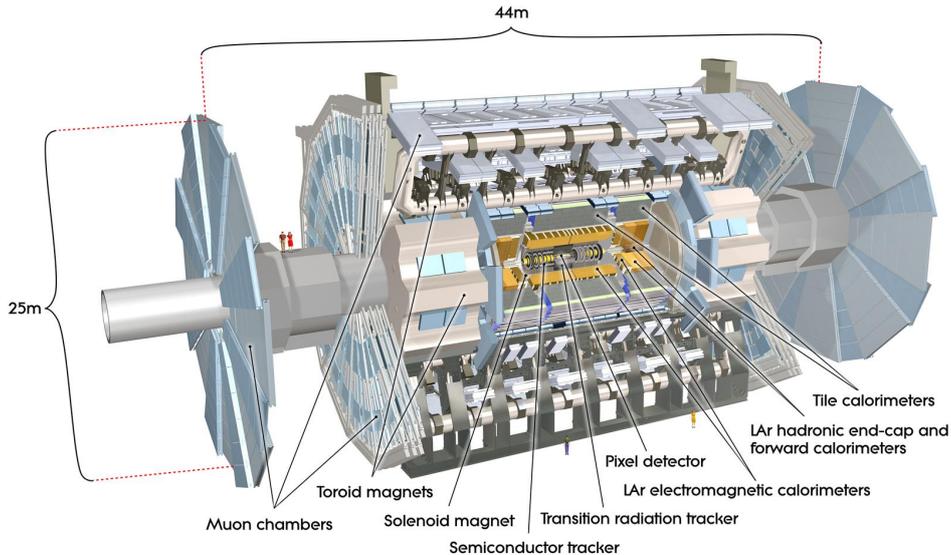


Figure 3.3: Atlas detector [51].

2015, 2016 and 2017. Therefore detailed explanations about the ATLAS detector are presented in the next sections. For more information about the other LHC detectors, see References [47–49].

## 3.2 The ATLAS detector

ATLAS (A Toroidal LHC ApparatuS) Collaboration comprises about 3000 physicists from over 181 institutions in 38 countries by the time this thesis is written [50]. The detector itself is as big as the collaboration; 44 meters long, 25 meters in height and the weight of the detector is about 7000 tonnes. See Figure 3.3 to visualise the size of the detector by comparing with the people standing on it [51].

In the ATLAS detector, the nominal interaction point is accepted as the origin of a right handed coordinate system, and the beam direction defines the  $z$ -axis. According to this, the  $x - y$  plane is the plane transverse to the beam direction; the  $x$ -axis points from the interaction point to the centre of the LHC ring, and the  $y$ -axis points upwards. The side with positive  $z$  is called the A side and the other side with negative  $z$  values is C side. Unless stated otherwise, all the transverse variables for momentum ( $p_T$ ), energy ( $E_T$ ) and the missing energy ( $E_T^{\text{miss}}$ ) are defined in the  $x - y$  plane.

The azimuthal angle  $\phi$  is measured around the beam axis, and the polar angle  $\theta$  is the angle from the beam axis. The pseudorapidity, which is used to describe the angle of a particle relative to the beam axis is defined as  $\eta = -\ln \tan(\theta/2)$ . The distance  $\Delta R$  in the pseudorapidity-azimuthal angle space is defined as  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ .

Five helix parameters are used to describe the trajectories of the charged particles in ATLAS. In the  $x - y$  plane these are:  $\frac{1}{p_T}$ ,  $\phi$ , and the transverse impact parameter ( $d_0$ ) that is defined as the transverse distance to the beam axis at the point of closest approach [52], signed according to the reconstructed angular momentum of the track about the axis. In the  $R - z$  plane these are:  $\cot \theta$  and the longitudinal impact parameter ( $z_0$ ), defined as the  $z$  coordinate at the point of closest approach.

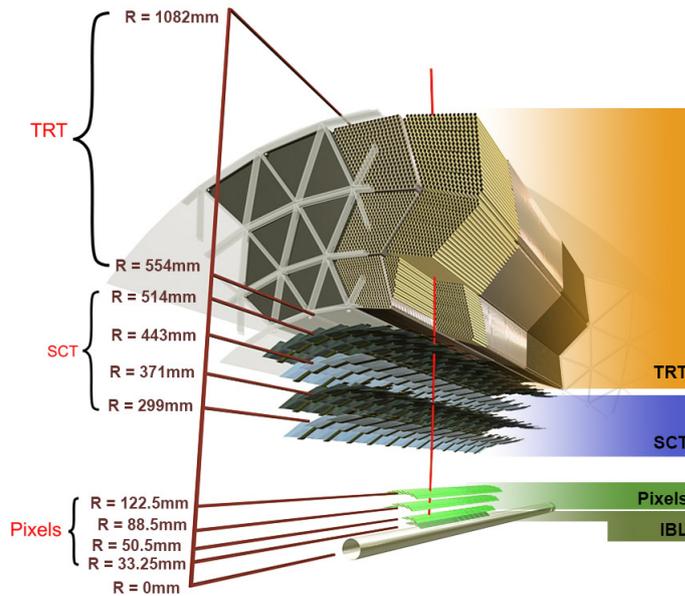


Figure 3.4: ATLAS inner detector [54].

### 3.2.1 Inner detector

The innermost part of the ATLAS detector, the Inner Detector (ID), is mainly responsible for the tracking of charged particles, their vertex identification and momentum measurements within the pseudorapidity range  $|\eta| < 2.5$ . In Table 3.2, the measured resolutions of tracking parameters are shown for both simulation and data.

Table 3.2: Track parameter resolutions for tracks with  $p_T > 30$  GeV in cosmic-ray 2008 data and simulation [53].

Parameter	Resolution in Data	Resolution in Simulation
$d_0$ [ $\mu\text{m}$ ]	$22.1 \pm 0.9$	$14.3 \pm 0.2$
$z_0$ [ $\mu\text{m}$ ]	$112 \pm 4$	$101 \pm 1$
$\phi_0$ [ $\text{mrad}$ ]	$0.147 \pm 0.006$	$0.115 \pm 0.001$
$\theta$ [ $\text{mrad}$ ]	$0.88 \pm 0.03$	$0.794 \pm 0.006$
$q/p$ [ $\text{GeV}^{-1}$ ]	$(4.83 \pm 0.16) \times 10^{-4}$	$(3.28 \pm 0.03) \times 10^{-4}$

The ID is located around the beam pipe and inside a solenoidal magnetic field of 2 Tesla. The outer radius of the ID cavity is 115 cm and its length is about 7 m. As can be seen in Figure 3.4, it has three sub-detectors which are from the inside out the pixel detector, the semiconductor tracker (SCT), and the transition radiation tracker (TRT), respectively. They are placed around the beam axis concentrically in the barrel region and perpendicular to the beam axis in the end-cap regions on disks. Below, the detailed description of these sub-detectors are given.

#### Pixel detector

Considering the requirements of the targeted physical processes in ATLAS, it is possible to understand the necessity of performing high-precision measurements in detectors. To

achieve that, fine granularity is required and pixel detector provides the highest granularity around the vertex region. High-resolution pattern recognition is achieved by discrete space points from silicon pixel layers crossed by each charged track. Pixel layers consist of identical pixel sensors with minimum pixel size in  $(R - \phi) \times z$  of  $50 \times 400 \mu\text{m}^2$  and have 80.4 million readout channels [51]. The expected hit resolution is  $10 \mu\text{m}$  in  $(R - \phi)$  and  $115 \mu\text{m}$  in  $z$ . Barrel and both end-caps have a total of 1744 modules. The pixel detector is the most important contributor in achieving a small impact parameter resolution and in identifying short-lived particles such as  $b$ -hadrons and  $\tau$  leptons.

In Run 1, the pixel detector had three barrels at radii of  $50.5 \text{ mm}$ ,  $88.5 \text{ mm}$  and  $122.5 \text{ mm}$ . However, with the increasing luminosity in Run 2, better performance of the detectors is required. Therefore, an extra silicon pixel detector, which is called inserted B layer (IBL), at  $33.2 \text{ mm}$  from the beamline was inserted during LS1 [55]. Apart from the modules mentioned above, IBL has also 280 silicon pixel modules arranged on 14 azimuthal carbon fiber staves, with the expected hit resolution of  $\sim 8 \mu\text{m}$  in  $R - \phi$  and  $\sim 40 \mu\text{m}$  in  $z$  and the size  $R - \phi \times z$  of  $50 \times 250 \mu\text{m}^2$ . As shown for  $d_0$  and  $z_0$  parameters resolution in Figure 3.5, IBL improves the tracking performances leading to better vertexing and  $b$ -hadron identifications, that will be clarified later in Section 5.1.6.

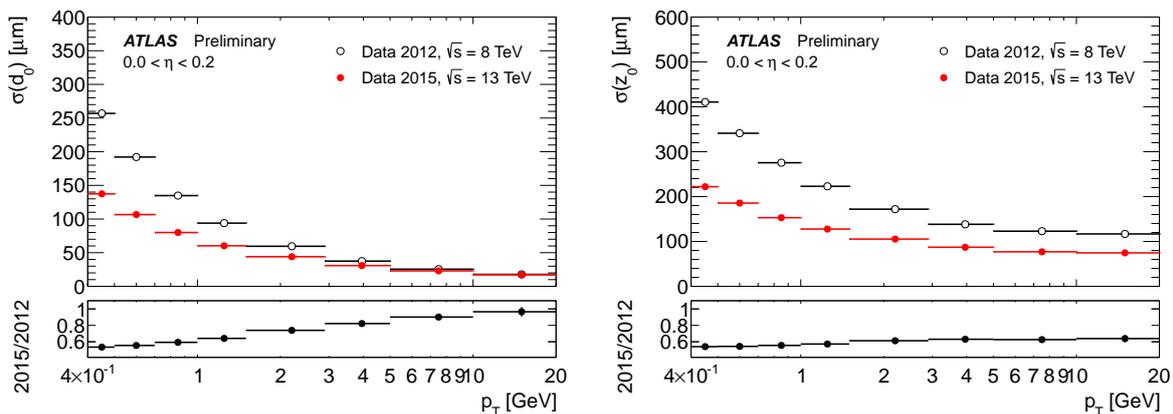


Figure 3.5: Comparison for transverse ( $d_0$ ) and longitudinal ( $z_0$ ) impact parameter resolutions measured from 2012 and 2015 data as a function of  $p_T$  [56].

### Semiconductor tracker

The semiconductor tracker (SCT) is placed between the pixel detector and TRT. It contributes to the measurement of momentum, track impact parameters, and vertex position. In the barrel region, SCT has four cylindrical layers at radii of 299, 371, 443 and 514 mm, consisting of 2112 modules in total. Additionally to the barrel region, there are 9 disks in the end cap regions contributing 1976 modules. Each module has 4 silicon sensors, two each on the top and bottom sides. These sensors are made from p-type silicon  $80 \mu\text{m}$  pitch microstrips implanted on n-type wafers. The top and bottom side sensors are aligned with a stereo angle of  $40 \text{ mrad}$ . This angle provides a measurement of the  $z$  coordinate, increasing the efficiency of track identification. These eight layers of silicon microstrip detectors provide precision points in the  $R - \phi$  and  $z$  coordinates [51]. SCT has nominal resolution of  $17 \mu\text{m}$  in the  $R - \phi$  plane and  $580 \mu\text{m}$  in the longitudinal plane ( $z$  or  $R$ ). The end-cap modules are also very similar to the barrel modules, only they have tapered strips with one set aligned radially.

### Transition radiation tracker

The outer-most region of the ID is the transition radiation tracker (TRT) which consists of 4mm diameter straw tubes filled with the nonflammable gas mixture of %70  $Xe$ , %27  $CO_2$  and %3  $O_2$ . Each straw tube has an anode sense wire at their centre to create high voltage difference with respect to the straw tube wall that serves as cathode, so that charged particles can ionise the gas as they pass through inside the tubes. Ionised gas is collected and converted to electrical signals. Using the information of arrival times of the ionised pairs, hits can be generated in TRT.

Due to the high cost of the materials used in the SCT and pixel detectors, precision layers are limited to three pixel layers and eight strip layers which are crossed by each track. On the contrary, TRT identifies large number of tracking points, about 36 per track, providing continuous track following up to  $|\eta| < 2.0$  with much less material per point and a lower cost. Lower precision of the TRT hits is compensated by the large number of measurements and TRT contributes significantly to the momentum measurements, electron identification and especially for the detection of photon conversions and  $V0$  decays [51].

In the barrel region, TRT has 52544 straws which are parallel to the beam axis while in the end-caps there are 122880 straws in each side arranged radially in wheels. Each straw gives 130 mm spatial resolution in  $R-\phi$  plane, with 420000 electronic channels. The drift-time information coming from the each electronic channel is used for the discrimination between the transition radiation hits and tracking hits.

In Run 1, several leaks are identified in the gas pipes which are mostly located in inaccessible areas and their repair was not possible. Considering the high cost of Xe-based gas mixture, significantly less expensive argon-based (Ar-based) gas mixture was used in the TRT modules with high leak rates in Run 2 [57].

### 3.2.2 Calorimeters

Placed between the ID and the muon spectrometer and outside of the 2T solenoid magnet, calorimeters are the ultimate destinations for many particles except neutrinos and muons. They provide accurate energy measurements for the particles which interact electromagnetically or hadronically with the calorimeters. Depending on the interaction type, separate techniques and materials are used. Therefore, two types of calorimeters exists in ATLAS; hadronic and electromagnetic calorimeters. Both of them are *sampling calorimeters*, meaning that they are composed of alternating layers made of different materials known as *active material* and *absorber material* [51]. While absorber materials are dense and initiate the showers, active parts are responsible for measuring the energies of the final particles produced showers. Using those energies, the energy of the initial particle is obtained by reconstructing the shower. It is important to stop the particle showers within the calorimeters before they escape to the muon system in order to perform a precise energy measurement. Therefore, the thickness of the calorimeters are determined as large radial depths so that both electromagnetic and hadronic showers can be fully contained within the calorimeter. In terms of pseudorapidity coverage, they also provide large range of measurement up to  $\eta < 4.9$ , having separate parts in the barrel region and in the end caps (see Figure 3.6). Lastly, the calorimeters are designed as segmented structures, divided into many small cells in order to reconstruct the direction and shape of the shower precisely using the individual cell information. Below, brief descriptions for

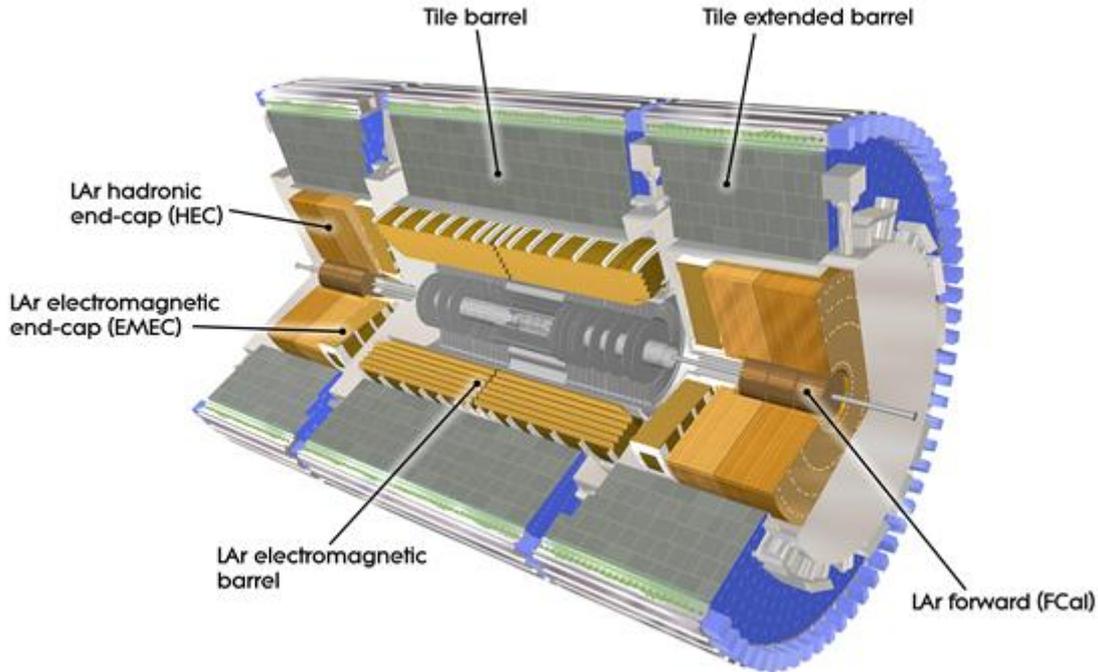


Figure 3.6: ATLAS calorimeters [51].

the ATLAS calorimeters, the innermost layer electromagnetic calorimeters and the outer layer hadronic calorimeters, are given.

### Electromagnetic calorimeter

Electromagnetic (EM) calorimeters are responsible for the energy measurements of the electromagnetically interacting particles such as electrons, positrons, photons as well as hadrons. These particles can interact with the matter in various ways creating electromagnetic showers. For energies above 100 MeV, electrons and positrons lose their energies almost entirely through *bremstrahlung*, producing energetic photons. For photons, the major interaction is the pair production process creating another energetic electron or photon.

In the ATLAS detector, the electromagnetic calorimeter is a Lead-liquid Argon (LAr) detector which uses LAr as an active material and Lead as absorber. To gain a complete and uniform coverage, the EM calorimeter is shape-wise designed similarly to the geometry of an accordion. The pseudorapidity range is between  $1.375 < |\eta| < 3.2$  in the two end caps and  $|\eta| < 1.475$  in the barrel region. The thickness of the calorimeters is  $> 22$  radiation lengths ( $X_0$ ) in the barrel region and  $> 24X_0$  in the end caps [51].

### Hadronic calorimeter

The hadronic calorimeters measure the energy of strongly interacting particles such as  $\pi^\pm$ ,  $K^\pm$ ,  $K^0$ , protons and neutrons. As a standard detector design, the hadronic calorimeter is placed behind the EM calorimeter to be able to absorb the energy of hadronic showers which is much larger than the electromagnetic showers.

The ATLAS hadronic calorimeter is divided into three different parts: Tile calorimeters, two hadronic end-cap calorimeters (HEC) and two forward calorimeters (FCal) (see Figure 3.6). Tile barrel calorimeters are located in the region up to  $|\eta| = 1.0$  and tile extended barrel calorimeters cover the region of  $0.8 < |\eta| < 1.7$ . While as an absorber material steel tiles are used, plastic scintillating tiles are chosen for the active layer which is much thinner than the steel layer. The total thickness of the tile calorimeters is approximately 1.97 m. Particles passing through the scintillating tiles produce light that is proportional to the incident energy and that light is detected via photo-multiplier tubes (PMTs). In the hadronic calorimeter, cell size is coarser than the EM calorimeter since hadronic showers are larger than the EM showers.

While the region between  $1.5 < |\eta| < 3.2$  is covered by the HEC calorimeters, the very forward region between  $3.1 < |\eta| < 4.9$  is spanned by FCals providing a large  $\eta$  coverage to the ATLAS Calorimeter system. Both the HEC calorimeters and the first layer of FCal are Copper-LAr sampling calorimeters, using copper as absorber and LAr as active material. Only the outer two layers of FCal are made of tungsten instead of copper. Those calorimeters are designed to be extremely dense to contain the showers since there is a limitation for space.

### 3.2.3 Muon spectrometer

The muon spectrometer (MS) is the outermost part of the ATLAS detector. It is responsible for the detection of charged particles escaping the calorimeters and measuring their momentum in the pseudorapidity region up to  $|\eta| < 2.7$ . Typically at the LHC, the only SM particles that are actually interacting with matter and can survive until MS are muons. Rarely in high energies, a jet is not fully absorbed by the calorimeters and they can reach until MS. These events are called punch through events and specifically treated [58].

MS has three large superconducting air-core toroid magnets; one is located in the barrel region and the other two are in the end caps. While the barrel toroid magnet provides 1.5 to 5.5 Tm of bending power in the range of  $0 < |\eta| < 1.4$ , the end-cap magnets give approximately 1 to 7.5 Tm in the region  $1.6 < |\eta| < 2.7$  [51]. They create a non-uniform magnetic field so that the particle trajectories are bent in the MS with respect to their  $p_T$  values. Measuring these trajectories allows us to find out the  $p_T$  of the particles with about 10% resolution.

Apart from the magnets, MS consists of a trigger system and high-precision tracking chambers. It has three concentric cylindrical layers around the barrel region with radii 5 m, 7.5 m, 10 m and two large wheels in the end-cap regions. There are four types of detectors in the MS. Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC) are responsible for high precision tracking. MDTs are cylindrical drift tubes filled with a gas mixture of Ar and CO<sub>2</sub> covering the pseudorapidity region of  $2 < |\eta|$ . Over the region of  $2 < |\eta| < 2.7$ , CSCs are active with a higher granularity. They provide higher rate capacity and better time resolution. CSCs are multiwire proportional chambers whose cathodes are segmented into strips to provide full 3D space points.

The other two detectors are Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC), used for triggering the events with muons. Their response time is shorter comparing to the other two detectors. RPCs are gaseous tracking chambers composed of two resistive parallel plates. Between those plates an electric field is formed so that an avalanche is created when a muon enters the region. The spacetime resolution is about

1 cm  $\times$  1 ns with digital readout. TGCs are responsible for triggering in the end cap regions providing time resolution of 4 ns. RPCs and TGCs are not only serving for triggering, but also contributing to the bunch-crossing identification and measurements for the coordinates of the muons.

During LS1, additional RPC-equipped MDT chambers are added to the MS as it has been initially designed [59]. The new layer is in the transition region between the barrel and the endcaps,  $1 < |\eta| < 1.4$  to improve the acceptance in that region compared to Run 1.

### 3.2.4 Trigger and data acquisition system

Considering the enormous number of collisions happening per second in LHC, the importance of the decision making process to save the interesting events can be understood. First of all, it is not suitable to record that much event to the disks in terms of storage spaces. Secondly, it is not needed to record all of events, since not all of them carry useful information for the physics analyses. Still, ATLAS is a general purpose detector and there is not only few physics signatures to be considered but lots of them. Therefore, ATLAS trigger and data acquisition system (TDAQ) is designed to select and save the interesting events for the physics analyses according to the existent resources of the experiment and it operates with an outstanding performance by decreasing the final event rate from about 40 MHz to 1 kHz.

In Run 2, ATLAS triggering system operated in two different trigger levels which are Level 1 (L1) and High Level Trigger (HLT). While L1 is a hardware based first level trigger with specifically designed electronics that are operating in runtime, the HLT is the software part of the TDAQ system, mostly built with the standard resources. In addition to these two levels, there are also some modules interfacing these two levels and making possible the data flow from the detectors to the disks as shown in Figure 3.7. The decision making process for an event is explained below in detail.

L1 trigger decisions are made in the Central Trigger Processor (CTP) using the inputs coming from the L1 calorimeter (L1 calo), L1 muon systems, and several other subsystems such as the Minimum Bias Trigger Scintillators (MBTS), the LUCID Cherenkov counter, and the Zero-Degree Calorimeter (ZDC) [60]. Each system are responsible from different type of trigger inputs and trigger decisions, physics analyses uses objects triggered by the L1Calo and L1Muon triggers as their name suggested the objects which deposits energy in Calorimeters or Muons recognised by the Muon Systems and Tile Calorimetry. But in addition to these triggers, there is also one new module responsible from the newly introduced topological triggers (L1 topo) using the geometric or kinematic association between the trigger objects coming from L1 Calo or L1 Muon systems. CTP handles the L1 acceptances as well as the concept of dead time, which is a protection mechanism to prevent overlaps between readout windows and overflows in front end buffers.

In L1 stage, specific units are defined in the detector as regions of interests (RoIs) and each of these regions carry the information of the accepted objects, the thresholds together with their location information. After the decision made in CTP and the L1 trigger is fired, Readout Drivers (RODs) carry the buffered data from CTP to the Readout System (ROS) and stored in the Readout Buffers (ROBs). HLT receives RoI information from L1 and reconstructs the region using more complex algorithms like the offline reconstruction algorithms and makes the final decision. If the HLT decision is also positive, events are transferred to the local storage at the site and then sent to the Tier-0 facility at CERN's

computing centre for offline reconstruction [60].

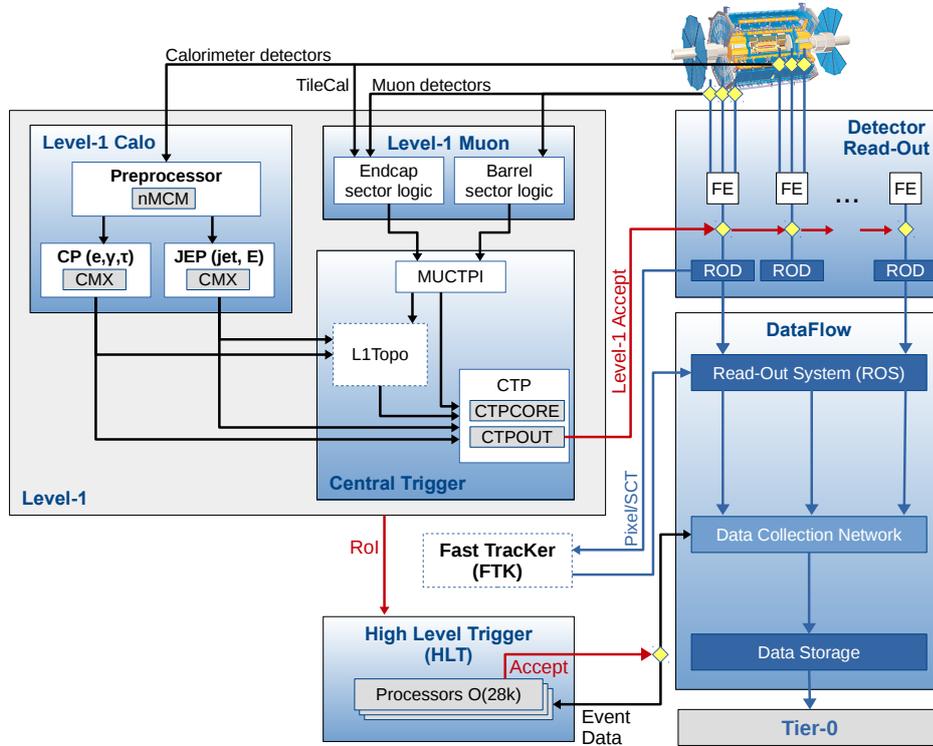


Figure 3.7: Schematic overview of the Trigger and DAQ system [61].

As mentioned above, there are lots of signatures to consider for different physics analyses and there are limitations in terms of total data storage and the event writing speed to the disks, total bandwidth of the system. Therefore, optimisation studies are done to allocate the resources between different triggers considering the operational circumstances, the needs of the different analyses groups, and the strategy of the experiment. In case the number of events passing a specific trigger is higher than the allocated bandwidth for that trigger, that trigger is *prescaled* by a certain number, meaning that only one event is recorded out of the number of triggered events defined by the prescale factor. Another important thing to mention here is the trigger threshold optimisations to handle the excessive rates in certain triggers. Especially with the increased luminosity values and centre-of-mass energies, the rates for the certain process increase rapidly. Therefore, the thresholds, minimum selection criteria for the quantities like  $p_T$  or energy, have to be increased in order to cope with the higher rates seen in those triggers. Figure 3.8 shows the increase in the L1 rate for some triggers with respect to the increased luminosity values.

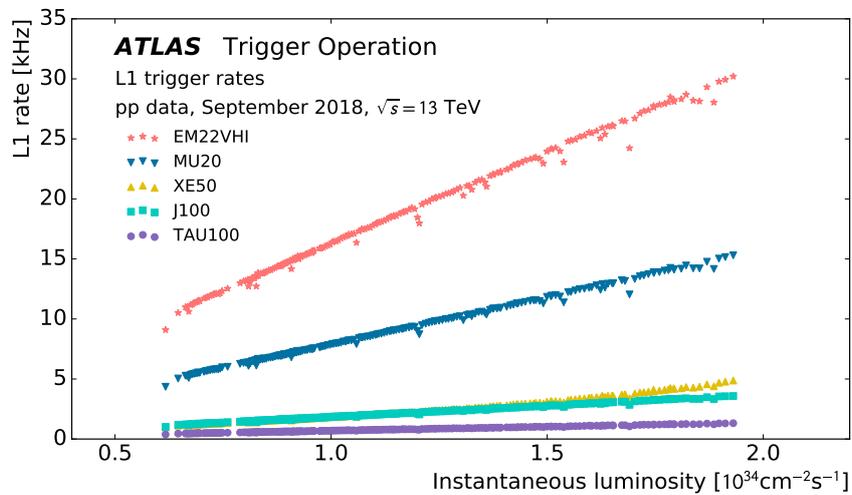


Figure 3.8: Level-1 (L1) physics trigger rates as a function of instantaneous luminosity in a fill taken in September 2018 with a peak luminosity of  $\mathcal{L} = 2.0 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$  and a peak average number of interactions per crossing of  $\langle \mu \rangle = 56$ . Presented are rates of some representative single-object trigger items, which have not been prescaled. These trigger items are based on such objects as electromagnetic clusters (EM), muon candidates (MU), jet candidates (J), missing transverse energy (XE) and tau candidates (TAU). The number in the trigger name denotes the trigger threshold in GeV. The other text refers to details of the selection: variable thresholds (V), hadronic isolation (H), and electromagnetic isolation (I). Dips in the rates are due to dead-time and spikes are caused by detector noise [61].

# Chapter 4

## Data and Simulations

This chapter describes the data and the Monte Carlo (MC) simulations used in this thesis. First, information on the datasets used in the searches presented in Chapter 7 and 8 is given. Then, the MC simulations that are used in the two searches and in the boosted Higgs identification studies presented in Chapter 6 are introduced.

Both the datasets and the simulated samples are processed with the same selections producing intermediate data formats, known as *derivations*, which are commonly used in ATLAS Run 2 analyses. While the two searches are using a common derivation processed for the new physics searches [62], the studies for Higgs identification uses another one optimised for flavour identification studies [63]. Each presented study is processed with the latest available ATLAS software at their time as given in Table D.1.

### 4.1 Data

Recalling from Section 3.1, the LHC operates in different centre-of-mass energies and provides different integrated luminosities per each year as shown in Figure 3.2. This thesis presents two data analyses based on the ATLAS Run 2  $pp$  dataset with the 13 TeV centre-of-mass energy. The detector records data in different periods to ease the separation of the datasets according to their quality and the quantity. Each period of dataset is reprocessed separately so that it is possible to remove the collisions recorded in poor detector conditions, using so-called *Good Run Lists* (GRLs) provided by the ATLAS data processing group [64] as given in Table D.1.

After removing collisions recorded in poor detector from 2015, 2016, and 2017, the total data corresponds to an integrated luminosity of  $\int Ldt = 80.5 \text{ fb}^{-1}$ . While the analysis presented in Chapter 8 uses the whole dataset, in Chapter 7, only the datasets from 2015 and 2016 are used, since the analysis was performed earlier. Summing the  $\int Ldt = 3.2 \text{ fb}^{-1}$  from 2015 and  $\int Ldt = 32.9 \text{ fb}^{-1}$  from 2016, the total data used for that search corresponds to an integrated luminosity of  $\int Ldt = 36.1 \text{ fb}^{-1}$ .

At the LHC, the collisions of  $pp$  bunches result not only in hard-scatter interactions, but also large amount of additional collisions accompanying the signal. Those collisions are referred as *pile-up* interactions. One important distinction between the different years of data taking is their pile-up profiles, which is expected due to the increased instantaneous luminosities in later years. While the mean number of  $pp$  interactions per bunch crossing ( $\mu$ ) was 14 and 25, respectively in 2015 and 2016 datasets, in 2017 it increased to 55, creating a more challenging data taking environment for the detector operation and physics analyses.

## 4.2 Monte Carlo simulations

This section introduces briefly all simulated samples used in this thesis, while more information is available in given references [65–67]. Firstly, the simulations that are used to model SM processes are introduced. Secondly, the simulations generated to model signals of new physics scenarios are introduced together with the relevant parameters for the corresponding BSM models.

The impact of pile-up events is considered in all simulated samples by overlaying soft  $pp$  collisions events generated with PYTHIA 8 [68]. The response of the ATLAS detector is modelled using the full detector simulation in GEANT 4 [69]. In all simulations, event reconstruction is performed using the same software that is used to reconstruct events in the data. Decays of  $b$ - and  $c$ -hadrons are modelled using EvtGen [70], except the events generated by Sherpa [71] that will be mentioned later in Section 4.2.1.

### 4.2.1 Simulations of Standard Model processes

Below, the description of the simulated MC samples for the SM processes are introduced.

#### Dijets

The most dominant background in the analyses presented in this thesis is multijet background from QCD interactions. In order to mimic this background, simulated QCD dijet events are generated by PYTHIA 8 [68] with the A14 tune [72] and the NNPDF 2.3 NLO parton distribution function (PDF) set [73].

Thirteen samples are generated for different  $p_T$  ranges of jets up to 5 TeV as shown in Table 4.1. The total number of events in each sample is maximised under the constraint of available computing power to have max statistics. Before using all the samples together, the number of events in each individual sample are brought into agreement to have a flat jet  $p_T$  spectrum applying additional event weights. However, total number of events in these simulations is not comparable with the number of events in the Run 2 datasets. Therefore, they are used mostly for optimisation of the signal selections, truth studies, and the background investigations.

#### Top-Antitop ( $t\bar{t}$ )

Production of  $t\bar{t}$  process is a significant background for the analyses presented in this thesis. Simulated events are split into different categories depending on the topology of the decay either both top quarks decay hadronically, only one decays hadronically, or both of them decay leptonically.

In Chapter 6 and 7, the simulated  $t\bar{t}$  samples are generated at tree-level using POWHEG-BOX v1 [75] and the CT10 PDF set [76]. The hadronisation is performed with PYTHIA 6 [77] with the CTEQ6L1 PDF set [78] and Perugia 2012 tune [79]. Apart from these, additional samples simulated for the  $Z' \rightarrow t\bar{t}$  BSM process are used to increase the number of events in the higher transverse momenta, so that the hadronically decaying top quarks can be modelled for the boosted Higgs identification studies. These simulations are generated over a range of  $Z'$  boson masses between 400 and 5000 GeV using Pythia8 with the LO NNPDF2.3 PDF set and the A14 underlying event tune.

In Chapter 8, POWHEG-BOX v2 [80] is used for the event generation with NNPDF2.3 NLO PDF set and hadronisation is simulated with PYTHIA 8 with the A14 tune and

Table 4.1: Pythia dijet samples are listed together with the corresponding cross-sections and  $p_T$  range information [74].

Sample	Cross-section(nb)	Number of events	Truth jet $p_T$ [GeV]
JZ0W	7.8420E+07	2000000	0-20
JZ1W	7.8420E+07	2000000	20 -60
JZ2W	2.4334E+06	1992000	60 -160
JZ3W	2.6454E+04	1767000	160-400
JZ4W	2.5464E+02	1997000	400-800
JZ5W	4.5536E+00	1995000	800-1300
JZ6W	2.5752E-01	1997000	1300-1800
JZ7W	1.6214E-02	1990000	1800-2500
JZ8W	6.2505E-04	2000000	2500-3200
JZ9W	1.9640E-05	2000000	3200-3900
JZ10W	1.1961E-06	2000000	3900-4600
JZ11W	4.2260E-08	1999000	4600-5300
JZ12W	1.0370E-09	1808000	5300-infinity

the NNPDF2.3 LO PDF set. Apart from these samples used for the final results of the analysis, a second set of  $t\bar{t}$  events is generated by SHERPA 2.2.1 [71] using the NNPDF30 NNLO PDF set [81] to estimate the modelling uncertainties in the analysis.

### V+jets

While  $Z$ +jets processes are background for both presented searches, the  $W$ +jets processes are important only for the di- $b$ -jet resonance search presented in Chapter 8.

The background contribution of  $Z$ +jets processes is minor for the analysis presented in Chapter 7. It is modelled using the samples generated with Pythia 8.186 with the NNPDF2.3 LO PDF set. For the di- $b$ -jet resonance search, the hadronically decaying  $W$  and  $Z$  events are generated using SHERPA 2.1.1 [71] with up to 4 additional partons at leading order and with PDF CT10 [76]. All these samples are separated into several orthogonal datasets based on the  $p_T$  of the vector boson. Alternative samples for hadronically decaying  $W$ +jets and  $Z$ +jets are also used and they are generated using Herwig++ 2.7.1 [82] with CTEQ6L1 PDF set [78]. They are also separated into several orthogonal datasets based on the  $p_T$  of the vector boson [67].

### Higgs

The SM Higgs processes are considered as both signal and backgrounds for the search presented in Chapter 8. As discussed in Section 2.1.3, there are four SM Higgs production mechanisms, however only three of them are considered: gluon gluon fusion ( $ggF$ ), vector boson fusion ( $VBF$ ), and Higgsstrahlung.

The events of the Higgs boson produced via  $ggF$  and  $VBF$  are generated using the MiNLO prescription with finite top mass [83, 84] using POWHEG-BOX v2 [80] with the NNPDF30 NNLO PDF set and they are showered using PYTHIA 8.212 with the AZNLO tune [85] and the CTEQ6L1 PDF set [78].

The Higgsstrahlung Higgs boson events are generated using PYTHIA 8.212 with the AZNLO tune and the CTEQ6L1 PDF set. Following the recommendations given in

Reference [86], the missing process of  $gg \rightarrow ZH$  in PYTHIA is corrected.

### SM Higgs boson pair production

These simulations are produced via the  $ggF$  production mode and simulated with MADGRAPH5\_aMC@NLO using form factors for the top-quark loop from HPAIR [87, 88] in order to take into account the finite top-quark mass effects. Following Reference [89], the cross section times branching ratio to the  $b\bar{b}b\bar{b}$  is evaluated at next-to-next-to-leading order (NNLO).

## 4.2.2 Simulations of Beyond the Standard Model processes

### $G_{KK}$ in Bulk RS Model

As introduced in Section 2.3.1, the  $G_{KK}$  particle used as benchmark in the context of di-Higgs production search presented in Chapter 7. 20 signal samples are produced with the mass points from 300 to 3000 GeV for the process of  $G_{KK} \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ . Madgraph generator [90] is used with the NNPDF2.3 LO PDF set [91]. The hadronic showers are produced in PYTHIA 8 with the A14 tune [72].

The Higgs mass has been set to 125.0 GeV for all the mass points and the model parameter  $c = k/\bar{M}_{P1}$  to 1.0 and 2.0, respectively. The values for the cross-section times branching ratio are given in Tables 4.2 and 4.3.

The samples produced with  $c = k/\bar{M}_{P1} = 1$  are also used for the boosted Higgs identification studies presented in Chapter 6. In order to assess the performance of the boosted Higgs identification strategies and optimise the methods for different topologies, all generated samples for different mass points are used together after applying a dedicated *reweighting* method [92].

### Narrow-width scalar

The simulations of the process  $gg \rightarrow \text{Scalar} \rightarrow HH \rightarrow b\bar{b}b\bar{b}$  are generated at LO in QCD using MADGRAPH5\_aMC@NLO with CT10 PDF sets. Hadronisation is performed with Herwig++ using CTEQ61 PDF sets. No specific model is used for computing the cross sections.

### $Z'$ mediator

$Z'$  mediator predicted by simplified Dark Matter model [93] is one of the benchmarks that is used in the di- $b$ -jet resonance search presented in Chapter 8, as discussed in Section 2.3.2. Considering the given model, signal samples are generated with MADGRAPH5\_aMC@NLO with the NNPDF 3.0 LO PDF at several  $Z'$  masses ( $m'_{Z'}$ ) spanning the search region between 100 GeV and 300 GeV in 25 GeV steps. Each simulated event contains a leptophobic  $Z'$  decaying to two quarks and produced in association with a jet. The showering is done using PYTHIA 8 with the A14 tune and the NNPDF23 LO PDF set [73]. In addition to these events, independent samples are generated containing events where  $Z'$  decaying to light and  $b$ -quarks, to enhance the number of simulated signal events. The absolute axial coupling defined by the model is chosen to be  $g_q = 0.25$ , same for all quark generations. This value does not significantly affect the kinematics of the reconstructed events due to the smaller natural width of the  $Z'$  than the large- $R$  jet mass

Table 4.2: Cross-section times branching ratio for RS graviton samples with  $k/\bar{M}_P = 1.0$  as a function of the graviton mass [74].

$m_{G_{KK}}$ (GeV)	$\Gamma_{G_{KK}}$ (GeV)	$\sigma \times \text{BR}(G_{KK} \rightarrow hh)$ (fb)	$\text{BR}(G_{KK} \rightarrow hh)$	$N_{events}$
300	8.365	$1319.9 \pm 1.0$	0.90	79000
500	18.43	$892.4 \pm 0.6$	6.43	93400
600	26.08	$410.4 \pm 0.3$	6.95	99000
700	33.65	$201.48 \pm 0.15$	7.19	54000
800	41.06	$105.49 \pm 0.07$	7.33	70000
900	48.30	$58.35 \pm 0.04$	7.41	85000
1000	55.40	$33.68 \pm 0.02$	7.47	100000
1100	62.38	$20.23 \pm 0.01$	7.51	99000
1200	69.27	$12.54 \pm 0.01$	7.54	99000
1300	76.09	$7.979 \pm 0.005$	7.56	19000
1400	82.84	$5.201 \pm 0.004$	7.58	98600
1500	89.54	$3.450 \pm 0.002$	7.59	99000
1600	96.20	$2.336 \pm 0.002$	7.60	99000
1800	109.4	$1.116 \pm 0.001$	7.62	15000
2000	122.5	$0.5559 \pm 3 \times 10^{-4}$	7.63	88800
2250	138.8	$0.2486 \pm 2 \times 10^{-4}$	7.64	99000
2500	155.0	$0.1158 \pm 1 \times 10^{-4}$	7.65	60000
2750	171.1	$0.05585 \pm 4 \times 10^{-5}$	7.66	58600
3000	187.2	$0.02772 \pm 2 \times 10^{-5}$	7.66	78000

Table 4.3: Cross-section times branching ratio for RS graviton samples with  $k/\bar{M}_P = 2.0$  as a function of the graviton mass [74].

$m_{G_{KK}}$ (GeV)	$\Gamma_{G_{KK}}$ (GeV)	$\sigma \times \text{BR}(G_{KK} \rightarrow hh)$ (fb)	$\text{BR}(G_{KK} \rightarrow hh)$	$N_{events}$
300	33.46	$9997 \pm 11$	0.90	90000
400	45.22	$8560 \pm 7$	4.99	60000
500	73.74	$3755 \pm 3$	6.43	100000
600	104.3	$1657 \pm 1$	6.95	98800
700	134.6	$789.9 \pm 0.6$	7.19	99000
800	164.2	$404.3 \pm 0.3$	7.33	99000
900	193.2	$219.3 \pm 0.2$	7.41	100000
1000	221.6	$125.1 \pm 0.1$	7.47	100000
1100	249.5	$74.19 \pm 0.05$	7.51	58600
1200	277.1	$45.48 \pm 0.003$	7.54	74000
1300	304.4	$28.72 \pm 0.02$	7.56	100000
1400	331.4	$18.55 \pm 0.001$	7.58	73800
1500	358.2	$12.27 \pm 0.001$	7.59	99000
1600	384.8	$8.254 \pm 0.005$	7.60	100000
1800	437.7	$3.913 \pm 0.003$	7.62	93400
2000	490.1	$1.951 \pm 0.001$	7.63	60000
2250	555.2	$0.8703 \pm 0.0006$	7.64	100000
2500	620.0	$0.4070 \pm 0.0003$	7.65	84000

resolution [67]. Cross section calculations are performed with the DM width calculator given in Reference [94], setting the DM particle mass to 10 TeV.

# Chapter 5

## Reconstruction

In the ATLAS detector, each sub-detector is responsible for providing different inputs to the reconstruction. With the input coming from those sub-detectors, not only the physical particles like electron, muon, photon etc. can be reconstructed as objects (see Figure 5.1), but also the signatures which do not directly represent particles, but a certain process like jets or event level quantities like transverse missing energy.

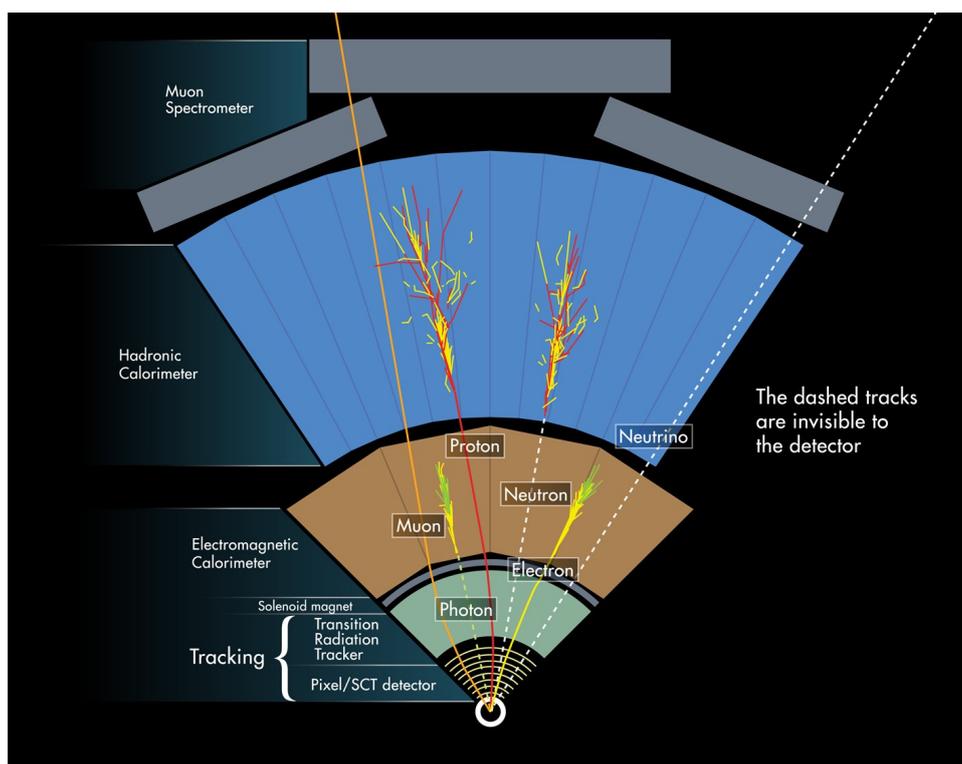


Figure 5.1: A representative sketch showing how the sub-systems of the ATLAS detector detects particles, taken from Reference [95]. While charged particles can be identified by tracks and energy deposits, neutral particles can only be detected by their energy deposits. Although showering starts already in the electromagnetic calorimeter for protons and neutrons, in the figure it is depicted only at the hadronic calorimeters to stress that the final identification of these objects requires the presence of the hadronic calorimeters.

The two searches presented in this thesis have only jets in their final state. In particular, both final states are characterised by *large-radius* ( $R$ ) jets with associated *b-jets* representing the boosted particles in the searches. Therefore, this chapter explains in

detail the reconstruction techniques for jets by focusing on the large- $R$  jets and the identification of  $b$ -jets. For completeness, small radius jets are also introduced to distinguish the difference between the boosted and non-boosted topologies. Furthermore, they are also used for the event cleaning procedure in both presented analyses, as described in Section 7.2.2 and 8.2.2. Muons are also considered in order to take into account the semileptonic  $b$ -hadron decays inside the large radius jets. Particularly for the Higgs identification, they are necessary to be able to calculate the jet mass observable more precisely (Section 6.2.2). Additionally, in di- $b$ -jet resonance search presented in Chapter 8, muons are used to create a dedicated control region for  $t\bar{t}$  background (Section 8.3.1). Hence, the reconstruction strategy for muons is also presented in this chapter.

## 5.1 Jets

As mentioned in Section 2.1.2, due to the colour confinement, quarks and gluons can not be observed in isolation but instead as colourless hadrons. In high energy physics experiments, they can be recognised only as *jets*, as a result of the fragmentation and hadronisation processes as schematically shown in Figure 5.2.

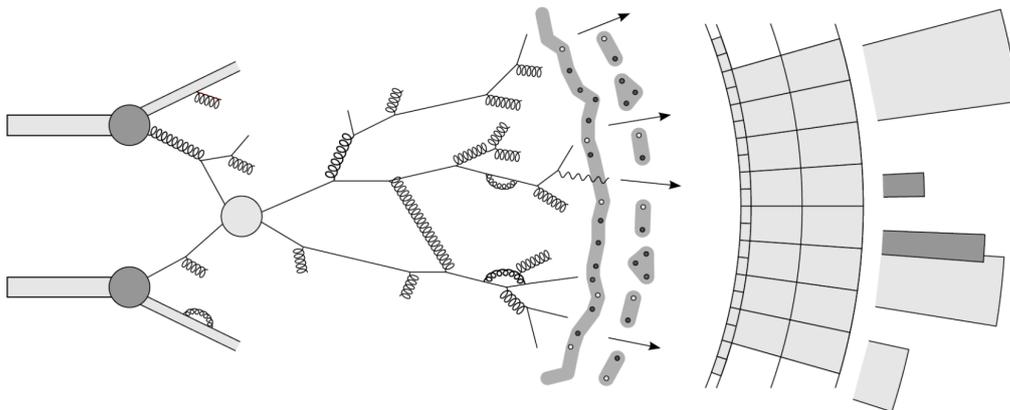


Figure 5.2: Schematic illustration of jet production and its energy deposits in the calorimeter cells [96].

At detector level, the hadrons are identified through the large energy deposits that they leave in the calorimeters and, for charged hadrons, the tracks reconstructed by the inner detector. In general, these individual signatures observed in the detector are called *constituents* of a jet, and they can be used as inputs for the jet finding algorithms. After identifying jets using these algorithms, additional stages may be applied to improve the reconstruction, so-called *jet grooming*, to suppress the undesirable contributions in the initially reconstructed jet.

This section describes the common jet finding (Section 5.1.1) and grooming algorithms (Section 5.1.2) used in the ATLAS detector, and then introduces the relevant jet reconstructions used in this thesis: small- $R$  jets (Section 5.1.3), large- $R$  jets (Section 5.1.4), and track jets (Section 5.1.5). Lastly, the identification of jets initiated by  $b$ -hadrons, *b-tagging*, is presented in Section 5.1.6.

### 5.1.1 Jet algorithms

The reconstruction of a jet is a challenging task since the correct constituents must be chosen out of many. Depending on the process or the particle which initiates the shower, different approaches can be preferred to have better reconstruction performance. Therefore, several algorithms were developed based on certain principles considering the features of jet formation. One of the key requirements for a jet finding algorithm is to take into account the property of *infrared* (IR) and *collinear safety* of a jet. These properties can be explained roughly as the invariance of the identified jets when one of the objects radiates a very soft object, or splits into two collinear objects [97]. While more detailed information can be found in the given reference, both situations can be seen schematically in Figure 5.3. While the IR safe algorithms prevent changes in the number of jets caused by soft emissions, the collinear safe algorithms keep the jets same after collinear splittings.

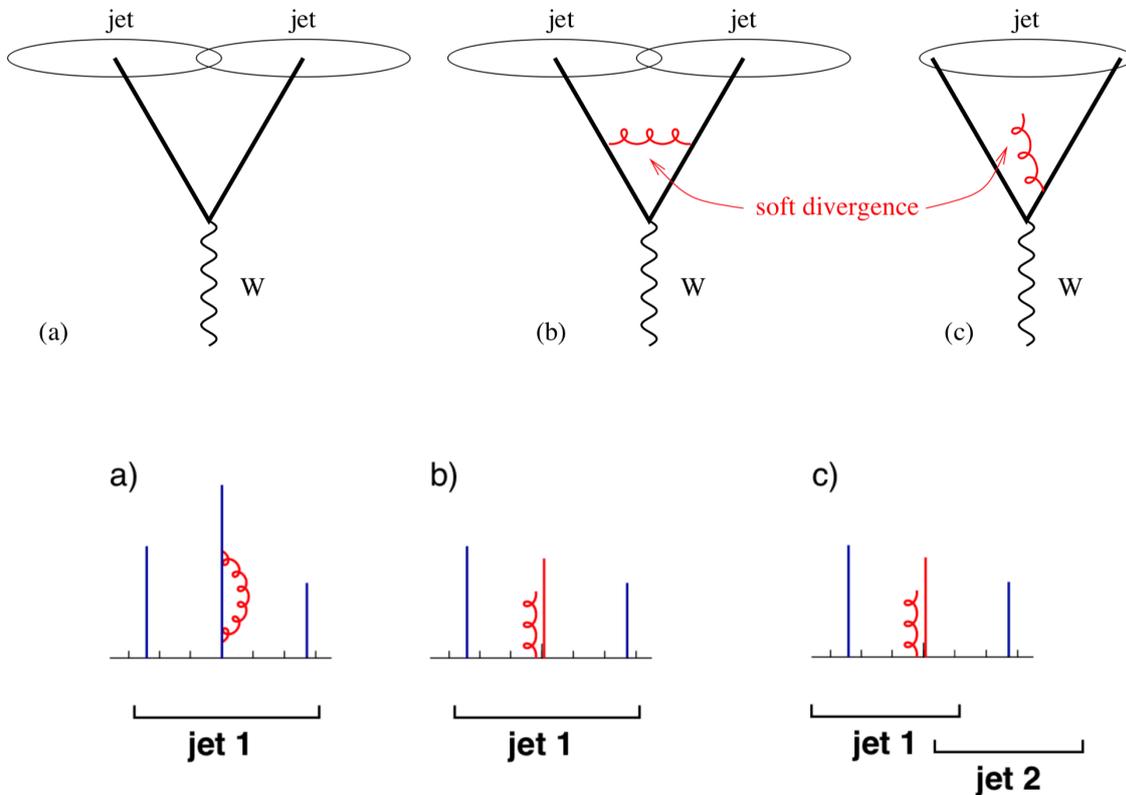


Figure 5.3: The upper row illustrates the IR safety. Subfigures shows respectively: (a)  $W$  boson decaying into two quarks forming two jets, (b) the IR safe algorithm forming still two jets despite the soft emission, (c) IR unsafe algorithm identifying only one jet. The bottom row illustrates the collinear safety where partons are indicated with vertical lines, their height is proportional to their transverse momentum, and the horizontal axis indicates rapidity. If the configuration in a) leads a single jet, b) the collinear safe algorithm with a split also leads a single jet c) the collinear unsafe algorithm leads two jets since the algorithm is heavily dependent on the initial hardest particle [97].

As mentioned, various algorithms are available to identify a jet, and they can be split into two main categories: cone and sequential recombination algorithms. Cone-type algorithms rely on the idea that QCD branching and hadronisation do not have a large impact on an event's energy flow. They follow a top-down approach while the sequential

recombination algorithms are opposite, following a bottom-up approach. The idea for the sequential recombination is to repeatedly recombine the closest pair of input constituents according to predefined distance measures [97]. They are by construction infrared and collinear-safe while the cone-type algorithms are typically not. At the LHC, mostly sequential recombination algorithms are used. Below, the common recombination scheme for obtaining  $k_T$ , anti- $k_T$ , and Cambridge-Aachen algorithms are explained. Then one of the relatively newer approaches used in ATLAS, the variable- $R$  algorithm is described.

### $k_T$ , Anti- $k_T$ and Cambridge-Aachen jets

Most common algorithms used not only in ATLAS but also in CMS are given here based on the following references [98–100]. First, a distance measure is introduced between the jet inputs  $i$  and  $j$ ,  $d_{ij}$ , and the jet input  $i$  and the beam axis  $B$ ,  $d_{iB}$ , according to the Equations 5.1 where  $R_0$  is a free parameter to determine the characteristic jet size,  $p$  is a parameter which indicates the power of  $p_T$  and the  $\Delta R_{ij}$  is defined as  $\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$  in terms of rapidity and azimuthal angle:

$$\begin{aligned} d_{ij} &= \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \Delta R_{ij}^2, \\ d_{iB} &= p_{T,i}^{2p} R_0^2. \end{aligned} \tag{5.1}$$

Based on the above definitions, the following steps are performed:

1. The algorithm computes all possible  $d_{ij}$  values, and determines the smallest value to start clustering.
2. Compares the  $d_{ij}$  and  $d_{iB}$ . If  $d_{ij} < d_{iB}$ , combines the 4-momentum of  $i$  and  $j$  into a single one and replaces that with the individual  $i$  and  $j$  vectors in the collection of the inputs, returning to step 1.
3. If  $d_{iB} < d_{ij}$ ,  $i$  is removed from the collection of the inputs for the entity  $j$ , and it is declared as a jet itself. The algorithm returns to step 1.
4. If there is no particle left, the algorithm stops.

Three common sequential recombination algorithms can be defined according to the value of the  $p$  parameter. For Cambridge-Aachen algorithm  $p = 0$ , for  $k_T$  algorithm  $p = 1$ , and for anti- $k_T$  algorithm  $p = -1$ . Considering the impact of this value to the  $p_T$ , the resulting jet shapes and characteristics are different. While the  $k_T$  algorithm clusters the softest inputs first, anti- $k_T$  jets grow outward around the highest  $p_T$  contributions. The Cambridge-Aachen algorithm is based completely on the angular separation. While the shape of anti- $k_T$  jets are mostly circular, Cambridge-Aachen and  $k_T$  jets can have more irregular boundaries as shown in Figure 5.5.

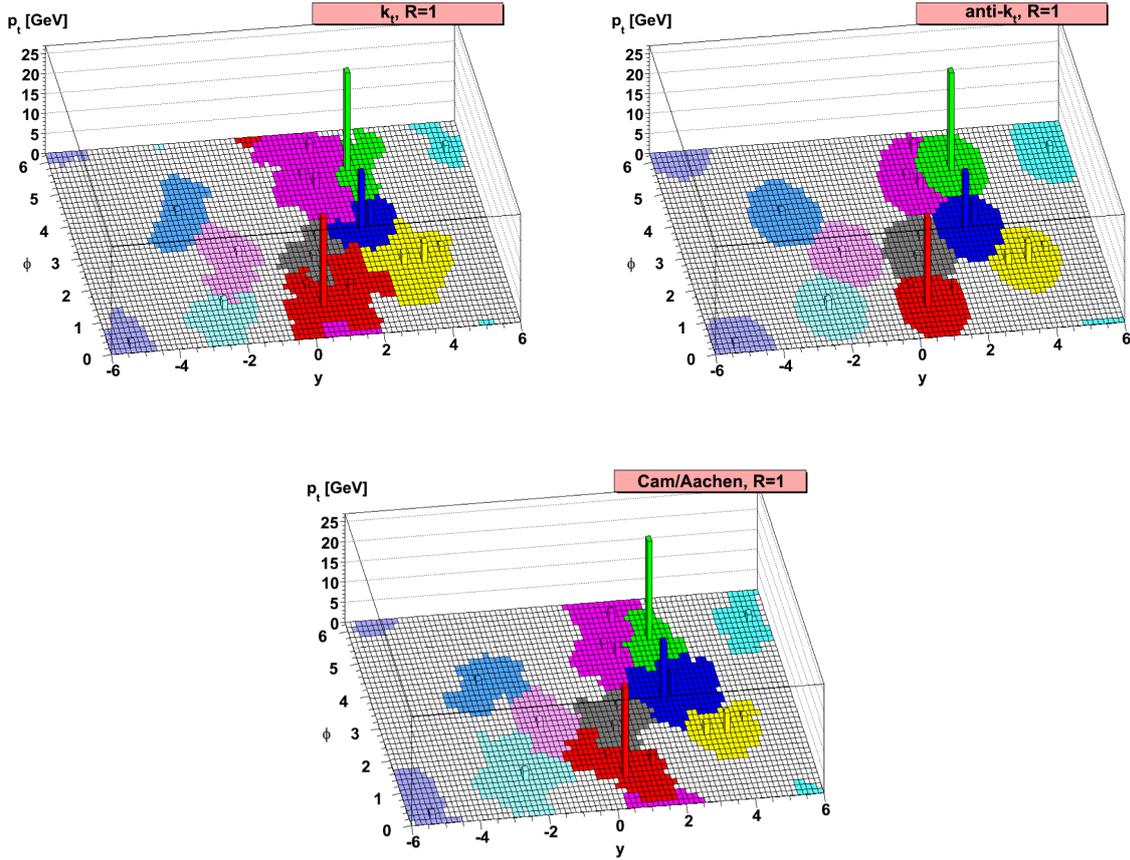


Figure 5.4: An illustrative figure showing the active catchment areas of the resulting hard jets clustered by the three jets algorithms:  $k_T$  (top left), anti- $k_T$  (top right), and Cambridge-Aachen (bottom) algorithm [98].

### Variable- $R$ jets

Another jet finding algorithm used in ATLAS detector is the variable- $R$  (VR) algorithm, where the radius of the jet is changing according to its  $p_T$  during the clustering procedure [101]. The defined distances in Equation 5.1 are rewritten as:

$$\begin{aligned} d_{ij} &= \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \Delta R_{ij}^2, \\ d_{iB} &= p_{T,i}^{2p} R_{eff}^2, \end{aligned} \quad (5.2)$$

where the  $R_0$  parameter is replaced with a  $p_T$  dependent function  $R_{eff}$ , called effective jet size:

$$R_{eff}(p_{T,i}) = \frac{\rho}{p_{T,i}}. \quad (5.3)$$

The parameter  $\rho$  simply defines how fast the effective jet size shrinks with the increased  $p_T$  of the jet. Besides  $\rho$ , algorithm requires two more parameters as  $R_{min}$  and  $R_{max}$ , lower and upper boundaries for the  $R_{eff}$ .

The variable- $R$  algorithm can be used in combination with all the other algorithms

mentioned in Section 5.1.1 using the different values of  $p$  parameter in Equation 5.2. However, it is most widely used in combination with  $p = -1$ , the distance defined for the anti- $k_T$  algorithm.

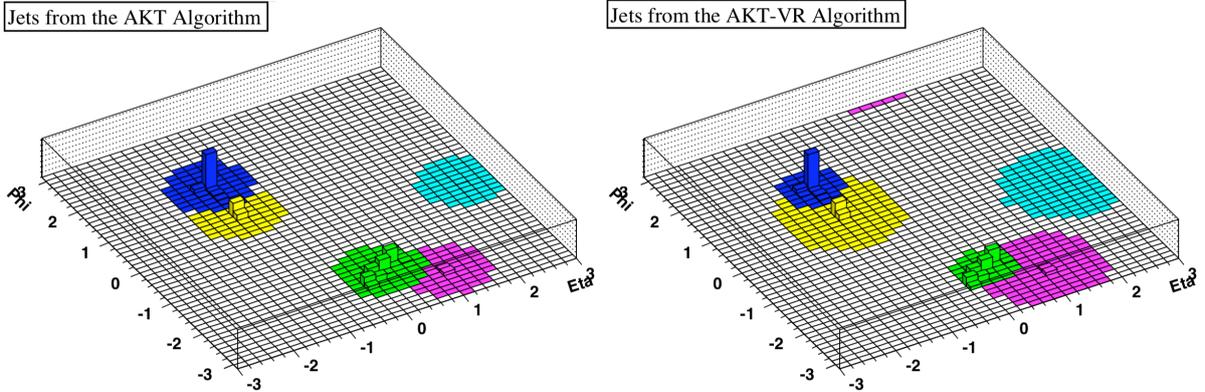


Figure 5.5: An illustrative figure showing the active catchment areas of the resulting hard jets clustered by the two jets algorithms: anti- $k_T$  (left), and variable- $R$  in combination with anti- $k_T$  algorithm (right). Note that in VR algorithm, the high- $p_T$  jets (dark blue, green) have been reduced in size while softer jets (yellow, purple, light blue) have grown [102].

### 5.1.2 Jet grooming

Jet finding algorithms are applied considering all the constituents seen in the event to identify the parton behind the observed hadron shower. However, often not only the real decay products coming from the parton, but also additional constituents like pile-up or underlying event are clustered inside the jet as depicted in Figure 5.6. These contributions are mostly softer than the hard scattering components of the jet. In order to obtain more precise jet reconstruction, they must be removed from the jet using grooming techniques. In this regard, grooming can be considered a cleaning procedure for jets via subtraction of the undesirable contributions, that do not originate from the hard scattering, from the reconstructed jet. Below, some of the most common grooming techniques are presented. Each grooming method has its own advantages and disadvantages depending on the topology of interest. Note that the default grooming method in ATLAS Run 2 new physics searches is trimming. Nonetheless, two other techniques are also briefly explained to provide a complementary insight to the topic and to be able to follow the studies performed for Higgs identification in Section 6.3.

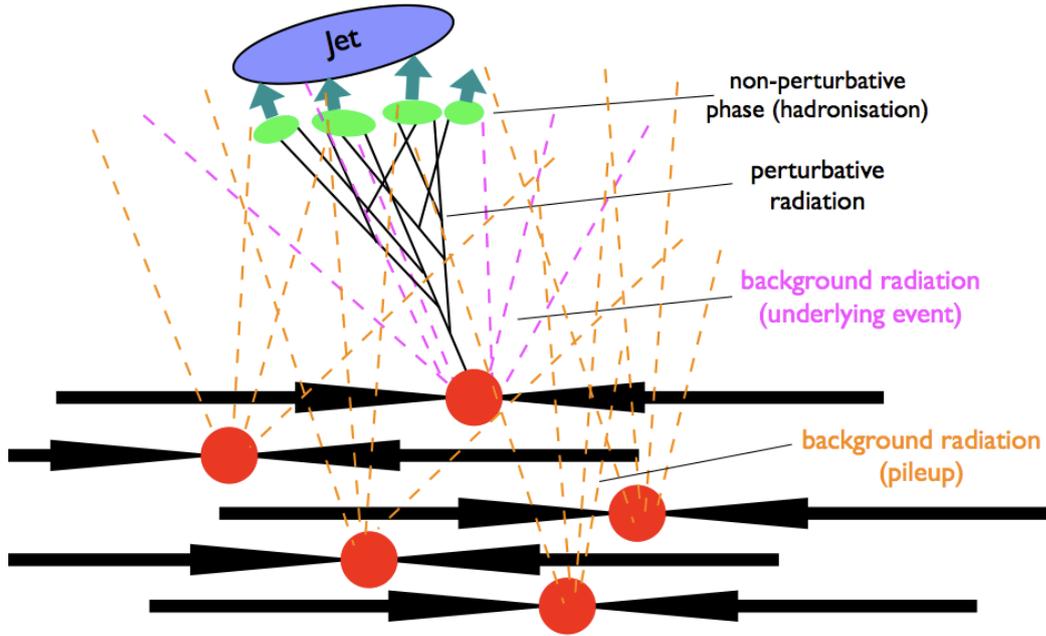


Figure 5.6: Sketch of an ungroomed jet showing the undesirable background radiation captured in the reconstructed jet area [103].

### Trimming

The idea of trimming [104] is based on the fact that the  $p_T$  of the decay products associated with the hard scattering and the subsequent final state radiations are most likely higher than the undesirable contaminations. Therefore, as a first step, the algorithm reclusters new subjets with smaller radius ( $R_{sub}$ ) than the large- $R$  jet size. The subjet finding algorithm can be different than the jet finding algorithm. However, the most common choice is the  $k_T$  algorithm to create subjets of size  $R_{sub}$ , since the soft components originated from background radiations can be well captured with the  $k_T$  algorithm. Then a requirement on the minimum momentum fraction  $f_{cut} > p_T^i/p_T^{jet}$  is applied on each subjet with index  $i$ . If the momentum fraction of the subjet is below the  $f_{cut}$  value, it is removed from the original jet constituents. The optimal values of the two specific parameters for this algorithm,  $R_{sub}$  and  $f_{cut}$ , are determined depending on the case. However, the expected values of  $f_{cut}$  is around the percentage levels.

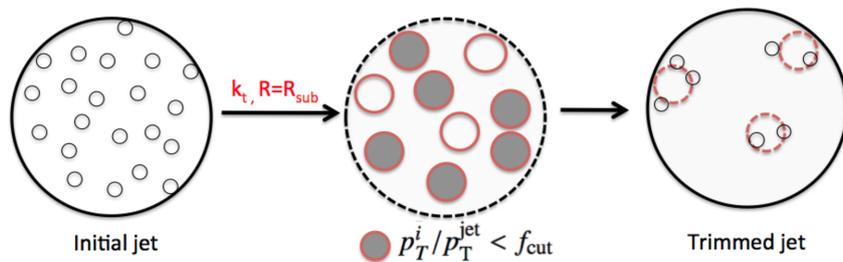


Figure 5.7: Sketch illustrating the trimming procedure as described in the text [105]. Constituents that are matched with the grey shaded subjets are groomed as a result of the trimming process.

## Pruning

Similar to trimming, the pruning algorithm discards the constituents with relatively small  $p_T$ , but in addition to that it applies additional rejections to wide-angle radiations  $\Delta R_{j_1, j_2} < R_{cut}$  [106]. The procedure also differs in terms of the execution of the algorithm. Instead of separate subjet reconstruction, it executes with the jet finding algorithm successively. In each step of the jet reconstruction, constituents are evaluated according to the given conditions and either accepted or removed from the jet constituents. Therefore, momentum fraction cut is applied as  $z_{cut} < p_T^i / p_T^{j_1 + j_2}$ , where  $j_1$  and  $j_2$  represent the two components that are in consideration for the jet finding algorithm in each step. This condition makes pruning a bit more tolerant to the soft components than the trimming.

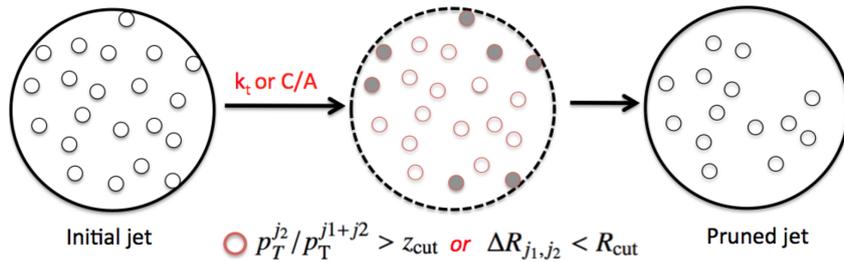


Figure 5.8: Sketch illustrating the pruning procedure as described in the text [105]. Grey shaded constituents are groomed as a result of the pruning process.

## Split (mass drop)-filtering

This method has two stages where the initial step is to split the initial jet, also known as *mass drop tagger* [107], and the following stage is to discard the constituents similarly to the trimming procedure. The method is designed to be used with Cambridge-Aachen algorithm and to reconstruct  $h \rightarrow b\bar{b}$  processes. In splitting, the algorithm tries to identify two symmetrical subjets whose mass values are considerably smaller than the initial jet mass ( $M^{jet}$ ) to ensure the requirement of having symmetrical subjets ( $j_1$  and  $j_2$ ). Assuming that the masses of the  $j_1$  and  $j_2$  are ordered such that  $j_1$  has larger mass than  $j_2$ , the ratio of  $m^{j_1} / M^{jet}$  should be smaller than the specified cut value for the mass sharing  $\mu_{frac}$ . Similarly to the mass symmetry, energy balance is also required with another cut value ( $y_{cut}$ ):

$$\frac{\min[(p_T^{j_1})^2, (p_T^{j_2})^2]}{M^{jet}} \times \Delta R_{j_1 j_2}^2 < y_{cut}. \quad (5.4)$$

Satisfying those criteria, the jet splits into two subjets, otherwise it is discarded. Then as a second step, the jet is reclustered again with Cambridge-Aachen algorithm using the constituents of  $j_1$ ,  $j_2$ , and the radius parameter  $R_{filt} = \min[R_{sub}, \frac{\Delta R_{j_1, j_2}}{2}]$ , where  $R_{sub}$  is a free parameter to be used as an upper boundary for the  $R_{filt}$ . At the end of the process, three hardest subjets are chosen and kept. The idea of having the third subjet in the two body decay process is to be able to capture possible final state radiation. The procedure shown in Figure 5.9.

