Search for pair production of Higgs Bosons decaying to four bottom quarks with data collected by the ATLAS detector

Lucas Santiago Borgna

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Physics and Astronomy
University College London.

1st April, 2022

Abstract

Using the full Run-2 data recorded by the ATLAS detector, a search for the elusive Higgs pair production decaying into four bottom quarks, $HH \to b\bar{b}b\bar{b}$, is presented in this thesis. The full Run-2 dataset corresponds to 126 fb^{-1} of integrated luminosity. The theoretical motivations for this search, which are summarized in this thesis, are clear as the search can probe the structure of the Higgs potential and Beyond the Standard Model physics.

To reconstruct the $HH \to b\bar{b}b\bar{b}$ events a combination of multi-b-jet triggers are used. Events are then selected if they contain at least four small radius jets that have passed the b-tagging selection. These jets are paired to reconstruct the Higgs candidates. As Monte Carlo simulations cannot reliably reproduce the $b\bar{b}b\bar{b}$ final state, a data-driven approach is taken to produce the background estimate. This makes use of a neural network to predict the background in the signal region. The data-driven approach is validated by the use of several orthogonal control samples.

The search is used to set exclusion limits at a 95 % confidence level for heavy resonances and non-resonant gluon-gluon fusion HH production. Two benchmark signals consisting of a spin-0 narrow width scalar and a spin-2 graviton were used for the resonant search. The upper limit on the cross-section of the non-resonant Standard Model HH production via gluon-gluon fusion was observed to be 5.1 times the Standard Model prediction. The trilinear Higgs self-coupling was constrained to the range of [-6.0, 15.0] times the Standard Model prediction. The improvements made to the $b\bar{b}b\bar{b}$ channel have made the search competitive with the other final states. These optimizations will be useful to maximize the potential of the HL-LHC program.

Impact Statement

The inception of particle physics has brought forth several ground-breaking discoveries in physics in the form of the Standard Model, which is used to describe our universe. Research in particle physics has also acted as a catalyst for technological advancements that are deeply entrenched in our society. The Standard Model has, so far, been hugely successful but there are still some questions about nature it cannot answer. The work presented in this thesis focuses on the search for pair production of Higgs Bosons that subsequently decay into four bottom quarks. This search can be used to probe deeply into the underlying mechanics of the Standard Model, which could then direct how it should be extended. To search for this decay, data recorded by the ATLAS detector at CERN's Large Hadron Collider is used. The results presented in this thesis are also compared with similar searches by both ATLAS and CMS collaborations to serve as a benchmark. The future projections of the search are also discussed.

To reconstruct the events measured by the ATLAS detector hundreds of petabytes of data needs to be processed. Naturally, this puts data analysis at the core of the work presented in this thesis. Unfortunately, the pair production of Higgs Bosons occurs quite infrequently and during the whole Run-2 data taking period of the ATLAS detector only about 100 of these events should occur. Conversely, the expected background of this search occurs with a frequency that is several orders of magnitude higher. To maximize the potential of this search, novel machine learning techniques are used. The usage of machine learning techniques has been a standard practice in particle physics for decades now but are now becoming of paramount importance in the data science field and in fact society. For example, machine learning has been used to model the number of COVID-19 patients in hospitals so that the appropriate policies could be implemented. Moreover, the statistical framework used in this work can be directly extrapolated to real world scenarios, such as testing for the efficacy of a new drug.

Lastly, the scale of the theoretical and technological challenges that particle physics presents demands a collaborative approach that unifies individuals, institutions, and states from all over the world. Without this level of collaboration research at the frontier of human understanding would not be possible. CERN and the wider particle physics community serve as an example to unify society to tackle the biggest questions.

Declaration

I, Lucas Santiago Borgna, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

It has just dawned on me that I am writing this in the same corner of the UCL science library that I was four years ago writing my master's thesis. Quite frankly, it still feels surreal that I am writing this now and I am eternally grateful for all of the people that helped me get here.

Firstly, I would like to thank my supervisor Nikolaos Konstantinidis for introducing me to the vast world of Higgs searches. It has been a pleasure to work alongside you and I am especially thankful for the support you have given me throughout the years and in particular during the pandemic. I would also like to extend my gratitude to David Sankey and Matt Warren for their expertise (and patience!) with the firmware design for my qualification task.

I do not think I could have asked for a better group of colleagues to tackle the 4b search and I am grateful to have been part of this group. Without many of them much of the work here would not have been possible. Essentially, I am not really sure where I would have ended up without the support from Maximilian Swiatlowski. Thank you for babysitting and on-boarding Charlie and I to the 4b group. Moreover, I am also thankful for all the advice and input from Rafael Lima and Liaoshan Shi to dig deeper into the many studies and problems we found along the way. At times the difficulties of the 4b search created a sense of hopelessness, but this was always quickly washed away by the positive outlook from Nicole Hartman. You have been absolutely magnificent to work with and I am thankful that you were a part of this group (I am so glad we ended up making the "cryptotuples"). I am also perpetually thankful to Sean Gasiorowski and I apologize for the multitude of bugs that made you mutter the phrase "Ugh, this guy" (I cleaned up the language). In a similar fashion I am also grateful to Beojan Stanislaus for all the technical support that enabled me to perform much of my studies. Although our overlap at CERN was short-lived, I'm also thankful to Dale Abbott, and his wisdom on the boosted search. Needless to say, I owe this whole group several rounds of drinks and I look forward to see what you come up with in Run-3.

Luckily I was joined by 17 other students to form the first ever CDT in Data Intensive Sciences at UCL. I enjoyed going through the course with every one of you. In particular, I want to thank Charlie Donaldson, sharing that flat in Geneva was awesome and one day I will beat you in snooker (or smash bros if that is even possible). I am also thankful to Greg Barbour, Davide Piras, Damien de Mijolla and Paddy Roddy for all the laughter and welcome distractions. I am also fortunate to have worked and skiied with Ioannis Xiotidis while out at CERN and I am glad that our paths have crossed once again. One of the upsides of growing up abroad is the friendships I have made. These friendships have kept me going through difficult times and I am incredibly thankful to Constantin von Stegmann, Kathryn Woolston-Thomas, Gabriella Andriesz and Olivia Buttar.

I am ever thankful for my parents, Armando Borgna and Diana Nocete, as well as my sister Ludmila Borgna. It has never been easy being spread across the globe in three different continents with over 10,000 km of distance between each other. However, even with this distance I have always had their support and sympathetic ears. This has probably been the longest we have gone without physically seeing each other and now that the world is opening up again I look forward to come home.

Ultimately, I cannot express how thankful I am for my partner Srishti Bhasin. The support and encouragement you provided me over these years have been transformative. Meeting you in R1 has been the most significant occurrence in my life, and I am glad that we had each other to endure this pandemic together. I cannot wait to see where we go to next.

Preface

This thesis describes the adventures of searching for the rare pair production of Higgs Bosons, which subsequently decay into a pair of bottom quarks each. To perform this search data collected by the ATLAS detector during the full Run-2 data taking period is used. The theoretical concepts surrounding this search are discussed in Chapter 2, which will provide strong motivations for the search. Machine learning aspects that are used in the search are introduced in Chapter 3. The statistical framework used to derive meaningful conclusions from the available data are described in Chapter 4. The physical details about the ATLAS detector are summarized in Chapter 5 and the relevant objects used in the $4b$ search are presented in Chapter 6. Having described all the relevant cogs of the search, the overall analysis strategy is outlined in Chapter 7. In depth details about the background estimation and trigger optimization are reserved for Chapters 8 and 9 respectively. Finally, the results of unblinding the signal region of interest in the $4b$ search are presented in Chapter 10. The conclusions of this thesis are put forth in Chapter 11.

My main contribution to the analysis has been the development and optimization of the trigger strategy. The trigger strategy was a critical development on the resonant search since this search covers a broad range of masses and a single trigger could not provide the sensitivity across the board. The combination of triggers through the trigger buckets scheme, which at first had a significant rate of events discarded, was a key aspect in this search. Similarly, the strategy developed for the non-resonant search allowed us to increase the sensitivity. Furthermore, the trigger matching algorithm was also another contribution that was needed in the 4b search. The implementation of the trigger strategy into the analysis software framework was aided by the support from Beojan Stanislaus and Maximilian Swiatlowski.

The background estimation strategy and uncertainty derivation were primarily developed by Sean Gasiorowski. To support their development, I contributed by validating the background estimation strategy with control samples such as the reversed $|\Delta\eta_{HH}|$. Since

the background estimation uses a full data driven approach this was one of the critical developments to provide us with the confidence to go forth with this approach. Not only that, but it was also needed to get the unblinding approval in the resonant search.

Along the journey of the search, I also contributed with several studies that influenced the event selection of both the resonant and non-resonant search. All of these studies have also been documented in the internal documentation of the search. Moreover, I have also been an active developer and maintainer of the code bases that make up the $HH \to b\bar{b}b\bar{b}$ search. One of these contributions was the development of Continuous Integration workflows in the code base, which may sound like technical jargon, but it quickly became an essential part of the software framework.

Ultimately, all of this work has led to the publication of the $HH \to b\bar{b}b\bar{b}$ resonant search using the full Run-2 data recorded by ATLAS. Additionally, the non-resonant search is currently undergoing the ATLAS review process, which should hopefully result in another publication.

Contents

List of figures

List of tables

Acronyms

2HDM Two-Higgs-Doublet-Model. 12

ASIC Application Specific Integrated Circuit. 52

ATLAS A Toroidal LHC ApparatuS. xxiii, 2, 13, 16, 18–21, 32, 39–42, 44–46, 48, 49, 51, 55, 56, 60, 62, 64, 69, 74, 75, 133, 163

BC Bunch Crossing. 39, 40

BDT Boosted Decision Tree. xii, 23, 64, 69, 79–81, 93, 98

BSM Beyond the Standard Model. 1, 12, 13, 15, 21, 163

C/A Cambridge/Aachen. 56

CKM Cabibbo-Kobayashi-Maskawa. 63

CMS Compact Muon Solenoid. 13, 16, 18, 21, 40, 45, 52

CR Control Region. 88, 89, 100, 193

CS Collinear Safe. 56

CSC Cathode Strip Chambers. 50

CTP Central Trigger Processor. 52

DAQ Data Acquisition. 51

DNN Deep Neural Network. 64

ECal Electromagnetic calorimeter. 46

EMD Earth-Mover's Distance. 101

FCal LAr Forward calorimeter. 46, 49

FPGA Field Programmable Gate Array. 51, 52

FTK Fast Tracker. 52, 53

ggF gluon-gluon fusion. 13, 15, 75

GPD General Purpose Detector. 40

GSC Global Sequential Calibration. 58

GSW Glashow-Weinberg-Salam. 5

HC Higgs Candidates. 78, 79, 81, 85

HL-LHC High Luminosity LHC. 21, 39, 42, 51–53

HLT High-Level Trigger. 51, 52, 67, 69

HTT Hardware Tracking for Trigger. 52

IBL Insertable B-layer. 43

ID Inner Detector. 42–45, 49, 57

IP Interaction Point. 40, 44, 48

IQR Interquartile Range. 106

IS Infrared Safe. 56

ITk Inner Tracker. 42

KF Kalman Filter. 63

KK Kaluza-Klein. 13

L1 Level-1. 51, 52, 69

L1Calo Level-1 Calo. 52

L1Muon Level-1 Muon. 52

L1Topo Level-1 Topological trigger. 52

LAr Liquid Argon Calorimeter. 45, 49

LHC Large Hadron Collider. 39–41, 56

LLR log-likelihood ratio. 63

LO Leading Order. 75

MC Monte Carlo. 61, 67, 75, 76, 93

MDT Monitored Drift Tubes. 49, 50

ML Machine Learning. 2

MLE Maximum Likelihood Estimation. 34

MLP Multi-Layer Perceptron. 27

MS Muon Spectrometer. 45, 49–52

MVA Multivariate Analysis. 64, 93

NLO Next-to-leading Order. 75

NN Neural Network. 23, 93, 99, 104

NP Nuissance Parameter. 32

NSW New Small Wheel. 51

PDF probability distribution function. 62, 63

PFlow Particle Flow. 57, 58

PID Particle Identification. 44

POI Parameter of Interest. 32

PU Processing Unit. 51

QCD Quantum Chromodynamics. 5, 84

QFT Quantum Field Theory. 3

ReLU Rectified Linear Unit. 28, 102

RMSE Root Mean Squared Error. 24

RNN Recurrent Neural Network. 64

ROI Regions of Interest. 52

RPC Resistive Plate Chambers. 50

RS Randall-Sundrum. 12, 13

SCT Silicon Microstrip Tracker. 42, 44

SGD Stochastic Gradient Descent. 30

SHAP SHapley Additive exPlanations. 30, 101, 103

SM Standard Model. xix, 1–5, 11, 12, 15–19, 21, 32, 73, 75, 76, 81, 158, 160

SMT Soft Muon Tagger. 64, 65

SR Signal Region. 88, 89, 91, 93, 100, 144

SSB Spontaneous Symmetry Breaking. 7

SV Secondary Vertex. 63

TDAQ Trigger and Data Acquisition. 51, 52

TGC Thin Gap Chambers. 50

TRT Transition Radiation Tracker. 42, 44, 50

VBF Vector Boson Fusion. 13

VEV Vacuum Expectation Value. 10, 15, 55

VR Validation Region. 88, 89, 91, 100, 193

WIMP Weakly Interacting Massive Particle. 11

Chapter 1.

Introduction

Since the genesis of humanity we have been wondering what matter is made up of and how it is bound together. The Standard Model (SM) of particle physics is currently our most complete answer to this age old question. Experiments carried out all over the earth using particle accelerators and detectors have been continually corroborating the SM since its inception. Although, the SM is one of the most remarkable scientific achievements there are several experimental observations that cannot be explained by the model. Perhaps most dauntingly is the omission of any description of gravity in the SM. Likewise, there is no description of candidate dark matter particles in the SM, meaning that it can only explain 5 % of the matter in the universe. Additionally, it is unclear if the SM can explain the matter-antimatter asymmetry in the universe. These are examples of key gaps in the SM that has propelled the high energy physics community to search for Beyond the Standard Model (BSM) processes that can explain these phenomena.

The measurement of the Higgs Boson mass in 2012 was a milestone achievement as it unveiled the missing cornerstone of the theory. This however did not provide a complete picture of the Higgs sector and all of its properties. Most notably, the Higgs potential is still largely unknown and this potential governs the exact dynamics of the Higgs interactions. To gain access to this potential, a constant known as the Higgs self-coupling, which describes the strength of interaction of the Higgs field with itself, needs to be experimentally constrained. The key that will give us access to this property is the pair production of two Higgs Bosons, HH. Fortunately, this production mode can additionally be used to directly search for BSM resonances that could answer some of the big questions in the SM. Unfortunately, the HH production mode is extremely rare and difficult to detect. This thesis will explore the search of HH decay to four bottom quarks, $b\bar{b}b\bar{b}$, as this has the largest branching ratio among the decay modes.

This thesis is organized as follows. Initially an overview of the theoretical aspects of the SM and the Higgs mechanism that motivates the search for HH production is given in Chapter 2. Additionally, Machine Learning (ML) concepts that are applicable to the $HH \to b\bar{b}b\bar{b}$ search are given in Chapter 3. The basics of how the statistical framework is used to extract meaningful conclusions of the data is outlined in Chapter 4. To assist the understanding of the objects used in the $HH \to b\bar{b}b\bar{b}$ event reconstruction the ATLAS detector and some of its event reconstruction algorithms are outlined in Chapters 5 and 6 respectively. A summary of the $HH \to b\bar{b}b\bar{b}$ analysis strategy is described in Chapter 7, while Chapter 8 provides further details into the background estimate and Chapter 9 discusses the optimization of the triggers used in the search. The results obtained by unblinding the signal region of interest are presented in Chapter 10 and concluding remarks are given in Chapter 11.

Chapter 2.

Theory

"Every particle physicist should write out the full Standard Model lagrangian at least once in their life."

— Nikolaos Mavromatos

In this chapter, the current theoretical framework of particle physics, known as the Standard Model (SM), is discussed. The main aspect of this intricate and detailed theory relevant to HH searches is the Higgs mechanism, which as we will see, is a cornerstone of the SM theory. Moreover, a summary of the phenomenology of HH searches and their results are summarized at the end of this chapter.

2.1. Introduction to the Standard Model

The SM of particle physics is the most successful theory to date that describes the smallest building blocks of matter and how they interact together. The SM is built upon the Quantum Field Theory (QFT) framework. This requires that a physical process is described by a Lagrangian density $\mathcal{L}$ that is invariant under a local transformation, which is a rather strict constraint. If this invariance is lost, then the theory becomes unrenormalizable and meaningful predictions cannot be made. The particles that make up the SM are excited states of quantum fields and are defined by a set of quantum numbers. The key quantum number of particles is their spin, which allows particles to be grouped into *fermions* and *Bosons*. All fermionic particles in the SM have half-integer spin ($\pm 1/2$) and obey Fermi-Dirac statistics. Meanwhile, Bosons are identified by their

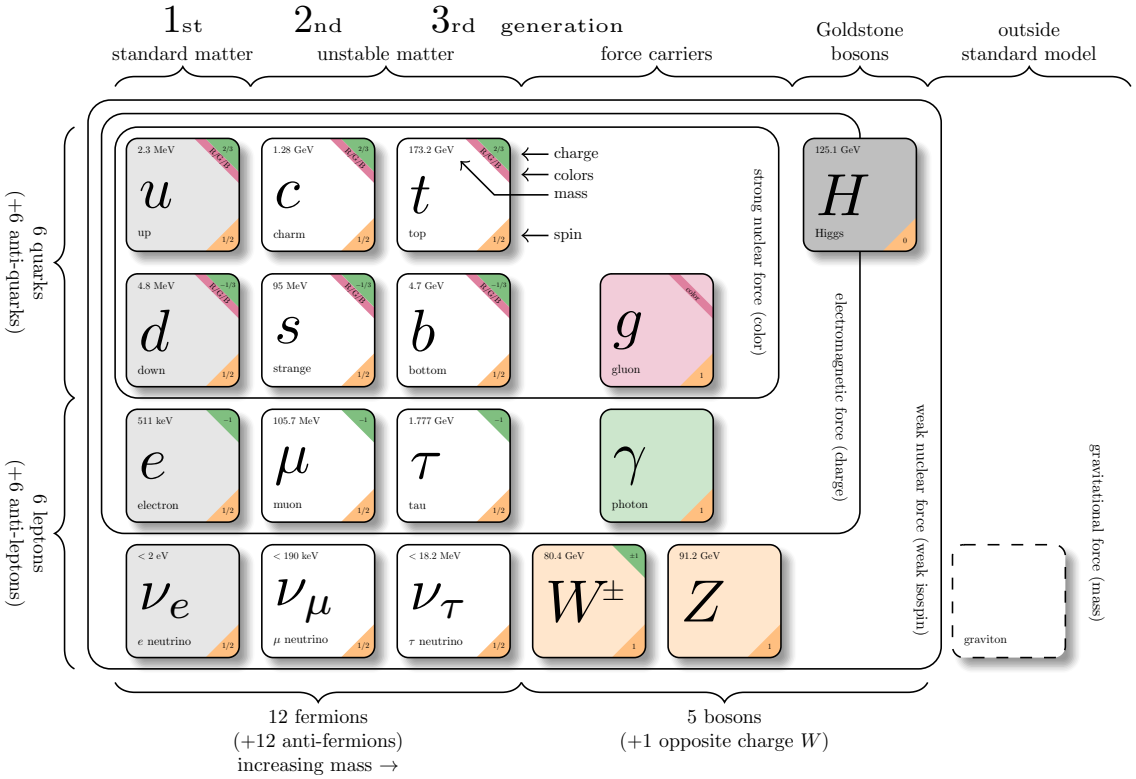


Figure 2.1.: Fundamental particles of the Standard Model [1]

integer spin ($0,\ \pm 1,\ \pm 2$) and follow Bose-Einstein statistics. A summary of all the fundamental SM particles can be found in Figure 2.1. The SM also describes interactions between these particles through the three fundamental forces, which are the strong, weak and electromagnetic forces.

One of the major successes of the SM was to unify the weak and electromagnetic theories through the *electroweak* interaction. The mediators of the electroweak force are the $W^{\pm}$, Z^0 and γ (photon) particles, while mediator of the strong force is the g (gluon). All of the mediators of these forces are spin-1 vector Bosons. Notably, gravity is omitted from the SM as it operates on a vastly different scale to these other forces and is more adequately described by General Relativity. There are however extensions to the SM that incorporate the effects of gravity through its mediator particle called the *Graviton*. Additionally, the scalar spin-0 Higgs Boson allows fermions and the electroweak gauge Bosons to acquire mass without breaking the local gauge invariance of the electroweak Lagrangian. Fermions are split into two subgroups, *quarks* and *leptons* and are grouped into 3 generations of increasing mass. Fermions also have anti-particle counterparts, each anti-particle has the opposite quantum numbers (like electric charge) of the corresponding particle. Fermions interact through the weak force as they carry weak charge, but only

electrically charged fermions can also interact with the electromagnetic force. Quarks also interact with the strong force as they carry an additional kind of charge known as the *colour* charge, which comes in three types, *red, green or blue*. The SM was constructed by local gauge symmetries using groups to generate all of the interactions and observations. The simplest group that the SM is made up of is shown in Equation 2.1.

$$SU(3)_C \times SU(2)_L \times U(1)_Y \tag{2.1}$$

Breaking this down, the $SU(3)_C$ group describes the interactions between quarks and gluons with the conservation of colour charge (C) being the objective. This interaction is described by the theory of Quantum Chromodynamics (QCD) and is described by eight gluon fields A^g_μ, where $g \in [1 \dots 8]$. Like the weak interaction, the non-abelian nature of the symmetry group of the strong force enables gluons to self-interact. This self-interaction, along with the conversation of colour charge, has important physical consequences. The first consequence is that quarks are always found in colourless states, because the energy to separate them linearly increases with separation distance and eventually it becomes favourable to create another colour-singlet state. This process is called *colour-confinement*, which is prevalent at low-energies or large distances. Conversely, the phenomenon of *Asymptotic freedom* occurs at high energies and this allows quarks and gluons to behave as free particles which will eventually get detected as jets of hadrons by the detectors. The remaining group $SU(2)_L \times U(1)_Y$ is responsible for the unified electroweak force, where the weak hypercharge Y is conserved and weak isospin only acts on left handed fields L. This will be described in more detail in the next section.

2.2. Electroweak unification

The observation of *charged currents* can only occur if the $W^\pm$ carry an electric charge, and if they do, then they must also interact electromagnetically via γ. Moreover, prior to unification, the theoretical cross-section of $e^- e^+ \to W^+ W^-$ diverges at high energy hinting that there is a real need to modify the weak theory. To fix this the solution proposed by Glashow-Weinberg-Salam (GSW) was to unify the weak and electromagnetic forces into the *electroweak force*. To unify the forces a left handed weak isospin gauge symmetry is imposed on the $SU(2)_L$ group that operates on the weak

isospin, I. Similarly, the $U(1)_Y$ symmetry group is constrained to operate on the weak hypercharge $Y = 2(Q - I_3)$, where I_3 is the third component of weak isospin. These modifications mean that the $SU(2)$ gauge symmetry is only imposed on *left-handed* doublets, which are defined in Equation 2.2. The *right-handed* fields are still present in the theory, but they are instead kept as *singlets* under the $SU(2)$ transformation. This creates a left-right asymmetry in the weak force.

$$\Psi_L = \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix} \text{ and } \Psi_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \tag{2.2}$$

The requirement that the electroweak Lagrangian is invariant under the local gauge transformations of $SU(2)_L \times U(1)_Y$ produces four, albeit massless, gauge Bosons W^1_μ, W^2_μ, W^3_μ and B_μ. The masslessness of these gauge Bosons is fixed by the Higgs mechanism which will be discussed later. The physical mass states of the electroweak gauge Bosons are obtained through a linear combination of the gauge fields. The exact combinations are shown in Equations 2.3 and 2.4

$$W^\pm_\mu = \frac{1}{\sqrt{2}} \left(W^1_\mu \pm W^2_\mu \right) \tag{2.3}$$

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} cos\theta_W & sin\theta_W \\ -sin\theta_W & cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix} \tag{2.4}$$

Through this electroweak unification, key issues were solved, for example the existence of the neutral Z^0 gauge Boson now allows the cross-section of the $e^-e^+ \rightarrow W^+W^-$ process to agree with experimental observations. The discovery of the neutral current process was evidence for the electroweak unification theory as this can only be explained if there is a Z^0 present. Figure 2.2 shows the cross-section of the $e^-e^+ \rightarrow W^+W^-$ process as a function of the centre of mass energy. The cross-section increases towards infinity as a function of the centre of mass energy when there is no Z exchange, violating unitarity. The introduction of the Z exchange destructively interferes and maintains a well behaved cross-section across the centre of mass spectrum.

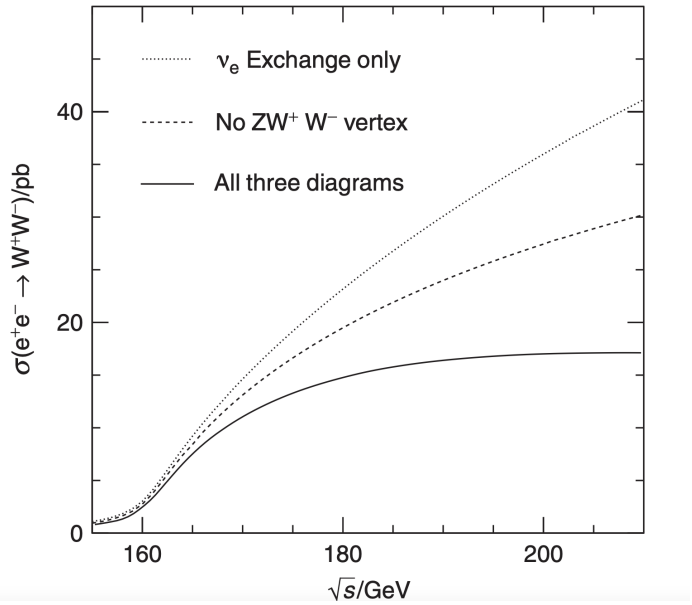


Figure 2.2.: Prediction of the $e^+e^- \to W^+W^-$ production cross-section for three different cases. The cross-section increases towards infinity for both cases that do not include a Z interaction [2].

2.3. The Higgs Mechanism

In the early 1960's the understanding of particle physics was essentially incomplete as the model was unable to correctly predict experimental observations. The source of the incompleteness comes from the stringent requirements of a locally gauge invariant Lagrangian, which allows us to keep the theory renormalizable and predictable. More formally, local gauge invariance on the $SU(2)_L \times U(1)_Y$, prohibits massive fermions and gauge Bosons. The fact that the weak interaction is only effective at a very short range indicates that it must be mediated by a massive particle, unlike the electromagnetic interaction which is effective at very long ranges due to its mediator, the photon, being massless. The Higgs mechanism, fortunately, fixes these problems by introducing a spin-0 scalar particle brought on by the Spontaneous Symmetry Breaking (SSB) of a gauge field. This section, will introduce this mechanism which will provide the theoretical grounding for Higgs pair production.

2.3.1. Broken symmetries

If we were to break a continuous global symmetry, we would introduce massless particles commonly known as *Goldstone* Bosons. However, when a *local* symmetry is spontaneously broken these massless particles disappear and the gauge Bosons are able to gain mass with the introduction of a massive scalar Boson. This massive scalar Boson is the elusive Higgs Boson. To demonstrate the details of this mechanism the local symmetry of the $SU(2)_L \times U(1)_Y$ needs to be spontaneously broken. To do this the covariant derivative of the symmetry group, shown in Equation 2.5, is used. The quantities $\boldsymbol{\tau}$ and Y are the generators of the $SU(2)_L$ and $U(1)_Y$ groups respectively. The generator $\boldsymbol{\tau}$ are the 2×2 Pauli spin matrices and is related to the isospin $\boldsymbol{T}$ by the relationship $\boldsymbol{T} = \boldsymbol{\tau}/2$. The electroweak coupling constants are represented by g and g'.

$$D_\mu^{EW} = \partial_\mu + \frac{i}{2} g \boldsymbol{\tau} \cdot W_\mu + i g' \frac{Y}{2} B_\mu \tag{2.5}$$

It is also important to introduce the $SU(2)_L$ complex scalar doublet of the Higgs field, with hypercharge $Y = 1/2$, as shown by Equation 2.6 below.