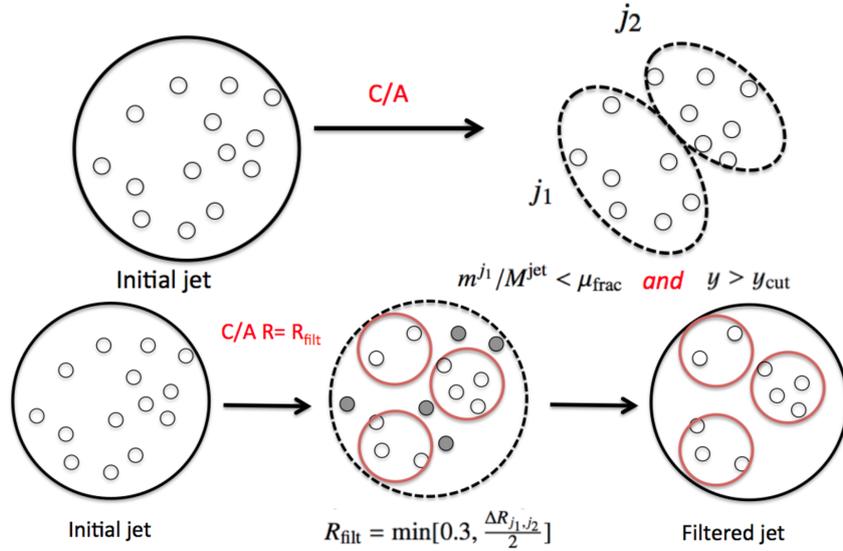


Figure 5.9: Diagram depicting the two stages of the mass-drop filtering procedure: (a) the mass-drop and symmetric splitting criteria, (b) filtering with  $R_{\text{sub}} = 0.3$  [105]. Grey shaded constituents are groomed as a result of the split-filtering process.

### 5.1.3 Small- $R$ jets

Although the standard small- $R$  jets are not used as the final state objects in this thesis, it is useful to know their features in order to follow the discussions regarding the choice of boosted topology. Besides, they are used in the preselection stages of the presented analyses. The most commonly used jets in ATLAS Run 2 new physics searches are jets reconstructed from topological clusters (topo-clusters) [108] using the anti- $k_T$  algorithm with radius parameter of  $R = 0.4$ . These topo-clusters are calibrated at the electromagnetic scale (EM) and passed through several corrections regarding the jet reconstruction as described in Reference [109]. After the jet reconstruction, additional in-situ calibrations are applied to the jets. All these stages are briefly summarised and shown in Figure 5.10.

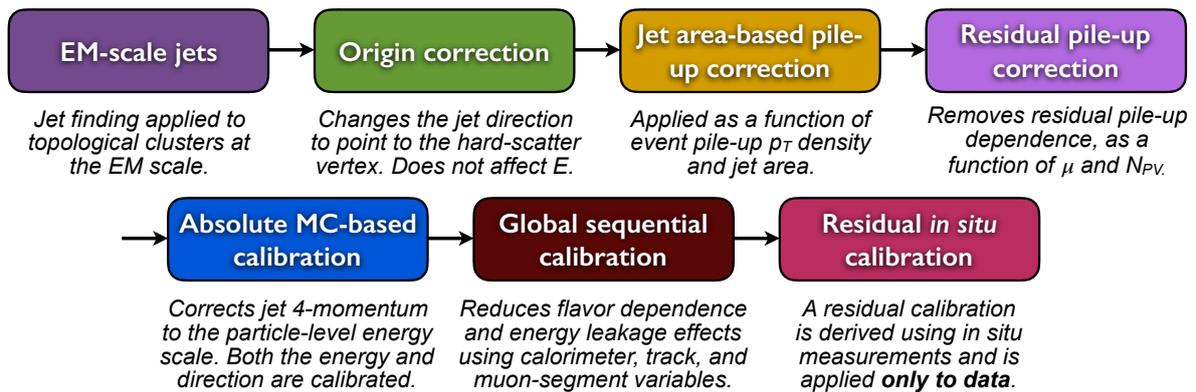


Figure 5.10: An overview of the ATLAS jet energy scale calibration chain [109]. All calibrations except the origin correction are applied to the four-momentum of the reconstructed jet.

Finally, using a multivariate algorithm, so-called *jet vertex tagger* (JVT) [110], jets arising from pile-up are discarded with a 92% efficiency for hard-scatter jets. Using the information about the primary vertex (PV), jet  $p_T$ , and track  $p_T$ , JVT constructs a two

dimensional likelihood to determine whether the tracks originate from the PV or not. An ordinary collision vertex is reconstructed from at least two reconstructed tracks with  $p_T > 0.5$  GeV, and the PV is defined as the one with the highest  $\sum p_T^2$  value.

This particular jet reconstruction described above is considered as the standard jet reconstruction method, and will be referred to as small- $R$  jets or standard jets in this thesis.

### 5.1.4 Large- $R$ jets

The ability to identify two close-by hadrons as two separate jets depends on the angular resolution of the detector and on the cone jet size. When the particles have a high  $p_T$ , their decay products become more and more collimated and it is more difficult to distinguish them from each other. For instance, electroweak scale particles are clustered as a single small- $R$  jet as soon as their  $p_T$  exceeds a few hundred GeV. Following Reference [97], for a quasi-collinear splitting into two decay products of a boosted particle, the Equation 5.5 can be written where  $m$  is the total mass of the boosted object,  $z$  and  $(1 - z)$  are the momentum fractions for each subjet and  $\Delta R$  is the distance between them:

$$m^2 \simeq z(1 - z)p_T^2 \Delta R^2. \quad (5.5)$$

Knowing that the electroweak bosons decay with a fairly uniform distribution in  $z$ , and especially the decay of the Higgs boson to  $b\bar{b}$  is almost flat [111], one can write the Equation 5.6 as the opening angle of the boosted object:

$$\Delta R \simeq \frac{2m}{p_T}. \quad (5.6)$$

In order to identify such boosted particles in ATLAS, jets are reconstructed with larger radius parameter than the radius of standard jets and called *large- $R$  jets*. Considering Equation 5.6, one can see that  $\Delta R$  depends on the process, i.e. the type of particle to be identified. Therefore, the optimum of  $\Delta R$  value for the large- $R$  jets could in principle be chosen depending on the process that is being studied. However, the whole chain of the jet reconstruction, calibration and uncertainty estimation requires lots of effort. Hence, instead of providing dedicated jets for every analysis, the experiment provides only a few sets of jets that are calibrated and reconstructed for certain type of topologies.

In Run 2, the recommended large- $R$  jets for the boosted topologies are reconstructed from topological clusters [108] using the anti- $k_T$  algorithm with radius parameter of  $R = 1.0$  and trimmed with the parameters of  $R_{\text{sub}} = 0.2$  and  $f_{\text{cut}} = 0.05$ . The corresponding jets and the impact of grooming to the jet mass is shown in Figure 5.11 before applying any calibration and correction. As can be seen from the figure, trimming causes significant amount of reduction in the large- $R$  jet mass.

The input clusters are calibrated using the local calibration (LC) method following Reference [112]. The jet energy, and pseudorapidity are corrected with the  $p_T$  and  $\eta$  dependent calibrations derived from the simulations using dijet balance measurements as described in given reference [113]. The jet mass calibration procedure is explained in the given paper [114], however more details on this topic are presented later in Section 6.2.1. Whole procedure of large- $R$  jet reconstruction and calibration is summarised in Figure 5.12.

In this thesis, all the presented search results in Chapter 7 and 8 use the recom-

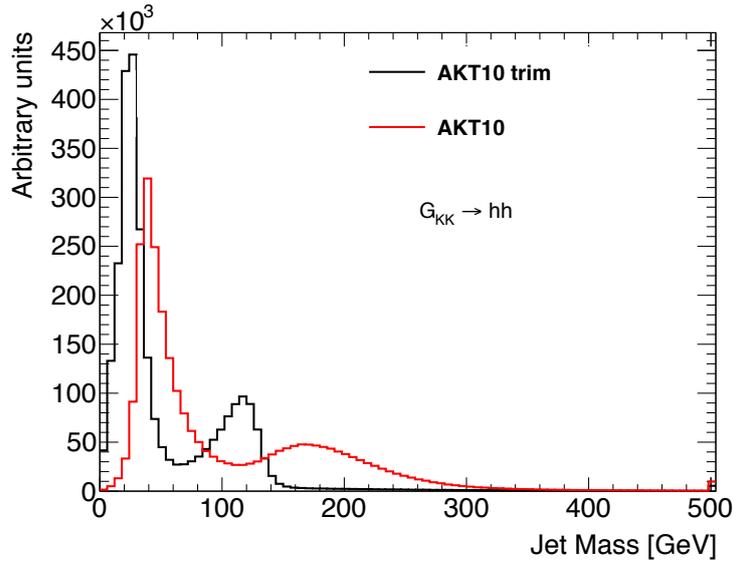


Figure 5.11: The impact of trimming on the jet mass using the simulated signal samples for di-Higgs production from  $G_{KK}$  decay. While red line (AKT10) represents the mass distribution of ungroomed anti- $k_T$  jets reconstructed with  $R = 1.0$ , black line shows the jet mass after trimming with  $R_{\text{sub}} = 0.2$  and  $f_{\text{cut}} = 0.05$ . Note that there is no calibration or correction applied on the shown jet mass distribution.

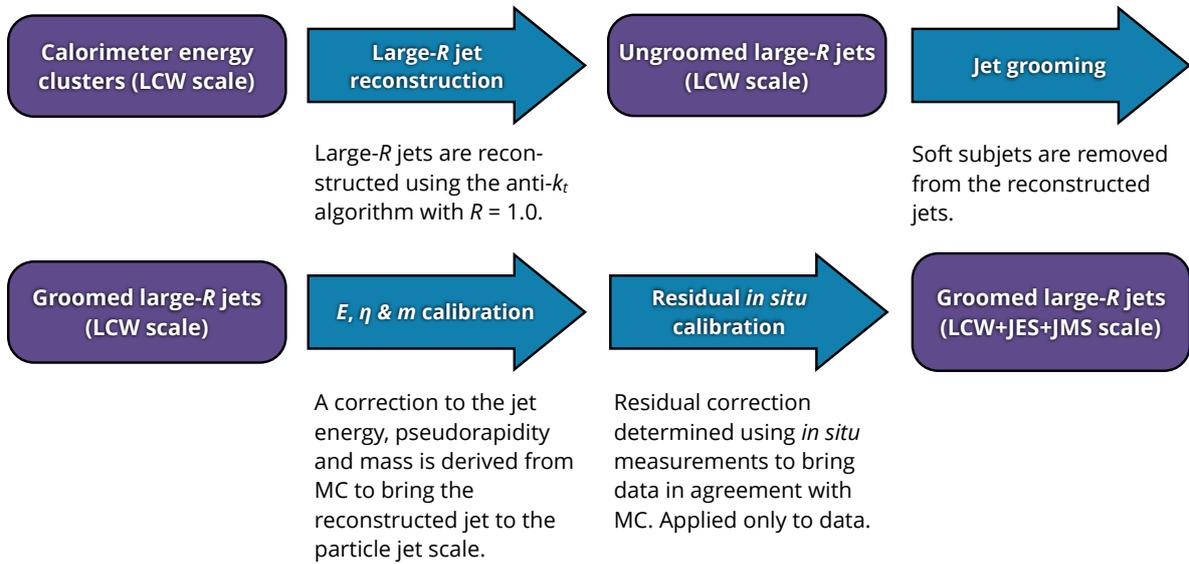


Figure 5.12: Overview of the large- $R$  jet reconstruction and calibration procedure is shown [113].

mended large- $R$  jets. The results shown for the boosted Higgs boson identification in Chapter 6 are mostly obtained with these jets. However, dedicated studies on alternative jet reconstructions to identify boosted  $h \rightarrow b\bar{b}$  process are performed, and they are presented later in Section 6.3. Unless specified otherwise, the above recommended large- $R$  jet reconstruction is used throughout the thesis and referred to as large- $R$  jets.

### 5.1.5 Track jets

Similarly to the calorimeter jets introduced above, track jets can also be reconstructed by the same jet algorithms. Different than the jets introduced so far, their inputs are the tracks provided by the ID. In ATLAS, track jets are mostly used together with the large- $R$  jets in the event to understand better the substructure of the large- $R$  jets. In those cases, the jet radius for the track jets is chosen to be much smaller than the radius of large- $R$  jets. Each reconstructed track jet is matched to the large- $R$  jets as explained later in Section 6.1.2. With this approach, collimated decay products of the boosted particle can be resolved better. As mentioned in the detector chapter, the angular resolution of the ID is higher than the calorimeter, therefore the track information can provide additional information about the inner structure of the jet.

Although there are various advantages to use track jets depending on the topology of the process, undoubtedly one of their evident benefits is to be able to identify better the flavour of the quarks behind the hadronic showers, using the information coming from tracks [115]. More detailed discussion about how the flavour identification works and why tracks are important for the process can be found in the next section 5.1.6. However, it is already clear that this ability is crucial for this thesis considering the interested final states containing the boosted  $b\bar{b}$  pair(s).

In this thesis, two different track jet collections are used. The first one is clustered with a standard anti- $k_T$  jet finding algorithm using  $R = 0.2$  as a radius parameter. That collection is referred as fixed radius size (fixed- $R$ ) track jets throughout the thesis. The other one is reconstructed with variable- $R$  jet algorithm with the parameters of  $\rho = 30$  GeV,  $R_{min} = 0.02$ ,  $R_{max} = 0.4$ , and in combination with anti- $k_T$  algorithm. Those jets are referred to as variable- $R$  or VR track jets. Both algorithms use subsets of the reconstructed tracks with minor selection differences as listed below.

- While for the fixed- $R$  track jets the track  $p_T$  requirement is  $p_T > 0.4$  GeV, for the variable- $R$  track jets it is  $p_T > 0.5$  GeV.
- For both track jet algorithms, all tracks must be in the pseudo-rapidity region  $|\eta| < 2.5$ , i.e. within the ID coverage.
- For both track jet algorithms, the track longitudinal impact parameter ( $z_0$ ) to be within  $|z_0 \cdot \sin \theta| < 3$  mm of the primary vertex (PV), implying that the accepted tracks originate from the PV or close to that point. That eliminates tracks from pile-up vertices.
- Tracks must have at least 7 hits in the SCT and pixel detector layers with the below conditions:
  - No more than one hit shared by multiple tracks in the pixel detector.
  - No more than one missing hit in the pixel detector.
  - No more than two missing hits in the SCT detector.

### 5.1.6 $b$ -jets

The identification of the jets originating from the hadronisation of  $b$ -quarks is possible due to the distinctive characteristics of  $b$ -quark decays. As the down-type quarks are relatively long lived, the lifetime of  $b$ -quarks is long enough to detect them before they

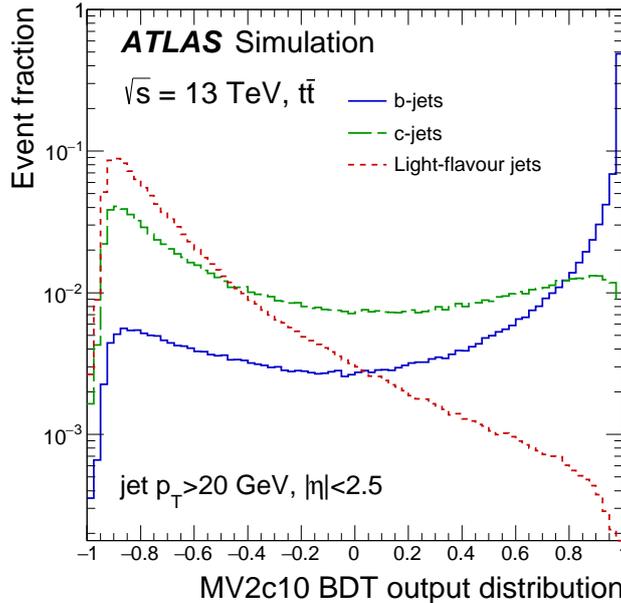


Figure 5.13: The distribution of the MV2c10 BDT output in  $t\bar{t}$  simulation is shown to illustrate the discrimination power of the algorithm between the standard  $b$ -,  $c$ -, and light jets [119].

decay into the other particles, of the order of 1.5 ps. Moreover, considering their large mass of 4.2 GeV, many orders of magnitude larger than the mass of the light quarks, they can travel several millimeters creating tracks with large impact parameters (IP) and a secondary vertex (SV). Using the features of the different shower profile, it is possible to identify them. This identification is called *b-tagging*.

Although there are several  $b$ -tagging algorithms available in the ATLAS, the presented results in this thesis are produced only using the multivariate  $b$ -tagging algorithm, that is called MV2c10 [116]. The optimisation studies using different algorithms showed that MV2c10 algorithm performs best for both of the searches presented in this thesis [74, 117]. Consequently, only a brief description of the MV2c10 algorithm is given in this section.

The MV2c10 is based on the three types of inputs which are actually the outputs of some other basic  $b$ -tagging algorithms. One of the inputs carries information about the impact parameters while the other one is about the probability of the presence of a secondary vertex. The last one investigates the topological structure of the jet and produces information based on the multi-vertices inside that [116]. The MV2c10 combines all these different information using a Boosted Decision Tree (BDT) algorithm [118] and produces a single value per jet which represents the probability of being a jet originated from  $b$ -hadron, as shown in Figure 5.13.

In order to apply MV2c10  $b$ -tagging for track jets, an additional step is required after the usual track jet reconstruction. With a looser track impact parameter constraint, an enhanced region is determined around the track jet axis to run the tagging algorithm. The tracks in this enhanced area are determined with an angular matching to the track jet axis depending on the jet  $p_T$ , wider for low  $p_T$  jets and narrower for high  $p_T$  jets. Using this larger set of tracks allows to catch better the SVs from the  $b$ -hadron decays [115].

There are certain threshold values to be able to declare a jet as  $b$ -jet. For instance, if the output of MV2c10 algorithm for a given jet exceeds a *weight cut*, which actually represents different selection efficiencies, that jet is identified as  $b$ -jet with the correspond-

ing *working point* (WP). Those values are determined using truth studies and changes with respect to the jet reconstruction. Since both variable- $R$  and fixed- $R$  track jets are the subjects of  $b$ -tagging with the MV2c10 algorithm in this thesis, the benchmarks for different WPs of MV2c10 algorithm are given in Table 5.1.

Track Jet	Release	WP	Weight Cut	$\epsilon_b\%$	$R_c$	$R_\tau$	$R_{light}$
Fixed- $R$	Rel 20	60	0.85	59.99	16.16	53.47	276.16
		70	0.65	70.00	7.09	17.25	119.69
		77	0.37	77.00	4.21	8.09	57.90
		85	-0.14	84.99	2.47	3.75	18.97
	Rel 21	60	0.86	59.94	15	110	480
		70	0.66	69.82	6	31	171
		77	0.38	76.75	4	13	73
		85	-0.15	84.72	2	5	21
Variable- $R$	Rel 21	60	0.92	59.99	23	207	1145
		70	0.79	69.97	9	55	304
		77	0.58	76.95	5	20	112
		85	0.05	84.92	2	7	28

Table 5.1: Track jet  $b$ -tagging working points (WP) are shown with the corresponding weight cuts,  $b$ -jet selection efficiencies ( $\epsilon_b$ ), and rejection factors ( $R$ ) based on  $t\bar{t}$  studies performed in ATLAS software in release 20 and release 21 [120].

## 5.2 Muons

Four muon reconstruction algorithms are used in ATLAS: Combined (CB), segment-tagged (ST), calorimeter-tagged (CT), and extrapolated (ME) [121]. Mostly CB muons are used in this thesis, therefore the underlying algorithm for this type of muons is explained, while the other algorithms are briefly introduced.

CB muons are reconstructed from the information coming from the two detector parts: ID and MS. ID and MS both have their own tracking algorithms and provide their tracks independently. After these tracks are reconstructed, a comparison between the two sets of tracks is performed. From the successful matches, combined tracks are produced. By using those tracks, the combined muon candidates can be reconstructed with highest muon purity among the other type of muons. Candidate muons are required to satisfy certain kinematic thresholds such as  $p_T$ ,  $|\eta|$  cuts and fulfill the track quality requirements. Apart from that, they have to satisfy also  $p_T$  dependent isolation requirements and the energy isolation criteria as given in Reference [121].

The ST and CT algorithms use a track in the ID as a muon candidate to start with. Then, while ST tries to match that track with MS, the CT looks for an energy deposit in the calorimeter associated to that track. And finally, the ME muons are purely based on the MS to extend the acceptance for muon reconstruction into the region  $2.5 < |\eta| < 2.7$ , which is not covered by the ID.

There are different muon identification working points according to their ability to fulfill certain *quality requirements*: Loose, Medium, Tight, and High- $p_T$  muons. Each category is inclusive, meaning that a looser category contains also the tighter category.

Throughout this thesis, unless specified otherwise, the default muon identification is medium with the requirements of  $p_T > 10$  GeV and  $|\eta| < 2.5$ . However, the loose identification is also used to suppress the non-prompt muons produced by the hadronic decays of  $b$ -mesons inside jets. Therefore, only loose and medium identifications are defined below. For more information, the given reference can be followed [121].

### Medium muons

This is the default selection for muons in ATLAS minimising the systematic uncertainties coming from reconstruction and calibration. Only CB and ME tracks are used. CB tracks are required to have  $\geq 3$  hits on at least two layers of MDT, except for the  $|\eta| < 0.1$  region where tracks with at least three hits in one single MDT layer are allowed. ME tracks are required to have at least three MDT/CSC hit layers in the region of  $2.5 < |\eta| < 2.7$  as motivated above. The difference between the ID and MS  $1/p$  measurements is required to be below  $7\sigma$  standard deviations.

### Loose muons

Loose muons selection provides high reconstruction efficiency and good-quality muon tracks. Their optimisation study is performed to reconstruct Higgs boson candidates in the four-lepton final state. All the muon types are used with certain restrictions for this identification. However in the central region,  $|\eta| < 2.5$ , about 97.5% of the loose muons are CB muons, approximately 1.5% are CT, and the remaining 1% are reconstructed as ST muons.



# Chapter 6

## Boosted $h \rightarrow b\bar{b}$ identification

As briefly mentioned in Chapter 1, searches performed in boosted topologies became more important with the increased centre-of-mass energy at LHC. Many BSM models predicting new heavy particles which were not easily accessible before LHC, are currently under investigation with the Run 2 dataset collected at  $\sqrt{s} = 13$  TeV. Moreover, new data that will be collected in Run 3 will contain collisions at even higher energies with  $\sqrt{s} = 14$  TeV. Therefore, improving the boosted analysis techniques carries great significance for LHC experiments to be able to handle these newly accessed kinematic regimes well. However, not only the new heavy particle searches get benefit from those techniques. Any analysis which predicts a high- $p_T$  massive particle within its topology can and should evaluate the benefits of using boosted techniques. As it will be seen later in the search presented in Chapter 8, to explore a larger kinematic phase space does not always require higher energies but improved techniques.

After the discovery of the 125 GeV Higgs boson [14, 15], both the measurements for SM Higgs boson properties and the searches with SM Higgs bosons moved into focus of the ATLAS and CMS physics programmes. With the highest branching fraction among all the Higgs decay modes (see Figure 6.2), the  $b\bar{b}$  channel is one of the natural choices to look for Higgs production. Despite this advantageous side of this hadronic final state, there is also a significant drawback. The enormous background coming from the QCD interactions (see Figure 6.1) is a challenge and more sophisticated techniques are required to deal with it. Therefore, the first observation of the  $h \rightarrow b\bar{b}$  process in association with a  $W$  or  $Z$  boson with a significance of more than  $5\sigma$  standard deviation could only be achieved recently, using the large Run 2 datasets [122, 123].

In the light of these statements, improving the techniques for the identification of the  $h \rightarrow b\bar{b}$  process in boosted topologies is crucial and interesting. This chapter is dedicated to reveal the characteristics of the boosted  $h \rightarrow b\bar{b}$  process and discusses the identification techniques used in the searches presented in Chapter 7 and 8. Beyond this, some results which cast additional light on the subject and provide references for the future studies are shown. The chapter starts with the description of common Run 2 ATLAS boosted  $h \rightarrow b\bar{b}$  tagging strategy. Then, one of the most crucial observables for boosted object tagging, the concept of *jet mass* is introduced and discussed within the context of the Higgs boson. Finally, alternative jet reconstruction and grooming algorithms for Higgs identification are presented.

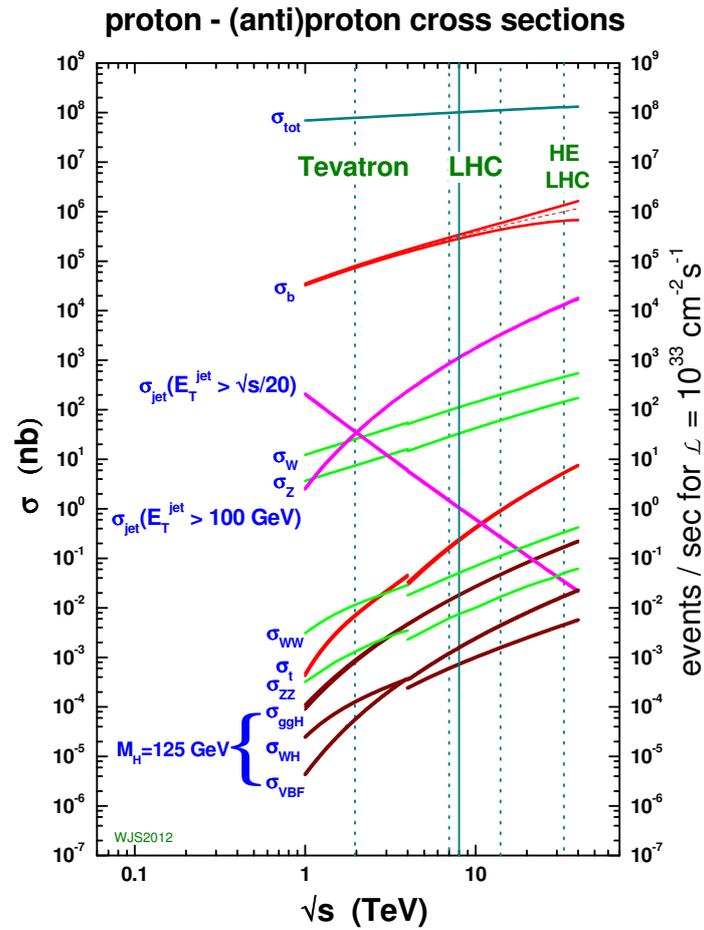


Figure 6.1: The discontinuity in some of the cross sections at 4 TeV is due to the switch from proton-antiproton to proton-proton collisions at that energy [124].

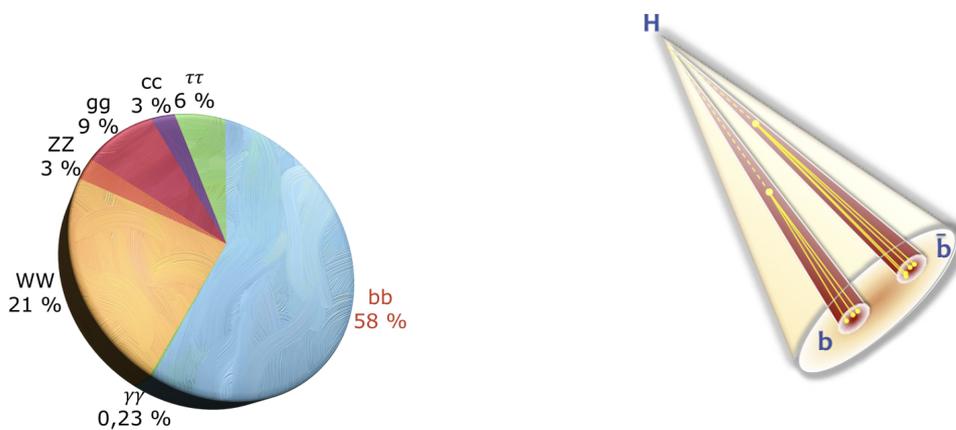


Figure 6.2: (a) Decay channels of the Higgs boson shown together with the branching fractions. (b) Sketch of a boosted Higgs jet, the large- $R$  jet is shown in orange, track jets are shown in red, and yellow lines represent the tracks.

## 6.1 Boosted $h \rightarrow b\bar{b}$ tagger

Similar to the identification of boosted  $W/Z$  bosons and boosted top quarks, the process of discriminating boosted Higgs bosons from the other boosted particles and jets initiated by QCD processes is known as *Higgs tagging*. In this section, the ATLAS Run 2 boosted  $h \rightarrow b\bar{b}$  tagger strategy is described. The tagging prescription is provided as a point of reference for the analysers, and they have been commonly used in many Run 2 analyses with boosted Higgs to  $b\bar{b}$  process in it.

Most of the optimisation and performance studies presented in this chapter use truth information to label jets obtained from the simulated  $G_{KK}$ ,  $t\bar{t}$ , and dijet samples as introduced in Section 4.2. In this context, a *Higgs jet* is defined as a reconstructed large- $R$  jet that is matched to at least two  $b$ -hadrons with  $p_T > 5\text{GeV}$ , and at least one truth Higgs boson. Similarly, jets originating from  $t\bar{t}$  and QCD processes are labelled using truth information. If a large- $R$  jet is matched with a truth top quark, it is labelled as *Top jet*. Finally, all large- $R$  jets that do not match with any truth boson or top quark are labelled as *QCD jet*. The matching is performed with a method, so-called *ghost association* which is introduced later in Section 6.1.2. Below, the baseline Higgs tagging strategy can be summarised in the following three steps:

- Large- $R$  jet selection for the Higgs candidate;
- Selection of  $b$ -tagged track jets associated to the Higgs candidate;
- Selection on the properties of the Higgs candidate jet using jet substructure/jet mass observable.

In this section, each item given above is summarised briefly to provide an overview of the tagger. Some of the crucial points are left to the subsequent sections to be discussed later in detail.

### 6.1.1 Large- $R$ jet selection

Large- $R$  jets are appropriate to identify Higgs bosons in boosted topologies, as discussed in Section 5.1.4. Recalling the boost condition from Equation 5.6 and inserting the Higgs boson mass,  $\Delta R$  between the decay products of the Higgs boson follows the relation:  $\Delta R \simeq \frac{2 \times 125 \text{ GeV}}{p_T}$ . Considering that the jet radius is 1 for the default large- $R$  jets, one can see that the  $p_T$  of a boosted Higgs jet candidate must be above 250 GeV in order to be fully reconstructed and captured in the jet. Below this value, the separation between the decay products, the  $b$ -hadrons, is expected to be larger than the total jet size and typically they will not be captured in the jet area. Therefore, the lowest  $p_T$  requirement for large- $R$  jets, 250 GeV, is naturally defined due to the kinematics of the topology.

On the other hand, the situation differs at higher  $p_T$  values of large- $R$  jets. For instance, one can see that the decay products of large- $R$  jets start to merge at higher  $p_T$  values as given in Figure 6.3 for fixed- $R$  track jets with  $R = 0.2$  used to reconstruct  $b$ -hadrons. Therefore, it is important to note that the internal structure of large- $R$  jets can be quite different depending on the  $p_T$  regimes.

In boosted topologies especially in hadronic final states, having a large number of QCD jets as background is unavoidable (Figure 6.1). Since QCD jets typically have smaller mass values than the boosted particles, applying a cut on the jet mass is a

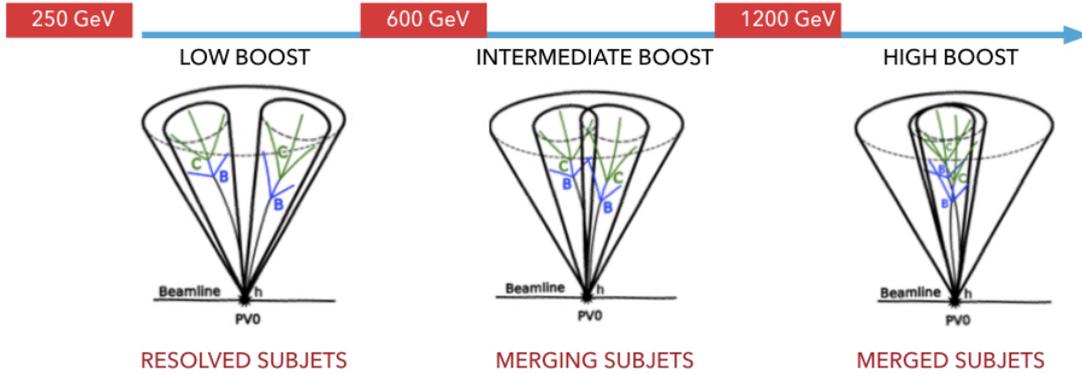


Figure 6.3: Higgs jet topologies in boosted environments. The given boundaries for the  $p_T$  values for the large- $R$  jets are calculated for fixed- $R$  track jets with  $R = 0.2$ , figure edited from [103].

reasonable option in order to avoid a large background contamination in a desired signal region. For this reason, the large- $R$  jets with a mass below 40 GeV are removed within the Higgs tagging analysis using the *combined mass* as a *jet mass* definition, which is discussed extensively in Section 6.2.1.

### 6.1.2 Track jet selection

For a well reconstructed Higgs jet candidate, the corresponding  $b\bar{b}$  pair is expected to be captured in the large- $R$  jet area. Each  $b$ -quark hadronises and forms a  $b$ -hadron, therefore typically two  $b$ -hadrons are expected to be identified as part of the reconstructed track jets described in Section 5.1.5. Therefore, these objects carry great significance to identify  $h \rightarrow b\bar{b}$  processes.

Since the track jet reconstruction is independent from the large- $R$  jet reconstruction, reconstructed track jets need to be associated separately to the large- $R$  jet area to make sure that they belong to the same boosted process. Thus, in this section, the track jet association methods, the choice of track jet reconstruction method, and the  $b$ -tagged track jet requirements used for the Higgs identification are explained.

#### Track jet association

Although the most common and straightforward association method used to match reconstructed objects in the ATLAS experiment is *geometrical ( $\Delta R$ ) matching* that requires the objects to be in certain distances in  $\eta \times \phi$  space, the association method used for track jets and large- $R$  jets is *ghost association* which results to be a more robust method than the geometrical matching [125, 126]. The idea behind ghost association is to match a constituent to a jet using the active area of that jet. In this case, the constituents which need to be associated to the jet area are track jets and they are referred to as *ghosts* during the association process. After reconstructing separately the two objects (large- $R$  jets and track jets) to be matched, the  $p_T$  of the ghosts are set to small values (e.g. 1eV) keeping  $\eta$  and  $\phi$  information unchanged. Then, the ghosts are added to the input constituents list of the jet finding algorithm. Jet finding algorithm (anti- $k_T$ ,  $R = 1$ ) is performed again to reconstruct large- $R$  jets, this time with the additional ghosts. The addition of the ghosts does not affect the main jet reconstruction due to their negligible

$p_T$ . Consequently, ghosts can be identified as one of the constituents in the jet area of the ungroomed anti- $k_T$  jet.

### Track jet choice

In this thesis, two track jet reconstructions are used: fixed- $R$  and variable- $R$  track jets as can be recalled from Section 5.1.5. Both sets of track jets are ghost associated to large- $R$  jets. In order to match the cut values of the  $b$ -tagging studies, both categories of the track jets are selected to satisfy the following kinematic requirements:  $p_T > 10$  GeV,  $|\eta| < 2.5$ . However, variable- $R$  track jets were only ready much later than the fixed- $R$  track jets to be used in the Run 2 analyses. Thus, the baseline  $h \rightarrow b\bar{b}$  tagger studies are performed using fixed- $R$  track jets. Variable- $R$  track jets are started to be investigated as an option for the boosted Higgs searches as soon as their reconstruction and calibration procedures were finalised. While the impact of the track jet choice has to be evaluated specifically for each individual analysis, Figure 6.4 shows that the usage of variable- $R$  track jets is advantageous over fixed- $R$  track jets in higher  $p_T$  regions, providing significantly higher efficiencies for the identification of two separate track jets with  $b$ -hadrons [127].

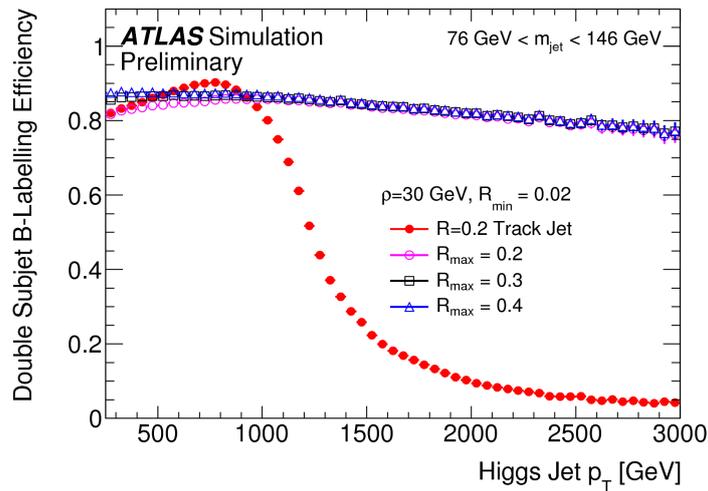


Figure 6.4: Efficiency is defined as the ratio of the number of Higgs jets with two leading track jets that are geometrically matched to the truth  $b$ -hadrons to the total number of Higgs jets in the  $G_{KK}$  samples. The efficiency is shown for VR track jets with  $\rho = 30\text{ GeV}$  and  $R_{min} = 0.02$  for varying values of  $R_{max}$ . The efficiency for  $R = 0.2$  fixed- $R$  track jets is also included. The error bars include statistical uncertainties only [127].

All results presented in this chapter use fixed- $R$  track jets. Additional comparison studies between fixed- $R$  track jets and variable- $R$  track jets are performed and shown for the analysis presented in Chapter 8.

### $b$ -tagged track jets

In the context of Higgs tagging studies, the identification of the Higgs jet from the  $b\bar{b}$  decay requires at least one or two ghost associated track jets depending on the benchmark used for the flavour tagging options. As shown in Figure 6.3, depending on the boost of the particle, two  $b$ -jets can either be resolved separately or they start to merge. Therefore, requiring two  $b$ -jets for the Higgs boson decay may not be always the best option for the

analyses. It is worthwhile requiring single  $b$ -tagged track jets to be able to consider also the cases where two  $b$ -jets are merged, especially for highly boosted regimes.

Not only the ability to resolve the track jets but also the impact of the  $b$ -tagging efficiencies must be considered for the  $b$ -tagged track jet selection. Clearly, the  $b$ -tagging WP choice has an impact on the final Higgs jet selection efficiency. Thus, to be able to consider all these scenarios, four benchmarks are evaluated in the context of the Higgs tagger. As explained in Section 5.1.6, the MV2c10 algorithm is used for  $b$ -tagging.

- *Double  $b$ -tagging*: The two highest  $p_T$  track jets must be  $b$ -tagged using the same  $b$ -tagging WP.
- *Asymmetric  $b$ -tagging*: The track jet with the largest  $b$ -tagging weight of the two highest  $p_T$  track jets should be  $b$ -tagged using a fixed  $b$ -tagging WP, while the other one is  $b$ -tagged with a varying  $b$ -tagging WP.
- *Single  $b$ -tagging*: At least one of the two highest  $p_T$  track jets is  $b$ -tagged.
- *Leading single  $b$ -tagging*: The highest  $p_T$  (leading) track jet is  $b$ -tagged.

Using the multijet and  $t\bar{t}$  MC simulations described in Chapter 4 as background, the tagger performance is evaluated for the above four benchmark scenarios designed for  $b$ -tagging track jets. To ensure good coverage of the kinematic phase space, all the simulated samples produced for different mass values of  $G_{KK}$  are used as Higgs samples. A two-step reweighting procedure is applied to obtain a physically meaningful  $p_T$  distribution among the different  $G_{KK}$  mass points. The details of the procedure can be found in Reference [92].

Figure 6.5 shows the Higgs signal selection, the rejection efficiencies for multijet,  $t\bar{t}$  backgrounds for Higgs jet  $p_T$  values above 250 GeV, and 1 TeV. The Higgs jet efficiency is defined as the ratio of the number of Higgs jets satisfying the benchmark  $b$ -tagged track jet selections to the total number of Higgs jets. The background rejection is defined as the ratio of the total number of background jets to the number of background jets satisfying the benchmark  $b$ -tagging selection. As can be seen from the figure, four benchmarks provide different advantages in terms of rejection and selection efficiencies in different  $p_T$  regimes. While the single  $b$ -tagging benchmark results in the highest Higgs jet efficiencies with low background jet rejection, double and asymmetric  $b$ -tagging provides similar good performances in comparison to the other benchmarks up to 70% Higgs efficiencies for  $p_T > 250$  GeV. One of the notable differences between the given  $p_T$  regimes is the performance drop observed in double and asymmetric  $b$ -tagging benchmarks caused by merging in boosted environments.

### 6.1.3 Jet substructure and mass

The internal structure of a jet can provide unique information on the underlying processes. Using the information coming from each individual jet constituent, many observables can be defined to gain discrimination power between signal and background jets. Therefore, it is convenient to introduce the concept of jet substructure (JSS) at this point and give a glimpse of some of the important JSS observables that can be used to discriminate Higgs jets from background jets: QCD, top or  $W/Z$  jets.

There are plenty of JSS observables exploiting the relations between the constituents using quantities like  $p_T$ , energy, or their relative position to each other. Using individual

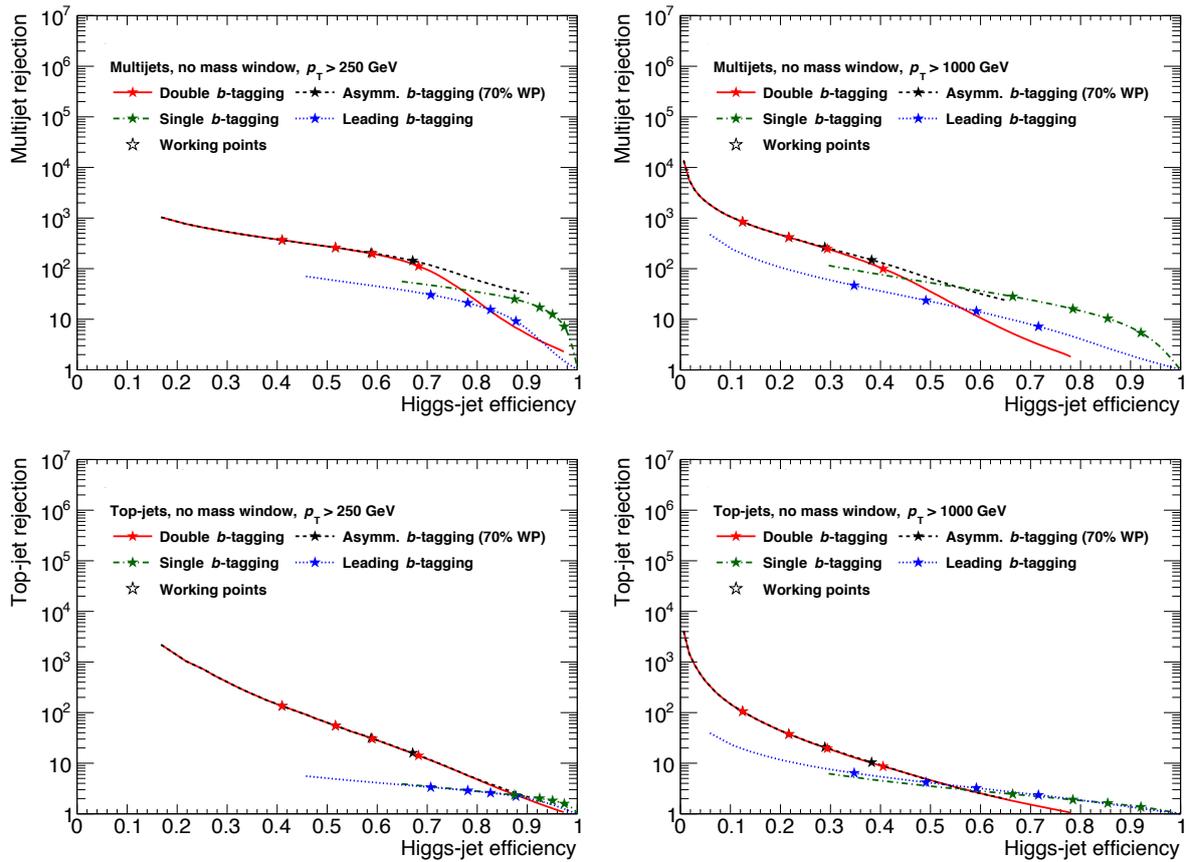


Figure 6.5: The multijet (top) and the top-jet (bottom) rejection as a function of the Higgs boson tagging efficiency for large- $R$  jet  $p_T$  above 250 GeV (left) and above 1000 GeV (right) for various  $b$ -tagging benchmarks defined in Section 6.1.2. The stars correspond to the 60%, 70%, 77%, and 85%  $b$ -tagging WPs (from left to right) [92].

information coming from each constituent, different features of the jets can be revealed as JSS observables like the jet shape, the total number of axes present in jet, the energy correlation functions indicating to the total number of subjets etc. In the context of Higgs tagger studies, the possible improvements that can come from the usage of JSS variables are studied and for most of the cases it is found that using additional JSS variables on top of the  $b$ -tagging requirements provides minor improvements for multijet and top rejection. The improvement is more pronounced in the region where the jet  $p_T$  is higher than 1 TeV [92].

However, another crucial jet observable for background discrimination, the *jet mass*, carries great significance for taggers. Since the jet mass observable gives a unique and well known information about the particles which initiate the jet shower, almost all of the taggers and searches use this observable either for increasing the background rejection or as a final discriminant to look for an excess. The observable is the resultant 4-vector of the reconstructed jet corresponding to the sum of the original hadrons comprising the jet after correcting for detector response and resolution effects. While a more detailed discussion about the observable can be seen in the following section, this section shows the impact of the usage of Higgs jet mass windows which are developed specifically for the Higgs identification.

Obviously, for the Higgs boson identification, the mass of the Higgs boson at 125 GeV is a crucial information. If the jets actually originate from the Higgs boson, the reconstructed jet mass value is expected to be close to the truth mass value of the boson. In the context of the Higgs tagger, two Higgs jet mass ranges are optimised as a function of Higgs jet  $p_T$  using the fit function given in Equation 6.1. The lower and upper boundaries of the mass windows are determined from twelve Higgs jet mass distributions in different Higgs jet  $p_T$  ranges.

$$\sigma_m(p_T) = \sqrt{(a + b/p_T)^2 + (c \cdot p_T + d)^2}. \quad (6.1)$$

Both of the mass windows are designed to have a fixed Higgs jet selection efficiency in a given  $p_T$  range. These windows are introduced below, however a more detailed discussion on the topic is left to Section 6.2. The impact of these mass selections on the tagger performance in terms of the Higgs jet selection and background rejection efficiencies are shown for single and double  $b$ -tagging benchmarks in Figure 6.6. The results obtained by using other benchmarks can be seen in Reference [92].

- Tight mass window, containing 68% of Higgs jets;
- Loose mass window, containing 80% of Higgs jets.

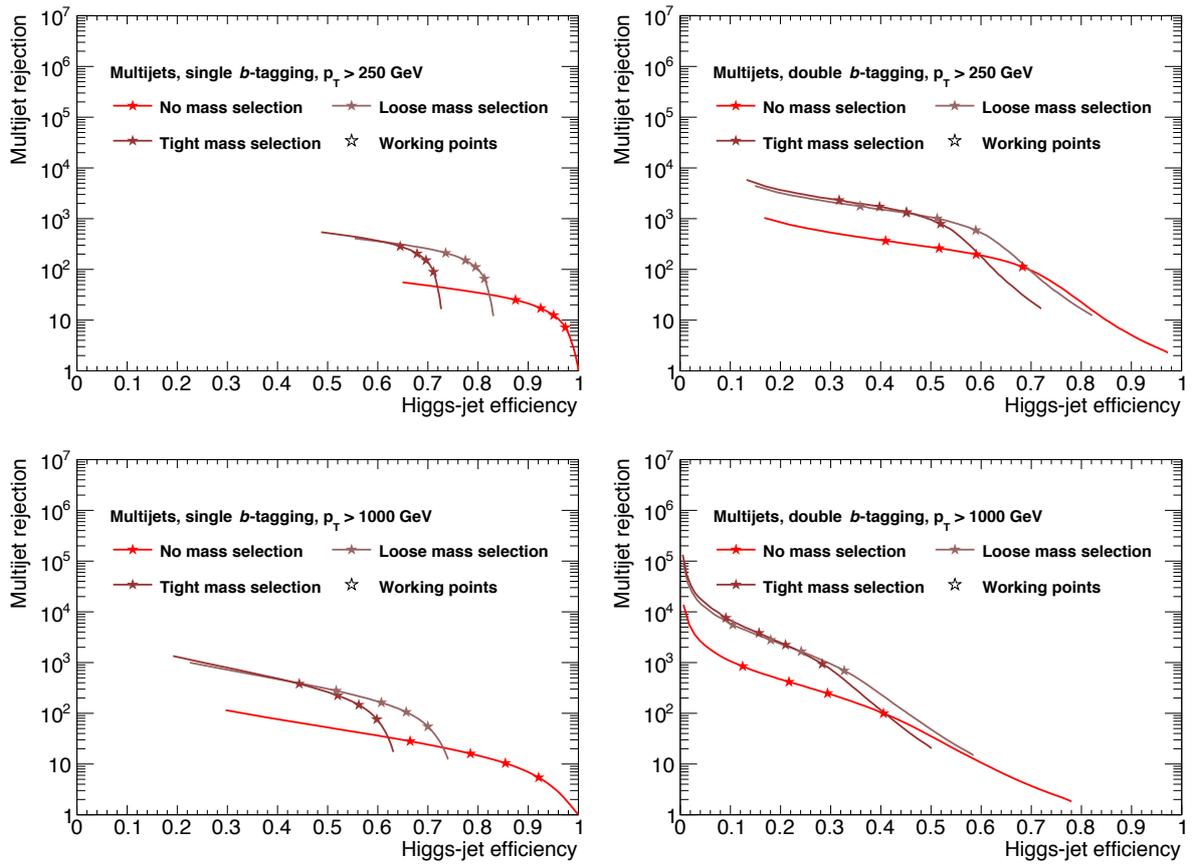


Figure 6.6: Multijet rejection as a function of Higgs jet selection efficiency for large- $R$  jet  $p_T$  above 250 GeV (top) and above 1000 GeV (bottom) for single  $b$ -tagging (left) and double  $b$ -tagging (right) benchmarks defined in Section 6.1.2 without, with loose, and tight mass windows selections. The stars correspond to the 60%, 70%, 77% and 85%  $b$ -tagging WPs (from left to right) [92].

## 6.2 Higgs jet mass

As already mentioned, the jet mass observable is invaluable to distinguish different processes from each other. The dominant background for a Higgs jet is the multijet background where jets are produced in QCD processes. Initially, QCD jets are massless but later they acquire mass through emissions during parton showering. Jet mass is an infrared and collinear safe quantity that can be calculated order by order in perturbation theory. As it is the case for most perturbative jet properties, first emission is dominant also for the jet mass and that can be calculated [128]. Due to the soft and collinear singularities of the QCD matrix element for gluon emission, the QCD jet mass distribution receives strong enhancement at low values of  $m_J$  resulting in a continuously steeply falling mass distribution. However, for the hadronic decays of heavy objects, final decay products are caught in the large- $R$  jet area building up the original mass of the heavy object [129], as shown in Figure 6.7.

Clearly, the jet mass observable plays a key role in many searches with jets. Therefore it is crucial to improve the reconstructed jet mass by enhancing its resolution and reducing any associated systematic uncertainties. In this section, the jet mass observable for Higgs boson is evaluated in various aspects. To begin with, one of the jet mass definition in ATLAS, the *combined jet mass* is discussed and the impact of the jet mass definition choice on the Higgs jet mass resolution is shown in Section 6.2.1. Then, an additional correction for the observable, so-called *muon (in jet) correction*, applied to the Higgs jet mass observable in order to consider the semileptonic  $b$ -decays of the Higgs boson is presented in Section 6.2.2. Finally, the optimisation of the  $p_T$  dependent Higgs jet mass window is presented together with the dedicated studies performed for the tails of the Higgs jet mass distribution in Section 6.2.3.

### 6.2.1 Combined mass for Higgs jets

The combined jet mass ( $m^{\text{comb}}$ ) is one of the major improvements for the field of boosted object tagging in ATLAS Run 2 [114]. While the method is developed mainly considering the  $W/Z$  tagging, the motivation behind its development also applies for Higgs tagging. The idea is born from the desire of improving the jet mass resolution in the highly boosted regimes.

The traditional jet mass definition, *calorimeter-based jet mass* ( $m^{\text{calo}}$ ), suffers from the finite angular resolution of the calorimeter at high  $p_T$  values where the decay products start to merge. As it has been mentioned earlier in Chapter 5, the large- $R$  jets are calorimeter jets reconstructed using topoclusters as input for the jet finding algorithm. For a large- $R$  calorimeter jet,  $J$ , with calo-cluster constituents  $i$  with energy  $E_i^{\text{topo}}$  and momentum  $\vec{p}_i^{\text{topo}}$ , the  $m^{\text{calo}}$  is defined as:

$$m^{\text{calo}} = \sqrt{\left(\sum_{i \in J} E_i^{\text{topo}}\right)^2 - \left(\sum_{i \in J} \vec{p}_i^{\text{topo}}\right)^2}. \quad (6.2)$$

Considering the better angular resolution of the tracking system to overcome the finite calorimeter resolution, an alternative jet mass definition is introduced, the so-called

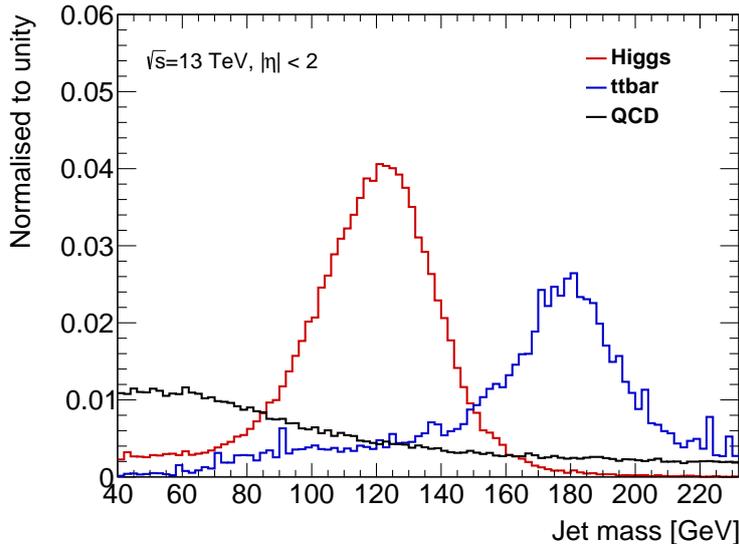


Figure 6.7: Normalised jet mass distributions for Higgs, Top, and QCD jets using the single  $b$ -tagging benchmark with the 77% WP. All distributions are obtained from simulated samples as described in Chapter 4.

*track-assisted jet mass* ( $m^{\text{TA}}$ ),

$$m^{\text{TA}} = \frac{p_T^{\text{calo}}}{p_T^{\text{track}}} \times m^{\text{track}}, \quad (6.3)$$

where  $p_T^{\text{calo}}$  is the transverse momentum of the large- $R$  jet after trimming and calibration,  $p_T^{\text{track}}$  is the transverse momentum of the sum of track four vectors associated to  $J$ ,  $m^{\text{track}}$  is the invariant mass of this four-vector sum where track masses are set to the pion mass. Since only the charged particles are reconstructed as tracks, and  $m^{\text{track}}$  alone clearly does not reflect the true jet mass value. The missed neutral components of the jet are taken care of by introducing the additional correction factor to the equation, which is the ratio of  $p_T^{\text{calo}}$  to  $p_T^{\text{track}}$ .

Using the resolutions of the jet mass responses ( $R_m$ ), where  $R_m$  is defined as the ratio between the reconstructed jet mass ( $m^{\text{reco}}$ ) to the true jet mass value ( $m^{\text{truth}}$ ), as a figure of merit to evaluate the performance of the mass definitions, the following observations are made for  $W/Z$  jets. While the resolution of  $R_m$  for  $m^{\text{TA}}$  is found significantly better than  $m^{\text{calo}}$  for the jets whose  $p_T$  are above 1 TeV, below this value  $m^{\text{calo}}$  performs better than  $m^{\text{TA}}$  [114]. Since both observables have their own strengths in specific kinematic regimes, a new observable that performs well in all  $p_T$  ranges is desired. Considering the low correlation between the  $m^{\text{calo}}$  and  $m^{\text{TA}}$  responses for  $W/Z$  jets, a simple weighted linear combination of the  $m^{\text{calo}}$  and  $m^{\text{TA}}$  observables is found to be optimal for this purpose, which is the combined mass ( $m^{\text{comb}}$ ):

$$m^{\text{comb}} = a \times m^{\text{calo}} + b \times m^{\text{TA}}, \quad (6.4)$$

The coefficients  $a$  and  $b$  are the weights for  $m^{\text{calo}}$  and  $m^{\text{TA}}$ , respectively, and they are defined as below using the jet mass resolutions ( $\sigma$ ) obtained from the corresponding jet

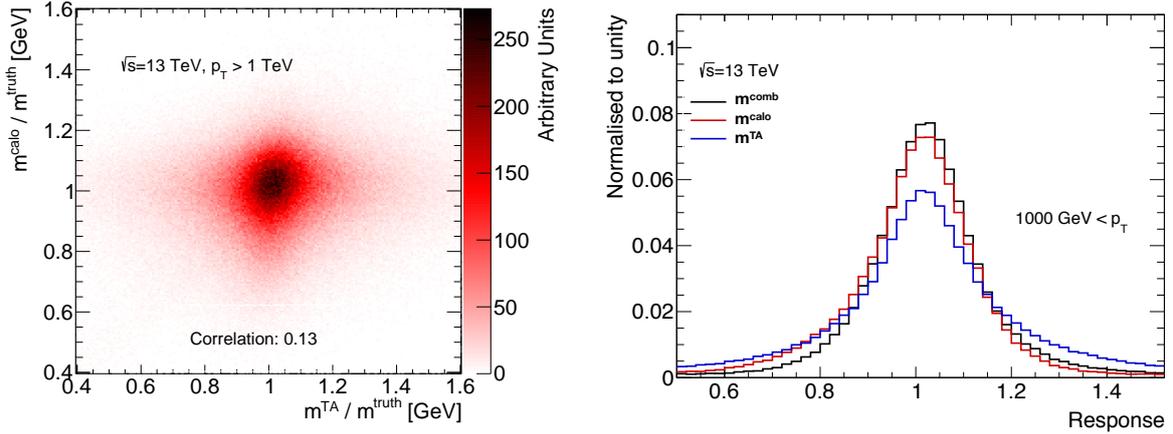


Figure 6.8: The correlation between the  $m^{\text{calo}}$  and  $m^{\text{TA}}$  observables for the mass of the large- $R$  jet from  $h \rightarrow b\bar{b}$  decays. The linear correlation is 0.13 indicates that additional gains in precision can be achieved by combining  $m^{\text{calo}}$  and  $m^{\text{TA}}$ .

mass responses  $R_m$ .

$$a = \frac{(\sigma_{\text{calo}})^{-2}}{(\sigma_{\text{calo}})^{-2} + (\sigma_{\text{TA}})^{-2}},$$

$$b = \frac{(\sigma_{\text{TA}})^{-2}}{(\sigma_{\text{calo}})^{-2} + (\sigma_{\text{TA}})^{-2}}.$$

In this section, the combined jet mass definition is implemented and adapted for the Higgs jets and its impact on the Higgs jet mass is shown. The selection of the Higgs jets is done by ghost association of the true Higgs bosons to the large- $R$  jet area before trimming using the  $h \rightarrow b\bar{b}$  decays in  $G_{KK} \rightarrow HH$  events from the MC simulations introduced in Section 4.2.2. In order to understand the applicability of the  $m^{\text{comb}}$  observable for Higgs jets, the correlation between the  $m^{\text{calo}}$  and  $m^{\text{TA}}$  responses for the Higgs jets is checked. The mass response per jet is obtained by applying a geometrical matching between the reconstructed jets and the truth jets in the simulation.  $R_m$  is a direct comparison between the reconstructed jet mass and the true jet mass value reflecting how good the reconstructed jet mass observable is. Obviously, for a nicely reconstructed jet mass the response is expected to be close to 1. Figure 6.8 shows the  $m^{\text{calo}}$  and  $m^{\text{TA}}$  responses for Higgs jets above 1 TeV revealing the low correlation between the two observables. This indicates that additional gains can be achieved by combining the two observables also for Higgs jets, therefore the  $m^{\text{comb}}$  observable is introduced for Higgs jets.

The resolution of the  $R_m$  distributions can be considered as a measure of the *goodness* of the mass reconstruction method. The lower the resolution, the better the jet mass reconstruction is. Since the shapes of the  $R_m$  distributions are irregular with their distinct and asymmetric shapes of the tails (Figure 6.8), the interquantile range (IQnR) method is chosen and used as a figure of merit to evaluate the resolution [114]. The resolution in  $R_m$  is defined as the half of the 68% interquantile range divided by the median value of the  $R_m$ . The 68% interquantile range corresponds to the difference between the 16th and 84th percentiles of a given distribution. This definition of response coincides with the standard deviation for an ideal Gaussian distribution.

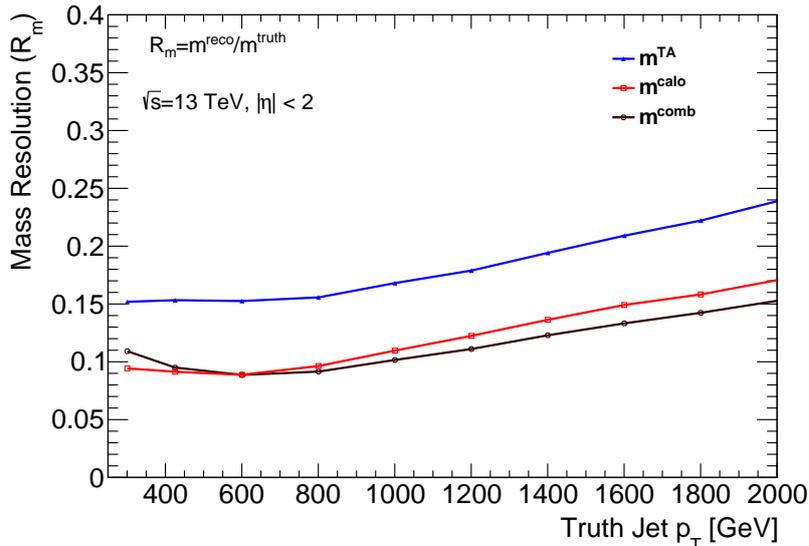


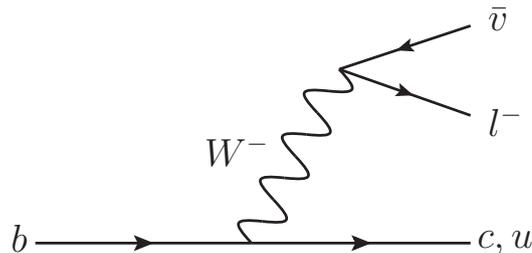
Figure 6.9: Resolution of the jet mass responses as a function of truth Higgs jet  $p_T$  is shown. As described in the text, resolution corresponds to the half of the 68%IQnR divided by the median of  $R_m$  distributions.

Both for the mass and the  $R_m$ , the shapes of the distributions differ depending on the  $p_T$  ranges of the corresponding jets. Therefore, the overall performance in terms of resolution should be considered in several  $p_T$  ranges. Figure 6.9 shows the resolution of the three mass definitions with respect to the truth Higgs jet  $p_T$  value. As it can be seen from the figure, the combined jet mass outperforms almost everywhere. Unlike the situation seen in  $W/Z$  jets,  $m^{\text{TA}}$  does not provide additional benefits for Higgs jets with respect to the  $m^{\text{calo}}$ . At this point it is important to note that, for the same  $p_T$ , the separation in the  $\eta$ - $\phi$  plane between the hadronic decay products of  $W/Z$  jets is smaller than the one for  $b\bar{b}$  pairs coming from Higgs jets. Hence, for the Higgs jets the problem of the merged calorimeter cells is not as crucial as for the  $W/Z$  jets in these  $p_T$  ranges. However, the resolution gain obtained from  $m^{\text{comb}}$  increases by going to the higher  $p_T$  regimes.

To summarise, it has been shown that the  $m^{\text{comb}}$  technique is beneficial for the Higgs jet mass reconstruction. While in the low  $p_T$  regimes the performance of the observable is comparable with the  $m^{\text{calo}}$ , above 350 GeV  $m^{\text{comb}}$  provides up to 6% improvement in resolution of the jet mass response function comparing to the other mass definitions. Both data analyses presented in this thesis use the combined jet mass. The impact of this jet mass definition to the search sensitivities is discussed later in Chapter 7, considering the search for pair produced boosted Higgs bosons both with the  $m^{\text{calo}}$  and  $m^{\text{comb}}$  observables.

## 6.2.2 Muon correction with combined mass

Approximately 33% of the  $b$ -hadron decays produce neutrinos and leptons in their final state as the Feynman diagram in Figure 6.10 depicts. Considering these semi-leptonic decays of the  $b$ -hadrons, the mass resolution of the identified Higgs jets can be further improved. Recalling the fact that neutrinos are not directly measured in ATLAS detector from Chapter 3, it is evident that the reconstructed jets misses the contribution of neutrinos.

Figure 6.10: Feynman diagram of a semileptonic  $b$ -quark decay.

While electrons and hadronically decaying taus leave their energies in the calorimeters, muons deposit only a small fraction of their energy. Together with the muons produced by the leptonic decays of taus, it is convenient to say that in  $\sim 12\%$  of the semileptonic  $b$ -hadron decays, jet reconstruction misses both the muon and the neutrino contributions. However, it is possible to correct the missing muon contribution in the reconstruction of calorimeter jets using the information coming from the MS and the ID.

As a first step, muons are searched within the distance of  $\Delta R < 0.2$  of the  $b$ -tagged track jets (fixed- $R$  jets). If there is more than one muon within a track jet, only the muon closest to the track jet axis is chosen for the correction. The correction is initially developed and used for  $m^{\text{calo}}$  [65, 130]. This section presents its implementation for  $m^{\text{TA}}$  and  $m^{\text{comb}}$  definitions together with its impact on the Higgs jet mass resolution that is obtained using  $m^{\text{comb}}$ .

In order to implement muon correction for  $m^{\text{calo}}$ , the four vector of the identified muon is added to the four vector representing the large- $R$  jet. Then, the energy loss of the muon (while it is traveling inside the calorimeters) is subtracted from this sum in order to prevent double counting. Finally, corrected versions of the  $p_T^{\text{calo}}$  and the  $m^{\text{calo}}$  are retrieved from the final four vector sum. Recalling the definition of  $m^{\text{TA}}$  from Equation 6.3, it can be seen that the only term influenced by the missing muon is the  $p_T^{\text{calo}}$ , since the tracks obtained from the ID already consider the muon contribution. Therefore for the muon correction of  $m^{\text{TA}}$ , it is enough to replace the nominal  $p_T^{\text{calo}}$  term with the corrected  $p_T^{\text{calo}}$  term.  $m^{\text{comb}}$  is the linear combination of these two mass definitions, thus no further correction is needed for  $m^{\text{comb}}$  as the corrected  $m^{\text{calo}}$  and  $m^{\text{TA}}$  observables are used in the combination. In Figure 6.11, the  $m^{\text{comb}}$  distributions of Higgs jets are shown before and after the muon correction in two large- $R$  jet  $p_T$  ranges. The mass resolution of the Higgs jets is improved in both cases, while the improvement is more pronounced for the low Higgs jet  $p_T$  ranges.

The improvement in resolution is quantified using the IQnR method (Section 6.2.1) throughout the entire  $p_T$  range. This time, the mass response function is defined as  $R_m = m^{\text{reco}}/m^{\text{H}}$  using the Higgs parton mass in the denominator instead truth jet mass since the truth jets do not contain muons. Another possible way to evaluate the impact of the correction is to check the mean value of the Higgs jet mass distributions. In Figure 6.12, both the mean and the resolution of  $R_m$  is presented with respect to the truth Higgs jet  $p_T$ .

The results show that while the mean of the  $R_m$  for the uncorrected  $m^{\text{comb}}$  changes from 0.84 to 0.98, the muon corrected  $m^{\text{comb}}$  is mostly stable and closer to unity, indicating that the correction works and reflects the true Higgs boson mass, as expected. The largest resolution improvement in the Higgs jet mass resolution is found to be 3% in the lowest  $p_T$  range of the analysis.

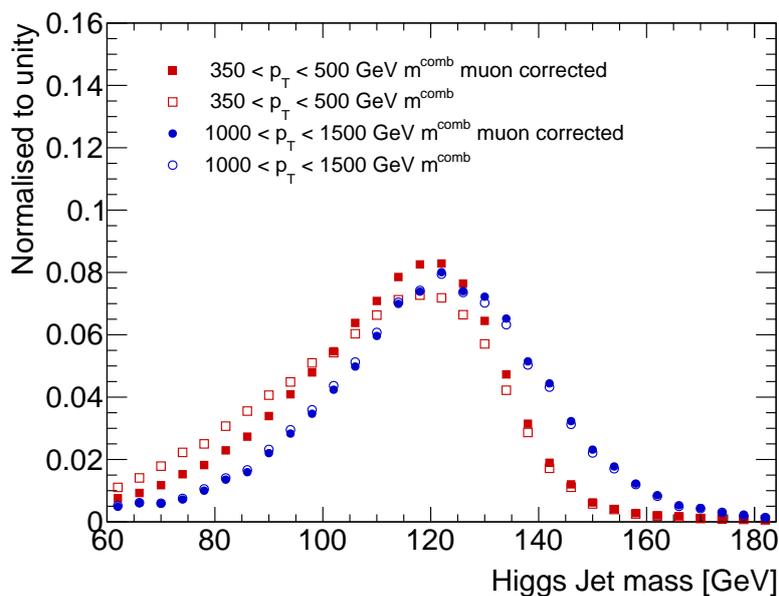


Figure 6.11: Higgs jet mass distribution using the combined mass definition is shown before and after the muon correction in the jet transverse momentum range of  $350 \text{ GeV} < p_T < 500 \text{ GeV}$  (red) and  $1000 \text{ GeV} < p_T < 1500 \text{ GeV}$  (blue).

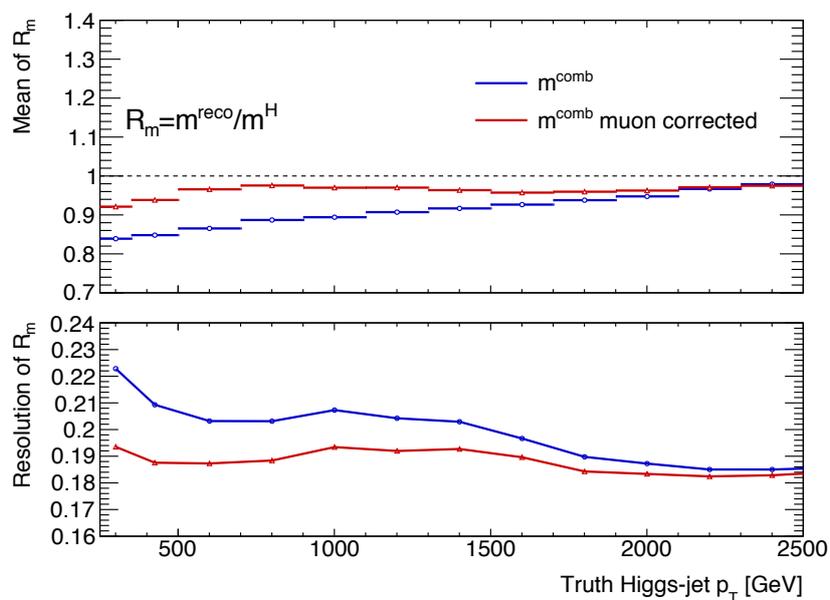


Figure 6.12: Comparison for the mean values (top) and resolutions (bottom) of the Higgs jet mass parton response as a function of truth Higgs jet  $p_T$  before and after the muon correction.

### 6.2.3 Higgs jet mass window

In earlier Higgs tagger approaches,  $m^{\text{calo}}$  was used as jet mass definition, and two different fixed mass windows were defined to select the Higgs jets: [93,134] GeV corresponding to the *tight mass window*, and [76,146] GeV, the *loose mass window* [131]. These ranges correspond to 68% and 90% of the groomed Higgs jet mass distributions, respectively.

The jet mass resolution studies revealed that both the shape and the resolution of the Higgs jet mass distributions are dependent on  $p_T$  of the Higgs jets (Figure 6.9) [65]. Therefore, the idea to develop a  $p_T$  dependent Higgs jet mass window is born in order to improve the performance of the boosted Higgs to  $b\bar{b}$  tagger presented in Section 6.1.

Before going into the details of the developed mass window, it is important to understand the content of the Higgs jet mass distribution so that a reasonable choice can be made for the upper and lower boundaries of the mass window. In the following parts, the underlying physics behind the high and the low mass tails of the distribution are investigated and discussed. Then the final mass window determination is presented.