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.6}$$

The doublet is composed of the ϕ^+ and ϕ^0 scalar fields which represent a positively charged particle and a neutral particle respectively. They are made up of the real scalar fields ϕ_i where $i \in [1, 2, 3, 4]$ as shown in Equations 2.7 and 2.8

$$\phi^+ = \frac{\phi_1 + i\phi_2}{\sqrt{2}} \tag{2.7}$$

$$\phi^0 = \frac{\phi_3 + i\phi_4}{\sqrt{2}} \tag{2.8}$$

Now the following Lagrangian of the scalar field, $\boldsymbol{\Phi}$, can be considered as a description of the electroweak sector.

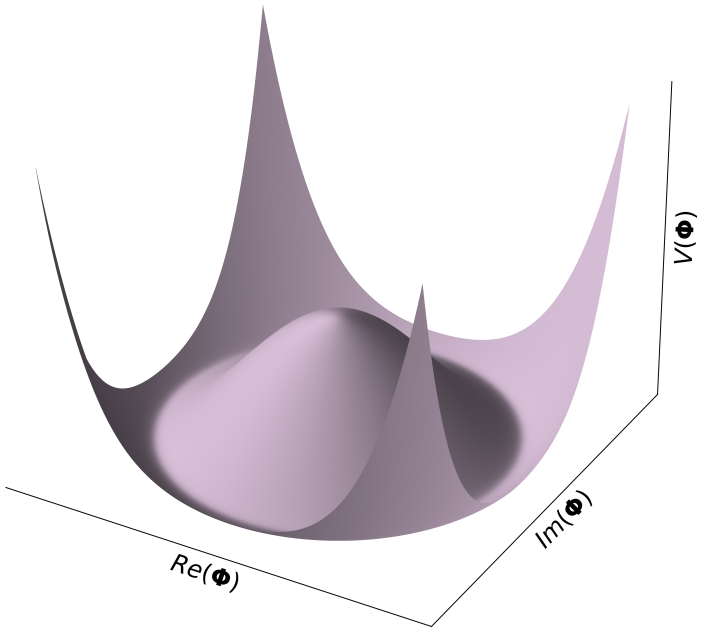


Figure 2.3.: Example of the Higgs potential, $V(\mathbf{\Phi})$, with $\mu^2 < 0$ and $\lambda > 0$. The minimum occurs at $-\mu^2/2\lambda$.

$$\mathcal{L}_{\mathbf{\Phi}} = (D_\mu \mathbf{\Phi})^\dagger D^\mu \mathbf{\Phi} - V(\mathbf{\Phi}) \tag{2.9}$$

To keep the theory invariant and renormalizable, for the $SU(2)_L \times U(1)_Y$ group, the scalar potential $V(\mathbf{\Phi})$ is shown in Equation 2.10.

$$V(\mathbf{\Phi}) = \mu^2(\mathbf{\Phi}^\dagger \mathbf{\Phi}) - \lambda(\mathbf{\Phi}\mathbf{\Phi})^2 \tag{2.10}$$

The symmetry can be spontaneously broken by constraining both $\mu^2 < 0$ and $\lambda > 0$. This constraint causes the potential to yield an infinite amount of non-zero minima, leading to the *Mexican hat* potential shape. An example of the Higgs potential is shown in Figure 2.3. These degenerate minima occur at value of $\mathbf{\Phi}^\dagger \mathbf{\Phi}|_{min} = -\mu^2/2\lambda$. This minima can then be expressed in terms of the real fields ϕ_i as shown in Equation 2.11.

$$\frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda} \tag{2.11}$$

The minima are chosen so that the Vacuum Expectation Value (VEV) of ϕ_1, ϕ_2 and ϕ_4 is zero, while $\langle 0|\phi_3|0\rangle = v^2 = -\mu^2/\lambda$. The choice of this particular particular minima is arbitrary, but with hindsight, this keeps the photon massless.

The doublet field $\mathbf{\Phi}$ can be expanded around this vacuum by setting $\phi_3 = H + v$, where H is a neutral scalar Higgs field.

$$\mathbf{\Phi} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ H + v \end{pmatrix} \tag{2.12}$$

The Goldstone modes with zero VEV can disappear by using the unitary gauge $\boldsymbol{\theta}(x)/v$ and the local gauge transformation $e^{i\boldsymbol{\tau}\cdot\boldsymbol{\theta}(x)/v}$. Through these transformations the terms of the Lagrangian involving the scalar field H can be written in terms of the mass eigenstates of the $W^\pm$ and Z Bosons. This is shown in Equation 2.13. From this Lagrangian several consequences of the Higgs mechanism arise, such as the mass terms for the $W^\pm$, Z and Higgs Bosons and can be summarized in Equations 2.14. This Lagrangian also indicates that there are 2 types of interactions between the vector Bosons (V) and the Higgs, the VVH and the $VVHH$ interactions. Moreover, there are also two self-interaction terms of the Higgs field, the triple HHH and quartic $HHHH$ interactions. These interactions are summarized in Figure 2.4. Now that these intricate details of the Higgs mechanism are laid out it is clear that a measurement of the pair production of HH enables a direct probe in the electroweak sector of the Standard Model.

$$\begin{aligned} \mathcal{L}_{\mathbf{\Phi}} = & \frac{1}{2}\partial_\mu H \partial^\mu H \\ & + \frac{1}{4}g^2(H^2 + 2vH + v^2)W_\mu^+ W^{-\mu} \\ & + \frac{1}{8}(g^2 + g'^2)(H^2 + 2vH + v^2)Z_\mu Z^\mu \\ & + \mu^2 H^2 + \frac{\lambda}{4}(H^4 + 4vH^3) \end{aligned} \tag{2.13}$$

$$M_{W^{\pm}} = \frac{1}{2}gv \tag{2.14}$$

$$M_{Z^0} = \frac{1}{2}\frac{gv}{cos\theta_w} \tag{2.15}$$

$$M_H = \sqrt{-2\mu^2} \tag{2.16}$$

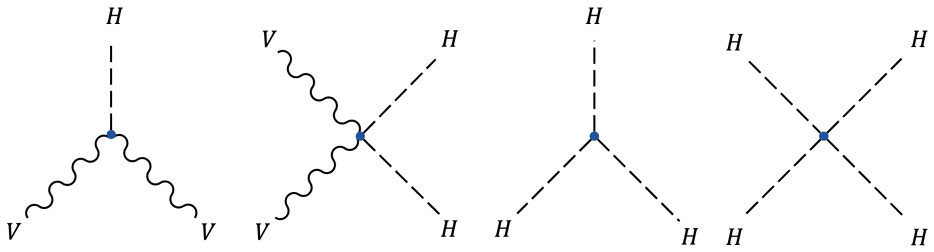


Figure 2.4.: Higgs self and vector Boson interactions as described by the Lagrangian shown in Equation 2.13.

2.4. Shortcomings of the Standard Model

Ever since 1960's there has been a plethora of experimental evidence in support of the Standard Model and in 2012 the measurement of a particle consistent with Standard Model Higgs was a particularly huge milestone for the theory [3]. There is however known physics that are not described by the theory. The first problem with the SM and perhaps the most glaring one is that gravity is completely omitted from the theory. This means that the theory is missing out on the fourth fundamental force and we cannot unify cosmological measurements with the theory. In the recent history of particle physics neutrinos have been shown to have mass through the neutrino oscillation experiments [4], [5]. This is yet another sector of the SM that needs to be modified. Cosmological measurements, quite shockingly, indicate that 95 % of the matter in the universe is unaccounted for and is made up of a combination of *dark matter* and *dark energy* [6]. There are no candidate particles in the Standard Model that are compatible with these observations and so currently the SM can only describe a disappointing 5% of our universe. One of the most popular SM extensions is to include a Dark Matter candidate particle called the Weakly Interacting Massive Particle (WIMP) that would tie together all of the cosmological observations [7]. Furthermore, there is contention on the requirement of

lepton flavour universality. The source of contention comes from the recent measurement of the decay rates of $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to K^0\mu^+\mu^-$, which have a deviation from the SM prediction at a significance of 3.1 σ [8]. Lastly, another observation in contention with the known SM physics is the $g-2$ measurement of the anomalous magnetic moment of a muon, which stands at a significance of 4.2σ from the SM prediction [9]. These observations are evidence of the fact that the SM may need to be modified.

One noteworthy BSM extension that includes the effects of gravity is the bulk Randall-Sundrum (RS) model. In this model, the existence of a spin-2 Kaluza-Klein graviton G^*_{KK} particle is predicted as a consequence of the warped extra dimensions. This model is parametrized by the value of c which is determined by the ratio $k/\bar{M}_{Pl}$, where k quantifies the curvature scale of the extra dimensions and $\bar{M}_{Pl}$ is the reduced Planck mass [10], [11]. Another particularly interesting BSM extension is the Two-Higgs-Doublet-Model (2HDM), where an additional Higgs doublet is introduced [12]. Having two complex doublets leads to a total of 8 degrees of freedom, 3 of which are eaten by the electroweak Bosons, leaving 5 degrees of freedom for physical fields. These additional 5 scalar particles, can help to bridge several of the previously mentioned gaps in the current understanding of the SM. For example this can provide a mechanism for neutrinos to acquire mass. It can also add a mechanism for CP violation in the Higgs sector, which can help to solve the problem of baryogenesis [13]. Moreover, it can provide a natural candidate for dark matter particles [14]. There are plenty of BSM extensions available, however these two mentioned here are not only able to solve quite a few of the current problems of the SM but they are also of relevance to the HH searches that will be discussed next.

2.5. Phenomenology of HH production

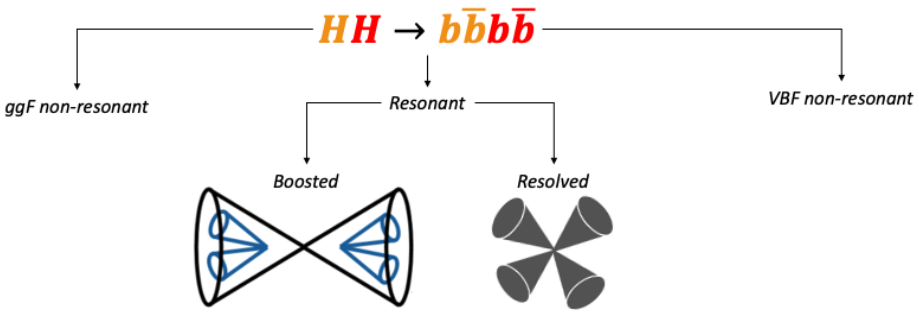


Figure 2.5.: Structure of the $HH \to b\bar{b}b\bar{b}$ searches. This thesis focuses on the resolved resonant search and the ggF non-resonant search.

The main focus of this thesis is on HH searches by using the $b\bar{b}b\bar{b}$ final state. In this section some of the main concepts of this search will be described. Figure 2.5 identifies the main categories of the HH search. Principally, the search is split into two categories: *resonant* and *non-resonant* searches. The difference between these two categories is the propagator particle that generated the HH. In *resonant* searches a heavy physical resonance is produced that then decays into a pair of Higgs Bosons, thus the search is dependent on the mass of the heavy resonance. The *non-resonant* search has no physical resonance and the HH production is instead constrained by the exchange of a *virtual* or *off-shell* particle. In this thesis only the gluon-gluon fusion (ggF) production mode will be considered. It is possible that the HH production occurs via Vector Boson Fusion (VBF), however the cross-section of this mode is much smaller than the ggF. Figure 2.6, shows the branching ratios for the main HH decay channels. The $b\bar{b}b\bar{b}$ final state has the largest branching ratio at roughly 34%, which provides a strong motivation for a search using this channel. Ultimately, both the ATLAS and Compact Muon Solenoid (CMS) collaborations will combine HH searches from various final states for a combined limit, which will be discussed more in more detail in section 2.6

2.5.1. Resonant HH production

A pair of Higgs Bosons can be produced through BSM processes, provided that the mass of the particle is $m_{\mathcal{X}} > 2m_H$. This leads to the resonant production mode through ggF, which can be probed by the $HH \to b\bar{b}b\bar{b}$ search. In this case two benchmark signal models are used. The bulk RS model allows for the existence of a spin-2 Kaluza-Klein (KK) graviton, G^*_{KK}. Similarly a narrow-width, spin-0, scalar resonance X, is used as a BSM benchmark model. Either of these resonances may decay into a pair of Higgs Bosons, which will subsequently decay into $b\bar{b}b\bar{b}$. As an example the spin-0 scalar Boson, X, is a heavy Higgs Boson, predicted by all two-Higgs-doublet models (2HDM).

2.5.2. Non-resonant HH production

After the spontaneous symmetry breaking of the electroweak Lagrangian, the Higgs potential, $V(H)$, yields a mass term and the triple and quartic self interactions as shown in Equations 2.17 and 2.18.

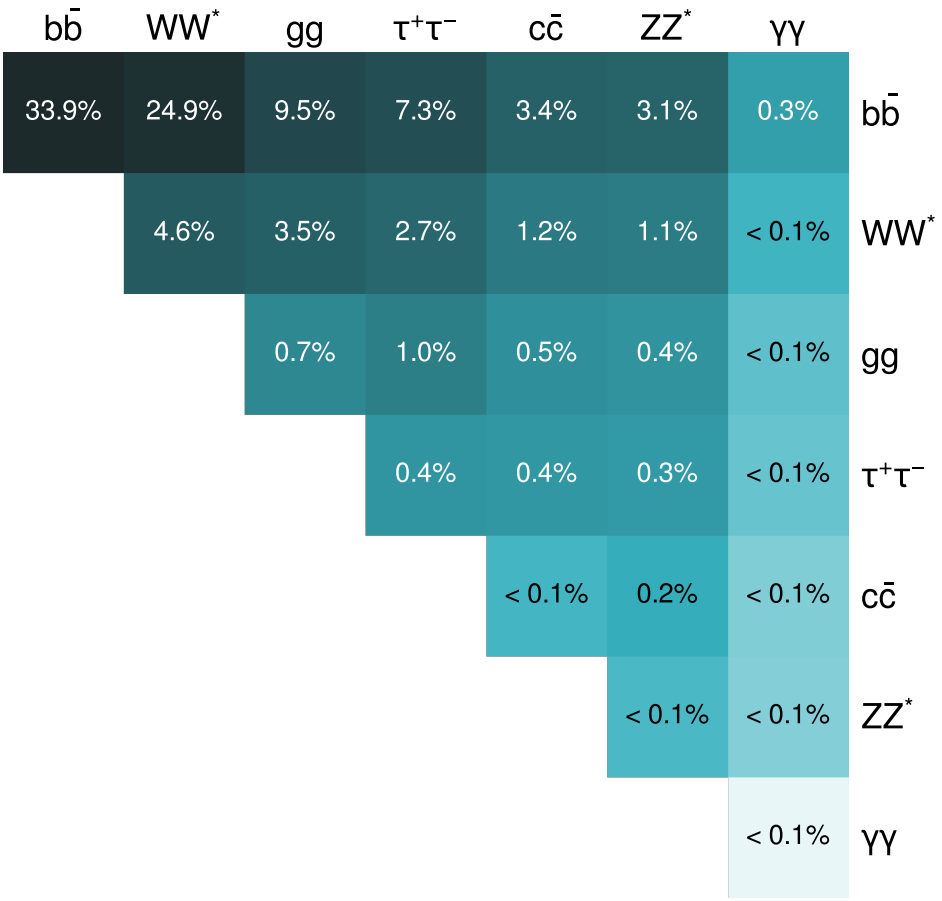


Figure 2.6.: Most common branching ratios of HH. The most common decay is $HH \to b\bar{b}b\bar{b}$ at roughly 34 % of the HH decays. This decay mode will be the focus of this thesis [15].

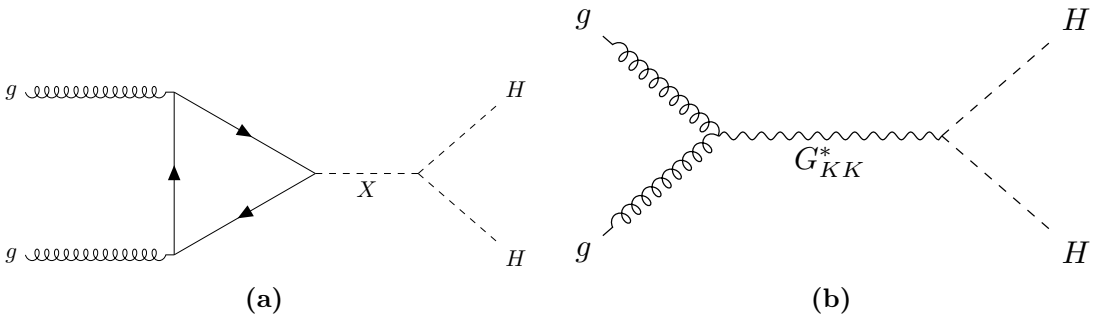


Figure 2.7.: Feynman diagrams for resonant HH production via gluon-gluon fusion. The production is mediated through two different resonances: a) Spin-0 scalar X resonance. b) Spin-2 Kaluza-Klein graviton G^*_{KK} resonance [16].

$$V(H) = -\lambda v^2 H^2 - \lambda v H^3 - \frac{1}{4}\lambda H^4 \tag{2.17}$$

$$V(H) = -\frac{1}{2}m_H^2 - \lambda_{HHH}^{SM} v H^3 - \lambda_{HHHH}^{SM} H^4 \tag{2.18}$$

The cross-section of the triple interaction is determined by the *trilinear* coupling, λ_{HHH}^{SM}. The coupling depends on the VEV, v, and the mass of the Higgs, m_H. The VEV has been measured experimentally by muon decay at $v = 246$ GeV, and the mass of the Higgs has been measured at $m_H = 125$ GeV. This means that the trilinear coupling and hence the cross-section of the non-resonant production can be accurately predicted by the SM, as shown by Equation 2.19. Therefore if an experimental measurement shows evidence of deviations from this value, then this is a sign of new physics and would indicate that the electroweak sector requires modification.

$$\lambda_{HHH}^{SM} = \frac{m_H^2}{2v^2} \approx 0.13 \tag{2.19}$$

For convenience the trilinear coupling is re-parametrized as κ_λ, which is defined in Equation 2.20, where λ_{HHH} is the coupling variation that may include BSM processes. The results of the search are often reported as a function of the κ_λ and κ_t variables, where a value of unity corresponds to the SM prediction.

$$\kappa_\lambda = \frac{\lambda_{HHH}}{\lambda_{HHH}^{SM}} \tag{2.20}$$

The cross-section for this process with $\kappa_\lambda = \kappa_t = 1$ at $\sqrt{s} = 13$ TeV is 31.05 fb. This is calculated at NNLO[1] precision with the finite top mass approximation [17]. Figure 2.8 shows the dominant diagrams that contribute to the non-resonant ggF production at leading order. In Figure 2.8b, there are four top quark propagators and the pair of Higgs Bosons are produced from two vertices of the box diagram. This results in the

[1]Next-to-next-to-leading order

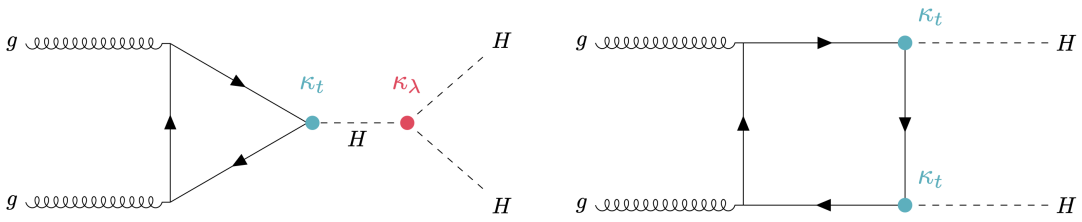


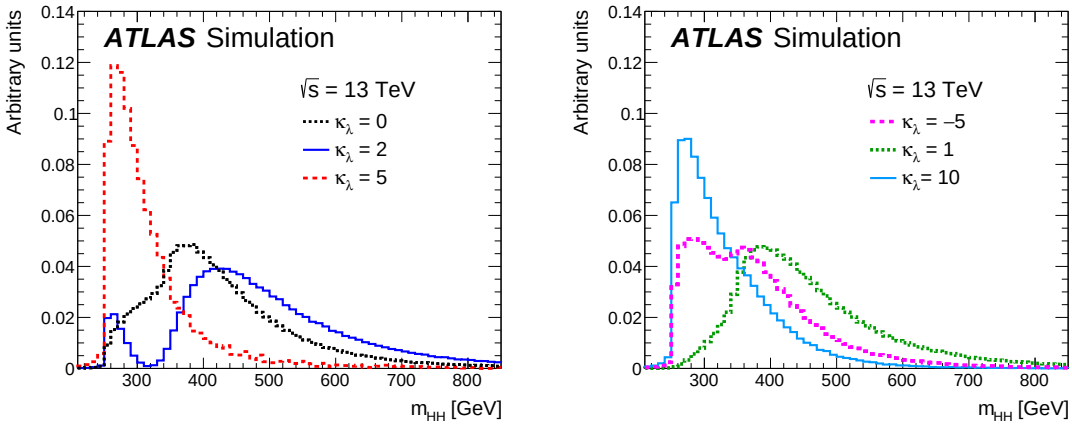
Figure 2.8.: Feynman diagrams for non-resonant gluon-gluon fusion production of HH. κ_t represents the ratio of the Higgs coupling to the top quarks relative to the SM value. Similarly, κ_λ is the ratio of the Higgs self-coupling relative to the SM value.

amplitude of this diagram to being dependent on the top coupling to the Higgs, k_t and hence it is constrained by the $t\bar{t}H$ production. The alternative leading order diagram is shown in Figure 2.8a, where the pair of Higgs Bosons arise from a far off-shell SM Higgs. The amplitude of this diagram depends on the triple Higgs vertex, or κ_λ variation. Unfortunately, these two diagrams destructively interfere, which is the reason why the SM cross section is minuscule and is the main challenge behind HH searches. Figure 2.9, shows how the distribution of the invariant mass of the HH system, m_{HH}, changes with different values of the κ_λ parameter. The κ_λ parameter controls how the two diagrams interfere with each other. Extreme values away from the $\kappa_\lambda = 1$ point create a very soft m_{HH} spectrum, for example at $\kappa_\lambda = 5$ or $\kappa_\lambda = 10$. The triangle diagram is dominant at low m_{HH}, while the box diagram is dominant at high m_{HH}.

Figure 2.10 shows HH ggF cross section $\sigma(gg \rightarrow HH)$ as a function of the κ_λ parameter. Varying the κ_λ parameter alters the interference between the box and triangle diagrams and the cross-section can be significantly altered. The cross-section for the SM $\kappa_\lambda = 1$ point is relatively low at $\sigma^{SM}_{ggF} = 31.05$ [fb], which was calculated at NNLO FTApprox [17, 19].

2.6. Recent efforts in *HH* searches

Given the strong theoretical motivations for HH searches there has already been several efforts to explore this production mode. Initially, both ATLAS and CMS used the data available from 2015 and 2016 to produce the *Early Run-2* results. The integrated luminosity of these searches was in the range of 27.5–36.1 fb^{-1} . Several final states have been combined using this limited amount of Run-2 data. In ATLAS the most sensitive



(a) (b)

Figure 2.9.: m_{HH} distributions for various κ_λ values obtained at the Monte Carlo generator level. The shape of these distributions is controlled by how the box and triangle diagrams interfere with each other. One particular point of interest is the $\kappa_\lambda = 0$, which essentially turns off the contribution of the triangle diagram. Conversely, the $\kappa_\lambda = 5$ distribution is dominated by the triangle diagram which results in a strong low m_{HH} peak, which also occurs to a lesser extent at a value of $\kappa_\lambda = 10$. The distribution at $\kappa_\lambda = 2$ results in a double peaked structure, which is near the point of maximum destructive interference [18].

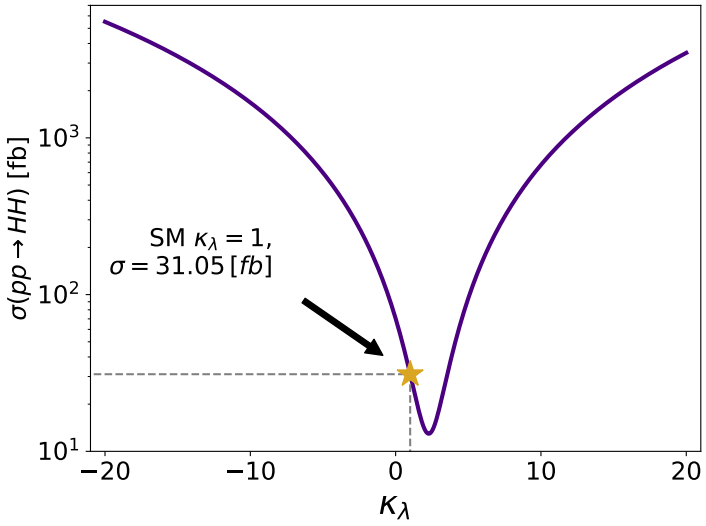
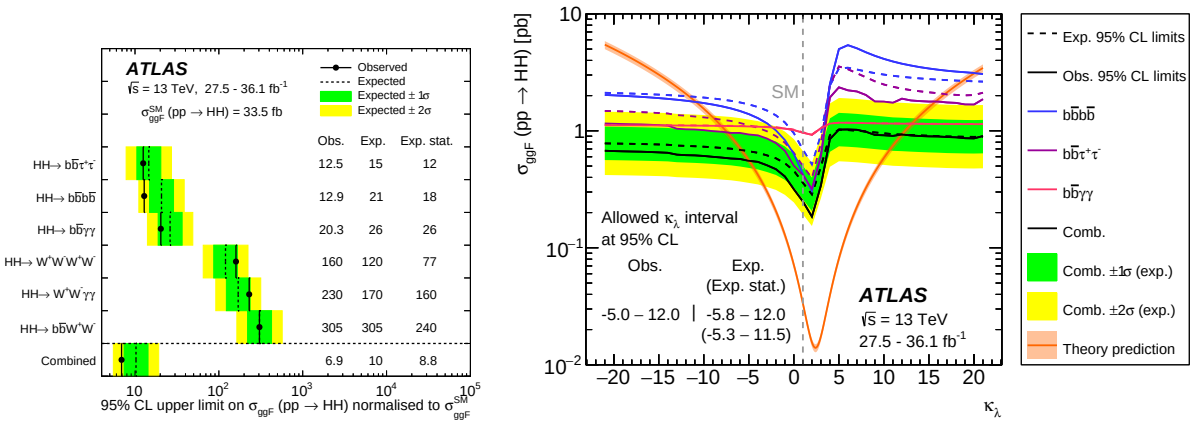


Figure 2.10.: HH ggF cross section $\sigma(gg \to HH)$ as a function of the κ_λ parameter. Varying the κ_λ parameter alters the interference between the box and triangle diagrams and the cross-section can be signficantly altered. The cross-section for the SM $\kappa_\lambda = 1$ point is indicated by the star on the Figure and has a value of $\sigma^{SM}_{ggF} = 31.05$ [fb] [19].

channels have been $b\bar{b}\tau^+\tau^-$ [20], $b\bar{b}b\bar{b}$ [21], $b\bar{b}\gamma\gamma$ [22]. Other final states, $W^+W^-W^+W^-$ [23], $W^+W^-\gamma\gamma$ [24] and $b\bar{b}W^+W^-$ [25] have also been explored. These searches have been combined by the ATLAS collaboration to produce early Run-2 limits [18]. The CMS collaboration has also searched for HH production in the same final states with the addition of $b\bar{b}ZZ$ [26]. More recently results using the full Run-2 dataset have been published. In ATLAS these were the boosted $b\bar{b}\tau^+\tau^-$ [27], VBF $b\bar{b}b\bar{b}$ [28], $b\bar{b}\ell\nu\ell\nu$ [29] and $b\bar{b}\gamma\gamma$ [30], while the CMS collaboration has already published their $b\bar{b}b\bar{b}$ results using the full Run-2 dataset [31].