#### Low mass tail of Higgs jets

As mentioned in 6.1, the presented mass performance studies use the truth information from simulations to identify Higgs jets by selecting large- $R$  jets that contain a ghost associated truth Higgs boson. On top of this requirement, two ghost associated  $b$ -hadrons are required so that all the decay products can be caught inside the large- $R$  jet. Figure 6.13 shows the impact of this requirement on the Higgs jet mass distribution in different kinematic regimes. The impact is visible in the low mass tail of the Higgs jet mass distributions, since some of the large- $R$  jets are not able to capture two  $b$ -hadrons, but only one. As a result, the mass of the reconstructed Higgs jets are only low mass values. The impact disappears in the high  $p_T$  regimes and becomes negligible for the Higgs jets with  $p_T > 1\text{TeV}$ , as both decay products can be captured inside the large- $R$  jet area in this regime.

However, even after applying the  $2b$ -hadron requirement, a non-negligible fraction of the Higgs jets are observed in the low mass region due to the *out-of-cone effects*. Although the truth  $b$ -hadrons are required to be contained within the Higgs jet, the parton shower of one of the  $b$ -hadrons is not fully captured in the jet area. Figure 6.14 shows the distances between the Higgs jet axis and the  $b$ -hadrons for different mass values. This plot suggests that one of the  $b$ -hadrons is identified very close to the Higgs jets axis in the low mass tail indicating that there is no significant contribution from another jet constituent [92].

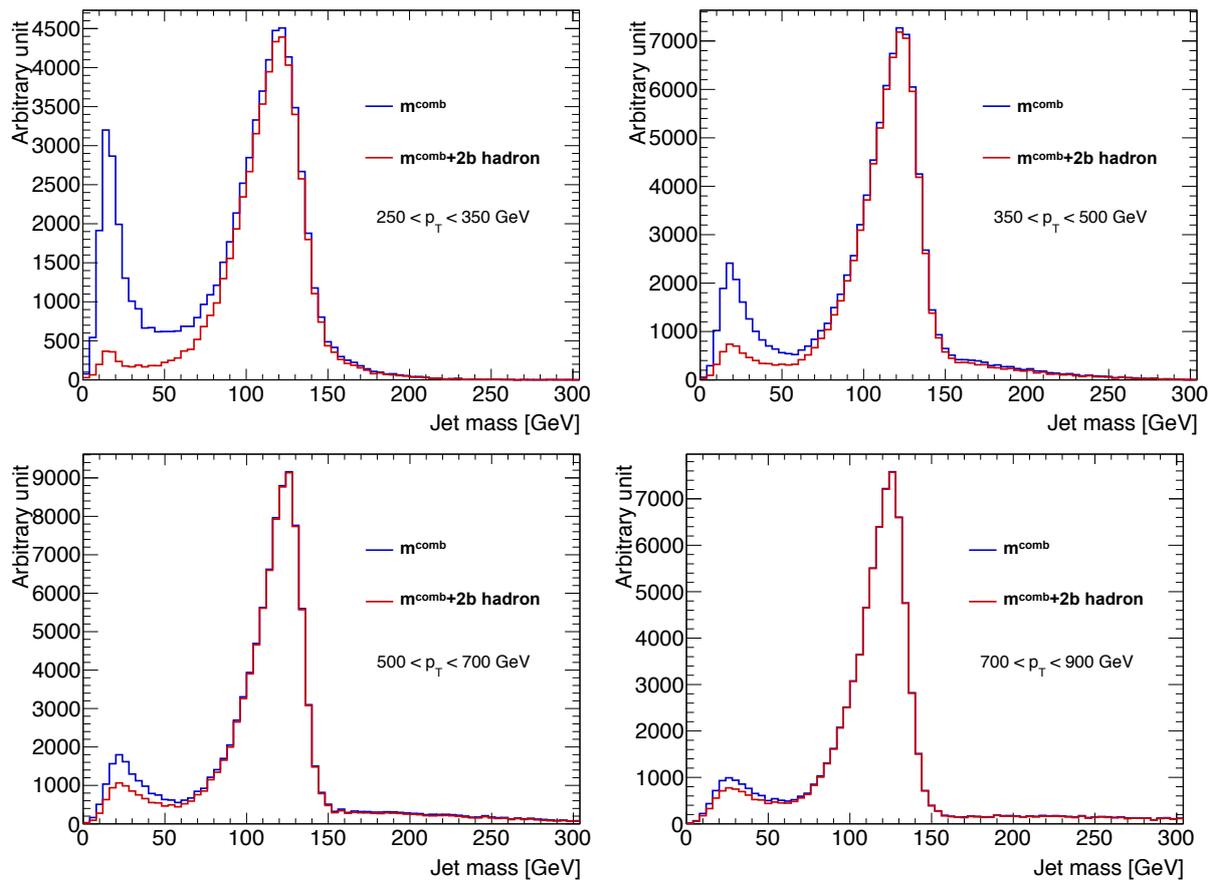


Figure 6.13: The impact of the  $2b$ -hadron requirement on the Higgs jet mass distribution is shown for different  $p_T$  regimes of truth Higgs jets.

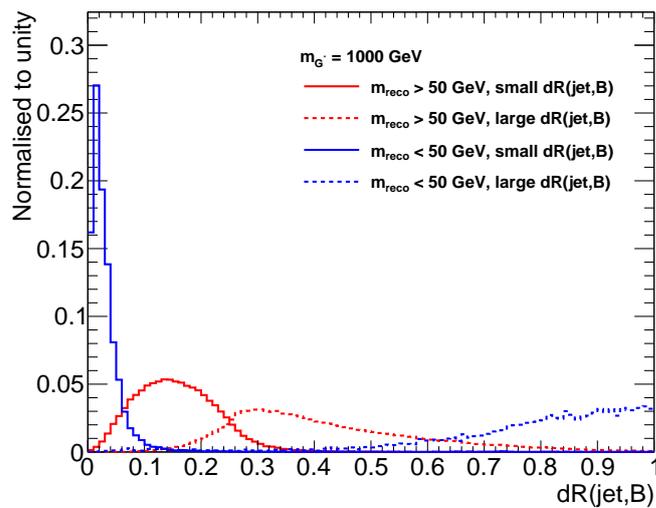


Figure 6.14: Distributions of the  $\Delta R$  between the reconstructed Higgs jet axis and the two  $b$ -hadrons,  $dR(\text{jet}, B)$ , are shown for the jets with their masses larger and smaller than  $m^{\text{reco}} = 50$  GeV [92].

### High mass tail of Higgs jets

Understanding the nature of the jets located in the high mass tail of the Higgs jets mass distribution is an important subject, since a sizeable tail at high invariant mass of the Higgs jets is identified as shown in Figure 6.15. As can be seen from the figure, the size of the tail varies in different  $p_T$  ranges of the Higgs jets, increasing with higher boosts. The presence of such a tail results in a loss of acceptance for the boosted Higgs boson signal if a Higgs boson mass requirement is applied. Therefore, it is important to reveal the underlying facts of this result.

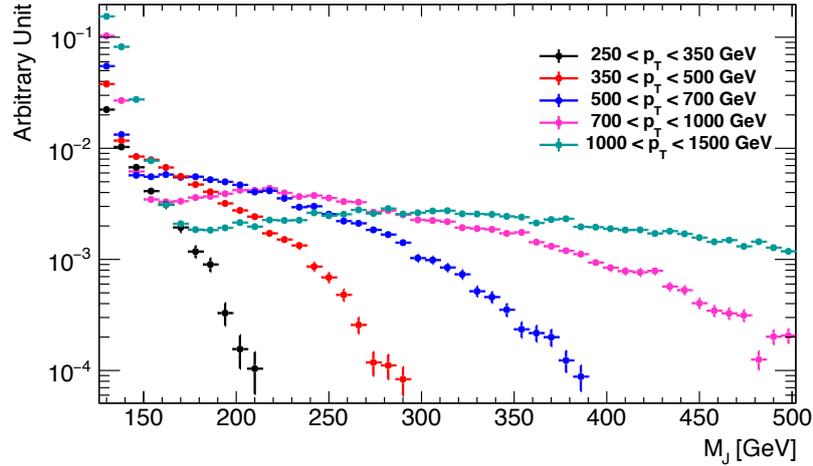


Figure 6.15: High mass tails of the Higgs jets mass distributions are shown in different  $p_T$  ranges of Higgs jets. The size of the tails increase in higher transverse momentum of Higgs jets.

Previous studies [130] excluded the possibility that the high mass tail of the Higgs jets is caused by the underlying events (UE) or pile-up events as this feature is accompanied by additional subjects. However, the same effect in Figure 6.15 has been observed even at the generator level, when there is no pile-up considered. Alternatively, initial state radiation (ISR) or final state radiation (FSR) are thought as other potential mechanisms, but FSR can be directly excluded from first principles, as it will not increase the Higgs jet masses. In contrast, ISR would contribute additional energy unrelated to the  $h \rightarrow b\bar{b}$  decay, which could in turn increase the mass of the Higgs jets.

To verify the ISR hypothesis, a truth level analysis has been performed. Higgs jets containing contributions from ISR are identified using particle level information obtained from simulations. The analysis selection is summarised below:

- a)  $b$ -quarks with  $p_T > 5$  GeV are ghost associated to the subjects that remain after the trimming procedure inside the Higgs jets. If there is at least one subject which does not contain a ghost associated  $b$ -quark, the Higgs jet is selected for further analysis. This category of jets are referred as *Higgs jets with light-jets*.
- b) Using the status code information that is available in Pythia simulations [68], particles coming from ISR are identified and ghost associated to each subjects in the category of *Higgs jets with light-jets*. A minimum requirement of  $p_T > 5$  GeV is imposed on the ISR particles to eliminate a potential contribution from the UE as well as the  $b$ -quarks. Higgs jets are selected if there is at least one light subject which is ghost associated to at least one ISR particle, and they are referred to as *Higgs jets with ISR*.

After applying the above selections to the Higgs jets, the invariant mass distributions are compared for the different  $p_T$  bins as shown in Figure 6.16. While Higgs jets after selection a) contain FSR resulting in a peak around the Higgs boson mass, the mass distribution of the ISR tagged Higgs jets is not resonant in the Higgs mass any more after selection b). While Figure 6.16 underlines already that the high invariant mass tail is dominated by Higgs jets with ISR, results in non logarithmic y scale can also be seen in Appendix A.

The same analysis is repeated using different jet reconstruction algorithms to validate the findings, and the results are found to be qualitatively identical to the above results (see Appendix A). This builds further confidence in the hypothesis that the high mass tail of Higgs jets is dominated by the contributions from ISR.

### $p_T$ dependent mass window and final fit

Providing a stable mass window with a fixed Higgs jet selection efficiency as a function of Higgs jet  $p_T$  is challenging as the tails of the mass distribution play a substantial role to define the boundaries of the windows. The boundaries which are determined with the IQnR method fluctuate significantly due to the very low statistics in the tails. That prevents a definition of a fit function for the mass window with respect to  $p_T$ . Thus, to mitigate the statistical fluctuation in the tails, the mass distribution is fitted with the following function which is a linear combination of a landau function for the low mass and a gauss function for the high mass regions of the distribution:

$$f(m) = \frac{a_l}{2\pi i} \cdot \int_{c-i\infty}^{c+i\infty} \exp(s \log s + (m - \mu_l) s) ds + \frac{a_g}{\sqrt{2\pi\sigma_g^2}} \cdot \exp\left(-\frac{(m - \mu_g)^2}{2\sigma_g^2}\right) \quad (6.5)$$

The parameters  $a_{l,g}$  define the amplitude,  $\mu_{l,g}$  the position on the mass axis, and  $c, \sigma_g$  are the scale parameters of the landau and gauss components, respectively. Figure 6.17 shows the two components of the fit function together with the Higgs jet mass distribution in the  $p_T$  range of 350 to 500 GeV. The landau function features a tail towards low masses which converges sufficiently quick to zero, and describes very well the lower edge of the distribution, as shown in the figure for this particular  $p_T$  bin. The gauss function fit is considered well enough for the higher edge of the distribution. For the lower  $p_T$  regimes, it agrees with the distribution up to 160 GeV and for the higher  $p_T$  regimes the agreement continues up to 170 GeV.

In general, the fit describes the core of the mass distribution well and is stable across the entire range of transverse momentum. Calculating the upper and lower boundaries of the mass windows to satisfy the 68% and 80% signal selection efficiencies, the final mass windows are determined for all  $p_T$  ranges. The results are fitted with Equation 6.1 through the entire  $p_T$  spectrum and shown in Figure 6.18.

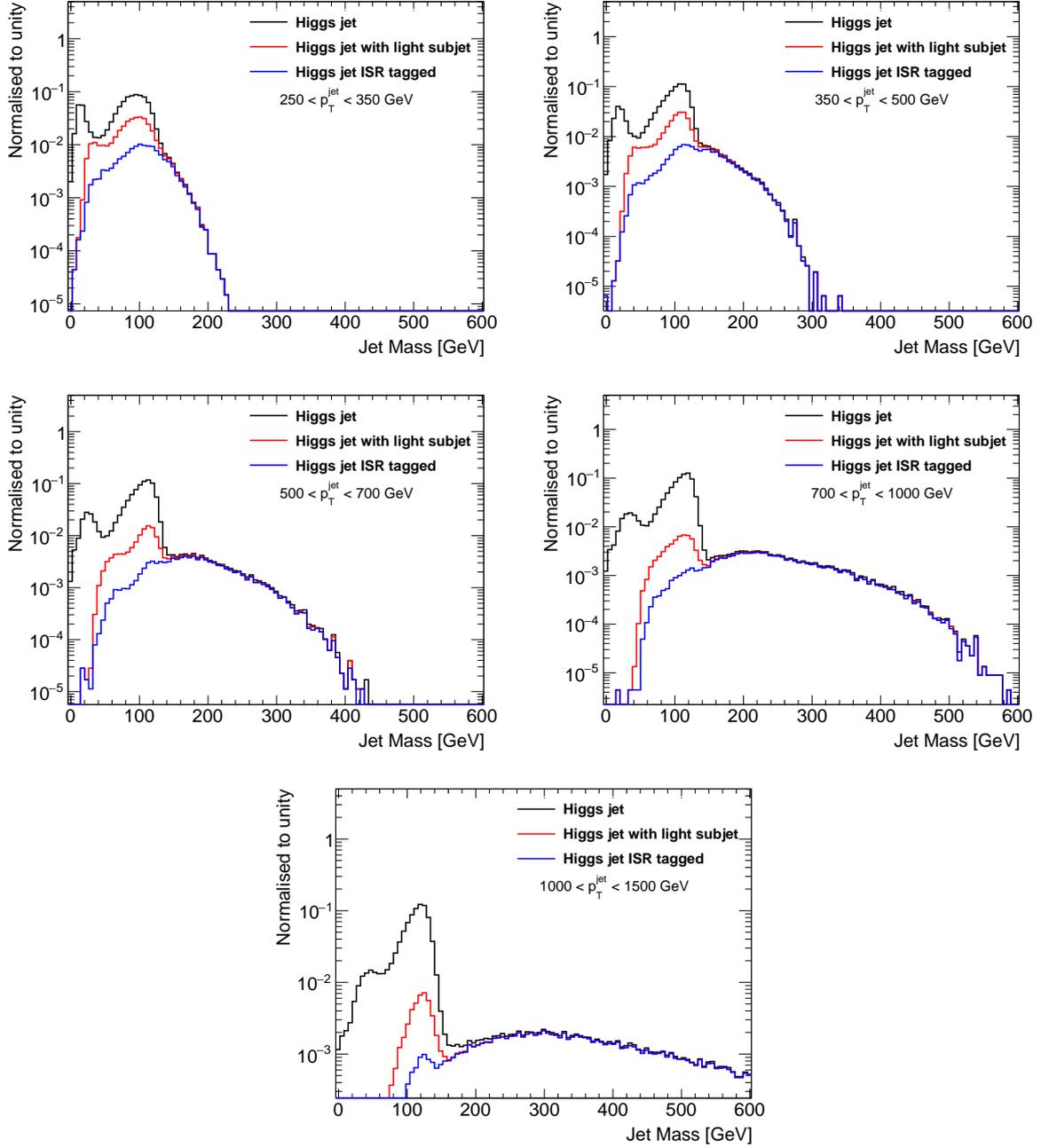


Figure 6.16: Invariant mass distributions of Higgs jets with  $p_T > 250$  GeV and  $|\eta| < 2.0$  before any selection (black), after the requirement that at least one of the subjets should not contain a ghost-associated  $b$  quark (red), and after the additional requirement of the subjets not associated to a  $b$  quark to contain ghost-associated particles from ISR (blue). The results are shown for different Higgs jets  $p_T$  regimes. In all kinematic regions, the high mass tail is dominated by ISR.

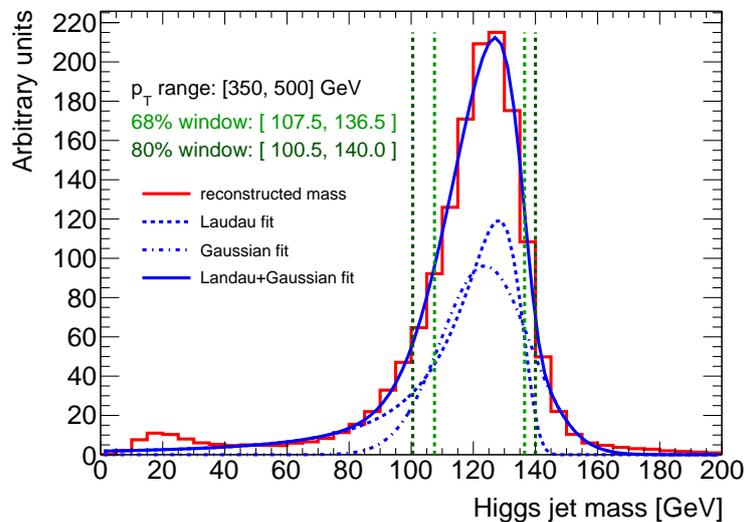


Figure 6.17: Reconstructed Higgs boson mass distribution for Higgs jet  $p_T$  in the range of 350 to 500 GeV. The dotted and dash-dotted blue lines correspond to the two components of the fit function, while the solid blue line shows their combination. The vertical lines indicate the boundaries of mass ranges for 68% (light green) and 80% (dark green) [92].

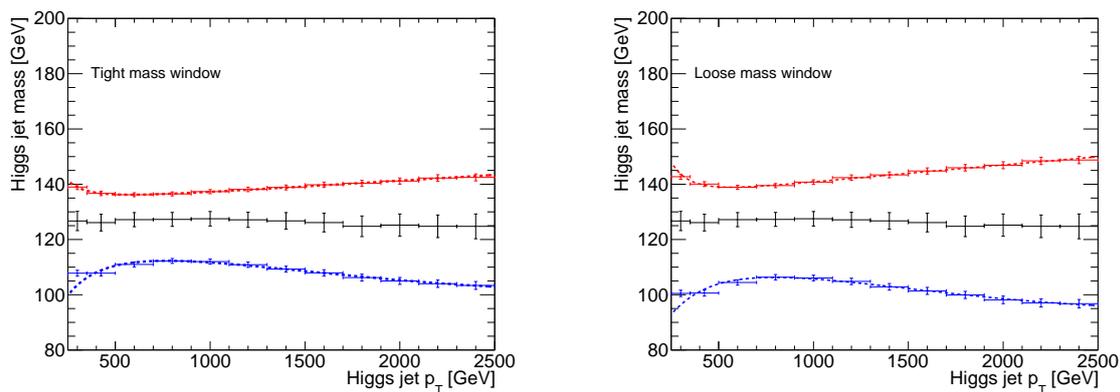


Figure 6.18: Upper (red) and lower (blue) boundaries of the tight and loose Higgs jet mass windows as a function of the truth Higgs jet  $p_T$  [92].

### 6.3 Alternative Higgs jet reconstruction techniques

Alternative jet reconstruction and grooming methods have been studied in detail, mostly in the context of  $W$  and  $Z$  boson tagging in late Run 1 as a preparation for Run 2 ATLAS analyses [132]. As a result of these studies, the default Run 2 large- $R$  jet reconstruction method was determined to be the anti- $k_t$  algorithm with  $R = 1.0$  trimmed with  $R_{\text{sub}}=0.2$   $f_{\text{cut}}=0.05$  (see Section 5.1). As the jet calibration and uncertainty estimations are developed for this reconstruction method, the Run 2 boosted Higgs analyses use this default jet reconstruction method. However, alternative jet reconstruction methods and parameters for the identification of the boosted  $h \rightarrow b\bar{b}$  process are investigated using the *FastJet* software [133] and presented in this section.

Note that the methods mentioned and used in this section are described extensively in Chapter 5, and they can be recalled if it is needed. All the alternative jet reconstructions presented here use LC calibrated calo-clusters as jet input and the results are compared using  $m^{\text{calo}}$  as jet mass observable. Similarly to the previous sections, simulated  $G_{KK} \rightarrow HH$  samples are used as to simulate the Higgs bosons. After the jet reconstruction, truth Higgs bosons are ghost associated to the large- $R$  jet area to obtain *Higgs jets* without requiring  $2b$  hadrons to be able to compare the reconstruction performances through the entire jet mass distribution. Besides, as shown in Figure 6.13, this requirement has an impact only in the low mass distribution. Table 6.1 presents the studied jet reconstruction methods and the total number of reconstructed jets in different stages of the reconstruction and selection procedure. According to this result, the

Algorithm	Radius	Grooming	Parameter1	Parameter2	ParentJets	GroomedJet	HiggsAssoc	Eta>1.6
AntiKt	1.0	Trimming	$R_{\text{sub}}=0.2$	$f_{\text{cut}}=5\%$	2477997	2477997	1024470	961888
AntiKt	1.0	Trimming	$R_{\text{sub}}=0.3$	$f_{\text{cut}}=5\%$	2477997	2477997	1024470	961035
AntiKt	1.0	Trimming	$R_{\text{sub}}=0.2$	$f_{\text{cut}}=8\%$	2477997	2477997	1024470	962390
AntiKt	0.7	Trimming	$R_{\text{sub}}=0.2$	$f_{\text{cut}}=5\%$	2127995	2127995	1022950	964234
C/A	1.2	Split Filter	$R_{\text{sub}}=0.3$	$y_{\text{cut}}=15\%$	2686074	2606818	975239	913824
C/A	1.2	Split Filter	$R_{\text{sub}}=0.3$	$y_{\text{cut}}=4\%$	2686074	2680372	1024060	959411
C/A	0.7	Split Filter	$R_{\text{sub}}=0.3$	$y_{\text{cut}}=4\%$	1982885	1972058	1009130	950160
C/A	1.0	Pruning	$R_{\text{cut}}=0.5$	$z_{\text{cut}}=15\%$	2477997	2477997	1024470	960565

Table 6.1: Cut flow for Higgs jet selection using alternative jet reconstruction methods using *FastJet* software and calo-clusters as an input for jet finding algorithms.

pruned Cambridge-Aachen and the trimmed anti- $k_T$  jet reconstructions using parameter of  $R = 1$  lead to the same maximum number of Higgs jets, implying that the radius parameter choice as 1 is can yield higher signal efficiencies for boosted  $h \rightarrow b\bar{b}$  processes. After requiring certain kinematic range, anti- $k_T$  trimmed jets with  $R = 0.7$  indicates higher yields. However, in order to evaluate the real performance of the reconstructed jet collections, the Higgs jet mass distributions must be compared. Figure 6.19 compares the Higgs jet mass distributions obtained from the alternative jet reconstruction methods. Note that the presented results are not calibrated, therefore the mean value of the distributions does not convey any message. Nonetheless, the shapes of the core mass distributions and the tails are two figures of merits to interpret the results.

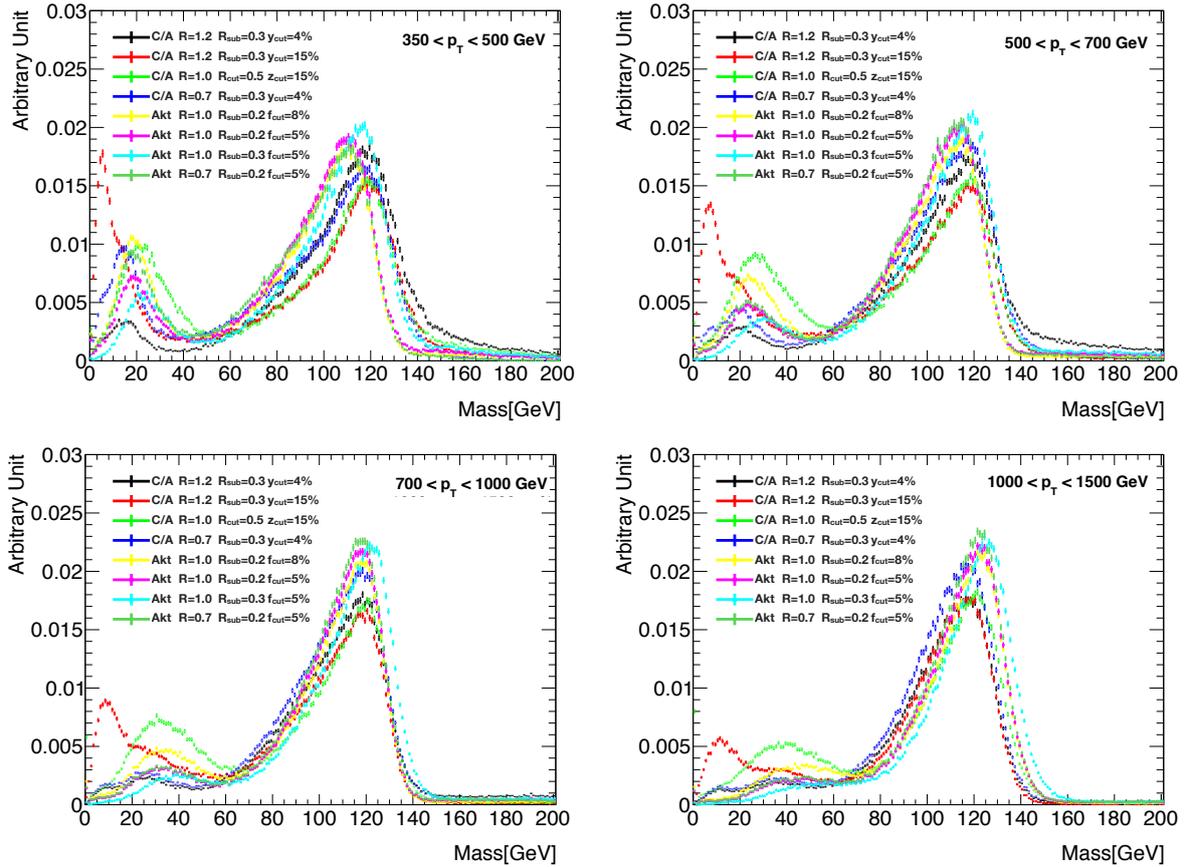


Figure 6.19: Higgs jet mass distributions obtained from alternative jet reconstruction methods are shown for different  $p_T$  regimes. The grooming methods are not indicated in the legends, however they can be recognised by their parameters. All Cambridge-Aachen algorithms except the the bright green are groomed with the split-filtering method and all the anti- $k_T$  algorithms are trimmed. Only the bright green Cambridge-Aachen algorithm is groomed with pruning.

Table 6.2: The RMS parameter and the integral of the mass distribution of Higgs jets above  $p_T > 140$  GeV for the studied jet reconstruction and grooming configurations in bins of Higgs jet  $p_T$ .

Algorithm	$1000 < p_T < 1500\text{GeV}$		$700 < p_T < 1000\text{GeV}$	
	Tail Integral	RMS	Tail Integral	RMS
Anti- $k_T$ $R=1.0$ $R_{\text{sub}}=0.2$ $f_{\text{cut}}=0.05$	0.122	22	0.092	21
Anti- $k_T$ $R=1.0$ $R_{\text{sub}}=0.3$ $f_{\text{cut}}=0.05$	0.124	21	0.091	21
Anti- $k_T$ $R=0.7$ $R_{\text{sub}}=0.2$ $f_{\text{cut}}=0.05$	0.066	21	0.039	20

Comparing the low mass tails, pruned Cambridge-Aachen jets, split filtered Cambridge-Aachen jets with  $R = 1.2$ ,  $y_{\text{cut}}=15\%$ , and anti- $k_T$  jets trimmed with  $f_{\text{cut}}=8\%$  are performing worse with the presence of a high amount of low mass jets. Looking at the core distribution, it can be seen that the remaining anti- $k_T$  jets are performing similarly good and better than Cambridge-Aachen jets. To quantify the further differences between them, root mean square (RMS) values of the core distributions and the integrals of the high mass tails are compared as presented in Table 6.2. The results indicate that anti- $k_T$  jets reconstructed with radius parameter of 0.7 have lower RMS values and smaller high mass tails.

In addition to the above studies, the performance of variable- $R$  jet finding algorithm (see Section 5.1.1) is also compared with the default large- $R$  jet reconstruction (AKT10 trim) method. Variable- $R$  jets are reconstructed with the parameters of  $\rho=600$  GeV,  $R_{\text{max}}=1.0$ , and  $R_{\text{min}}=0.2$  and they are trimmed similarly to the default collection (VR600trimmed). Comparisons are performed in terms of signal selection and background rejection efficiencies and it is found that variable- $R$  jets perform better in the regime where the jet  $p_T > 1$  TeV [101]. Moreover, undesirable ISR contributions that are captured in the jet area can be reduced by using variable- $R$  jet algorithm. In order to prove that, jets with ISR contributions are tagged as described earlier in Section 6.2.3. Figure 6.20 compares the mass distributions obtained from both reconstruction methods indicating the jets with ISR contributions. As it can be clearly seen from the results, the high mass tail of the Higgs jet mass distribution caused by the ISR contributions can be reduced with variable- $R$  jet reconstruction.

In summary, the studies performed in this section showed that default ATLAS Run 2 large- $R$  jet reconstruction is one of the optimum choices for boosted  $h \rightarrow b\bar{b}$  identification among the studied alternatives. However, lowering the jet radius parameter can offer better resolution for the Higgs jet mass peak in boosted regions and it reduces the observed high mass tail of Higgs jets as it is observed both for anti- $k_T$  jets with  $R=0.7$  and variable- $R$  jets.

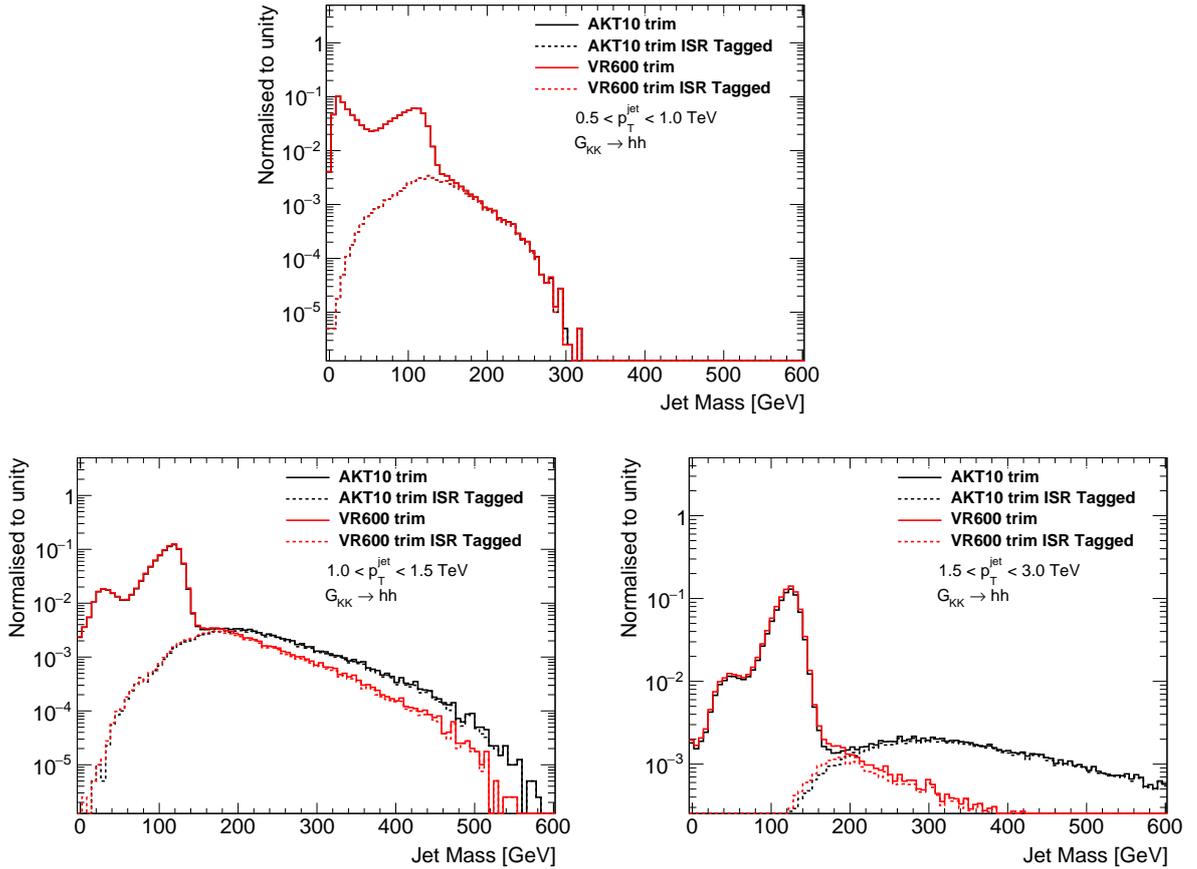


Figure 6.20: The invariant mass distributions for Higgs jets are compared for trimmed VR (red) and trimmed anti- $k_T$   $R = 1.0$  jets (black) in various Higgs jet  $p_T$  regions. The distributions are shown for all Higgs jets (solid lines) and for Higgs jets that are ISR-tagged (broken lines), i.e. jets where at least one of the subjets does not contain a ghost-associated  $b$  quark, but at the same time contains at least one ghost-associated particle from ISR. The shoulder at Higgs jet masses above 150 GeV, dominated by the contributions from ISR, is significantly reduced for variable- $R$  jets for  $p_T \gtrsim 1$  TeV. The impact is more pronounced in higher jet  $p_T$ .

## 6.4 Conclusion

In this chapter, the ATLAS Run 2 baseline boosted  $h \rightarrow b\bar{b}$  tagging strategy is presented together with the performed studies in this subject. Significant improvements and results are shown in terms of the Higgs jet mass reconstruction. The combined mass for the Higgs jets is studied and implemented for the first time in the context of the boosted  $h \rightarrow b\bar{b}$  tagger as well as the implementation of muon in jet correction for  $m^{\text{comb}}$ . It has been shown that the resolution improvement for the jet mass response is increasing for higher Higgs jet  $p_T$  with  $m^{\text{comb}}$ , although the impact is not as significant as it is for boosted  $W/Z$  boson decays.  $m^{\text{comb}}$  is used in the physics analyses presented in the following chapters, as well as in many other ATLAS Run 2 analyses with the involvement of a boosted  $h \rightarrow b\bar{b}$  process.

In order to understand better the nature of the Higgs jet mass distribution, and to introduce a mass window, several studies are performed and their results are presented in this chapter. Using truth level information obtained from simulations, ISR contributions are tagged inside Higgs jets, proving the hypotheses that the high mass tail of Higgs jet mass is caused by ISR, especially for highly boosted Higgs jets. A  $p_T$  dependent mass window is developed and implemented as ATLAS Run 2 boosted  $h \rightarrow b\bar{b}$  tagger recommendations.

Finally, alternative jet mass reconstruction methods are studied for the Higgs boson identification. It has been shown that default Run 2 large- $R$  jet reconstruction method performs similarly, or better than the studied alternative jet reconstruction methods. However, reducing the jet radius parameter in the default jet reconstruction method and variable- $R$  jet finding algorithm are found to be promising alternatives, particularly to avoid from the undesirable ISR contribution inside the Higgs jets.

# Chapter 7

## Search for boosted di-Higgs production in the $b\bar{b}b\bar{b}$ final state

### 7.1 Overview

This chapter presents the search for pair production of boosted Higgs bosons in the  $b\bar{b}b\bar{b}$  final state with  $36.1 \text{ fb}^{-1}$  of ATLAS  $pp$  data. The aim of this chapter is to explain the basic methodology of the analysis and discuss the background estimation method in detail. The analysis uses several methods which are introduced in Chapter 6 in order to identify boosted Higgs bosons. Therefore, one of the aims of this chapter can be also considered as to show the functionality of the methods shown in Chapter 6.

The analysis is mainly performed to search for new physics, but it also carries great significance for the future SM measurements. In this section, the motivation of the search and the analysis strategy are given. In Section 7.2, event selection and final event categorisation is presented. Section 7.3 discusses the background modelling of this analysis, particularly the multijet background. As can be recalled from Figure 6.1 and Chapter 6, multijet production is substantially overwhelming background for Higgs jets, that is why it is studied and presented extensively. Finally in Section 7.5, results of the search is briefly presented together with the systematic uncertainties, final data-background comparisons, and the statistical evaluation of the search.

#### 7.1.1 Searching for new physics

This search looks for a possible new resonance decaying into two Higgs bosons as shown in Figure 7.1. There are several new physics models predicting significantly higher rates for the Higgs boson pair production than the SM rate [134–136]. As introduced in Section 2.3.1, the Kaluza Klein (KK) excitation of the graviton ( $G_{KK}$ ) predicted in the bulk Randall-Sundrum (RS) model [137, 138], and extensions of the Higgs sector (see Section 2.3.3) such as heavy neutral scalar as predicted in two-Higgs-doublet models (2HDM) [139] are two possible candidates for such a resonance. Beside of these BSM theories, there are also some cases which are not suggesting a new resonant particle but an enhanced Higgs boson pair production, such as new, light, coloured scalars [140], direct  $t\bar{t}hh$ ,  $thh$  vertex modifications [141, 142] or modified  $\lambda_{hhh}$  parameter, which can result in higher production rates for the process.

In this search, both the KK excitation of the spin-2 graviton predicted within a Bulk RS model and a narrow-width scalar resonance decaying into a Higgs boson pair are

used as benchmark signals for the resonant di-Higgs production. Furthermore, for the enhanced non-resonant Higgs boson pair production, upper limit is set on the production cross section.

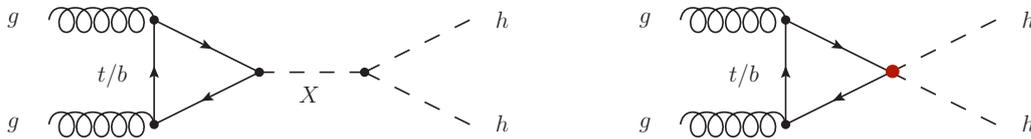


Figure 7.1: Example Feynman diagrams for BSM di-Higgs production illustrating the production of a Higgs boson pair via the decay of an intermediate resonance ( $X$ ) produced through a heavy-quark loop (on the left) and via a new vertex or modified  $\lambda_{hhh}$  parameter (on the right).

### 7.1.2 SM non-resonant Higgs boson pair production

Apart from the new physics motivations, the other exciting aspect of this channel is to measure the SM di-Higgs production. The SM processes leading to double Higgs production at hadron colliders can be categorised into four main classes: gluon fusion, vector boson fusion, double Higgsstrahlung and associated production with a top quark pair [143]. The dominant production mode is the gluon fusion process in analogy to single Higgs production and is shown in the Feynman diagrams in Figure 7.2. The cross section of the gluon fusion production mode is about one order of magnitude larger than the second largest process which is vector boson fusion.

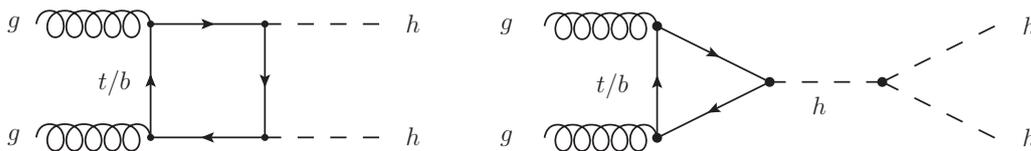


Figure 7.2: Feynman diagrams for SM di-Higgs production via gluon fusion

The measurement of double Higgs production is one of the central physics goals both for the current LHC program and also for the HL-LHC. These processes are sensitive to Higgs trilinear coupling which can provide valuable information on the electroweak symmetry breaking (Section 2.1.3). The third term in the Higgs scalar potential, given in Equation 2.13, represents the Higgs self coupling ( $\lambda_{hhh}$ ), and provides information about the shape of the Higgs potential. Since  $\lambda_{hhh}$  also plays a role in di-Higgs production, the SM di-Higgs measurement is a direct way to obtain information on the Higgs sector and it is very crucial. However, the cross section of the SM di-Higgs production at 13 TeV LHC is around 33 fb [144, 145], which is too small to measure in the current dataset and energies of Run 2. Therefore, this analysis can not be sensitive for such a measurement. In this regard, while the main motivation behind performing the presented analysis is to search for new physics, it is possible to set upper limits for the di-Higgs production, providing information and reference for future SM di-Higgs production measurements.

### 7.1.3 Analysis strategy

As motivated above, performing a search for di-Higgs production is promising. Considering the wealthiness of the possible final states produced by the di-Higgs decays, various search channels can be studied. Nevertheless, with the largest branching fraction,  $b\bar{b}b\bar{b}$  final state becomes one of the natural choices among the others providing higher yields for such a rare process. Although, this comes with a cost due to the overwhelming multijet backgrounds seen in fully hadronic final states (see Figure 6.1), the combination of the different final states shows that, the highest sensitivity among the other channels for resonance masses above 500 GeV is still achieved in the  $b\bar{b}b\bar{b}$  channel [146]. The previous ATLAS di-Higgs production search in the  $b\bar{b}b\bar{b}$  final state has been performed with a  $3.2 \text{ fb}^{-1}$  dataset [147], excluding certain resonant mass ranges and setting upper limits on the di-Higgs production cross section. This analysis aims to repeat the aforesaid search with a larger dataset and improved techniques.

Preliminary results [148, 149] showed that using different approaches to the reconstruction of the two Higgs decay significantly changes the sensitivity of the search. As can be seen from Figure 7.3, while the search for four resolved  $b$ -jets in the final state is the more sensitive approach in the region where resonant mass is below  $\sim 1 \text{ TeV}$ , the boosted approach which looks for two back-to-back large- $R$  jets in the final state provides better sensitivity than the resolved one for the resonant mass values above  $\sim 1 \text{ TeV}$ . In order to have the best sensitivity throughout the whole mass range, both approaches were used together.

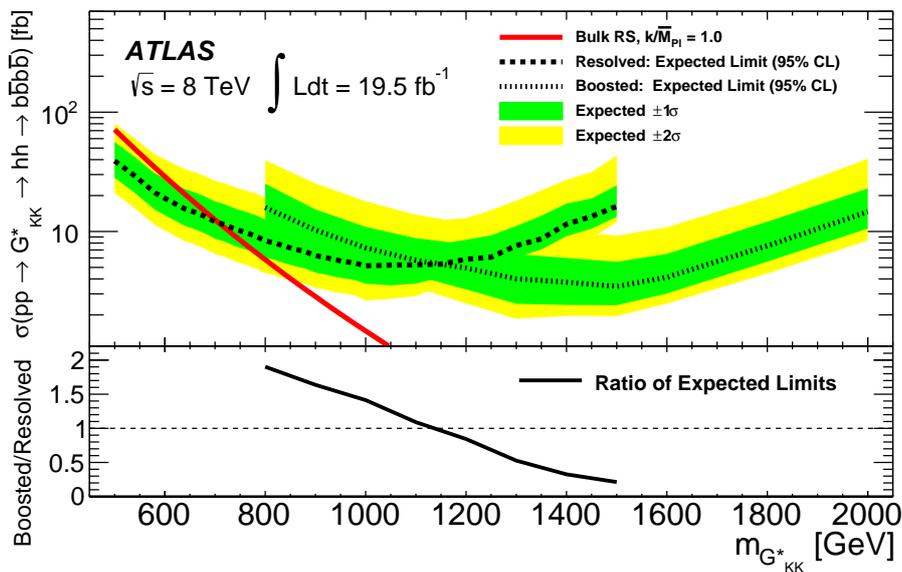


Figure 7.3: The overlay of the expected limits for the boosted and the resolved analyses for the bulk RS model with  $k/\bar{M}_{Pl} = 1$ . The red curves show the predicted cross-section as a function of resonance mass for the model considered [149].

In analogy to that, also in this round of the analysis, both approaches are performed in parallel, employing orthogonal selections. In case there is no sign of new physics, a combination of the results from the resolved and the boosted searches can provide the strongest limits and exclusions for the searched signals. While the results obtained from the resolved analysis are for a resonance mass between 260 and 1400 GeV, the boosted analysis is performed for masses between 800 and 3000 GeV. In the mass ranges where two analyses overlap, a statistical combination is performed.

In this thesis, only the boosted analysis is presented. Two back to back large- $R$  jets are required to represent each Higgs boson as shown in Figure 7.4. The invariant mass of the di-Higgs ( $m_{HH}$ ) system is used as the final discriminant of the analysis, which corresponds to the invariant mass of the di-jet system ( $m_{JJ}$ ). This chapter presents the optimised event selection and background estimation methods for the boosted analysis. More details about the boosted analysis, resolved analysis and the statistical analysis performed for the combination of the results can be found in References [66, 74, 150], respectively.

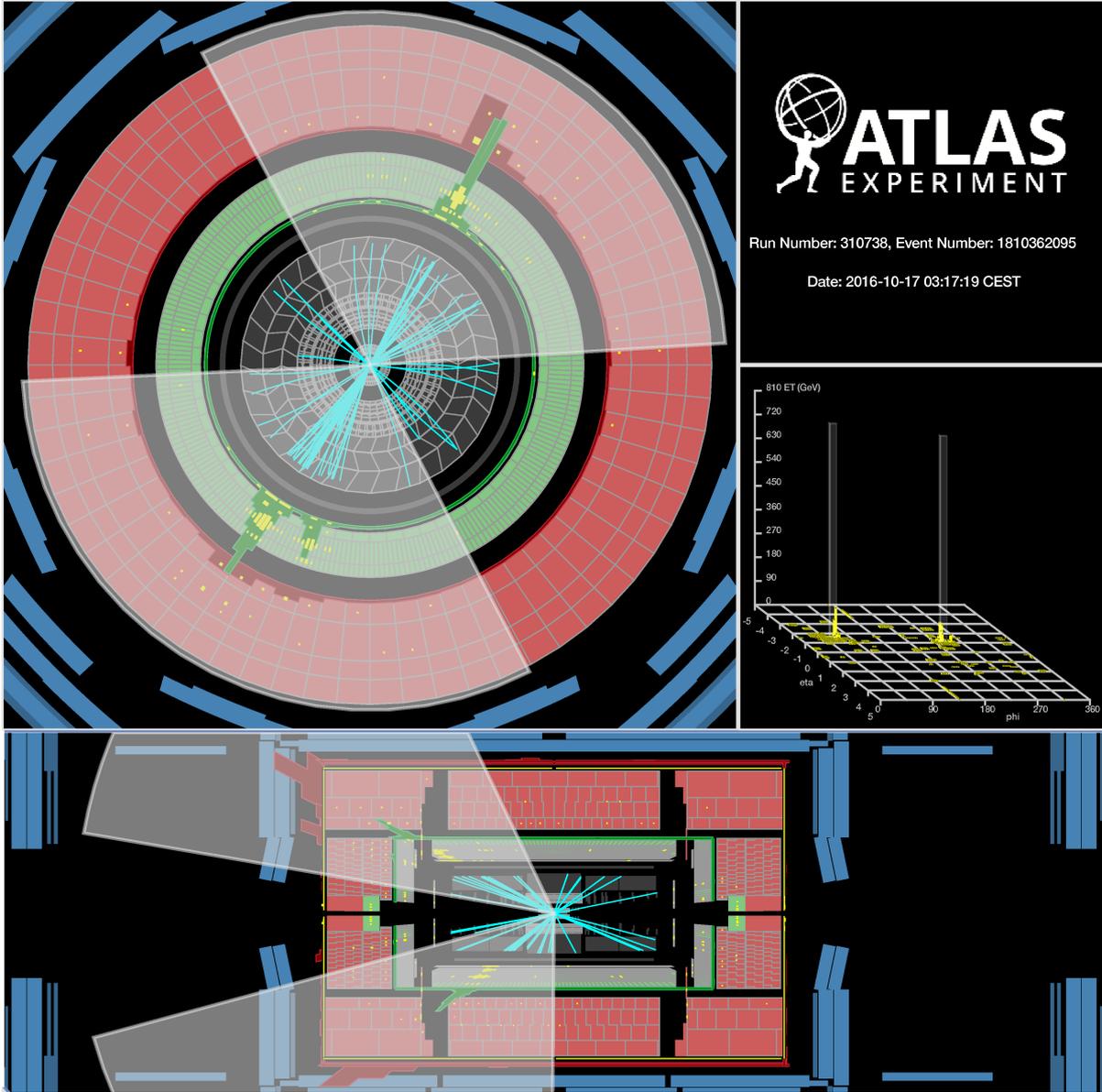


Figure 7.4: The event shown passes the boosted signal region in the two-tag sample, i.e. it contains two large- $R$  jets with one  $b$ -tagged track jet associated to each. The event was recorded during 2016 in 13 TeV  $pp$  collision data. The calibrated large- $R$  jets have transverse momenta of 748 GeV and 747 GeV. The value of the invariant mass of the two selected large- $R$  jets is 3.89 TeV (after rescaling the four-vectors of the Higgs boson candidates to match the Higgs boson mass). There are two track jets associated to each large- $R$  jet, with values of  $p_T$  of 38 GeV, 193 GeV, 259 GeV and 301 GeV. One of the two track jets in each large- $R$  jet is  $b$ -tagged. The tracks shown have transverse momenta above 2 GeV, and the energy deposits in the calorimeters exceed 0.5 GeV [74].

## 7.2 Event Selection

In this section, the trigger requirement, event preselection, and final event categorisation of the boosted analysis are presented.

### 7.2.1 Trigger

The events are required to be triggered by the presence of a large- $R$  jet reconstructed by the HLT and seeded by the lowest unrescaled L1 jet trigger which corresponds to a 100 GeV  $p_T$  threshold. Both in data and simulation, events must pass the lowest unrescaled large- $R$  jet trigger which are different for different years of data taking:

- 2015 - HLT ungroomed large- $R$  jet with  $p_T > 360$  GeV,
- 2016 - HLT ungroomed large- $R$  jet with  $p_T > 420$  GeV,

For the signal masses above 1200 GeV, the efficiencies for both triggers are found to be  $> 98\%$  [74].

### 7.2.2 Preselection

Before going into the final event selection to obtain signal events, there are some preselection requirements applied to each event passing the trigger requirement. The preselection criteria can be considered as natural choices either due to the topology that is being studied in the search like kinematic cuts, or the experimental necessities which have to be applied to ensure the data quality such as event or object cleaning procedures. Below they are listed according to their order in the actual analysis selection.

- As explained in Section 4.1, the analysis uses the *good dataset* that is selected with respect to data taking conditions. In addition to this selection, special data cleaning procedures are applied following the recommendations given in Reference [151], like removing the events affected by the individual problems seen in sub-detectors such as LAr calorimeters, Tile calorimeters, and SCT, and the events that are not fully reconstructed.
- A jet cleaning procedure is applied to the events both in data and simulations in order to remove the events that contain significant calorimeter noise, such as an unclean jet originated from non-collision background processes [152]. The existence of such a jet may have a negative impact on the event reconstruction. These jets are identified from small- $R$  jets using certain quality criteria referred as *LooseBad* in given reference. If a jet is identified as *bad jet* and fulfills any of the following kinematic conditions, the event is removed:
  - $p_T > 60$  GeV,
  - $20 < p_T < 60$  and  $|\eta| \geq 2.4$ ,
  - $20 < p_T < 60$ ,  $|\eta| < 2.4$  and the jet is not marked as pile-up jet by JVT algorithm (see Section 5.1.3).
- At least two large- $R$  jets are required in each event with  $p_T > 250$  GeV,  $|\eta| < 2$ ,  $m^{\text{comb}} > 50$  GeV and arbitrary number of ghost associated fixed- $R$  track jets, as described in Section 5.1.4, 5.1.5, and 6.1.2.

- The leading large- $R$  jet is required to have  $p_T > 450$  GeV so that both of the triggers are fully efficient.
- Track jets are required to have  $p_T > 10$  GeV.
- The leading and subleading large- $R$  jets are considered as the Higgs candidates and must satisfy  $|\Delta\eta| = |\eta_{\text{leadjet}} - \eta_{\text{subjet}}| < 1.7$ .
- In order to have an orthogonal selection with the resolved analysis, events that pass the resolved signal region selection are rejected. The resolved selection requires for any event at least four  $b$ -tagged small- $R$  jets with  $p_T > 40$  GeV,  $|\eta| < 2.5$ . MV2c10 algorithm with 70% working point is used for  $b$ -tagging the small- $R$  jets. Further requirements are the cuts on the  $X_{hh}$  variable, that is used also for the boosted analysis and described later in Section 7.2.3, and the minimisation on the  $D_{hh}$  variable to select the correct combination of  $b$ -jet pairs for Higgs reconstruction out of 4  $b$ -jets. The definition of the  $D_{hh}$  variable and a more detailed explanation of the resolved analysis can be found in Reference [150].
- As a last step of the preselection, muon correction for  $m^{\text{comb}}$  is applied for each Higgs candidate jet as explained in Section 6.2.2 using CB muons with  $p_T > 4$  GeV,  $|\eta| < 2.5$ , and passing at least the *medium* quality requirement introduced in Section 5.2.

### 7.2.3 Final Event Categorisation

After applying the preselection, events are categorised according to the  $b$ -tagged track jet multiplicities in the Higgs candidates. Note that there is no requirement on the minimum number of track jets in this analysis. Each of the track jets are checked to identify their flavours and then depending on the total number of  $b$ -tagged track jets, the following categorisation is performed. Moreover, two variables are introduced in order to select an optimum mass window for both of the Higgs candidates. Using these variables, each track jet category is further divided into the signal region, the sideband region, and the control region.

#### Signal Region

For the signal region (SR) selection, the presence of  $b$ -tagged track jets associated to the candidate large- $R$  jets, and a large- $R$  jet mass ( $m_J$ ) around 125 GeV (as discussed earlier in Chapter 6) are required. As shown in Section 6.1.2, requiring different numbers of  $b$ -tagged track jets for the Higgs candidate culminates in different signal selection efficiencies. Therefore, also for this analysis, various approaches are considered for the choice of the number of  $b$ -tagged track jets per Higgs candidate to maximise the signal efficiency. Although the naive expectation is to have 4  $b$ -tagged track jets in the final state of the given process, 3  $b$ -tagged track jets and 2  $b$ -tagged track jets are also selected for the SR, since  $b$ -tagged track jets merge in highly boosted regimes as discussed in Section 6.1.1 and depicted in Figure 6.3. Below, these three categories are given. Sketches of the events in each category are shown in Figure 7.5.

- 4-tag: For each Higgs candidate jet, 2  $b$ -tagged track jets are required.  $b$ -tagging WPs are selected as 70% for all the track jets in the analysis. There is no limitation for the additional numbers of track jets that are not  $b$ -tagged.

- 3-tag: One Higgs candidate jet has 2  $b$ -tagged track jets while the other one has only 1  $b$ -tagged track jet.
- 2-tag: Each Higgs candidate jet has only 1  $b$ -tagged track jet.

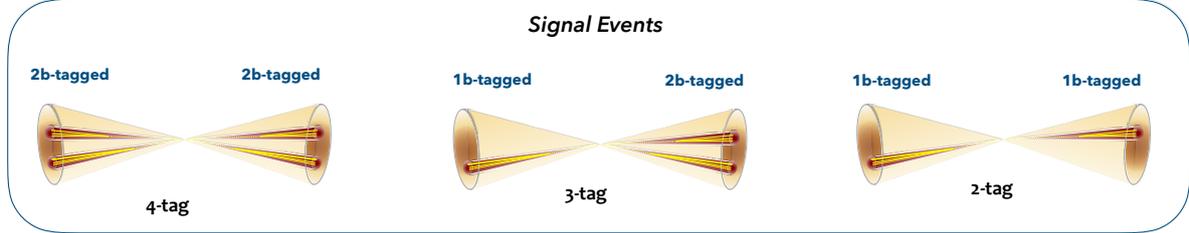


Figure 7.5: Signal event categories depending on the  $b$ -tagged track jet multiplicities for each Higgs candidate.

Obviously, not all the events falling in these track jet categories are signal events. As mentioned above, further selection on both the leading and the subleading Higgs candidate masses ( $m_J^{\text{lead}}$ ,  $m_J^{\text{subl}}$ ) is required in order to distinguish the di-Higgs events from background events. Therefore, the  $X_{hh}$  variable is introduced to define the final signal region of the boosted and resolved analyses constraining the two dimensional mass distribution plane for Higgs candidates.

$$X_{hh} = \sqrt{\left(\frac{m_J^{\text{lead}} - 124 \text{ GeV}}{0.1 (m_J^{\text{lead}})}\right)^2 + \left(\frac{m_J^{\text{subl}} - 115 \text{ GeV}}{0.1 (m_J^{\text{subl}})}\right)^2} < 1.6. \quad (7.1)$$

The denominator of each term in the definition can be interpreted as a resolution on the reconstructed mass, which is 10% for both large- $R$  jets. In this way, the  $X_{hh}$  can be considered as a  $\chi^2$  compatibility with the  $hh$  hypothesis. In order to decide the values used in the given equation, optimisation studies are performed using  $G_{KK}$  signal simulations. The central values of 124 GeV and 115 GeV correspond to the median values of the narrowest intervals that contain 90% of the simulated signal events [74]. The  $X_{hh}$  expression defines a circular boundary in the two dimensional jet mass plane and the requirement for the events to be in the signal region is then defined as  $X_{hh} < 1.6$ , as shown in Figure 7.7.

After applying the  $X_{hh}$  requirement individually to the 4 $b$ -tag, 3 $b$ -tag, and 2 $b$ -tag categories, the final signal regions can be obtained. Figure 7.6 shows the signal efficiency as a function of signal resonance mass. For higher resonance mass values, 2 $b$ -tag category becomes more efficient than the other categories that have higher  $b$ -tagged track jet multiplicities. The efficiency drop can be caused by the merging of track jets in boosted environments as well as the drop in  $b$ -tagging efficiency. Since, each  $b$ -tagged track jet requirement introduces additional  $\epsilon$  factor to the overall efficiency, where  $\epsilon$  is the  $p_T$  dependent efficiency of the tagging algorithm, the overall efficiency for 4 $b$ -tag category is roughly  $\epsilon^4$  explaining significantly higher contributions coming from 3 $b$ -tag and 2 $b$ -tag categories.

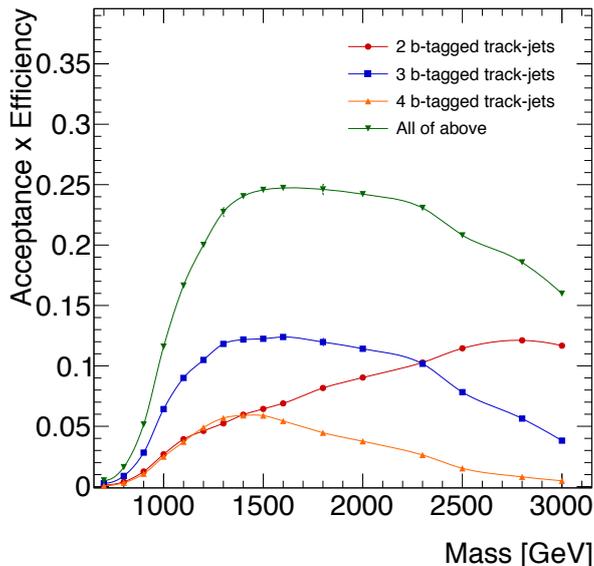


Figure 7.6: Signal efficiency in the three categories of SR as a function of signal  $G_{KK}$  resonance mass. The efficiencies are relative to the total number of events in the signal mass region and the green curve indicates that signal events passed all the preselection and the  $X_{hh}$  requirement[74].

### Sideband and Validation Regions

Similarly to the  $X_{hh}$  variable, additional variables,  $R_{hh}$  and  $R_{hh}^{\text{high}}$ , are introduced to define the sideband (SB) and the validation region (VR) in the two-dimensional mass plane. Both variables are circular and defined as below:

$$R_{hh} = \sqrt{(m_J^{\text{lead}} - 124 \text{ GeV})^2 + (m_J^{\text{subl}} - 115 \text{ GeV})^2} \quad (7.2)$$

$$R_{hh}^{\text{high}} = \sqrt{(m_J^{\text{lead}} - 134 \text{ GeV})^2 + (m_J^{\text{subl}} - 125 \text{ GeV})^2} \quad (7.3)$$

While the  $R_{hh}$  variable uses the same centre as  $X_{hh}$ ,  $R_{hh}^{\text{high}}$  has a central value shifted by 10 GeV. Using these variables, VR is defined as the region where  $R_{hh} < 33 \text{ GeV}$  and  $X_{hh} > 1.6$ , and SB is defined as  $33 \text{ GeV} < R_{hh}$  and  $R_{hh}^{\text{high}} < 58 \text{ GeV}$ . These cut values are optimised in order to have enough statistics to perform an estimation of systematic uncertainties and background modelling studies. Furthermore, an additional sideband region, the *extended sideband region*, is defined using the lower bound of SB and  $1 \text{ TeV} < R_{hh}$ .

While VR is chosen close to the SR to share similar features with jets, SB is chosen as shifted towards higher jet mass values to contain enough statistics of the  $t\bar{t}$  background expected around 175 GeV. Figure 7.7 shows the area defined by the VR and the SB selections inclusively before categorising the events according to the track jet multiplicities, using the signal simulation generated for  $G_{KK}$  mass 1500 GeV.

#### 7.2.4 Impact of jet mass choice

Previous ATLAS searches in this channel performed using the  $m^{\text{calo}}$  observable as the jet mass definition, as well as the preliminary studies performed for this analysis. Later

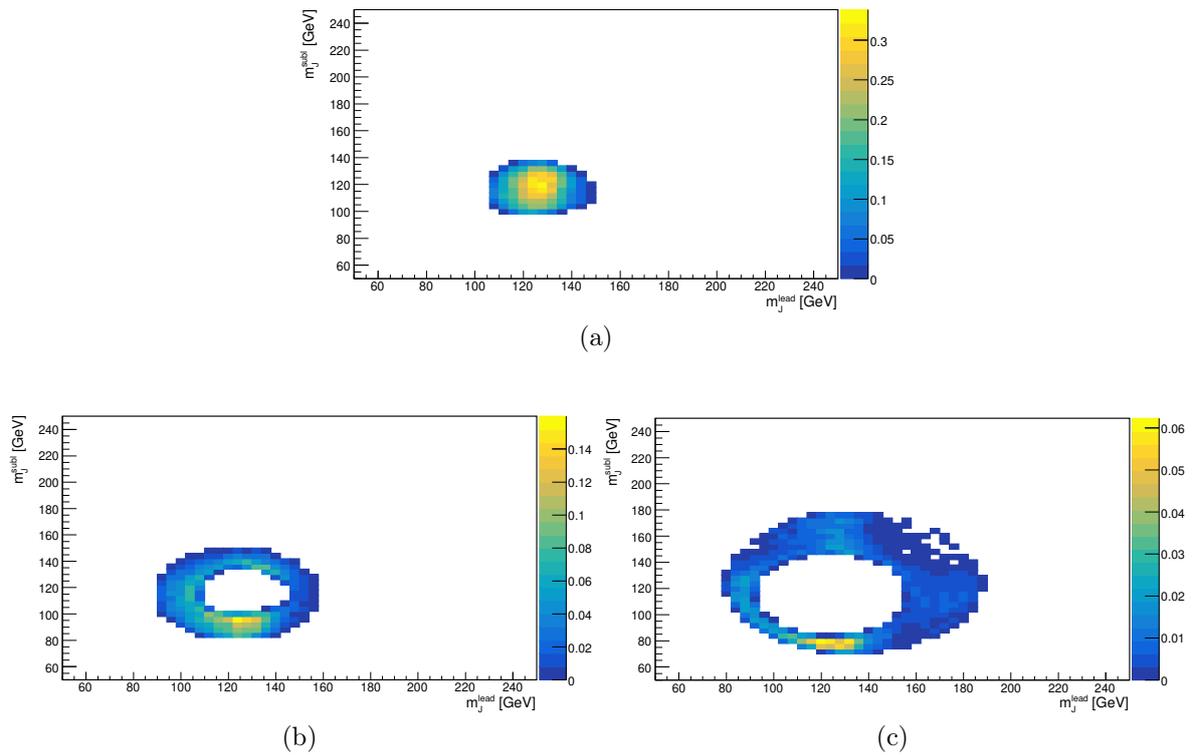


Figure 7.7: The number of events in the a) SR, b) VR and c) SB are shown in two dimensional mass plane for the leading and the subleading Higgs candidate jet using the  $G_{KK}$  signal sample with 1500 GeV mass. This plot shows inclusively all the events passing the preselection and  $X_{hh}$  requirement without dividing them into the subcategories according to the numbers of track jets.

when the  $m^{\text{comb}}$  observable was introduced as explained in Section 6.2.1, further studies were performed in order to assess the impact of the jet mass choice on the signal selection efficiency, the background rejection and the resolution of the Higgs candidate mass distributions.

Figure 7.8 compares the mass distributions of the leading Higgs candidate jet obtained from the events in  $3b$ -tag signal region using the simulated samples for  $G_{KK}$  with 5 TeV mass. Similar comparisons are performed in all categories using different signal mass points and they are given in Appendix B.1. For the resolution comparisons, two figures of merits are evaluated, being the standard deviation ( $\sigma$ ) obtained from Gaussian fits to each distribution, and the resolution definition using the inter-quantile range method as explained before in Chapter 6. It is observed that the resolutions of the  $m^{\text{comb}}$  distributions are better than the resolutions of  $m^{\text{calo}}$  distributions, both for the leading and the subleading Higgs candidates.

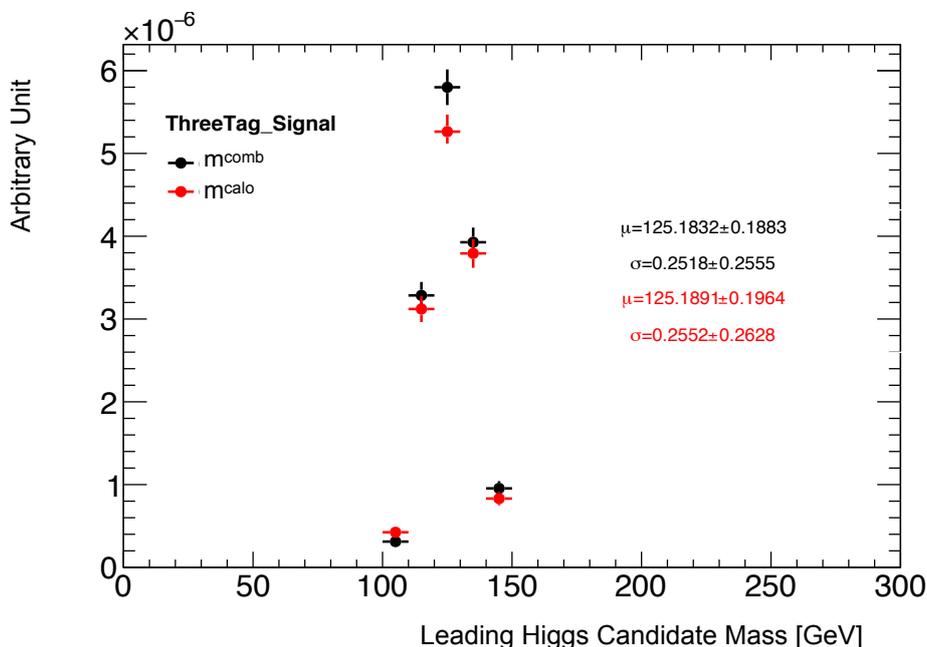


Figure 7.8:  $m^{\text{comb}}$  and  $m^{\text{calo}}$  distributions of the leading Higgs jet in  $3b$ -tag SR using a  $G_{KK}$  signal at 5 TeV mass. The given mean ( $\mu$ ) and sigma ( $\sigma$ ) values on the figure are obtained from the Gaussian fits performed for each distribution. Another resolution comparison is done using the inter-quantile definition,  $\frac{68\%IQnR}{2 \times \text{median}}$  and it indicates 2% improvement with the usage of  $m^{\text{comb}}$ .

In addition to the resolution checks, total yields in SR are compared for signal and background events using different  $G_{KK}$  signal mass points and dijet simulations. In these studies, it has been shown that for all signal mass points, the selection using the  $m^{\text{comb}}$  provides higher yields than the selection using the  $m^{\text{calo}}$ . The amount of the increase varies depending on the signal mass points as shown in Table 7.8. As discussed in Section 6.2.1, the gain is more pronounced for higher graviton masses which leads to higher boosts in the Higgs bosons. At the same time, the SR yields obtained from multijet simulations (as given in Chapter 4) shows that using the  $m^{\text{comb}}$  observable results in  $\sim 7\%$  less background events compared to the SR selection with the  $m^{\text{calo}}$ .

Signal Yields for $G_{KK}^* \rightarrow hh$					
$G_{KK}^*$ mass	6 TeV	5 TeV	4 TeV	3 TeV	2 TeV
Combined mass	13010	12358	15010	16834	25877
Calorimeter mass	11579	10757	13352	15109	24414
Yield increase	12 %	14 %	12 %	11 %	5 %

Table 7.1: Total event yields in SR obtained from  $G_{KK}$  signal samples generated at different mass values for both  $m^{\text{comb}}$  and  $m^{\text{calo}}$ .

In conclusion, the  $m^{\text{comb}}$  observable improves the signal to background ratio, and thus provides better sensitivity for this search. Therefore, it is used as the jet mass definition of this analysis.

## 7.3 Background modelling

Three processes contribute as background to the selected event samples in the signal region of this analysis: multijets (QCD),  $t\bar{t}$ , and  $Z + \text{jets}$ . In each subcategory of the signal region that is defined according to the  $b$ -tagged track jet multiplicities, the relative contribution of the backgrounds are different. However in all categories, the multijet background is dominant, and the contribution of the  $Z + \text{jets}$  is insignificant below 1%. Fractions of the backgrounds with respect to the total background in each category are given in the following. In the 4-tag signal region, QCD  $\sim 95\%$ ,  $t\bar{t} \sim 5\%$ . In the 3-tag signal region, QCD  $\sim 90\%$ ,  $t\bar{t} \sim 10\%$ , and in the 2-tag signal region, QCD  $\sim 80\%$ ,  $t\bar{t} \sim 20\%$ .

While the  $Z + \text{jets}$  background is obtained from MC simulations (Section 4.2.1) and found negligible, the estimation of the other backgrounds requires data driven techniques in order to be modelled well. For the  $t\bar{t}$  background, MC simulations are used to estimate the shape of the background contribution, while the normalisation is obtained from the data. On the other hand, for the multijet background, neither shape nor normalisation estimations can be obtained from the simulations, since the statistics in simulations are very limited compared to the huge amount of background events expected in the analysed phase space (recall Figure 6.1). Therefore, more sophisticated techniques are required to estimate the multijet background to be able to sensitive for a possible excess in the  $m_{HH}$  spectrum. In order to develop such techniques, better understanding of the multijet background is obligatory and crucial, and the following parts of this section explore that in detail. In the light of the presented findings about the background, the developed data driven method, *kinematic reweighting*, for the background estimation is introduced (Section 7.3.1). And finally, the yields of the both backgrounds, the  $t\bar{t}$  and multijets, are estimated using a second data driven method that is explained briefly in Section 7.3.2.

### 7.3.1 Data driven multijet background estimation

The idea is to estimate the dominant multijet background from the data, since the statistics in the MC is not comparable with the number of events expected to be in a dataset of  $36 \text{ fb}^{-1}$ . The normalisation of the simulated MC samples would have to be scaled up significantly, leading to large uncertainties and loss of the sensitivity of the analysis. This is why using the available simulations is not an option for this analysis, and data driven approaches are needed for the multijet background estimation.

Typically, to estimate a background from data, the initial step would be to create a *control region* in data, where a negligible signal contamination is expected. A *good control region* means that the relevant observables, particularly the final discriminant, behaves similarly to the signal region, so that the information obtained from the control region can also be used in the signal region.

In this analysis, the contribution of the multijet background has to be estimated in the three signal regions, 4-tag, 3-tag, and 2-tag. The below samples are defined as control regions (CR) using different number of  $b$ -tagged track jets distributed to each Higgs candidates, as depicted in Figure 7.9.

- 2-tagCR: One large- $R$  jet has 2  $b$ -tagged track jets, while the other one does not have any  $b$ -tagged track jets.
- 1-tagCR or 1-tag: Only 1  $b$ -tagged track jet in one of the large- $R$  jet.
- 0-tagCR or 0-tag: There is no  $b$ -tagged track jet in any of the large- $R$  jets.