To serve as a benchmark comparison for the results of the $HH \to b\bar{b}b\bar{b}$ search presented in this thesis, some of the early Run-2 combined results from ATLAS are presented in this section. Figure 2.11a shows the early Run-2 expected and observed upper limits on the SM HH cross section via ggF production (σ_{ggF}^{SM}), normalized to the SM cross section. The $b\bar{b}b\bar{b}$ channel yields an observed upper limit of 12.9 and an expected limit of 21. The combination of all the channels in ATLAS yields an observed limit of 6.9 and an expected limit of 10. Similarly, Figure 2.11b shows the upper limit on the cross-section as a function of the κ_λ coupling parameter. Table 2.1 summarizes the constraints on κ_λ that each channel provides to the combination. The combination constrains the value of κ_λ to be within $\kappa_\lambda \in [-5, 12.0]$. The $b\bar{b}b\bar{b}$ channel has the least constraints on the κ_λ parameter, at $\kappa_\lambda^{b\bar{b}b\bar{b}} \in [-10.9, 20.1]$. This is due to the softer m_{HH} spectrum at in the extreme κ_λ values, as seen in Figure 2.9, where the $b\bar{b}b\bar{b}$ search lacks sensitivity. The results of the resonant early Run-2 searches are shown in Figure 2.12 for the three bench mark signals: X, G^*_{KK} with $c = 1$ and $c = 2$. The results of the $b\bar{b}b\bar{b}$ channel combines both the resolved and boosted topologies to obtain these limits and in general this channel provides the strongest constraint above resonance masses of 500 GeV. In the $b\bar{b}b\bar{b}$ channel an excess is observed in the spin-0 limits at 280 GeV. The excess has a global (local) significance of 2.5 (3.6). The $b\bar{b}b\bar{b}$ channel also excludes spin-2 $c = 1$ gravitions within a mass range of 313–1362 GeV, while the $c = 2$ models have been excluded below 1744 GeV.



(a) (b)

Figure 2.11.: Upper limits at 95 % CL on the cross-section of HH ggF production obtained by the ATLAS collaboration from the early Run-2 search. a) Shows the limits obtained for each channel for the SM prediction normalized to the expected SM cross-section ($\sigma^{SM}_{ggF}(pp \to HH)$. For the SM value the combination of all the channels can exclude cross-sections higher than 6.9 times the SM value of 33.5 fb. b) Shows the upper limit on the cross-section obtained for each channel as a function of the κ_λ coupling parameter. Although the $b\bar{b}b\bar{b}$ channel is sensitive to the SM point relative to the other channels, it provides the least constraint to the κ_λ coupling [18].

Final State	**Allowed κ_λ interval at 95 % CL**		
	Observed	**Expected**	**Expected (stats only)**
$b\bar{b}b\bar{b}$	[-10.9, 20.1]	[-11.6, 18.8]	[-9.8, 16.3]
$b\bar{b}\tau^+\tau^-$	[-7.4, 15.7]	[-8.9, 16.8]	[-7.8, 15.5]
$b\bar{b}\gamma\gamma$	[-8.1, 13.1]	[-8.1, 13.1]	[-7.9, 12.9]
combined	[-5.0, 12.0]	[-5.8, 12.0]	[-5.3, 11.5]

Table 2.1.: Early Run-2 constraints on κ_λ obtained by the ATLAS collaboration by combining the three most sensitive final states: $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$. The $b\bar{b}b\bar{b}$ channel provides the least constraints to the κ_λ coupling parameter. This is due to the fact that the extreme κ_λ values yield a softer signal m_{HH} distributions, which is where the background of the $b\bar{b}b\bar{b}$ channel peaks. [18]

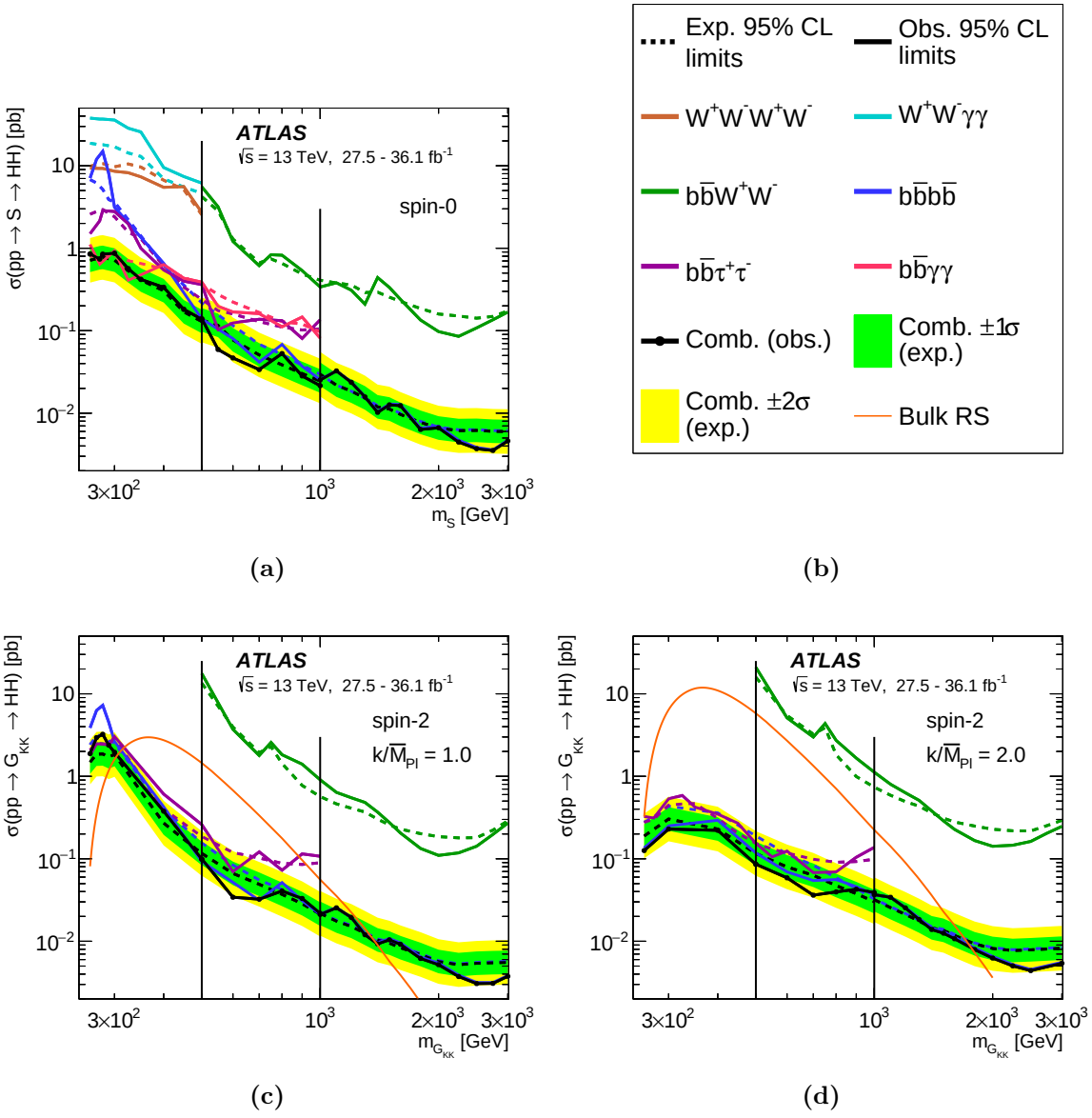


Figure 2.12.: Early Run-2 upper limits at 95 % CL on the cross-section of the resonant HH production as a function of resonance mass obtained by the ATLAS collaboration. Three different benchmark signals were tested in this search. a) Spin-0 narrow width heavy scalar particle, c) spin-2 Kaluza-Klein graviton with $k/\overline{M}_{Pl} = 1$ and d) spin-2 Kaluza-Klein graviton with $k/\overline{M}_{Pl} = 2$. Only $b\bar{b}b\bar{b}$, $b\bar{b}W^+W^-$ and $b\bar{b}\tau^+\tau^-$ have been combined to produce the spin-2 limits. The limits obtained for the $b\bar{b}b\bar{b}$ search are the result of combining both the resolved and boosted event topologies [18].

Assuming that the SM is correct, a single HH event is about 1500 times rarer than a single Higgs event [32], and after the detector inefficiencies are applied only about 100 HH events are expected in the full Run-2 dataset. Even with the combined ATLAS and CMS searches, this is still not likely to be significant enough to reach a 3 σ level required for evidence. It is therefore important to understand how the search scales up with the full luminosity of the HL-LHC, 3000 fb^{-1}. In this dataset the combined significance of all the different channels in both ATLAS and CMS will reach a level of 4.0 σ when systematic uncertainties are included [33]. This is still far off the 5σ level required for a discovery. Moreover, Figure 2.13, shows the combined constraint on the κ_λ parameter which can get down to a range of $\kappa_\lambda \in [0.1, 2.3]$. Although both of these projections are likely to be on the conservative end of the spectrum, it clearly outlines that our understanding of the HH search needs to improve to fully exploit the potential of the HL-LHC.

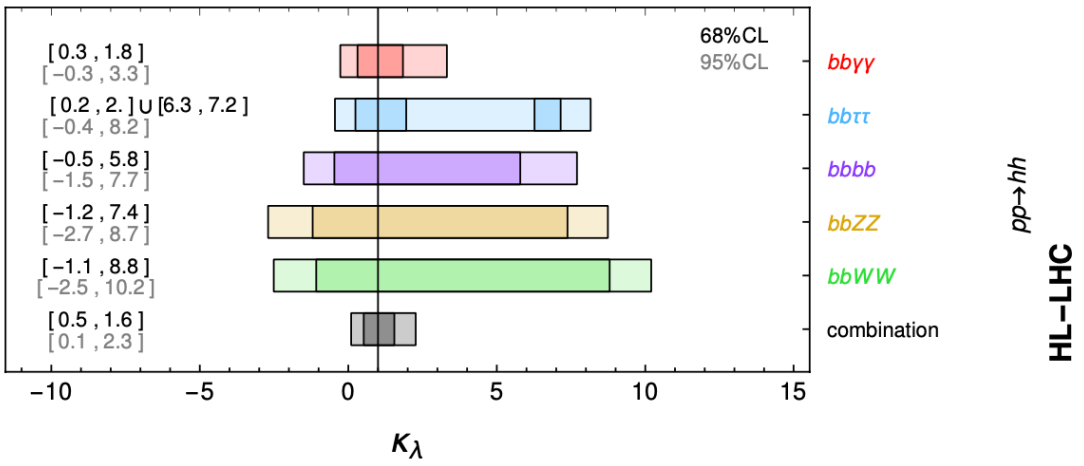


Figure 2.13.: Combined ATLAS and CMS High Luminosity LHC (HL-LHC) projections on the Higgs self-coupling κ_λ assuming a luminosity of 3000 fb^{-1} at $\sqrt{s} = 14$ TeV. The projections are made using the early Run-2 searches as a baseline. The projections indicate that the κ_λ can be constrained to values between $\kappa_\lambda \in [0.1, 2.3]$ at 95% CL. Constraining the κ_λ parameter down to these values can help exclude a number of BSM models. Figure edited from [33].

Chapter 3.

Machine Learning Methods

Since the $HH \to b\bar{b}b\bar{b}$ analysis makes extensive use of multivariate techniques, some of the key concepts are outlined in this section. All of the models used fall under the fully supervised category, which is characterized by having a dataset $\boldsymbol{x}$ that maps onto an output $\boldsymbol{y}$. This mapping is done by minimizing the estimator function θ to describe the probability density $p(\boldsymbol{y}|\boldsymbol{x}, \theta)$. Supervised learning models can be further sub-categorized into classification or regression models depending on the nature of the output target vector $\boldsymbol{y}$. If $\boldsymbol{y}$ can only take categorical or discrete values, then the model is a classification model. An example classifier model can be used to classify events into having originated from Higgs Bosons from a $\tau\tau$ decay (signal) and a generic QCD background, which can improve the sensitivity of the search. If $\boldsymbol{y}$ is able to take continuous numerical values, then the model is considered to be a regression model. An example use case of a regression model has been applied to accurately calibrate the energy of b-jets by using information of the jet composition as an input feature. This increased accuracy and energy resolution enhances the sensitivity of analyses that depend on b-jets such as $HH \to b\bar{b}$ and $HH \to b\bar{b}b\bar{b}$ [34].

The models used in the $HH \to b\bar{b}b\bar{b}$ search are parametrized as Boosted Decision Tree (BDT) and a Neural Network (NN), which will be described in the next sections. One prominent issue with all supervised learning models is that of *under-fitting* or *over-fitting*. Under-fitting occurs when the model is too simple for the given task. In the opposite case over-fitting is when the model's complexity enables it to perform remarkably well in the given dataset, but is unable to generalize to unseen data. These two issues are the primary reason why the total dataset is split into at least two subsets used for training and testing. The ratio of training and testing is up to the user to decide. The model is not exposed to the testing set and that is used to evaluate the model's performance,

giving a good approximation for how well the model can generalize to unseen datasets. An example of overfitting is illustrated in Figure 3.1, which shows the error of a linear regression model with increasing complexity (the degree of the polynomial). The training error in this case continues to decrease with increasing complexity, but the error of the testing set starts to increase after a degree of 4. This is a very clear indication that the model is overfitted at this point. Decision trees and neural networks will exhibit the exact same behaviour and this point of divergence will be used to halt the training procedure.

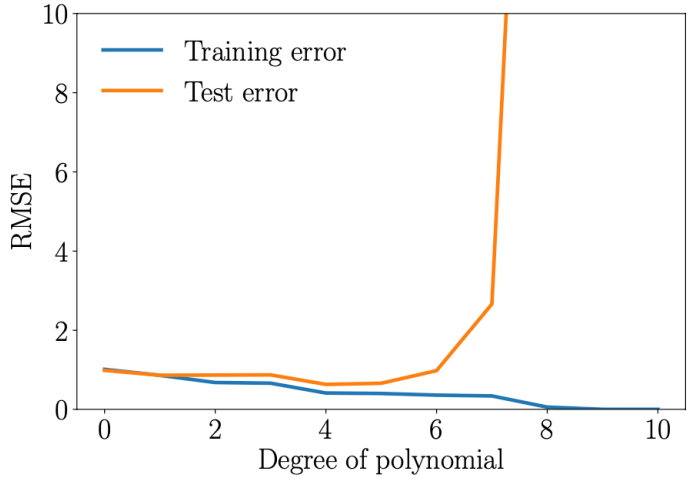


Figure 3.1.: Overfitting example using a linear regression model with varying polynomial degrees (i.e. the model's complexity). The y-axis shows the Root Mean Squared Error (RMSE), for both the training and testing datasets. When the testing and training error diverge, the model is considered to be overfitted [35].

3.1. Decision trees

Decision trees are one the simplest and most interpretable models that can be used for supervised learning tasks. The structure of the decision tree model is shown in Figure 3.2. In this case the tree is used to classify events into *signal* (S) or *background* (B). Each event has input features $\boldsymbol{x}$ and target label y so that, $(\boldsymbol{x}, y) = (x_1, x_2, \ldots, x_n, y)$. The target label y, will be predicted by the model. A decision tree is constructed by splitting the input space with arbitrary cuts. For example at the root node events are split into two nodes depending on whether they pass or fail events with a cut c_i on variable x_i. The tree can continue to split nodes by applying further cuts on the feature space $\boldsymbol{x}$. Ideally each cut enhances the signal to background separation of the decision tree model and the effectiveness of this depends mostly on the feature input space, $\boldsymbol{x}$.

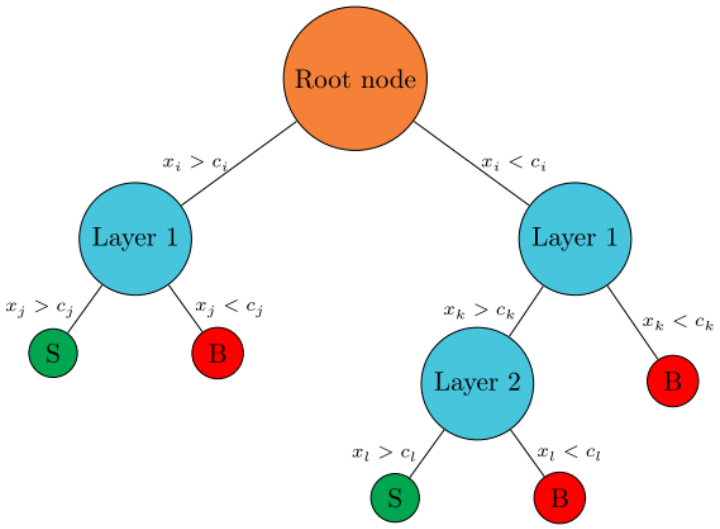


Figure 3.2.: Visualization of a decision tree with a depth of 3. The construction of the tree model begins at the *Root node*. The input features are represented with the variable x and the threshold cuts are represented with the variable c. The nodes denoted with "S" and "B" are for signal-like and background-like events respectively [36].

To construct an optimal tree the *Gini* purity index can be used as metric to determine the optimal cut. Equation 3.1, shows the definition of the Gini index, where p is the signal purity [37]. The signal purity is computed by $n_S/(n_B + n_S)$, where n_S and n_B are the number of signal and background events respectively. Events also carry an associated weight W_i, which is typically normalized so that the sum of event weights is unity.

$$Gini = \left(\Sigma_{i=1}^{N} W_i\right) p(1-p) \tag{3.1}$$

The Gini index tells us the purity at a specific node, but it can be used to determine whether a node should be split and how. To do this consider a parent node that may be split into two child nodes. In this case the separation gain G (sometimes also called the information gain) can be computed by Equation 3.2.

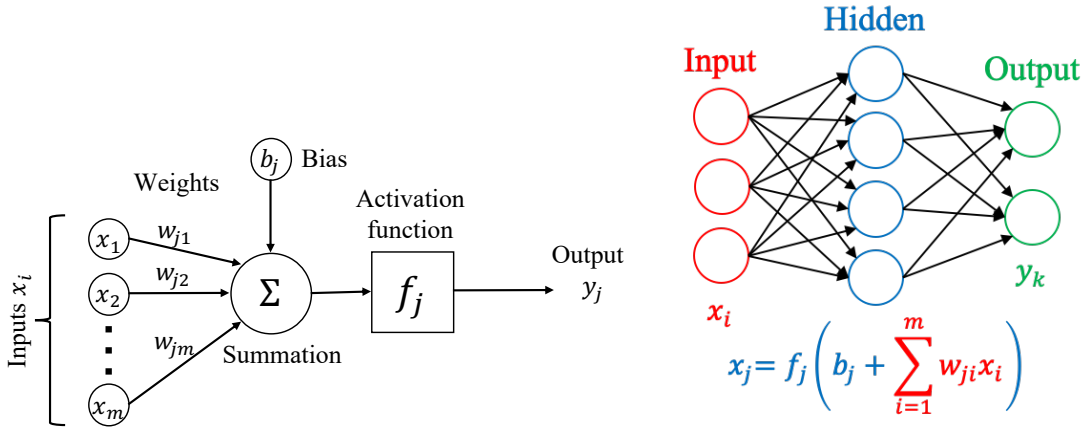
$$G = Gini_{parent} - (Gini^1_{child} + Gini^2_{child}) \tag{3.2}$$

The separation gain is computed by scanning over the input space $\boldsymbol{x}$ and the variable and cut with the largest gain is selected. The tree construction is terminated when any of the three stopping criteria are met:

1. Number of events in a node is below a threshold.
2. The maximum tree depth has been reached (hyperparameter set by the user).
3. The scan across all input features yields a negative separation gain, $G < 0$.

Decision trees in general perform relatively poorly compared to neural networks and would have become obsolete without *Boosting* [38].To improve the performance of decision trees the Boosting method creates an ensemble of *weak learners* to create a single *strong learner*. This is done by sequentially retraining the weak learner on weighted versions of the dataset, where the larger weights are given to the incorrectly classified events. Boosting can be done in many different ways and there is an abundance of intricate details, but it has been streamlined in three popular software packages that yield state of the art performance: XGBoost [39], LightGBM [40] and CatBoost [41].

3.2. Neural networks



(a) Perceptron Model **(b)** Single layer network

Figure 3.3.: a) Perceptron model of a neural network node b) Multi-Layer Perceptron model with a single hidden layer [42].

Neural networks are inspired from biological learning systems made up of complex webs of interconnected neurons. Neural networks can also be used for regression and

classification tasks, but the architecture and training procedures are significantly different from that of a decision tree. Figure 3.3a, shows the most basic unit of a neural network, commonly known as a *perceptron*. In a perceptron, all entries from the input vector $\boldsymbol{x}$ are mapped into the transfer function node, where a bias b_j can be added, through tuneable weights W_{ij}. This is then passed to an *activation function*, which introduces key non-linearities to the system[1].

The result then gets passed to the next layer, which could be the output layer. By combining several layers of perceptrons the neural network can be built up as is shown in Figure 3.3b, this is then known as a Multi-Layer Perceptron (MLP) model. The layers that connect the *input* and *output* layers are called *hidden* layers and typically if the network has more than 1 hidden layer it is called a *deep neural network*. The depth of the network (number of layers) and width of each layer are tunebale hyperparameters that need to be optimized for the task to solve. Equation 3.3 shows how the output of the first hidden layer x_j relates to the input x_i of a single layer neural network with weights w_{ji}, biases b_j and the activation function f_j, as shown in Figure 3.3b. The sum in Equation 3.3, is performed over the number of input features m. Similarly if the network has an additional layer then the output of the second hidden layer x_k is described by Equation 3.4, which encapsulates Equation 3.3 and adds an extra set of weights w_{kj}, biases b_k and activation function f_k. The sums in Equation 3.4 run over the number of input features m_1 and the number of neurons in the second layer m_2.

$$x_j = f_j \left(b_j + \sum_{i=1}^{m} w_{ji} x_i \right) \tag{3.3}$$

$$x_k = f_k \left(b_k + \sum_{j=1}^{m_2} w_{kj} f_j \left(b_j + \sum_{i=1}^{m_1} w_{ji} x_i \right) \right) \tag{3.4}$$

In a classification task, a sigmoid activation function, shown in Equation 3.5, is used in the output layer to provide a smoothly varying output between 0 and 1, which can be

[1] It is important to note that without activation functions a neural network boils down to linear regression.

interpreted as the probability to belong to a certain class.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.5}$$

The *tanh* and Rectified Linear Unit (ReLU) activation functions are usually used in hidden layers, with ReLU often preferred as it helps the training to converge faster. Example of these activation functions are shown in Figure 3.4. Moreover, the *softmax* activation function, defined in Equation 3.6, is also useful as it is a multidimensional sigmoid function. The softmax function maps any unbounded input vector $\vec{z}$ into a probability space, such that $\mathcal{S}(z)_i \in [0, 1]$ and $\sum_{i=1}^{N} \mathcal{S}(z)_i = 1$, where N is the number of output classes. These aspects make the softmax function suitable for multi-class classification tasks.

$$\mathcal{S}(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} \tag{3.6}$$

In order to estimate how well the network is performing during training, a *loss* function, L, is defined for the given problem. During the training procedure, the loss function is minimised to steer the network towards a given solution. In a classification task, a *categorical cross-entropy* loss function is used as the standard, which is defined in Equation 3.7. The truth labels are represented by t_i and the sum is performed over the number of classes N. As an example, in binary classification the truth labels can have values of 0 and 1. The categorical cross-entropy loss function, (CCE), is essentially the same as a cross-entropy loss but with the added softmax function, $\mathcal{S}(z)_i$, as the input.

$$CCE = -\sum_{i}^{N} t_i \log\left(\mathcal{S}(z)_i\right) \tag{3.7}$$

Having defined all the key aspects of a neural network model, the concepts of training the network can be introduced. When the network is trained this means that the set of tunable weights across all layers has been tuned to optimal values for the task at hand. Training even a small network has been a major challenge for the past decades due to their computational complexity. Neural networks are trained through the *back-propagation* technique, which uses the labels $\boldsymbol{y}$ of the training dataset to calculate the loss between

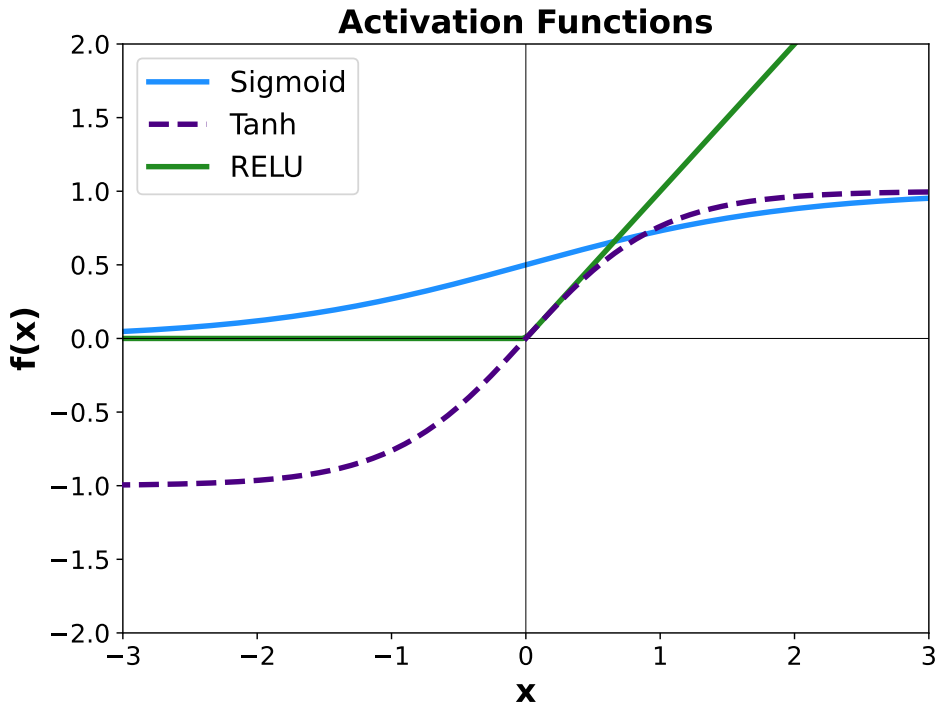


Figure 3.4.: Example of the most common activation functions: sigmoid, tanh and RELU. The output of the sigmoid function is bounded between 0 and 1 and can be interpreted as a probability. The tanh function is similar, but is instead bounded between -1 and 1. The RELU function is slightly different to these as it replicates the input for positive values but it is zero everywhere else [43].

the networks current output and its true value. The gradient of the loss function, ∇L is computed for all weights and biases. This information is used through the network to update these trainable parameters by a determined step-size in the direction of the gradient. The step-size is known as the *learning rate*. This process is iterated many times over until the loss function is minimized. Usually this will require multiple pass-throughs the whole training dataset and a single pass through is referred to as an *epoch*. There are many different slight variations on the training procedure, with one of the most popular being Stochastic Gradient Descent (SGD) [44].

3.3. Explainable AI - SHAP values

Machine learning models are relied upon across society to make automated decisions. It is essential to know why a particular model made a decision. For linear models, this is relatively simple to investigate as the weight of each feature can directly provide this information. In the case of a neural network, we can no longer do this as the weights of each layer have intricate connections. A model agnostic method to explain individual decisions by the model is to use Shapley values, which stem from cooperative game theory [45]. To translate this back into the field of machine learning, consider the case of a regression model that outputs a single score for a given set of input features. In this case, the contribution of a specific feature can be determined by evaluating the model score for all possible combinations of input features, excluding the feature of interest. This information is needed to calculate the Shapley value for the given feature. In more technical terms, the Shapley value represents the average marginal contribution of a specific feature to the model's output. Unfortunately, this computation of Shapley values quickly becomes an intractable calculation. The SHapley Additive exPlanations (SHAP) framework introduces some approximations to solve this issue [46]. For example, it uses the average value of a feature to calculate the marginal distribution of the feature. One advantage of the faster computation of the SHAP framework is that it enables the model to be interpreted globally across the entire dataset. Global interpretability is crucial to building an explainable model because it can indicate how important a feature is [47].

Chapter 4.

Statistical Analysis

"There are two possible outcomes: if the result confirms the hypothesis, then you've made a measurement. If the result is contrary to the hypothesis then you've made a discovery."

— Enrico Fermi

To make meaningful statements about the experimental data and its relation to the theory, a statistical framework is needed. These frameworks are at the core of data analysis in particle physics. Furthermore, several decisions in the architecture of the $HH \to b\bar{b}b\bar{b}$ search were driven by the expected results of the statistical framework. The statistical frameworks in particle physics are a common source of confusion and this chapter is dedicated to demystify the concepts that are relevant to the $HH \to b\bar{b}b\bar{b}$ efforts.

4.1. Introduction

In the $HH \to b\bar{b}b\bar{b}$ search, like in all analyses searching for new physics, we are interested first to make a discovery and, failing that, in knowing the extent to which our signal model can be excluded based on the data that has been observed. More specifically, we would like to know what the upper limit would be on the cross-section of the signal process we are interested in, such that we would not be able to detect it with some confidence. This is typically quoted at a 95 % confidence level. In order to do this, the $HH \to b\bar{b}b\bar{b}$ event reconstruction *predicts* two key pieces of information, the number of

signal events S and the number of background events B. This can then be compared to what we *observe* in data through our detector, $m = S + B$, with the same event selection. When dealing with a search, the number of events measured is reparametrized as a function of μ, which is defined as $\mu \equiv \sigma/\sigma_{theory}$. This parametrization allows for direct comparison with the theoretical cross-section prediction, which is typically taken to be the SM prediction and would correspond to $\mu = 1$. Similarly the background-only hypothesis corresponds to the case when $\mu = 0$. Given this parametrization the statistical search boils down to finding the permissible values of μ that would be consistent with the signal prediction at 95 % confidence. This can then be used to exclude theoretical models that are beyond this limit, which helps to constrain the search for new physics. In light of this, μ is sometimes referred to as the Parameter of Interest (POI) [48].

To a large extent each collision recorded in ATLAS can be considered statistically uncorellated from other collisions. This allows collisions to be treated as separate experiments, equivalent to multiple flips of a coin. The number of events measured in data can thus be considered as a Poisson counting experiment. For example, Equation 4.1 describes the probability of measuring m_i events when the predicted number of events is $\mu s_i + b_i$ in each bin i of a distribution:

$$P_i(m_i) = \frac{(\mu s_i + b_i)^{m_i}}{m_i!} e^{-(\mu s_i + b_i)} \tag{4.1}$$

In particle physics, we rarely have the luxury of knowing things with absolute certainty and hence systematic uncertainties need to be considered in the statistical framework. Both the signal and background predictions can have sources of uncertainty. To encapsulate all of the sources of uncertainties relative to each prediction the vector $\boldsymbol{\theta}$ is used to describe all of the Nuissance Parameters (NPs). The signal and background predictions are then described as a function of this parameter, $s(\boldsymbol{\theta})$ and $b(\boldsymbol{\theta})$. It is important to note that the NPs can be different between the signal and background predictions. In fact, it is the systematic uncertainties related to the background prediction that will have the largest impact on the sensitivity of the $HH \to b\bar{b}b\bar{b}$ search. Equation 4.1, can be modified to incorporate the effect of the NPs as shown in Equation 4.2

$$P_i(m_i, \boldsymbol{\theta}) = \frac{(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{m_i}}{m_i!} e^{-(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))} \tag{4.2}$$

4.2. Building the likelihood

Given that we are interested in comparing our model with the data that we decided to measure, the probability of observing the entire dataset needs to be computed. This is typically called a *likelihood* and it is a fundamental construct of the framework. The likelihood, $\mathcal{L}$, can be constructed by taking the product of the probabilities of all the bins in the histogram as shown in Equation 4.3. In this equation, the $f(x_i|\mu;\boldsymbol{\theta})$ term is used to describe a generic probability distribution function of a variable x_b, such as the invariant mass of the system, which is dependent on both μ and $\boldsymbol{\theta}$. In our case, this generic distribution can be replaced by the Poisson distribution shown in Equation 4.2.

$$\mathcal{L}(\mu,\boldsymbol{\theta}) = \prod_{b\in bins}^{N_{bins}} f(x_b|\mu;\boldsymbol{\theta}) = \prod_{b\in bins}^{N_{bins}} \frac{[\mu s_b(\boldsymbol{\theta}) + b_b(\boldsymbol{\theta})]^{m_b}}{m_b!} e^{-[\mu s_b(\boldsymbol{\theta})+b_b(\boldsymbol{\theta})]} \tag{4.3}$$

The likelihood described in Equation 4.3 has two parameters μ and $\boldsymbol{\theta}$. The parameter of interest, μ, is typically left unconstrained but the nuissance parameters are constrained as they are interpreted to be $\pm 1\sigma$ variations about the nominal estimate. This interpretation naturally leads to a Gaussian constraint added to the likelihood for each of the systematic variation, $\mathcal{G}(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$. The exception to this rule is when the variation can be interpreted as a statistical uncertainty, in which case a bin-by-bin Poisson constraint is added, $\mathcal{P}(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$. To make matters slightly more complicated, the non-resonant search utilizes different orthogonal channels, which also need to be incorporated into the likelihood. Putting all of these components together the full likelihood is described by Equation 4.4. Although the likelihood shows all of the main components, some complexity is hidden for simplicity. For example, the signal and background predictions have different sources of nuissance parameters. These likelihoods can be built using the *HistFactory* framework [49], however the $HH \to b\bar{b}b\bar{b}$ search uses the new pythonic binding of HistFactory called *pyhf* [50].

$$\mathcal{L}(\mu,\boldsymbol{\theta}) = \prod_{c\in channel}^{N_{channels}} \cdot \prod_{b\in bins}^{N_{bins}} \Big[\mathcal{P}(m_{b,c}|\mu s_{b,c}(\boldsymbol{\theta}) + b_{b,c}(\boldsymbol{\theta})) \times \mathcal{P}(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})\Big] \cdot \prod_{s\in systematic}^{N_{systematic}} \mathcal{G}(\tilde{\boldsymbol{\theta}}_s|\boldsymbol{\theta}_s) \tag{4.4}$$

4.2.1. Maximum likelihood estimation

Building the likelihood is essential to start building the statistical tests that will inform us about our model's compatibility with the experimental data. The parameter of interest can be fixed to a specific value or left unconstrained. If it is fixed to a specific value then a p-value test can be performed to check the compatibility of the data with that specific μ value. Alternatively, if μ is left unconstrained then the likelihood can be used to determine the best fit value given the data available via the process of Maximum Likelihood Estimation (MLE). The MLE search is performed over the parameters μ and $\boldsymbol{\theta}$, which results in their *unconditional* estimators, $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$. The MLE procedure can also be performed when one or more parameters are set to a specific value. In this case, the output of the optimization would lead to the *conditional* maximimum likelihood estimates, which are referenced as $\hat{\hat{\mu}}$ or $\hat{\hat{\boldsymbol{\theta}}}$. The conditional estimates become crucial in the hypothesis testing of the measured data. For example setting $\mu = 0$ would evaluate the background-only scenario relative to the measured data.

4.3. Testing the hypothesis

In order to test the compatibility of the model and the measured data, p-values are needed. In this case the p-values can be loosely interpreted as the probability that the measurement of $\hat{\mu}$ can fluctuate to a new value of μ given the observed data. Typically a cut-off is made at $\alpha = 0.05$, which indicates that if $\hat{\mu}$ is true we would observe μ only 5% of the time. The distribution of $\hat{\mu}$ is assumed to be Gaussian distributed. To compute the p-values first the test statistic needs to be built up, which begins with the profile likelihood ratio shown in Equation 4.5. This is typically then constructed into a test statistic by $\tilde{t}_{\mu} = -2 \ln \tilde{\lambda}(\mu)$, which is done only for numerical convenience. It can be shown that in the limit of large number of data samples the test statistic t_{μ} follows a *non-central chi-squared* distribution [51].

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} & \hat{\mu} \geq 0 \\ \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(0, \hat{\boldsymbol{\theta}}(0))} & \hat{\mu} < 0 \end{cases} \tag{4.5}$$

The parameter of interest, μ, is constrained to be $\mu > 0$ as a negative value would lead to an unphysical process. This constrain is reflected on the test statistic, $\tilde{q}_\mu$, shown in Equation 4.6.

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} = \begin{cases} -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(0,\hat{\boldsymbol{\theta}}(0))} & \hat{\mu} \leq 0, \\ -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\boldsymbol{\theta}})} & 0 \leq \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \tag{4.6}$$

Now that the test statistic is defined, the p-values can be determined by Equations 4.7 and 4.8. Equation 4.7 corresponds to the p-value for the signal+background hypothesis and 4.8 corresponds to the background only or null hypothesis.

$$p_{S+B} = \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu|\mu,\boldsymbol{\theta}(\mu))\, d\tilde{q}_\mu \tag{4.7}$$

$$p_B = 1 - \int_{\tilde{q}_{\mu,obs}}^{obs} f(\tilde{q}_\mu|\mu=0,\boldsymbol{\theta}(\mu=0))d\tilde{q}_\mu \tag{4.8}$$

The problem with calculating the aforementioned p-values is that the distribution of the test statistic, $\tilde{q}_\mu$ is needed to integrate over. Fortunately, at the asymptotic limit this has an analytical distribution that can be used for the required computations [51]. In cases where the analytical form cannot be used, then a toy-method based on sampling from the background and background+signal models can be performed. This is however a computationally exhaustive endeavour. In the $HH \to b\bar{b}b\bar{b}$ search the asymptotic formula can be used due to the large number of background events.

4.4. Confidence levels

Knowing the best fit values of μ is useful, however it is more useful to know the range of values that it can take while still being consistent with the measured data. Typically, upper limits of μ are quoted along $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals. The CL_S method

is employed to determine these limits [52]. The use of this technique is appropriate for the signals in the $HH \to b\bar{b}b\bar{b}$ search since the signal and background distributions are not well separable [53]. This method excludes values of μ if the confidence level is below a certain threshold of α, as depicted in Equation 4.9. The exclusion threshold is set at a 5 % level, much more relaxed than the threshold required for a discovery. In Equation 4.9, p_{S+B} refers to the p-value of the *signal+background* hypothesis, while p_B is the p-value for the *background-only* hypothesis. These p-values are computed under the test statistic $\tilde{q}_\mu^{observed}$ from the observed data. Recall that these p-values are obtained from the Equations showed previously, Equations 4.7 and 4.8

$$CL_S = \frac{p_{S+B}}{1 - p_B} < \alpha \tag{4.9}$$

Lastly, in order to gauge the level of expected sensitivity to a particular signal process prior to the unblinding of the signal region an Asimov dataset is used. This dataset is created by the expected background-only contribution, where μ and all the nuisance parameters are set to zero. Figure 4.1, shows an example of the CL_S values as a function of the μ parameter space, which allows us to obtain our upper limits on μ with its uncertainty bands.

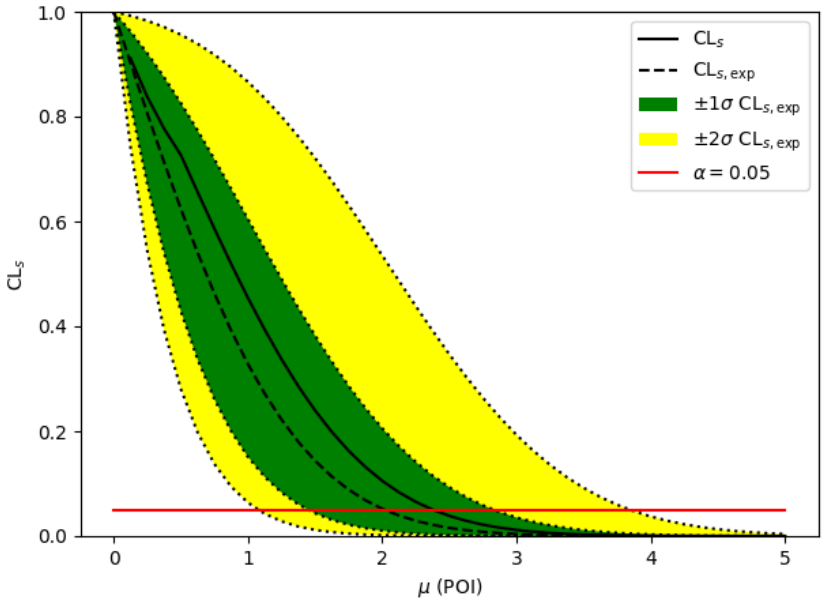


Figure 4.1.: Single bin example CL_S values and confidence intervals. The y-axis indicates the CL_S values for different upper limits on μ, which is the parameter of interest. The solid black line is the *observed* CL_S value, while the dashed black line is the *expected* CL_S value. The green and yellow bands correspond to the $\sim$68% and $\sim$95% uncertainty band on the estimated upper limit of μ. Since a 95% confidence level on the upper limit of μ is desired, the solid red line indicates the range of μ values at this level [54].

Chapter 5.

The LHC and ATLAS

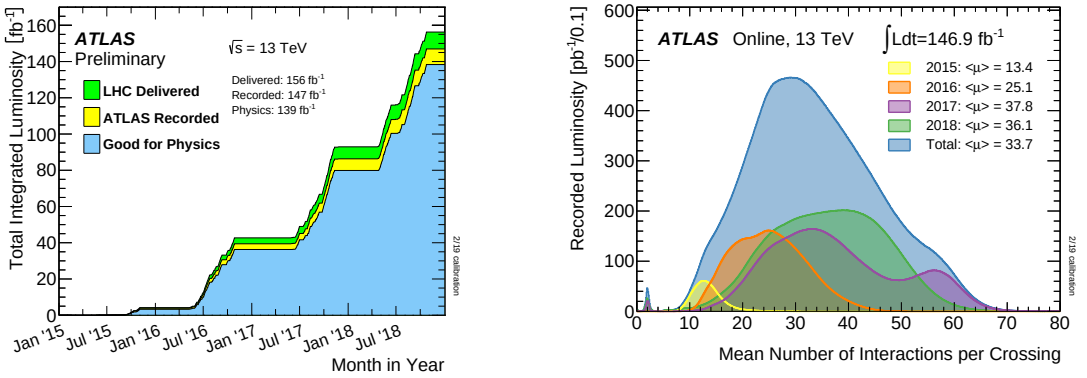
5.1. The LHC and Run-2 conditions

By the end of 2018 the Large Hadron Collider (LHC) completed its second data taking period, commonly referred to as Run-2. During this run, the machine was run closer to its design specifications by reaching a centre of mass energy of $\sqrt{s} = 13$ TeV. The The time between proton bunches that the Large Hadron Collider (LHC) delivers is know as the Bunch Crossing (BC). During Run-2 the BC was halved from the Run-1 final limit down to 25 ns. The instantaneous luminosity describes the number of collisions per second and relates to the total number of events when the cross-section (σ) for a process is factored in, as shown by Equation 5.1, where L_{INT} is the total integrated luminosity over a time period.

$$N = \sigma \times \int_{\Delta T} \mathcal{L}(t)dt = \sigma \times L_{INT} \tag{5.1}$$

Figure 5.1a, shows the total integrated luminosity as a function of time, the total luminosity delivered by the LHC was 156 fb^{-1} but due to data quality and operational constraints only 139 fb^{-1} are deemed *good for physics* by the ATLAS collaboration. Similarly, Figure 5.1b shows the distribution of the mean number of interactions per BC, where the expected values can be computed by $\langle\mu\rangle = \mathcal{L}(t) \times \sigma_{inel} \times 25\,\text{ns}$. The maximum instantaneous luminosity, $\mathcal{L}(t)$, delivered by the LHC during Run-2 was $2 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$. In the near future the LHC will restart with its third run of collisions (Run-3), which will deliver a similar amount of luminosity as Run-2 at roughly 150 fb^{-1}. After this, the LHC will undergo a long-term shutdown to begin the upgrades for the HL-LHC, which is designed to deliver a ten-fold increase in luminosity to 3000 fb^{-1} by increasing

the instantaneous luminosity to approximately 5–7.5 $\times 10^{34}$ cm^{-2}s^{-1} . This increase in instantaneous luminosity will increase $\langle\mu\rangle$ to 140–200 , which will present huge challenges for the collaboration [55].



(a) Run-2 Luminosity delivered.

(b) Distribution of the number of interactions per BC.

Figure 5.1.: Run-2 luminosity and pile-up conditions in ATLAS [56].

5.2. The ATLAS experiment

"In one of the world's least helpful analogies, I have often described this as like 'a sort of high-tech cylindrical onion'."

— Jon Butterworth, Smashing Physics, 2014

At the LHC there are four main experiments, two of them are designed to operate as General Purpose Detectors (GPDs), A Toroidal LHC ApparatuS (ATLAS) and CMS. ATLAS is a hermetic[1] detector that is used to study proton-proton and heavy ion collisions. GPDs are designed to have 3 different components, a tracker, a calorimeter and a muon spectrometer. Each of these subsequent components are further out from the Interaction Point (IP). This chapter will discuss the main aspects of the ATLAS detector. A cutaway of the ATLAS detector is shown in Figure 5.2, which illustrates both the scale and positioning of the main subcomponents.

[1]Hermetic detectors are designed to encompass the interaction point as much as possible. These are also called 4π detectors as they are designed to cover nearly all of the solid angle.

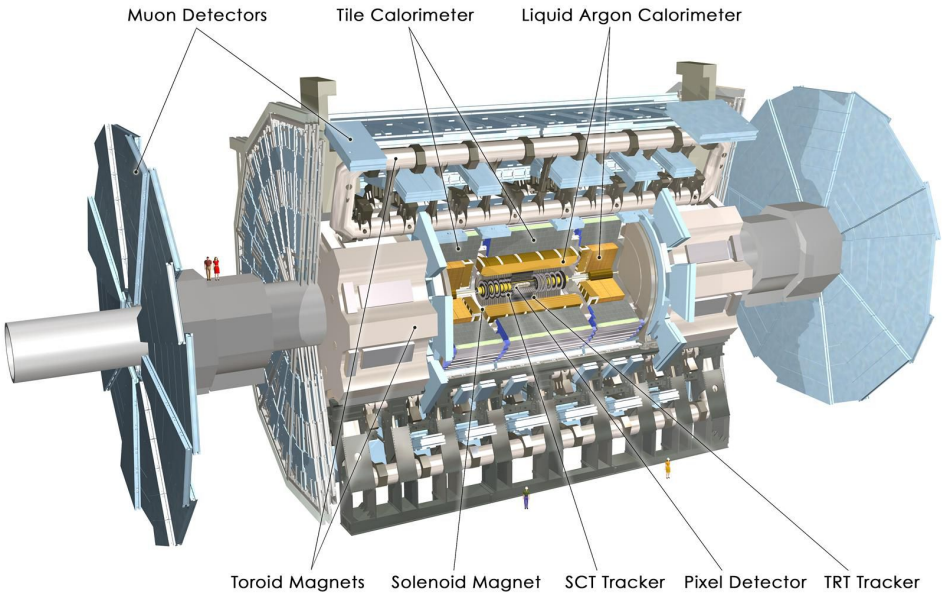


Figure 5.2.: Cut away of the ATLAS detector [57].

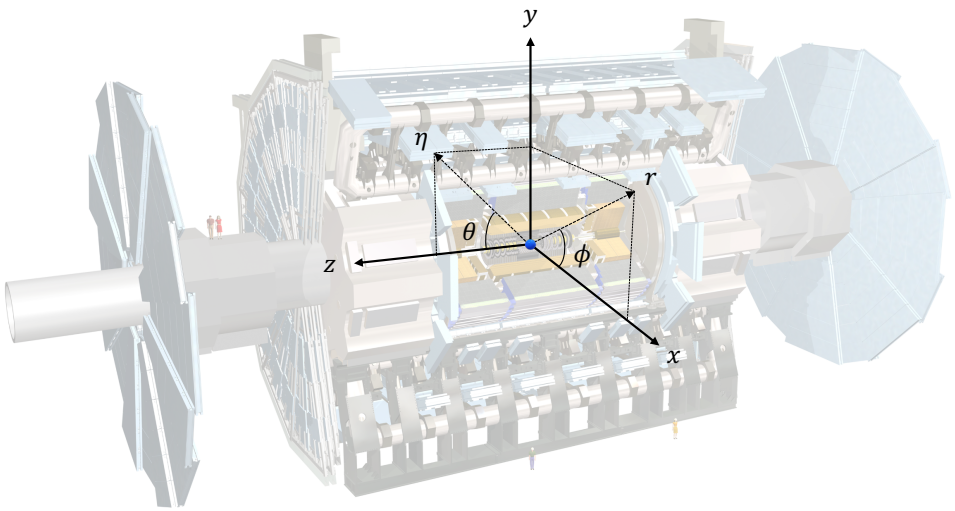


Figure 5.3.: ATLAS coordinate system. The x axis points towards the centre of the LHC ring, the z axis is in the direction of the beam and the y axis points vertically upwards.

5.2.1. ATLAS coordinate system

To aid the understanding of the detector components, the ATLAS coordinate system is outlined in this section. To describe the detector components and physics events a right-handed coordinate system is used with the ATLAS detector, where the z-axis follows the beam. The x-axis points towards the centre of the LHC ring and y-axis points vertically upwards. The azimuthal angle, $\phi \in (-\pi, \pi]$ describes the position along the

transverse $(x - y)$ plane and is measured from the x-axis. Similarly, the polar angle $\theta \in [0, \pi]$, describes the position along the longitudinal plane $(x - z)$. The polar angle is commonly replaced by the more useful *pseudo-rapidity*, which is defined as

$$\eta = -\ln(\tan(\theta/2)) \tag{5.2}$$

The pseudorapity is used instead of θ because in this space the difference between two objects is invariant under a Lorentz boost along the z-axis. Figure 5.3 shows an overlay of the coordinate system on the ATLAS detector, showing the variables described above. The distance of two objects is measured by the quantity ΔR, which is defined as,

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{5.3}$$

Lastly, the momentum of an object can be defined in the usual Cartesian coordinate system $\boldsymbol{p} = \big(p_x, p_y, p_z\big)$, where x, y, z refer to the direction in the axes shown in Figure 5.3. Typically the *transverse momentum* (p_{T}) of an object is used to describe its behaviour as it is Lorentz invariant in the z direction. This quantity can be determined by projecting the momentum onto the transverse plane, $x - y$, $p_T = |\boldsymbol{p}| sin(\theta)$. Similarly, the *transverse energy*, E_T, is another quantity that is used by the triggers and event reconstruction algorithms.

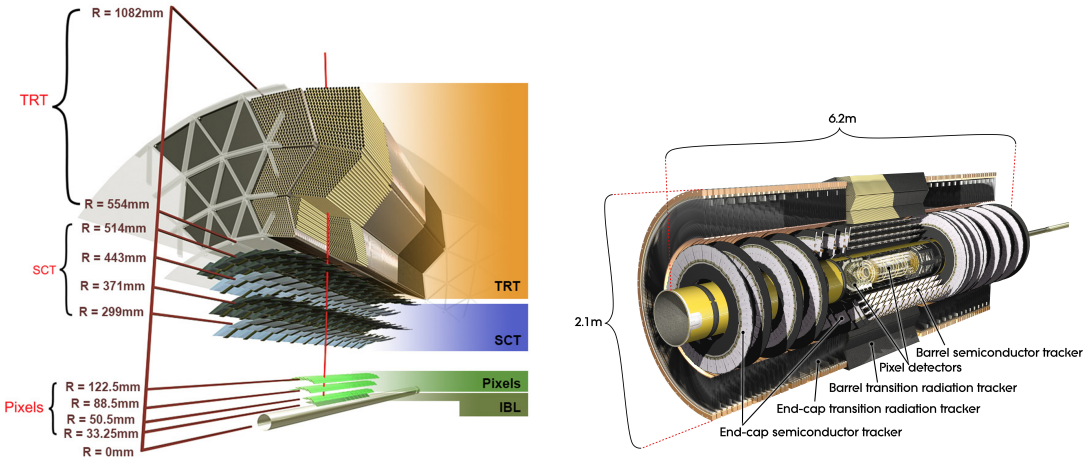
5.2.2. Inner detector

The Inner Detector is designed to measure the trajectory of charged particles without stopping them, up to $|\eta| < 2.5$. The ID is composed of 3 different sub-sections: The Pixel Detector, Silicon Microstrip Tracker (SCT) and Transition Radiation Tracker (TRT). Each of these sub-systems is more distant from the interaction point with respect to the previous one. Moreover, the whole ID is surrounded by a 2 T magnetic field provided by a solenoid magnet. The purpose of this magnetic field is to bend the charged particle tracks in order to measure the particle momenta. The p_T resolution of the whole ID system is given by Equation 5.4. The current ID design will be completely revamped for the HL-LHC with the all-silicon detector called Inner Tracker (ITk), which will help to deal with the harsher environment [60].

$$\frac{\sigma_{p_T}}{p_T} = 0.05\,\% \cdot p_T \oplus 1\,\% \tag{5.4}$$

Pixel detector

The innermost part of the ID is known as the Pixel Detector, which consists of a gargantuan 80.4 million independent channels with a pixel size of 50×400 μm. This sub-detector provides the most critical information for tracking in this dense environment and flavour tagging. In the barrel region, the Pixel Detector consists of four cylindrical layers where the closest layer to the beamline is known as the Insertable B-layer (IBL) [61]. In each end-cap there are 3 layers of pixels arranged in a disk shape and placed perpendicularly to the beam line. This layout can be seen in Figure 5.4b. The addition of the IBL helps to mitigate the reduction in performance of the original Pixel Detector from the radiation damage caused from Run-1. This damage resulted in a significant number of dead pixels. Moreover, the additional layer provides resiliency in the tracking algorithms when exposed to higher instantaneous luminosity conditions. Lastly, the IBL improves the resolution of the distance of closest approach in the $x - y$ plane of the track to the beam axis, d_0. This measurement is essential to identify B-hadron decays as their relatively long lifetime creates displaced secondary vertices from the primary interaction



(a) **(b)**

Figure 5.4.: Layout of the Inner Detector (ID) a) Shows the layout of the ID within the barrel region of the detector [58]. b) Shows the full layout of the ID, which includes the end-cap regions [59].

point. Improving this measurement has the most impact in b-tagging algorithms, which is of paramount importance to the $HH \to b\bar{b}$ and $HH \to b\bar{b}b\bar{b}$ analyses [61].

Silicon microstrip tracker

The SCT also relies on silicon semiconductor technology to detect the passage of charged particles through its volume. It extends the Pixel Detector by adding 4 cylindrical layers in the barrel region and 9 disk layers in each end cap. The number of channels in the SCT is reduced by an order of magnitude relative to the Pixel Detector to about 6.3 million strips. Each of these strips are 6.5 cm long and have a pitch of 80 μm. The reason for the SCT design is to reduce the cost of production and material used since the volume to cover is now larger but has lower particle multiplicity as it is further away from the IP relative to the Pixel Detector.

Transiton radiaton tracker

The TRT is the only part of the ID that does not rely on silicon technology to provide tracking information. Instead it relies on proportionally drift tubes, commonly referred to as straws, that are filled with a gas mixture (mostly Xenon). A gold-plated tungsten wire runs along the centre of the cylindrical straws. Each straw is put under a high voltage by using the conductive outer shell as a cathode and the inner wire as the anode. A charged particle passing through the straw will ionize the gas and due to the potential difference the electrons will drift towards the anode. The straws are interleaved with materials that have different dielectric constants to induce the transition radiation effect. Due to this design, charged particles will emit low energy photons that will ionize the gas in tubes, which creates an electron avalanche effect. The intensity of the transition radiation is proportional to the charged particle's Lorentz factor γ. This property can be exploited, along with careful calibrations, to provide initial Particle Identification (PID) between charged pions and electrons.