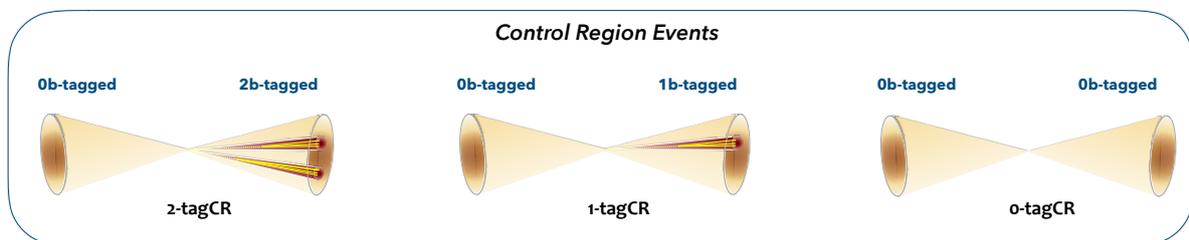


Figure 7.9: Control region event categories depending on the  $b$ -tagged track jet multiplicities for each Higgs candidate jet.

In order to study the possibility to use these CRs as *good control region*, the normalised distributions of the most important kinematic variables are compared using the data obtained from the validation region and the extended sideband region, which was introduced earlier in Section 7.2.3. In Figure 7.10, the comparison of the  $m_{HH}$  distributions in the VR is shown between the 2-tag region and the three CRs defined above. It can be seen that the shapes of the distributions are not identical between the 2-tag signal events and the CR samples, similarly for the other observables.

Further comparisons are performed using the extended sideband region. Figure 7.11 shows the normalised  $p_T$  distributions of Higgs candidates in all categories. Due to the limited statistics comparisons are performed using large bins. Still, the results clearly indicate that the shape of the  $p_T$  distributions of Higgs candidates significantly differs in the categories with higher  $b$ -tagging multiplicities. Note that these comparisons are

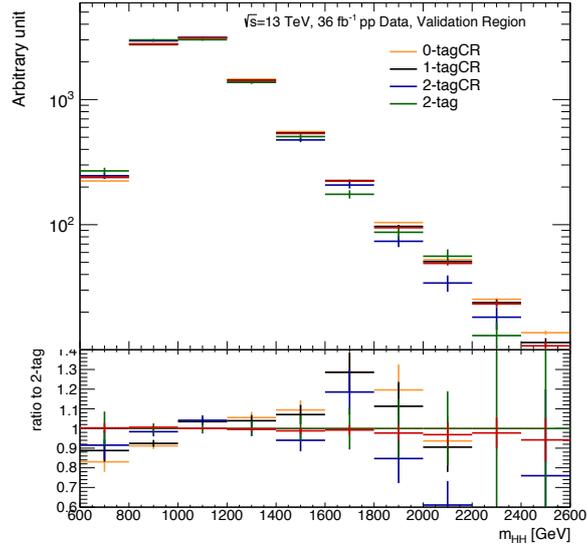


Figure 7.10: Shape comparison for the  $m_{HH}$  observable in different  $b$ -tagged event categories obtained from the extended sideband data. All distributions are normalised according to the  $2b$ -tag distribution (each large- $R$  jet with  $1b$ -tagged track jet) since the statistics of the higher  $b$ -tagged multiplicity events are dramatically smaller than the control region samples in the VR. Ratio is obtained from the normalised distributions.

performed in the extended sideband region data. Even if the dominant portion of these events comes from the multijet background, there is also non-negligible amount of  $t\bar{t}$  contribution. Therefore, more detailed studies of the impact of the  $b$ -tagging on the kinematic distributions are performed using QCD MC simulations to understand the reason of these modelling differences.

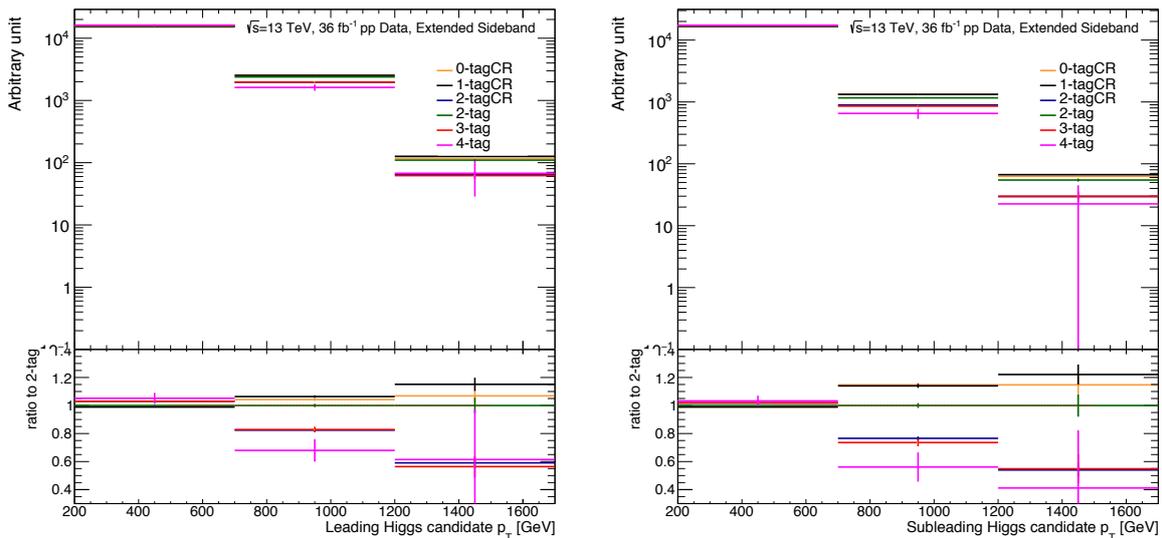


Figure 7.11: Shape comparison for the  $p_T$  of the leading and the subleading Higgs candidates in different  $b$ -tagged event categories obtained from the extended sideband data. All distributions are normalised according to the  $4$ -tag distribution.

### Impact of $b$ -tagging on kinematic observables

By comparing the relevant distributions both in the MC and the data, the aim is to understand if the shape discrepancies observed in the data can also be observed in the MC or not. If such an effect exists in the MC simulations, and the impact is similar with the observed trend in the data, it can support the argument that  $b$ -tagging sculpts the distributions.

Six variables have been studied separately, two of them are the  $p_T$  distributions of the Higgs candidates, and the other four are  $p_T$  of the first two leading track jets of each Higgs candidates. Distributions are compared in the 0-tag, 1-tag and 2-tag categories. Note that there is no requirement on the minimum number of associated track jets to the large- $R$  jets in the event selection. Thus, the statistics of the leading and subleading track jets can be different than the total number of events expected in their category. However, it does not have an impact on the studies directly, since the comparisons are performed for the normalised distributions.

In order to have a fair understanding between the different topologies of the events, 1-tagCR is further divided into two orthogonal sets of events according to the association of the  $b$ -tagged track jet. One of the group is composed of the events where the  $b$ -tagged track jet is associated to the leading Higgs candidate (1-tag on Lead), and in the other one it is associated to the subleading Higgs candidate (1-tag on Subl), while the 0-tag and the 2-tag regions are used as they are. As a result of the performed comparisons using the extended SB data and the MC, in both of them similar results are obtained as shown for the leading track jets on the leading Higgs candidate in Figure 7.12 and subleading Higgs candidates in Figure 7.13. Note that for the simplicity 1-tagCR and 0-tagCR are referred as 0-tag and 1-tag.

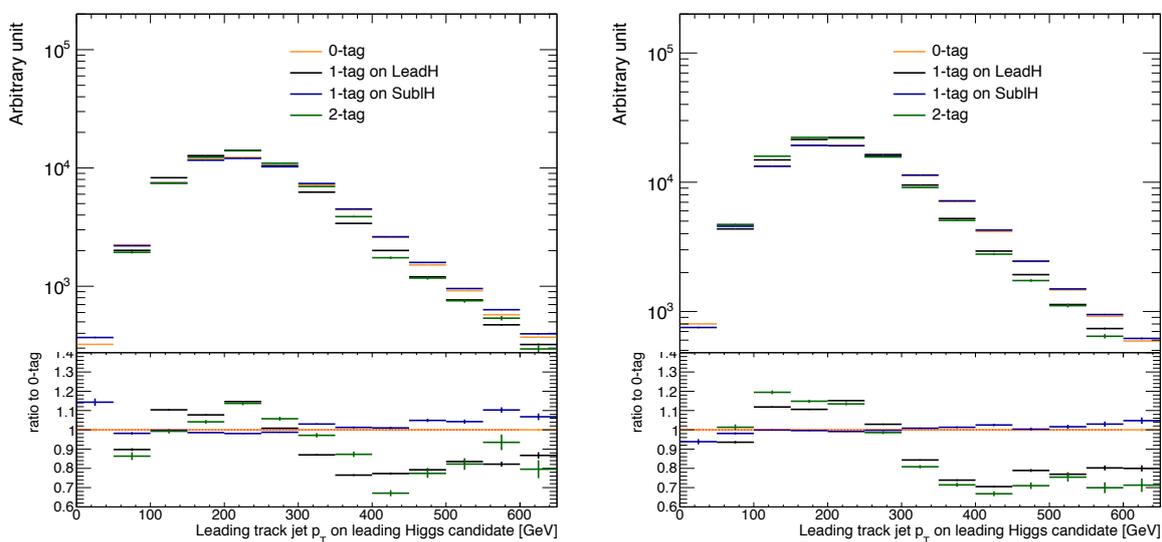


Figure 7.12:  $p_T$  distributions of the leading track jet on the leading Higgs candidate are shown for different event categorisation obtained from the extended SB in the data (right) and in the simulations (left). All distributions are normalised according to the number of events in the 2-tag region.

As can be seen from the  $p_T$  distributions of leading track jet on the leading Higgs candidate in Figure 7.12, both in data and MC, *0-tag* and *1-tag on Subl* categories behave similar to each other forming one group, while *1-tag on Lead* and *2-tag* categories form another group. Considering that the given distribution is for the leading track jet on

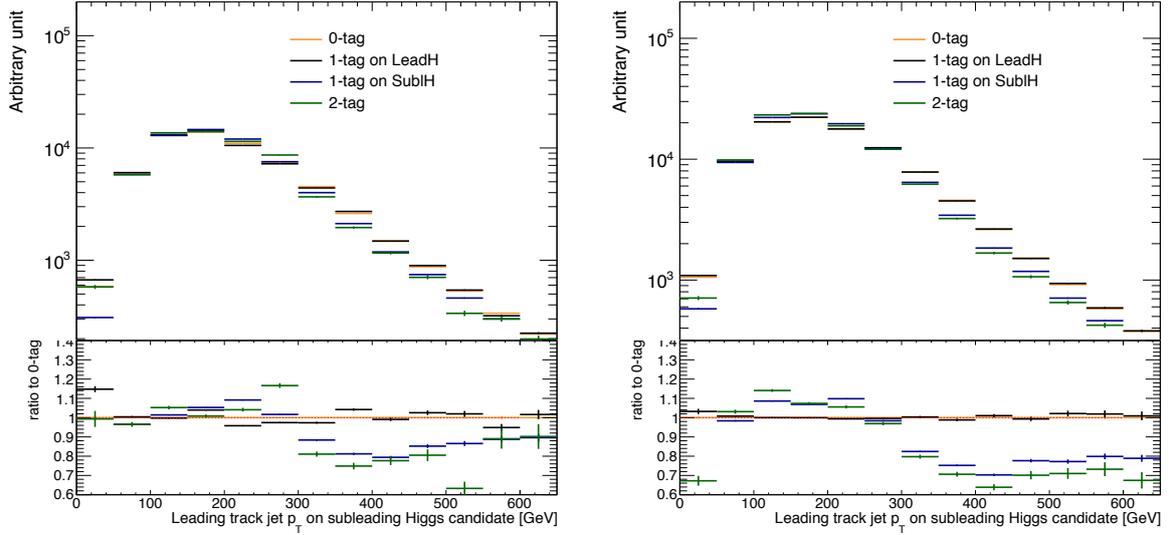


Figure 7.13:  $p_T$  distributions of the leading track jet on the subleading Higgs candidate are shown for different event categorisation obtained from the extended SB in the data (right) and in the simulations (left). All distributions are normalised according to the number of events in the 2-tag region.

leading Higgs candidate, the track jets shown for  $0\text{-tag}$  and  $1\text{-tag on Subl}$  are not  $b$ -tagged by construction. On the other hand, for the other two categories, the situation is different. Significant portion of the distribution of  $1\text{-tag on Lead}$  and  $2\text{-tag}$  composed of the  $b$ -tagged track jets, since by construction in both categories there must be one  $b$ -tagged track jet in the leading Higgs candidate. The exact same reasoning also applies to the  $p_T$  distributions of the leading track jet on the subleading Higgs candidate shown in Figure 7.13. While more details on the topic is given in Appendix B.1, the results can be summarised as follows:

- For the track jet comparisons on the **leading** Higgs candidate:  $2\text{-tag}$  and  $1\text{-tag on Lead}$  is one group,  $0\text{-tag}$  and  $1\text{-tag on Subl}$  is another group.
- For the track jet comparisons on the **subleading** Higgs candidate:  $2\text{-tag}$  and  $1\text{-tag on Subl}$  is one group,  $0\text{-tag}$  and  $1\text{-tag on Lead}$  is another group.
- Both the leading and the subleading track jet distributions follow similar trends if they are associated to the same Higgs candidate.
- In all four track jet  $p_T$  distributions, it is observed that the shape of the distributions changes around  $\sim 300$  GeV, if the shown track jet has the possibility to be  $b$ -tagged.

These results indicate that the kinematic properties of the  $b$ -tagged track jets are different than the untagged track jets. In order to prove this statement conclusively,  $b$ -tagged track jets are selected from the inclusive track jet distributions creating two sub-categories,  $b$ -tagged track jets and un-tagged track jets and further comparisons are performed. These results are shown in Figure 7.14 and Figure 7.15.

While Figure 7.14 indicates that  $b$ -tagged track jet distribution follows the same shape as expected with the track jets that have possibility to be  $b$ -tagged, the other comparison in Figure 7.15 reveals that the un-tagged track jets do not follow exactly a similar trend with the other un-tagged track jet distribution in  $0\text{-tag}$  category, implying

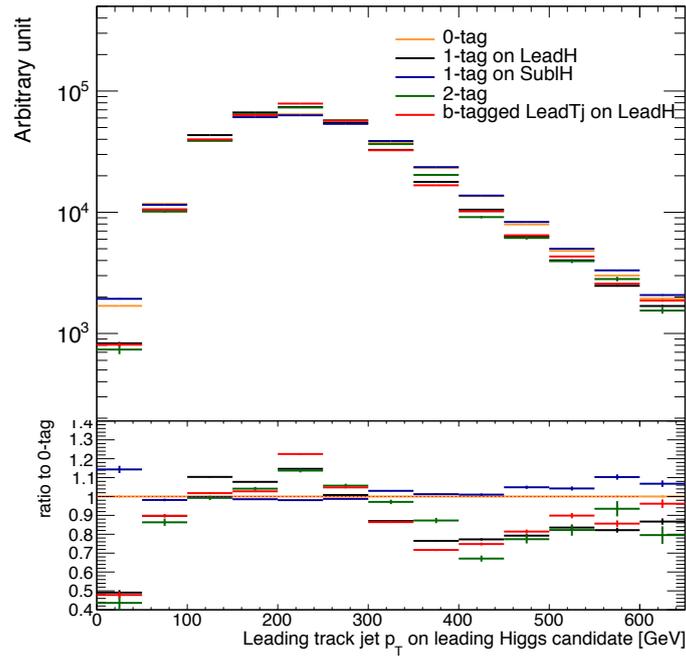


Figure 7.14:  $p_T$  distributions of the leading track jet on the leading Higgs candidate are shown in different categories using multijet MC simulations. Plot shows that the kinematic properties of the  $b$ -tagged leading track jets on the leading Higgs candidate ( $b$ -tagged  $LeadTj$  on  $LeadH$ ) is similar to the  $2$ -tag and the  $1$ -tag on  $LeadH$  categories as expected.

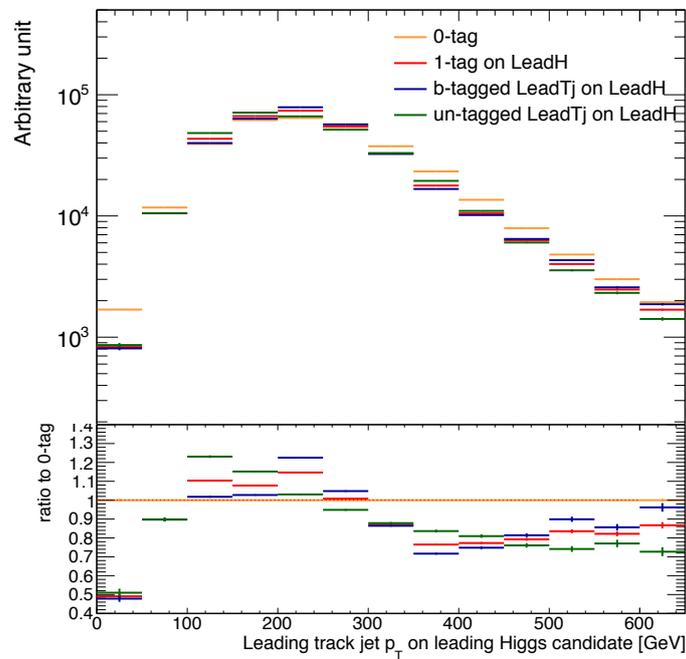


Figure 7.15:  $p_T$  distributions of the leading track jet on the leading Higgs candidate are shown in different categories using multijet MC simulations. Plot shows the comparison of the two orthogonal subsets of the  $1$ -tag on  $LeadH$ ,  $b$ -tagged and un-tagged leading track jets on the leading Higgs candidate. All distributions are normalised with respect to the sub-category  $b$ -tagged  $LeadTj$  on  $LeadH$ .

that not only the  $b$ -tagged track jet kinematics are different, but also the kinematics of the entire large- $R$  jet is different when one track jet is  $b$ -tagged, including the un-tagged track jet.

### Kinematic reweighting

Using the similarities summarised above, it is possible to construct good CRs for the multijet background estimation in the three SR. For simplicity, the method is explained for the  $2$ -tag category, and then it is generalised to the other categories. As shown above the kinematic properties of the track jets associated to the leading Higgs candidate in  $2$ -tag category is similar to the  $1$ -tag on *Lead* category, therefore the shape of these variables in  $2$ -tag category can be modelled from the  $1$ -tag on *Lead* category. Similarly, for the track jets in the subleading Higgs candidate,  $1$ -tag on *Subl* category can be used. The idea is reweighting the relevant kinematic variables of the non-tagged Higgs candidate to the variables of the tagged Higgs candidate in each category, creating similar events expected in the signal region and combining them as depicted in Figure 7.16. Since changing the  $p_T$  of the track jet in the large- $R$  jet has a direct impact on the  $p_T$  of the corresponding large- $R$  jet and the other associated track jet, reweighting uses simultaneously three  $p_T$  variables for each Higgs candidate: two track jet  $p_T$  and the large- $R$  jet  $p_T$ .

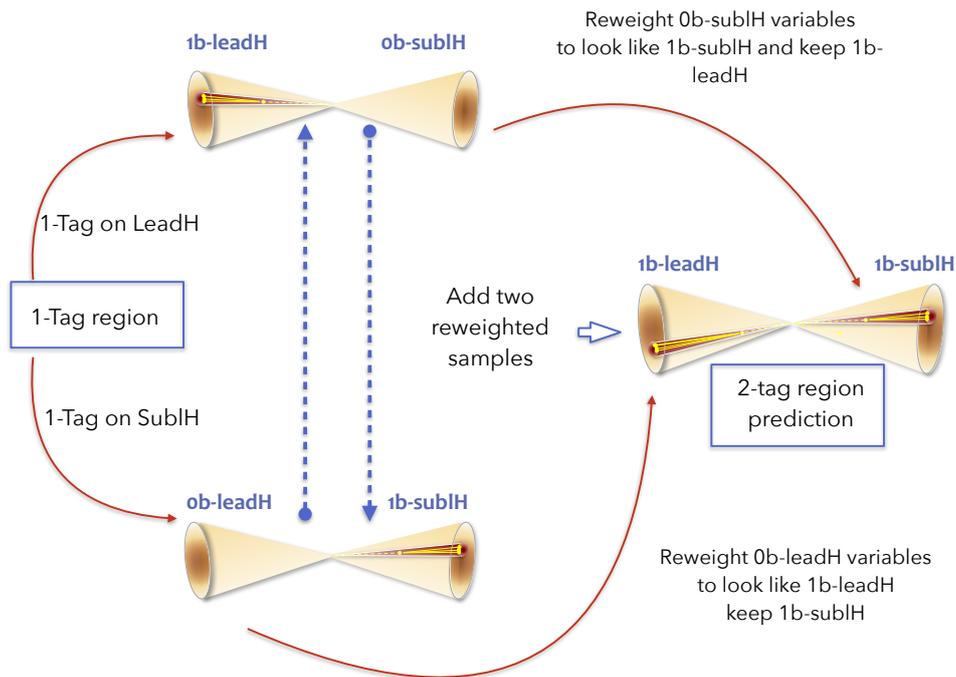


Figure 7.16: Sketch of the multijet background estimation for the 2-tag category using the subcategories of 1-tag CR.

Similarly to the 2-tag category, the method is generalised and used for the 3 and 4-tag categories. For each category, CR categories with fewer  $b$ -tags (lower tagged) are used by separating to two orthogonal regions depending on the track jet distributions per leading and the subleading Higgs candidates, as explained below:

- Background estimation in 2-tag:  $1$ -tag CR is used with its subcategories  $1$ -tag on *Subl* and  $1$ -tag on *Lead*, as described above in detail.

- Background estimation in 3-tag:  $2\text{-tagCR}$  and  $1\text{-tagCR}$  are used depending on the multiplicity distributions in the leading and subleading Higgs candidates. The Higgs candidate with 0  $b$ -tagged track jet in  $2\text{-tagCR}$  is reweighted to look like 1-tagCR sample.
- Background estimation in 4-tag:  $2\text{-tagCR}$  is used with its two subcategories, where 2  $b$ -tagged track jet is on the leading or on the subleading Higgs candidate.

The weights are derived iteratively from the ratios of the subsets of the lower-tagged data samples. These samples are used inclusively, meaning that there is no selection on the two dimensional jet mass plane. However, the  $t\bar{t}$  and the  $Z + \text{jets}$  samples are subtracted from the data before deriving the weights. The ratios are obtained simultaneously from the three  $p_T$  distributions and then fitted with a spline function. Three weights are multiplied together and used as event weights. The iterative reweighting procedure continues until the weights stabilise. Figure 7.17 shows the reweighting procedure for 2-tag category. More details can be seen in Reference [74].

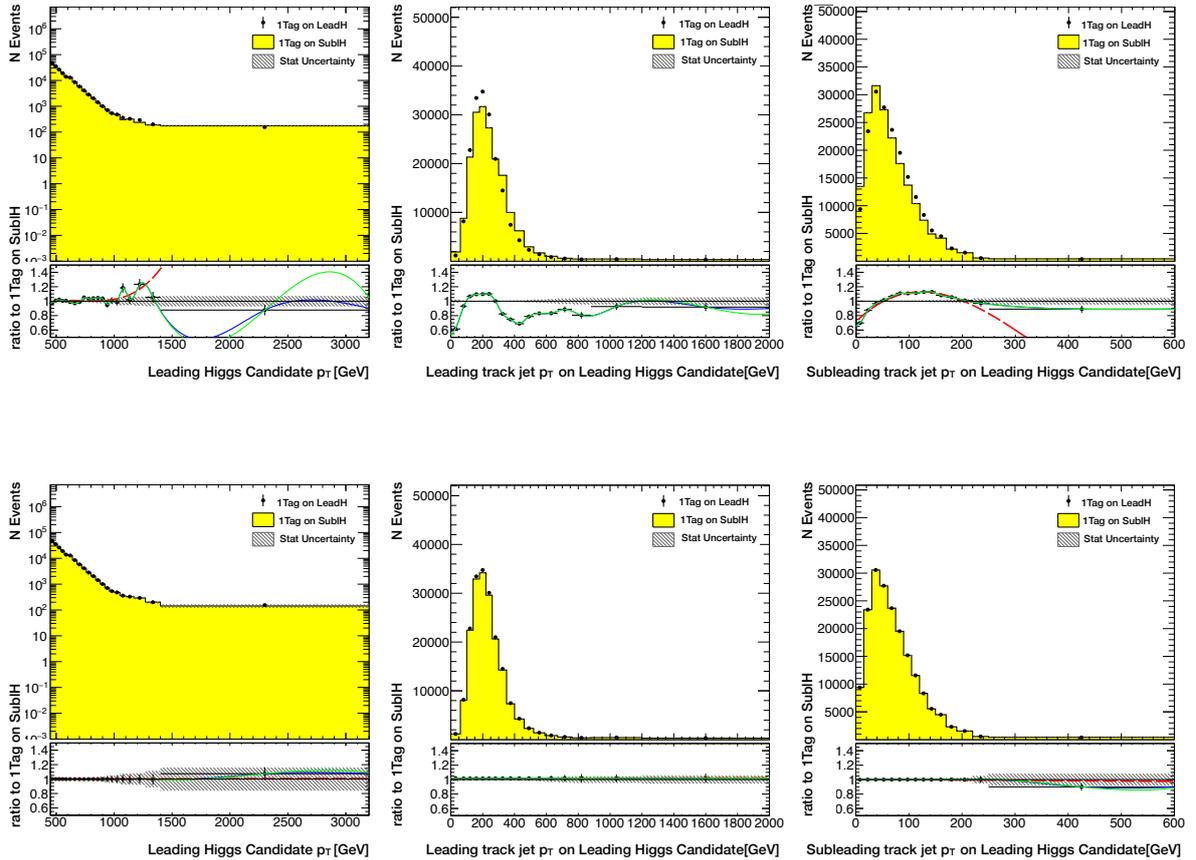


Figure 7.17: First and last iterations of the 2-tag background estimate is shown for the leading Higgs candidate observables. The performed spline fits to the ratios of the leading Higgs candidate observables in  $1\text{-tag on LeadH}$  (black point), over the  $1\text{-tag on Subl}$  (yellow) are shown. Upper plots show the distributions and the fit function before reweighting and the bottom plots show them after the last, tenth iteration. The green line is the spline extrapolation; and the red line is a polynomial fit [74].

### Background composition and track jet $b$ -tagging

In order to have a better understanding of the differences seen in the kinematic properties of the  $b$ -tagged track jets, additional truth analyses are performed with multijet simulations. As shown above in Section 7.3.1, both in the data and in the MC, the kinematic properties of the Higgs candidates with  $b$ -tagged track jets are different than the un-tagged ones. The physical processes with a  $b$ -quark involvement represents a different topology than the other QCD processes with the involvement of only light quarks or gluons. Especially for the events with a presence of large- $R$  jet and two associated  $b$ -tagged track jets, the process of gluon splitting to a  $b\bar{b}$  pair is a significant background, and has different features than the  $g \rightarrow qq$ , as well as the  $g \rightarrow bq$  processes [65, 153]. On the other hand, the process of  $b$ -tagging track jets using the MV2c10 algorithm (Section 5.1.6) may have an impact on the  $p_T$  modelling of the track jet distributions.

It is possible to use the truth information obtained from the simulations to reveal whether the observed differences in the track jet  $p_T$  distributions in this analysis are actually representing the different physics of  $b$ -hadrons, or the dependence of the  $b$ -tagging performance on the large- $R$  jet topology. Below the steps of the truth analysis are outlined:

- Each  $b$ -tagged track jet is checked using the geometrical matching (see Section 6.1.2) with  $\Delta R = 0.2$  whether they are matched to at least 1  $b$ -hadron.
- If there is at least 1  $b$ -tagged track jet that does not match a  $b$ -hadron in the event, the event is labelled as a *fake* event. The remaining events are labelled as *real* events.
- Two orthogonal subsets are created in each categories.

The comparisons shown in Figure 7.12 are repeated with the fake and the real event categories separately. The results revealed that the trends seen earlier are similar to the results obtained from the fake events category, more than the real events as shown in Figure 7.18. This indicates that the shape difference is not caused by the different physics of the track jets initiated by  $b$ -hadrons, but it is an artefact of the performance of the  $b$ -tagging algorithm. Both the  $b$ -tagging and the mis-tagging efficiency has an impact on the modelling of the track jets.

Performing the truth studies also showed that in each category, the number of fake events coming from the mis-tagged track jets are dominant than the contribution of the real events with the truth matched track jets. Figure 7.19 shows the fractions of truth matched  $b$ -tagged track jets in the 2-tag extended sideband region. While the fraction of the truth matched  $b$ -tagged track jets fluctuates between 30 to 50% for the  $b$ -tagged track jets in the leading Higgs candidate, it is between 30 to 40% for the  $b$ -tagged track jets in the subleading Higgs candidate.

Considering that the amount of the mis-tagged events depends on both the performance of the  $b$ -tagging algorithm and the purity of the multijet samples, this result is expected. Even if the mis-tagging rate is not so high, the purity of the multijet sample in terms of  $b$ -hadrons is low enough to result in larger amount of fake event contributions. Checking the fractions of the truth  $b$ -tagged track jets and mis-tagged track jets,  $p_T$  dependencies of the  $b$ -tagging algorithm in the analysis topology can be obtained for the  $b$ -tagged track jets. For more detailed discussion, Appendix B.1 can be seen.

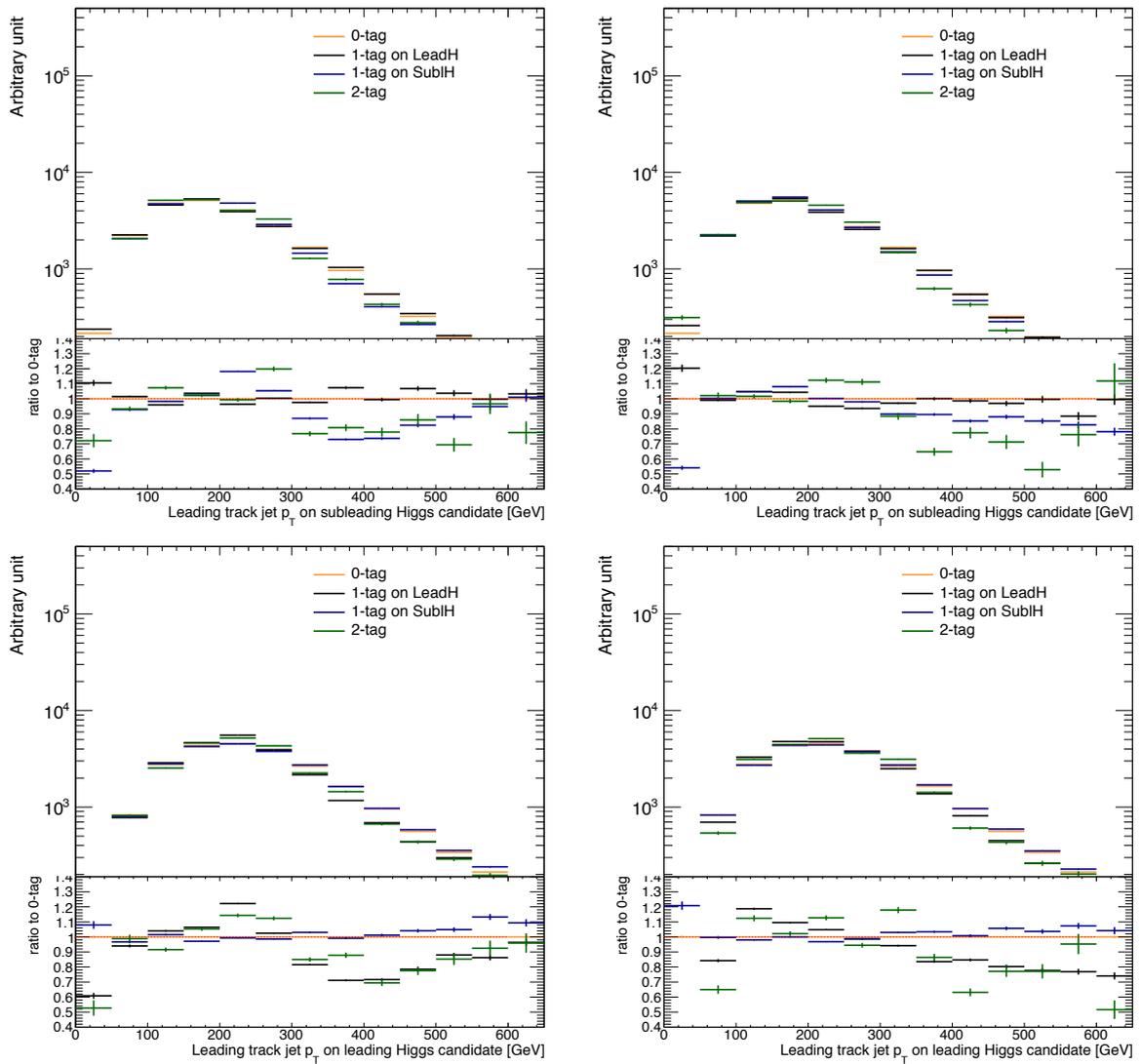


Figure 7.18: Impact of  $b$ -tagging in different event categories for the fake and real events as explained in the text. On the left fake events, on the right real  $b$ -tagged events.

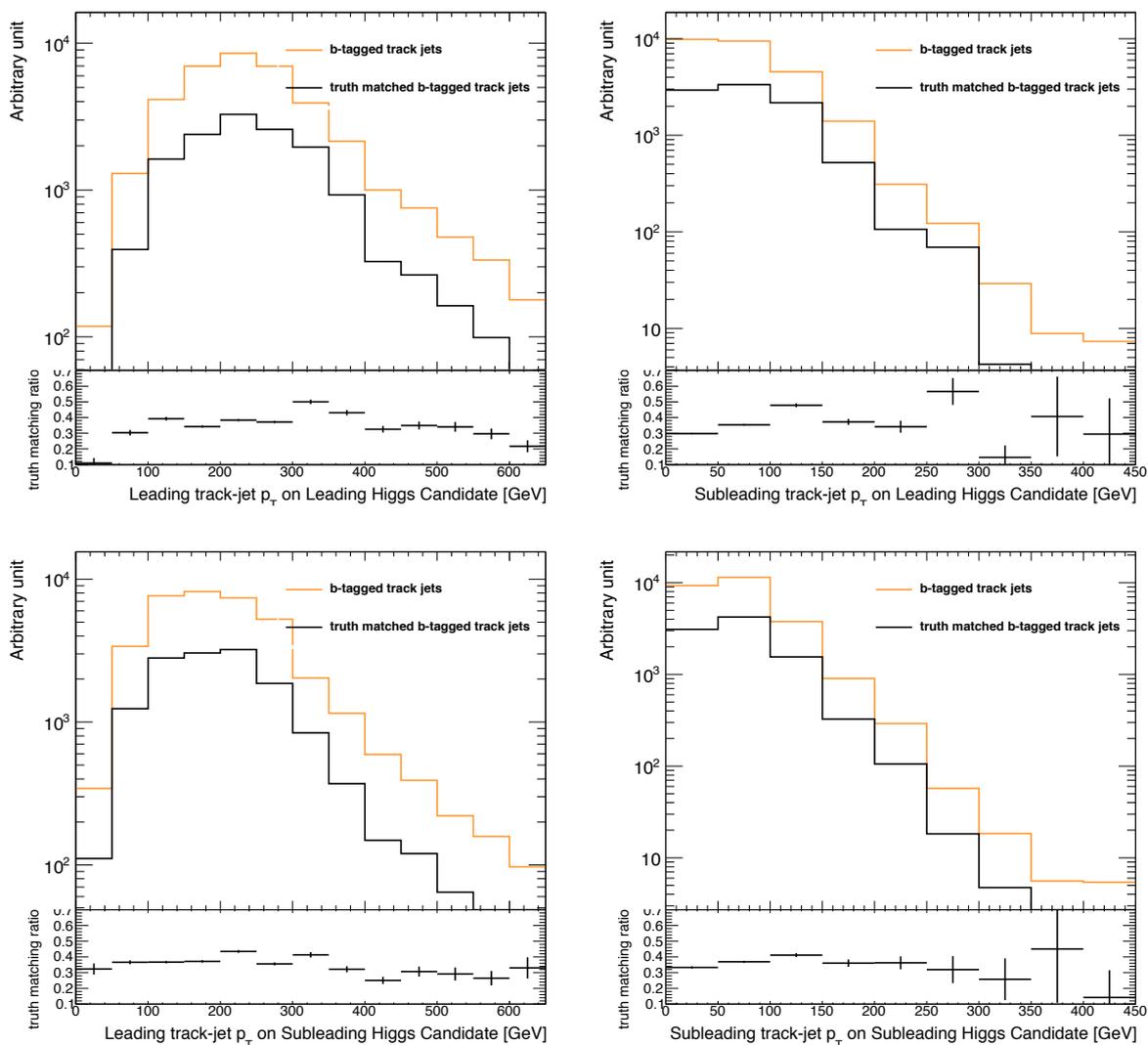


Figure 7.19:  $b$ -tagged track jets (yellow) together with their truth matched subsets (black) are shown with respect to track jet  $p_T$ . Different plots show the distributions obtained from the two leading track jets in each Higgs candidate in extended sideband region of 2-tag category. Truth matching ratio is defined as the ratio of truth matched  $b$ -tagged track jets to the total  $b$ -tagged track jets in each  $p_T$  bin.

### 7.3.2 Data driven normalisation of $t\bar{t}$ and multijet backgrounds

The background yield in the 4-,3-, and 2-tag event categories ( $N_{\text{background}}^{n\text{-tag}}$ ) can be expressed as:

$$N_{\text{background}}^{n\text{-tag}} = \mu_{\text{multijet}}^{n\text{-tag}} N_{\text{multijet}}^{x\text{-tag}} + \alpha_{t\bar{t}}^{n\text{-tag}} N_{t\bar{t}}^{n\text{-tag}} \quad (7.4)$$

where  $x$  is 1 for the 2-tag category, and 2 both for the 3-tag and the 4-tag categories.  $\mu_{\text{multijet}}$  corresponds to the ratio of the number of multijet events in the  $n$ -tag category to the  $x$ -tag category.  $N_{\text{multijet}}^{x\text{-tag}}$  is obtained from the corresponding data, after subtracting the  $t\bar{t}$  contributions in the  $x$ -tag category.  $N_{t\bar{t}}^{n\text{-tag}}$  is the number of  $t\bar{t}$  events obtained from the MC simulations and scaled to the total integrated luminosity.  $\alpha_{t\bar{t}}$  is the parameter introduced as a correction factor for  $N_{t\bar{t}}^{n\text{-tag}}$ . In order to obtain the values of  $\mu_{\text{multijet}}$  and  $\alpha_{t\bar{t}}$ , binned likelihood fits are performed to the leading large- $R$  jet mass distributions in each category using the sideband region data. Extracted values of the parameters are given in Table 7.2. More details on the fit studies are found in the given references [66, 74].

Table 7.2: Background scaling parameters estimated from the fits to the leading large- $R$  jet mass distributions in the 4-, 3-, and 2-tag sideband regions [66].

Category	$\mu_{\text{multijet}}$	$\alpha_{t\bar{t}}$
2-tag	$0.06273 \pm 0.00057$	$0.986 \pm 0.019$
3-tag	$0.1626 \pm 0.0043$	$0.800 \pm 0.073$
4-tag	$0.0332 \pm 0.0043$	$0.89 \pm 0.60$

In order to mitigate the statistical fluctuations in the tails of the  $m_{HH}$  distribution in the 3- and 4-tag categories, a smoothing procedure is applied in the range of  $1.2 \text{ TeV} < m_{HH} < 3.0 \text{ TeV}$  [66]. Lower tagged data regions are fitted with the following function:

$$y(x) = \frac{a}{\left(\frac{x}{\sqrt{s}}\right)^2} \left(1 - \frac{x}{\sqrt{s}}\right)^{b-c \log\left(\frac{x}{\sqrt{s}}\right)}, \quad (7.5)$$

where  $a$ ,  $b$ ,  $c$ , and  $s$  are free parameters in the fit. Due to the very limited number of  $t\bar{t}$  events in the 3- and 4-tag regions, the shape of the  $t\bar{t}$  distribution in the 2-tag region is used for the modelling of these regions.

## 7.4 Systematic Uncertainties

This section briefly points out the important systematic uncertainty sources for this analysis and presents their impact on the background and signal event yields in the three SRs as shown in Table 7.3.

The theoretical uncertainties that have an impact on the signal acceptance, such as the PDF set choice, the modelling differences in MC generators, and the uncertainties on the renormalisation and factorisation scales, are considered in the statistical analysis as normalisation factors. Besides the theoretical uncertainties, the detector and reconstruction related uncertainties are estimated and found to be the most significant uncertainties for the signal processes. Regarding the detector and reconstruction modelling uncertainties, large- $R$  jet related uncertainties (as being the only objects of this analysis) and

Source	Two-tag			Three-tag			Four-tag		
	Background	$G_{KK}$	Scalar	Background	$G_{KK}$	Scalar	Background	$G_{KK}$	Scalar
Luminosity	-	2.1	2.1	-	2.1	2.1	-	2.1	2.1
JER	0.25	0.74	1	1.4	0.93	0.93	0.45	1.1	1.5
JMR	0.52	12	12	1.4	12	13	7.9	13	14
JES/JMS	0.43	1.7	2.1	2.0	1.9	2.2	1.3	3.7	5.7
$b$ -tagging	0.83	27	29	0.48	2	2.9	1.1	28	28
Bkgd estimate	2.8	-	-	5.8	-	-	16	-	-
Statistical	0.6	1.2	1.3	1.3	1.0	1.1	3.1	1.6	1.9
Total Syst	3.1	30	32	6.6	13	14	18	31	32

Table 7.3: Summary of systematic uncertainties (expressed in percentage) on the total background and signal event yields in the signal region of the boosted analysis. Uncertainties are provided for each of the three samples for the background, a 2 TeV scalar, and a  $G_{KK}$  with  $k/\bar{M}_{P1} = 1$  and  $m = 2.0$  TeV [66].

the uncertainties caused by the  $b$ -tagging algorithm are evaluated. Both uncertainties influence the shape and the normalisation of the signal and the  $t\bar{t}$  background. Since the multijet background is estimated with a data-driven method, these uncertainties have a small impact on the background estimation.

However, the background estimation method has its own uncertainties. As given in Table 7.2, one uncertainty is the statistical uncertainty on the fit parameters extracted for the normalisation of the  $t\bar{t}$  and multijet backgrounds. Moreover, the choice of the SB and the CR influences  $\mu_{\text{multijet}}$  significantly as the multijet background modelling is based on these regions, which affects the normalisation fit in Section 7.3.2. In order to assign an uncertainty on the region choices, alternative regions are defined and their yields are compared in each category. The largest differences between the data and the expected background is assigned as normalisation uncertainty [26].

In order to assess the uncertainty on the data driven multijet background estimation, the smoothed CR data and the smoothed background predictions are compared in each category. The difference between the shapes is assigned as systematic uncertainty on the background modelling [66].

## 7.5 Results

In this section, the results of the boosted analysis are briefly presented. The validations performed in the SB and in the CR resulted in good agreement in the yields as well as the kinematic distributions obtained from the data and the background prediction [74]. Hence, the number of background events and the yields for two benchmark signals are compared with the data in three SRs as given in Table 7.4. The results are in an agreement within the statistical uncertainties.

The  $m_{HH}$  distributions are compared to data in the three SRs after considering all statistical and systematic uncertainties. Figure 7.20 shows the results together with the signal predictions. No significant deviation from the SM prediction is observed.

Following the statistical analysis in Reference [14], exclusion limits are produced for the graviton and the narrow scalar resonance signals using only the combined boosted analysis results. They are shown in Figure 7.21.

While the 4-tag region is most sensitive results between the resonance masses of 1200 GeV-1800 GeV, the 3-tag and 2-tag regions are most sensitive for the resonance masses

	Two-tag		Three-tag		Four-tag	
Multijet	3390	$\pm 150$	702	$\pm 63$	32.9	$\pm 6.9$
$t\bar{t}$	860	$\pm 110$	80	$\pm 33$	1.7	$\pm 1.4$
Total	4250	$\pm 130$	782	$\pm 51$	34.6	$\pm 6.1$
$G_{KK}$ (2 TeV)	$0.97 \pm 0.29$		$1.23 \pm 0.16$		$0.40 \pm 0.13$	
Scalar (2 TeV)	$28.2 \pm 9.0$		$35.0 \pm 4.6$		$10.9 \pm 3.5$	
Data	4376		801		31	

Table 7.4: The number of predicted background events in the signal region compared to the data, for the two-tag, three-tag, and four-tag samples. The yields for a 2 TeV scalar and a 2 TeV  $G_{KK}$  with  $k/\bar{M}_P = 1.0$  are also shown. The scalar is normalised to a cross section times branching ratio of 12 fb. The quoted uncertainties include both the statistical and systematic uncertainties. The anti-correlation between the multijet and  $t\bar{t}$  yields is accounted for in the uncertainty in the total background yield [66].

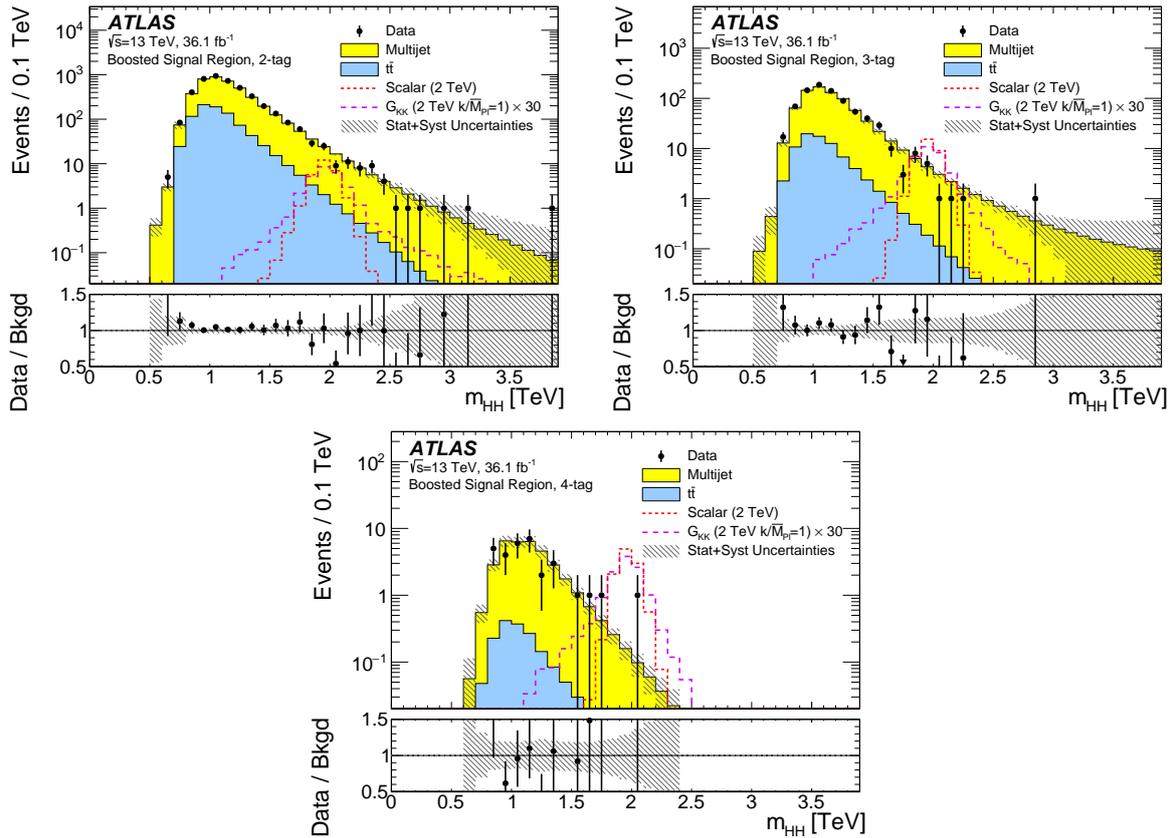
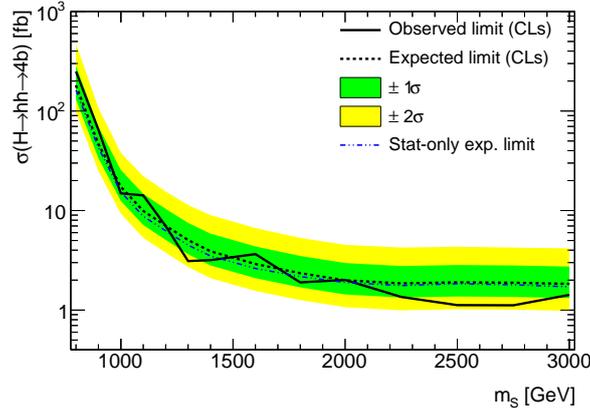


Figure 7.20: The distributions of  $m_{HH}$  in the signal regions of the boosted analysis for the 2-tag sample, the 3-tag sample, and the 4-tag sample, compared to the predicted backgrounds. The data-to-background ratio (bottom panels) shows also the combination of statistical and systematic uncertainties as the grey hatched band. The expected signal for a 2 TeV  $G_{KK}$  resonance with  $k/\bar{M}_P = 1$  and a scalar with the same mass is also shown. The scalar has an arbitrary cross section times branching ratio of 12 fb [66].



(a) Narrow-width scalar

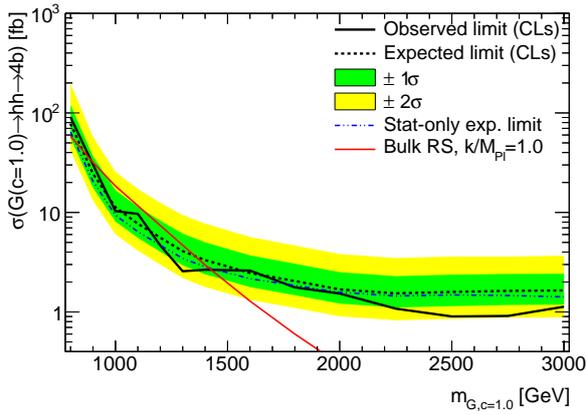
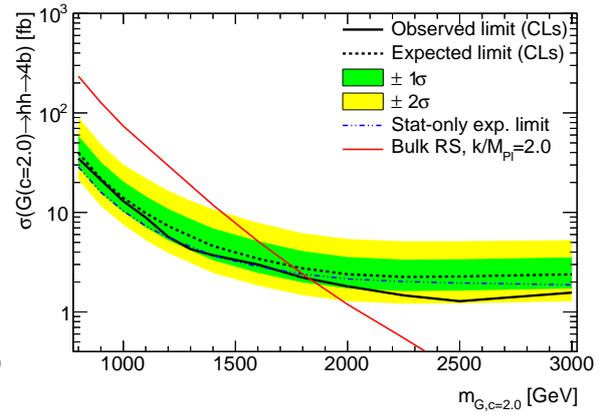
(b) Bulk RS Graviton with  $k/\bar{M}_{Pl} = 1$ (c) Bulk RS Graviton with  $k/\bar{M}_{Pl} = 2$ 

Figure 7.21: The observed and expected 95% CL upper limits on the production cross section times branching ratio for the (a) narrow-width scalar, (b) bulk RS model with  $k/\bar{M}_{Pl} = 1$  and (c) bulk RS model with  $k/\bar{M}_{Pl} = 2$  considering only the boosted analysis including all systematic uncertainties. The dot-dashed line shows the expected limit when only statistical uncertainties are included. An additional (red) curve shows the predicted cross section as a function of resonance mass for each of the graviton models. The limits are derived within the asymptotic approximation [26].

between 2000 GeV-2500 GeV, and 2500 GeV-3000 GeV, respectively.

In order to assess the impact of the jet mass choice on the sensitivity of the boosted analysis, the expected limit is reproduced using  $m^{\text{calo}}$  instead of  $m^{\text{comb}}$  for  $G_{KK}$  with  $k/\bar{M}_{\text{Pl}} = 1$ . In Figure 7.22, the comparison of the two results are shown. Using  $m^{\text{comb}}$  instead of  $m^{\text{calo}}$  improves the boosted exclusion limit across the mass range up to 10% in highly boosted regimes.

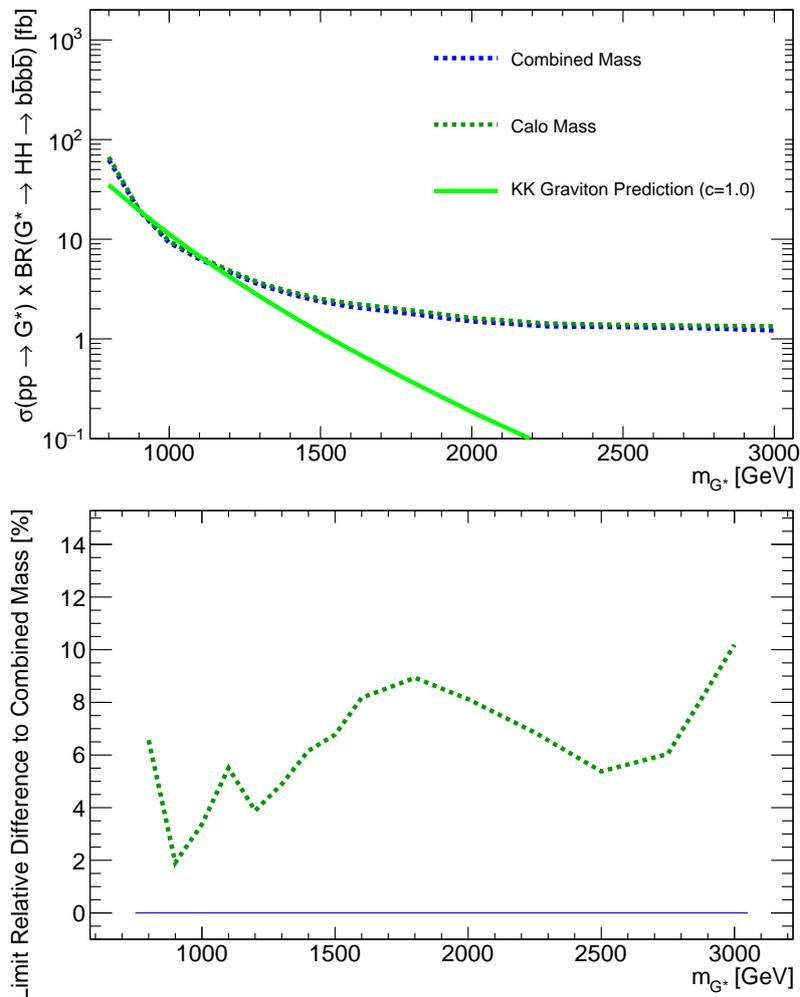
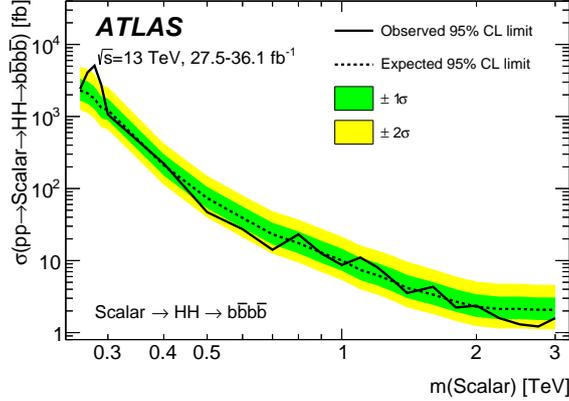


Figure 7.22: The expected limits obtained with  $m^{\text{comb}}$  and  $m^{\text{calo}}$  jet mass definitions using the combined boosted signal regions. The top plot shows both expected limits for a  $G_{KK}$  with  $k/\bar{M}_{\text{Pl}} = 1$ . The bottom plot shows the ratio of the exclusion limit obtained with  $m^{\text{calo}}$  (green) relative to the limit obtained with  $m^{\text{comb}}$  (blue) [74].

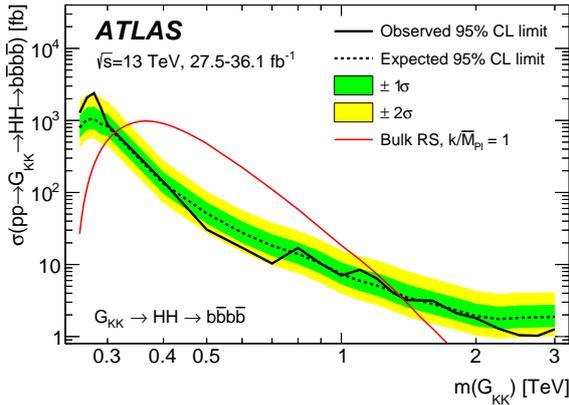
### 7.5.1 Combined results with the resolved analysis

As introduced in Section 7.1.3, the boosted and resolved analyses are performed separately using orthogonal selections. A statistical analysis is performed to combine the boosted and the resolved analyses to search for resonant Higgs pair production, while for the non-resonant search, only the results of the resolved analysis are used since the sensitivity to non-resonant signals is found much better in the resolved analysis compared to the boosted analysis [66]. As the search ranges of the two analyses are different, the combination of the results is only performed in the overlapping mass range, between 800

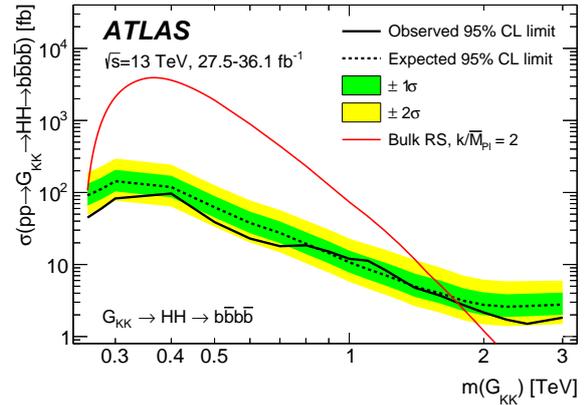
GeV and 1400 GeV. In Figure 7.23, the combined analysis results for the resonant Higgs pair production are shown.



(a) Narrow-width scalar



(b) Bulk RS Graviton with  $k/\bar{M}_{\text{Pl}} = 1$



(c) Bulk RS Graviton with  $k/\bar{M}_{\text{Pl}} = 2$

Figure 7.23: The observed and expected 95% CL upper limits on the production cross section times branching ratio for the (a) narrow-width scalar, (b) bulk RS model with  $k/\bar{M}_{\text{Pl}} = 1$  and (c) bulk RS model with  $k/\bar{M}_{\text{Pl}} = 2$ . An additional (red) curve shows the predicted cross section as a function of resonance mass for each of the graviton models [66].

The largest local deviation is found at 280 GeV in the resolved analysis mass range with  $3.6\sigma$  significance for the narrow-width scalar signal and  $2.5\sigma$  significance for the  $k/\bar{M}_{\text{Pl}} = 1$  graviton signal. The graviton with  $k/\bar{M}_{\text{Pl}} = 2$  is found to be too wide to explain the deviation. For the narrow-width signal the global significance is  $2.3\sigma$ .

As a result of the combination of the two analyses, the bulk RS model with  $k/\bar{M}_{\text{Pl}} = 1$  is excluded for the masses between 313 GeV and 1362 GeV and the bulk RS model with  $k/\bar{M}_{\text{Pl}} = 2$  is excluded below 1744 GeV. The search sensitivity is significantly improved between 1 and 1.5 TeV due to the combination of the two analyses. For the non-resonant Higgs boson pair production 95% CL upper limits are set on the production cross section of 147 fb [66].

## 7.6 Conclusion

This chapter presents a boosted analysis performed in the context of the search for pair produced Higgs bosons in the  $b\bar{b}b\bar{b}$  channel within the ATLAS experiment. The analysis uses  $36.1 \text{ fb}^{-1}$  of  $pp$  collision data collected by ATLAS in 2015 and 2016 at  $\sqrt{s} = 13 \text{ TeV}$ . No significant excess is observed and upper limits are set on the production cross section times branching ratio for the  $b\bar{b}b\bar{b}$  final state.

The analysis uses the techniques introduced in Chapter 6 to identify the boosted Higgs bosons. The muon corrected  $m^{\text{comb}}$  observable is used as the default jet mass definition of the analysis as a result of the selection optimisation studies presented in Section 7.2.4, leading to a significant improvement in the search sensitivity. The analysis uses a sophisticated data driven multijet background estimation method combining the control regions with lower  $b$ -tagged track jet multiplicities. In this chapter, this method is motivated and validated with the presented MC simulation studies as shown in Section 7.3.1. The kinematic properties of the  $b$ -tagged track jets are investigated and the impact of the tagging algorithm on the kinematic properties of the large- $R$  jets is revealed.

The combination of the presented boosted analysis and the performed resolved analysis excluded Kaluza Klein Graviton masses between 313 GeV and 1362 GeV predicted in the bulk RS model with  $k/\bar{M}_{\text{Pl}} = 1$ , and masses below 1744 GeV in the bulk RS model with  $k/\bar{M}_{\text{Pl}} = 2$ . The 95% CL upper limit on the non-resonant di-Higgs production is set at 147 fb corresponding to 12.9 times the SM expectation.

# Chapter 8

## Search for boosted di- $b$ -jet resonances associated with a jet

### 8.1 Overview

In this chapter, a search for boosted dijet resonances decaying to two  $b$ -quarks and produced in association with an initial state radiation (ISR) jet, using  $80.5 \text{ fb}^{-1}$  ATLAS  $pp$  data collected in 2015, 2016 and 2017 is presented [67]. The aim of the chapter is to explain the search strategy and optimised event selections stressing the connection with Chapters 6 and 7. Similar to the presented search in Chapter 7, this analysis suffers from large multijet background. Moreover, using the facts revealed previously for the modelling of boosted  $b\bar{b}$  pairs, a new data driven multijet background estimation method is explored and its results are discussed in this chapter to shed light for future studies. The analysis looks for a resonance in the invariant mass distribution of boosted  $b\bar{b}$  pairs within 70 and 230 GeV. The physics motivation as to why that particular region and topology is investigated is discussed in the following two sections, while the analysis strategy is explained in Section 8.1.3.

#### 8.1.1 Searching for boosted Higgs boson production

After the discovery of the Higgs boson at 125 GeV in  $\gamma\gamma$  channel, efforts have been increased to look for the Higgs boson in the remaining decay channels and better understand the nature of this newly discovered particle. Although obtaining  $5\sigma_{\text{sig}}$  significance in the fully hadronic decay channels was a challenging task due to the large multijet background, a greater dataset and sophisticated analysis techniques made this possible. Recently, measuring 5.4 and 5.6  $\sigma_{\text{sig}}$  significances, first observations from the ATLAS and the CMS for the  $h \rightarrow b\bar{b}$  process have been published in the context of a  $VH(\rightarrow b\bar{b})$  analysis [122, 123]. Despite the fact that the dominant production mechanism of the Higgs boson is gluon fusion production mode (ggF) as shown in Section 2.3.3, searches in this production mode of the Higgs are still recent and in development. Due to the overwhelming multijet background in this channel, the expected search sensitivities are rather low. While CMS published an inclusive search for the boosted  $h \rightarrow b\bar{b}$  process with an observed local significance of  $1.5\sigma_{\text{sig}}$  standard deviation [154], the presented search in this chapter is the first ATLAS search to observe boosted  $h \rightarrow b\bar{b}$  production in association with an additional jet, providing an opportunity to elucidate the structure of gluon fusion production. The corresponding Feynman diagram for the process is shown in Figure 8.1.

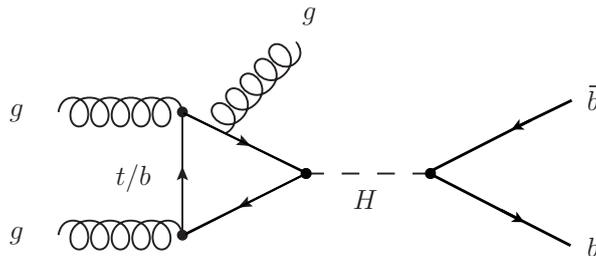


Figure 8.1: Feynman diagram for Higgs production associated with an initial state radiation

However, the aim of searching for the boosted Higgs production is not only to measure the SM  $h \rightarrow b\bar{b}$  coupling, it is also to search for new physics beyond the SM. Since the Higgs boson production starts being sensitive to the top quark loop at high  $p_T$  regimes, possible new physics contributions from new resonances can be expected in this channel [42]. Moreover, the anomalous couplings can lead to effective gluon-gluon-Higgs interaction. In case these scenarios are true, increases in the boosted Higgs production is foreseen as high as 50% higher than the SM production [41] and searches for boosted Higgs production offer an invaluable way to discover new physics.

### 8.1.2 Searching for a mediator to dark matter

As previously described in Section 2.3.2, there are several ways to search for a dark matter candidate in colliders. In the context of simplified dark matter models dijet resonance searches are one of the strongest approaches, with the possibility to target a significant portion of phase spaces, as shown earlier in Figure 2.7. Standard dijet resonance searches are successfully exploring the high mass regions for mediators of the simplified dark matter models. Following the increased centre-of-mass energy at the LHC, unprecedented mass limits at TeV scales have been reached. However, the situation is different for resonant mass values at sub-TeV scales. As one approaches lower mass values, the amount of the multijet background becomes increasingly overwhelming and searches lose their sensitivities. To overcome this limitation, both the ATLAS and the CMS experiments developed and used a new technique [155, 156], which allows to collect more data in the low mass regions by saving less information per each event. These events are recorded to dedicated data streams, and receive a separate treatment. Using this technique, searches referred to as a *Trigger Level Analysis* (TLA) are performed in ATLAS, extending the dijet resonance mass limits down to 400 GeV [155]. However, below that, the limitation for a TLA is simply the L1 thresholds, which are optimised according to the capability of the DAQ system of the experiment and the available resources for storing data. Therefore, another approach is needed to be able to reach the lower resonance mass values. At this point, the idea of triggering events using additional ISR contribution is born. In this case, the ISR recoils against to dijet system and the energy of ISR can be shared by the decay products of the resonances, allowing to explore much lower resonance mass values than the threshold energy to trigger the event. These searches are referred as *dijet+ISR* searches, and performed in two channels depending on the radiation type; either with photon or jet as ISR contribution [157].

In Figure 8.2, the summary of the above searches is shown for a  $Z'$  axial-vector mediator model in terms of upper exclusion limits in the coupling-mediator mass plane. As can be seen from the plot, *dijet+ISR* searches lose their sensitivity at resonant masses

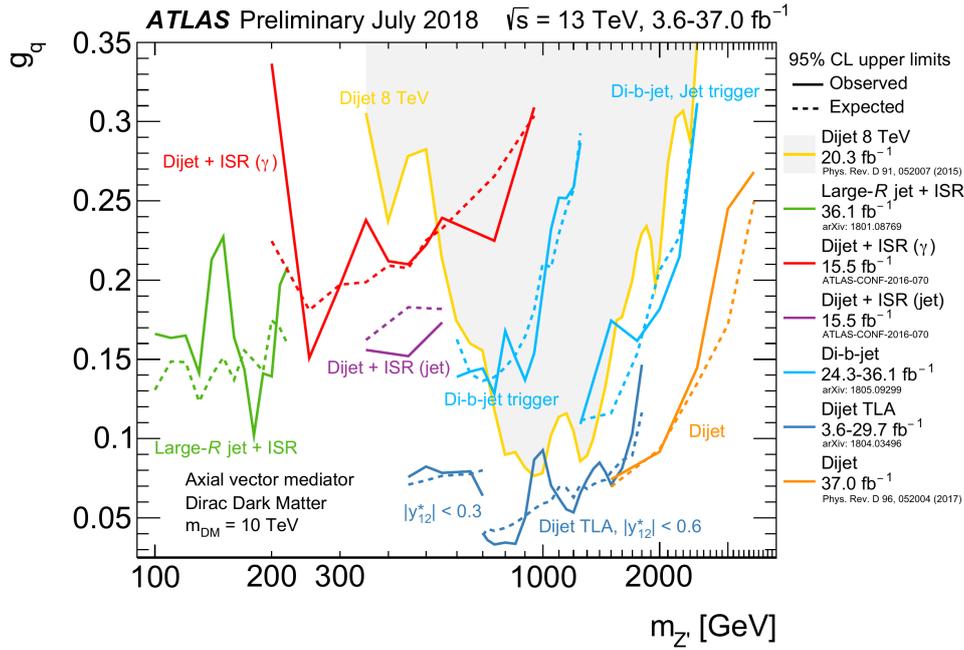


Figure 8.2: Summary plot of ATLAS bounds in the coupling-mediator mass plane from dijet resonance searches using 2012, 2015 and 2016 data [39]. The 95% CL upper limits are obtained from seven ATLAS resonance searches on the coupling  $g_q$  as a function of the resonance mass  $m_{Z'}$  for the leptophobic  $Z'$  model described in Section 2.3.2.

below 250 GeV. The reason for this sensitivity loss is simply due to the boost of the dijet system in the recorded events at low resonant mass values. Since the decay products of the resonance get higher  $p_T$  values in order to balance the energetic ISR at low masses, it is not possible to identify them separately as two regular small- $R$  jets. As discussed in Section 5.1.4, the natural choice when reconstructing these particular events in boosted topologies is to require large- $R$  jets instead of two small- $R$  jets. Performing this analysis [158], referred to *Large- $R$  jet+ISR*, provides the lowest mass limits down to 100 GeV as can be seen in Figure 8.2. Thus, it is evident that boosted topologies are advantageous in targeting low-mass dijet resonances.

Another low-mass dijet resonance channel is the search for di- $b$ -jets with ISR instead of inclusive dijets in the final state, as shown in Figure 8.3. Many physics models predict low-mass dijet resonances decaying to  $b$ -quarks. Since Yukawa couplings are proportional to the mass of the decay particle,  $b$ -quark decays are favoured with their higher mass values. However, even if the new resonance decays democratically to all quarks, requiring  $b$ -jets in the final state of the analysis can reduce the amount of multijet background significantly. Therefore, searching for boosted di- $b$ -jet resonances associated with an ISR can improve limits further, reaching lower TeV mediator mass values and coupling parameters for the simplified dark matter model.

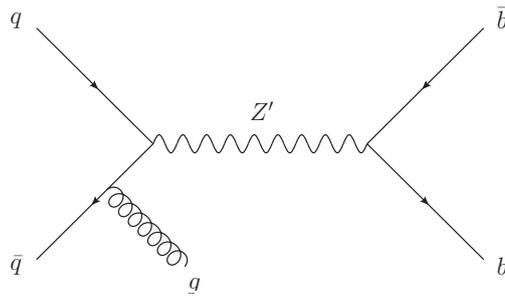


Figure 8.3: Feynman diagram for a di- $b$ -jet resonance with a hard gluon radiated from the initial state radiation.

### 8.1.3 Analysis strategy

This search investigates the process of a boosted di- $b$ -jet pair that recoils against an energetic ISR jet with the two physics motivation explained above. As extensively discussed throughout the thesis, in boosted topologies the  $b\bar{b}$  pair is expected to be reconstructed as a single large- $R$  jet. Since there is no major difference between ISR jets reconstructed as large or small- $R$  jets, large- $R$  jets are used to identify the ISR jet in the final state to simplify the analysis and the propagation of the related uncertainties.

The analysis selection and strategies are optimised considering both the  $Z'$  and the Higgs boson as resonance candidates using simulated samples described in Chapter 4. The signature of this search is shown in Figure 8.4. It involves two large- $R$  jets that are recoiling against each other, one of which contains two  $b$ -hadrons.

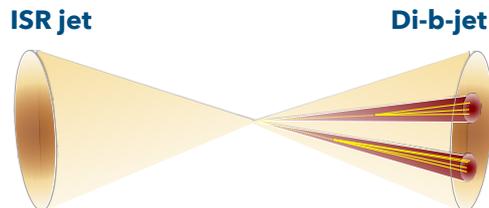


Figure 8.4: Sketch of the process expected to be seen in the transverse plane of the detector.

## 8.2 Event selection

### 8.2.1 Trigger

The events are required to be triggered by the presence of a large- $R$  jet reconstructed by the HLT. Each year of data taking, the thresholds for the lowest unscaled large- $R$  jet trigger has changed. Therefore, the analysis uses three triggers for different datasets from each year. Below they are listed together with the  $p_T$  thresholds and recorded integrated luminosities:

- 2015 - HLT ungroomed large- $R$  jet with  $p_T > 360$  GeV,  $3.2 \text{ fb}^{-1}$ ,
- 2016 - HLT ungroomed large- $R$  jet with  $p_T > 420$  GeV,  $33.0 \text{ fb}^{-1}$ ,
- 2017 - HLT trimmed large- $R$  jet with  $p_T > 460$  GeV,  $44.3 \text{ fb}^{-1}$ .

For simulated events, the requirement is to accept the events which pass at least one of the above triggers to simplify the combination of the three years. In order to be on the plateau of all three triggers, events are required to have a trimmed large- $R$  jet with  $p_T > 480$  GeV.

### 8.2.2 Preselection

Similarly to the preselection presented in Section 7.2.2, before going into the detailed selection and optimisation, the following common requirements are applied to each event passing the trigger selection. They are listed below according to their order in the actual analysis selection.

- Event cleaning procedure is applied as introduced in Section 7.2.2, in order to remove the events that have significant calorimeter noise. Unclean jets are identified from small- $R$  jets with  $p_T$  above 25 GeV satisfying *LooseBad* criteria [159]. Events are removed depending on the kinematic features of the bad jet following Reference [160].
- At least two large- $R$  jets are required in each event. Leading and subleading large- $R$  jets have to fulfill minimum  $p_T$  requirements of  $p_T > 480$  GeV and  $p_T > 250$  GeV, respectively. While a leading large- $R$  jet  $p_T$  cut is chosen to ensure that all three triggers are fully efficient, a subleading large- $R$  jet  $p_T$  cut is chosen to mimic the analysis topology, with the presence of considerable hadronic initial state radiation.
- Note that the signal candidate large- $R$  jet, either being a  $Z'$  or a Higgs, is supposed to contain the decay products of the boosted particle, the  $b\bar{b}$  pair. It is chosen among all large- $R$  jets present in each event as described here. Firstly, large- $R$  jets that satisfy the criteria given below are determined, and then, they are ordered according to their  $p_T$ . As a signal candidate jet, the one with the highest  $p_T$  is selected.
  - The signal candidate jet has to have at least 2 ghost associated variable- $R$  track jets with  $p_T > 10$  GeV independent of its flavour.
  - The distance between the two leading variable- $R$  track jets is supposed to be larger than the minimum radius size of the variable- $R$  track jets,  $\Delta R_{jj} / \min R_{VR} > 1$  where  $R_{VR}$  is the radius of variable- $R$  track jets and  $\Delta R_{jj}$  is the distance between them. This requirement prevents the situations where one track jet contains the other one as depicted in Figure 8.5.

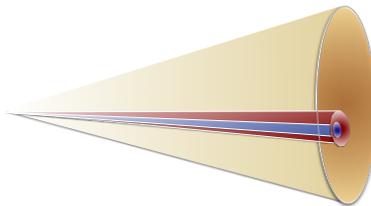


Figure 8.5: Sketch of a large- $R$  jet where one of the variable- $R$  track jets is reconstructed inside the other one.

- The signal candidate jet is supposed to satisfy the boost requirement  $2m_J/p_{T,J} < 1$  as introduced in Equation 5.6. Fulfilling this condition, the signal candidate jet can fully contain the 2  $b$ -hadrons within  $R = 1$ , and partially reconstructed large- $R$  jets are removed.
- To remove possible  $t\bar{t}$  background contamination (ttbarCRveto), events with muons on the opposite side of the signal candidate jet are removed. These muons are selected with the following kinematic requirements,  $p_T > 40$  GeV and  $\Delta\phi > \frac{2}{3}\pi$  with respect to the signal candidate.
- The  $p_T$  of the signal candidate is required to be above 480 GeV to ensure a smooth  $p_T$  distribution in order to avoid possible bias in the mass spectrum, which could occur due to different  $p_T$  requirements on the leading large- $R$  jet and the signal jet.
- As a final selection, the invariant mass of the signal candidate jet is chosen to be above 40 GeV, reducing large number of events that are not in the region of interest of this search. The reason why this range is selected will be clarified later in Section 8.3.4. Moreover, large- $R$  jet mass calibration is optimised for jets with masses above 50 GeV. The calibration would carry large uncertainties for jets with masses below 50 GeV [161].

### 8.2.3 Final event categorisation

After preselection and the determination of the signal candidate jet, further categorisation is applied to the selected events. Events that have the desired final state with  $b$ -tagged track jets inside the signal candidate is categorised as the signal region (SR), while the remaining events are selected for the control region (CR) or validation region (VR) depending on varying  $b$ -tagging criteria. While the corresponding  $b$ -tagging WPs and efficiencies for variable- $R$  track jets with MV2c10 algorithm can be recalled from Section 5.1.6, in this analysis, tight  $b$ -tagged track jets satisfy the WP with a 77%  $b$ -tagging efficiency, and loose ones correspond to the WP with a 85%  $b$ -tagging efficiency. Note that tight  $b$ -tagged track jets are also loose  $b$ -tagged track jets by definition. Event categorisation is done considering only the signal candidate jet as described below.

- Signal Region ( $2b$  SR): Signal candidate has exactly 2 tight  $b$ -tagged track jets.
- Validation Region ( $1/2b$  VR): Signal candidate can not have a tight  $b$ -tagged track jet, but it can contain 1 or 2 loose  $b$ -tagged track jet(s).
- Control Region ( $0b$  CR): Signal candidate jet can not have a loose  $b$ -tagged track jet.

While the benefits of defining a CR and VR becomes more evident in the following sections, it should be briefly stated that these regions are mainly used to test and validate the developed methods for the analysis, before analysing the data in the SR.

Table 8.1 shows the relative event yields passing the preselections and the final event categorisation for the simulated backgrounds and the data.

	Cut	QCD	W+jets	Z+jets	$t\bar{t}$	Data
<b>Preselection</b>	Trigger	1.000	1.000	1.000	1.000	1.000
	Jet Cleaning	1.000	0.999	0.999	0.999	0.999
	Leading large- $R$ jet $p_T > 480$ GeV	0.200	0.301	0.290	0.263	0.484
	Subleading large- $R$ jet $p_T > 250$ GeV	0.181	0.284	0.275	0.218	0.446
	At least one signal candidate	0.160	0.272	0.263	0.210	0.402
	No opposite muon (ttbarCR veto)	0.160	0.272	0.263	0.198	0.402
	signal candidate $p_T > 480$ GeV signal candidate mass $> 40$ GeV	0.135 0.099	0.246 0.208	0.236 0.198	0.177 0.159	0.338 0.248
<b>CR</b>	0 $b$ -tagged track jet (85% WP)	0.076	0.154	0.134	0.068	0.187
<b>VR</b>	0 $b$ -tagged track jet (77% WP)	0.086	0.180	0.156	0.085	0.214
	$\geq 1$ $b$ -tagged track jet (85% WP)	0.009	0.025	0.021	0.016	0.026
<b>SR</b>	2 $b$ -tagged track jets (77% WP)	0.001	0.001	0.010	0.006	0.003

Table 8.1: The cutflow relative to the number of events after the trigger requirement in different regions using simulated background events and data. QCD contribution is scaled by a factor of 0.74 to normalise Pythia samples. This factor is obtained from fitting studies performed with CR data, more details of which can be found in Reference [117].

### 8.2.4 Impact of the track jet choice

As already explained in Chapter 6, variable- $R$  track jets became available for ATLAS Run 2 physics analyses only after 2017. Although there were few results available demonstrating the performance gain in using variable- $R$  track jets over fixed- $R$  track jets by the time this analysis started, the impact may vary depending on the topologies and event selections of each individual analysis. Therefore, in order to optimise the final signal region selection of this analysis, the comparisons are performed using the two alternative track jet reconstruction methods.

Figure 8.6 shows that, as expected from the variable- $R$  jet finding algorithm, the ability to resolve two separate track jets is higher with variable- $R$  track jets than fixed- $R$  track jets. The total number of  $b$ -tagged fixed- $R$  track jets are checked for each signal candidate jet obtained from the event selection using variable- $R$  track jets. Results show that considerable amount of the large- $R$  jets in the SR, that is created with variable- $R$  track jets, have only one associated  $b$ -tagged fixed- $R$  jet. Furthermore, the distances between the two leading track jets obtained from the two SR selections are shown in Figure 8.6, indicating that the amount of the variable- $R$  track jets which are reconstructed with smaller radius than 0.2 is non-negligible.

Apart from the gain in signal selection efficiency, the modelling of the kinematic variables is checked for each track jet reconstruction method. As extensively discussed in Section 7.3.1,  $b$ -tagging the track jets sculpts the distributions. Therefore, also for this analysis, it is important to understand the modelling differences between the selected SR, VR, and CR in order to estimate the multijet background. Figure 8.7 compares the shapes of the inclusive track jet  $p_T$  distributions and the  $b$ -tagged track jet  $p_T$  distributions for both reconstruction methods in different event categories. The impact of  $b$ -tagging on the  $p_T$  modelling is also visible with the selections using variable- $R$  track jets.

In summary, the impact of the track jet selection on the signal region selection efficiency is evaluated and it is found that using variable- $R$  track jets increasing the signal selection efficiency. The modelling differences between the two methods are also investigated particularly for  $b$ -tagging, and the results are found similar to each other.

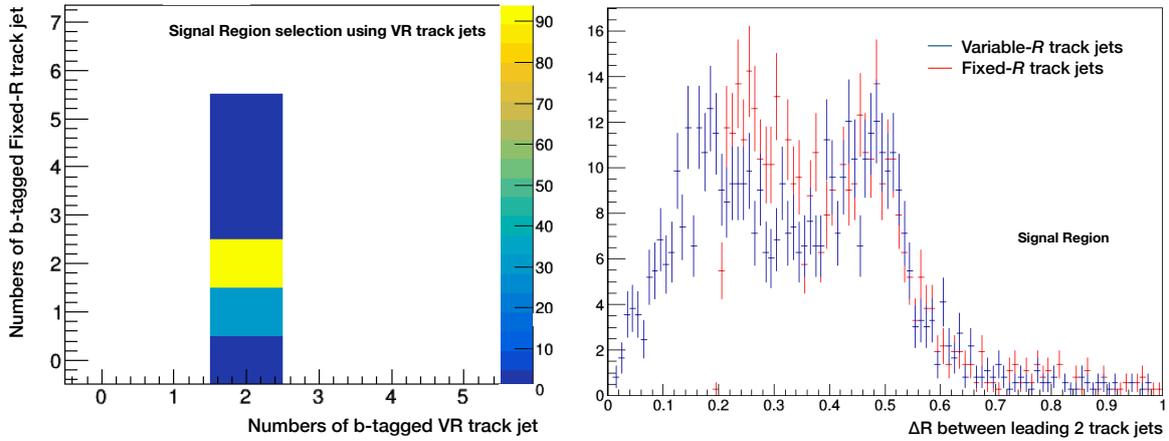


Figure 8.6: (Left)  $b$ -tagged fixed- $R$  track jet multiplicity obtained from the signal candidates in the SR created with variable- $R$  track jets. (Right)  $\Delta R$  distributions between the two leading track jets associated to the signal candidate large- $R$  jet. Distributions are obtained from the events in the two different SR created with alternative track jet reconstruction methods.

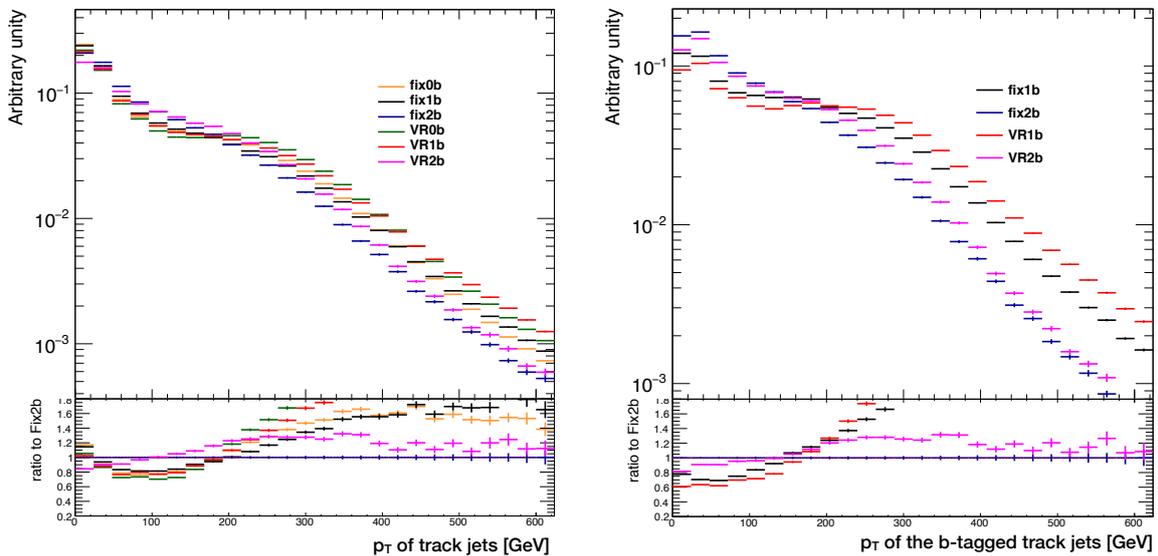


Figure 8.7:  $p_T$  distributions of variable- $R$  and fixed- $R$  track jets in different event categorisations. In order to avoid a possible bias that can come from the usage of different  $b$ -tagging efficiency WPs during event categorisation, same  $b$ -tagging WP, 77%, is used for event categorisation. While the left plot shows the two leading track jet  $p_T$  distribution in each event selection category, the right plot shows only the  $b$ -tagged track jet distributions in  $1b$  and  $2b$  categories.

## 8.3 Background modelling

As stated before, this analysis aims to explore the low mass region of the coupling-mediator mass plane, especially below 200 GeV, where the resolved topology resonance searches lose their sensitivity (recall Figure 8.2). Hence searching for an excess in this specific region is quite interesting from a physics perspective, but also challenging not only because of the existence of large multijet background, but also the existence of several other resonant SM particles within the same range. In this section, the estimation methods both for the resonant and non-resonant backgrounds are explained.

### 8.3.1 Resonant backgrounds

With a selection optimised for  $b\bar{b}$  pairs in the final state, one can get almost the entire  $Z + \text{jets}$  events in the  $Z \rightarrow b\bar{b}$  decay channel as background. In this analysis the contribution of  $Z + \text{jets}$  processes is also quite large. As another resonant SM process peaking at the similar mass values as the  $Z$  boson,  $W + \text{jets}$  processes can pass the SR selection in case one of the track jets is mis-tagged as a  $b$ -jet or additional  $b$ -hadrons from background radiations are caught in the jet area. In this analysis, both of the backgrounds coming from  $Z + \text{jets}$  and  $W + \text{jets}$  processes are treated together as  $V + \text{jets}$ , since their contributions are inseparable and kinematically similar. With the presence of  $b$ -tagged track jets associated to the large- $R$  jets that are reconstructed around the top quarks,  $t\bar{t}$  decays can also produce final states that are similar to the final state of this analysis, especially in their full hadronic decay channels. Lastly, when this search considers  $Z'$  mediator as a benchmark signal candidate,  $h \rightarrow b\bar{b}$  decays also contribute as a background to the SR. Three of the Higgs production mechanisms are considered for this analysis: ggF, VBF, and Higgsstrahlung (recall Figure 2.4). The biggest contribution comes from the ggF as shown in Table 8.2. However, the impact of the total Higgs contribution on the total background in the SR is found insignificant compared to other backgrounds, therefore it is not included for the final background estimation that is discussed later in Section 8.3.4.

The contributions of the SM resonant backgrounds in SR as a function of the signal candidate jet mass are shown in Figure 8.8, using the MC simulations given in Chapter 4. As can be seen from the figure,  $t\bar{t}$  and  $Z + \text{jets}$  backgrounds are clearly dominant among the resonant backgrounds. Based on previous studies [162], it is known that  $t\bar{t}$  MC simulations are not able to model the cross section properly, especially in the high  $p_T$

		CR	VR	SR
$V + \text{jets}$	$Z + \text{jets}$	0.28	0.27	0.80
	$W + \text{jets}$	0.72	0.73	0.20
$t\bar{t}$	All Hadronic	0.58	0.61	0.63
	Semi-Leptonic	0.38	0.35	0.34
	Dileptonic	0.04	0.03	0.03
$H \rightarrow b\bar{b}$	ggF	0.49	0.48	0.53
	VBF	0.17	0.20	0.25
	$WH$	0.21	0.19	0.12
	$ZH$	0.12	0.12	0.10

Table 8.2: The fractional composition of the different resonant contributions in the analysis regions for the three categories:  $V + \text{jets}$ ,  $t\bar{t}$ , and Higgs production. Results are obtained from simulations.

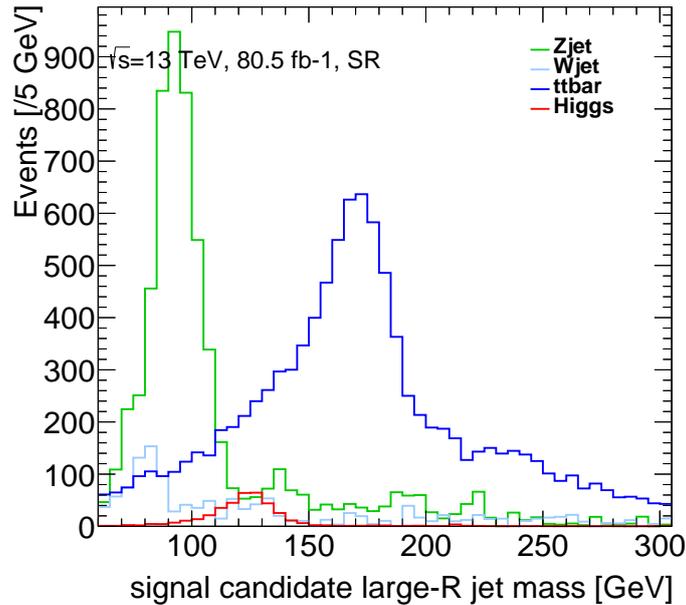


Figure 8.8: Resonant SM backgrounds in the signal region are shown as a function of the signal candidate large- $R$  jet mass. The absolute normalisation allows a comparison of the yields.

regions. Thus, in order to predict the  $t\bar{t}$  normalisation correctly, a data driven method is used. All other SM resonant backgrounds are modelled by MC simulations. Below, the data driven method to estimate the yield of  $t\bar{t}$  events is explained.

### Data-driven normalisation for $t\bar{t}$

This section briefly describes the event selection to define  $t\bar{t}$  enriched control region in the data ( $CR_{t\bar{t}}$ ), and the procedure of extracting the normalisation scale factor from this region in order to correct the  $t\bar{t}$  yield in the SR. The selection to create  $CR_{t\bar{t}}$  is given below:

- Signal candidate large- $R$  jets are identified applying the same preselections as described in Section 8.2.2, except the muon veto requirement.
- At least 1 tight  $b$ -tagged track jet is required to be associated to the signal candidate jet.
- 1 loose, isolated muon is required in the opposite hemisphere of the signal candidate large- $R$  jet with  $p_T > 40$  GeV and  $\Delta\phi > \frac{2}{3}\pi$ .
- A large- $R$  jet with a tight  $b$ -tagged leading track jet is required within the distance of  $\Delta R < 1.5$  from the muon.

After applying these selections,  $t\bar{t}$  enriched control regions are obtained from the data and the MC. Comparing the signal candidate large- $R$  jet mass distributions in both, it can be seen that 90% of the events are  $t\bar{t}$  events, as intended. From these distributions, it is possible to extract a normalisation scale factor which brings the data and the MC into agreement in the kinematic regime of this analysis. In order to do that, data is fitted with templates for background processes within the signal candidate mass region

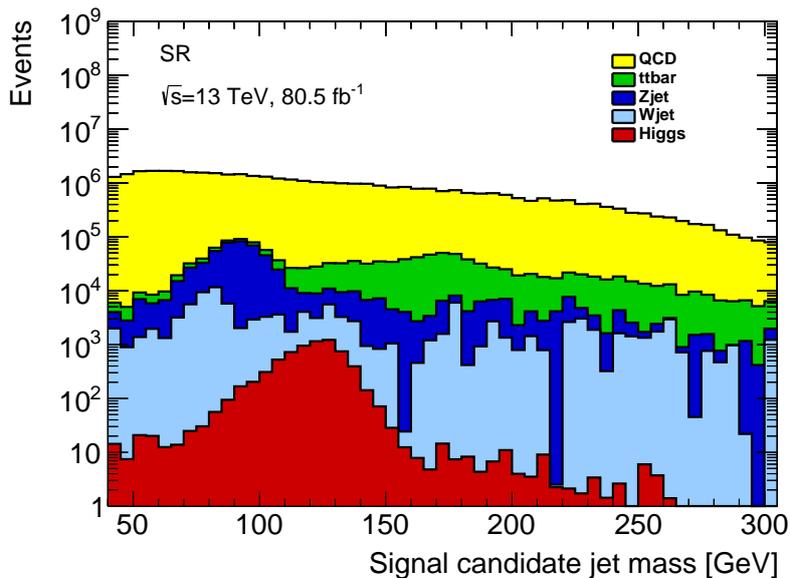


Figure 8.9: Expected composition of the event yield in the signal region as a function of the signal candidate large- $R$  jet mass.

[100 GeV, 200 GeV], allowing only the  $t\bar{t}$  template normalisation to float, and keeping the other template normalisations constant. The normalisation scale factor of 0.84 is extracted from the fit, and it is used to constrain the amount of  $t\bar{t}$  events in the final background estimation.

### 8.3.2 Non-resonant multijet background

The dominant background of this analysis is the non-resonant multijet background that is shown in Figure 8.9. As can be recalled from the search presented in Chapter 7, multijet simulations (4.2.1) can not be used for the background estimation due to the statistical limitations compared to the large jet production cross section at the LHC as indicated in Figure 6.1. Therefore, more sophisticated methods are needed to model this background to be sensitive for any kind of excess, and for a valid statistical interpretation of the analysis. However, the modelling of the kinematic variables for the  $b$ -tagged track jets and the signal candidate jet can be investigated using MC simulations, as the similarity between the data and the MC in terms of modelling has been shown earlier for the search presented in Chapter 7.

In the following sections, two different data driven background estimation approaches are discussed. While the first approach explores the possibility to estimate the multijet background using the data in the CR, *CR reweighting*; the second one uses a suitable parametric fit function, *direct fitting*, to describe the signal candidate large- $R$  jet mass distribution in the SR, which composed of the expected backgrounds and the possible signal contributions. Both of the approaches provide their own benefits and also disadvantages. While the final results shown for this analysis use the direct fitting method, the CR reweighting method stands as an alternative approach for future studies, especially with its advantage of being able to explore the lower mass regions of the coupling- $Z'$  mediator mass plane. In the following section, the CR reweighting method is discussed. The presented studies and the results that are shown in the context of this approach give

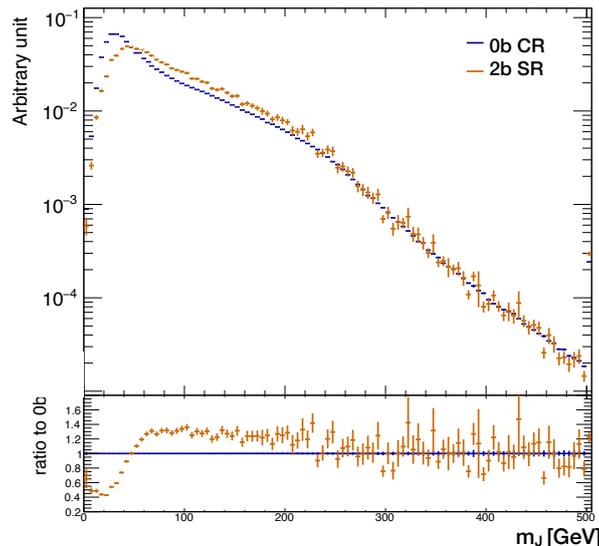


Figure 8.10: Shape comparison for the signal candidate jet mass distributions in the SR and in the CR obtained from the simulated multijet samples.

an insight into the multijet background characteristics of this analysis, and will be useful afterwards.

### 8.3.3 Data-driven multijet background: CR Reweighting

Recalling the detailed discussions in Chapter 7, it can be said that the estimation of the multijet background is not so straightforward with the involvement of  $b$ -tagging. As shown in Section 7.3.1 and also in Figure 8.7, the modelling of the kinematic observables related to  $b$ -tagged track jets is affected by the process of tagging. Therefore, more sophisticated approaches are needed to create a *good control region* (Section 7.3.1), which reflects the corresponding distributions in the SR.

As described in Section 8.2.3, the events with zero  $b$ -tagged track jets are selected as the CR of this analysis. Even if the requirement on the  $b$ -tagged track jet multiplicity is different than in the SR, it is still desirable to use this region for the background estimation method due to larger number of events and smaller uncertainties in the CR compared to the SR. In order to do this, the differences between the two regions must be understood and corrected for.

After the event categorisation, the shapes of the signal candidate large- $R$  jet mass ( $m_J$ ) distributions are compared in both regions in Figure 8.10. The result shows that the agreement between the distributions is limited, particularly in the mass region around 70 GeV. There is a sharp turn-on in the SR distribution with respect to the CR. Note that these low mass values are in the region of interest of this analysis. Therefore, the impact of this turn-on is non-negligible.

In order to understand the differences observed in the mass distributions, several other kinematic observables relevant to the  $m_J$  distribution are introduced and studied. For instance,  $p_T$  of the two leading track jets  $p_{T,1}$  and  $p_{T,2}$ , the opening angle between the two leading track jets ( $\Delta R$ ) associated to the signal candidate large- $R$  jet, the ratio ( $\mathcal{R}_{p_T}$ ) of the leading track jet  $p_{T,1}$  to the sum of the leading and subleading track jet  $p_T$ :  $\mathcal{R}_{p_T} \equiv p_{T,1}/(p_{T,1} + p_{T,2})$ , are four of the studied observables. In Figure 8.11, these two observables and the  $p_T$  distributions of the leading two track jets are shown both for the

SR and the CR. As can be seen from the figures, there are significant differences between the two regions. However, considering the different physics processes contributing to each compared region, the observed shapes are understandable. While the SR is dominated by  $g \rightarrow b\bar{b}$  splitting, the CR contains contributions from  $g \rightarrow q\bar{q}$  and  $g \rightarrow g\bar{g}$  splitting, as well as  $q \rightarrow q\bar{q}$  radiation. Thus, the kinematics of these processes are also expected to be different besides the impact of  $b$ -tagging algorithm efficiencies which is discussed in Chapter 7. The  $\mathcal{R}_{p_T}$  distribution is quite flat for the SR compared to the CR, since typically  $b\bar{b}$  pairs are sharing similar  $p_T$  values. It also explains the differences observed in the individual track jet  $p_T$  distributions. While the leading track jets are harder in the CR than the SR, it is the opposite for the subleading track jets. Finally, the topologies are also different in both region considering the heavier mass of the  $b$ -quarks, as the distance between the two track jets is observed to be higher for the  $b\bar{b}$  pairs compared to the non  $b$ -tagged track jets in the CR.

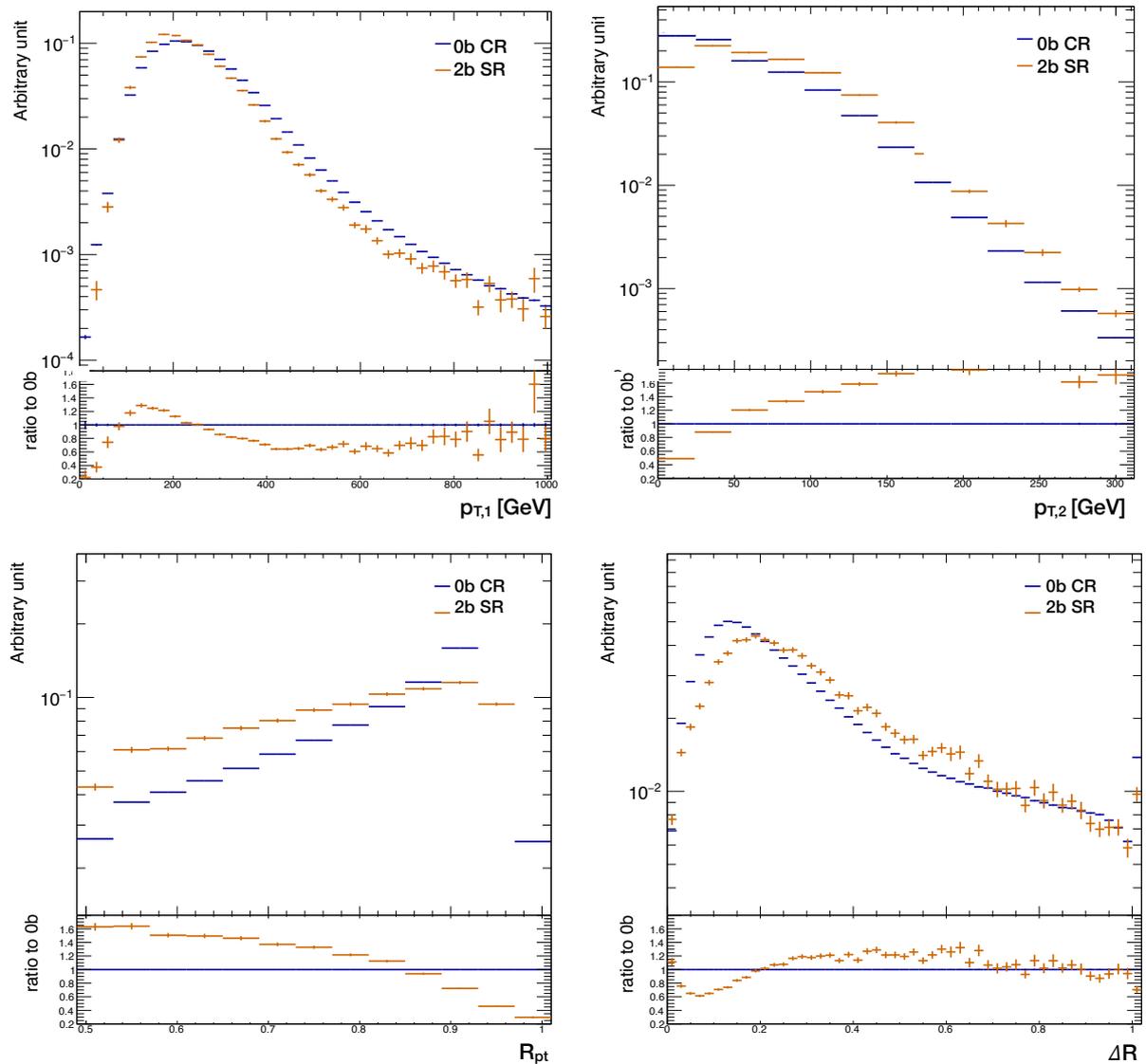


Figure 8.11: Kinematic observable distributions relevant to the signal candidate jet:  $p_T$  of the leading (top left) and subleading (top right) track jets,  $\mathcal{R}_{p_T}$  (bottom left), and the  $\Delta R$  between the two track jets (bottom right).

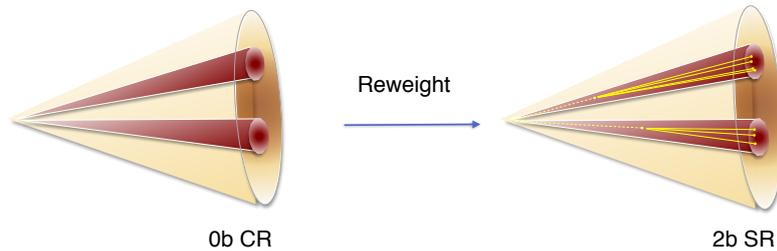


Figure 8.12: Sketch of the control region reweighting procedure. Signal candidate large- $R$  jets in the CR are reweighted to mimic the kinematics of the signal candidate large- $R$  jets with 2  $b$ -tagged track jets in the SR.

Using the observed differences between QCD jets in both categories, the main motivation for the CR reweighting approach is to bring the  $m_J$  distribution in the CR into agreement with the SR. In other words, the signal candidate large- $R$  jets in the CR are reweighted to mimic the signal candidate large- $R$  jets in the SR as depicted in Figure 8.12. A direct reweighting in  $m_J$  distribution between the CR and the SR is avoided since it would result in large statistical uncertainties being directly propagated bin-by-bin and because the potential signal contamination in the SR. However, using different observables other than  $m_J$  minimises the bias from a potential signal contamination, as it is not resonant in these observables.

Several kinematic observables correlated with  $m_J$  are considered in order to choose an appropriate set of observables for the reweighting procedure. Multijet simulations are used to develop and test the method as a first step. Technically, the reweighting is derived as a 2- or 3-dimensional ratio of the relevant distributions in the SR to those in the CR using simulation. Although using this ratio introduces a systematic uncertainty from the limited statistics of the events used in the reweighting, after propagation to the SR, it is about 1/10 of the statistical uncertainty in the SR. The procedure can be summarised as following:

- Obtain the 2 or 3 dimensional observable distributions both in the CR and the SR.
- For each bin in the given distribution, calculate the reweighting factor ( $R_F$ ) as the ratio of the number of jets in the CR and the SR  $R_F = N_{SR}^{QCD} / N_{CR}^{QCD}$ .
- Apply the  $R_F$  as an additional event weight.

Consequently, the reweighted CR distributions of the observables that are used to derive the  $R_F$  should be similar to the SR distributions.

Several combinations have been studied before finding the best candidate reweighting schemes. In this section, the best performing 2- and 3-dimensional combinations are described, where the figure of merit is the agreement in the  $m_J$  distribution between the multijet background in the SR and the reweighted multijet background prediction from the CR. The results of the other reweighting schemes can be found in Appendix C.2.

**Reweighting in 2 dimensions** The best performing reweighting in 2 dimensions is in  $(\Delta R, \mathcal{R}_{p_T})$ . The result of the reweighting is shown in Fig. 8.13. The sanity check is fulfilled for both reweighted distributions in  $\Delta R, \mathcal{R}_{p_T}$  that are brought into exact agreement with the SR distributions. A good agreement in  $m_J$  in the relevant range  $40 < m_J < 225$  is achieved, however, there is a deficit in the region  $40 < m_J < 70$ , where several data points of the reweighted CR distribution are below the SR.

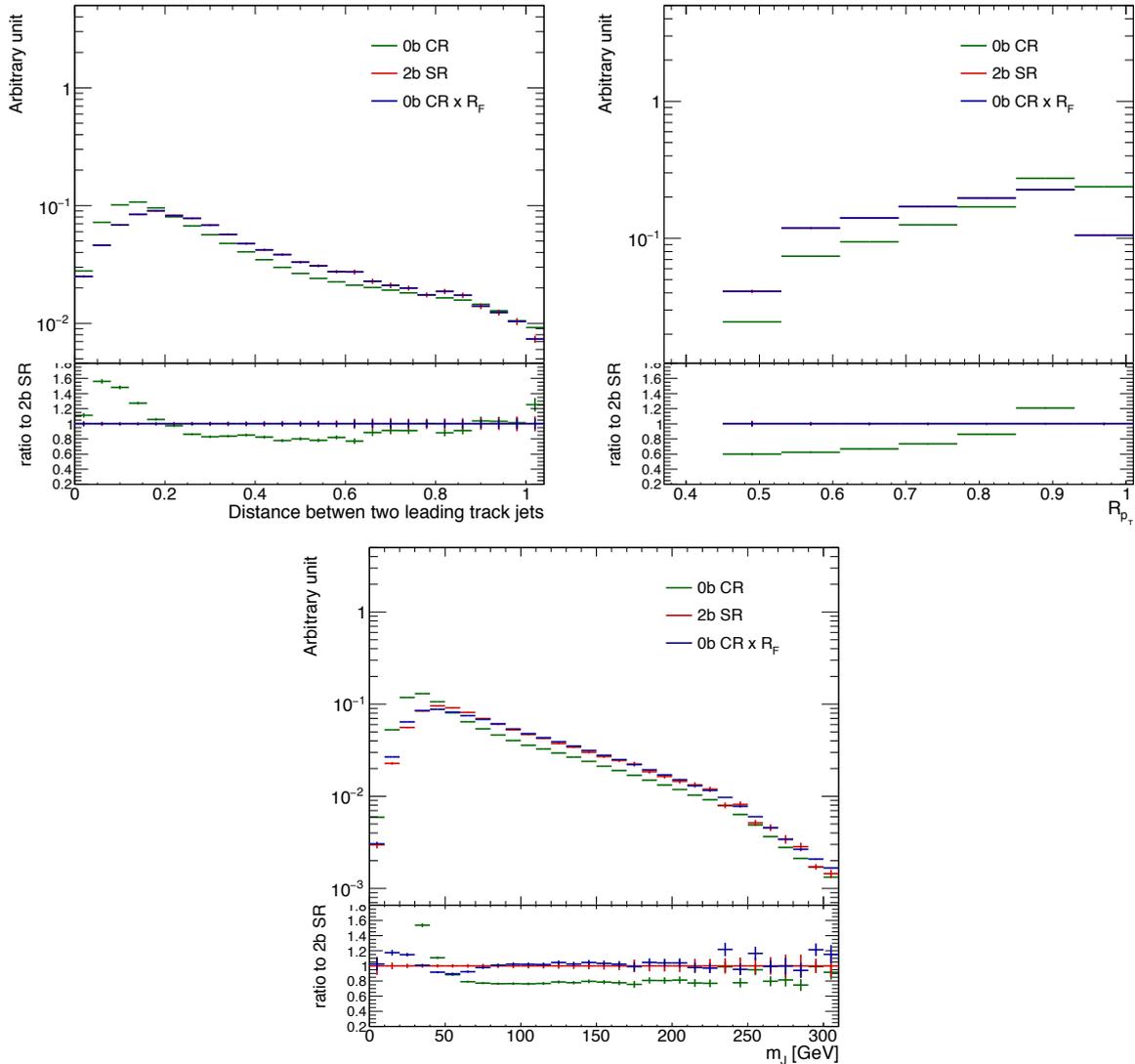


Figure 8.13: The effect of the 2-dimensional reweighting of the CR to the SR on the observables  $(\Delta R, \mathcal{R}_{p_T})$  for the distributions of  $\Delta R$ ,  $\mathcal{R}_{p_T}$ , and  $m_J$  of the signal candidate. The distribution in the SR (2b) is compared to the distribution in the CR before (0b) and after ( $\times R_F$ ) reweighting. The bottom panel shows the ratio relative to the SR. The two distributions,  $(\Delta R, \mathcal{R}_{p_T})$ , represent sanity checks that the reweighting is applied correctly, while the agreement in the  $m_J$  distribution (bottom) corresponds to the figure of merit for the performance.

**Reweighting in 3 dimensions** The best performing reweighting in 3 dimensions is in  $(p_{T,J}, \Delta R, \mathcal{R}_{p_T})$ , where  $p_{T,J}$  is the transverse momentum of the large- $R$  jet identified as the signal candidate. The result of the reweighting is shown in Figure 8.14. As a sanity check, all three reweighted distributions of  $p_{T,J}, \Delta R, \mathcal{R}_{p_T}$  are brought into exact agreement with the SR distributions. A good agreement in  $m_J$  in the relevant range

$40 < m_J < 225 \text{ GeV}$  is achieved, which is somewhat better than for the 2-dimensional reweighting shown in Figure 8.13. In the region  $40 < m_J < 70$  several data points of the reweighted CR distribution are below the SR. While this effect is consistent with no disagreement within statistical uncertainties, this region can be closely monitored in the VR using data, where more statistics is available.

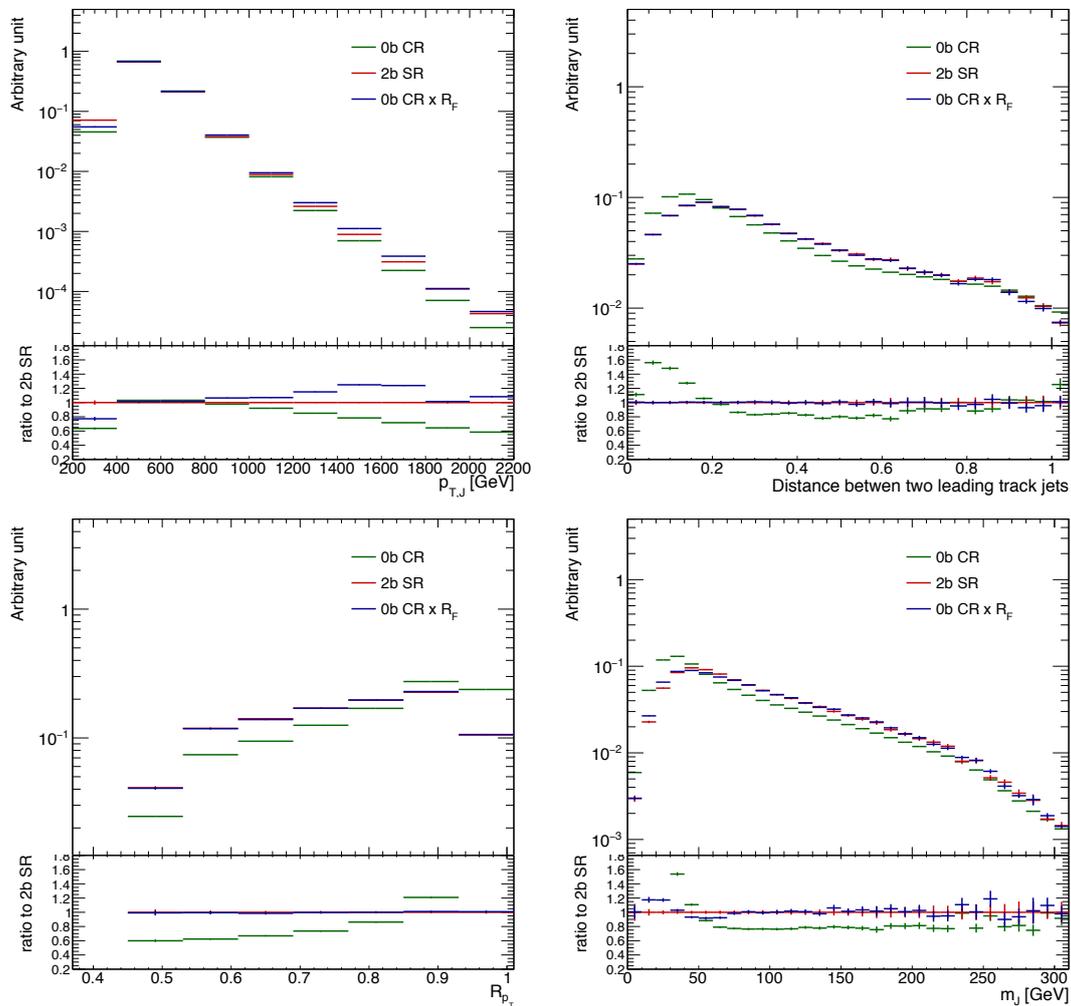


Figure 8.14: The effect of the 3-dimensional reweighting of the CR to the SR on the observables ( $p_{T,J}$ ,  $\Delta R$ ,  $\mathcal{R}_{p_T}$ ) for the distributions of  $p_{T,J}$ ,  $\Delta R$ ,  $\mathcal{R}_{p_T}$ , and  $m_J$  of the signal candidate. The distribution in the SR (2b) is compared to the distribution in the CR before (0b) and after ( $\times R_F$ ) reweighting. The bottom panel shows the ratio relative to the SR. The three distributions, ( $p_{T,J}$ ,  $\Delta R$ ,  $\mathcal{R}_{p_T}$ ), represent sanity checks that the reweighting is applied correctly, while the agreement in the  $m_J$  distribution (bottom right) corresponds to the figure of merit for the performance.

The presented studies using the multijet simulation constitute the proof of principle of the CR reweighting method is demonstrated. In order to assess the validity of the method in the presence of additional SM background processes or a potential resonance signal, signal contamination bias tests are performed.

### Signal Contamination Bias Test

The studies presented so far are based on MC simulations which contain the QCD process only. Yet, in data the CR and the SR will contain (semi-)resonant contributions from  $Z + \text{jets}$ ,  $W + \text{jets}$ ,  $H + \text{jets}$ ,  $t\bar{t}$ , and a potential  $Z'$  signal. Therefore, it is important to

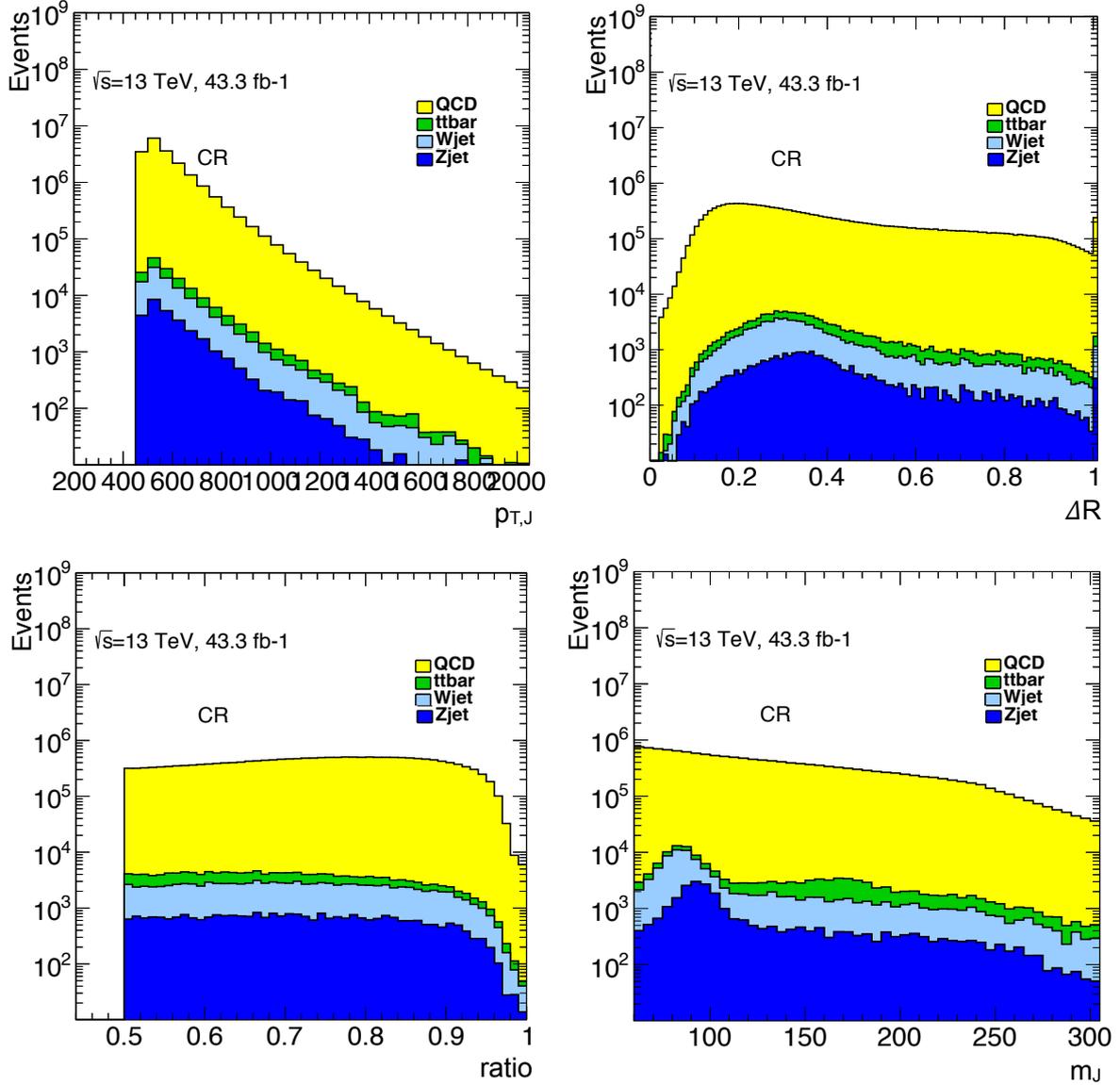


Figure 8.15: The distribution of the  $V + \text{jets}$  and QCD MJ backgrounds in the CR of the observables (a)  $p_{T,J}$ , (b)  $\Delta R_{bb}$ , (c)  $\mathcal{R}_{p_T}$ , and (d)  $m_{b\bar{b}}$  of the Higgs candidate.

verify that the contamination of the CR and the SR with these processes can not bias the reweighting procedure significantly. Therefore, two validation tests are performed with additional SM background processes and  $Z'$  resonance signal.

The first test is performed by considering the resonant contribution from  $V + \text{jets}$  and  $t\bar{t}$  processes in addition to the multijets, meaning that the  $R_F$  is derived as,

$$R_F^{\text{Signal}} = \frac{N_{SR}^{\text{QCD}} + N_{SR}^{V+\text{jets}} + N_{SR}^{t\bar{t}}}{N_{CR}^{\text{QCD}} + N_{CR}^{V+\text{jets}} + N_{CR}^{t\bar{t}}}.$$

The resulting composition of the CR in  $m_J$  and other kinematic variables is shown in Figure 8.15, while the corresponding distributions in the SR are shown in Fig. 8.16. The relative contribution of the  $V + \text{jets}$  and  $t\bar{t}$  process accounts for a non-negligible amount and its resonant contribution to  $m_J$  is clearly visible in the figures. Yet, its contribution to the reweighting observables shown in panels is non-resonant, and results in small bias.

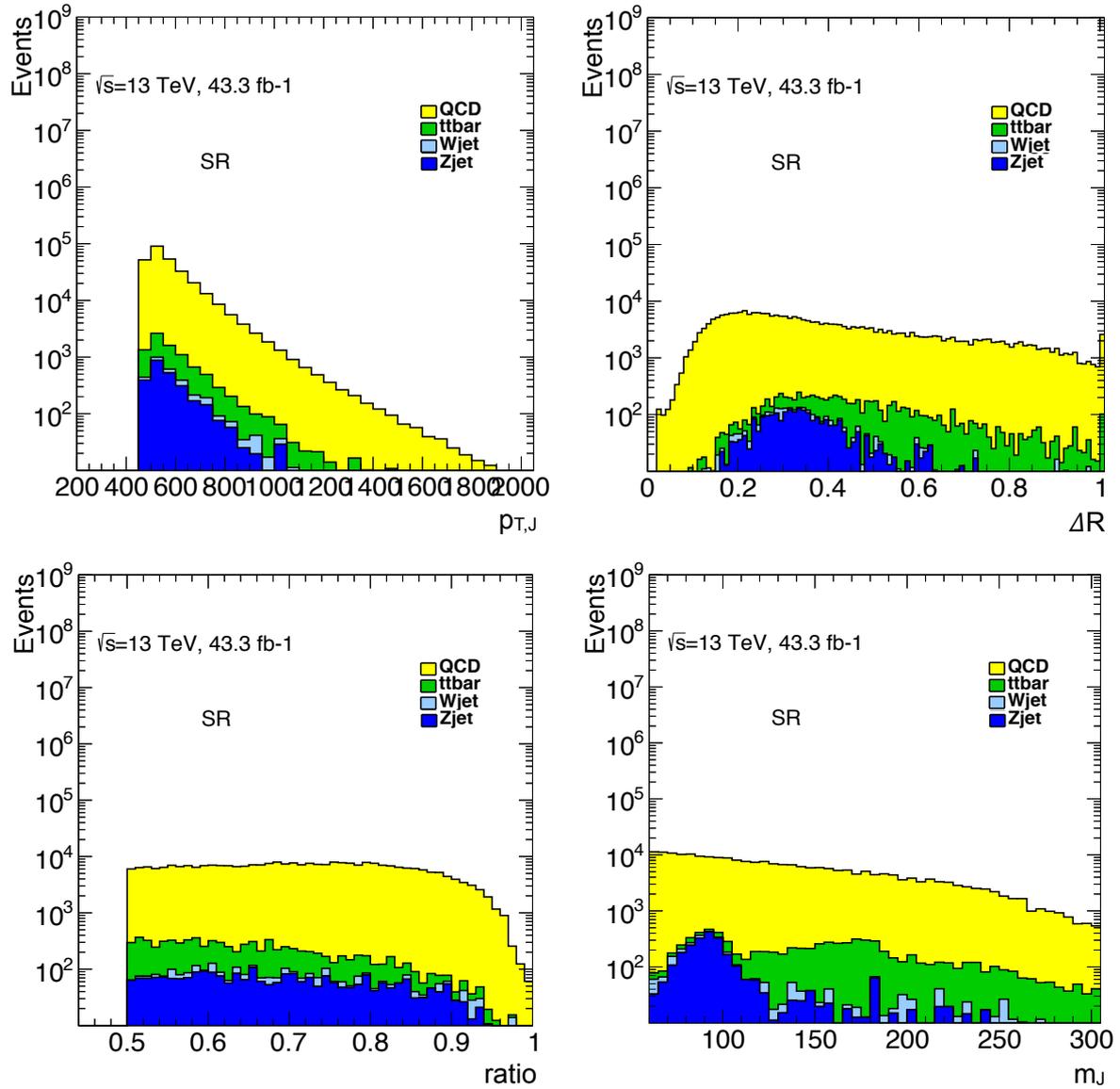


Figure 8.16: The distribution of the  $V + \text{jets}$  and QCD MJ backgrounds in the control region of the observables (a)  $p_{T,J}$ , (b)  $\Delta R_{bb}$ , (c)  $\mathcal{R}_{p_T}$ , and (d)  $m_{b\bar{b}}$  of the Higgs candidate.

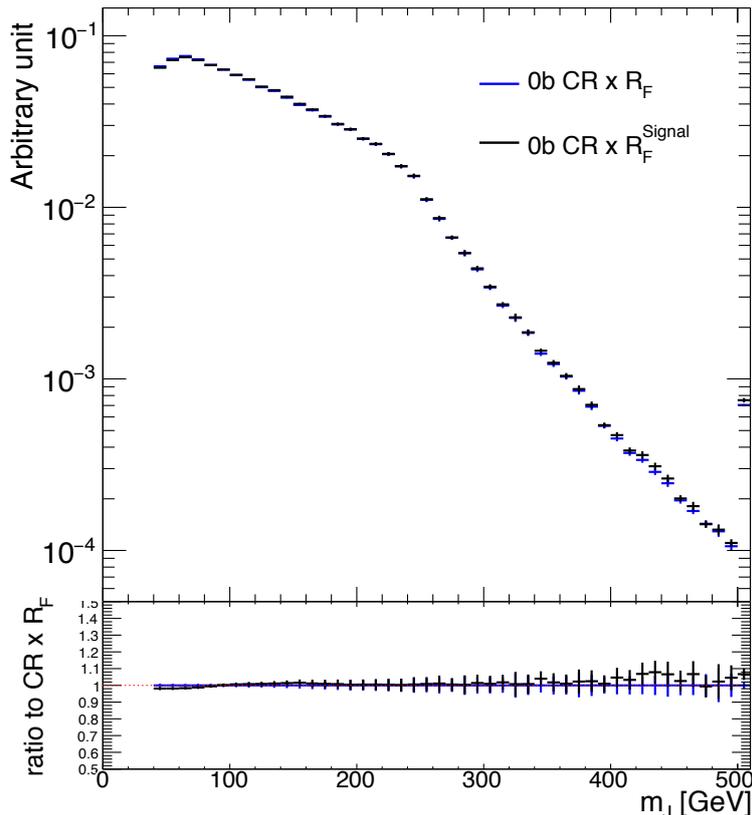


Figure 8.17: The multijet background prediction using the CR considering only multijet production when deriving the 3-dimensional reweighting factors ( $CR \times R_F$ ) is compared to the prediction where the reweighting is derived including the contribution from  $V + \text{jets}$  and  $t\bar{t}$  processes both in the CR and in the SR ( $CR \times R_F^{\text{Signal}}$ ). The bottom panel shows the ratios of the distributions with respect to the nominally reweighted  $CR \times R_F$  distribution.

Applying the same procedure with the  $R_F^{\text{Signal}}$  for CR reweighting as described earlier, the bias to the reweighting approach from the contamination of the resonant  $V + \text{jets}$  and  $t\bar{t}$  production is quantified and the result is shown in Fig. 8.17, where the multijet background prediction using the CR is derived as above, i.e., considering only multijet production when deriving the 3-dimensional reweighting factors in  $(p_{T,J}, \Delta R, \mathcal{R}_{pT})$ , is compared to the prediction where the reweighting is derived including the contributions from  $V + \text{jets}$  and  $t\bar{t}$  both in the CR and in the SR. The difference between the two multijet background predictions is very small at small  $m_J$ , and grows up to 6% with increasing mass values. This difference is attributed to the statistical fluctuations from the finite size of the MC samples, since there is no clear trend in the ratio of the two background predictions. However, the regions of  $m_J > 230$  GeV are not in the kinematic regime of this analysis.

Overall, the bias from the resonant  $V + \text{jets}$  and  $t\bar{t}$  contamination typically stays well below the size of the statistical uncertainty on the multijet background prediction indicated by the error bars of the distributions, and never exceeds it in the relevant kinematic range of  $40 \text{ GeV} < m_J < 225 \text{ GeV}$ . It can be concluded that the bias can be caused by the additional SM background processes is sufficiently small.

In order to check the bias that can come from a potential BSM signal, the same test is repeated with the simulated  $Z'$  with mass of 175 GeV. The result of the comparison for the nominal reweighting, and the reweighting derived from the merged sample of multijets

and  $Z'$  is shown in Figure 8.18.

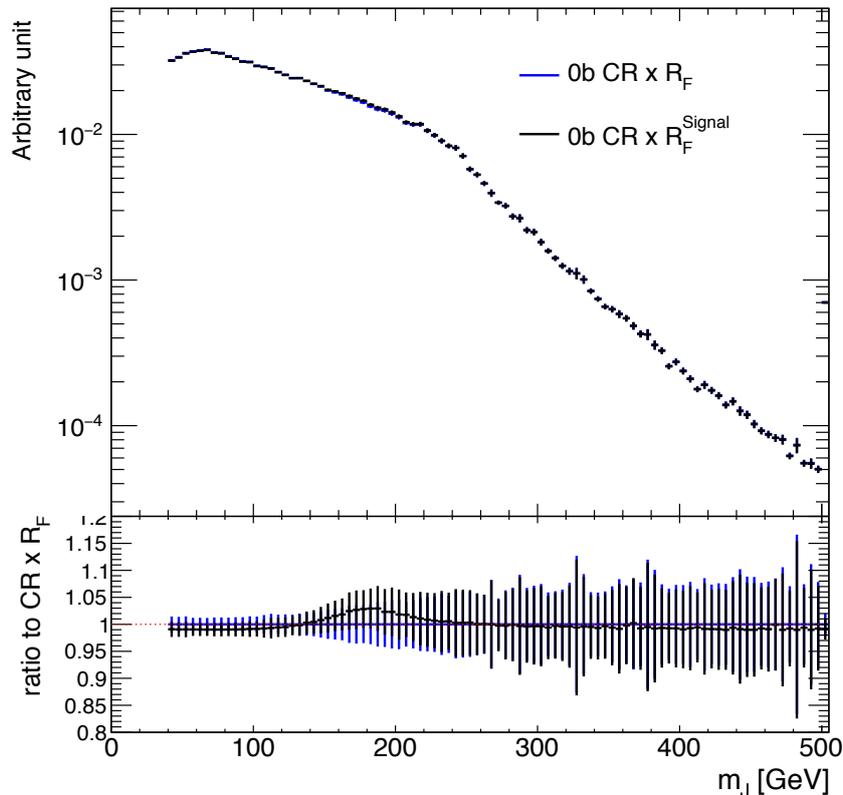


Figure 8.18: The multijet background prediction using the CR considering only multijet production when deriving the 3-dimensional reweighting factors ( $CR \times R_F$ ) is compared to the prediction where the reweighting is derived including the contribution from  $Z'$  both in the CR and in the SR ( $CR \times R_F^{Signal}$ ). The bottom panel shows the ratios of the distributions with respect to the nominally reweighted  $CR \times R_F$  distribution.

The injected  $Z'$  signal has an impact on the background estimation which is up to 5% relative to the nominal (QCD only) CR reweighting. Therefore, before proceeding any further, the impact of the signal contamination should be prevented. One of the alternative approaches in order to avoid bias from the signal contamination would be to remove the events in the  $(p_{T,J}, \Delta R, \mathcal{R}_{p_T})$  distributions that fall into the  $m_J$  region contaminated by a potential signal for a given mass hypothesis. Repeating the procedure for each mass hypothesis, it is possible to obtain several multijet  $m_J$  distributions in the sidebands of the each mass hypothesis. Although this approach is not studied in the context of this thesis due to time limitations, it can be said that once the potential biases are avoided, the CR reweighting method represents as a significant alternative to obtain the shape of  $m_J$  distribution of multijet background down to low mass values of around 50 GeV.

### 8.3.4 Direct Fitting Method

This analysis is performed for the first time in ATLAS, but it is very similar to the other dijet analyses as mentioned in Section 8.1.2. The direct fitting method is one of the reliable methods used before in those searches, and it has been found an appropriate choice for the first round of this analysis, mainly for practical reasons i.e. time. The major

challenge of this method is to find a suitable fit function describing the data distribution of the final analysis discriminant in the signal region. The idea is to treat this fit function as the combination of the different background and possible signal processes in the signal region. With this approach and the permit of the statistical analysis tools, it is possible to search for an excess over the spectrum of interest. Below, a more precise description of the fit procedure used in this analysis is given.

The main idea is to fit the data distribution of the signal candidate large- $R$  jet mass in the SR with the best possible fit functions among the studied ones. This fit describes the whole background contributions coming from multijets,  $W + \text{jets}$ ,  $Z + \text{jets}$ , and  $t\bar{t}$  processes in the SR of this analysis, as well as a possible new physics signal. In order to perform such a fit, several constraints are used as a result of the individual background studies. The final fit procedure used in the analysis is briefly summarised below:

- $W + \text{jets}$  and  $Z + \text{jets}$  processes are modelled using a single combined  $V + \text{jets}$  template that is obtained by summing the simulated  $W + \text{jets}$  and  $Z + \text{jets}$  samples.
- As described in Section 8.3.1, the  $t\bar{t}$  background is constrained with the correction factor 0.84.
- Multijet contribution is modelled using a parametric fit function. The polynomial exponential function and formal Laurent series, given below, are investigated as a candidate for the fit function.

– Polynomial exponential (nominal)

$$f_n(x|\vec{\theta}) = \theta_0 \exp\left(\sum_{i=1}^n \theta_i x^i\right), \quad x = \frac{m_J - 150 \text{ GeV}}{80 \text{ GeV}}, x \in [-1, 1], \quad (8.1)$$

– Formal Laurent series (alternate)

$$f_n(x|\vec{\theta}) = a \sum_{i=0}^n \frac{\theta_i}{x^{i+1}}, \quad a = 10^5, x = \frac{m_J + 90 \text{ GeV}}{160 \text{ GeV}}, x \in [1, 2], \quad (8.2)$$

– In order to test the candidate multijet fit functions and validate the background estimation method, 60 data slices are created from the CR containing approximately  $1.2 \text{ fb}^{-1}$  of data each. The size of the data slices is chosen such that each slice has a similar number of events as the SR. To take into account only the multijet background,  $V + \text{jets}$  and  $t\bar{t}$  MC templates are removed from the total data after scaling them correctly according to the luminosities of each slice and the extra correction factor for  $t\bar{t}$  template. Using the  $\chi^2$  test statistics to compare the agreement between the data and the fit functions, and performing a likelihood ratio test, in combination with Wilk's theorem [163], and the F-test [164], the exponential polynomial functions with five parameter model is chosen for the multijet background modelling. Although test results suggest that both the four and the five parameter models are equally good, the five parameter model is preferred in order to be conservative [117]. Figure 8.19 shows the corresponding fit results, indicating good agreement with the modelled dataset.

- $V + \text{jets}$  and  $t\bar{t}$  normalisations are kept as floating parameters in the final fit.

- The final fit function consists of 7 parameters: 5 parameters are for the multijet modelling, 1 parameter is for the  $V + \text{jets}$  normalisation, and 1 parameter is for the  $t\bar{t}$  normalisation. Note that the Higgs template is neglected for the final fit, as its contribution found to be small comparing to total uncertainties. Using the above procedure, the CR data is fitted with 7 parameters using  $1.2 \text{ fb}^{-1}$  dataset as shown in Figure 8.19.

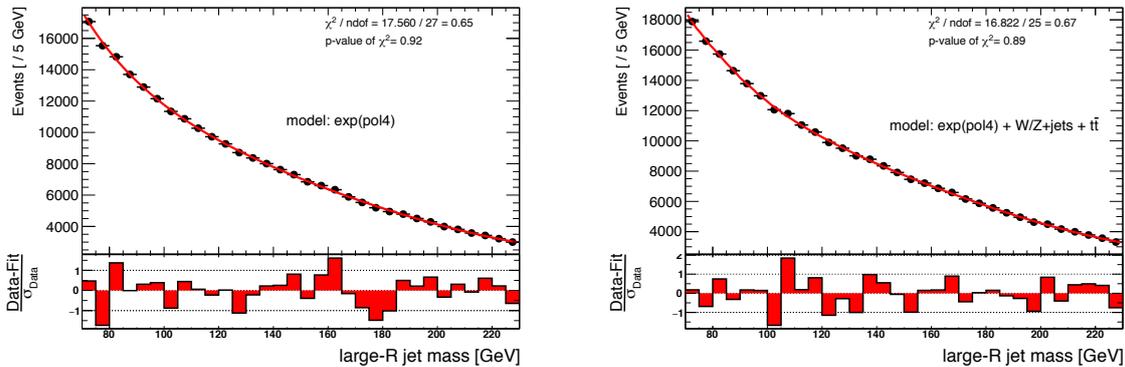


Figure 8.19: Multijet background fit with the exponential polynomial functions with five parameters (left), and the final template fit with 7 parameters (right) using a  $1.2 \text{ fb}^{-1}$  CR data. Both fits exhibit a low reduced  $\chi^2$  value and a high  $p$ -value for the  $\chi^2$ , indicating that the fit models are in an agreement with their fitted data [117].

As a result of these studies, obtained fit functions are further tested in the VR dataset of  $9.5 \text{ fb}^{-1}$ , which has a similar number of events as the SR. Figure 8.20 shows both the multijet only fit and the final fit function. As can be seen from the results, while final fit is in an agreement with its dataset, the multijet fit indicates an excess around  $V + \text{jets}$  mass values. As expected, the performed multijet fit with the CR data is less sensitive with respect to the VR data, since the amount of the  $V + \text{jets}$  background significantly differs in both region.

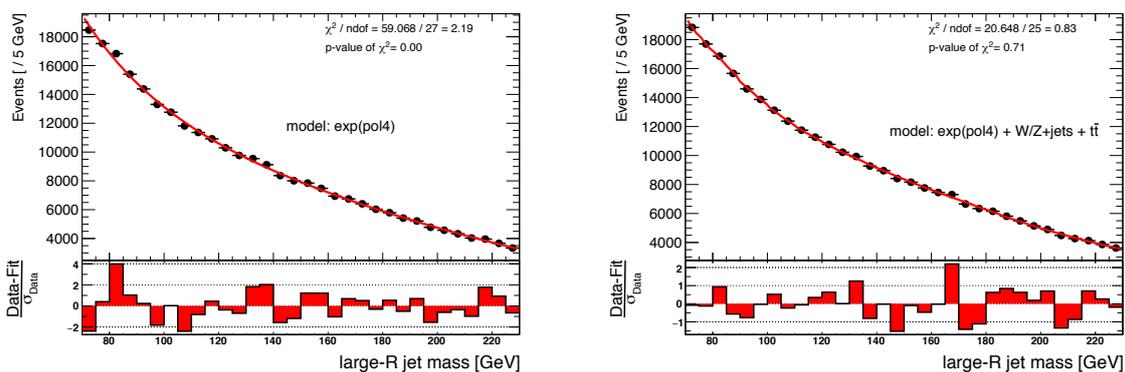


Figure 8.20: Multijet background fit with the exponential polynomial functions with five parameters (left), and the final template fit with 7 parameters (right) using a  $9.5 \text{ fb}^{-1}$  VR data. VR Fits with  $1.2 \text{ fb}^{-1}$  datasets. While final with indicates a goos agreement between data and the fit model, multijet fit shows an excess around  $V + \text{jets}$  [117].

The direct fitting method is further validated with spurious signal and signal injection tests, and the results found to be convincing since no significant bias or uncertainty due

to the model selection is observed [117]. Therefore, direct fitting is used as the final background estimation method of this analysis in this iteration. Although it is truly convenient to use this method since it yields information simultaneously both for the multijet background and the other SM processes, the method has a significant drawback: the limited fit range. Due to the non-smooth shape of the multijet distribution as shown in Figure 8.21, the fits fail below 70 GeV. The upper bound is chosen as 230 GeV due to the shape difference observed in the spectra around this value. This is a consequence of the boosted topology, and the applied  $p_T$  cut on the leading large- $R$  jets. Recalling the boost condition from Equation 5.6, the presence of the kink in the spectrum can be understood. For instance, a large- $R$  jet with  $p_T$  of 450 GeV can only be reconstructed as a large- $R$  jet if its mass is lower than the 225 GeV. Above that value, there would be 2 different small- $R$  jets creating the observed kink in the SR of this analysis. Consequently, range of the fit is determined as [70, 230] GeV.

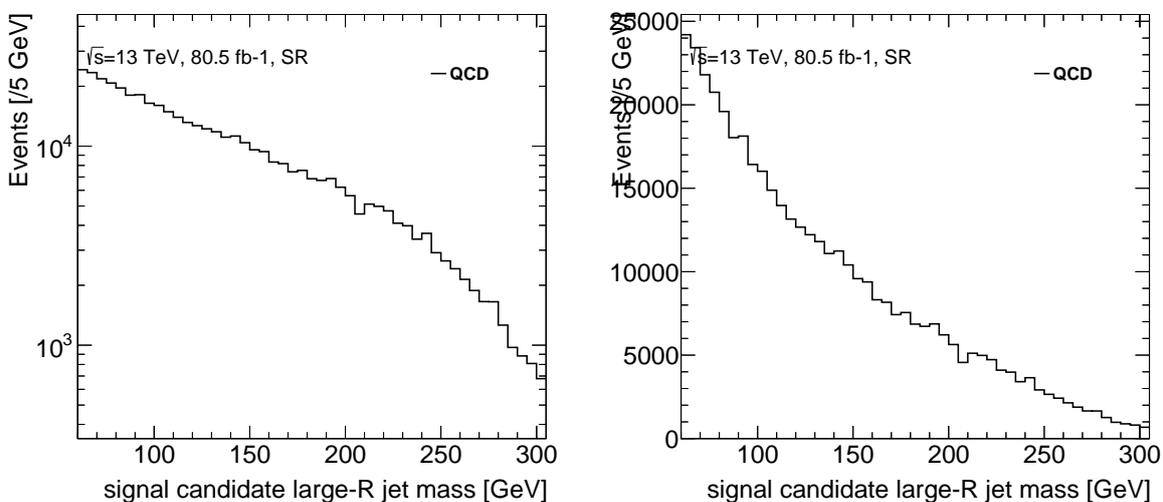


Figure 8.21: Multijet distribution obtained from the simulation is shown on the left in logarithmic scale and on the right in non-logarithmic scale, to reveal the non-smooth shape of the distribution.

## 8.4 Systematic Uncertainties

In high energy physics, systematic uncertainties arise from several sources like the measurement precision of the detector, the reconstruction and calibration techniques, the background estimation methods, the theoretical assumptions, the MC simulations etc. These uncertainties are as significant as the statistical ones and they must be handled with a great care.

In this analysis signal candidate large- $R$  jet mass distribution is measured, that is why the uncertainties on the measurements which alter this spectrum must be evaluated. Two types of impacts are expected on the measured spectrum due to the systematic uncertainties: the shape (modelling) uncertainties that contribute to the differential shape of the distribution, and the normalisation uncertainties that affects the overall yield of the distribution. Some of the uncertainty sources may have an impact on both of them as well. In Table 8.3, the important sources of systematic uncertainties are presented together with their impact types. Moreover, these impacts are propagated in the measurement of the signal strength ( $\mu$ ) and the impact on the signal strength is quantified as the difference

in quadrature between the uncertainty in  $\mu$  computed when all the other uncertainties are considered with their nominal values. In the following, the important sources of the systematic uncertainties are briefly explained.

Table 8.3: Summary of the impact of the main systematic uncertainties on the total uncertainty on the measurement of the signal strength  $\mu$  for the  $V + \text{jets}$ , Higgs boson and  $Z'$  signals. The impact of a systematic uncertainty is defined as the difference in quadrature between the uncertainty  $\sigma$  in  $\mu$  computed when all other uncertainties are considered and when are fixed to their pre-fit values. The total systematic uncertainty is then defined as the difference in quadrature between the total uncertainty in  $\mu$  and the total statistical uncertainty, denoted as  $\sqrt{\Delta\sigma^2}$  in the table [67].