5.2.3. Calorimeters

The calorimeter in ATLAS is used to measure the energy of electrons, photons and hadronic particles. Except for the neutral pions measured by the TRT, the calorimeter is the main source of information about neutral particles in an event. The calorimeter

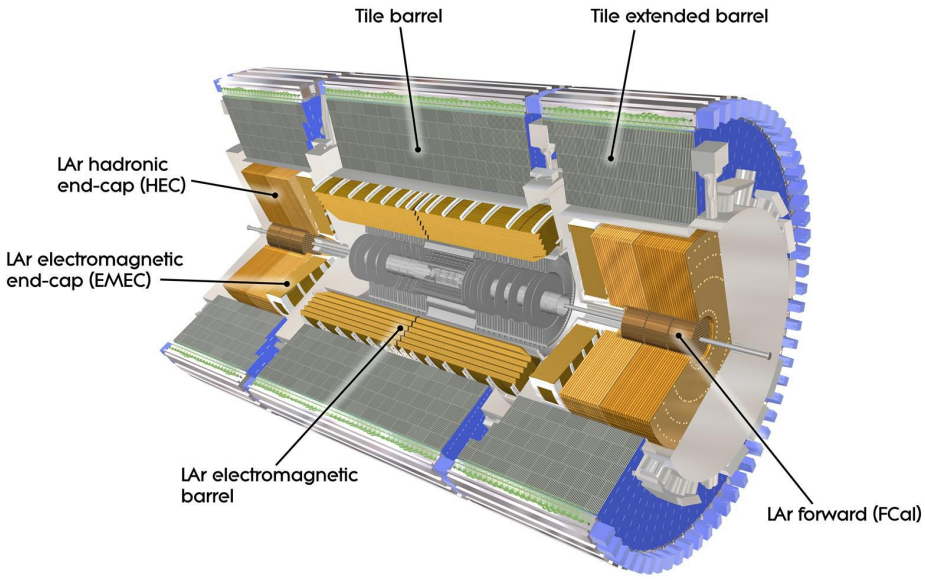


Figure 5.5.: ATLAS calorimeters systems [62].

system however is different to the tracker as it is designed to stop particles (except for muons) in order to measure their energy. Muons will only deposit a small amount of energy and will eventually make their way to the Muon Spectrometer (MS). ATLAS uses two types of calorimeter technologies to measure electromagnetic and hadronic showers. These are the Liquid Argon Calorimeter (LAr) and the Tile calorimeters. These two types of calorimeter designs are further split into 6 subsystems in total, which are described in Table 5.1. Figure 5.5, shows how these 6 subsystems are arranged around the ID. All of these calorimeter systems are designed as a sampling calorimeter, which is a fundamental difference with the calorimeter design used by the CMS experiment.

Name	**Shower type**
LAr Electromagnetic barrel (EMB)	electromagnetic
Tile Barrel	hadronic
Tile Extended Barrel	hadronic
LAr electromagnetic endcap (EMEC)	electromagnetic
LAr Hadronic endcap (HEC)	hadronic
LAr Forward Calorimeter (FCal)	hadronic/electromagnetic

Table 5.1.: Calorimeter subsystems of the ATLAS detector

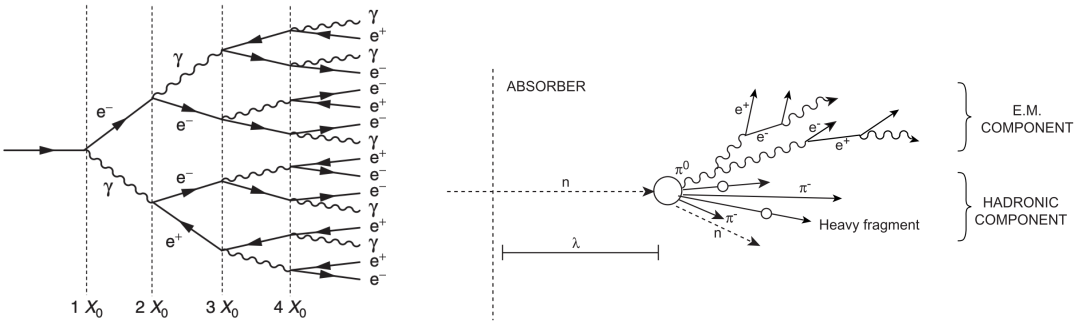
Electromagnetic calorimeters

The electromagnetic calorimeter is responsible for measuring the energy of electromagnetic showers produced in collisions, which can originate from particles such as electrons, photons and π^0 meson. The measurement of π^0 particles is due to the fact that it preferentially decays to 2 γ, with a branching ratio of approximately 98.8 % [63]. The electromagnetic calorimeter is also able to provide some discriminative power between these different particles. Electromagnetic showers occur through *bremsstrahlung* $e \to e\gamma$ and *pair-production* $\gamma \to e^+e^-$. These processes continue as they produce a cascade of photons, electrons and positrons, which is referred to as an electromagnetic shower as shown in Figure 5.6a. These showers are described by the *radiation-length* X_0, which is the average distance by which the energy is reduced by a factor of $1/e$ and is heavily dependent on the atomic number of the material. The value can be approximately calculated by Equation 5.5, where r_e is the classical electron radius, n is the number density of nuclei, Z is the atomic number of the material and α is the fine-structure constant ($\alpha = 1/137$) [2].

$$X_0 \approx \left(4\alpha n Z^2 r_e^2 \ln(287/Z^{1/2})\right)^{-1} \tag{5.5}$$

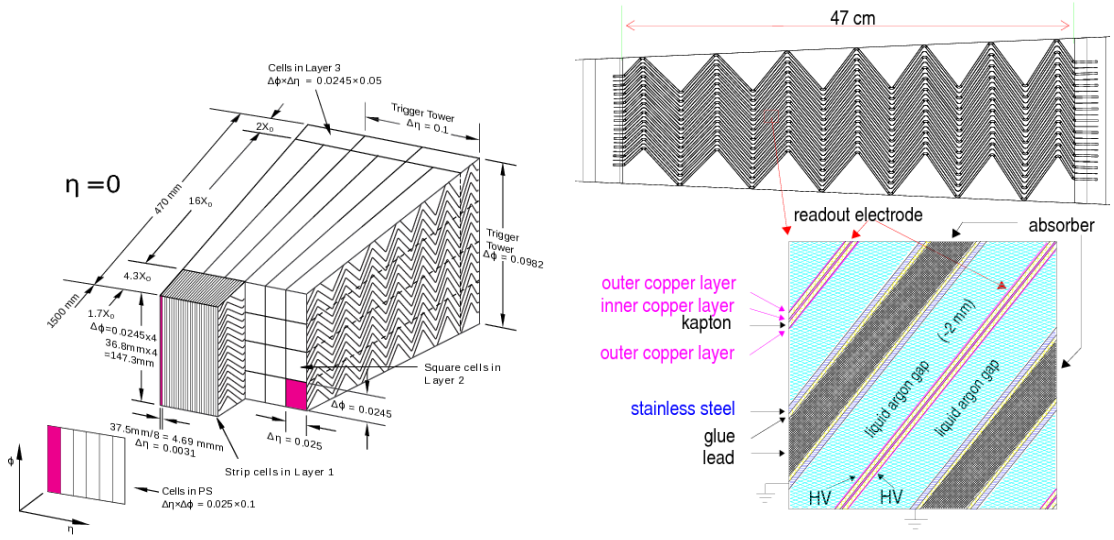
The electromagnetic calorimeter in ATLAS utilizes inert argon gas in liquid state (at 87K) as the active material and lead as the absorber material [64]. The Electromagnetic calorimeter (ECal) is split into two regions, the barrel regions which cover a range of $|\eta| < 1.475$ (24–32 X_0) and the two endcap regions covering a range of $1.475 < |\eta| < 3.2$ (24–38 X_0). The barrel sits between 2.8–4 m from the beamline and has the iconic *accordion* shape that is designed to reduce signal drift time. The end caps sit closer to the beam line between 33–209.8 cm. The radiation lengths (X_0) for the liquid argon and lead are 142.24 mm and 5.56 mm respectively. The first part of the LAr Forward calorimeter (FCal) is designed to measure electromagnetic showers in the range of $3.1 < |\eta| < 4.9$ and absorber material has been switched to copper. The resolution of the electromagnetic calorimeter in ATLAS is approximately given by Equation 5.6.

$$\frac{\sigma_E}{E} \simeq \frac{10\,\%}{\sqrt{E\,[\mathrm{GeV}]}} \oplus 0.7\,\% \tag{5.6}$$



(a) Electromagnetic shower [2] **(b)** Hadronic Shower [65]

Figure 5.6.: Examples of electromagnetic and hadronic showers and their characteristic properties.



(a) LAr module layout [66] **(b)** Accordion structure of the LAr barrel and the structure of electrodes and absorber plates [67]

Figure 5.7.: Structure of LAr modules.

Hadronic calorimeters

Hadronic particles, except for π^0, do not create electromagnetic showers as seen in Figure 5.6a. The only way for hadronic particles to be detected is through the nuclear interaction with the calorimeter material, that subsequently fragments the initial hadron into several secondary hadrons. These secondary hadrons can then undergo secondary interactions with the detector and so forth, giving rise to a hadronic shower. Hadronic showers are vastly more complicated than electromagnetic showers due to the nature of hadronic interactions. Hadronic showers do not develop in a uniform manner and the large multiplicity of final states gives them an inherently more varied nature. Moreover, any π^0 produced can decay electromagnetically via $\pi^0 \rightarrow \gamma\gamma$ which then causes an electromagnetic shower component of the full hadronic shower chain. Hadronic showers are typically parametrized by the *interaction length* λ, which is the mean free path of an incoming hadron given the nuclear cross-section (σ) and detector material density (A) and is determined by Equation 5.7, where N_A is Avogadro's constant [65]. The main components of a hadronic shower are shown in the example illustrated in Figure 5.6b.

$$\lambda = \frac{A}{N_A \sigma} \propto A^{1/3} \tag{5.7}$$

Ultimately, there is a significant component of energy lost due to nuclear break ups. In ATLAS, the full hadronic calorimeter system is able to provide a relative resolution on the energy of a single particle as described by Equation 5.8 below.

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E[GeV]}} \oplus 3\% \tag{5.8}$$

There are four detector subsystems responsible for recording hadronic interactions, their placement is critical since at this distance from the IP will be mostly hadrons. The Tile subsystems are designed by having alternating layers of steel and plastic scintillating tiles. Similar to the electromagnetic calorimeters, the detector operates by hadrons interacting with the steel tiles and producing showers of lower energy particles. These showers then pass by the scintillating plastic tiles and excite the organic molecules in the plastic which will emit scintillation light. This light is collected by readout fibers

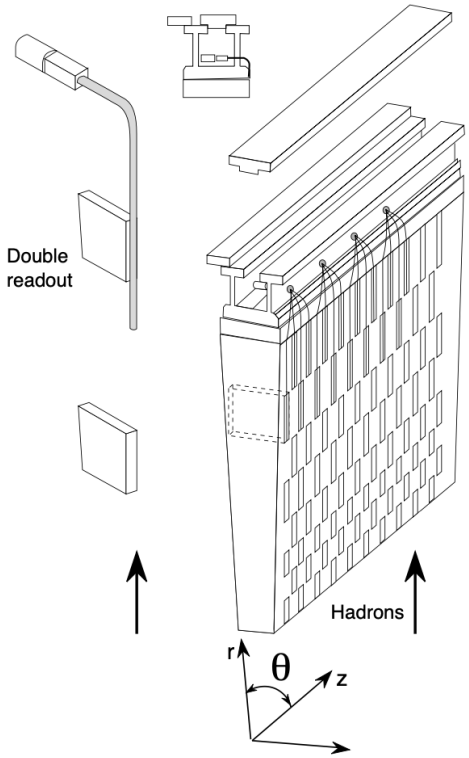


Figure 5.8.: Slice of Tile Hadronic Calorimeter and its coordinate system [68]

and read out by photomultiplier tubes. The choice of material selection is due to cost constraints for the large volume required. The outer region of the FCal is also used to provide hadronic event information in the forward region, to do this the absorber material has been switched to tungsten. A schematic of the LAr modules are shown in Figure 5.7 and a slice of the Tile calorimeter is shown in Figure 5.8.

5.2.4. Muon spectrometer

The outermost part of detector hardware in ATLAS is taken up by the Muon Spectrometer (MS), which provides tracking for muon reconstruction for $|\eta| < 2.7$. The layout of the MS system is shown in Figure 5.9. The MS system lies inside a large magnetic field that is provided by a set of toroid magnets (8 in the barrel and each end-cap), which give ATLAS its name. This magnetic field is used for the same purposes as in the ID, to bend the trajectory of muons and reconstruct their associated momentum. There are four different types of detection chambers used in the MS, which are aranged in the barrel region ($|\eta| < 1.05$) and two end-cap sections ($1.0 < |\eta| < 2.7$). The first technology are the Monitored Drift Tubes (MDT), which span the full η range of the MS and provide

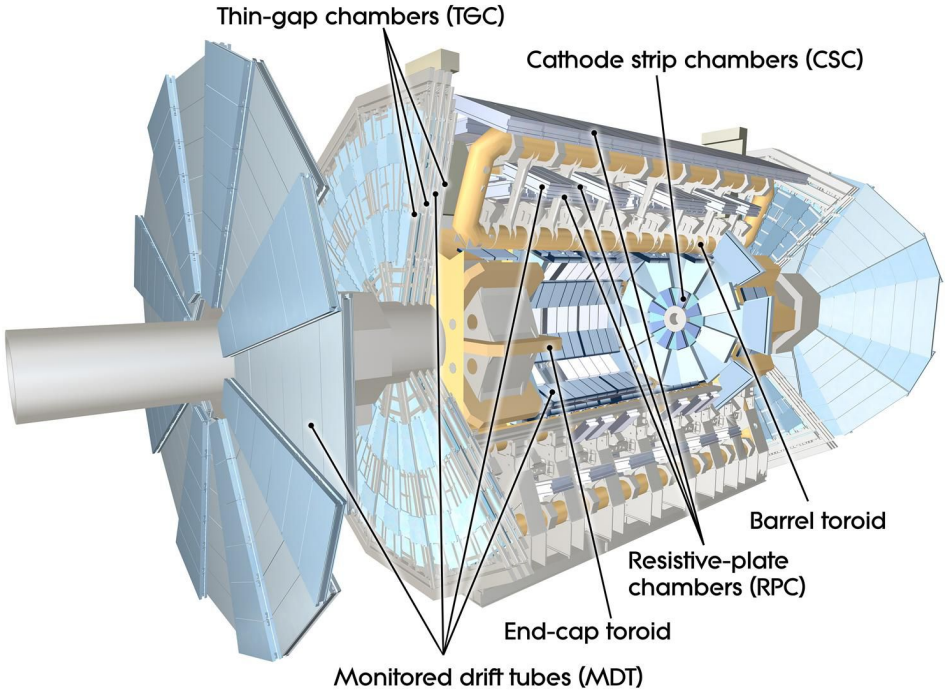


Figure 5.9.: Components of the Muon Spectrometer system. [66]

the best resolution for muon tracking. The MDT use the same detection principle as the TRT, except these tubes are made out of aluminum and filled with an argon-based gas mixture. This composition tunes the MDT for the detection of muons. In the innermost region of the endcaps the MDT are replaced by a slightly different technology called Cathode Strip Chambers (CSC). The CSC are essentially the same as the MDT, but the cathodes are replaced by long strips instead of tubes with several anode wires in each chamber's plane. This design is to provide finer granularity over the MDT, which is critical in the innermost region due large rates and backgrounds. The other two parts of the MS are the Resistive Plate Chambers (RPC) and the Thin Gap Chambers (TGC), which are deployed in the barrel and endcap respectively. The RPC and TGC have a time resolution of 1.5 ns and 4 ns respectively, which enables these chambers to be used for triggering purposes. They can also be used to provide additional tracking information for the offline reconstruction algorithms. The RPC are essentially a capacitor, with a narrow gap of 2 mm and the dielectric material is a gas mixture base on tetrafluoroethane. A high voltage of 8.9 kV is used in the RPC to drive the primary ionization into an avalanche for fast readouts to help the trigger decision. Lastly, the TGC are multi-wire proportional chambers that are similar to the CSC and are used in the endcap regions. During Run-1 and Run-2 most of the forward TGC muon triggers are fakes (at a rate

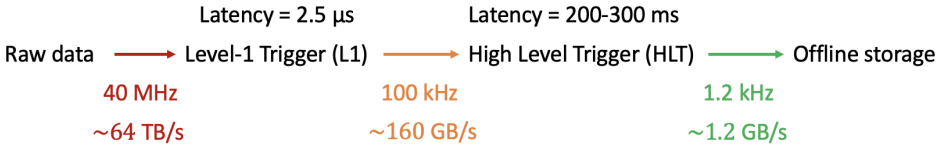


Figure 5.10.: ATLAS Run-2 dataflow, average event size 1.6 MB

of about 90%), which are typically low energy protons that are back scaterring from detector material elsewhere. To reduce the rate of fake triggers an upgrade to the MS system is being installed for Run-3 and HL-LHC, this is known as the New Small Wheel (NSW) [69]. The NSW, will ensure that trigger rates are maintained at an acceptable level without increasing the muon trigger p_T thresholds that would worsen the physics goals.

5.2.5. Data acquisition and triggering

Having described the main components of the detector, one can now appreciate the complexity required to read out the data, reconstruct it and then store it for physics analysis. To read the raw detector data there are custom systems designed for each part of the sub-detector and are beyond the scope of this discussion, more information can be found here [66]. For this Data Acquisition (DAQ) system to operate it needs to work in tandem with the instructions that the trigger system provides. This is because the sheer scale of data being read-out and processed at a 25 ns bunch crossing interval is unfeasible for current technologies as shown in Figure 5.10. The trigger system in ATLAS had a 2-level design during Run-2. The first stage is the Level-1 (L1) and the second stage is the High-Level Trigger (HLT). An example of the dataflow of the Trigger and Data Acquisition (TDAQ) system is shown in Figure 5.10. The L1 system is a completely custom hardware design that makes use of Field Programmable Gate Arrays (FPGAs) to process the 40 MHz raw data down to 100 kHz. Due to the limited buffer size in the detector's DAQ components, the L1 system has a strict limit of 2.5 μs to decide whether the event is kept or not. The events that are given the Level-1 accept signal are then passed down to the HLT system. The HLT system is made up of 40,000 Processing Unit (PU), which are essentially commodity components and it has a more generous latency of about 200–300 ms for a trigger decision. The longer latency for the HLT system, allows for more complicated algorithms to process the data which in turn allows for further suppression of the data rate. It is at this point in which the complicated b-jet reconstruction algorithms are applied. To reconstruct candidate b-jet events information

from the tracker needs to be used, which is available at the HLT level albeit with some restrictions due to high CPU usage. The events passing the HLT decision are then kept and passed to Tier-0 resources for permanent storage and offline reconstruction at a more manageable rate of roughly 1.2 GB/s rather than the gargantuan 64 TB/s from the raw data [70].

A more in depth description of the Run-2 TDAQ system is shown in Figure 5.11. The L1 system is split up into three major components: Level-1 Calo (L1Calo), Level-1 Muon (L1Muon) and the Central Trigger Processor (CTP). The L1Calo and L1Muon are initial processing stages of the L1 system by looking at specific Regions of Interest (ROI) in the calorimeter and muon detectors respectively. L1Calo performs rudimentary energy clustering and with rough calibrations checks if the energies are above the given trigger threshold. It can also perform scalar and vector sums of the calorimeter energy for triggers that depend on total energy or missing energy. Similarly, the L1Muon system uses the MS to trigger on muons that are above a certain p_T threshold. Additionally, Level-1 Topological trigger (L1Topo) provides additional rate suppression by being able to calculate more complicated quantities such as invariant mass and the angular separation between objects [71]. Ultimately, the CTP combines the information from the different ROI to provide the Level-1 Accept signal to the detector read-out using object multiplicity and energy thresholds.

It is important to note that during Run-2 no tracking information was used in the L1 system. This because the nature of track reconstruction is prohibitively expensive to compute since a naive combinatorial approach scales as a power law with particle multiplicity. In fact, the Kalman-Filter based approach used in CMS has a computational complexity of $\mathcal{O}(n^6)$, where n represents the number of particles [72]. Given the strict latency requirements, these are difficult to reconstruct even for the custom FPGA-based systems. To expand the tracking capabilities of the HLT system, the Fast Tracker (FTK) project was tested during Run-2 [73]. This ambitious project made use of Application Specific Integrated Circuits (ASICs) to store large pattern banks to then perform pattern recognition on FPGAs. This system was designed to provide tracks to the HLT system with a latency of 100 ms. The denser the environment of collisions, the more difficult it will be to suppress rates without tracking information. This will become a key development of the TDAQ system for the HL-LHC era, where the average number of interactions $\langle\mu\rangle$ will range between 140–200 The Hardware Tracking for Trigger (HTT) project was an attempt to design a system that will provide tracking information to assist in the trigger decision and was based on the same technologies as the FTK project, although

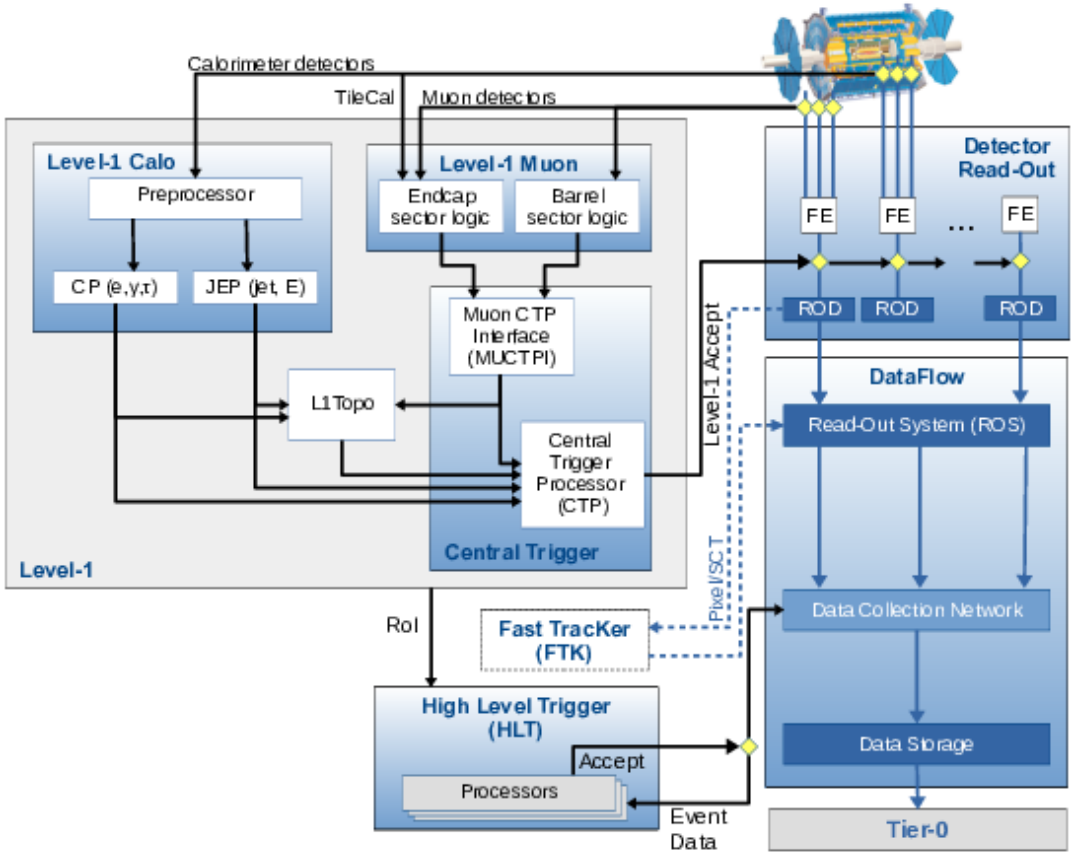


Figure 5.11.: ATLAS Run-2 TDAQ system [76]. The FTK system was used for testing purposes only.

its organization was simpler [74]. Alternative methods using commodity hardware such as CPU or GPU are also potential candidates for the HL-LHC era [75].

Chapter 6.

Event Reconstruction

The ATLAS detector is made up of many sub-components as described in Chapter 5. The information obtained from the detectors needs to first be reconstructed into more useful objects. Fortunately, the $HH \rightarrow b\bar{b}b\bar{b}$ search only requires four b-quarks in the final state, which means that the main objects used in the analysis are *b-tagged jets*. This section will describe how jets are reconstructed and flavour tagged.

6.1. Jets

The interaction strength between quarks is slightly more complicated than the QED counterpart. This is due to properties of the SU(3) symmetry group, which causes the potential energy between two quarks to increase linearly at larger separation distances, i.e. $V_{q\bar{q}} \propto r$ [77]. The consequences of this potential is that as a quark pair gets separated from each other, the field increases in density, comprising an endothermic process. When the energy associated to this field approaches the VEV of the gauge field, it is now energetically advantageous to create an extra quark and anti-quark pairs spontaneously from the vacuum. This process is known as hadronization and it serves to preserve the colourless singlet state due to quark colour charge confinement [78]. As illustrated by Figure 6.1, this process can repeat several times, leading to the previously mentioned hadronic showers until the final state hadrons are produced. This production of high energy quarks or gluons is then observed as highly collimated beams. These collimated beams are reconstructed into objects known as *jets*, which many physics analyses depend on.

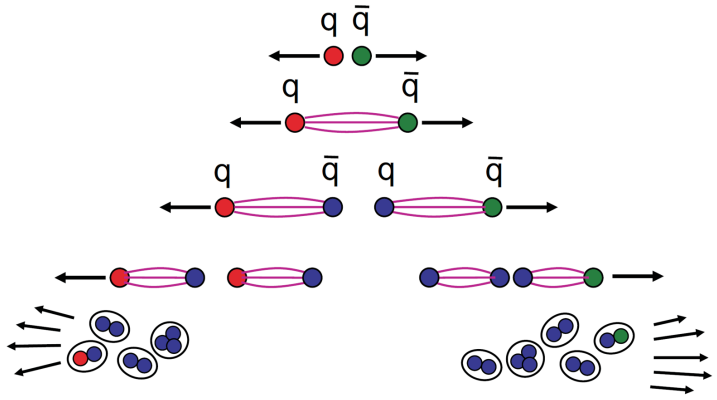


Figure 6.1.: Example of hadronization process that leads to the formation of *jets* [2].

6.1.1. Jet reconstruction

Reconstruction of jets is a critical component of the ATLAS event reconstruction and there are three main ways to do this. A jet reconstruction algorithm maps the momenta of final state particles (i) into the momenta of a certain number of jets (k), such that: $\{p_i\} \mapsto \{j_k\}$. There are two main categories of jet reconstruction algorithms: cone algorithms and sequential recombination algorithms. Cone algorithms create clusters of objects based on their proximity. Cone algorithms, except for the *SISCone* algorithm [79], are not Infrared Safe (IS) nor Collinear Safe (CS), but they are however fast to execute. Recombination algorithms work instead by backwards propagating through all the possible combinations of objects and choosing the combination that minimizes their distance d_{ij} as shown by Equation 6.1, where R is the radius parameter. $k_{ti(j)}$ refers to the transverse momentum of either object i or j. The quantity Δ_{ij}^2 refers to the distance between the objects i and j in terms of rapidity y and azimuthal angle ϕ as defined by Equation 6.3. The distance between object i and the beam B is quantified by d_{iB} in equation 6.2. The exponent p in Equation 6.1, can take values of -1, 0 and 1. This value corresponds to the three main types of recombination algorithms available: the Cambridge/Aachen (C/A) algorithm ($p = 0$) [80, 81], the k_t algorithm ($p = 1$) and the anti-k_t algorithm ($p = -1$) [82]. The anti-k_t algorithm is used as the default jet reconstruction algorithm by all of the LHC collaborations. This preference is due to the fact that the anti-k_t algorithm maximises resilience against underlying events and pile-up and their regular shapes allow for simpler detector-related corrections. Different reconstruction algorithms may be used for specific purposes and they are all available in the *FastJet* package [83]. The $HH \to b\bar{b}b\bar{b}$ uses two collections of jets: anti-k_t with

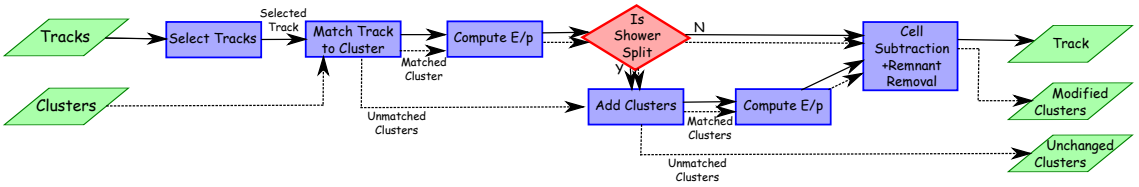


Figure 6.2.: Schematic of the inner workings of the PFlow algorithm. The algorithm starts with tracks and clusters as the input. The output collection is made up of tracks and clusters that have been geometrically matched and subsequently modified. Clusters which have not been matched to a particle track are also part of the output collection as these can be formed by neutral particles. Further details on each specific section of the PFlow algorithm can be found in [84].

$R = 0.4$ for the resolved channel and anti-k_t with $R = 1.0$ for the boosted channel. These jet collections are typically referred to as small-R and large-R jets respectively.

$$d_{ij} = min\left(k_{ti}^{2p}, k_{tj}^{2p}\right) \frac{\Delta_{ij}^2}{R^2} \tag{6.1}$$

$$d_{iB} = k_{ti}^{2p} \tag{6.2}$$

$$\Delta_{ij}^2 = \left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2 \tag{6.3}$$

6.1.2. Particle flow objects

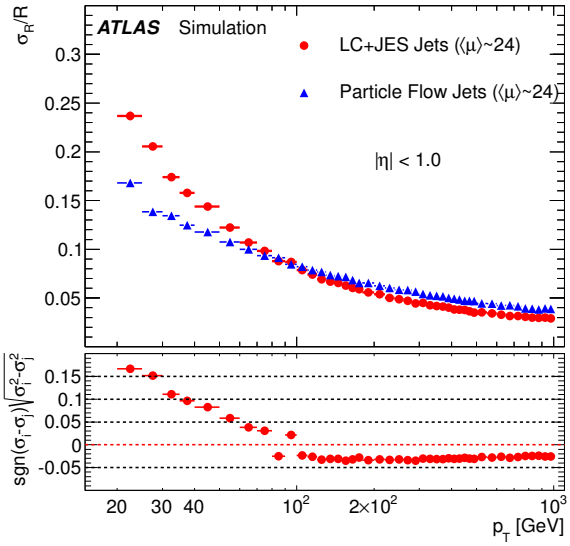
As jet algorithms are designed to reconstruct hadronic events, the information from the calorimeters is used to form *topological clusters* (topo-clusters). These topological clusters are a three-dimensional reconstruction of particle showers [84]. Topo-clusters were used throughout Run-1 and early Run-2 as the input to jet reconstruction algorithms. To improve on the performance from the topo-clusters, the Particle Flow (PFlow) algorithm combines the information from both the inner tracker and calorimeters which is then used by the jet algorithms. The PFlow method improves the energy resolution of low energy jets since the larger deflection in the ID allows for higher resolution of energy reconstruction. Typically calorimeters suffer in resolution at low energies due to the stochastic nature and noise level. However, when the energy of the event is higher the tracker suffers in momentum resolution performance as the deflection by the magnetic

field is lower but now the calorimeters have an improved resolution since the level of stochastic noise becomes less significant. Moreover, PFlow objects offer better pileup suppression from soft particles that the calorimeter alone is unable to distinguish.

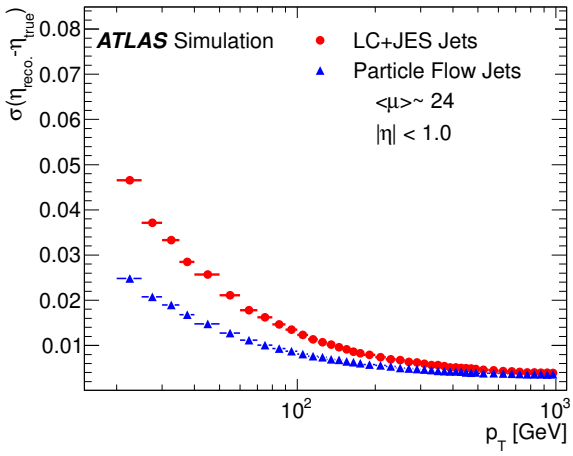
Although PFlow algorithm offers superior performance by using additional information from the tracker, it adds a complication. That complication is that the track to be used needs to correctly be assigned to its signal in the calorimeter. This is needed to avoid double-counting the energy during reconstruction. To do this, particle tracks are geometrically matched to a calorimeter cluster. This allows for the expected energy deposited by the track to be subtracted from the calorimeter cluster. The schematic shown in Figure 6.2 shows the various steps involved in the reconstruction of PFlow objects. A comparison of the performance between the PFlow jets and jets using only calorimeter information is shown in Figure 6.3. In general, PFlow jets offer better resolution at lower p_T, however at high p_T PFlow performs slightly worse. At high p_T, the dense core of the jet becomes more difficult for the tracking algorithms resulting in a lower tracking efficiency and accuracy, which ultimately results in the performance degradation of PFlow at high p_T.

6.1.3. Jet calibrations and corrections

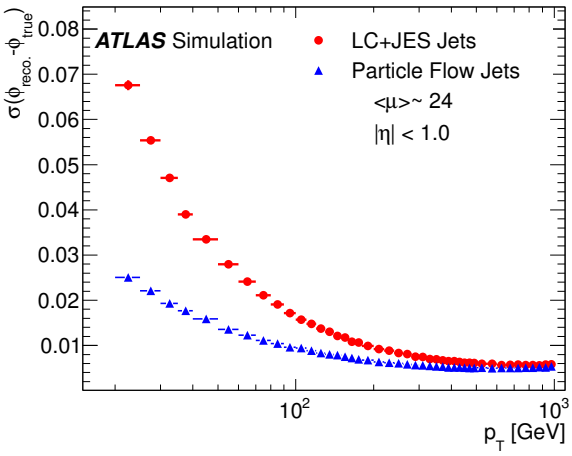
To improve the accuracy and precision of the reconstructed objects several calibrations and corrections steps are taken. Most of the inaccuracies come from both *in-time pile-up* and *out-of-time pile-up* as well as any detector related inefficiencies and defects. These inaccuracies will cause the reconstructed jets to have an energy that is different to the true energy of the jet, as obtained from the Monte Carlo simulations. Figure 6.4 shows a summary of the steps taken to calibrate the energy, mass and p_T of jets. The first correction stage addresses the out-of-time pile-up caused by residual energy left in the calorimeter from previous bunch crossings. The next stage removes the in-time pile-up which depends on the mean number of particles per bunch crossing (μ) and the number of reconstructed primary vertices (N_{PV}). In the next part of the chain information derived from Monte Carlo simulations is used to re-scale the four momentum to the truth particle level, which compensates for detector defects. The Global Sequential Calibration (GSC) then applies six multiplicative correction factors to correct for further discrepancies of the detector. Specifically, these correction factors improve the jet energy resolution and reduce sensitivity to the flavour content in the shower. Lastly, a residual *in-situ* calibration is applied to data to minimize the differences between data and simulations.



(a) p_T resolution



(b) η resolution



(c) ϕ resolution

Figure 6.3.: Performance comparison between PFlow jets and calorimeter jets (LC+JES) as a function of jet p_T. a) Shows the p_T resolution. b) shows the η resolution and c) shows the resolution in ϕ [84].

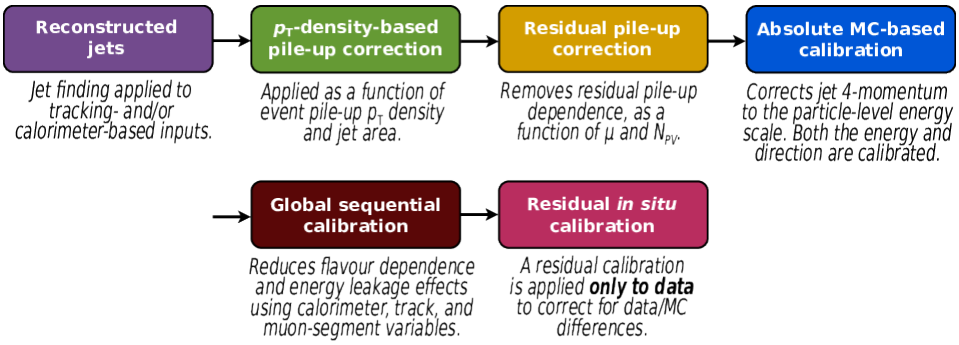


Figure 6.4.: Jet energy calibration stages which are applied to the four-momentum of the jet [85].

This calibration is derived from reference objects that have been measured with high precision, such as photons, Z Bosons and calibrated jets. More in-depth description of all the calibrations performed on jets in ATLAS can be found in [85].

When dealing with b-quark initiated jets a further correction needs to be applied, which arises from the fact that B hadrons can decay semi-leptonically. If the B hadron decays semi-leptonically by producing an electron then this energy will be accurately captured by the calorimeter and reconstructed in the jet. However when the semileptonic decay of the B-mesons produces a muon, this muon will pass through the calorimeter and deposit a minuscule fraction of their energy. To correct for this, b-jets have a *muon-in-jet* procedure applied to them that adds back the portion of the b-quark energy that was taken away by the muon. This is only done if there is a muon matched to a b-jet by checking if it is within $\Delta R = 0.4$ of the jet axis; if so, then the muon four-vector is added to the jet four-vector.

6.2. Flavour tagging

The reconstruction of jets is also followed by classifying the flavour of partons that the jet originates, a major aspect of analyses involving quarks in their final state. Typically we refer to the process of identifying jets as originating from a b-quark as b-tagging. The classification task is grouped into three classes: b-jets, c-jets and light-flavoured jets (u, d, s or gluon initiated). c-hadrons are particularly difficult to discriminate from b-hadrons due to comparable lifetimes. The efficiency of a b-tagging algorithm, ε_j, is defined as the fraction of jets with flavour j passing a certain selection criterion. The rejection

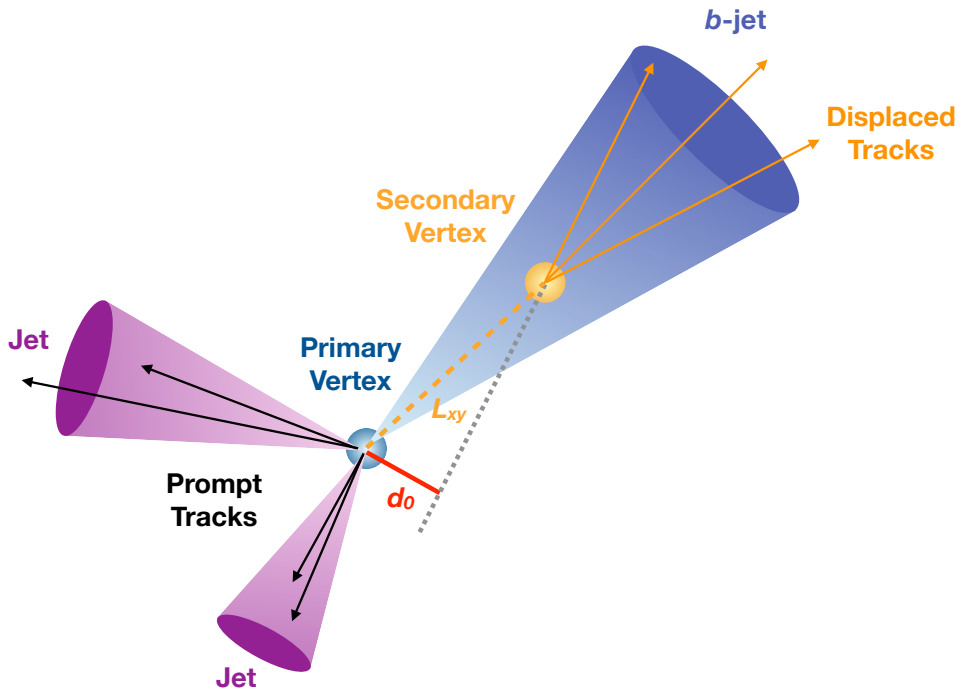


Figure 6.5.: Interaction producing a single b-jet and two light-flavoured quarks in the transverse plane. b-hadrons will typically travel a few mm before decaying, L_{xy} represents the decay length in the transverse plane. The transverse impact parameter d_0 is also shown. Note that the light-flavoured jets are reconstructed promptly on the primary vertex. [86]

is then the inverse, $1/\varepsilon_j$. b-tagging algorithms are designed to provide a pre-defined b-jet tagging efficiency, typically called the working point. This way the performance of different b-tagging algorithms can be compared is by looking at the c- and light-flavour jet rejection as a function of the b-jet efficiency. Moreover, working points are used to calibrate the b-tagging algorithms to data from the Monte Carlo (MC) simulations.

Figure 6.5 shows an interaction where two light-flavoured jets are produced alongside a single b-jet. Due to its lifetime, the b-hadron travels a distance L_{xy} before it decays (this is typically a few mm). This distance will create displaced tracks that originate from a secondary vertex. The transverse impact parameter, d_0, is known as the distance of closest approach of a displaced track relative to the primary vertex in the transverse plane. Typically the impact parameter, d_0, is large for tracks that originate from a b-hadron and small for light-flavoured quarks or gluons as their tracks will be closer to the primary vertex. This feature is used to distinguish between jets originating from light-flavoured quarks or gluons and b-hadrons. Additionally b-hadrons will also exhibit large mass, large number of tracks and carry a large fraction of the original parton's momentum, which aids

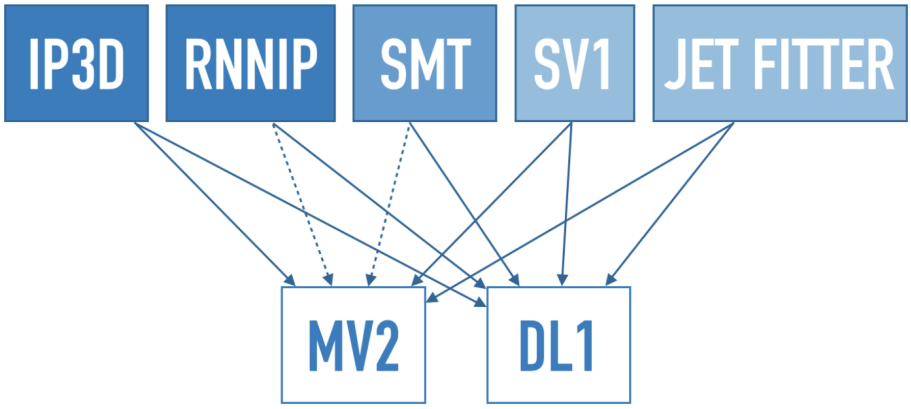


Figure 6.6.: Map of both low-level and high-level b-tagging algorithms. The high-level algorithms use the output of the the low-level algorithms as an input[87].

in the b-jet identification. Naturally, precise tracking information is required to accurately identify b-jets from other flavours. This is because both primary and secondary vertices need to be identified to determine the impact parameter. The b-tagging algorithms in ATLAS are classified into low-level taggers and high-level taggers. Low-level taggers were designed using physics based knowledge to create highly discriminating features that are detector specific. High-level taggers are then used to enhance the performance by using the information of the low-level taggers as an input and are detector agnostic. This interlink between low and high level taggers can be seen in Figure 6.6. The following sections will summarize these types of taggers.

6.2.1. Low-level taggers

Impact parameter tagging algorithms

As mentioned previously, the impact parameters and their resolutions are critical measurements to identify b-hadrons. d_0 is the impact parameter in the transverse $(r-\phi)$ direction and z_0 is the longitudinal impact parameter along the z direction. The significance of both the transverse and longitudinal impact parameters can be determined by $S_{d_0} = d_0/\sigma_{d_0}$ and $S_{z_0} = z_0/\sigma_{z_0}$ respectively, where $\sigma_{d_0(z_0)}$ represents the resolution of the impact parameter. The impact parameter taggers IPxD use smoothed and normalized distributions of the impact parameter significances and they are treated as a probability distribution function (PDF) for each of the hadron flavours (b, c or light flavoured jets). These PDFs, are described in Equation 6.4, where IP_m is the impact

parameter significance of track m. The PDFs are computed over all possible flavour combinations where $i, j \in b, c, light - flavour$. Ultimately, these PDFs are used to evalute the log-likelihood ratio (LLR), which is described in Equation 6.5, where N is the number of tracks associated to the jet.

$$PDF_{i(j)}(IP_m) = \begin{cases} PDF_{i(j)}(S_{d_0}, S_{z_0}) & IP3D \\ PDF_{i(j)}(S_{d_0}) & IP2D \end{cases} \tag{6.4}$$

$$log(p_i/p_j) = \log\left(\frac{\Pi_{m=1}^{N} PDF_i(IP_m)}{\Pi_{m=1}^{N} PDF_j(IP_m)}\right) \tag{6.5}$$

Secondary vertex taggers

A different approach to flavour tagging can be done by reconstructing the Secondary Vertex (SV), this can then provide several discriminating features. The SV tagger is an iterative process that first begins by building all possible two-track vertices using all of the tracks associated to a jet. Then these are used to construct a single vertex and an iterative process is applied to remove tracks with the lowest compatibility until the χ^2 value indicates that the vertex fit is of suitable quality.

Decay chain algorithms

From the measured values of the Cabibbo-Kobayashi-Maskawa (CKM) matrix we know that b-hadrons will preferentially decay to c-hadrons, as $|V_{cb}|^2 >> |V_{ub}|^2$. This provides an additional signature that allows for the identification of b-jets, since a b-hadron decay to a c-hadron will create a decay chain with a secondary and a tertiary vertex. To reconstruct this chain an iterative process based on a Kalman Filter (KF) is used, assuming both the b and c hadrons are along the same path. This procedure provides several variables that are used to provide discriminating power between the different flavours.

Recurrent neural network impact parameter taggers

To expand on the capabilities of the IP3D tagger a Recurrent Neural Network (RNN) approach is used to develop a b-tagging classifier, this is commonly referred to as RNNIP. These types of networks are suitable for problems where the input space is a sequence of events with an arbitrary length. RNNIP exploits this feature of RNN so that jets are represented as sequence of tracks that are ordered by the absolute transverse impact parameter significance. Each track has a vector of features consisting of: S_{z_0}, S_{d_0}, fraction of jet p_T carried by the track, distance between track and jet axis, track quality. An evolution to this approach has been explored by using *deep sets*, which yields faster training and prediction time over the current RNNIP approach [88].

Soft muon tagger

The Soft Muon Tagger (SMT) is designed to reconstruct muons arising from the semileptonic decay of b or c hadrons. The branching ratio for these semileptonic decays is $BR(b \to \mu\nu X) \approx 11\%$. These muons are considered to be *soft* with respect to the p_T of leptons arising from electroweak Boson decays. Although this is a low-level tagger, it is parametrized as a BDT to discriminate between b-, c- and light-flavoured jets containing a muon candidate [89].

6.2.2. High-level taggers - multivariate algorithms

The performance of the low-level taggers can be enhanced by combining their output through the use of Multivariate Analysis (MVA) techniques. In ATLAS, the two high-level taggers used for Run-2 data were parametrized as either a BDT or a Deep Neural Network (DNN). The BDT tagger is typically referred to as MV2c, while the DNN model is referred to as DL1. Both of these models were trained on a hybrid $t\bar{t} + Z'$ sample, where the b-jets are considered to be signal candidates and the others are the background. The number of c-jets present in the training sample of the MV2 tagger can be altered to improve the c-jet rejection at the cost of lowering the light-flavoured jet rejection. As an example, a common value of 20 % of c-jets in the background is used as an optimal level of impurity and these algorithms are referred to as *MV2c20*. Similarly to the MV2c tagger, the DL1 tagger is used to provide a multi-class discrimination between b, c and light-flavoured hadrons. In this architecture, the output nodes calculate

the probabilities associated with each jet flavour, i.e. p_b, p_c or $p_{light-flavour}$. As a final discriminant of b-jet classification, the outputs are combined into a tunable function that depends on the fraction of c-jets in the background, f_c, as shown in Equation 6.6. Additionally, there exist slight variations on the MV2 and DL1 taggers that use the information from either RNNIP and or SMT as input feature. These are named MV2Mu (MV2 + SMT), MV2Rnn (MV2+RNNIP), MV2MuRnn (MV2+SMT+RNNIP) for the MV2 based taggers and DL1mu (DL1+SMT), DL1r (DL1 +RNNIP), DL1rmu (DL1 + RNNIP + SMT) for the DL1 based taggers [87]. As an example, the performance of the two MV2 and DL1 taggers is shown for different working points in Table 6.1. A full list of input features, architecture and hyperparamaters of both MV2 and DL1 models can be found in [90].

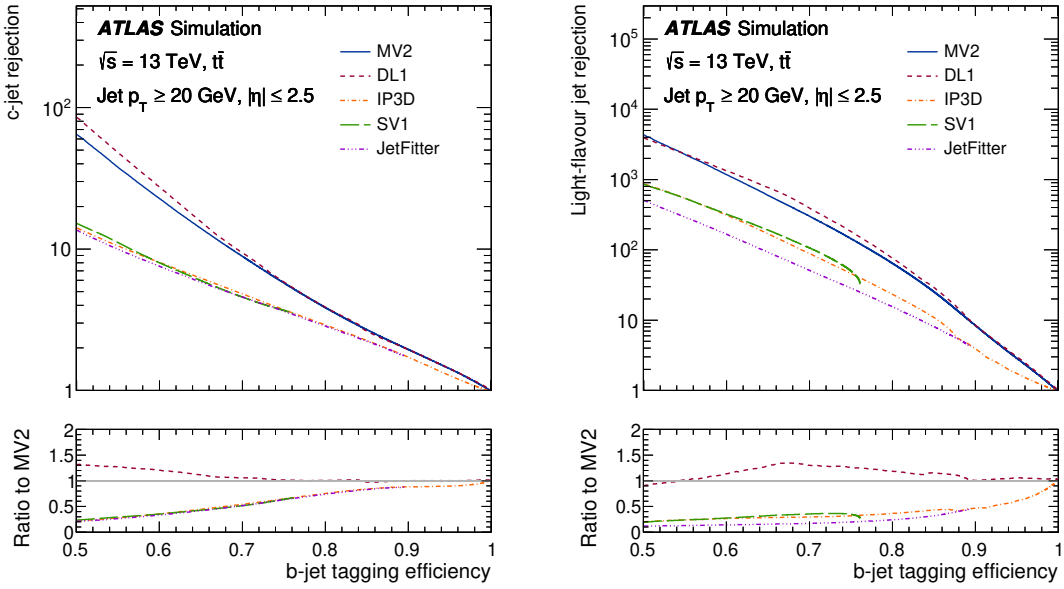
$$DL1(f_c) = \log\left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{light-flavour}}\right) \tag{6.6}$$

6.2.3. Performance comparison

To compare the performance of the different b-tagging algorithms the rejection power of c-jets and light-flavoured jets for a specific b-jet efficiency working point is evaluated. The rejection of the various low and high level taggers is shown in Figure 6.7. Table 6.1 compares the rejection power of both MV2 and DL1 for specific b-jet tagging efficiency points.

ε_b (%)	**MV2**				**DL1**			
	Selection	**Rejection**			**Selection**	**Rejection**		
		***c*-jet**	***t*-jet**	**light-flavour jet**		***c*-jet**	***t*-jet**	**light-flavour jet**
60	>0.94	23	140	1200	>2.74	27	220	1300
70	>0.83	8.9	36	300	>2.02	9.4	43	390
77	>0.64	4.9	15	110	>1.45	4.9	14	130
85	>0.11	2.7	6.1	25	>0.46	2.6	3.9	29

Table 6.1.: Selection and rejection performance different high-level taggers (MV2 and DL1), with different working points. DL1 offers superior rejection power in tighter working points [90].



(a) c-jet rejection **(b)** light-jet rejection

Figure 6.7.: Performance of low and high level b-tagging algorithms. Performance is measured by the rejection power of c and light-flavoured jets as a function of b-jet tagging efficiency. The performance is evaluated on $t\bar{t}$ events. High-level taggers offer a significant improvement over the low-level taggers. In general DL1 performs better than MV2 across the whole b-jet tagging efficiency spectrum [90].

6.2.4. Offline *b*-tagging calibrations

In general, MC simulations are not able to perfectly describe the performance observed in real data. This means that the performance of b-tagging algorithms, which is evaluated by using MC samples, needs to be corrected. This deviation between data and MC is corrected by applying a set of *scale-factors* to the MC events. To calculate these scale-factors the performance of the b-tagging algorithms needs to be measured in data. The performance of the b-taggers is dependent on the chosen working point and so a scale factor for each working point needs to be derived. In addition the scale factor also depends on each of the tagged flavours. The scale factor is defined in Equation 6.7 below

$$SF_j = \frac{\varepsilon_j^{data}}{\varepsilon_j^{MC}} \tag{6.7}$$

where SF_j represents the scale factor for a jet of tagged flavour j. The measured efficiency for tagging jets of flavour j is ε_j^{data}, while the efficiency in simulation is ε_j^{MC}. In a similar fashion an *inefficiency* scale factor, $\overline{SF}_j$ is also defined in Equation 6.8 below, which corresponds to the case where the jet is untagged.

$$\overline{SF}_j = \frac{1 - \varepsilon_j^{data}}{1 - \varepsilon_j^{MC}} = \frac{1 - SF_j \cdot \varepsilon_j^{MC}}{1 - \varepsilon_j^{MC}} \tag{6.8}$$

The measured efficiencies and hence the scale-factors are typically calculated as a function of p_T. Ultimately, an event-level scale-factor is derived by taking the product of all the individual tagged or untagged scale-factors. Figure 6.8 shows an example of the scale factors obtained for the M2V tagger.

6.3. Triggers and *b*-tagging

The previously mentioned b-tagging algorithms were all discussed from the perspective of offline event reconstruction, where precision tracking information across the whole detector is available. Running these algorithms in the online trigger environment (HLT level), is computationally expensive but possible provided that the single event reconstruction latency is below 500 ms. The benefit of running these algorithms at the trigger level is

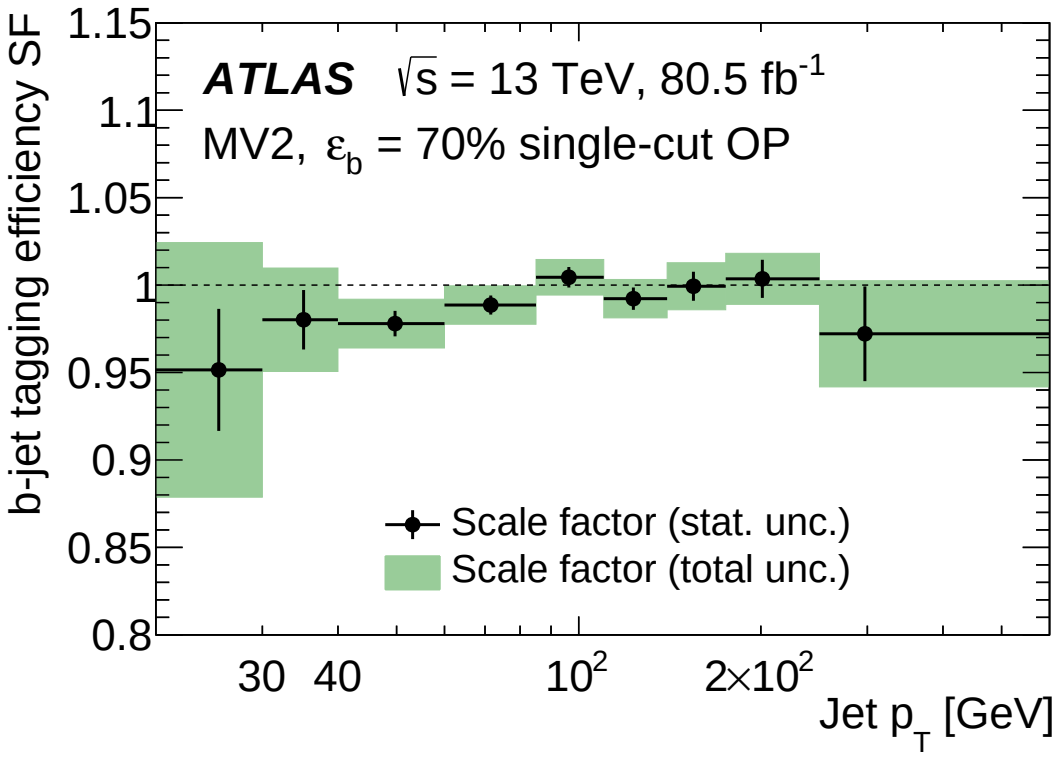


Figure 6.8.: Example scale factors to correct the b-tagging efficiency of Monte Carlo simulation to the measured efficiency in data. These scale factors are for the MV2 tagger at the 70 % b-tagging efficiency working point. The scale-factors are calculated as a function of the jet p_T [90].