Source	Type	Impact on Signals( $\sqrt{\Delta\sigma^2}/\mu$ )			
		V+jets	Higgs	Z' (100 GeV)	Z' (175 GeV)
Jet energy and mass scale	Norm. & Shape	15%	14%	23%	18%
Jet mass resolution	Norm. & Shape	20%	17%	30%	20%
$V + \text{jets}$ modeling	Shape	9%	4%	4%	< 1%
$t\bar{t}$ modeling	Shape	< 1%	1%	< 1%	11%
Flavor-tagging ( $b$ )	Normalisation	11%	12%	11%	15%
Flavor-tagging ( $c$ )	Normalisation	3%	1%	3%	5%
Flavor-tagging ( $l$ )	Normalisation	4%	1%	4%	7%
$t\bar{t}$ scale factor	Normalisation	2%	3%	2%	58%
Luminosity	Normalisation	2%	2%	2%	3%
Alternative QCD function	Norm. & Shape	4%	4%	3%	17%
$W/Z$ and QCD (Theory)	Normalisation	14%	–	–	–
Higgs (Theory)	Normalisation	–	30%	–	–

### 8.4.1 Luminosity

The total integrated luminosity is used to normalise the MC simulations. The uncertainty on the measurement of the total integrated luminosity is taken to be 2.1% following the recommendations of the ATLAS luminosity group [165].

### 8.4.2 Large- $R$ jet energy and mass uncertainties

#### Scale uncertainties

The systematic uncertainties related to the large- $R$  jet energy (JES) and mass (JMS) scale calibrations are derived by propagating uncertainties from the individual in-situ response measurements using a statistical combination [113]. The uncertainties are estimated following the  $R_{trk}$  approach, that is based on the comparisons of the kinematic observables measured in the calorimeter to those measured in the tracker [125]. While the details on this method can be found in the given reference, below the sources of systematic uncertainties that are obtained using this method are introduced in four categories [166].

- *Baseline*: the base difference between data and Pythia8.
- *Modelling*: the maximum difference between Pythia8 and Herwig7 or Pythia8 and Sherpa.

- *Tracking*: the uncertainties on the tracks being used as a reference.
  - Tracking1 : tracking efficiency and related uncertainties.
  - Tracking2 : tracking fake rate uncertainties.
  - Tracking3 : tracking  $q/p_T$  bias uncertainties.
- *TotalStat*: the statistical uncertainty on the measurement.

In Figures 8.22 and 8.23, the impact of  $2\sigma$  up and down variations of these uncertainties on the signal candidate large- $R$  jet mass distribution is shown for the simulated Higgs and  $Z'$  samples. The results indicate that JES+JMS uncertainties alter both the shape and the normalisation of the distribution. The *modelling* component is by far the most dominant uncertainty source. The second and third most dominant uncertainty source are the *tracking* and *baseline* uncertainties. Finally, the impact of *totalstat* component is observed to be almost negligible. The impact of  $1\sigma$  and  $3\sigma$  variations can be seen in Appendix C.4.

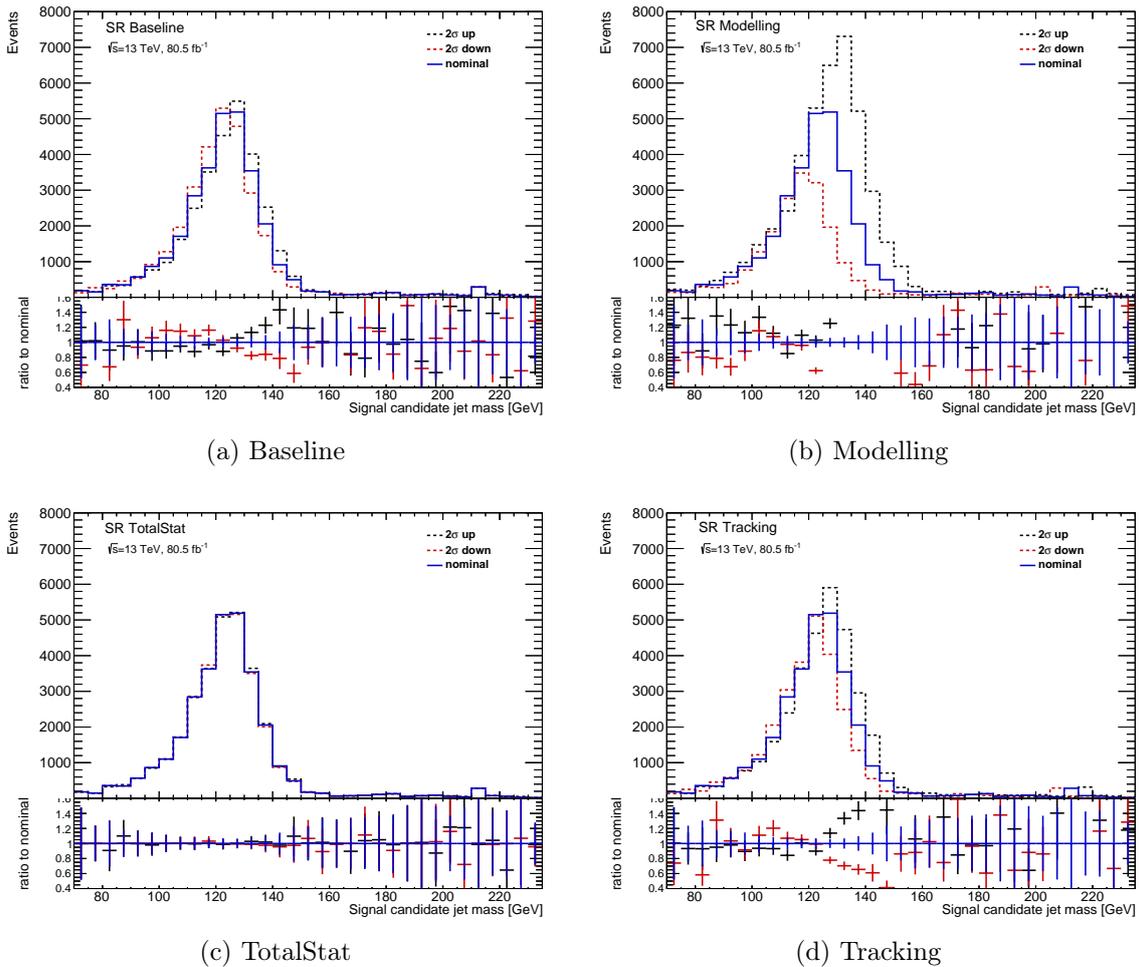


Figure 8.22: The impact of the  $2\sigma$  varied JES and JMS uncertainties on the signal candidate large- $R$  jet mass distribution for simulated Higgs samples.

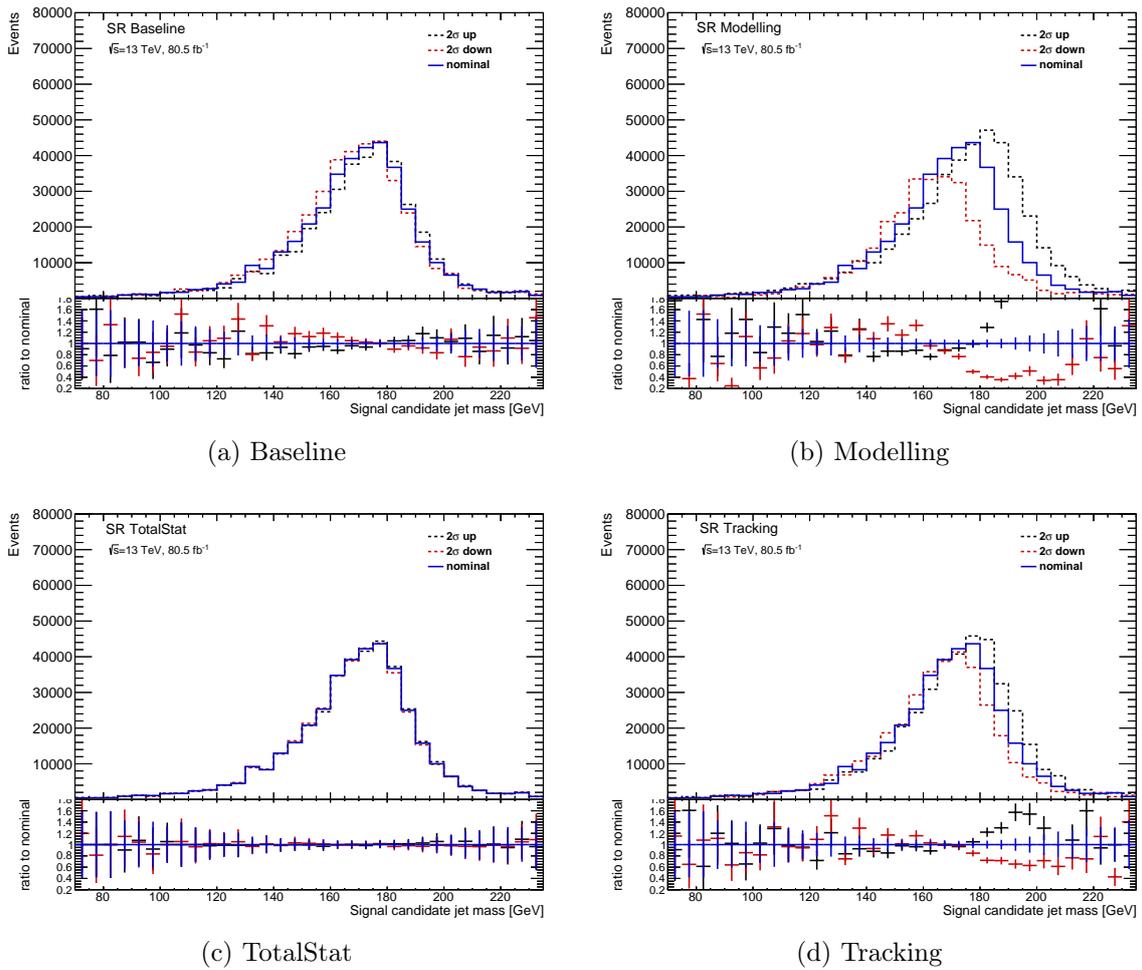


Figure 8.23: The impact of the  $2\sigma$  varied JES and JMS uncertainties on the signal candidate large- $R$  jet mass distribution for simulated  $Z'$  samples generated at 175 GeV mass.

### Resolution uncertainties

In order to take the correct modelling of the resolution into account, two sources of uncertainty, the jet energy resolution (JER), and the jet mass resolution (JMR) are introduced. Following the recommendations obtained from the ATLAS performance groups, a Gaussian smearing is applied using absolute 2% JER uncertainty and relative 20% JMR uncertainty [166]. In order to smear the energy resolution, the jet energy is multiplied with random values using a normalised Gaussian distribution with a mean of  $\mu = 1$  and a width of  $\sigma = 0.02$ . The resulting difference between the smeared and nominal distribution is taken as one sided systematic uncertainty and symmetrized. The effect is tested and found to be negligible.

The smearing for JMR requires the knowledge of the resolution in the MC simulation before smearing,  $\sigma_{\text{nominal}}$ . The procedure to obtain jet mass resolution can be recalled from Chapter 6. To illustrate the resolution difference between different processes,  $t\bar{t}$ , Higgs, and multijet resolutions are shown in Figure 8.24. Smeared JMR ( $\sigma_{\text{new}}$ ) with a relative uncertainty is obtained using the equation given below:

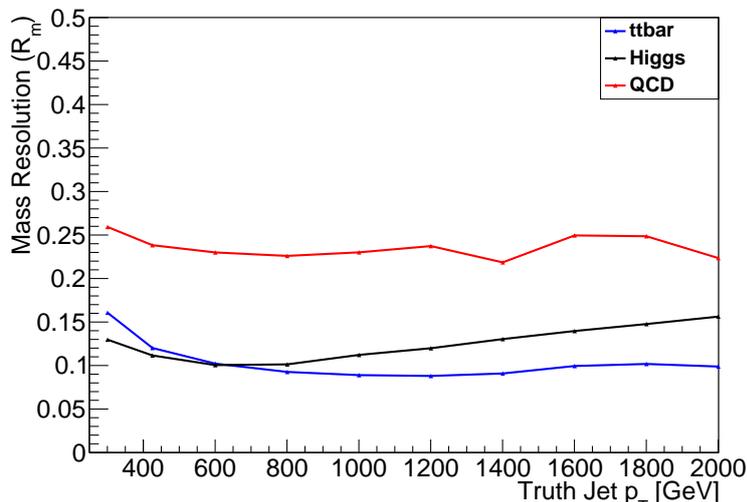


Figure 8.24: Combined jet mass resolutions for  $t\bar{t}$ , Higgs, and multijet processes are shown. The results are produced in the context of the  $h \rightarrow b\bar{b}$  tagging studies using single  $b$ -tagging benchmark. While the SM Pythia MC simulation is used for the multijet processes, BSM signal simulations are used to obtain  $t\bar{t}$  and Higgs processes as described in Chapter 4. Resolution, that is indicated as  $\sigma$  in the text, is obtained using the IQnR method.

$$\sigma_{\text{new}}^2 = \sigma_{\text{nominal}}^2 + (x \cdot \sigma_{\text{nominal}}^2). \quad (8.3)$$

In order to obtain a 20% increase in the mass resolution, the  $x$  parameter must be chosen as  $x = 0.66$ .

### 8.4.3 Flavour tagging uncertainties

The flavour tagging uncertainties arise from the calibration of the MV2c10 algorithm are parametrised with separate nuisance parameters for  $b$ ,  $c$  and light track jets. 25 nuisance parameters have been taken into consideration with their up and down variations, 9 for  $b$ -jets, 3 for  $c$ -jets and 11 for light quarks. The impact of each individual variation on the

$m_J$  spectrum is checked. In Figure 8.25, the most dominant component of the  $b$ -flavour uncertainty sources is presented for the different signal and background processes. As can be seen from the ratio plots in the figure, the ratios of the varied distributions to the nominal distributions are mostly flat. Similar results are observed for the other nuisance parameters as well, and they are presented in Appendix C.4. Therefore, the impact of the flavour tagging uncertainties are considered only for normalisation.

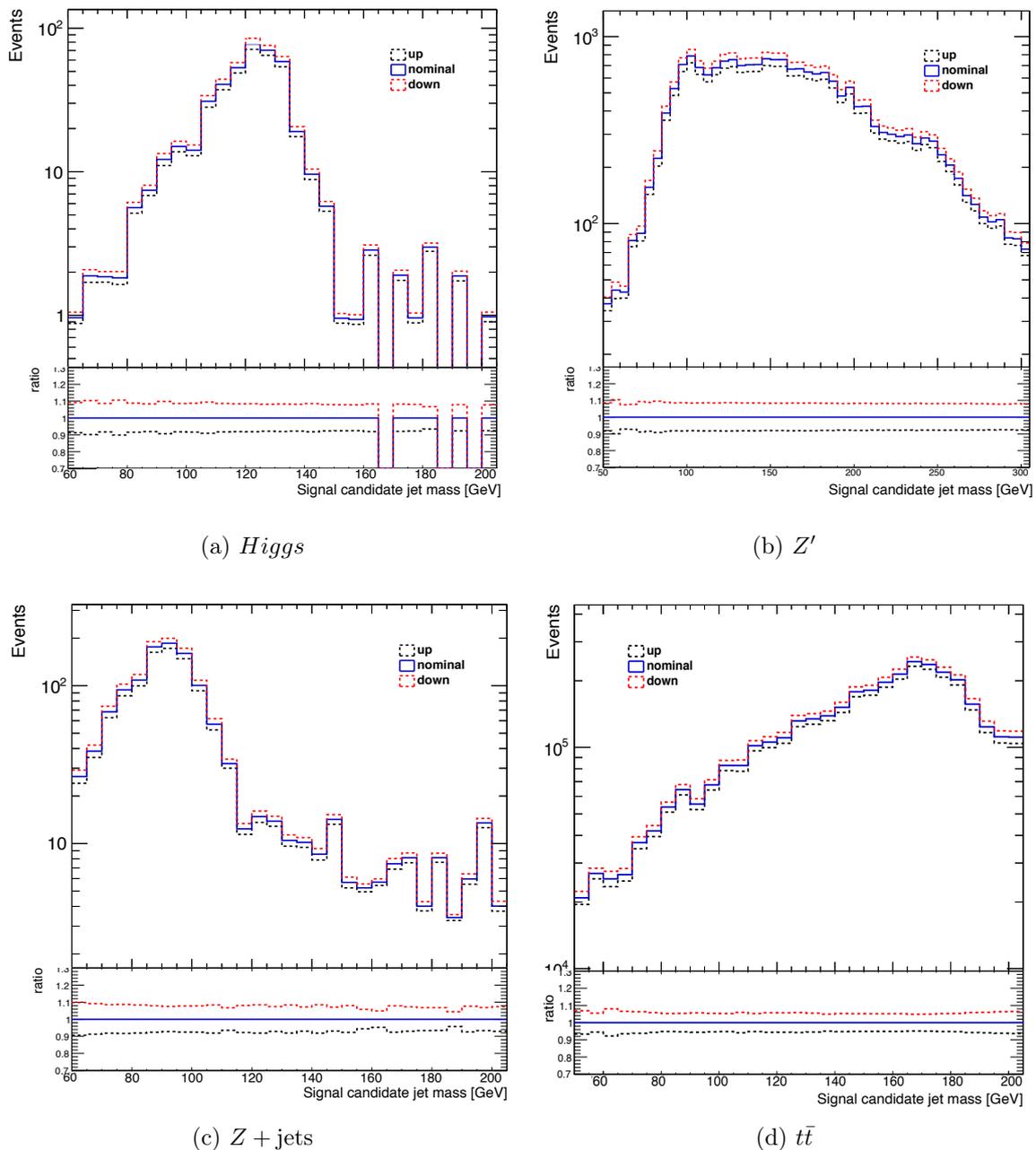
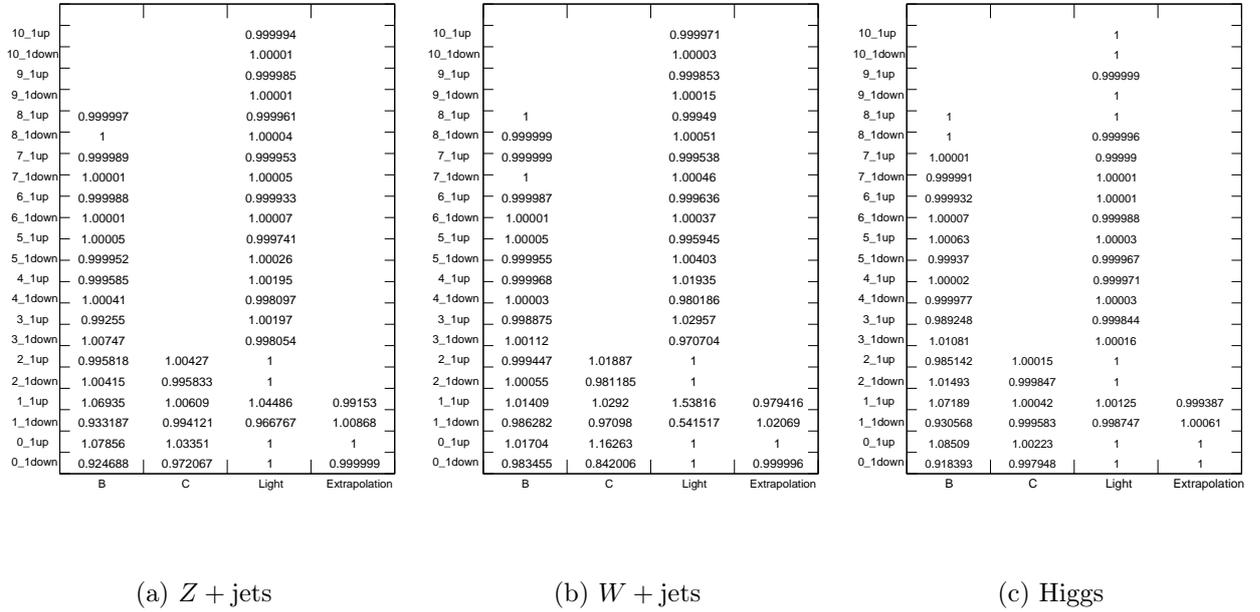


Figure 8.25: The impact of the  $1\sigma$  up/down variations of the most dominant  $b$ -flavour uncertainties on the signal candidate large- $R$  jet mass distribution is shown for the signal and resonant background samples. The impact on the  $Z'$  signal is shown after merging all the signal mass points.

For each nuisance parameter, the impact of the  $1\sigma$  up/down variations are quantified by calculating the ratio of the integrals for each systematically varied distribution to the nominal distribution, between the range of 70 GeV to 230 GeV. The results are shown

in Figure 8.26. The contributions of the nuisance parameters for each jet flavour are added in quadratures producing three normalisation uncertainties per flavour to use in the statistical analysis.

(a)  $Z + \text{jets}$ (b)  $W + \text{jets}$ 

(c) Higgs

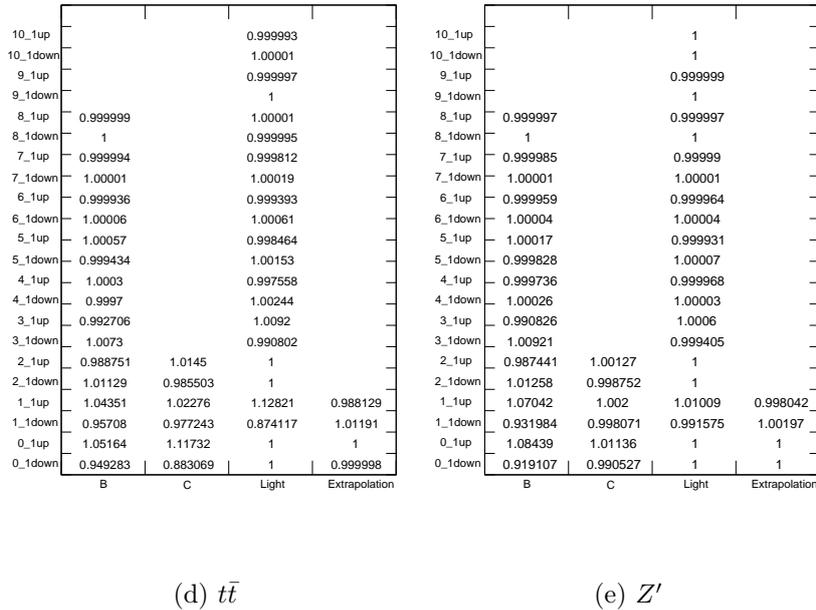
(d)  $t\bar{t}$ (e)  $Z'$ 

Figure 8.26: The impact of the  $1\sigma$  up/down variations of the nuisance parameters on the signal candidate large- $R$  jet mass distribution. Results are quantified by calculating the ratio of the integrals for each systematically varied distribution to the nominal distribution, between the range of 70 GeV to 230 GeV. The impact on the  $Z'$  signal is shown after merging all the signal mass points.

#### 8.4.4 Background and signal modelling uncertainties

For the multijet background estimation method, two different uncertainties are introduced. The first one is the uncertainty on the choice of the function to model the back-

ground, and the second one is the uncertainty on the parameters of the fit [117]. The uncertainty on the fit function is estimated using an alternative parametric fit function as given in Equation 8.2 instead of the polynomial function used in the analysis. Poissonian toys are generated from the nominal fit function and refitted with the alternative function. The impact of the uncertainties on the fit parameters are obtained by varying each parameter individually, while keeping the others constant. More details on this subject can be obtained from Reference [117].

The  $t\bar{t}$  scale factor that is extracted from the fit described in Section 8.3.1, has been measured with an uncertainty of 13%. This uncertainty is treated as a systematic uncertainty on the  $t\bar{t}$  normalisation.

In order to take into account the possible mismodelling of the large- $R$  jet mass shapes obtained from  $V$  + jets and  $t\bar{t}$  MC simulations, alternative MC generators are used. While for the  $V$  + jets samples generated using SHERPA 2.1.1, Herwig++ 2.7 is used as an alternative MC generator, for the  $t\bar{t}$  samples, SHERPA 2.2.1 is used instead of POWHEG-BOX 2.

Mismodelling of the top  $p_T$  is a known issue observed in several MC generators. Therefore, a  $p_T$  reweighting method which is derived using the available NNLO theory calculations [167] is explored for some ATLAS analyses [168], to improve the  $p_T$  modelling of the top jets. The effects of this approach is also investigated for this analysis in order to assess the necessity to introduce an additional uncertainty. Following the recommendations from the given reference [168], events are reweighted depending on their truth top jet  $p_T$  values. In Figure 8.27, the impact of this reweighting is shown for the signal candidate jet  $p_T$  and  $m$  distributions. As can be seen from the results, such a reweighting alters only the shape of the  $p_T$  distribution in the high  $p_T$  regimes, and has no impact on the shape of the mass distribution in the region of interest of this analysis. Thus, no additional systematic uncertainty is introduced.

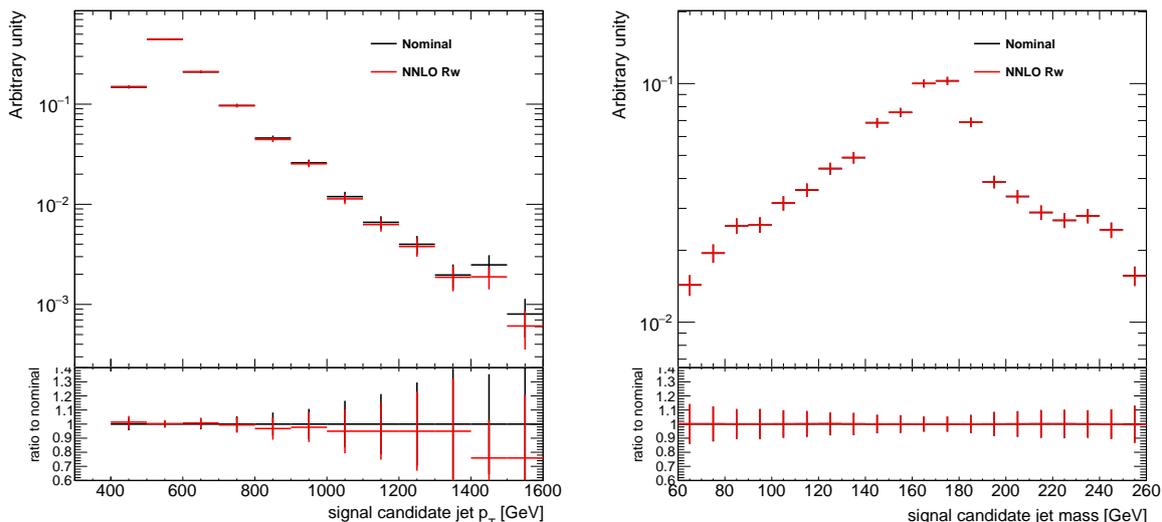


Figure 8.27: The impact of the NNLO top  $p_T$  reweighting [168] on the normalised signal candidate jet  $p_T$  and mass distributions in SR is shown. Results are obtained from the  $t\bar{t}$  MC simulations generated using POWHEG-BOX 2.

### 8.4.5 Theoretical uncertainties

The theoretical uncertainties for different Higgs production mechanisms are checked separately. As being the dominant one, the uncertainty prediction of 30% for the ggF production mode is taken as the total uncertainty on the cross section [67, 117]. The VBF and VH productions are obtained from Reference [86], and the ggF prediction is based on the incoming recommendations of the LHC Higgs Cross-Section Working Group.

The theoretical uncertainties on the  $V + \text{jets}$  cross sections are also considered following Reference [169], which occur due to the higher order electroweak and QCD corrections.

## 8.5 Statistical Framework

Primary goal of this analysis is to search for a new resonance peak in data, which could be observed as an excess over the expected background. In order to search for such an excess, data and the expected background must be compared and the differences between them should be quantified statistically. There are various methods and frameworks to perform such an evaluation, but this analysis uses an algorithm, so-called *BumpHunter* (*BH*) [170], which is designed to find possible excesses over the given spectrum and provide a measure of the significance of the largest excess using a *frequentist approach* [171, 172]. That stage of the statistical analysis is referred as *search phase* and described in Section 8.5.1.

In order to quantify a certain excess with respect to given parameters of a theory, a *Bayesian approach* [171, 172] is used within the framework called *Bayesian Analysis Toolkit* (*BAT*) [173]. Recalling that there are several expected peaks in the SR of this analysis, it is clear that this kind of an evaluation is needed, at least to validate the SM peaks and the background modelling. For instance using this approach and treating  $V + \text{jets}$  as signal, a statistical evaluation is performed and presented in Section 8.6.1.

In case there is no significant excess found in the analysed spectrum during the search phase with *BumpHunter*, the goal of the analysis becomes to narrow down the possible parameter space for the benchmark signal models, the  $Z'$  mediator, and to set upper limits on the quantity  $\sigma \times A \times BR$  as a function of resonant mass, where  $\sigma$  is the production cross-section of the signal,  $A$  is the signal acceptance and  $BR$  is the branching fraction for the  $b\bar{b}$  process. In addition to this quantity, results can also be interpreted in the plane of  $g_q - m_{med}$  by excluding certain  $g_q$  values for given signals generated with different  $m_{med}$  values. This part of the analysis is referred as *limit setting*, which is also based on the Bayesian approach using the *BAT* framework.

In this section, both statistical frameworks are briefly introduced. For more details, the given references can be followed [170, 173].

### 8.5.1 Search phase with *BumpHunter*

A test statistic is a variable computed from the data sample in order to discriminate between the two hypotheses [174]. The *BumpHunter* algorithm is also a test that searches an excess in the given spectrum by comparing the agreement between the data and the expected background using certain windows that consist of neighbouring bins. The window size gradually increases from two bins to half of the given spectrum, and the total excess or deficit for each possible window is calculated using the Poisson probability:

$$p - \text{value} = \begin{cases} \sum_{n=d}^{\infty} \frac{B^n e^{-B}}{n!}, & \text{for } b \geq d \\ \sum_{n=0}^d \frac{B^n e^{-B}}{n!}, & \text{for } d < b, \end{cases} \quad (8.4)$$

where  $d$  is the total number of data events, and  $b$  is the total number of background events in the window. Using the smallest local  $p$ -value, the test statistics  $t$  is calculated:

$$t = -\log(p - \text{value}_{\min}). \quad (8.5)$$

Using the  $t$  value, the global  $p$ -value is calculated in order to find the largest discrepancy at any position in the given spectrum. With the ability to find the exact location of the largest excess and to quantify its significance, this test statistics provides a model independent evaluation for a possible excess. For a significant excess, the global  $p$ -value should be smaller than 0.01, meaning that this discrepancy could be caused by statistical fluctuations of the background, only with the probability of less than 1% [175].

### 8.5.2 Limit setting with BAT framework

The Bayesian Analysis Toolkit (BAT) is used in order to evaluate the significance of the existent peaks and setting limits for the  $Z'$  mediator mass. In the Bayesian approach, the aim is to assign probabilities to the parameters of the theory considering that the data is fixed, unlike the frequentist approach where it calculates the probability of obtaining a certain data under the fixed parameters of the theory.

Providing each parameters with their prior distributions and the likelihood functions ( $\mathcal{L}(\nu, \theta | \text{Data})$ ), the marginalised posterior distributions ( $p(\nu, | \text{Data})$ ) produced within the BAT framework as a function of number of signal events,  $\nu$  [173]. Nuisance parameters, indicated as  $\theta$ , represent the sources of the systematic uncertainties on the signal hypothesis. Using the Bayes theorem, the posterior probability density can be written as following:

$$p(\nu, \theta | \text{Data}) = \frac{\mathcal{L}(\nu, \theta | \text{Data}) \pi(\nu, \theta)}{p(\text{Data})}, \quad (8.6)$$

where prior probability density for the signal normalisation is  $\pi(\nu)$ , and the prior probability densities for the nuisance parameters are  $\pi(\theta)$ . Integrating the the posterior  $p(\nu, \theta | \text{Data})$  over  $\theta$  parameters, one can get the marginalised posterior  $p(\nu, | \text{Data})$ .

Using the integrated marginalised prior over the number of signal events, the credibility level (C.L.) is determined. Depending on the chosen C.L., the upper limit for the possible number of signal events in data can be calculated. For most of the particle physics analyses, C.L. is chosen as 95%, which provides 95% assurance that the number of signal events in data are less or equal then  $\nu_{upper}$  [175].

$$0.95 = \int_{-\infty}^{\nu_{upper}} p(\nu, | \text{Data}) d\nu. \quad (8.7)$$

## 8.6 Results

This section briefly presents the analysis results. Section 8.6.1 shows the results of the background validation studies performed using the  $V + \text{jets}$  peak. Then Section 8.6.2 and Section 8.6.3 present the results of the measurement of the boosted SM Higgs boson production and the search for a leptophobic  $Z'$  mediator in the context of the simplified dark matter models, respectively.

### 8.6.1 Validation of the background modelling with $V + \text{jets}$

It is important to validate the background estimation method in the SR before proceeding for the search phase with the BumpHunter algorithm and the boosted Higgs measurement, besides the tests performed within the VR and CR. In the following, the background modelling is validated through observation of the SM  $V + \text{jets}$  process in the SR. If the test result is positive, meaning that at least  $3\sigma$  significance is achieved for  $V + \text{jets}$  peak, the background modelling can be considered as reliable and the analysis can proceed to the search phase.

Using the Bayesian approach presented in Chapter 8.5, and considering the  $V + \text{jets}$  as signal, a combined fit is performed with the parametric QCD fit and the  $t\bar{t}$  background constrained with the scale factor of 0.84 as described in Section 8.3.1. The corresponding fit results are shown in Figure 8.28. The ratio of number of observed events to the number of expected events is defined as signal strength ( $\mu$ ), and it is obtained as  $\mu_V = 1.3 \pm 0.2$  [117]. Under the assumption of 10% uncertainty on the  $V + \text{jets}$  production cross section, the observed statistical significance is estimated to be around  $6.5\sigma$ , indicating that the background modelling is reliable.

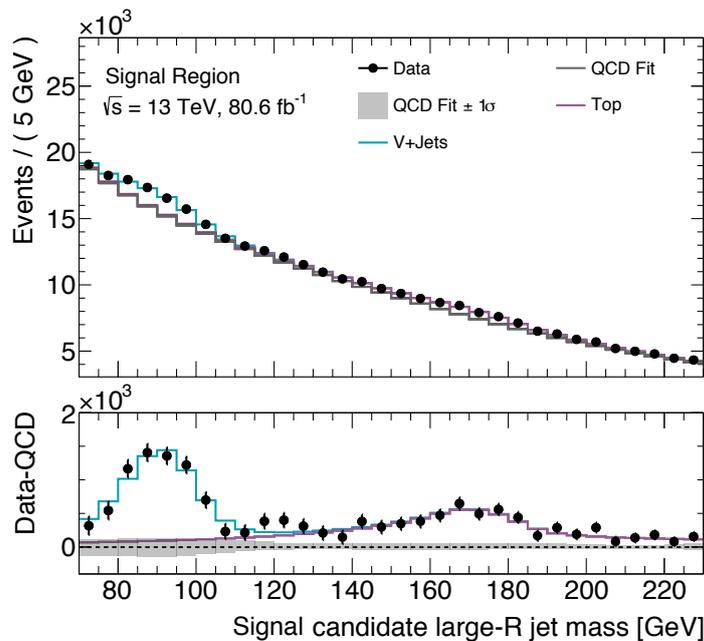


Figure 8.28: Postfit plots of the  $V + \text{jets}$ ,  $t\bar{t}$  and QCD fit comparison to data [117].

### 8.6.2 Measuring the boosted Higgs production

In order to measure the SM boosted Higgs production, a combined fit is performed with the  $V + \text{jets}$ ,  $H + \text{jets}$ ,  $t\bar{t}$ , and multijet components. While the  $t\bar{t}$  normalisation is constrained, both the  $V + \text{jets}$  and  $H + \text{jets}$  normalisations are kept as floating parameters and obtained simultaneously from the fit. Considering the systematic uncertainties as nuisance parameters for the likelihood function as introduced in Equation 8.6, the marginalised posterior distributions are calculated. In Figure 8.29, the results are shown after marginalisation.

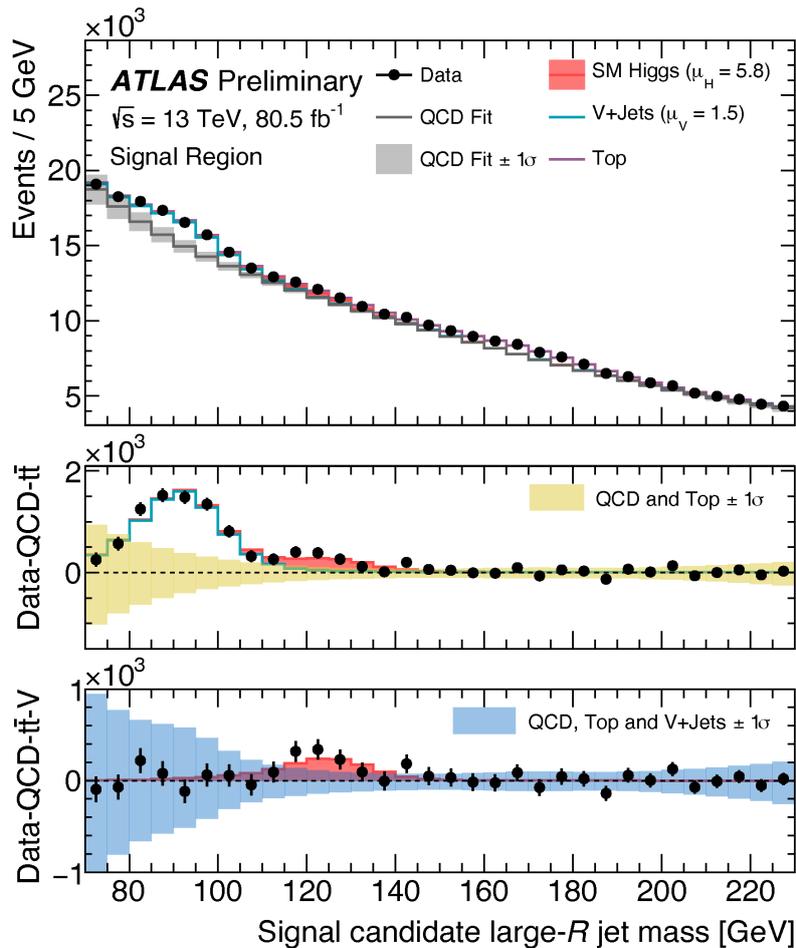


Figure 8.29: Postfit plot of the SM Higgs boson,  $V + \text{jets}$ ,  $t\bar{t}$  and QCD fit comparison to data. The middle panel shows the postfit and data distributions with the QCD and  $t\bar{t}$  components subtracted. The lower panel shows the same distributions when also the  $V + \text{jets}$  component is subtracted [67].

Following results are obtained for the  $V + \text{jets}$  and Higgs process:

- The observed signal strength is estimated as  $\mu_V = 1.5 \pm 0.22$  (stat.)  $^{+0.29}_{-0.25}$  (syst.)  $\pm 0.18$  (th.) with  $5\sigma$  statistical significance for  $V + \text{jets}$ .
- For the boosted Higgs boson production, the observed signal strength is measured as  $\mu_H = 5.8 \pm 3.1$  (stat.)  $\pm 1.9$  (syst.)  $\pm 1.7$  (th.) which corresponds to a significance of  $1.6\sigma$ . Result is consistent with the background-only hypothesis.

### 8.6.3 Searching the $Z'$ mediator

As described in Section 8.5.1, the BumpHunter search phase is performed to look for possible new physics signals in the large- $R$  jet mass spectrum. Since the impact of the Higgs boson is found negligible with respect to the expected uncertainty on the  $Z'$  limits,  $H + \text{jets}$  processes are not included to the performed fit for the search phase. Including the full set of systematic uncertainties, best fit values are obtained from the BumpHunter algorithm [117]. In Figure 8.30, the result is shown. The largest deviation from the background is found around 125 GeV with a small significance and global  $p$ -value of 0.54.

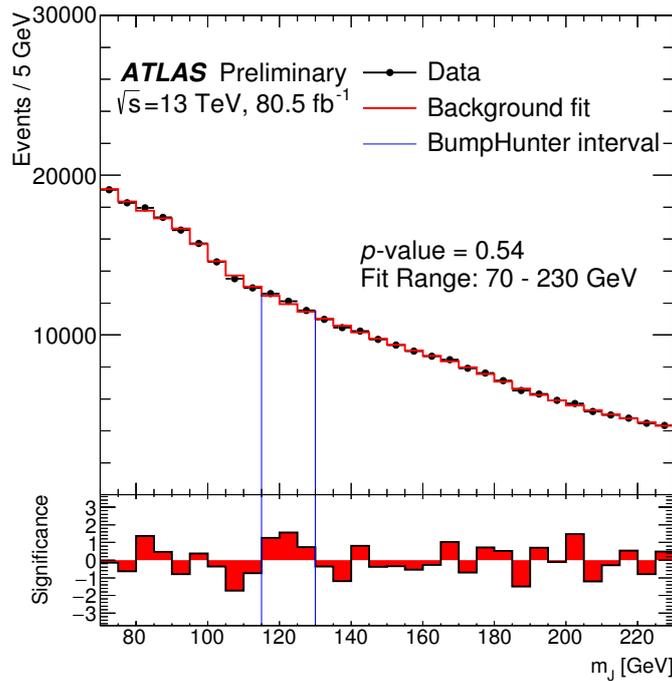


Figure 8.30: The result of the search phase with BumpHunter algorithm (see Section 8.5.1) is shown for the reconstructed mass distribution  $m_J$ . The solid red line depicts the background prediction, consisting of the non-resonant dijet,  $V + \text{jets}$  and  $t\bar{t}$  processes. The vertical blue lines indicate the most discrepant interval identified by the BumpHunter algorithm. Without including systematic uncertainties, the probability that fluctuations of the background model would produce an excess at least as significant as the one observed in the data anywhere in the distribution, the BumpHunter probability, is 0.54. The low panel shows the bin-by-bin significances of the differences between the data and the fit, considering only statistical fluctuations [67].

No significant excess is observed, therefore upper limits are set on the quantity  $\sigma \times A \times BR$ , as well as the quark coupling,  $g_q$ , parameter as a function of the  $Z'$  resonant masses between 100 GeV and 200 GeV, as presented in Figure 8.31. The analysis excludes  $Z'$  masses between 100 GeV and 200 GeV for the leptophobic  $Z'$  mediator with  $g_q = 0.25$ .

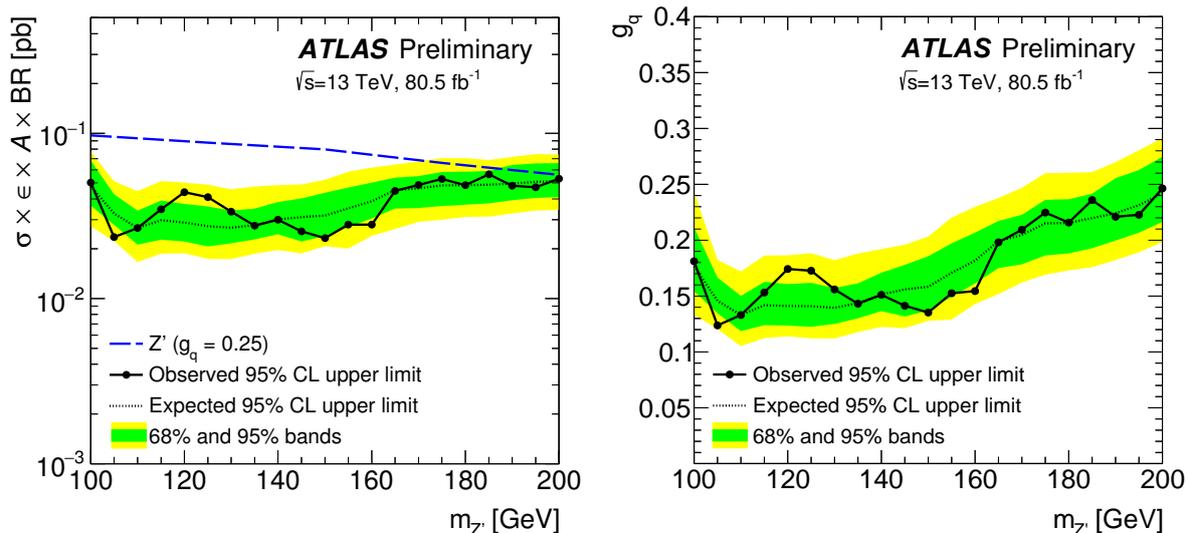


Figure 8.31: The 95% credibility-level upper limits obtained from the invariant mass distribution on (left) the cross-section times acceptance times branching ratio times efficiency for the leptophobic  $Z'$  with  $g_q = 0.25$  and on (right) the  $g_q$  parameter [67].

## 8.7 Conclusions

In this chapter, a search for boosted  $b\bar{b}$  pairs in association with an additional energetic initial state radiation jet using 80.5 fb $^{-1}$  ATLAS  $pp$  collision data at  $\sqrt{s} = 13$  TeV is presented. The analysis searches for an excess in the invariant mass spectrum of  $b\bar{b}$  pair, between the range of 70 GeV and 230 GeV. No significant excess is observed over the search range and upper limits are set at 95% CL on the production cross section times branching fraction times acceptance times efficiency and on the coupling parameter  $g_q$  for the leptophobic  $Z'$  mediator model.

Moreover, the analysis measured the SM boosted Higgs boson production in the gluon fusion mode for the first time in ATLAS. The observed signal strength is measured as  $\mu_H = 5.8 \pm 3.1$  (stat.)  $\pm 1.9$  (syst.)  $\pm 1.7$  (th.), corresponding to a significance of  $1.6\sigma$ .

The analysis validates the effectiveness of the used methods by measuring the signal strength of SM  $V + \text{jets}$  process. The signal strength of  $V + \text{jets}$  is obtained as  $\mu_V = 1.5 \pm 0.22$  (stat.)  $^{+0.29}_{-0.25}$  (syst.)  $\pm 0.18$  (th.), corresponding to a significance of 5 standard deviations.

Similar to the analysis presented in Chapter 7, this analysis uses boosted identification techniques. The combined large- $R$  jet mass as introduced in Chapter 6 serves as the final discriminant of this analysis playing the most crucial role to identify a potential excess. The positive impact of using this observable has already been studied and presented in previous chapter. One of the other improvements for the boosted particle identification in ATLAS Run 2, the variable- $R$  track jets are used in this analysis to identify  $b\bar{b}$  pairs. It has been shown that the usage of these jets improves the sensitivity of this analysis by increasing the signal selection efficiencies in high  $p_T$  regions.

# Chapter 9

## Conclusions and outlook

In this work, focusing on the process of boosted Higgs boson decays in  $b\bar{b}$  channel, two data analyses are presented using ATLAS Run 2  $pp$  collision data at  $\sqrt{s} = 13$  TeV. While both of the searches are motivated primarily to find new physics, they also convey significant information for future Standard Model measurements.

The presented search for boosted Higgs boson pair production in a  $b\bar{b}b\bar{b}$  final state explored the new physics possibilities for extra dimensions and additional new scalars. No sign of new physics observed and the analysis constrains a considerable portion of the targeted phase space by exceeding the search sensitivity of the previous analyses performed in this channel. Significant sensitivity improvements are obtained especially in the high mass resonance regions of the analysis, thanks to the improved analysis techniques and increased dataset. The combination of the boosted and resolved analysis provide stronger limits on the production cross section times branching ratio for the  $4b$  final state.

In addition to the BSM aspect, the measurement of the SM nonresonant Higgs boson pair production carries great significance in obtaining information on the Higgs sector. The resolved analysis provided limits which corresponds to 12.9 times the SM expectation. This result is further used in combination with the other  $HH$  channels in different final states and led to the improved limits as 6.7 times the SM expectation [145]. While the boosted analysis is statistically limited in the high mass regions, both analyses have systematic uncertainties arising from the data driven background estimation methods. Therefore, the increased datasets will improve the search sensitivities significantly.

The second analysis presented within this thesis is a search for boosted resonances decaying to two  $b$ -quarks in association with an additional jet, addressing to one of the biggest mysteries of today's particle physics, the nature of dark matter. Targeting the lower mass regions of the dijet resonance mass spectrum to find new physics signatures, this heavy-flavoured search provides complementary results to dijet resonance searches. No sign of new physics is observed and limits are set on the quark couplings to a new leptophobic dark matter mediator particle,  $Z'$ .

Furthermore, using the gluon fusion production mode of the Higgs boson, the signal strength of the boosted Higgs boson production in the  $b\bar{b}$  channel is measured for the first time in ATLAS. The  $W$  and  $Z$  bosons are measured with a significance of 5 standard deviations validating the SM expectations. The presented results can be further improved using different analysis techniques for the future iterations of the analysis: An alternative background estimation method was presented in this thesis as a reference for future studies.

Both of the presented analyses greatly benefit from the improved boosted identification techniques although they target different phase spaces. As shown in this thesis, these techniques are not only used to reach the higher mass regions but also to probe the lowest accessible kinematic regimes.

# Appendix A

## Boosted $h \rightarrow b\bar{b}$ identification

### A.1 High mass tail identification

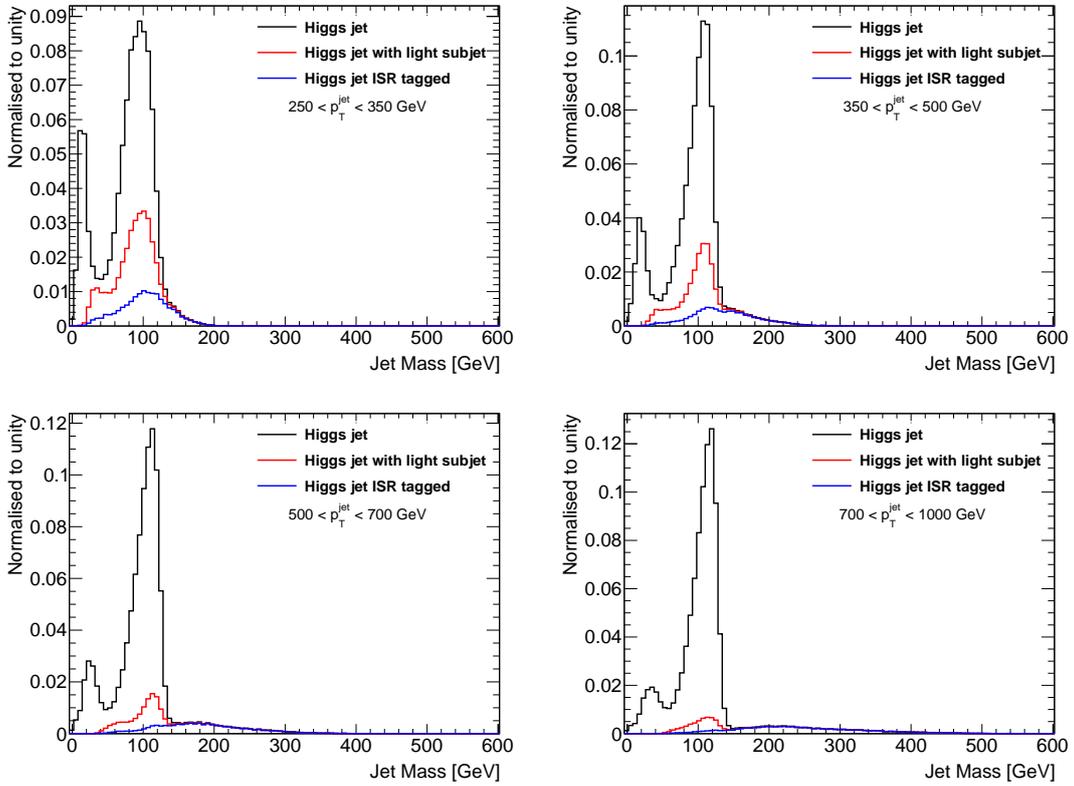


Figure A.1: Invariant mass distributions for Higgs jets with  $p_T > 250$  GeV and  $|\eta| < 2.0$  before any selection (black), after the requirement that at least one of the subjets should not contain a ghost-associated  $b$  quark (red), and after the additional requirement of the subjets not associated to a  $b$ -quark to contain ghost-associated particles from ISR (blue). The results are shown for different Higgs jets  $p_T$  regimes. In all kinematic regions shown the high mass tail is dominated by ISR.



# Appendix B

## Search for boosted di-Higgs production in the $b\bar{b}b\bar{b}$ final state

### B.1 Performance of $m^{\text{calo}}$ and $m^{\text{comb}}$

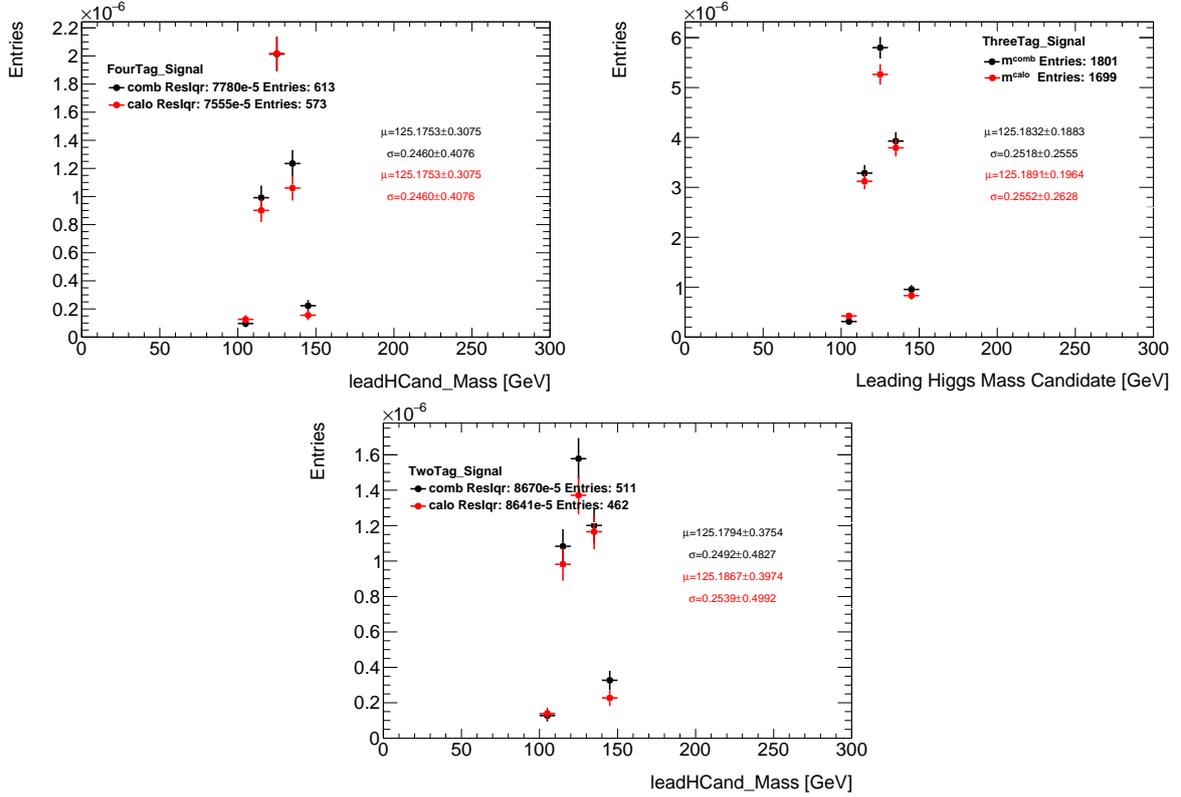


Figure B.1: Leading Higgs candidate jet mass distribution comparison in three SRs for  $m^{\text{comb}}$  and  $m^{\text{calo}}$  jet mass, shown for simulated  $G_{KK}$  signal at 5 TeV. ResIqr: resolution of the jet mass peak measured with IQnR method as described in Chapter 6.

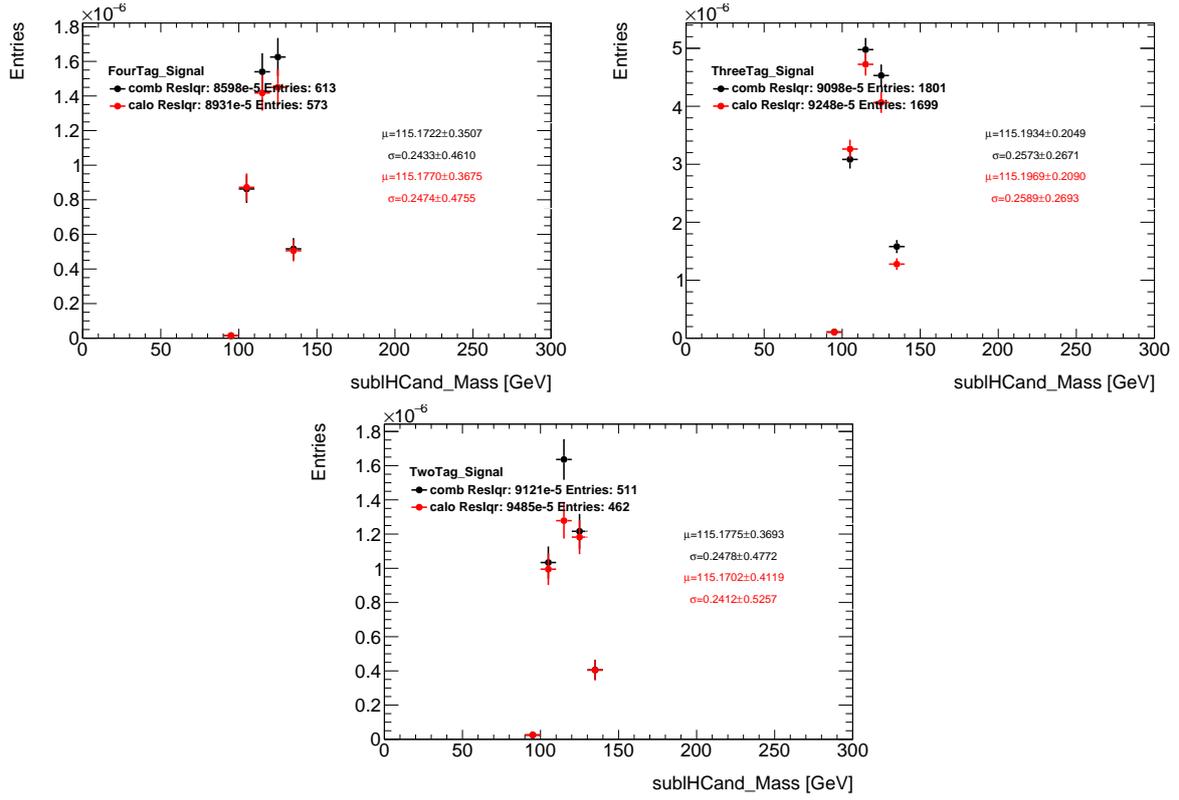


Figure B.2: Subleading Higgs candidate jet mass distribution comparison in three SRs for  $m^{\text{comb}}$  and  $m^{\text{calo}}$  jet mass, shown for simulated  $G_{KK}$  signal at 5 TeV. Reslqr: resolution of the jet mass peak measured with IQnR method as described in Chapter 6.

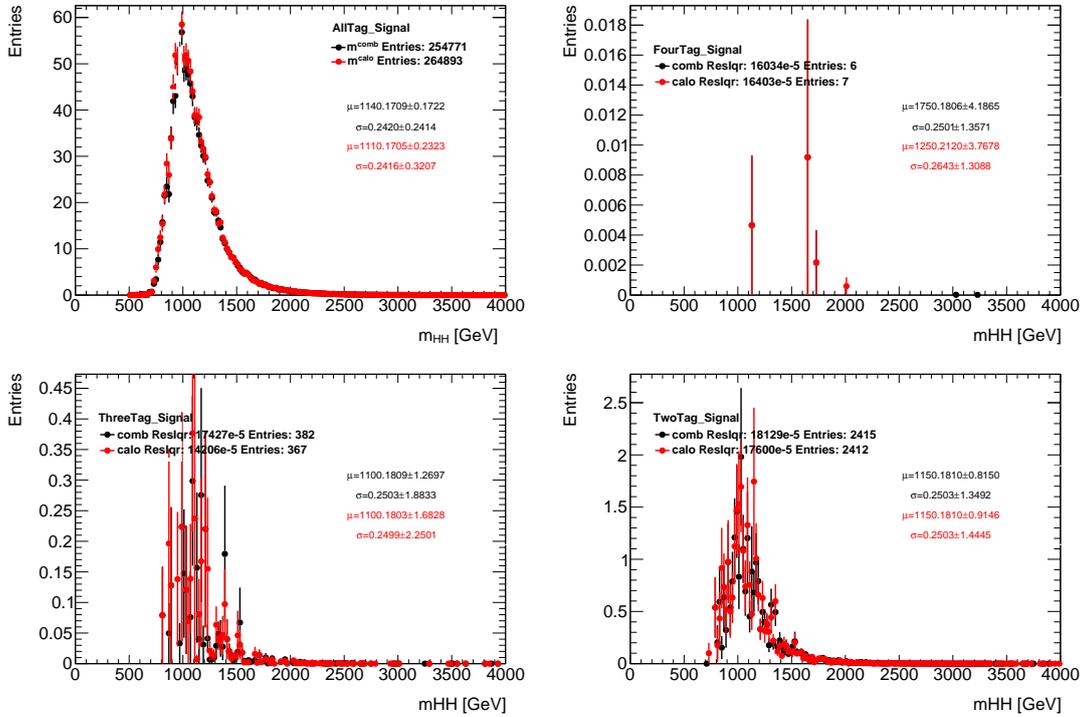


Figure B.3:  $m_{HH}$  observable comparison in three SRs and in an inclusive (Alltag) category for  $m^{\text{comb}}$  and  $m^{\text{calo}}$  jet mass, shown for multijet simulation. Alltag category shows all the events passing the SR selection without making any  $b$ -tagged track jet requirement.

## B.2 Background prediction using MC simulations

Note that the 2bs distributions refer to 2-tag region in the following. The background predictions using simulation for the 2-tag region before applying any reweighting have been shown in the Figure B.6 for all the track jet  $p_T$  distributions. The predictions stand between the 0b region and the 2bs regions, as expected from the background modelling method. Addition of the 1bL and 1bSL provides a partial recovery for the shape differences between 0b and 2bs regions. However, the predictions still need to be reweighted. As a result of those studies, it has been confirmed that similar reweighting is needed for dijet MC, in a similar manner as data.

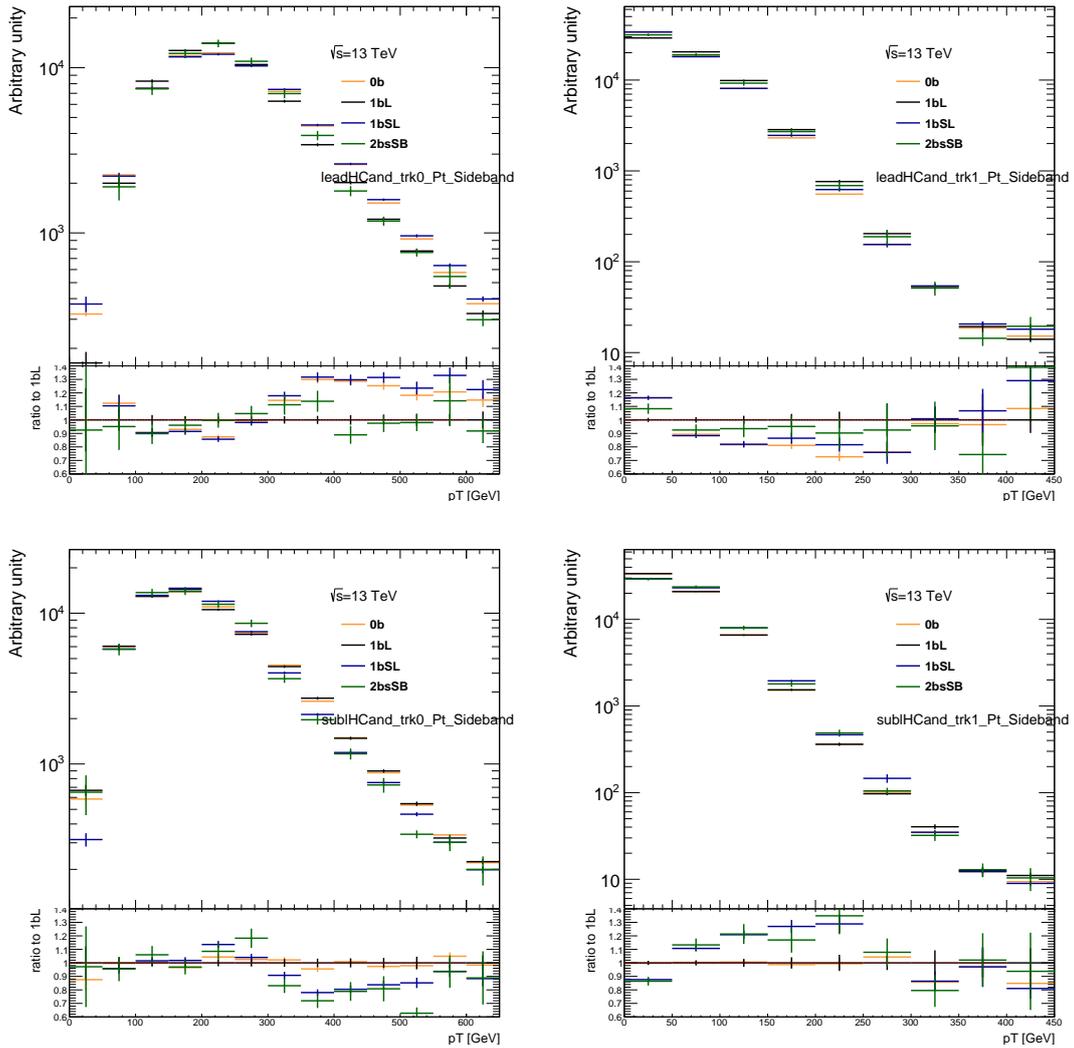


Figure B.4: Track jet  $p_T$  distributions in Dijet MC. (Top left) Leading track jet on leading Higgs candidate, (top right) subleading track jet on leading Higgs candidate, (bottom left) leading track jet on subleading Higgs candidate and (bottom right) subleading track jet on subleading Higgs candidate have been shown for the SB. All distributions are normalised according to the 2bsSB region, which corresponds to 2-tag category in the sideband.

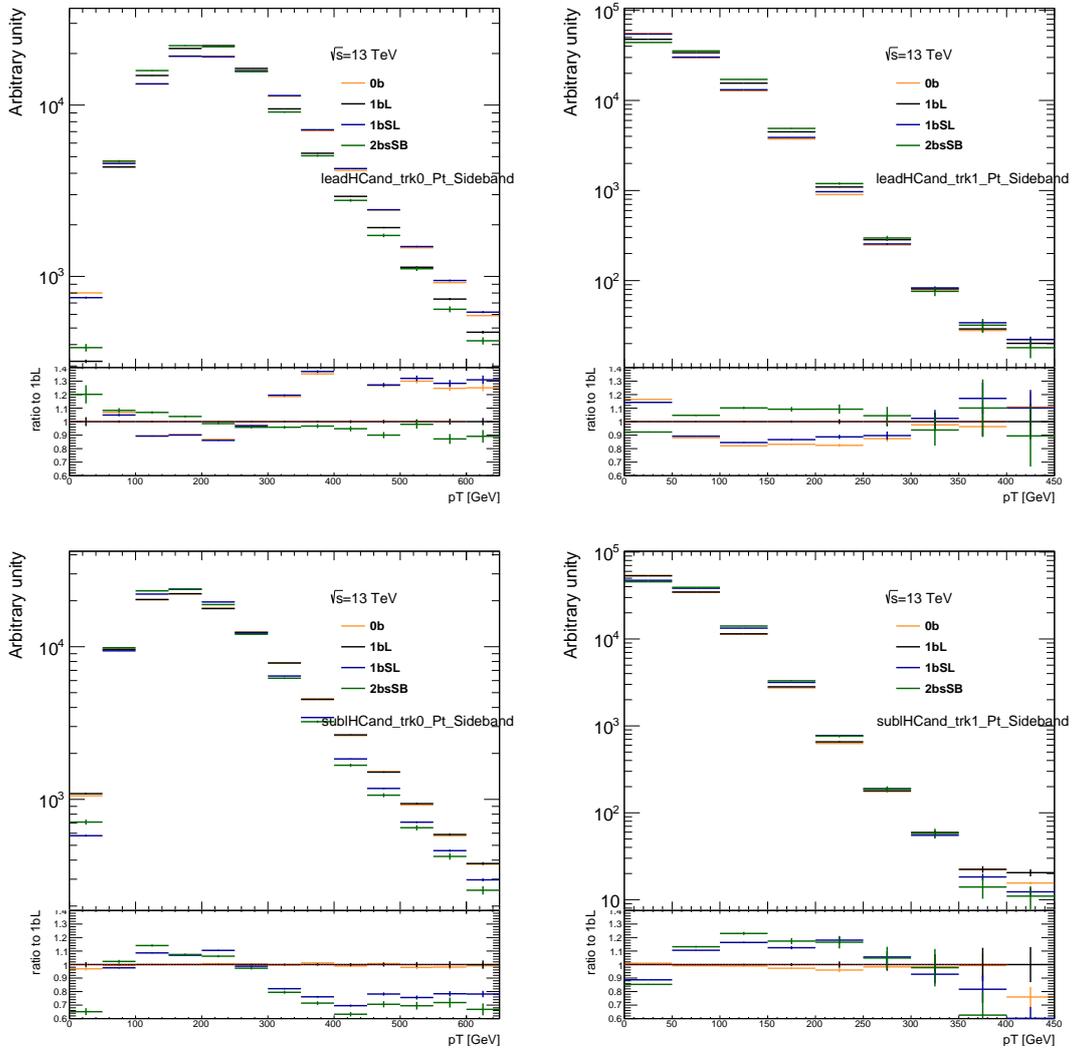


Figure B.5: Track jet  $p_T$  distributions in Data. (Top left) Leading track jet on leading Higgs candidate, (top right) subleading track jet on leading Higgs candidate, (bottom left) leading track jet on subleading Higgs candidate and (bottom right) subleading track jet on subleading Higgs candidate have been shown for the different regions as described in the text. All distributions are normalised according to the 2bs region.

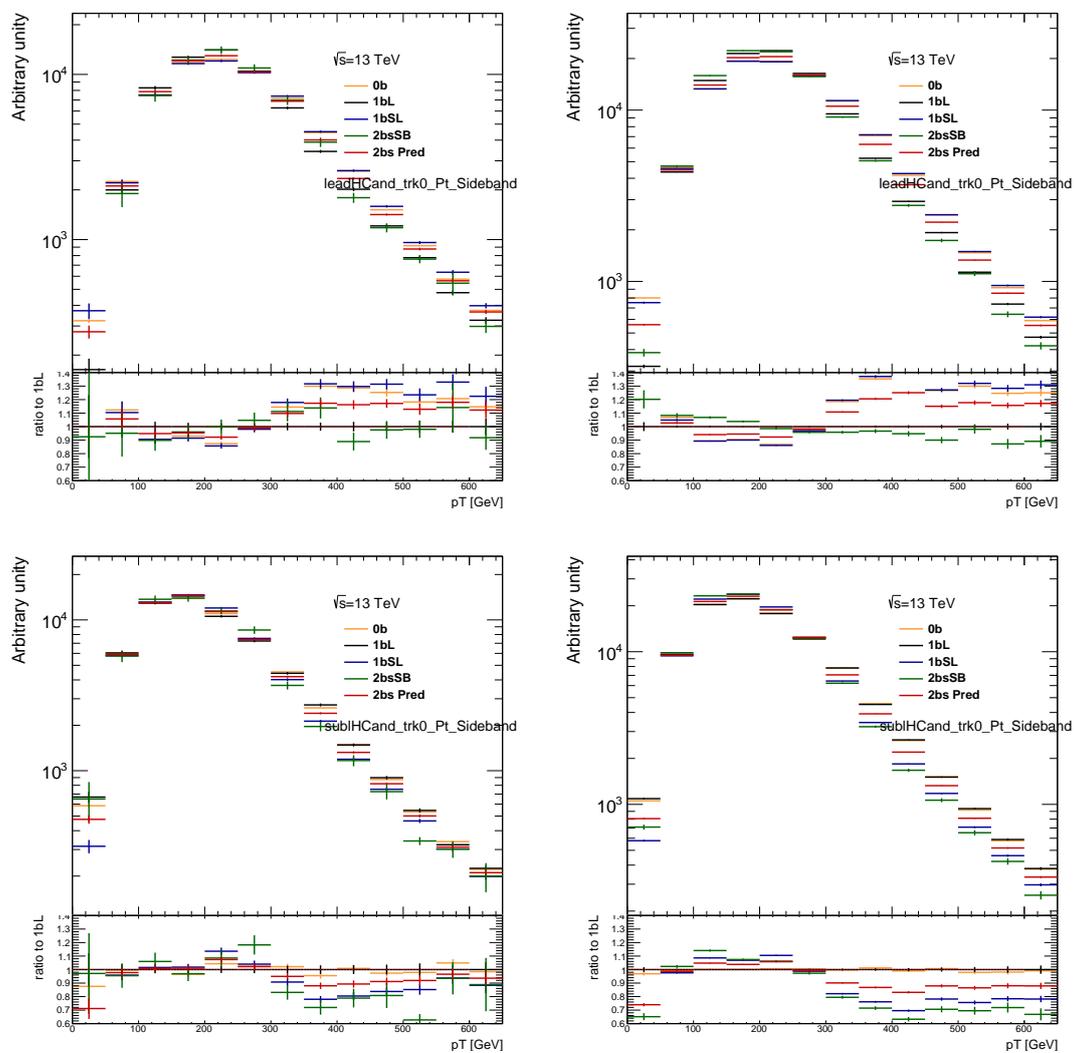


Figure B.6: Leading track jet  $p_T$  distributions together with the 2bs prediction as a comparison for the dijet MC (left) and the data (right). (First row) Leading track jet on leading Higgs candidate, (second row) subleading track jet on leading Higgs candidate, (third row) leading track jet on subleading Higgs candidate and (fourth row) subleading track jet on subleading Higgs candidate have been shown for the different regions as described in the text. All distributions are normalised according to the 2bs region.