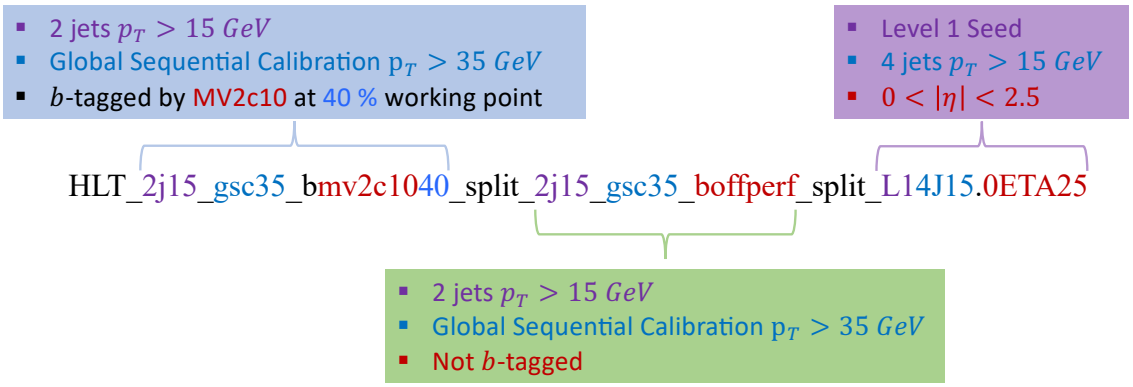


Figure 6.9.: Deciphering the naming scheme of HLT triggers to identify the trigger requirements. This trigger is an example of a $2b + 2j$ trigger used during data taking in 2017. This trigger uses the *mv2c1040* tagger, which is the BDT-based tagger with 10% *c*-jet background and it uses a rather tight 40 % *b*-jet efficiency working point. It also requires that all jets have an $p_T > 35GeV$ after the global sequential calibration (*gsc35*). Moreover, the HLT jets in this trigger example are constructed from a L1 seed that requires 4 jets with $p_T > 15GeV$ and are within $0 < |\eta| < 2.5$ as indicated by *L14J15.0ETA25*.

to record events with *b*-jets efficiently. More specifically, applying the *b*-tagging selection to the trigger can help to lower jet E_T thresholds, which ultimately affects the sensitivity of physics analyses that depend on *b*-quark final states. The *b*-tagging algorithms in the online environment are essentially identical to those in the offline environment. The only difference is that in the online environment, these algorithms are tuned to work with lower quality tracks and jets. During Run-2, only the MV2 tagger was used in the trigger as DL1 was being developed during Run-2. In ATLAS, a trigger menu is a collection of all the triggers used during data taking. These triggers must specify the chain of algorithms used at the L1 and HLT stages. Typically, a single trigger that maximizes acceptances is used for a physics analysis. In the $HH \to b\bar{b}b\bar{b}$ search, multiple of these triggers will be combined in a non-trivial way to increase the signal acceptance.

Deciphering *b*-jet trigger names

Each *b*-jet trigger has a specific kinematic signature that is used to reduce the readout. This signature can be obtained from the naming scheme, but it can be sometimes confusing to decipher. To explain all of the components that build the HLT trigger, a trigger from 2017 is used as an example. This trigger is shown in Figure 6.9 and because it requires 2 *b*-tagged jets and two non *b*-tagged jets it is referred to as *2b2j*.

6.3.1. Calibrating *b*-jet triggers

Understanding the efficiency of triggers is essential for evaluating the correct yield of signal and background in any physics analysis. The previously mentioned offline b-tagging algorithms do not use any b-jet trigger requirements for calibration. This requires us to calculate a conditional b-jet trigger efficiency relative to the offline b-tagging efficiency. This is defined as the fraction of b-jets that are b-tagged offline and matched to an HLT jet, while also passing the b-tagging requirements of the HLT trigger. With the efficiencies known, a scale factor (SF_b) can be derived to correct the simulation MC performance of the b-jet trigger from what has been observed in data. This correction is needed due to potentially unaccounted detector discrepancies. These efficiencies and scale factors need to be measured and computed for all possible combinations of b-tagging working points between the online and offline taggers. They also need to be applied in conjunction to the offline b-tagging scale factors [90]. Scale factors for light-jets and c-jets are not needed for the $HH \to b\bar{b}b\bar{b}$ analysis because, as will be described in Chapter 8, the analysis uses a data-driven method for the estimation of the background and hence it does not rely on the modeling of the c-jet and light-jet rejection in the simulation.

The conditional efficiency, $\varepsilon_b^{Trig|Off}$ is defined as the efficiency of a jet to be tagged as a b-jet by the online b-tagging algorithm given that it has also passed the offline b-tagging. The *Trig* notation refers to the jet that has passed the trigger (online) decision and the *Off* refers to the jet passing the offline decision. Similarly the *MC* notation refers to an efficiency measured from simulation and conversely the *Data* notation refers to an efficiency measured from data. Equation 6.9 shows the formal definition of how the scale factor is defined. All of the efficiencies and scale factors are calculated as a function of the *jet* p_T, but this notation is omitted for simplicity. Moreover, the scale factors at event level are derived by the product of all the jet-level scale factors such that $SF_{b,event}^{Trig|Off} = \prod_{i=1}^{N} \frac{\varepsilon_{b,i}^{Trig|Off,Data}}{\varepsilon_{b,i}^{Trig|Off,MC}}$, where the product is iterated over each jet i.

$$SF_b^{Trig|Off} = \frac{\varepsilon_b^{Trig|Off,Data}}{\varepsilon_b^{Trig|Off,MC}} \tag{6.9}$$

Scale factors are also needed to correct for cases when a jet has failed either the online and/or offline b-tagging requirements. To calculate the efficiency of these three regions Bayes' theorem can be applied to express these quantities with only ε_b^{Off}, ε_b^{Trig}

and $\varepsilon_b^{Trig|Off}$. Equation 6.10 is used to calculate the efficiency when the jet fails the trigger b-tagging requirement but passes offline b-tagging requirement. Equation 6.11 is the opposite, where the jet passes the trigger b-tagging requirement but fails the offline requirement. Lastly Equation 6.12 is used to calculate the efficiency when the jet fails both the trigger and offline requirements. A more detailed discussion into the calculation of these scale factors will be reserved for Chapter 7, when some of the key elements of the $HH \to b\bar{b}b\bar{b}$ search have been introduced.

$$\varepsilon_b^{\overline{Trig}\cap Off} = \left(1 - \varepsilon_b^{Trig|Off}\right)\varepsilon_b^{Off} \tag{6.10}$$

$$\varepsilon_b^{Trig\cap\overline{Off}} = \varepsilon_b^{Trig} - \varepsilon_b^{Trig|Off}\varepsilon_b^{Off} \tag{6.11}$$

$$\varepsilon_b^{\overline{Trig}\cap\overline{Off}} = 1 - \varepsilon_b^{Off} - \varepsilon_b^{Trig} + \varepsilon_b^{Trig|Off}\varepsilon_b^{Off} \tag{6.12}$$

Chapter 7.

Introduction to the $HH \rightarrow b\bar{b}b\bar{b}$ analysis strategy

"Looking for a silver needle in a stack of aluminium needles"
— Beojan Stanislaus.

7.1. Search introduction

There are two main categories of event reconstruction in the $HH \rightarrow b\bar{b}b\bar{b}$ search: *resolved* and *boosted*. The resolved category is used when the overall transverse energy of the dihiggs system is relatively small and the four individual jets of the b-quarks can be identified. In this category, the jets are then reconstructed as four small radius (R=0.4), anti-k_t jets. In the boosted category, the overall transverse energy of the system is relatively high and the two Higgs Bosons system receives a significant boost. This boost causes the jets of each Higgs to become more collimated and difficult to resolve individually. In this category the jets are thus reconstructed as large radius (R=1.0), anti-k_T jets. The search then exploits the rich substructure of these large-R jets to make further event selections. The resonant search uses a combination of the resolved and boosted categories and when events pass both categories the priority is given to the resolved category. These two categories are shown in Figure 7.1. Since the non-resonant SM signal peaks at about 400 GeV and most κ_λ variations have a softer spectrum, the boosted category is not needed in this search. The focus of this thesis will be on the resolved event reconstruction, but the awareness of the boosted category is important

for interpreting the combined limit results. This chapter will outline the main aspects of the resolved analysis strategy, which are: trigger selection, offline event selection, jet pairing strategy, background suppression, top-veto, kinematic region boundaries and the discriminating variable. Only a brief summary of the background estimation and trigger selection will be given in this chapter, with more intricate details reserved for Chapters 8 and 9.

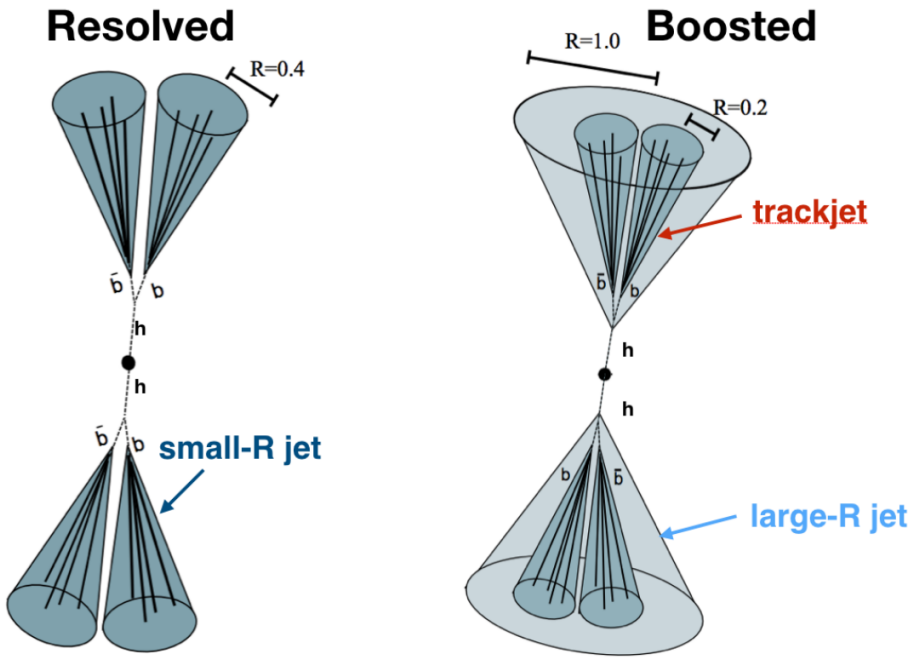


Figure 7.1.: Resolved and boosted event topology. The boosted category utilizes jets made up of tracks (trackjet) to enhance the offline selection [91].

7.2. Data samples

This analysis makes use of almost all of the available data recorded in Run-2 at 13 TeV in the ATLAS detector. Only the 2015 dataset has been excluded due to the lack of available b-jet trigger calibrations, which corresponds to a loss of 3.22 fb^{-1} of data. This loss corresponds to about 2.5 % of the total luminosity and because in 2015 the inferior IP3D+SV1 b-taggers were used this is likely to have an even smaller impact to the overall analysis sensitivity than what would be expected based on its fraction of the total Run-2 luminosity. Moreover, due to complications in 2016 related to the b-jet triggers and online vertex reconstruction around 9 fb^{-1} of data cannot be used for this analysis. This

Year	**Integrated Luminosity fb^{-1}**
2016	24.6
2017	43.65
2018	58.45
Total	126.7

Table 7.1.: The integrated luminosity for each year used in the resolved $HH \to b\bar{b}b\bar{b}$ analysis

ultimately leaves 126.7 fb^{-1} of data available for this analysis and is roughly 3.5 times the amount of data used in the early run-2 search. Other HH channels in ATLAS, such as $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$, are able to exploit the full Run-2 luminosity of 139 fb^{-1}, which is 9 % more than the available data for $b\bar{b}b\bar{b}$.

7.2.1. Monte carlo samples

Due to the final state in the $HH \to b\bar{b}b\bar{b}$ analysis, only the signals can be reliably reproduced via MC methods. This section describes the MC samples used for both non-resonant and resonant searches.

7.2.2. Resonant searches: gravitons and scalars

As mentioned before, the resonant searches use spin-0, narrow width scalar and spin-2 gravitons to serve as a benchmark signal for the search. The scalar and graviton samples were produced with a resonance mass of 251–5000 GeV. These samples were simulated at Leading Order (LO) using MadGraph [92], with HERWIG 7 [93, 94] to provide the parton shower part of the simulation. In all signal samples, EvtGen [95] is used to model the properties of bottom and charm hadron decays. Similarly, the spin-2 gravitons were also simulated at LO with MadGraph, but PYTHIA 8 [96] is used over HERWIG 7.

7.2.3. Non-resonant searches: SM signal and κ_λ variations

For the non-resonant ggF searches the SM ($\kappa_\lambda = 1.0$) and $\kappa_\lambda = 10.0$ signal samples are produced using the POWHEG BOX v2 generator at Next-to-leading Order (NLO) with a

finite top mass approach. Parton showers are also handled via PYTHIA 8, but HERWIG 7 samples are used to estimate shower uncertainties.

Due to the large computational expenditure of MC sample generation only the $\kappa_\lambda = 1.0$ and $\kappa_\lambda = 10.0$ points have been generated. A reweighting technique is used to generate the cross-section and kinematic distributions that allows the signal hypothesis to be modelled across the κ_λ parameter space of $\kappa_\lambda \in [-20, 20]$. To do this the expression for the differential cross section can be utilized as shown in Equation 7.1, and for simplicity the top coupling has been set to unity.

$$\frac{d\sigma(\kappa_\lambda)}{dm_{HH}} = |A(\kappa_\lambda)|^2 = |\kappa_\lambda \mathcal{M}_{\triangle}(m_{HH}) + \mathcal{M}_{\square}(m_{HH})|^2 \tag{7.1}$$

In the above Equation , $\mathcal{M}_{\triangle}$ is the matrix element of the triangle diagram and $\mathcal{M}_{\square}$ is the matrix element of the box diagram. By expanding out the matrix elements in Equation 7.1, the differential cross section for the the $gg \rightarrow HH$ process can be written as shown in Equation 7.2:

$$\frac{d\sigma(\kappa_\lambda)}{dm_{HH}} = \kappa_\lambda^2 a_1(m_{HH}) + \kappa_\lambda a_2(m_{HH}) + a_3(m_{HH}) \tag{7.2}$$

Here $a_1(m_{HH})$ and $a_3(m_{HH})$ correspond to the amplitude/matrix element of the triangle and box diagrams respectively. $a_2(m_{HH})$ corresponds to the amplitude of the interference between the two terms. The values of a_i are determined by solving a set of three linear Equations for the κ_λ values of 0, 1 and 20. A set of weights are then derived by the ratio of the given distribution to that of the $\kappa_\lambda = 1.0$ SM point. With these weights the SM $\kappa_\lambda = 1.0$ signal sample can be transformed to a sample with a different κ_λ value. This enables the search to set sensitivity limits as a function of the κ_λ and constrain the parameter space.

7.3. Event selection

7.3.1. Trigger selection

One of the most unique aspects of the $HH \to b\bar{b}b\bar{b}$ analysis is the use of multiple b-jet triggers. This combination allows the search to maximize the signal acceptance across the full kinematic phase space. This combination of triggers is unique to the $HH \to b\bar{b}b\bar{b}$ analysis but it presents a big challenge to correctly calculate the trigger scale factors which are needed to accurately estimate the signal efficiency. To circumvent this issue events are put into orthogonal categories by using offline kinematic selections that mimic the trigger decision. This categorization procedure is called *trigger buckets*. The design and optimization of the trigger buckets is reserved for section 9.4. As will be explained later the background estimation strategy requires the usage of events with exactly two b-tagged jets, which constrains the selection of triggers.

1b - single b-jet trigger requires only one b-tagged jet.

2b + 1j - requires two b-tagged jets with an additional non-b-tagged jet.

2b + HT - requires two b-tagged jets with an additional constraint that the value of H_T must be above some threshold. H_T is determined by the scalar sum of p_T of all the jets in the event.

2b + 2j - requires two b-tagged jets with two extra non-b-tagged jets.

Due to the characteristics of the non-resonant signal only the $2b + 2j$ and $2b + 1j$ triggers are used. Further details about the selection and optimization of the triggers used in the $HH \to b\bar{b}b\bar{b}$ search can be found in Chapter 9

7.3.2. Offline selection

The resolved $HH \to b\bar{b}b\bar{b}$ analysis selection requires at least 4 anti-k_t jets with a radius of $R = 0.4$ (small-R jets). These jets are required to have $p_T > 40\,\text{GeV}$, to move away from the trigger turn-on point, and $|\eta| < 2.5$, to ensure that they can be b-tagged within the tracker acceptance. Events are considered to be b-tagged if they pass the DL1r 77 % working point offline b-tagging decision. The working point refers to the b-jet reconstruction efficiency of the b-tagging algorithm as explained in section 6.2. This tagger and working point was shown to yield optimal results and is the nominal tagger

and working point used throughout this analysis. The jets are selected by sorting them using the binary b-tagging decision and then by p_T.

7.3.3. b-tagging categories

Events can be classified into different b-tagging categories depending on how many of the jets have passed the offline b-tagging conditions. The resonant analysis uses two categories, the *four-tag* **(4*b*)** and the *two-tag* **(2*b*)** categories. The non-resonant analysis builds on these two categories by adding two more, the $3b + 1loose$ **(3*b*1*l*)** and the $3b + 1notloose$ **(3*b*1*f*)** categories. The definitions and purpose of each category is explained below.

4*b* - At least four jets need to be b-tagged by the offline b-tagger. This is the main $HH \rightarrow b\bar{b}b\bar{b}$ signal region.

2*b* - Exactly two jets need to be b-tagged by the offline b-tagger. This is utilized for the data-driven background model, which is discussed in detail in Chapter 8.

3*b*1*l* (non-resonant only) - Exactly three jets need to be b-tagged at the 77 % working point but with an additional jet that fails the 77 % working point but passes the looser 85 % working point. This is used as an additional signal region for the non-resonant analysis.

3*b*1*f* (non-resonant only) - This category also requires exactly three b-tagged jets at the nominal 77 % working point but any additional jet has to also fail the 85 % working point. This is used as a validation sample for the non-resonant analysis.

7.4. Higgs candidate reconstruction

After the basic selections mentioned above, the next critical task is to pair the jets to reconstruct the two Higgs Candidates (HC). To visualize the impact of a particular pairing procedure a two dimensional histogram of the reconstructed Higgs Candidates masses is used. This $m_{H1} - m_{H2}$ plane is commonly referred to as the *massplane*. The great challenge behind the jet pairing procedure is that it can affect the massplane quite significantly and it therefore impacts the background estimation procedure. Ideally, the pairing procedure should result in a smooth massplane while maintaining a high signal efficiency. Prior to finding the correct pair of jets for each HC, events are preprocessed

to simplify the pairing procedure. The preprocessing selects four jets to then choose the best pairing out of the 3 possible configurations, which are illustrated in Figure 7.2. In the case of the $min\text{-}D_{HH} + \Delta R$ and BDT pairing, the scalar sum of jet p_T of the chosen pair is used to determine whether the HC is leading (HC1) or subleading (HC2). This section will discuss the three pairing strategies considered for the $HH \to b\bar{b}b\bar{b}$.

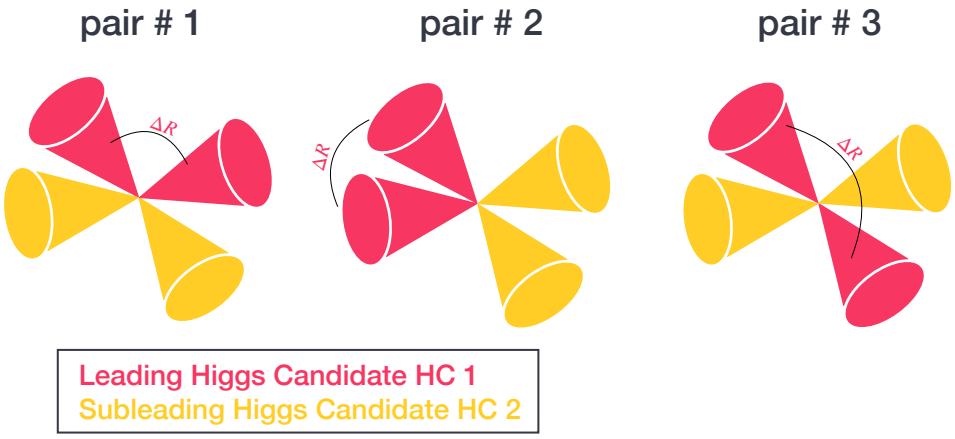


Figure 7.2.: Three possible pairing options when 4 jets are selected to be reconstructed into 2 Higgs candidates.

7.4.1. $min\text{-}D_{HH} + \Delta R$ pairing

This is one of the most intuitive pairing methods and it creates the candidates by picking the combination that yields the most similar invariant masses. Initially, the combinatoric background is reduced by requiring that any paring satisfies the conditions shown in Equations 7.3 and 7.4. These conditions were optimized to maximize the signal sensitivity of the Early Run-2 search. If an event does not contain a pairing that satisfies these conditions then the event is eliminated.

$$\frac{360}{m_{HH}/\text{GeV}} - 0.5 < \Delta R_{jj}^{lead} < \frac{653}{m_{HH}/\text{GeV}} + 0.47 \tag{7.3}$$

$$\frac{235}{m_{HH}/\text{GeV}} - 0.02 < \Delta R_{jj}^{sublead} < \frac{875}{m_{HH}/\text{GeV}} + 0.35 \tag{7.4}$$

If multiple pairings satisfy the ΔR_{jj}^{lead} and $\Delta R_{jj}^{sublead}$ conditions, then the pairing is chosen so that the distance in the massplane from the line (0,0) to (120, 110), is

minimized. This distance is referred to as D_{HH} and is depicted in Figure 7.3 and is mathematically defined in Equation 7.5. The values of 120 GeV 110 GeV were chosen in the Early Run-2 analysis to account for the the energy loss from semileptonic b-hadron decays. This pairing method unfortunately sculpts the background to peak along the diagonal line. The sculpting can be seen in Figure 7.5a, which shows the massplane of the $2b$ category after the $min\text{-}D_{HH} + \Delta R$ pairing. This pairing method was used in the early Run-2 analysis.

$$D_{HH} = \frac{\left| m_{H1} - \frac{120}{110} m_{H2} \right|}{\sqrt{1 + \left(\frac{120}{110}\right)^2}} \tag{7.5}$$

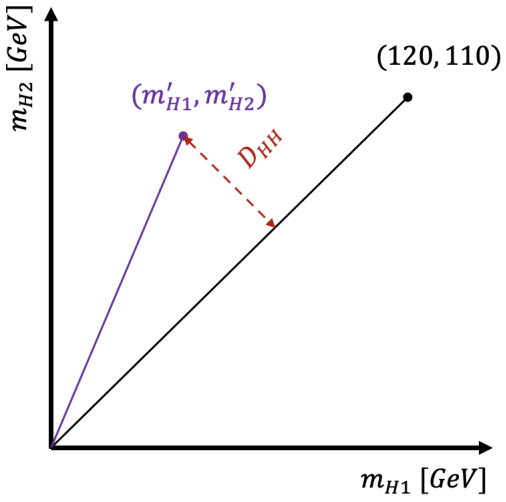


Figure 7.3.: $min\text{-}D_{HH} + \Delta R$ pairing

7.4.2. Boosted Decision Tree (BDT) pairing

To improve upon the sculpting issue present in the D_{HH} pairing method, a classification BDT was used for the resonant analysis. The model uses several kinematic variables to provide a *pairing-score* for each possible jet pairing, and the pairing with the highest score is selected. The score can be interpreted as the probability that it is a correct pair. The features of this model are: the total separation between jets ΔR_{jj}, the pseudo-rapidity difference between the jets $\Delta\eta_{jj}$ and the angular separation in the $x-y$ plane $\Delta\phi_{jj}$. The BDT model is also parametrized by the dihiggs mass, m_{HH}, to sustain a good pairing

efficiency across all of the resonance mass spectrum without using this feature as a discriminating variable, as this would ultimately lead to sculpting of the m_{HH} spectrum. Further details on the BDT Pairing architecture can be found in [97].

7.4.3. *min*-ΔR pairing

The pairing of jets into the HCs can also be done by another intuitive algorithm that utilizes the angular separation. To do this all the possible HC pairings are sorted by the vector sum of the constituent jet p_T. Then the angular separation, ΔR, is computed between jets in the leading HC, which is shown in Figure 7.2. The pairing is chosen so that the separation between jets is minimized, hence this method is referred to as *min*-ΔR pairing. This pairing method results in a smooth massplane while maintaining a high efficiency for the SM non-resonant signal. The disadvantage of this pairing method is that it discards low m_{HH} events, which yields low sensitivity for signal samples that peak in this region. These low mass signals occur frequently in low resonance mass signals and in κ_λ variations where the triangle diagram dominates.

7.4.4. Summary of pairing methods

The main problem that the pairing method presents is the sculpting of the massplane as this will result in a poor background estimate and lower the sensitivity of the analysis. The *min*-$D_{HH} + \Delta R$ pairing method used in the early Run-2 search showed severe sculpting of the massplane peaking in the signal region, as shown in Figure 7.5a. The BDT pairing method yields a smoother massplane with a peak that is largely outside of the signal region as seen in Figure 7.5b. Lastly, the *min*-ΔR pairing method yields the smoothest massplane overall without any peaks in the signal region, see Figure 7.5c. The efficiency of choosing the correct pair as a function of the invariant dihiggs mass m_{HH} for the three pairing methods is shown in Figure 7.4. The efficiency of the three methods above 500 GeV are very similar and close to 100 %, however below this mass the efficiency drops quite significantly. In particular, the *min*-ΔR pairing method has the steepest drop, while the BDT pairing method sustains the highest efficiency between all the pairing methods in this mass range. Since the BDT pairing method retains the highest signal efficiency in the low m_{HH} spectrum and the massplane sculpting is less severe than the *min*-$D_{HH} + \Delta R$ method, the BDT pairing method is used for the resonant analysis. The efficiency loss at high masses seen in the *min*-$D_{HH} + \Delta R$ pairing

is due to the ΔR constraints. The smooth massplane obtained by the *min*-ΔR pairing method makes it the optimal choice for the non-resonant analysis. The *min*-ΔR pairing method optimizes the analysis for the $\kappa_\lambda = 1$ point where the analysis has a competitive performance relative to other channels instead of focusing on the full κ_λ scan.