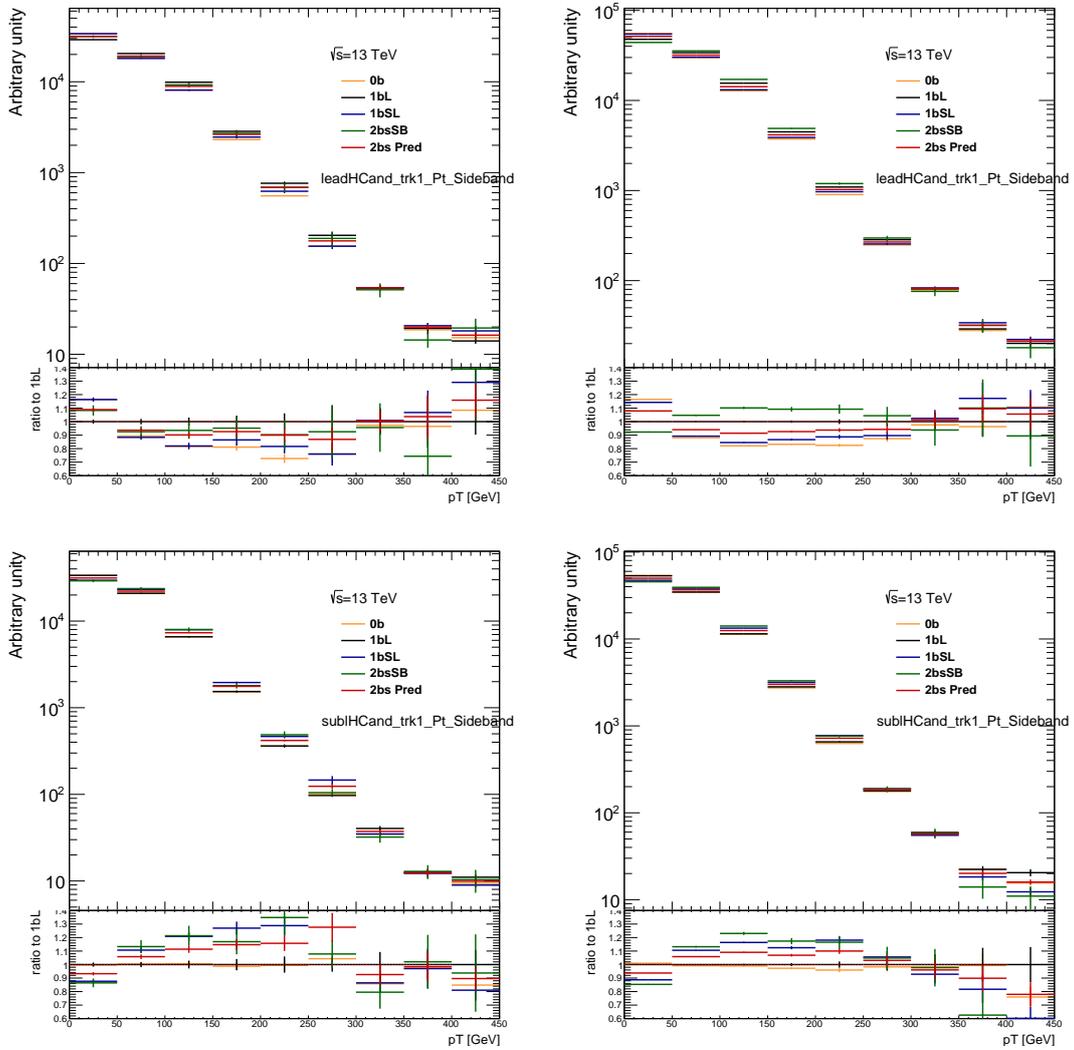


Figure B.7: Subleading track jet  $p_T$  distributions together with the 2bs prediction as a comparison for the dijet MC (left) and the data (right). (First row) Leading track jet on leading Higgs candidate, (second row) subleading track jet on leading Higgs candidate, (third row) leading track jet on subleading Higgs candidate and (fourth row) subleading track jet on subleading Higgs candidate have been shown for the different regions as described in the text. All distributions are normalised according to the 2bs region.

## B.3 Performance of the $b$ -tagging algorithm in different kinematic regimes

Figure B.8 and B.9 shows the kinematic dependencies of the  $b$ -tagging efficiencies.

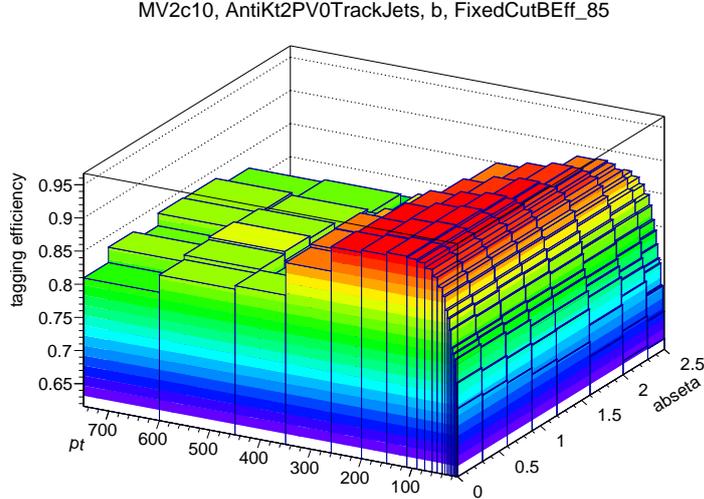


Figure B.8:  $b$ -tagging efficiencies with respect to eta and  $p_T$ .

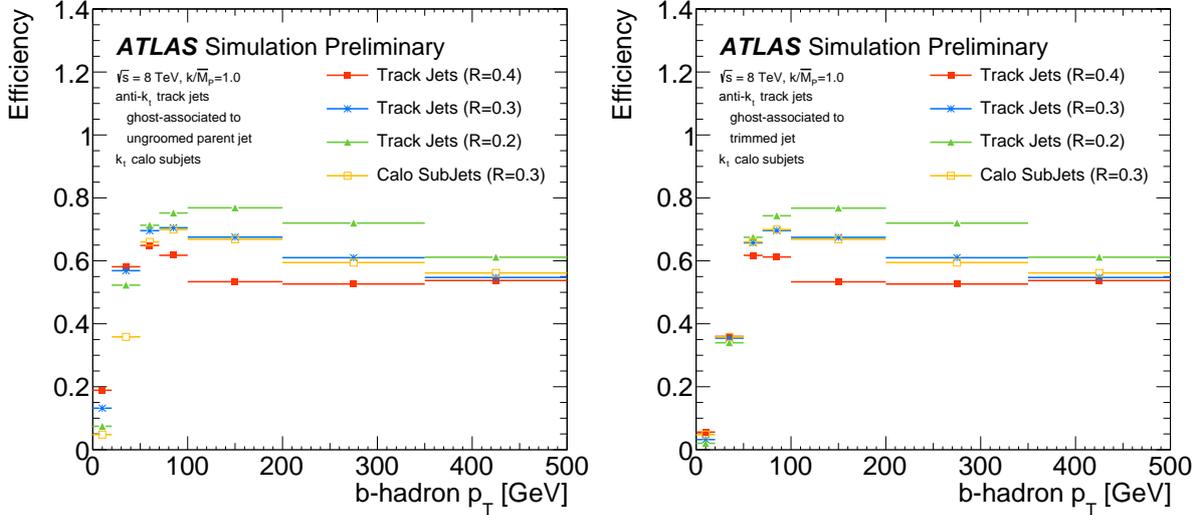


Figure B.9:  $b$ -identification efficiency when using track jets with various  $R$  parameters ghost-associated to trimmed and ungroomed large- $R$  jets as well as when using the trimmed subjects of the large- $R$  jet. The calorimeter subjects are the same in both figures. A sum of RSG samples with masses in 1000, 2400 GeV is used, with an equal contribution from each mass. Statistical uncertainties are included on the efficiencies but are negligible [115].

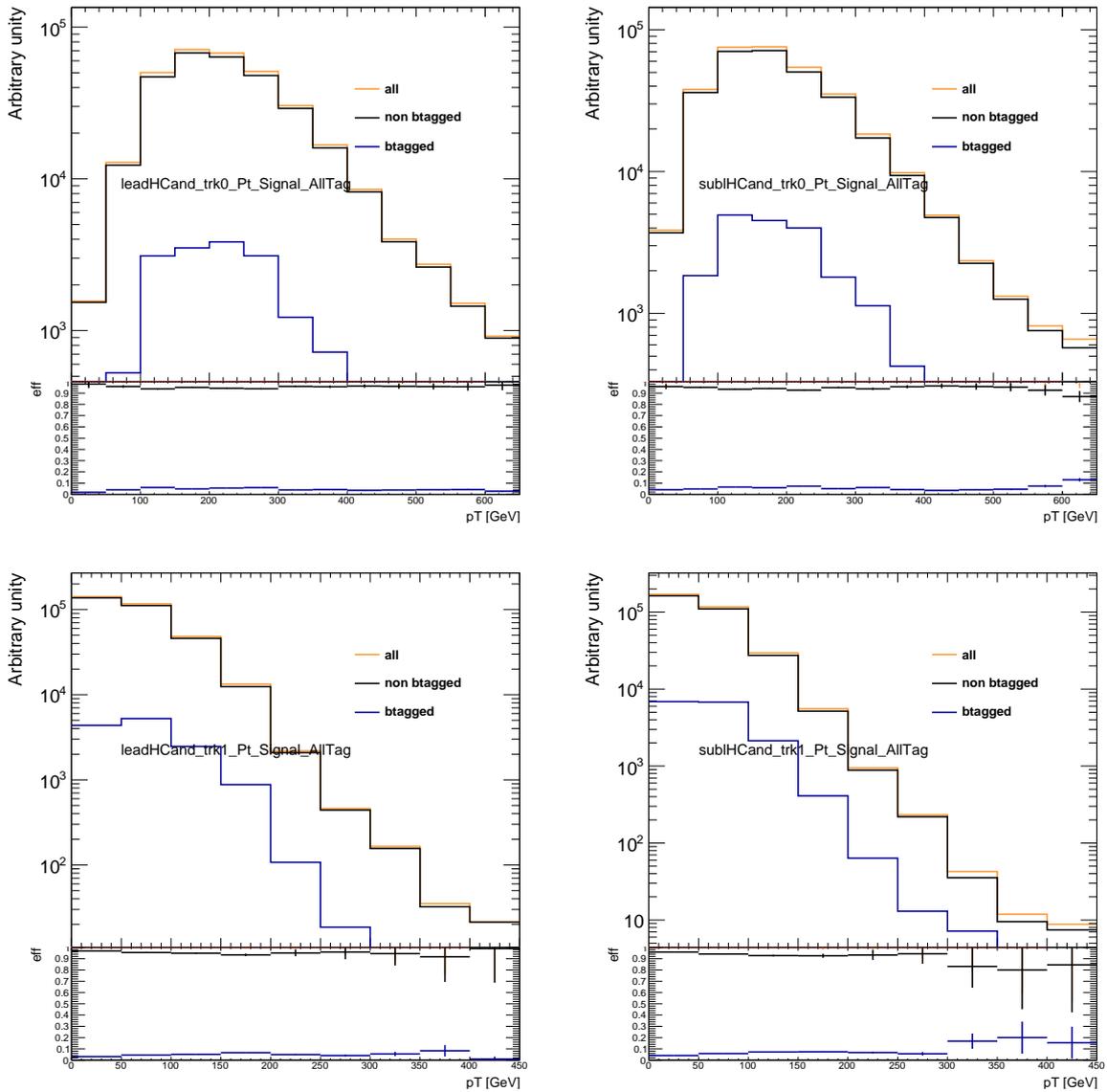


Figure B.10:  $b$ -tagged and un-tagged track jet fractions are shown with respect to  $p_T$  of track jets as indicated on the figures. Results are shown in the SR inclusively, without applying a requirement on the  $b$ -tagged track jet multiplicities.

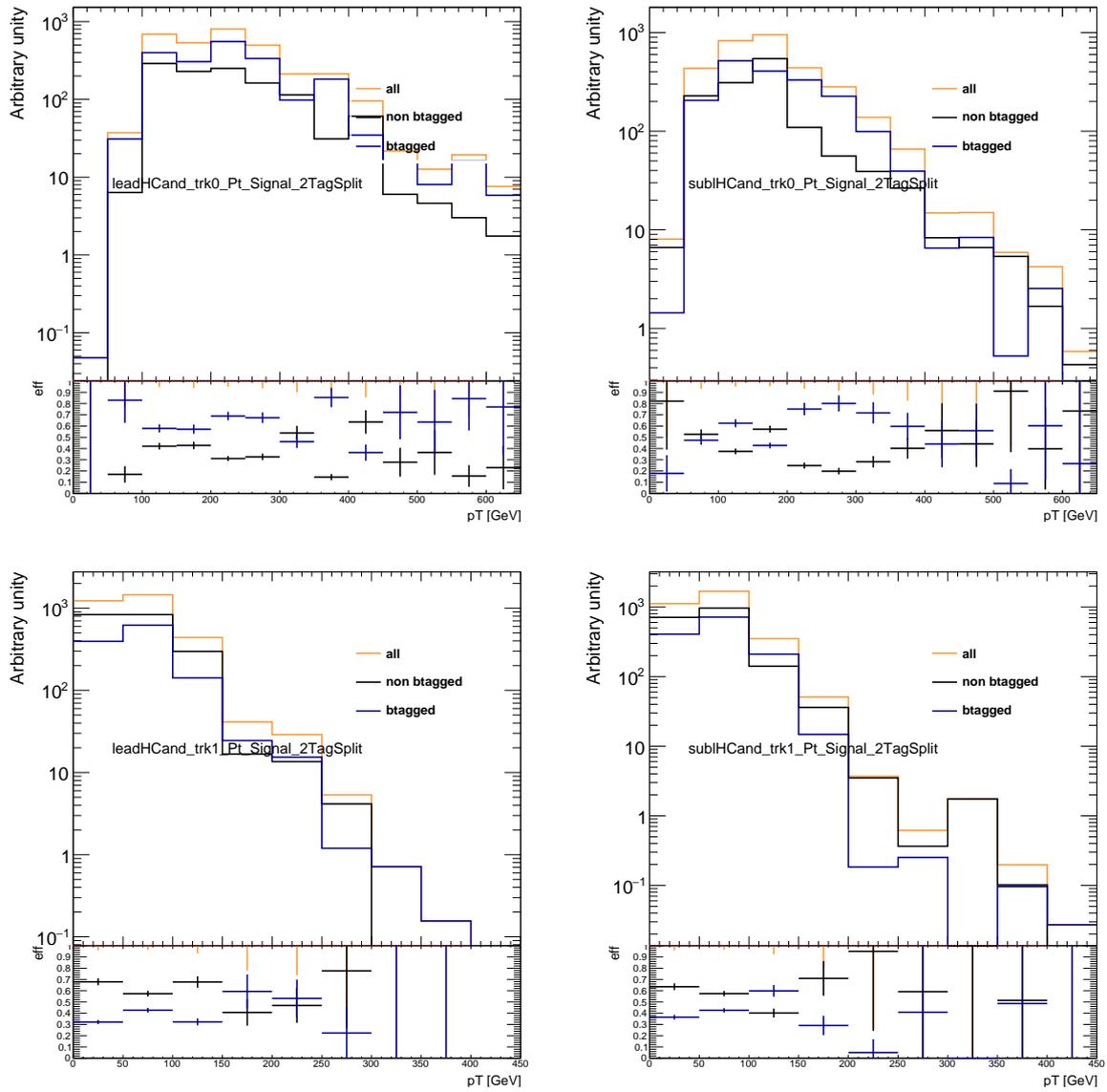


Figure B.11:  $b$ -tagged and un-tagged track jet fractions are shown with respect to  $p_T$  of track jets as indicated on the figures. Results are shown in the 2-tag SR.

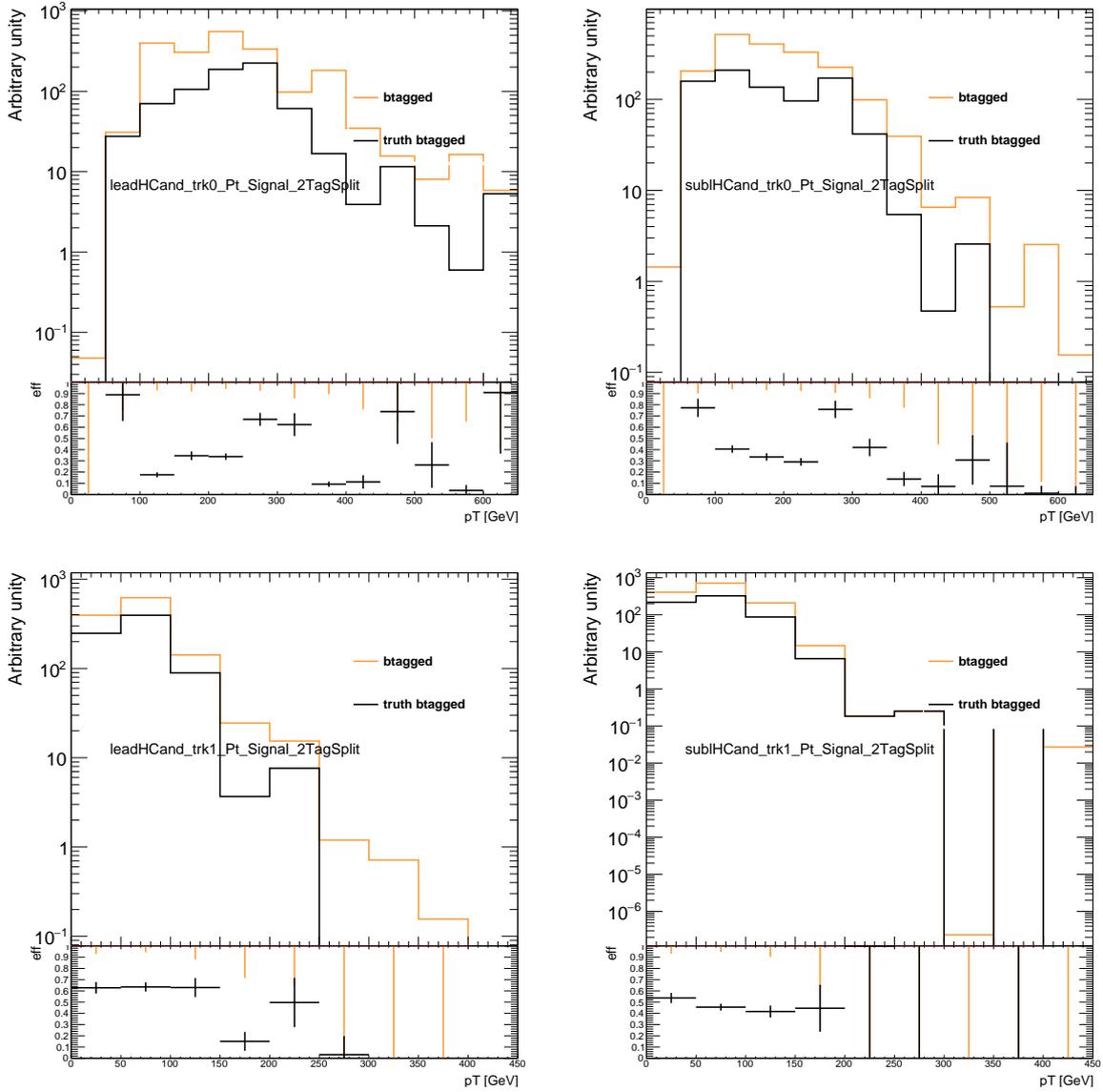


Figure B.12:  $b$ -tagged and truth matched  $b$ -tagged track jets are shown with respect to  $p_T$  of track jets as indicated on the figures. Results are shown in the 2-tag SR.

# Appendix C

## Search for boosted di- $b$ -jet resonances associated with a jet

### C.1 Fixed- $R$ vs variable- $R$ track jets

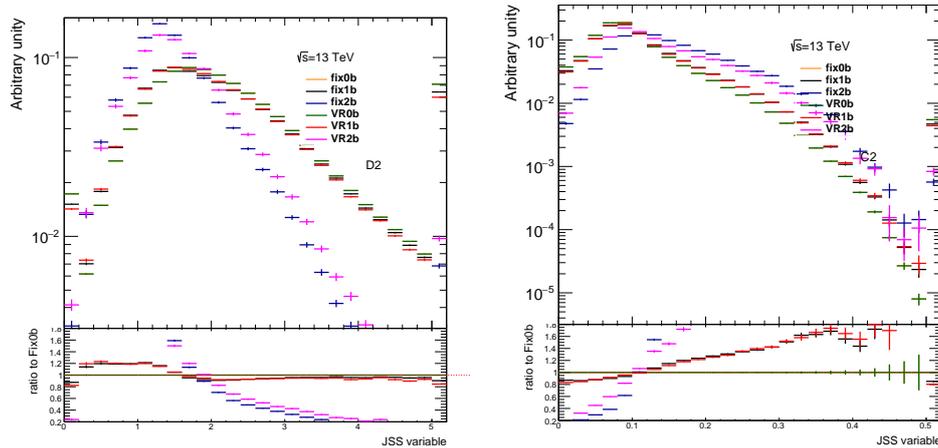


Figure C.1: Comparison for D2 and C2 JSS observables for fix- $R$  and variable- $R$  track jets.

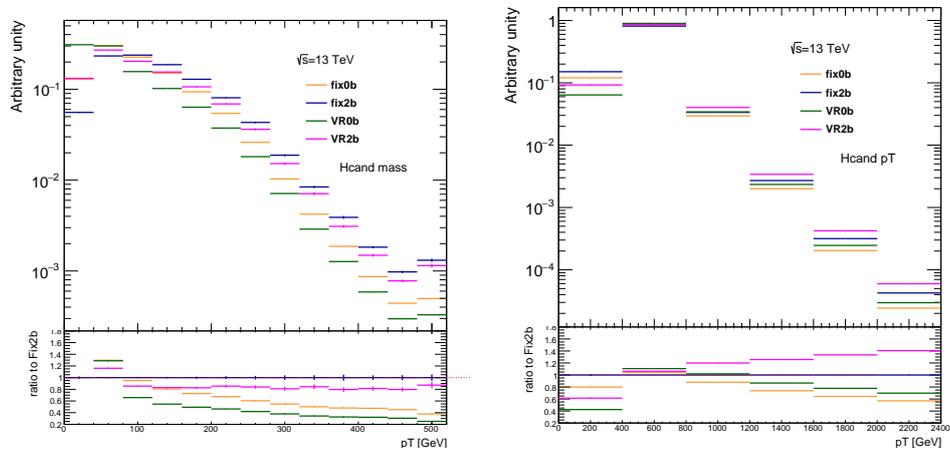


Figure C.2: Mass (left) and  $p_T$  (right) comparison for fix- $R$  track jets and variable- $R$  track jets.

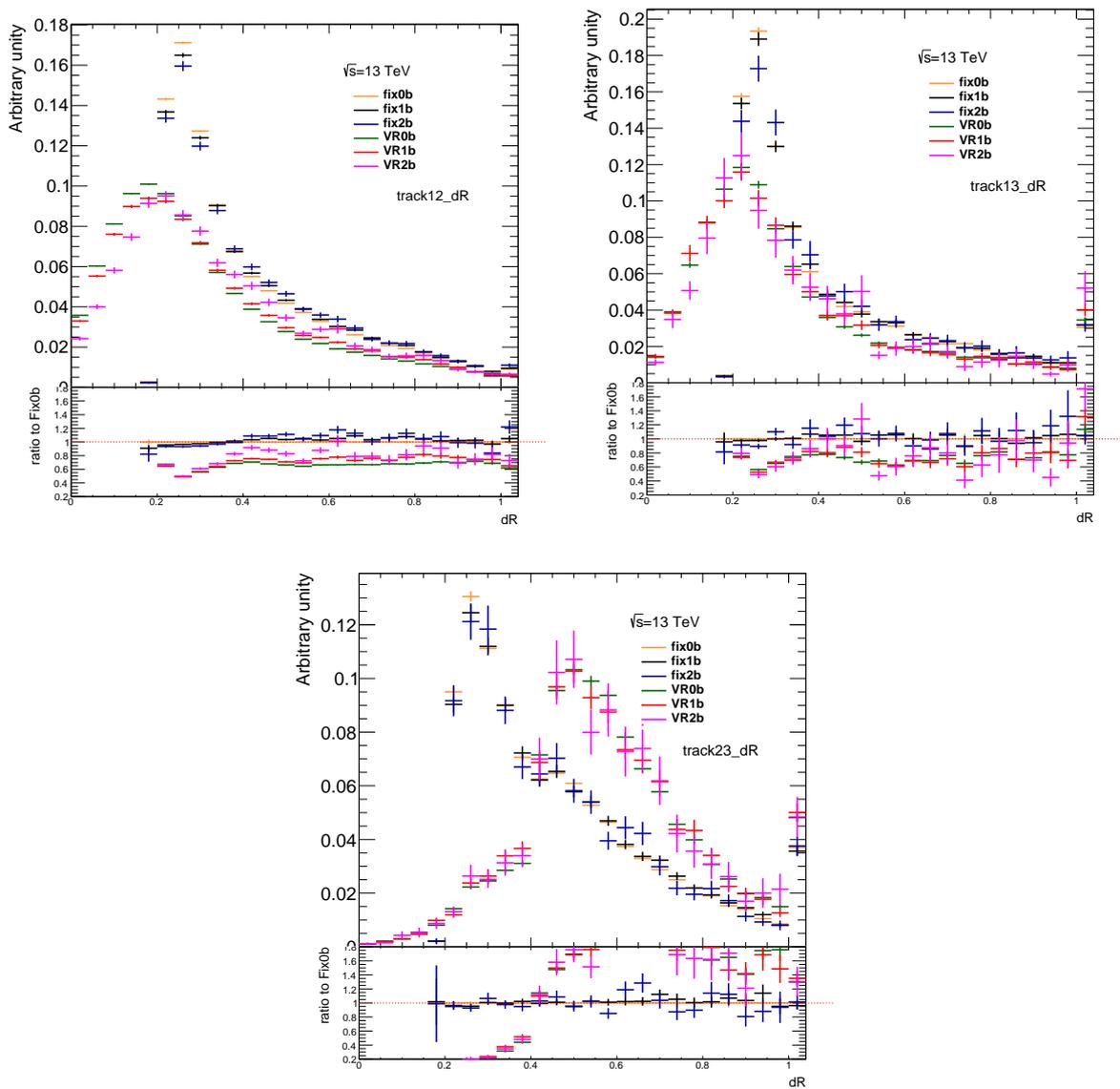


Figure C.3: Distances between leading three track jets in fix- $R$  and variable- $R$  track jets.

## C.2 CR reweighting

This section presents CR reweighting results obtained with alternative reweighting schemes, and signal contamination test with  $V + \text{jets}$ .

### C.2.1 40 GeV cut on 3D reweighting $p_{T,J}, \Delta R_{bb}, \mathcal{R}_{p_T}$

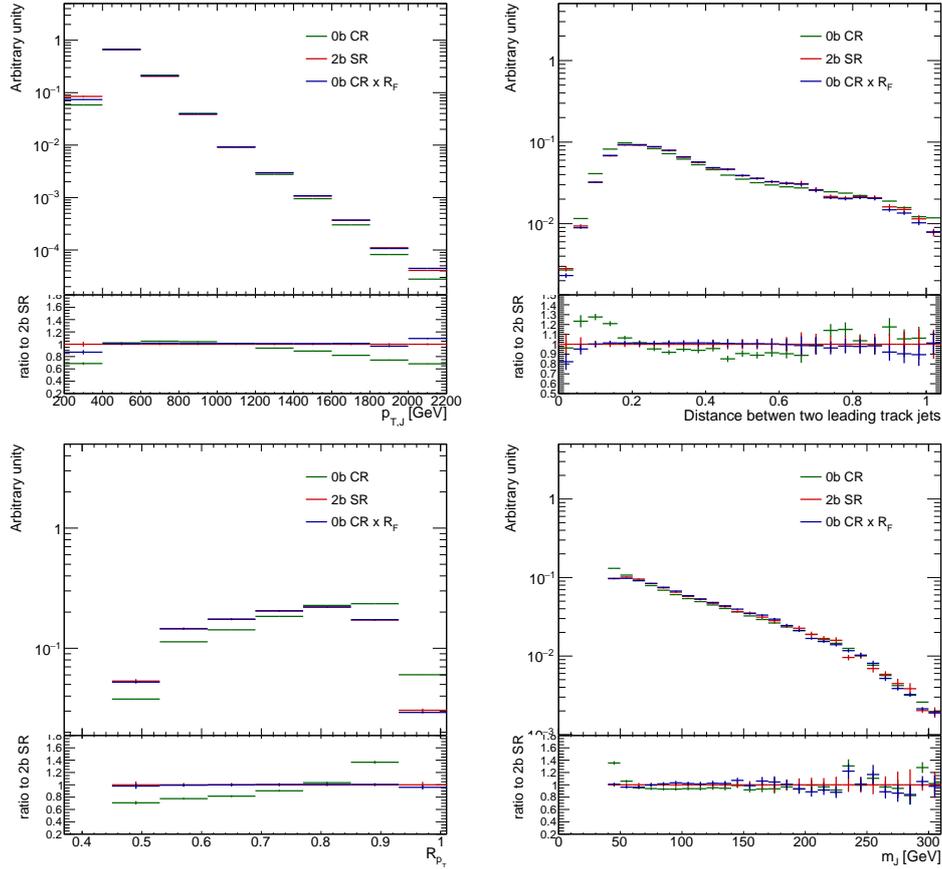


Figure C.4: 3-dimensional reweighting of the CR to the SR in the observables  $p_{T,J}, \Delta R_{bb}, \mathcal{R}_{p_T}$  after applying 40 GeV mass cut.

### C.2.2 3D reweighting with $p_{t,1}, \Delta R_{bb}, \mathcal{R}_{p_T}$

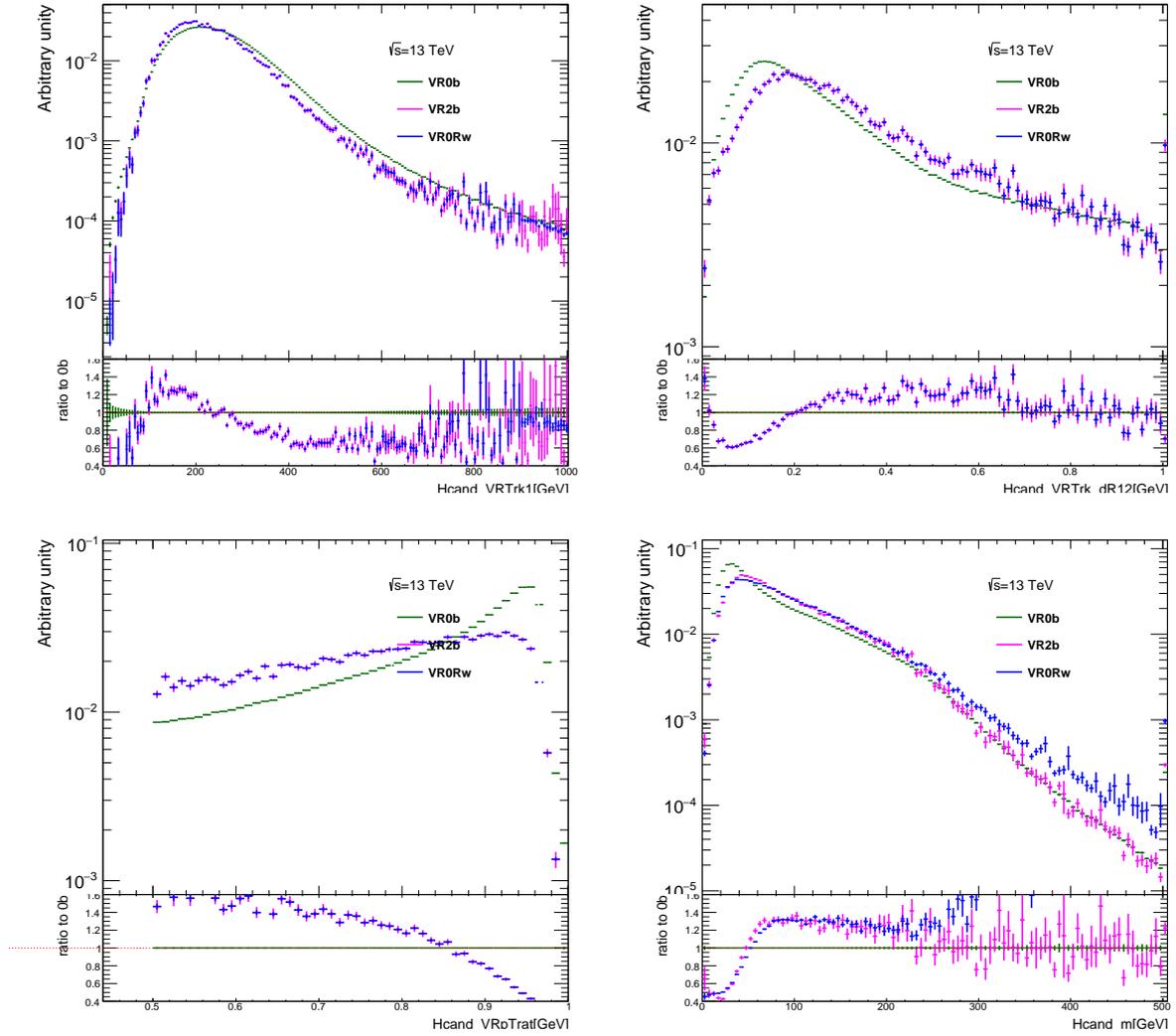


Figure C.5: The effect of the 3-dimensional reweighting of the CR to the SR in the observables  $p_{t,1}$ ,  $\Delta R_{bb}$ ,  $\mathcal{R}_{p_T}$  for the distributions. VR stands for variable- $R$  track jets.

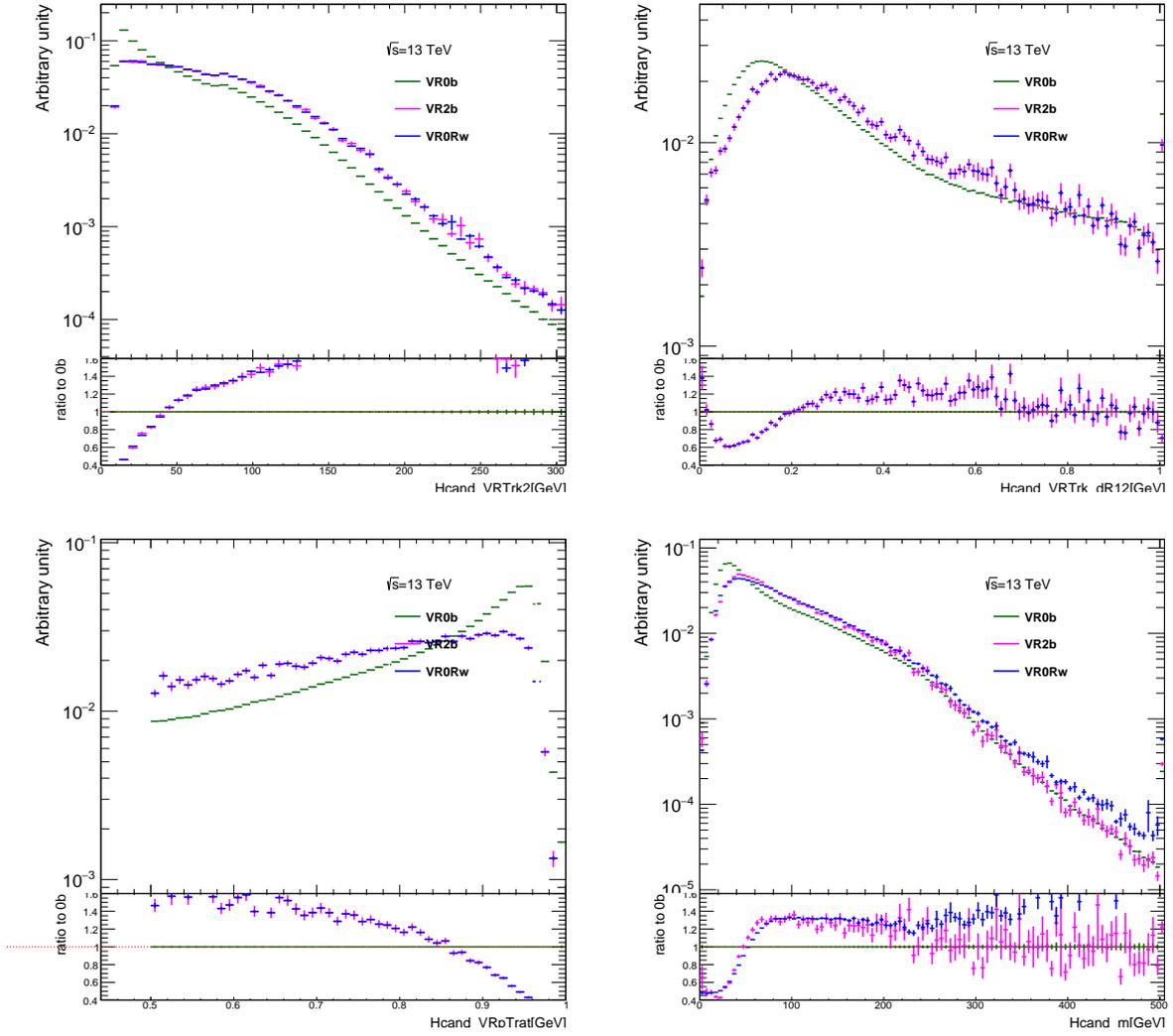
C.2.3 3D reweighting with  $p_{t,2}$ ,  $\Delta R_{bb}$ ,  $\mathcal{R}_{p_T}$ 

Figure C.6: The effect of the 3-dimensional reweighting of the CR to the SR in the observables  $p_{t,2}$ ,  $\Delta R_{bb}$ ,  $\mathcal{R}_{p_T}$ . VR stands for variable- $R$  track jets.

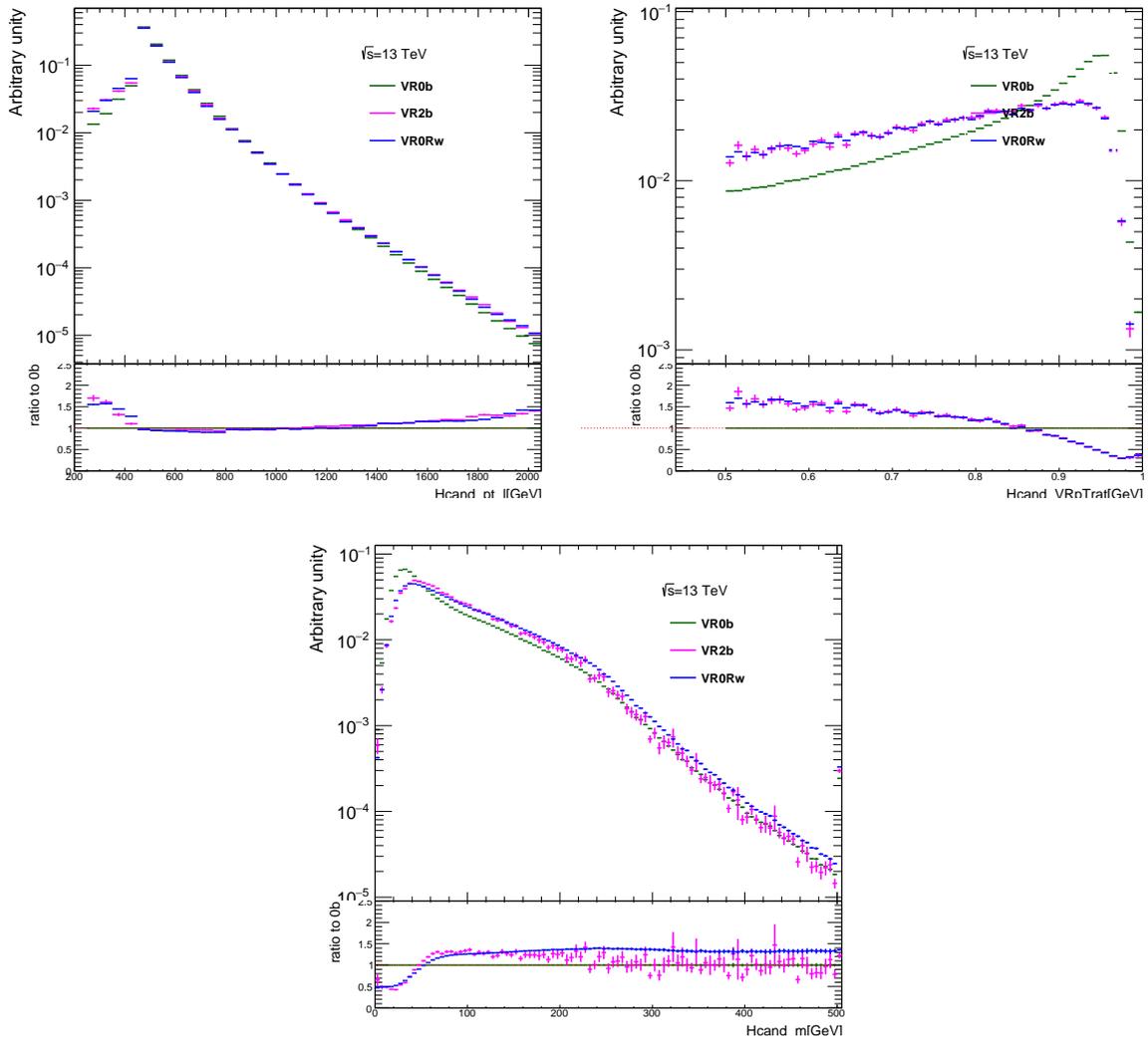
C.2.4 2D reweighting with  $p_{T,J}, \mathcal{R}_{p_T}$ 

Figure C.7: The effect of the 2-dimensional reweighting of the CR to the SR in the observables  $p_{T,J}, \mathcal{R}_{p_T}$ . VR stands for variable- $R$  track jets.

### C.2.5 2D reweighting with $\Delta R_{bb}, \mathcal{R}_{p_T}$

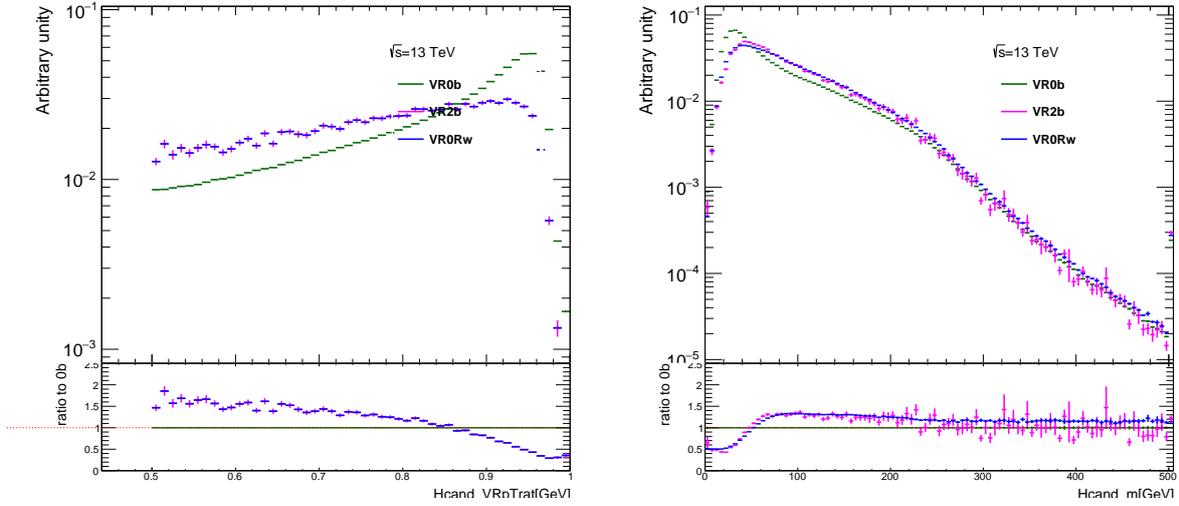


Figure C.8: The effect of the 2-dimensional reweighting of the CR to the SR in the observables  $\Delta R_{bb}, \mathcal{R}_{p_T}$ . VR stands for variable- $R$  track jets.

## C.3 Signal contamination tests

Below figure presents the injected  $Z'$  process for the signal contamination test.

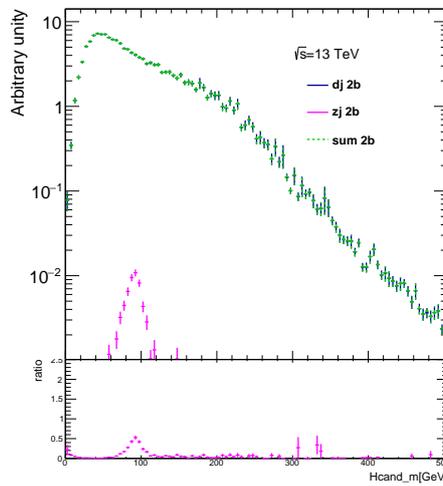


Figure C.9: Injected  $Z'$  signal is shown together with QCD background.

Below figures present the results of the signal contamination test performed with the injected SM  $V + \text{jets}$  process.

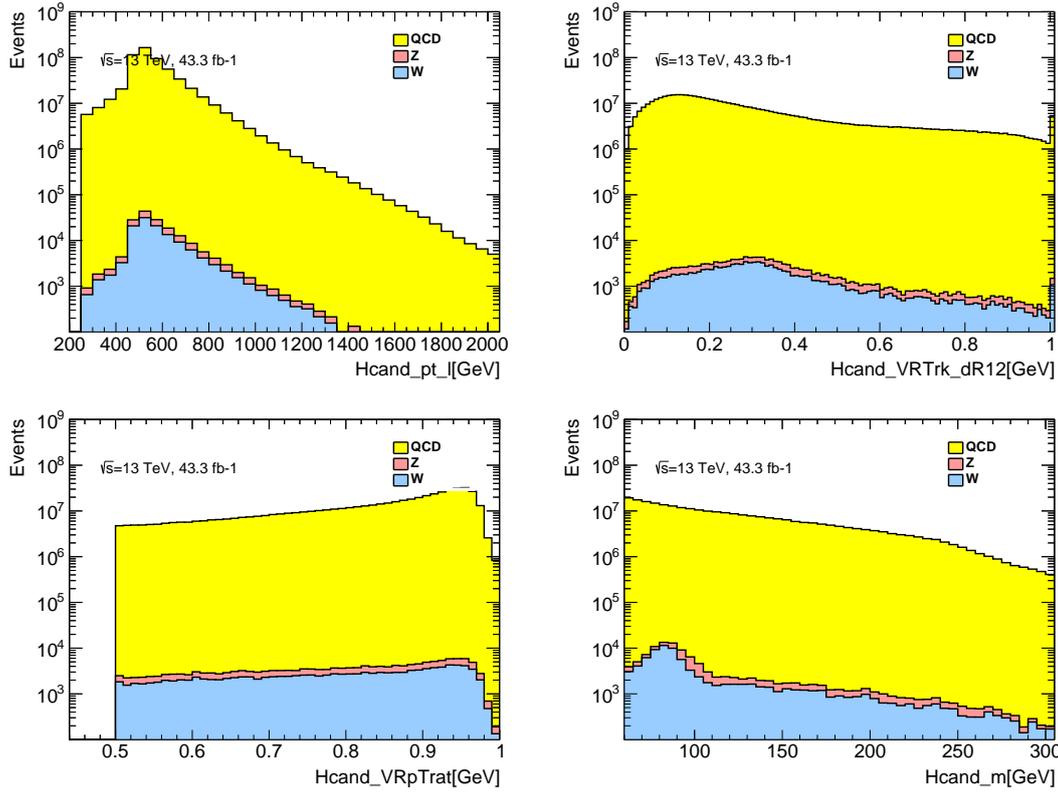


Figure C.10: The contribution of the  $V + \text{jets}$  and QCD MJ backgrounds to the CR in the observables (a)  $p_{T,J}$ , (b)  $\Delta R_{bb}$ , (c)  $\mathcal{R}_{p_T}$ , and (d)  $m_{bb}$  of the Higgs candidate.

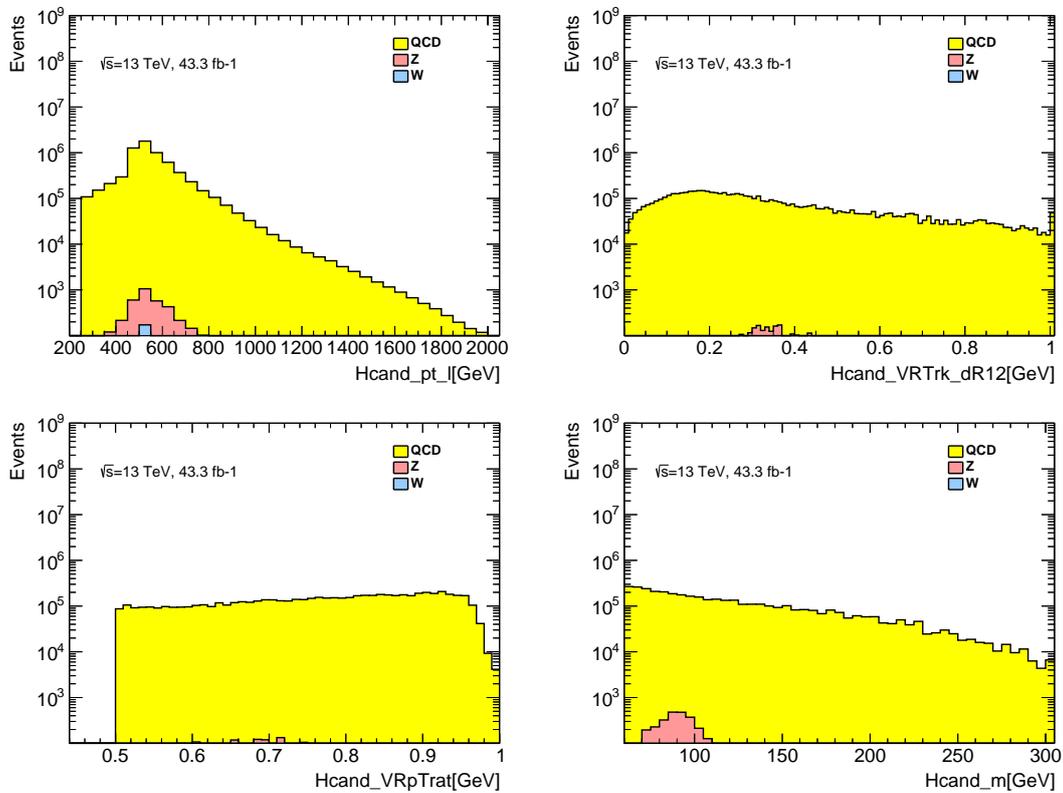


Figure C.11: The contribution of the  $V + \text{jets}$  and QCD MJ backgrounds to the SR in the observables (a)  $p_{T,J}$ , (b)  $\Delta R_{bb}$ , (c)  $\mathcal{R}_{p_T}$ , and (d)  $m_{bb}$  of the Higgs candidate.

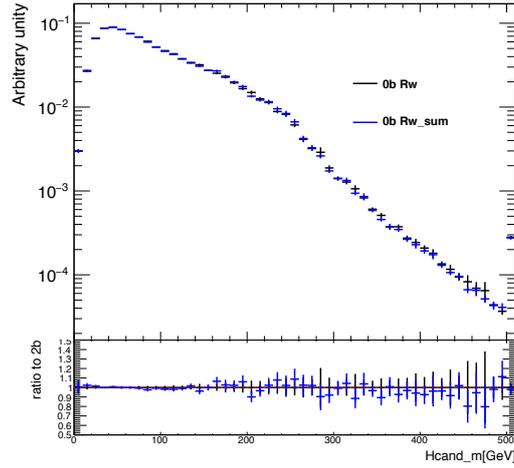


Figure C.12: The QCD MJ background prediction using the CR considering only QCD MJ production when deriving the 3-dimensional reweighting factors ( $R_w$ ) is compared to the prediction where the reweighting is derived including the contribution from  $V$  + jets both in the CR and in the SR ( $R_{w\_sum}$ ).

## C.4 Systematic uncertainties

### C.4.1 Scale uncertainties

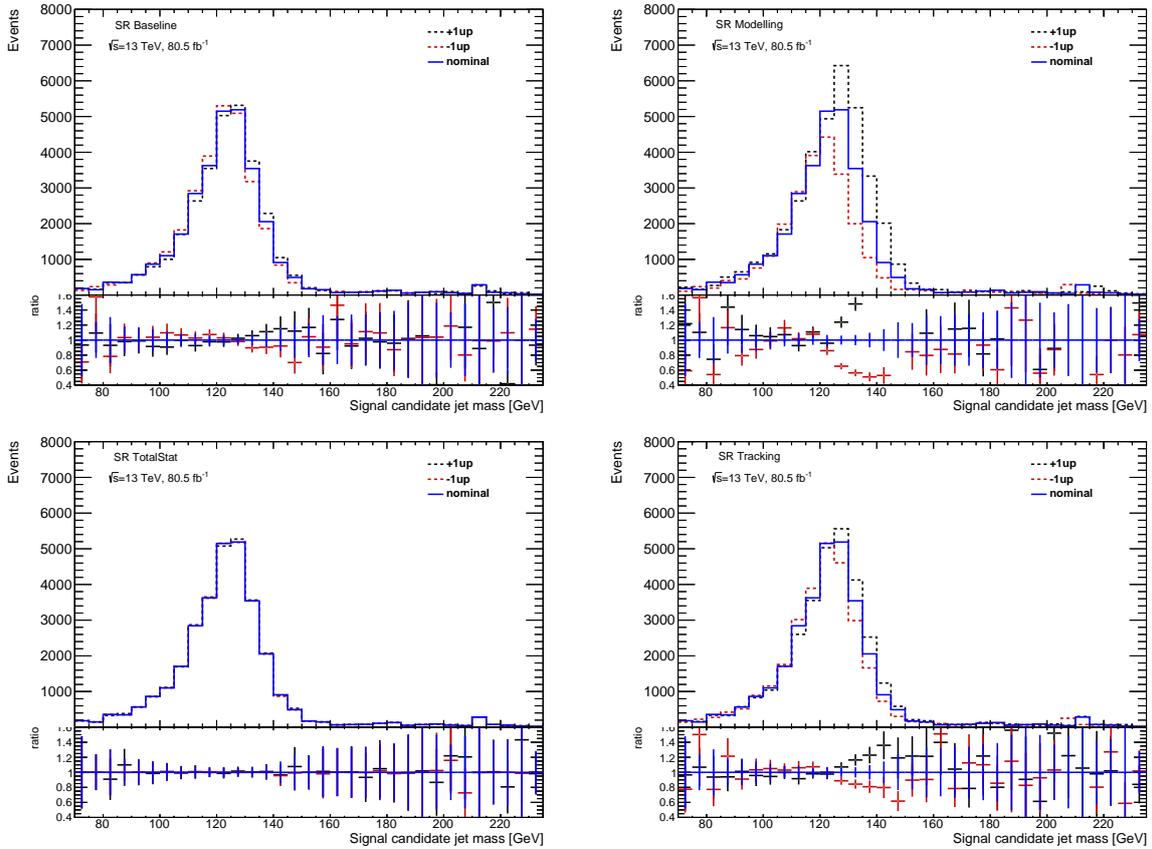


Figure C.13: The impact of the  $1\sigma$  varied JES and JMS uncertainties on the signal candidate large- $R$  jet mass distribution in simulated Higgs samples.

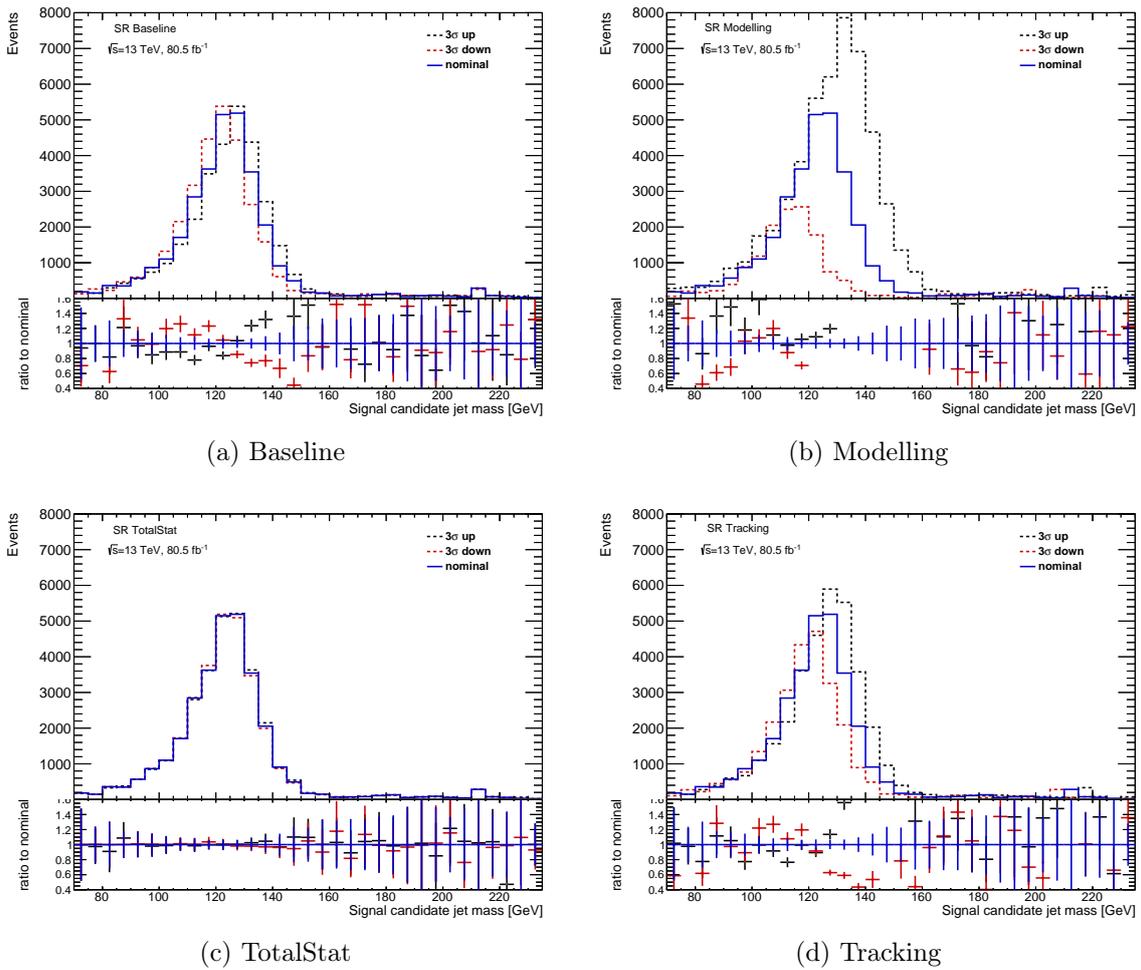


Figure C.14: The impact of the  $3\sigma$  varied JES and JMS uncertainties on the signal candidate large- $R$  jet mass distribution in simulated Higgs samples.

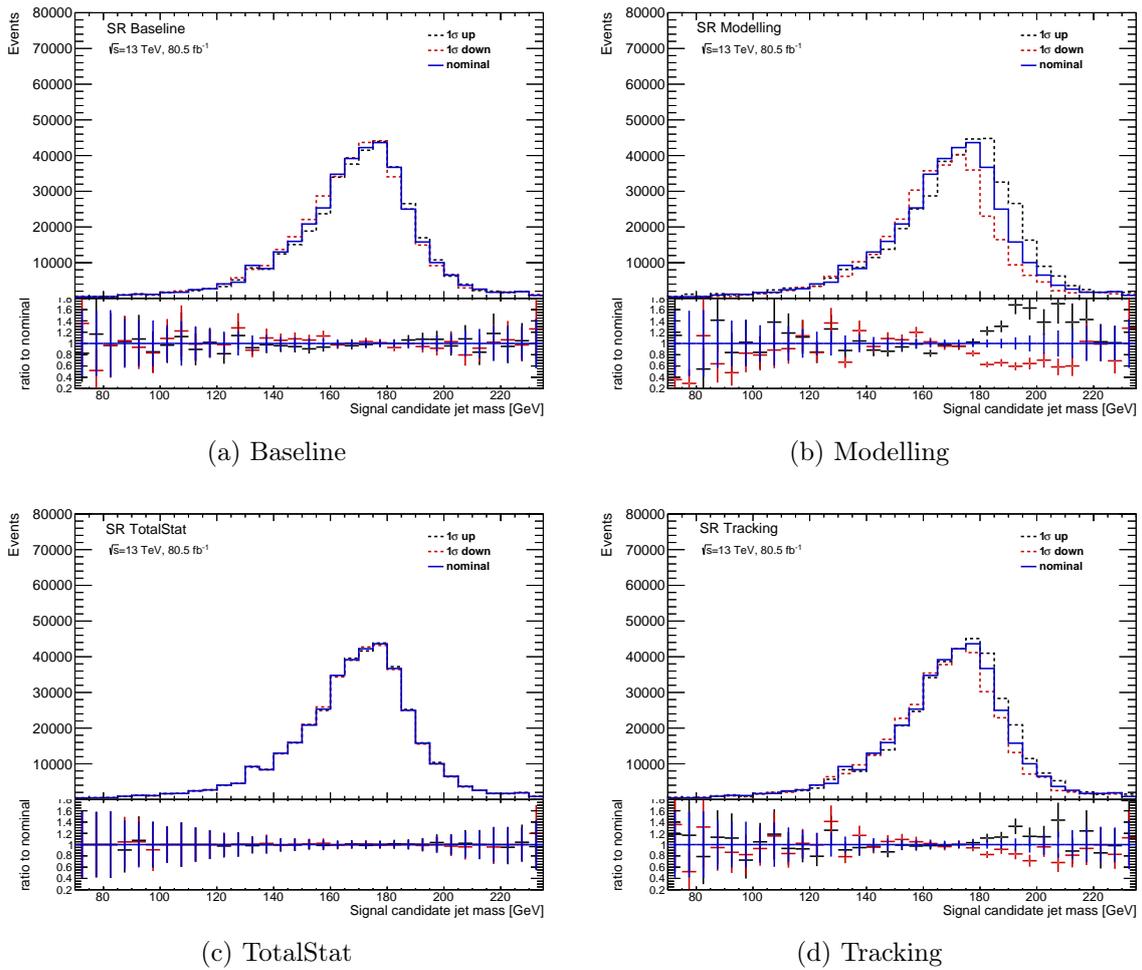


Figure C.15: The impact of the  $1\sigma$  varied JES and JMS uncertainties on the signal candidate large- $R$  jet mass distribution in simulated  $Z'$  samples generated at 175 GeV mass.

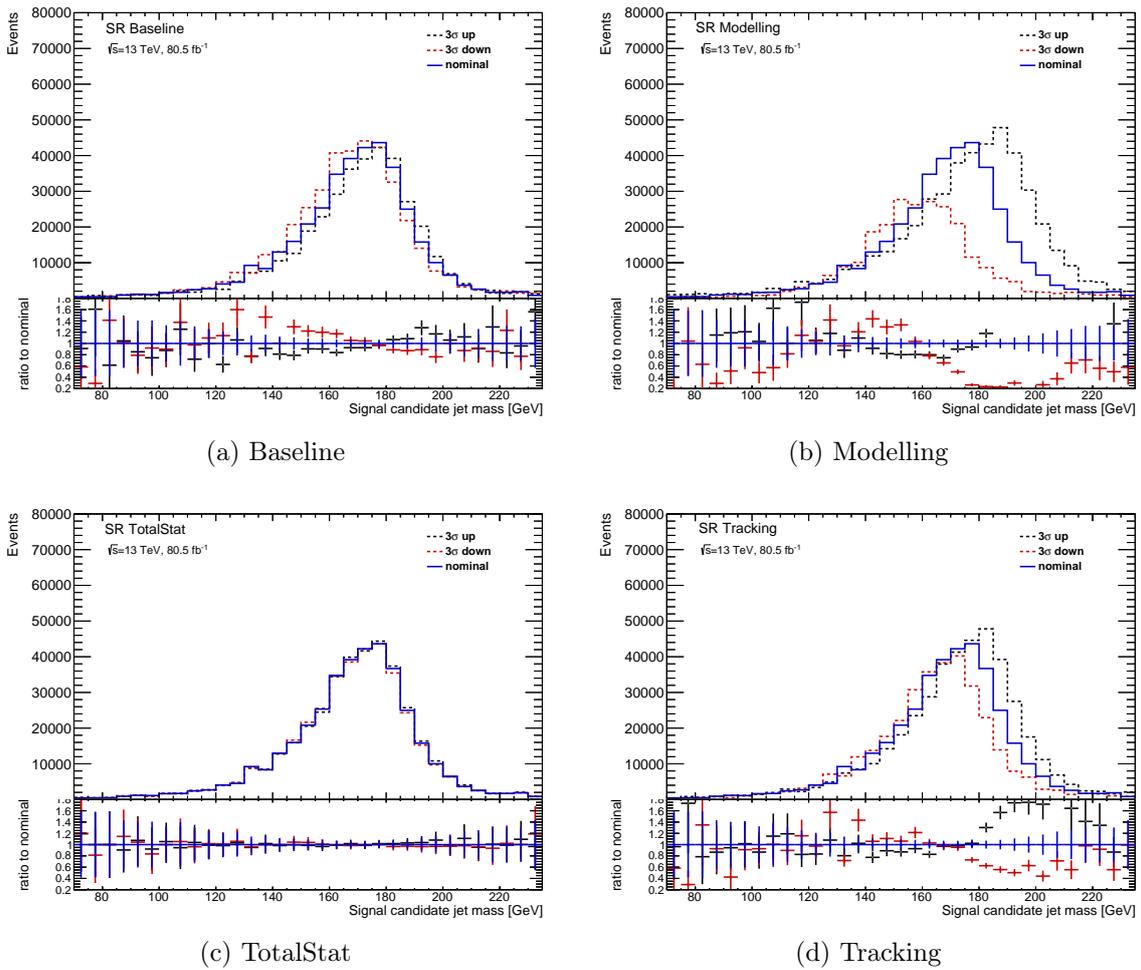


Figure C.16: The impact of the  $3\sigma$  varied JES and JMS uncertainties on the signal candidate large- $R$  jet mass distribution in simulated  $Z'$  samples generated at 175 GeV mass.

### C.4.2 Flavour tagging uncertainties

The assumption and size of a normalisation-only uncertainty has been checked by using the *SFEigen* method developed by flavour tagging group and the following CDI file:

- 2017-21-13TeV-MC16-CDI-2018-05-04\_v1.root,

25 nuisance parameters have been taken into consideration with their up and down variations, 9 for  $b$ -flavour, 3 for  $c$ -flavour and 11 for light flavours. The impact of the each source to the different samples shown in the Figures C.17, C.18 for  $Z$  + jets, Figures C.19, C.20 for the Higgs, Figures C.21, C.22 for  $t\bar{t}$  and Figures C.23, C.24 for  $Z'$  samples using all generated mass points as a single template. Numbers on the figures correspond to different sources of systematics.

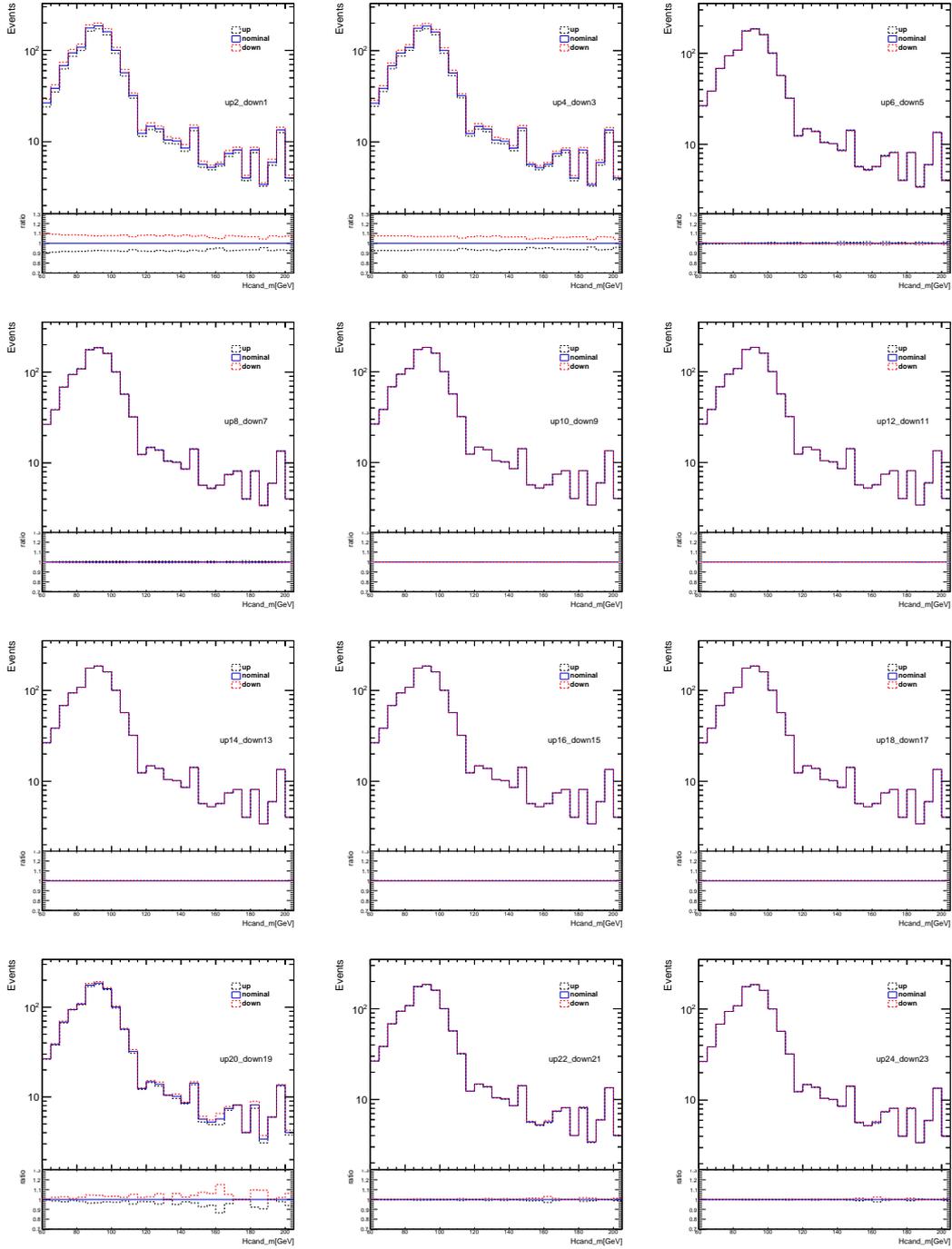


Figure C.17: The impact of the  $1\sigma$  up/down variations of  $b$ - and  $c$ - flavour uncertainties on signal candidate large- $R$  jet mass distribution in  $Z + \text{jets}$  samples.

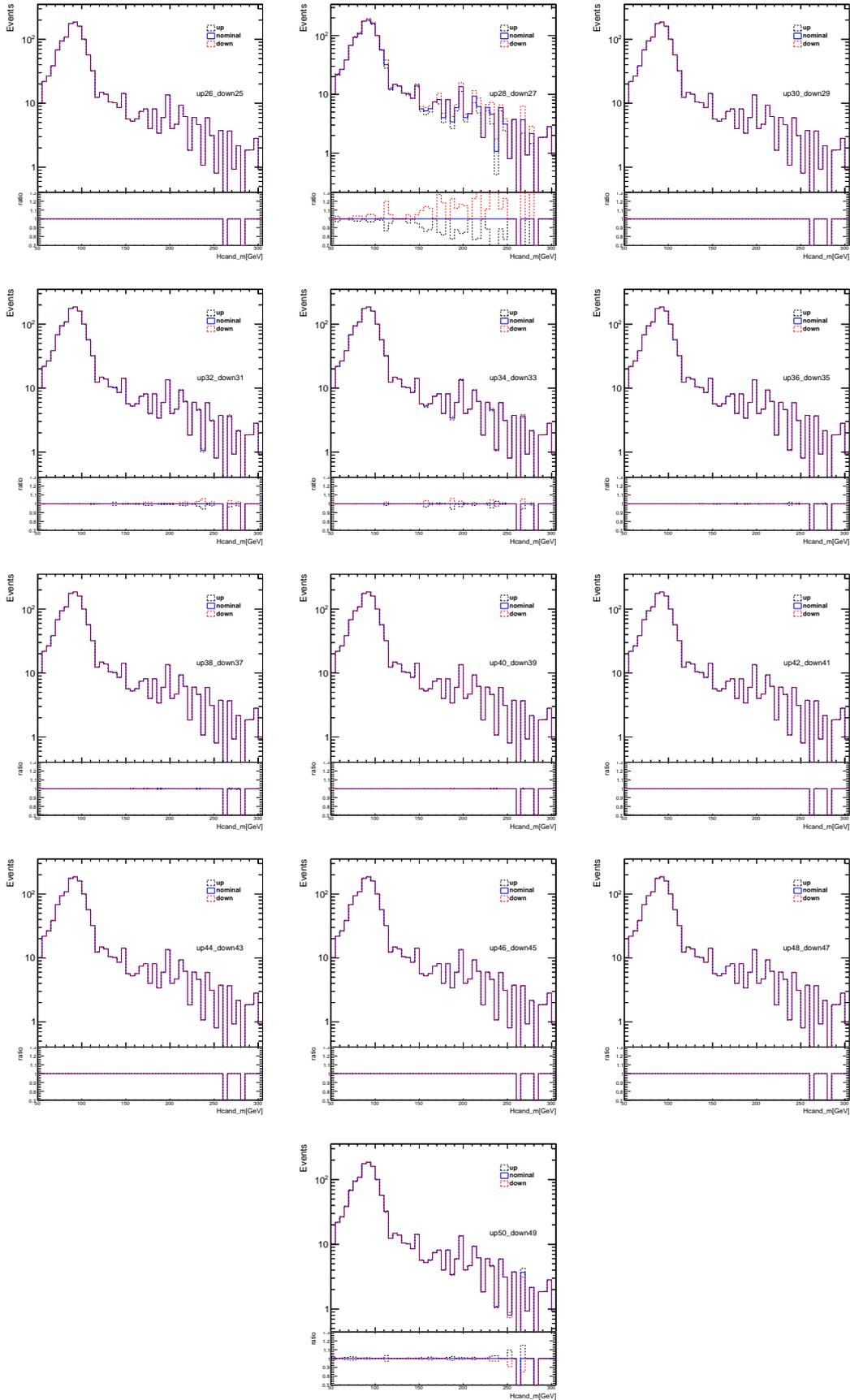


Figure C.18: The impact of the  $1\sigma$  up/down variations of light flavour uncertainties and extrapolation uncertainties on signal candidate large- $R$  jet mass distribution in  $Z + \text{jets}$  samples.

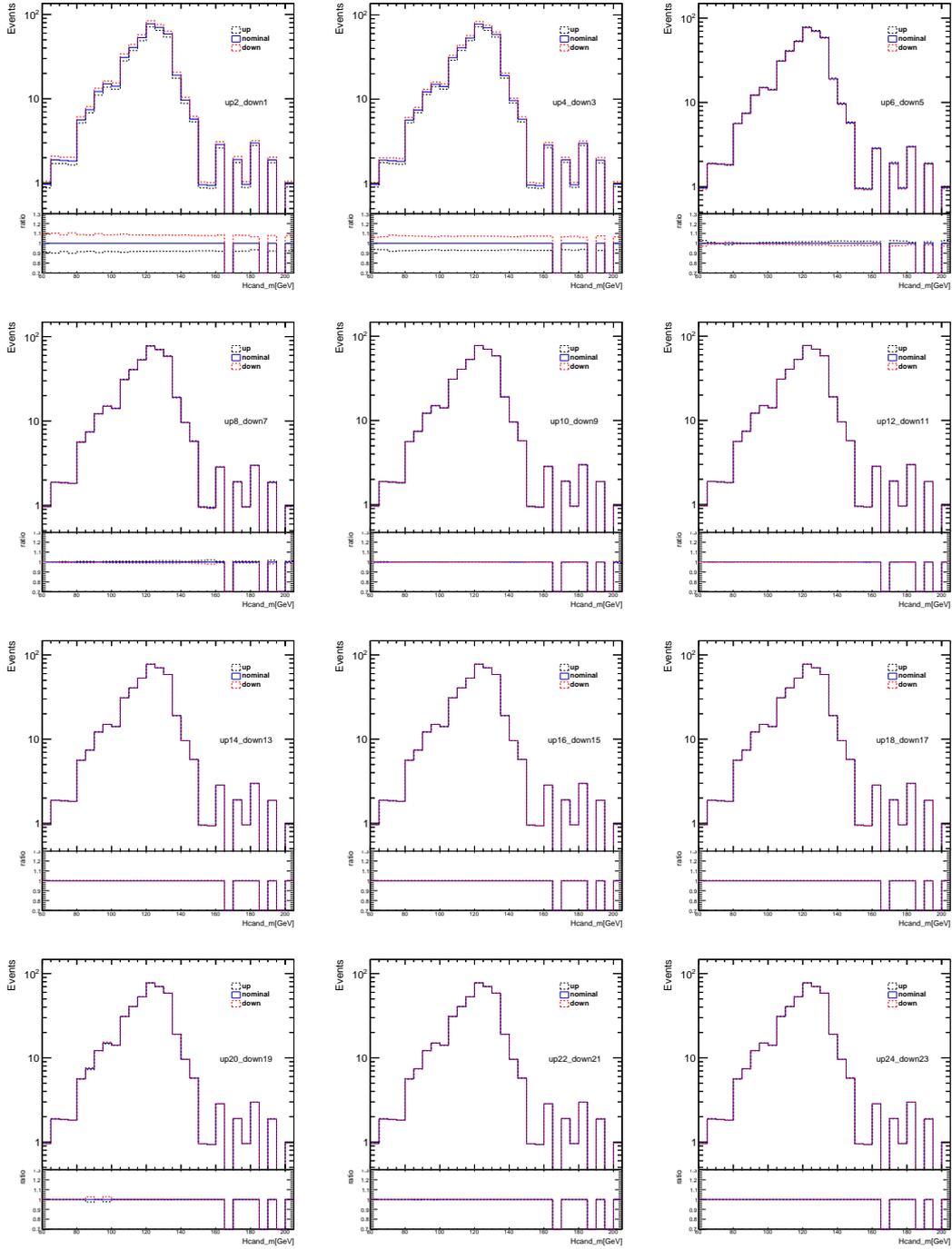


Figure C.19: The impact of the  $1\sigma$  up/down variations of  $b$ - and  $c$ -flavour uncertainties on signal candidate large- $R$  jet mass distribution in Higgs samples.

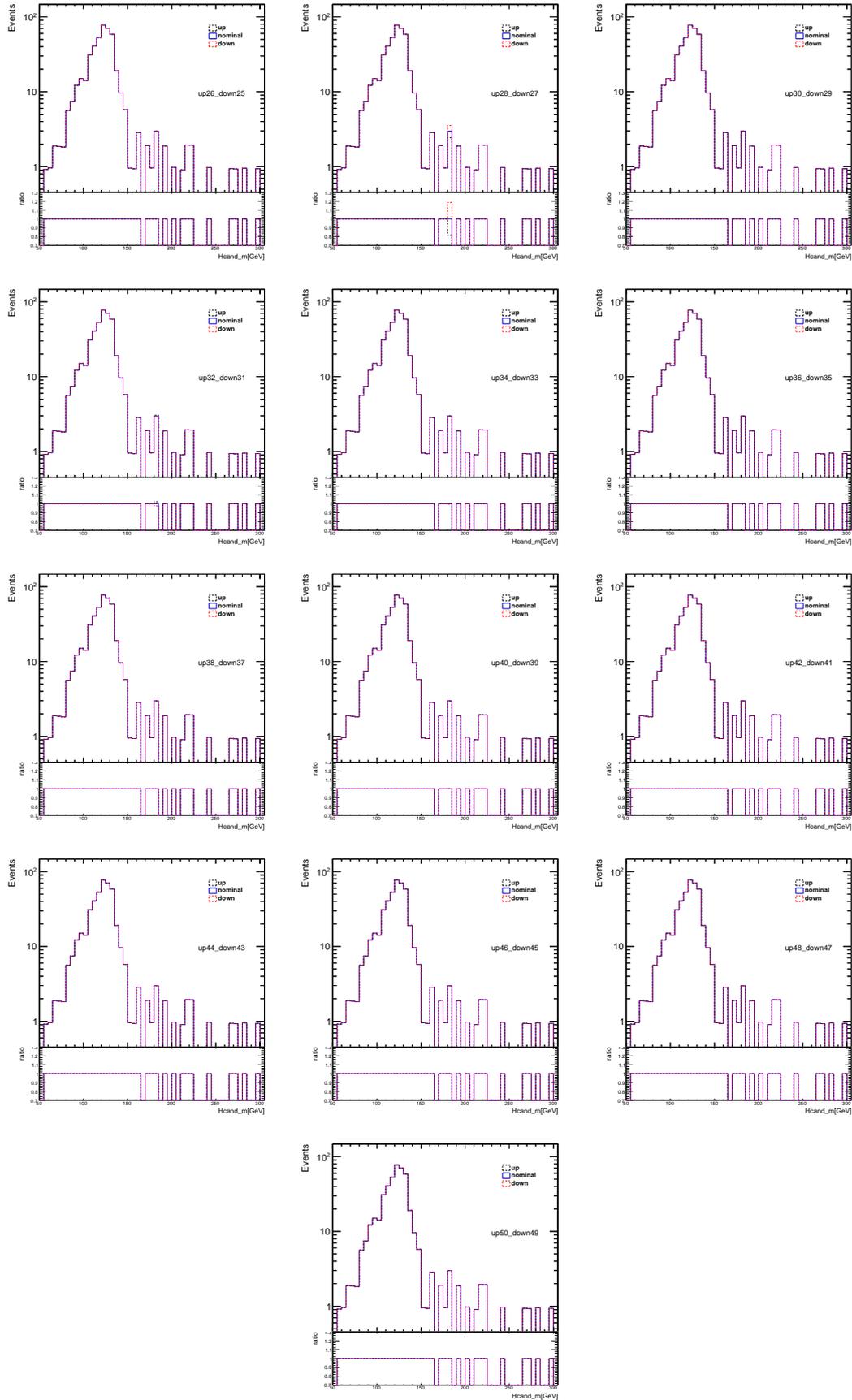


Figure C.20: The impact of the  $1\sigma$  up/down variations of light flavour uncertainties and extrapolation uncertainties on signal candidate large- $R$  jet mass distribution in Higgs samples.

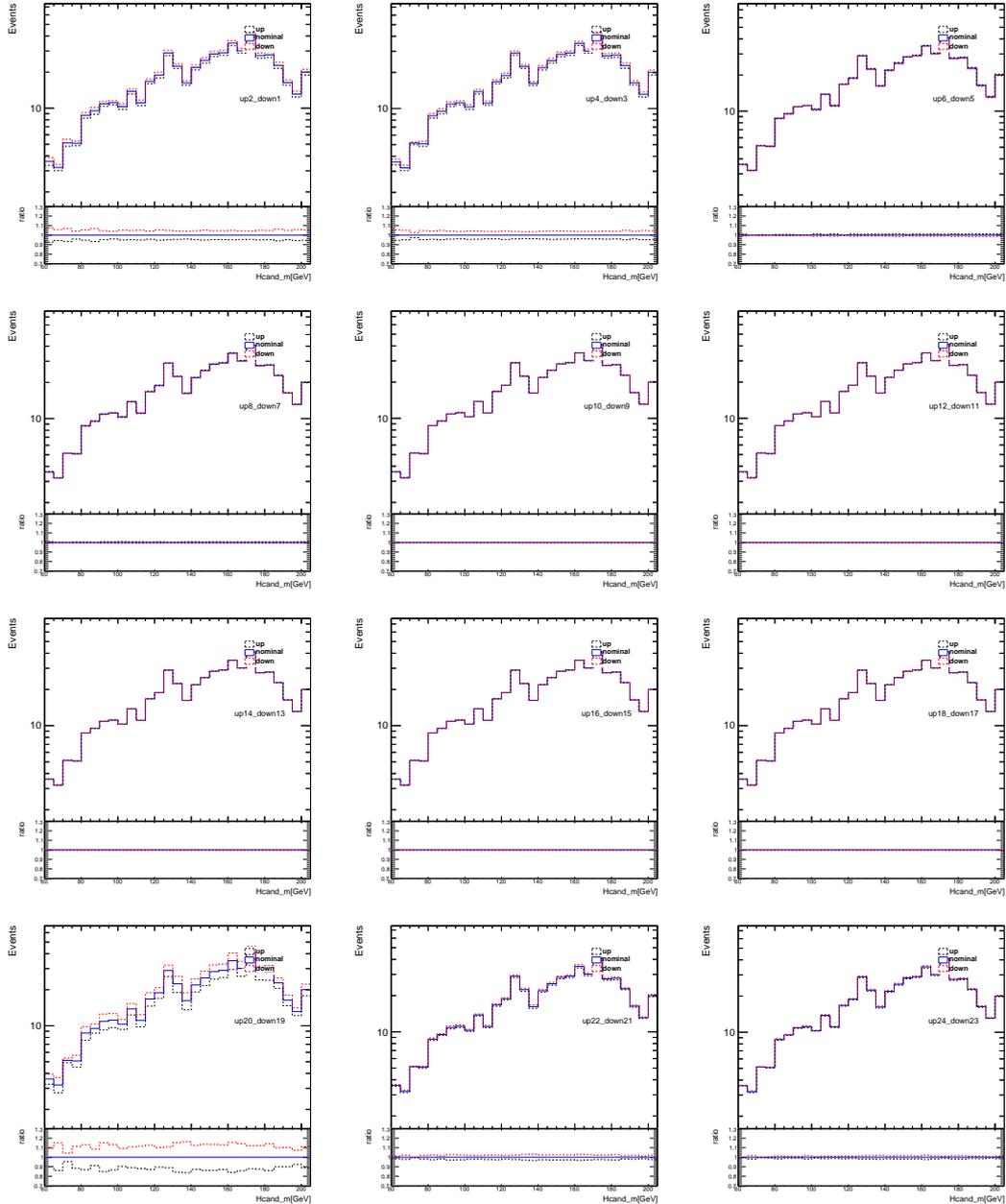


Figure C.21: The impact of the  $1\sigma$  up/down variations of  $b$ - and  $c$ - flavour uncertainties on signal candidate large- $R$  jet mass distribution in  $t\bar{t}$  samples.

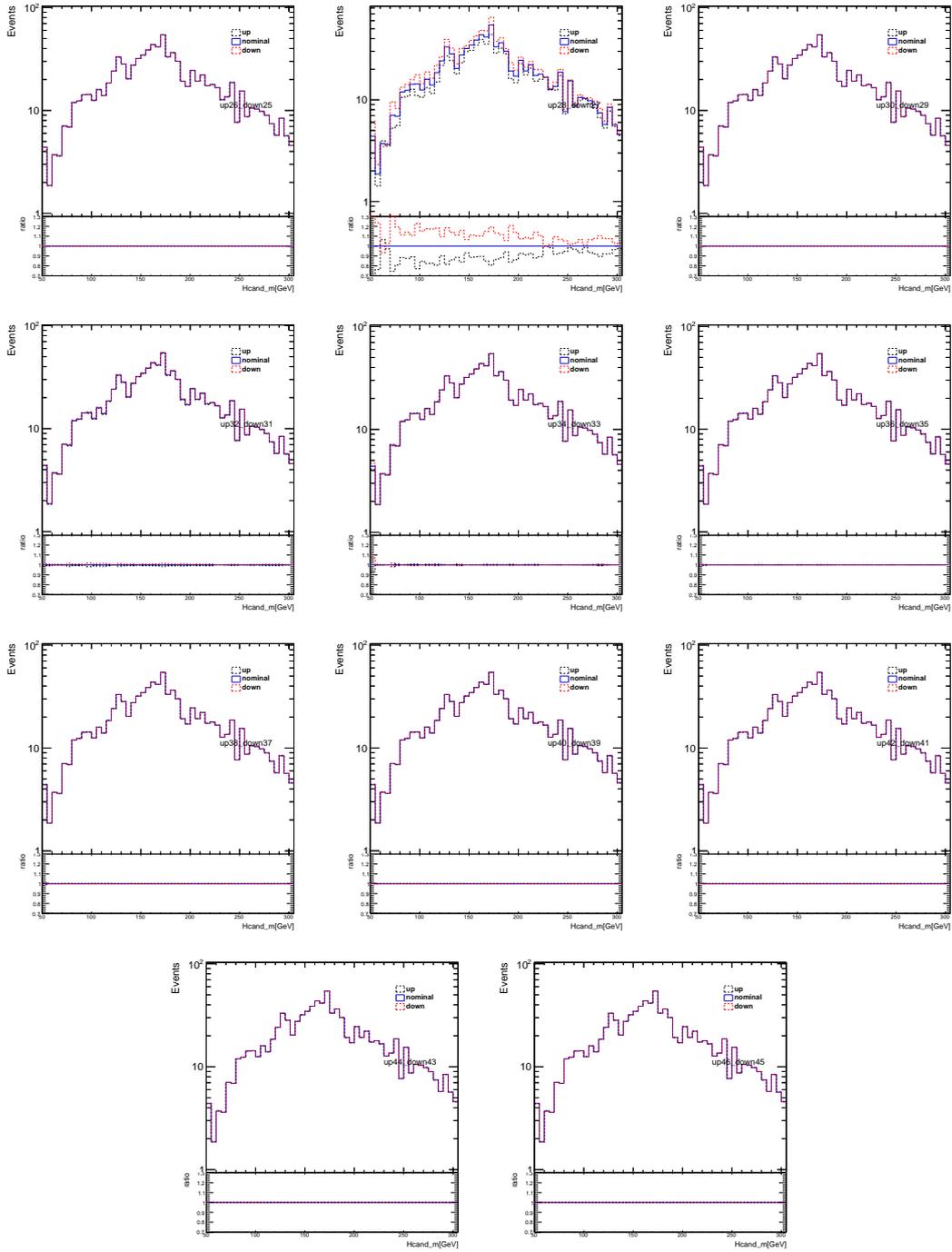


Figure C.22: The impact of the  $1\sigma$  up/down variations of light flavour uncertainties and extrapolation uncertainties on signal candidate large- $R$  jet mass distribution in  $t\bar{t}$  samples.

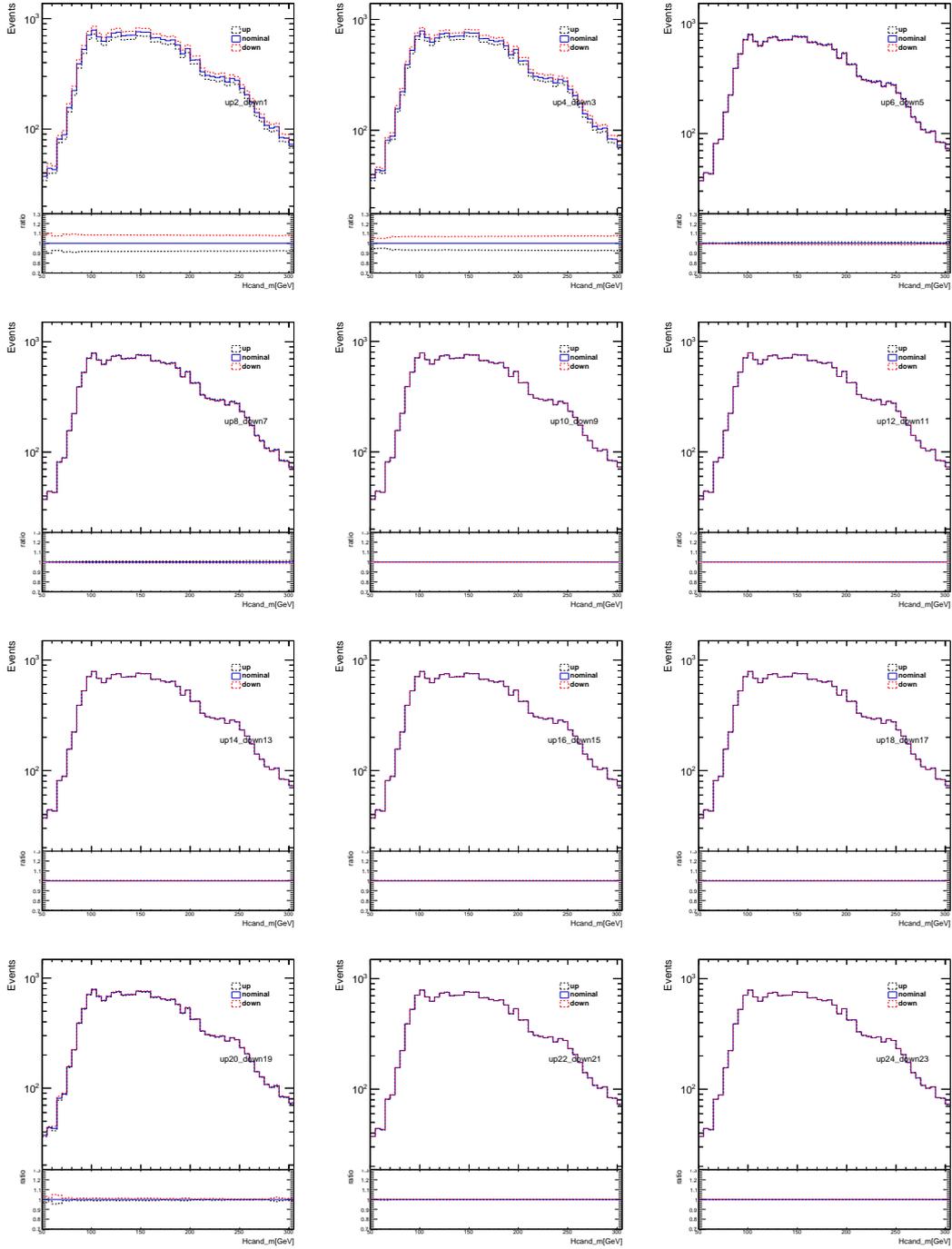


Figure C.23: The impact of the  $1\sigma$  up/down variations of  $b$ - and  $c$ -flavour uncertainties on signal candidate large- $R$  jet mass distribution in  $Z'$  samples, using all  $b$ -generated  $c$  mass points as a single template.

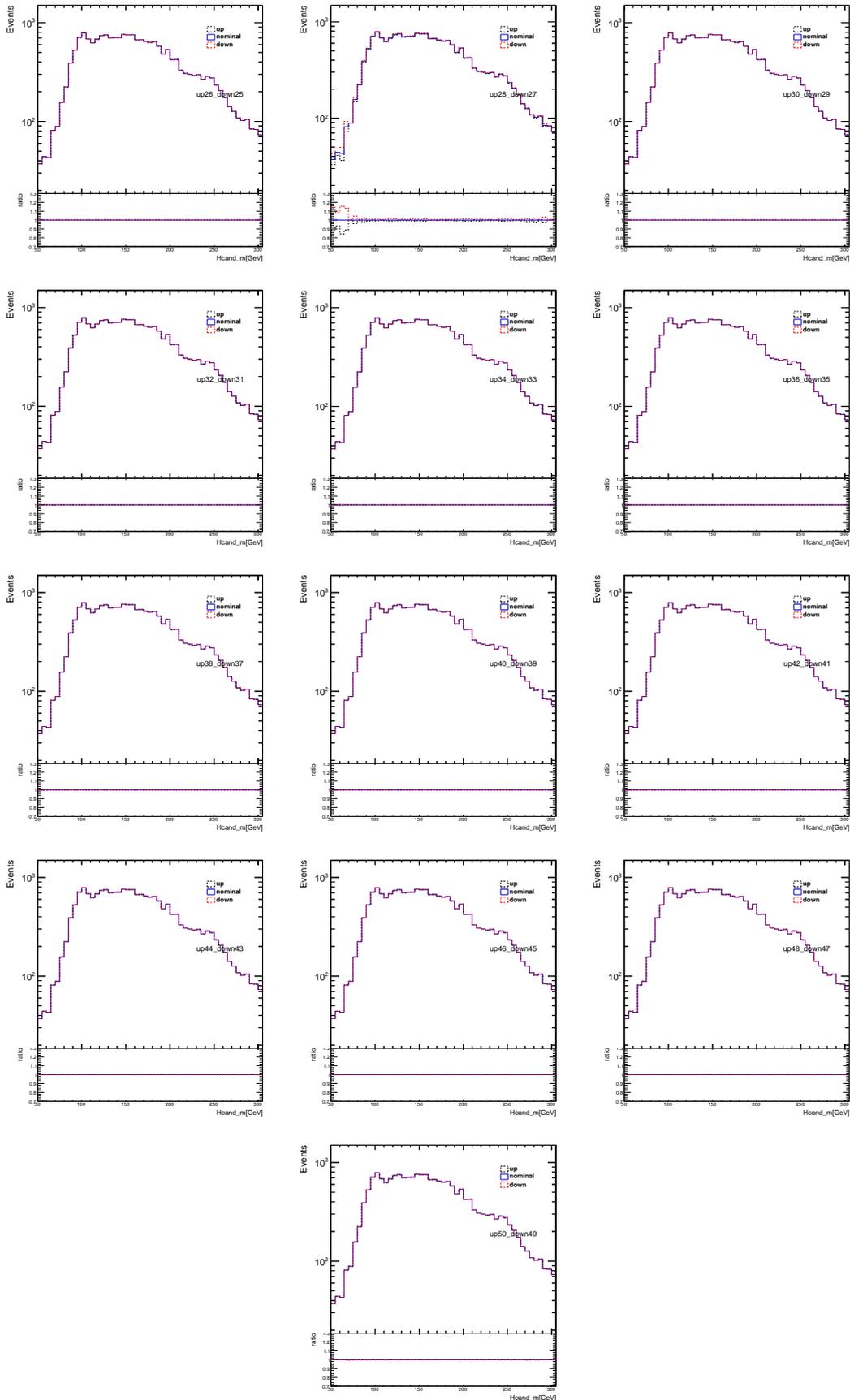


Figure C.24: The impact of the  $1\sigma$  up/down variations of light flavour uncertainties and extrapolation uncertainties on signal candidate large- $R$  jet mass distribution in  $Z'$  samples, using all generated mass points as a single template.

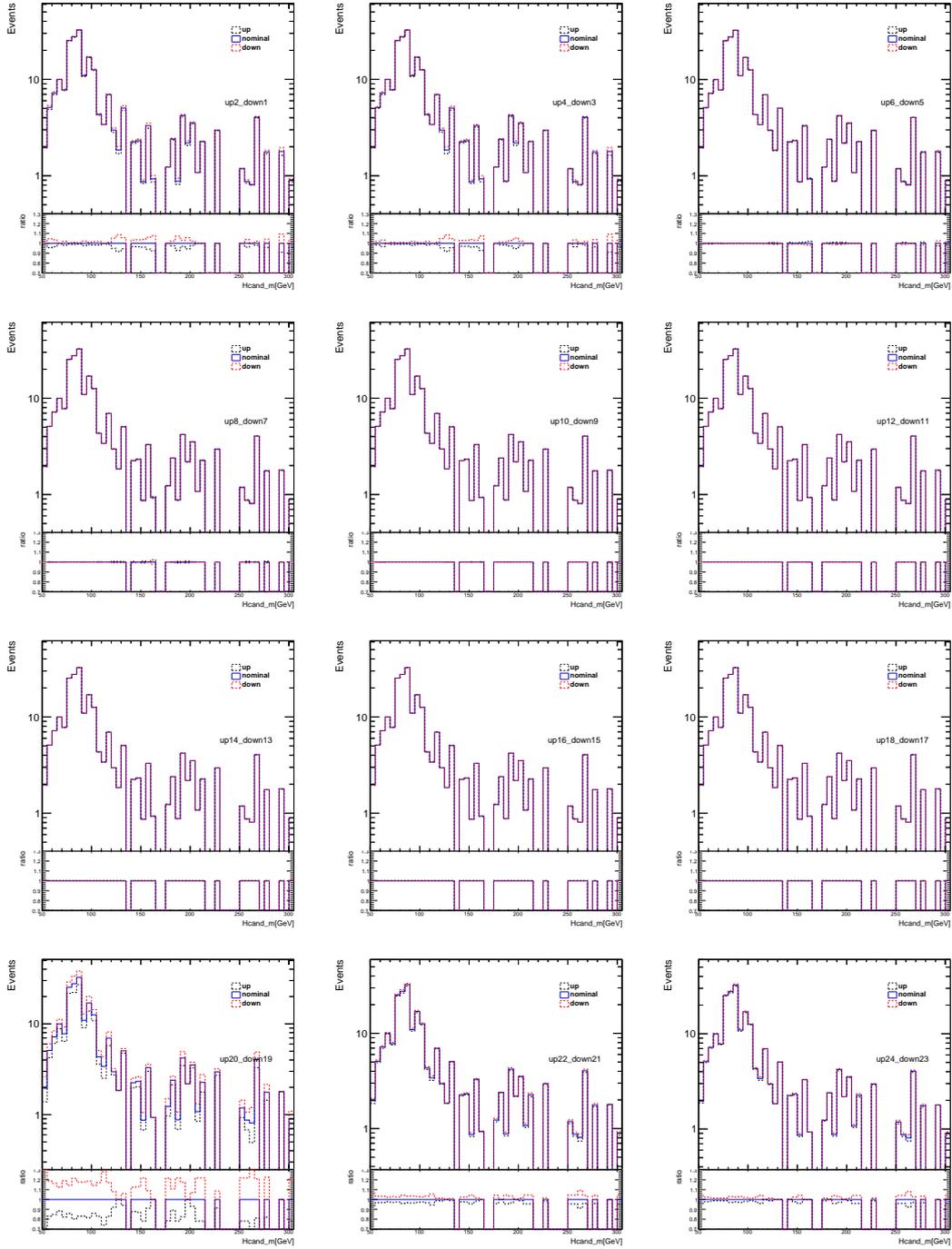


Figure C.25: The impact of the  $1\sigma$  up/down variations of  $b$ - and  $c$ - flavour uncertainties on signal candidate large- $R$  jet mass distribution in  $W + \text{jets}$  sample.

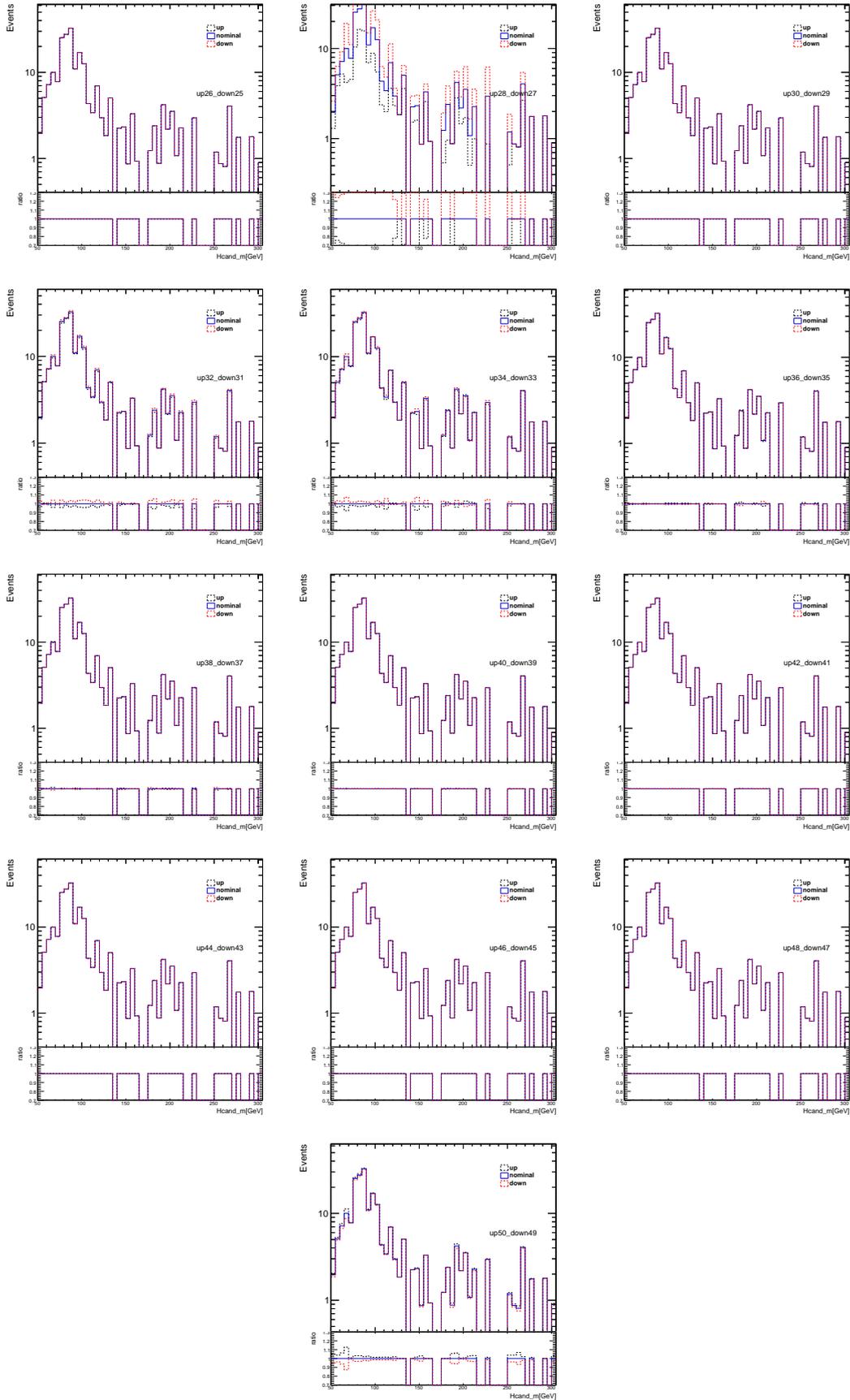


Figure C.26: The impact of the  $1\sigma$  up/down variations of light flavour uncertainties and extrapolation uncertainties on signal candidate large- $R$  jet mass distribution in  $W + \text{jets}$  sample.



# Appendix D

## Release and GRL

Release	Analysis	GRL
Rel 21	Chapter 8 (Di- <i>b</i> -jet)	data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Unknown_ PHYS_StandardGRL_All_Good_25ns.xml
		data16_13TeV.periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_ PHYS_StandardGRL_All_Good_25ns_ignore_TOROID_STATUS.xml
		data17_13TeV.periodAllYear_DetStatus-v99-pro22-01_Unknown_ PHYS_StandardGRL_All_Good_25ns_TriggerNo17e33prim.xml
Rel 20	Chapter 7 (Di-Higgs)	data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02_ PHYS_StandardGRL_All_Good_25ns.xml
		data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04_ PHYS_StandardGRL_All_Good_25ns.xml

Table D.1: The Good Run Lists and ATLAS Software information for the searches presented in this thesis.



# Bibliography

- [1] J. J. Thomson, *XL. Cathode Rays*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44** (1897) pp. 293–316, URL: <https://doi.org/10.1080/14786449708621070>.
- [2] Professor E. Rutherford, *LXXIX. The scattering of alpha and beta particles by matter and the structure of the atom*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **21** (1911) pp. 669–688, URL: <https://doi.org/10.1080/14786440508637080>.
- [3] D. Griffiths, *Introduction to Elementary Particles*, John Wiley&Sons, 1987 p. 55.
- [4] Wikipedia, “Standard Model of elementary particles”, [https://commons.wikimedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles\\_dark.svg](https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles_dark.svg), Accessed: 1.12.2018, 2018.
- [5] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 9781107034266, URL: <http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.2.T46::2013>.
- [6] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, *Prog. Theor. Phys.* **49** (1973) pp. 652–657.
- [7] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, *Phys. Rev. Lett.* **10** (12 1963) pp. 531–533, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.10.531>.
- [8] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, *Phys. Rev. Lett.* **30** (1973) pp. 1343–1346, [271(1973)].
- [9] Particle Data Group, *Review of Particle Physics*, *Phys. Rev. D* **98** (3 2018) p. 030001, URL: <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [10] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (16 1964) pp. 508–509, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [11] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (9 1964) pp. 321–323, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [12] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, *A Historical Profile of the Higgs Boson. An Updated Historical Profile of the Higgs Boson*, (2015) 22 p, URL: <https://cds.cern.ch/record/2012465>.

- [13] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*, Includes exercises, Westview, 1995, URL: <https://cds.cern.ch/record/257493>.
- [14] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) pp. 1–29, arXiv: [1207.7214](https://arxiv.org/abs/1207.7214) [hep-ph].
- [15] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) p. 30, arXiv: [1207.7235](https://arxiv.org/abs/1207.7235) [hep-ph].
- [16] URL: <http://www.grab-group.ethz.ch/research/higgs-boson.html>.
- [17] J. R. Ellis, “Limits of the Standard Model”, tech. rep. hep-ph/0211168. CERN-TH-2002-320, CERN, 2002, URL: <https://cds.cern.ch/record/591472>.
- [18] A. H. Broeils, K. G. Begeman, and R. H. Sanders, *Extended rotation curves of spiral galaxies: dark haloes and modified dynamics*, *Monthly Notices of the Royal Astronomical Society* **249** (1991) pp. 523–537, ISSN: 0035-8711, eprint: <http://oup.prod.sis.lan/mnras/article-pdf/249/3/523/18160929/mnras249-0523.pdf>, URL: <https://doi.org/10.1093/mnras/249.3.523>.
- [19] R. Massey, T. Kitching, and J. Richard, *The dark matter of gravitational lensing*, *Reports on Progress in Physics* **73** (2010) p. 086901.
- [20] Planck Collaboration, *Planck 2015 results - I. Overview of products and scientific results*, *A&A* **594** (2016) A1.
- [21] “ATLAS Physics Briefing”, URL: <https://atlas.cern/updates/physics-briefing/continuing-search-extra-dimensions>.
- [22] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, *The hierarchy problem and new dimensions at a millimeter*, *Physics Letters B* **429** (1998) pp. 263–272.
- [23] L. Randall and R. Sundrum, *Large Mass Hierarchy from a Small Extra Dimension*, *Phys. Rev. Lett.* **83** (17 1999) pp. 3370–3373, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.83.3370>.
- [24] I. Antoniadis, *A possible new dimension at a few TeV*, *Physics Letters B* **246** (1990) pp. 377–384.
- [25] J. Kretzschmar, *Searches for extra dimensions with the ATLAS and CMS detectors*, *Nuclear and Particle Physics Proceedings* **273-275** (2016) pp. 541–545.
- [26] T. Tong and M. Franklin, “Search for pair production of Higgs bosons in the four b quark final state with the ATLAS detector”, Presented 27 Apr 2018, 2018, URL: <https://cds.cern.ch/record/2634867>.
- [27] G. Angloher et al., *Commissioning run of the CRESST-II dark matter search*, *Astroparticle Physics* **31** (2009) pp. 270–276.

- [28] R. Bernabei et al., *The DAMA/LIBRA apparatus*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **592** (2008) pp. 297–315.
- [29] P. Belli, *Dark Matter direct detection: crystals*, Journal of Physics: Conference Series **718** (2016) p. 022003.
- [30] D. Akerib et al., *The Large Underground Xenon (LUX) experiment*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **704** (2013) pp. 111–126.
- [31] E. Aprile et al., *The XENON100 dark matter experiment*, Astroparticle Physics **35** (2012) pp. 573–590.
- [32] M. Ardid, *Dark Matter Searches with the ANTARES Neutrino Telescope*, Nuclear and Particle Physics Proceedings **273-275** (2016) pp. 378–382.
- [33] F. Halzen and S. R. Klein, *Invited Review Article: IceCube: An instrument for neutrino astronomy*, Review of Scientific Instruments **81** (2010) p. 081101.
- [34] M. Aguilar et al., *First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–350 GeV*, Phys. Rev. Lett. **110** (14 2013) p. 141102, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.110.141102>.
- [35] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001, URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08001>.
- [36] T. M. Hong, “Dark matter searches at the LHC”, tech. rep. ATL-PHYS-PROC-2017-100, Proceedings of the Fifth Annual LHCP: CERN, 2017, URL: <https://cds.cern.ch/record/2281114>.
- [37] E. Morgante, “Simplified Dark Matter Models”, tech. rep., 2018, arXiv: 1804.01245 [hep-ph], URL: <https://arxiv.org/pdf/1804.01245.pdf>.
- [38] A. Albert et al., “Recommendations of the LHC Dark Matter Working Group: Comparing LHC searches for heavy mediators of dark matter production in visible and invisible decay channels”, tech. rep., 2017, URL: <https://arxiv.org/pdf/1703.05703v2.pdf>.
- [39] ATLAS Collaboration, “Exotics Physics Public Results”, 2017, URL: <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/EXOTICS/>.
- [40] P. T. C. E. Pandini, “Theoretical studies for the 2HDM  $A \rightarrow Zh \rightarrow, h \rightarrow b\bar{b}$  search and the SM  $Vh \rightarrow b\bar{b}$  search”, tech. rep. ATL-COM-PHYS-2015-1474, CERN, 2015, URL: <https://cds.cern.ch/record/2111370>.
- [41] C. Grojean, E. Salvioni, M. Schlaffer, and A. Weiler, *Very boosted Higgs in gluon fusion*, Journal of High Energy Physics **2014** (2014) p. 22, ISSN: 1029-8479, URL: [https://doi.org/10.1007/JHEP05\(2014\)022](https://doi.org/10.1007/JHEP05(2014)022).

- [42] M. Schlaffer, M. Spannowsky, M. Takeuchi, A. Weiler, and C. Wymant, *Boosted Higgs shapes*, *The European Physical Journal C* **74** (2014) p. 3120, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-014-3120-z>.
- [43] 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>, 2016, URL: <https://cds.cern.ch/record/2227475>.
- [44] “Combined measurements of Higgs boson production and decay using up to 80 fb<sup>-1</sup> of proton–proton collision data at  $\sqrt{s} = 13$  TeV collected with the ATLAS experiment”, tech. rep. ATLAS-CONF-2018-031, CERN, 2018, URL: <http://cds.cern.ch/record/2629412>.
- [45] “CERN Accelerator Complex”, 2017, URL: [https://www.lhc-closer.es/taking\\_a\\_closer\\_look\\_at\\_lhc/0.cern\\_accelerator\\_complex](https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.cern_accelerator_complex).
- [46] ATLAS Collaboration, “ATLAS Luminosity Public Results”, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>, Accessed: 29.11.2018.
- [47] *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*, Technical Design Report CMS, There is an error on cover due to a technical problem for some items, CERN, 2006, URL: <https://cds.cern.ch/record/922757>.
- [48] LHCb Collaboration, *LHCb reoptimized detector design and performance: Technical Design Report*, Technical Design Report LHCb, CERN, 2003, URL: <http://cds.cern.ch/record/630827>.
- [49] ALICE Collaboration, *ALICE: Physics Performance Report. ALICE physics performance : Technical Design Report*, ed. by B Alessandro, vol. 32, Technical Design Report ALICE, revised version submitted on 2006-05-29 15:15:40, CERN, 2005, URL: <https://cds.cern.ch/record/879894>.
- [50] “ATLAS Collaboration”, URL: <https://atlas.cern/discover/collaboration>.
- [51] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [52] D. Sunday, “Distance between 3D Lines and Segments”, URL: [http://geomalgorithms.com/a07-\\_distance.html](http://geomalgorithms.com/a07-_distance.html).
- [53] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, *The European Physical Journal C* **70** (2010) pp. 787–821, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-010-1366-7>.
- [54] “Track Reconstruction Performance of the ATLAS Inner Detector at  $\sqrt{s} = 13$  TeV”, tech. rep. ATL-PHYS-PUB-2015-018, CERN, 2015, URL: <https://cds.cern.ch/record/2037683>.

- [55] ATLAS IBL collaboration, *ATLAS IBL Pixel Upgrade*, Nuclear Physics B **215** (2011) p. 147, arXiv: [1104.1980](https://arxiv.org/abs/1104.1980).
- [56] ATLAS Collaboration, “Impact Parameter Resolution”, URL: <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2015-007/>.
- [57] ATLAS Collaboration, *Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties*, Journal of Instrumentation **12** (2017) P05002–P05002, URL: <https://doi.org/10.1088%2F1748-0221%2F12%2F05%2Fp05002>.
- [58] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Phys. Rev. D **96** (7 2017) p. 072002, URL: <https://link.aps.org/doi/10.1103/PhysRevD.96.072002>.
- [59] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at  $\sqrt{s} = 13$  TeV*, The European Physical Journal C **76** (2016) p. 292, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-016-4120-y>.
- [60] M. Aaboud et al., *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. **C77** (2017) p. 317, arXiv: [1611.09661](https://arxiv.org/abs/1611.09661) [hep-ex].
- [61] ATLAS Collaboration, “ATLAS Trigger Operation Public Results”, [https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults#Trigger\\_rates\\_and\\_bandwidth](https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults#Trigger_rates_and_bandwidth), Accessed: 6.3.2019.
- [62] ATLAS Collaboration, “EXOT8 Derivation”, <https://svnweb.cern.ch/trac/atlasoff/browser/PhysicsAnalysis/DerivationFramework/DerivationFrameworkExotics/trunk/share/EXOT8.py>.
- [63] ATLAS Collaboration, “FTAG5 Derivation”, <https://svnweb.cern.ch/trac/atlasoff/browser/PhysicsAnalysis/DerivationFramework/DerivationFrameworkFlavourTag/trunk/share/FTAG5.py>.
- [64] ATLAS Collaboration, “Good Run Lists For Analysis Run 2”, <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/GoodRunListsForAnalysisRun2>.
- [65] “Boosted Higgs ( $\rightarrow b\bar{b}$ ) Boson Identification with the ATLAS Detector at  $\sqrt{s} = 13$  TeV”, tech. rep. ATLAS-CONF-2016-039, CERN, 2016, URL: <http://cds.cern.ch/record/2206038>.
- [66] ATLAS Collaboration, *Search for pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Journal of High Energy Physics **2019** (2019) p. 30, ISSN: 1029-8479, URL: [https://doi.org/10.1007/JHEP01\(2019\)030](https://doi.org/10.1007/JHEP01(2019)030).
- [67] “Search for boosted resonances decaying to two b-quarks and produced in association with a jet at  $\sqrt{s} = 13$  TeV with the ATLAS detector”, tech. rep. ATLAS-CONF-2018-052, CERN, 2018, URL: <https://cds.cern.ch/record/2649081>.

- [68] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) p. 852, arXiv: [0710.3820](https://arxiv.org/abs/0710.3820).
- [69] S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) p. 250.
- [70] D. J. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) pp. 152–155, BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S0168900201000894>.
- [71] T. Gleisberg et al., *Event generation with SHERPA 1.1*, (2008), eprint: [arXiv:0811.4622](https://arxiv.org/abs/0811.4622).
- [72] “ATLAS Run 1 Pythia8 tunes”, tech. rep. ATL-PHYS-PUB-2014-021, CERN, 2014, URL: <https://cds.cern.ch/record/1966419>.
- [73] S. Carrazza, S. Forte, and J. Rojo, “Parton Distributions and Event Generators”, 2013, eprint: [arXiv:1311.5887](https://arxiv.org/abs/1311.5887).
- [74] M. Sahinsoy et al., “Search for boosted pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”, tech. rep. ATL-COM-PHYS-2017-785, 2017, URL: <https://cds.cern.ch/record/2270230>.
- [75] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) p. 070, arXiv: [0709.2092](https://arxiv.org/abs/0709.2092).
- [76] H.-L. Lai et al., *New parton distributions for collider physics*, (2010), eprint: [arXiv:1007.2241](https://arxiv.org/abs/1007.2241).
- [77] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) p. 026, arXiv: [hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175).
- [78] J. Pumplin et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, (2002), eprint: [arXiv:hep-ph/0201195](https://arxiv.org/abs/hep-ph/0201195).
- [79] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Phys. Rev. D **82** (7 2010) p. 074018, URL: <https://link.aps.org/doi/10.1103/PhysRevD.82.074018>.
- [80] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **02** (2010) p. 037, arXiv: [0911.5299](https://arxiv.org/abs/0911.5299) [hep-ph].
- [81] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) p. 040, arXiv: [1410.8849](https://arxiv.org/abs/1410.8849) [hep-ph].
- [82] M. Bahr et al., *Herwig++ Physics and Manual*, (2008), eprint: [arXiv:0803.0883](https://arxiv.org/abs/0803.0883).
- [83] K. Hamilton, P. Nason, and G. Zanderighi, *MINLO: multi-scale improved NLO*, Journal of High Energy Physics **2012** (2012) p. 155.

- [84] K. Hamilton, P. Nason, and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, Journal of High Energy Physics **2015** (2015) p. 140.
- [85] ATLAS Collaboration, *Measurement of the  $Z/\gamma^*$  boson transverse momentum distribution in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, Journal of High Energy Physics **2014** (2014) p. 145.
- [86] B. Mellado Garcia, P. Musella, M. Grazzini, and R. Harlander, *CERN Report 4: Part I Standard Model Predictions*, (2016), URL: <https://cds.cern.ch/record/2150771>.
- [87] T. Plehn, M. Spira, and P. Zerwas, *Pair production of neutral Higgs particles in gluon-gluon collisions*, Nuclear Physics B **479** (1996) pp. 46–64.
- [88] S. Dawson, S. Dittmaier, and M. Spira, *Neutral Higgs-boson pair production at hadron colliders: QCD corrections*, Phys. Rev. D **58** (11 1998) p. 115012, URL: <https://link.aps.org/doi/10.1103/PhysRevD.58.115012>.
- [89] S. Borowka et al., *Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading Order with Full Top-Quark Mass Dependence*, Phys. Rev. Lett. **117** (1 2016) p. 012001, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.117.012001>.
- [90] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) p. 079, arXiv: [1405.0301](https://arxiv.org/abs/1405.0301).
- [91] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) p. 244, arXiv: [1207.1303](https://arxiv.org/abs/1207.1303).
- [92] C. Issever et al., “X  $\rightarrow$  bb Tagging: Baseline Tagger Support Documentation”, tech. rep. ATL-COM-PHYS-2017-1068, CERN, 2017, URL: <https://cds.cern.ch/record/2275335>.
- [93] D. Abercrombie et al., *Dark matter benchmark models for early LHC Run-2 searches: report of the ATLAS/CMS dark matter forum*, (2015), ed. by A. Boveia, C. Doglioni, S. Lowette, S. Malik, and S. Mrenna, arXiv: [1507.00966](https://arxiv.org/abs/1507.00966) [hep-ex].
- [94] Andrew Nelson, URL: <https://github.com/LHC-DMWG/DarkMatterWidthCalculation>.
- [95] J. Pequeno and P. Schaffner, “An computer generated image representing how ATLAS detects particles”, 2013, URL: <https://cds.cern.ch/record/1505342>.
- [96] F. Müller, “Jet production measurements at the ATLAS experiment”, PhD thesis: Ruperto-Carola-University of Heidelberg, 2013.
- [97] G. P. Salam, *Towards jetography*, The European Physical Journal C **67** (2010) pp. 637–686, URL: <https://doi.org/10.1140/epjc/s10052-010-1314-6>.

- [98] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti- $k(t)$  jet clustering algorithm*, *JHEP* **04** (2008) p. 063, arXiv: 0802.1189 [hep-ph].
- [99] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, *Longitudinally invariant  $K_t$  clustering algorithms for hadron hadron collisions*, *Nucl. Phys.* **B406** (1993) pp. 187–224.
- [100] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better jet clustering algorithms*, *Journal of High Energy Physics* **1997** (1997) p. 001, URL: <http://stacks.iop.org/1126-6708/1997/i=08/a=001>.
- [101] “Boosted Object Tagging with Variable- $R$  Jets in the ATLAS Detector”, tech. rep. ATL-PHYS-PUB-2016-013, CERN, 2016, URL: <https://cds.cern.ch/record/2199360>.
- [102] D. Krohn, J. Thaler, and L.-T. Wang, *Jets with variable  $R$* , *Journal of High Energy Physics* **2009** (2009) pp. 059–059.
- [103] G. Salam, “G. Salam ICTP Lectures, Basics of QCD: jets jet substructure”, 2015, URL: <https://gsalam.web.cern.ch/gsalam/repository/talks/2015-SaoPaulo-lecture4.pdf>.
- [104] D. Krohn, J. Thaler, and L. T. Wang, *Jet Trimming*, *JHEP* **02** (2010) p. 084, arXiv: 0912.1342 [hep-ph].
- [105] ATLAS Collaboration, *Performance of jet substructure techniques for large- $R$  jets in proton-proton collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *Journal of High Energy Physics* **2013** (2013) p. 76, ISSN: 1029-8479, URL: [https://doi.org/10.1007/JHEP09\(2013\)076](https://doi.org/10.1007/JHEP09(2013)076).
- [106] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches*, *Phys. Rev. D* **81** (9 2010) p. 094023, URL: <https://link.aps.org/doi/10.1103/PhysRevD.81.094023>.
- [107] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider*, *Phys. Rev. Lett.* **100** (24 2008) p. 242001, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.100.242001>.
- [108] W. Lampl et al., *Calorimeter clustering algorithms: Description and performance*, (2008).
- [109] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Rev.* **D96** (2017) p. 072002, arXiv: 1703.09665 [hep-ex].
- [110] “Tagging and suppression of pileup jets with the ATLAS detector”, tech. rep. ATLAS-CONF-2014-018, CERN, 2014, URL: <https://cds.cern.ch/record/1700870>.
- [111] S. Schätzel, *Boosted top quarks and jet structure*, *The European Physical Journal C* **75** (2015) p. 415, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-015-3636-x>.

- [112] The ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at  $\sqrt{s} = 7$  TeV*,  
*The European Physical Journal C* **73** (2013) p. 2304, ISSN: 1434-6052,  
URL: <https://doi.org/10.1140/epjc/s10052-013-2304-2>.
- [113] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in 13 TeV proton-proton collisions with the ATLAS detector*,  
*The European Physical Journal C* **79** (2019) p. 135.
- [114] ATLAS Collaboration,  
“Jet mass reconstruction with the ATLAS Detector in early Run 2 data”,  
tech. rep. ATLAS-CONF-2016-035, CERN, 2016,  
URL: <https://cds.cern.ch/record/2200211>.
- [115] “Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector”, tech. rep. ATL-PHYS-PUB-2014-013, CERN, 2014,  
URL: <http://cds.cern.ch/record/1750681>.
- [116] “Optimisation and performance studies of the ATLAS  $b$ -tagging algorithms for the 2017-18 LHC run”, tech. rep. ATL-PHYS-PUB-2017-013, CERN, 2017,  
URL: <http://cds.cern.ch/record/2273281>.
- [117] K. Krizka et al., “Search for boosted dijet resonances decaying to two  $b$ -quarks and produced in association with a jet”, tech. rep. ATL-COM-PHYS-2018-300, CERN, 2018, URL: <https://cds.cern.ch/record/2310645>.
- [118] “Optimisation of the ATLAS  $b$ -tagging performance for the 2016 LHC Run”, tech. rep. ATL-PHYS-PUB-2016-012, CERN, 2016,  
URL: <https://cds.cern.ch/record/2160731>.
- [119] ATLAS Collaboration, *Measurements of  $b$ -jet tagging efficiency with the ATLAS detector using  $t\bar{t}$  events at  $\sqrt{s} = 13$  TeV*,  
*Journal of High Energy Physics* **2018** (2018) p. 89, ISSN: 1029-8479,  
URL: [https://doi.org/10.1007/JHEP08\(2018\)089](https://doi.org/10.1007/JHEP08(2018)089).
- [120] ATLAS Collaboration, “BTaggingBenchmarksRelease20-21”,  
<https://twiki.cern.ch/twiki/bin/view/AtlasProtected/FlavourTagging>,  
2019.
- [121] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at  $\sqrt{s} = 13$  TeV*,  
*The European Physical Journal C* **76** (2016) p. 292, ISSN: 1434-6052,  
URL: <https://doi.org/10.1140/epjc/s10052-016-4120-y>.
- [122] ATLAS Collaboration,  
*Observation of  $H \rightarrow b\bar{b}$  decays and  $VH$  production with the ATLAS detector*,  
*Physics Letters B* **786** (2018) pp. 59–86, ISSN: 0370-2693, URL:  
<http://www.sciencedirect.com/science/article/pii/S0370269318307056>.
- [123] *Observation of Higgs Boson Decay to Bottom Quarks*,  
*Phys. Rev. Lett.* **121** (12 2018) p. 121801,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.121.121801>.
- [124] W.J.Stirling, “Parton luminosity and cross section plots”,  
<http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>, 2019.

- [125] ATLAS Collaboration, *Performance of jet substructure techniques for large- $R$  jets in proton-proton collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *Journal of High Energy Physics* **2013** (2013) p. 76, ISSN: 1029-8479, URL: [https://doi.org/10.1007/JHEP09\(2013\)076](https://doi.org/10.1007/JHEP09(2013)076).
- [126] M. Cacciari, G. P. Salam, and G. Soyez, *The catchment area of jets*, *Journal of High Energy Physics* **2008** (2008) pp. 005–005, URL: <https://doi.org/10.1088%2F1126-6708%2F2008%2F04%2F005>.
- [127] “Variable Radius, Exclusive- $k_T$ , and Center-of-Mass Subjet Reconstruction for Higgs( $\rightarrow b\bar{b}$ ) Tagging in ATLAS”, tech. rep. ATL-PHYS-PUB-2017-010, CERN, 2017, URL: <https://cds.cern.ch/record/2268678>.
- [128] J. Shelton, *TASI Lectures on Jet Substructure*, (2013), arXiv: [1302.0260v2](https://arxiv.org/abs/1302.0260v2) [[hep-ph](https://arxiv.org/abs/1302.0260v2)].
- [129] S. Sapeta, *QCD and jets at hadron colliders*, *Progress in Particle and Nuclear Physics* **89** (2016) pp. 1–55, ISSN: 0146-6410, URL: <http://www.sciencedirect.com/science/article/pii/S014664101600034X>.
- [130] “Expected Performance of Boosted Higgs ( $\rightarrow b\bar{b}$ ) Boson Identification with the ATLAS Detector at  $\sqrt{s} = 13$  TeV”, tech. rep. ATL-PHYS-PUB-2015-035, CERN, 2015, URL: <http://cds.cern.ch/record/2042155>.
- [131] M. Sahinsoy, *Boosted  $H \rightarrow b\bar{b}$  Tagger in Run II*, *PoS ICHEP2016* (2017) p. 1129.
- [132] “Impact of Alternative Inputs and Grooming Methods on Large- $R$  Jet Reconstruction in ATLAS”, tech. rep. ATL-PHYS-PUB-2017-020, CERN, 2017, URL: <http://cds.cern.ch/record/2297485>.
- [133] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C72** (2012) p. 1896, arXiv: [1111.6097](https://arxiv.org/abs/1111.6097) [[hep-ph](https://arxiv.org/abs/1111.6097)].
- [134] S. Dawson, S. Dittmaier, and M. Spira, *Neutral Higgs-boson pair production at hadron colliders: QCD corrections*, *Phys. Rev. D* **58** (11 1998) p. 115012, arXiv: [hep-ph/9703387](https://arxiv.org/abs/hep-ph/9703387).
- [135] J. Grigo, J. Hoff, K. Melnikov, and M. Steinhauser, *On the Higgs boson pair production at the LHC*, *Nucl. Phys. B* **875** (2013) p. 1, ISSN: 0550-3213, arXiv: [1305.7340](https://arxiv.org/abs/1305.7340).
- [136] D. de Florian and J. Mazzitelli, *Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD*, *Phys. Rev. Lett.* **111** (20 2013) p. 201801, arXiv: [1309.6594](https://arxiv.org/abs/1309.6594).
- [137] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, *Warped Gravitons at the LHC and Beyond*, *Phys. Rev. D* **76** (2007) p. 036006, arXiv: [hep-ph/0701186](https://arxiv.org/abs/hep-ph/0701186).
- [138] L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, *Searching for the Kaluza-Klein graviton in bulk RS models*, *JHEP* **09** (2007) p. 013, arXiv: [hep-ph/0701150](https://arxiv.org/abs/hep-ph/0701150).
- [139] G. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, *Phys. Rept.* **516** (2012) p. 1, arXiv: [1106.0034](https://arxiv.org/abs/1106.0034).

- [140] G. D. Kribs and A. Martin,  
*Enhanced di-Higgs production through light colored scalars*,  
Phys. Rev. D **86** (9 2012) p. 095023, arXiv: [1207.4496](https://arxiv.org/abs/1207.4496).
- [141] R. Grober and M. Muhlleitner,  
*Composite Higgs Boson Pair Production at the LHC*, JHEP **06** (2011) p. 020,  
arXiv: [1012.1562](https://arxiv.org/abs/1012.1562).
- [142] R. Contino et al., *Anomalous Couplings in Double Higgs Production*,  
JHEP **08** (2012) p. 154, arXiv: [1205.5444](https://arxiv.org/abs/1205.5444).
- [143] J. Baglio et al.,  
*The measurement of the Higgs self-coupling at the LHC: theoretical status*,  
Journal of High Energy Physics **2013** (2013) p. 151, ISSN: 1029-8479,  
URL: [https://doi.org/10.1007/JHEP04\(2013\)151](https://doi.org/10.1007/JHEP04(2013)151).
- [144] M. Grazzini et al.,  
*Higgs boson pair production at NNLO with top quark mass effects*,  
Journal of High Energy Physics **2018** (2018) p. 59, ISSN: 1029-8479,  
URL: [https://doi.org/10.1007/JHEP05\(2018\)059](https://doi.org/10.1007/JHEP05(2018)059).
- [145] “Combination of searches for Higgs boson pairs in  $pp$  collisions at 13 TeV with the ATLAS experiment.”, tech. rep. ATLAS-CONF-2018-043, CERN, 2018,  
URL: <https://cds.cern.ch/record/2638212>.
- [146] ATLAS Collaboration, *Searches for Higgs boson pair production in the  $hh \rightarrow bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$  channels with the ATLAS detector*,  
Phys. Rev. D **92** (9 2015) p. 092004,  
URL: <https://link.aps.org/doi/10.1103/PhysRevD.92.092004>.
- [147] ATLAS Collaboration,  
*Search for pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*,  
Phys. Rev. D **94** (5 2016) p. 052002,  
URL: <https://link.aps.org/doi/10.1103/PhysRevD.94.052002>.
- [148] D. C. Abbott et al., “Search for pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”,  
tech. rep. ATLAS-COM-CONF-2016-087, CERN, 2016,  
URL: <https://cds.cern.ch/record/2202750>.
- [149] ATLAS Collaboration, *Search for Higgs boson pair production in the  $b\bar{b}b\bar{b}$  final state from  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*,  
The European Physical Journal C **75** (2015) p. 412.
- [150] J. Alison et al., “Search for pair production of Higgs bosons in the resolved  $bbbb$  final state using proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector: (Analysis of 2015 + 2016 data) ”,  
tech. rep. ATL-COM-PHYS-2017-781, CERN, 2017,  
URL: <https://cds.cern.ch/record/2270210>.
- [151] ATLAS Collaboration,  
“ATLAS Data Preparation Checklist for Physics Analysis”,  
<http://twiki.cern.ch/twiki/bin/viewauth/Atlas/DataPreparationCheckListForPhysicsAnalysis>, Accessed: 6.3.2019.

- [152] ATLAS Collaboration, “How To Clean Jets 2016, Jet-Based Event Cleaning”, <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HowToCleanJets2016>.
- [153] “Identification of double-b quark jets in boosted event topologies”, tech. rep. CMS-PAS-BTV-15-002, CERN, 2016, URL: <https://cds.cern.ch/record/2195743>.
- [154] *Inclusive Search for a Highly Boosted Higgs Boson Decaying to a Bottom Quark-Antiquark Pair*, *Phys. Rev. Lett.* **120** (7 2018) p. 071802, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.120.071802>.
- [155] ATLAS Collaboration, *Search for Low-Mass Dijet Resonances Using Trigger-Level Jets with the ATLAS Detector in pp Collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. Lett.* **121** (8 2018) p. 081801, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.121.081801>.
- [156] The CMS collaboration, *Search for narrow and broad dijet resonances in proton-proton collisions at  $\sqrt{s} = 13$  TeV and constraints on dark matter mediators and other new particles*, *Journal of High Energy Physics* **2018** (2018) p. 130, ISSN: 1029-8479, URL: [https://doi.org/10.1007/JHEP08\(2018\)130](https://doi.org/10.1007/JHEP08(2018)130).
- [157] “Search for low-mass resonances decaying into two jets and produced in association with a photon using *pp* collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”, tech. rep. arXiv:1901.10917, \* Temporary entry \*: CERN, 2019, URL: <https://cds.cern.ch/record/2655302>.
- [158] ATLAS Collaboration, *Search for light resonances decaying to boosted quark pairs and produced in association with a photon or a jet in proton proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Physics Letters B* **788** (2019) pp. 316–335, ISSN: 0370-2693, URL: <http://www.sciencedirect.com/science/article/pii/S037026931830830X>.
- [159] “Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector”, tech. rep. ATLAS-CONF-2015-029, CERN, 2015, URL: <https://cds.cern.ch/record/2037702>.
- [160] ATLAS Collaboration, “How To Clean Jets 2017, Jet-Based Event Cleaning”, <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HowToCleanJets2017>.
- [161] ATLAS Collaboration, “Release 21 Large-*R* Jet MC Calibrations”, <https://twiki.cern.ch/twiki/pub/AtlasProtected/ApplyJetCalibrationR21/Rel21large-RjetMCcalibrationsJESJMS.pdf>.
- [162] “Improvements in  $t\bar{t}$  modelling using NLO+PS Monte Carlo generators for Run2”, tech. rep. ATL-PHYS-PUB-2018-009, CERN, 2018, URL: <http://cds.cern.ch/record/2630327>.
- [163] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *Annals Math. Statist.* **9** (1938) pp. 60–62.
- [164] T. Archdeacon, *Correlation and Regression Analysis: A Historian’s Guide*, University of Wisconsin Press, 1994, ISBN: 9780299136543, URL: <https://books.google.de/books?id=QmFFpQW8V8EC>.

- [165] ATLAS Collaboration, “Luminosity Uncertainties”, <https://twiki.cern.ch/twiki/bin/viewauth/Atlas/LuminosityForPhysics>, 2019.
- [166] ATLAS Collaboration, “ATLAS JetEtmis Recommendations”, <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetEtmisRecommendationsR21>, Accessed: 25.3.2019.
- [167] M. Czakon, D. Heymes, and A. Mitov, *Dynamical scales for multi-TeV top-pair production at the LHC*, *Journal of High Energy Physics* **2017** (2017) p. 71, ISSN: 1029-8479, URL: [https://doi.org/10.1007/JHEP04\(2017\)071](https://doi.org/10.1007/JHEP04(2017)071).
- [168] ATLAS Collaboration, “tth(bb) Analysis Run 2 studies”, <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TtHbbRun2studies>, Accessed: 1.11.2018.
- [169] J. M. Lindert et al., *Precise predictions for  $V$ +jets dark matter backgrounds*, *The European Physical Journal C* **77** (2017) p. 829, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-017-5389-1>.
- [170] G. Choudalakis, “On hypothesis testing, trials factor, hypertests and the BumpHunter”, *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, CERN, Geneva, Switzerland 17-20 January 2011, 2011, arXiv: [1101.0390](https://arxiv.org/abs/1101.0390) [physics.data-an], URL: <https://inspirehep.net/record/883244/files/arXiv:1101.0390.pdf>.
- [171] L. Lyons, *Bayes and Frequentism: a particle physicist’s perspective*, *Contemporary Physics* **54** (2013) pp. 1–16, eprint: <https://doi.org/10.1080/00107514.2012.756312>, URL: <https://doi.org/10.1080/00107514.2012.756312>.
- [172] G. Bohm and G. Zech, *Introduction to statistics and data analysis for physicists*, DESY, 2010, ISBN: 9783935702416, URL: <https://books.google.de/books?id=aQBHcgAACAAJ>.
- [173] A. Caldwell, D. Kollár, and K. Kröninger, *BAT – The Bayesian analysis toolkit*, *Computer Physics Communications* **180** (2009) pp. 2197–2209, ISSN: 0010-4655, URL: <http://www.sciencedirect.com/science/article/pii/S0010465509002045>.
- [174] L. Lista, *Practical Statistics for Particle Physicists*, CERN Yellow Reports: School Proceedings **5** (2017) p. 213, ISSN: 2519-805X, URL: <https://e-publishing.cern.ch/index.php/CYRSP/article/view/354>.
- [175] L. A. Beresford, “Searches for Dijet Resonances using  $\sqrt{s} = 13$  TeV proton–proton collision data recorded by the ATLAS detector at the Large Hadron Collider”, PhD thesis: Oxford U., 2017, URL: <https://ora.ox.ac.uk/objects/uuid:c516b04b-2763-487a-a633-6c506cec93ad>.



# Acknowledgements

I am much obliged to ATLAS Collaboration and HGSFP to make this study possible.

This thesis has greatly benefited from the support of the people whom I worked with in KIP ATLAS group and at CERN. First of all, I would like to thank my advisor Oleg Brandt for his academic guidance, support and trust. It has been a valuable experience to work with him. I would like to thank Prof. Hans-Christian Schultz-Coulon for the great working environment and passionate physics discussions he created in the ATLAS KIP group. I also thank all members of the ATLAS KIP group for their feedback in all stages of my doctoral studies. I would like to present my special thanks to Prof. Stephanie Hansmann-Menzemer for being the second referee of this thesis, being so understanding and supportive during this period.

I would like to thank my colleagues from CERN that I worked with in the presented analyses in this thesis, Felix Müller, Baoija Tong, Francesco Armando di Bello and Matthew Feickert, Stefan Maschek. I would like to thank also the analyses contacts that I worked with, Reina Coromoto Camacho Toro, Tatjana Lenz, Dan Guest, Çiğdem İşsever, Michael Kagan, Rafael Teixeira De Lima and Karol Krizka, for the fruitful discussions and their useful comments. I am grateful to Reina, Baoija and Karol for the technical support they provided for the analysis frameworks.

Not only performing physics analyses contributed to my knowledge, but also I learned a lot from my experiences in the ATLAS detector operation team in these years. Being a part of the devoted Trigger Operations group was an invaluable experience for me. Each person that I met in this group gave their unrequited support. I thank Catrin Bernius, Martin Zur Nedden, Jiri Masik, Joana Machado Miguens, Savanna Marie Shaw, Joerg Stelzer and Mark Stockton for training me as a Trigger Operation Expert.

Without their priceless friendship, it would not be possible for me to enjoy my time in Heidelberg. I would like to thank all of my friends in the KIP group and lovely Petra Pfeifer. I thank my great office mates Bastian Bernarding, Lars Henkelmann, Fabrizio Napolitano, Thomas Spieker and Sebastian Weber, and my dearest friends Claire Antel, Falk Bartels, Alessandra Baas, Nadia Bludau, Julia Djiuvslund, Sascha Dreyer, Manuel Geisler, Jan Jongmanns, Martin Klassen, Hanna Klungenmeyer, Valerie Lang, Philipp Ott and Daniel Narrias Villar. I am grateful to Stanislav Suchek, Hanno Meyer zu Theenhausen, Şafak Chasan, Ezgi and Taylan Kaya, Ecem Uludağ, and Katja Ovchinnikova for being there for me whenever I need.

I would like to thank my family for their endless support and love, Sevgi and Rifat Şahinsoy, my sister Melike Aslı Sim, and my nephew Aren Uzay Sim who managed to cheer me up everyday during the writing process of this thesis. Finally, I would like to thank specially Semih Karacasu who suffered with me in this process and supported me. I am deeply grateful for his contributions to this thesis and my doctoral studies in far too numerous ways to mention.