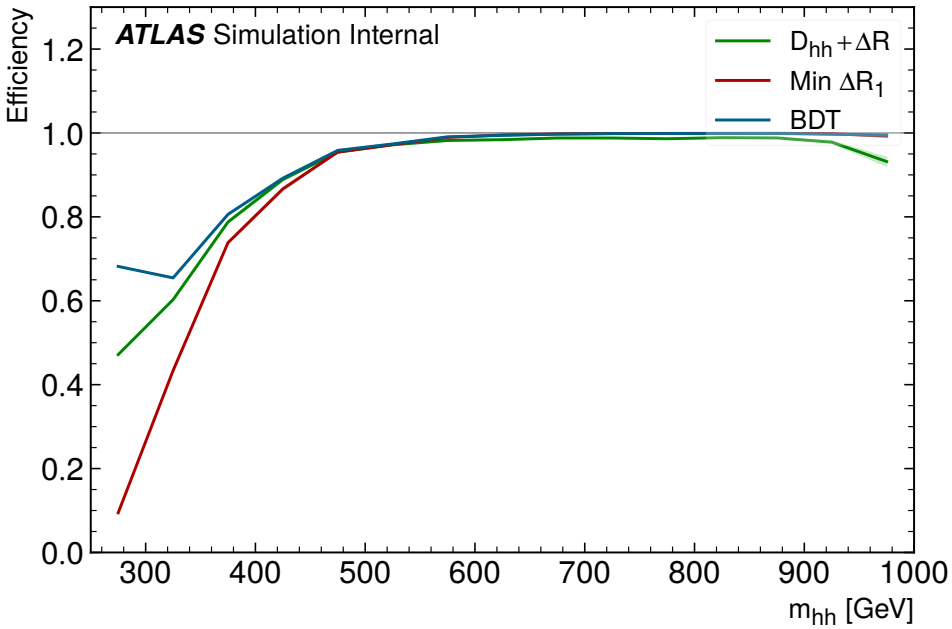
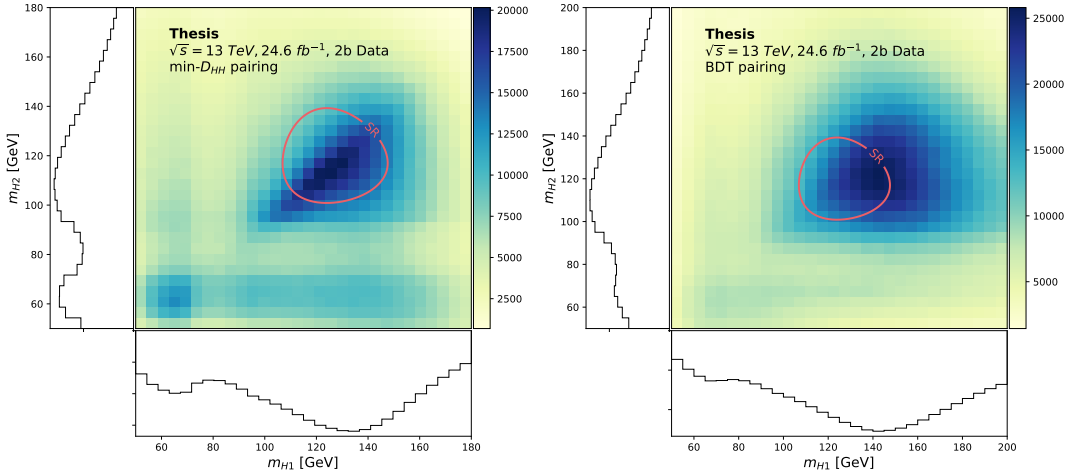
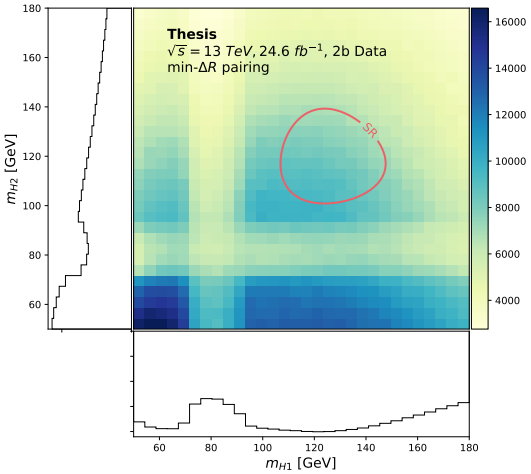


Figure 7.4.: Efficiency of choosing the correct pairs as a function of invariant dihiggs mass m_{HH} for the three pairing methods. [97]



(a) *min*-$D_{HH} + \Delta R$ **(b)** BDT

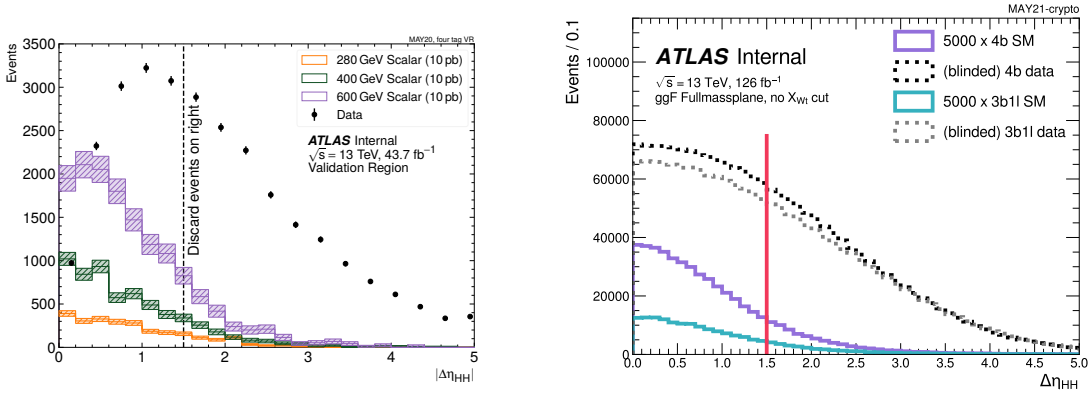


(c) *min*-ΔR

Figure 7.5.: m_{H1}-m_{H2} massplane of the $2b$ category with the three different pairing methods. The ring indicates the boundary of the signal region. The deficiencies seen as a stripes that are centered at about $m_{H1} = m_{H2} = 80\,\text{GeV}$ are due to the top veto. These massplanes were formed by using only data available from 2016, which is $24.6\,\text{fb}^{-1}$ of integrated luminosity. The *min*-$D_{HH} + \Delta R$ pairing method results in a significant amount of sculpting that peaks in the centre of the signal region. The BDT pairing method lessens the amount of sculpting while the *min*-ΔR pairing method completely eradicates it.

7.5. Background suppression

The main background source in the $HH \to b\bar{b}b\bar{b}$ search is the multijet QCD background, $2 \times (g \to b\bar{b})$. To significantly diminish this background the absolute pseudo-rapidity separation between the two Higgs candidates, $|\Delta\eta_{HH}|$, is constrained to be less than 1.5. The physical reason for this is that the HH signals are an s-channel diagram, whereas the background is mixture of s and t-channels. The two body t-channel diagrams typically give more forward topologies, or larger $|\Delta\eta|$ between the outgoing partons. The region defined by reversing this cut to $|\Delta\eta_{HH}| > 1.5$ is used to validate the background estimation model and is defined as the *reversed* $|\Delta\eta_{HH}|$ *region*. The $|\Delta\eta_{HH}|$ distribution for the resonant and non-resonant searches are shown in Figure 7.6.



(a) resonant **(b)** non-resonant

Figure 7.6.: $|\Delta\eta_{HH}|$ distribution for both resonant [97] and non-resonant analysis. Cutting out events with $|\Delta\eta_{HH}| > 1.5$ is used to suppress the background for both the resonant and non-resonant search.

7.6. Top veto

A portion of the background component in the $HH \to b\bar{b}b\bar{b}$ search comes from $t\bar{t}$. To mitigate the effect of this background a veto is applied on events consistent with the $t \to b(W \to q_1\bar{q}_2)$ decay. The branching ratio of a top quark decaying to Wb is very close to 100 %, this means that $t\bar{t}$ events will then contain at least 2 b-jets. The leptonic component from the W decay is reduced by vetoing events with isolated electrons or muons. To veto $t\bar{t}$ events, a physics inspired discriminating variable is used, X_{Wt}, which is defined in Equation 7.6. In this Equation , m_W and m_t are the masses of the W

Boson and t quark candidates. The value of 0.1 in the denominators is chosen as an approximation of the mass resolution. The top quark and W Boson masses are calculated for every possible triplet of jets and the combination yielding the lowest X_{Wt} value is chosen. Events with $X_{Wt} < 1.5$ are then discarded as they are attributed to $t\bar{t}$ decays. Prior to selecting the three jets associated to the b and W, the jets in an event are separated into HC jets and non-HC jets. The HC jets are the four jets that were used to build the Higgs Candidates by the pairing procedure, while the non-HC jets are any other jets that still pass the kinematic requirements. To build the W candidate two jets from all of the available jets are selected and the b candidate jet is selected from any of the remaining HC jets. Note that all of the W and b candidate jets must be distinct, so that jets that were used to build the W candidate cannot also be used to build the b candidate. To correctly select the b-jet an explicit b-tagging requirement should be enforced. In the resonant analysis this was not in place, but was subsequently fixed in the non-resonant analysis. The addition of this constraint has no ramifications on the $4b$ signal since all of these jets are b-tagged by definition. However, the constraint changes the distribution of lower tag regions such as $2b$. The corrected X_{Wt} definition leads to improved systematic uncertainties in the background estimation for the non-resonant analysis. In general the top-veto has a minimal effect on signal acceptance, but it can be considerable in low mass signal samples. For the resonant analysis, this definition was computed for all possible jets, whereas in the non-resonant analysis this was done for all b-tagged jets. This difference in the X_{Wt} definition leads to improved systematic uncertainties in the background for the non-resonant analysis. The distribution of X_{Wt} for both resonant and non-resonant searches are shown in Figure 7.7 and 7.8 respectively.

$$X_{Wt} = \sqrt{\left(\frac{(m_W - 80.4 \text{ GeV})}{0.1 m_W}\right)^2 + \left(\frac{(m_t - 172.5 \text{ GeV})}{0.1 m_t}\right)^2} \tag{7.6}$$

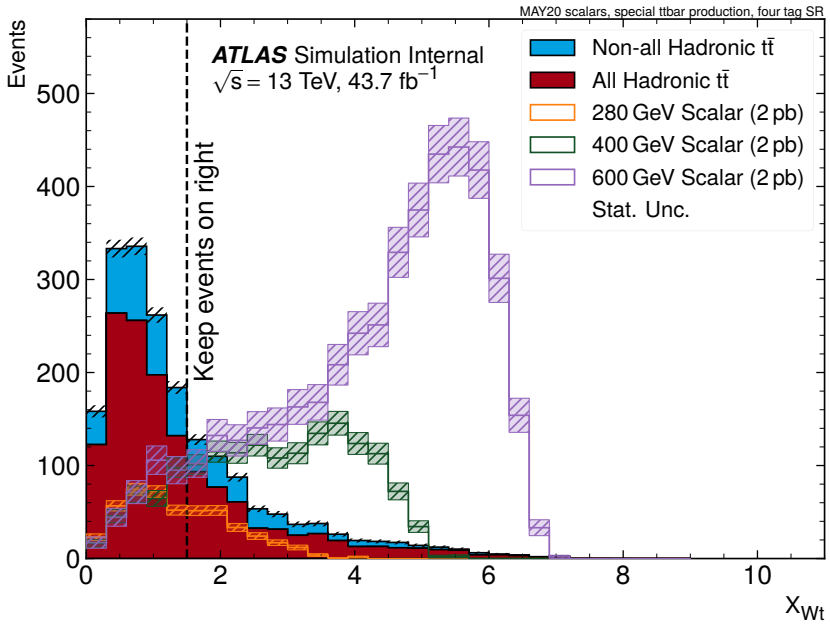
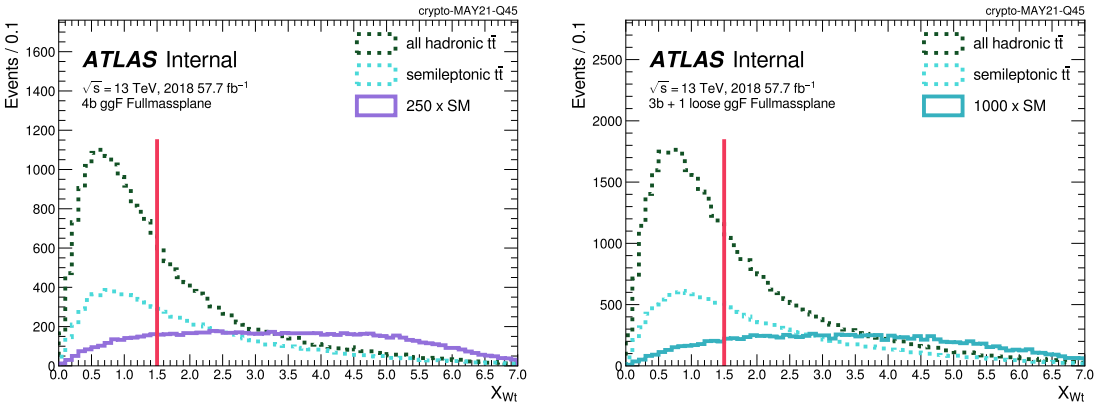


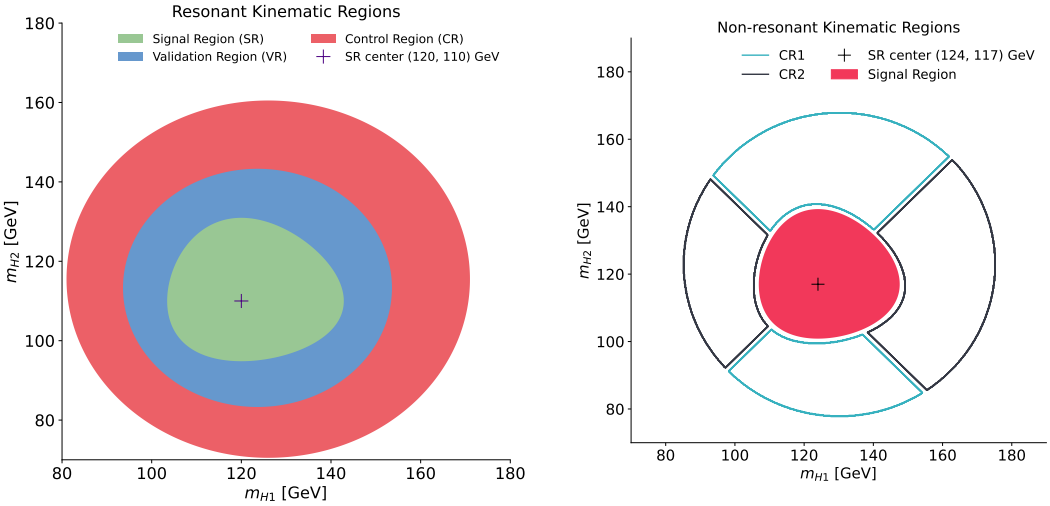
Figure 7.7.: Distribution of top-veto X_{Wt} variable used in the resonant analysis for $t\bar{t}$ and scalar signal samples. Vetoing events with $X_{Wt} < 1.5$ can reduce the contamination from $t\bar{t}$ events while maintaining a high signal efficiency, however low resonance mass signals can be significantly impacted by this veto. After the top veto, the remaining fraction of $t\bar{t}$ events in 2017 data is 5.39 % [97].



(a) 4b **(b)** 3b1l

Figure 7.8.: Distribution of top-veto X_{Wt} variable used in the non-resonant analysis for $t\bar{t}$ and both the $4b$ and $3b1l$ signal region categories in the standard model signal sample. In the non-resonant analysis, the b-jet coming from the t decay must explicitly be b-tagged. The top veto reduces the $t\bar{t}$ contamination in 2017 data from 17.6 % to 9.5 % in the $4b$ category and from 39.6 % to 23.5 % in the $3b1l$ category.

7.7. Kinematic regions



(a) Resonant kinematic regions **(b)** Non-resonant kinematic regions

Figure 7.9.: Kinematic regions for $HH \to b\bar{b}b\bar{b}$ searches

One of the defining aspects of the analysis is the formation of the *Higgs candidate massplane*, m_{H1}-m_{H2}. The construction of this plane is affected by the pairing method in particular. In this massplane, a Signal Region (SR) is defined to maximize the sensitivity of the analysis. The form of this region is taken from the early Run-2 analysis and is parametrized by the X_{HH} variable. The definition for this variable is shown in Equation 7.7, where c_1 and c_2 represents the centre of the SR. For the resonant analysis, the centre of the SR was (120, 110) GeV and the boundary of the SR was defined by $X_{HH} < 1.6$. In the non-resonant analysis the centre of the SR was shifted to (124, 117) GeV as this was found to be more optimal for the *min*-ΔR pairing method. To complement the SR, the Validation Region (VR) and Control Region (CR) are defined in Equations 7.8 and 7.9. These two regions take on elliptical shapes. The CR is used to derive a background estimate that is void of any potential signal contamination. The VR is used to derive an alternate background estimate that is used to estimate an extrapolation uncertainty. An example of the kinematic regions used in the resonant analysis is shown in Figure 7.9a.

The kinematic regions as defined in the resonant analysis suffer from significant signal contamination in the VR. Signal contamination has the potential to diminish sensitivity in the analysis as this information would be learned by the background model, creating a bias in the search. On the other hand, the fact that the VR is closer to the SR leads

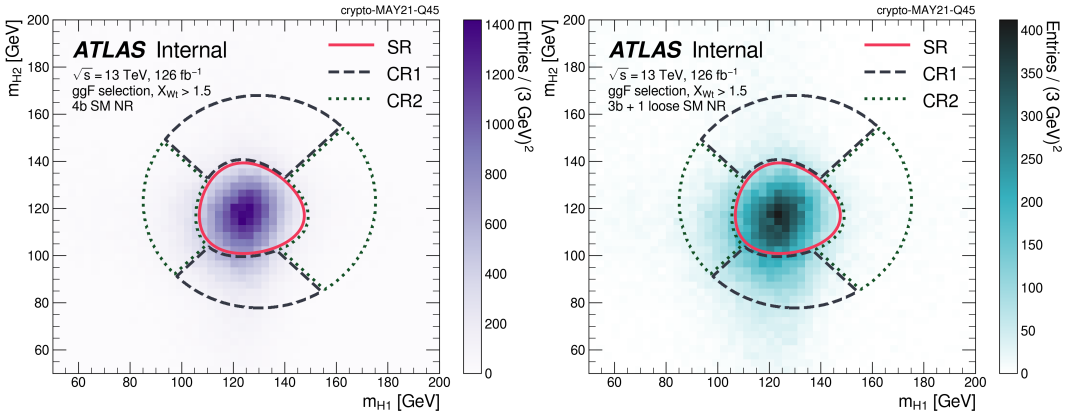
to a better estimate of the SR distribution. This is problematic if the VR is to be used to assign an extrapolation systematic uncertainty, since it is likely to introduce a bias. Since the resonant signals are narrow peaks, this is not a concern in the resonant analysis. In the non-resonant search this problem is alleviated by splitting the kinematic regions into quadrants as shown in Figure 7.9b. The overall boundary of this kinematic region is still constrained by the $CR_{HH} < 45$ GeV. The new regions CR1 and CR2 are used in the same way as the CR and VR are used in the resonant analysis. The two advantages of these new regions is that the signal contamination is significantly diluted and equally spread between CR1 and CR2. This eventually leads to a more accurate extrapolation uncertainty on the background estimate, which is the purpose of these regions.

$$X_{HH} = \sqrt{\left(\frac{10\,(m_{H1} - c_1\,)}{m_{H1}}\right)^2 + \left(\frac{10\,(m_{H2} - c_2\,)}{m_{H2}}\right)^2} \tag{7.7}$$

$$VR_{HH} = \sqrt{(m_{H1} - 1.03 \times c_1\,)^2 + (m_{H2} - 1.03 \times c_2\,)^2} < 30\,\text{GeV} \tag{7.8}$$

$$CR_{HH} = \sqrt{(m_{H1} - 1.05 \times c_1\,)^2 + (m_{H2} - 1.05 \times c_2\,)^2} < 45\,\text{GeV} \tag{7.9}$$

For the non-resonant search, the shape and size of the signal region were extensively studied. This was motivated by the fact that the analysis has significant deviations from the Early Run-2 analysis such as a different jet pairing method and b-tagging working points. Various elliptical boundaries were studied as a new signal region since they capture more of the simulated signals . In addition enlarged X_{HH} rings were also studied for potential improvements. These different shapes degraded the performance of the expected limits and the background model. The increase in the expected limits is attributed to the fact the background yields increased more rapidly than the increase in signal yield with these different signal regions. Ultimately, the results of these studies indicated that a signal region defined by $X_{HH} < 1.6$ was still the optimal choice as it yielded the best expected limits and background modelling. Figure 7.10 shows how the non-resonant signal samples in both $4b$ and $3b1l$ categories are captured by signal region.



(a) 4b **(b)** 3b1l

Figure 7.10.: Non-resonant massplane of the SM HH signal sample with both $4b$ and $3b1l$ categories.

7.8. Background estimation primer

The two main sources of background in this search are the QCD background and the $t\bar{t}$ backgrounds. The QCD background is roughly 95 % of the total background and $t\bar{t}$ is the remaining 5 %. Other backgrounds such as single Higgs Boson processes were found to be negligible. In the Early Run-2 analysis the two main backgrounds were modelled separately, however these are now modelled together through a data-driven approach. The data-driven model relies on the kinematic reweighting method, which allows us to derive a mapping between events that are kinematically similar to the background outside of the SR. This procedure is commonly referred to as reweighting and is parametrized as a deep neural network. This mapping can then be used to extrapolate into the SR to obtain the background estimate. The extrapolation uncertainty is evaluated by utilizing an alternative region such as the VR (or CR2 in the case of the non-resonant). As the the $2b$ (two-tag) region is kinematically similar to the $4b$ signal region and is abundant in statistics, this region is used as the starting point to model the $4b$ (four-tag) region. Additionally, in the non-resonant search the $3b1l$ and $3b1f$ are used as additional targets to model. The region created by reversing the $|\Delta\eta_{HH}|$ cut to $|\Delta\eta_{HH}| > 1.5$ and $3b1f$ are used as validations regions of the reweighting procedure. Validations of the neural network reweighting were also performed using simulated QCD and $t\bar{t}$ samples and were found to have adequate closure. Further details about the background estimation and validation procedures are reserved for Chapter 8.

7.9. Discriminating variable

The discriminating variable used in the resonant analysis is the *corrected* dihiggs mass, m_{HH}^{cor}. The variable is determined by re-scaling the higgs candidate four vectors so that the mass of the Higgs is reconstructed to be $m_H = 125$ GeV. The sum of the rescaled four vectors is then the m_{HH}^{cor}. The purpose behind this rescaling is to sharpen the m_{HH} peak and it correctly centres them on the resonance mass of the signal. The effect of this rescaling is illustrated in Figure 7.11 and the distributions indicate that the correction on m_{HH} improves the resolution, which enhances the sensitivity of the resonant search. On average the correction improves the mass resolution by 40 %, which is indicated by the percentage improvements in Table 7.2. Since the non-resonant signal is broad in m_{HH}, this correction is not used in the non-resonant analysis.

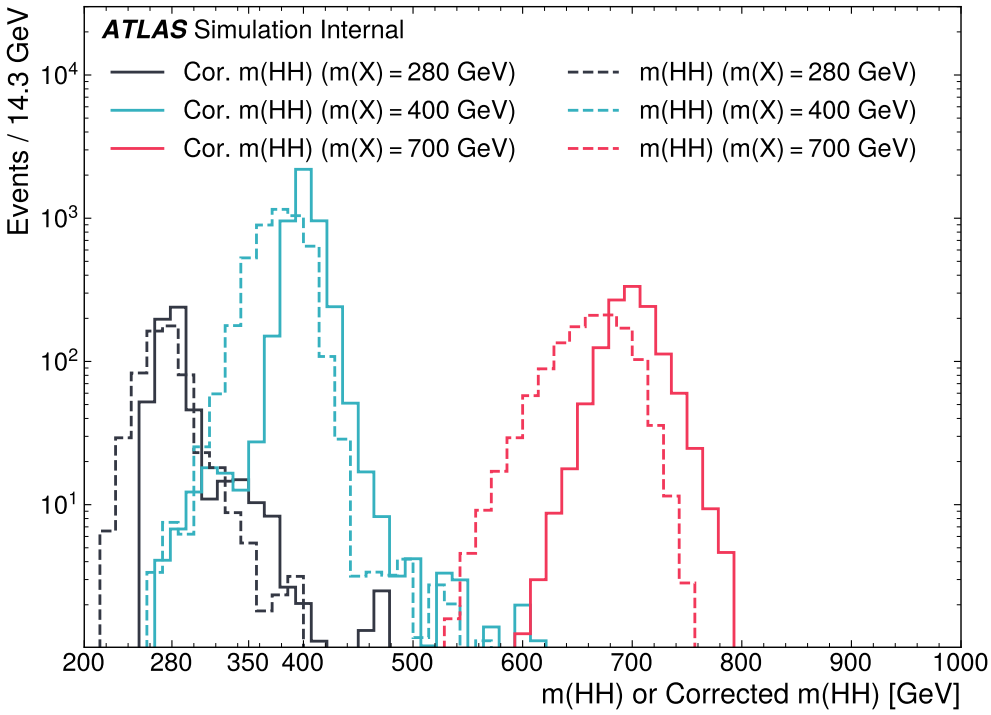


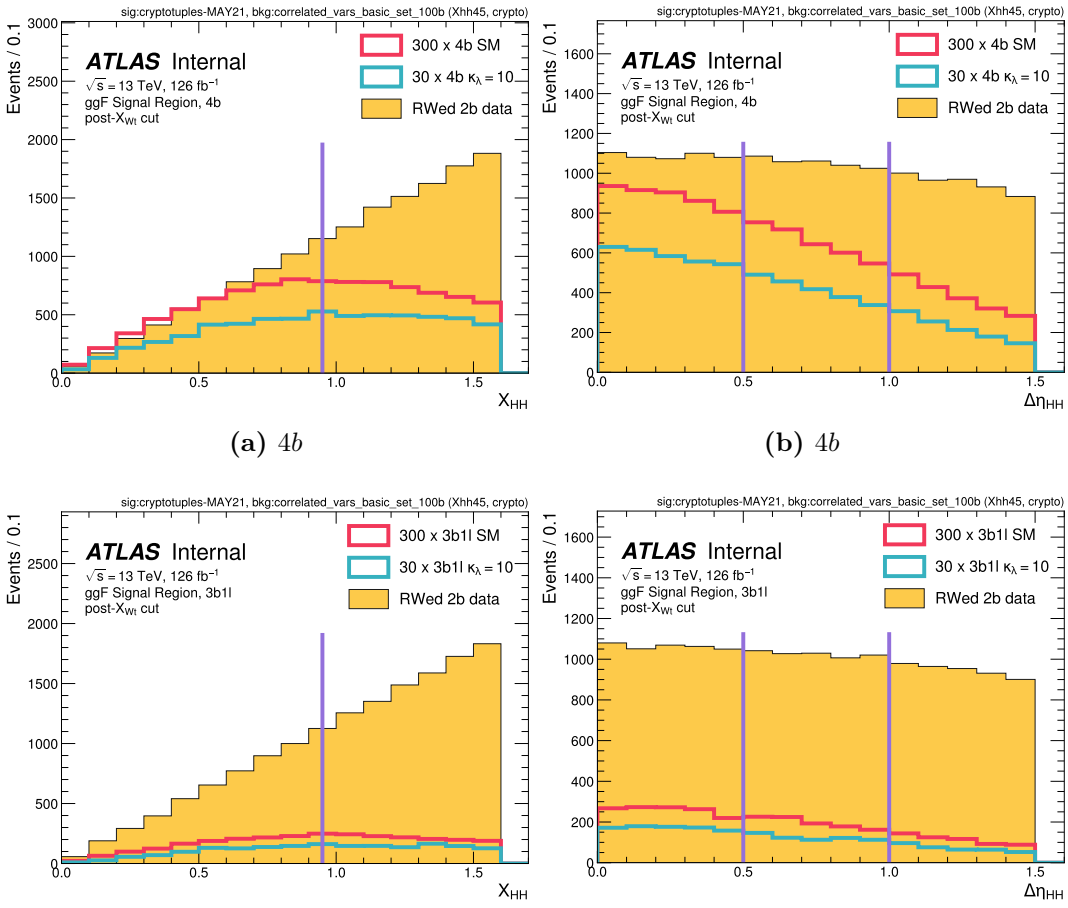
Figure 7.11.: Comparison between m_{HH} and the corrected m_{HH}^{cor} for scalar signal samples.

Scalar mass [GeV]	mass resolution uncorrected [GeV]	mass resolution corrected [GeV]	Difference (%)
280	23.8	14.5	39.0
400	31.0	17.4	43.8
700	45.8	28.0	38.8

Table 7.2.: Comparison of mass resolution between m_{HH} and m_{HH}^{cor} distributions. The percentage difference shows the increase in mass resolution gained by the corrected distribution relative to the uncorrected distribution.

7.9.1. Non-resonant categorization

In an ideal world, the power of discriminative MVA techniques like a BDT or a NN would be employed to improve the discrimination power. In fact, this is quite a common procedure in analysis where the signal and backgrounds are reliably produced by MC simulations. Since the non resonant search uses an entirely data-driven background estimate, the usage of an MVA technique affect the validation of the modelling of the input variables correlations. An intermediate solutions is employed by creating orthogonal categories which leads to high and low $S/\sqrt{B}$ purity bins instead. The variables X_{HH} and $|\Delta\eta_{HH}|$ are used to create a a total of six categories. There are two categories in X_{HH}: $X_{HH} \leq 0.95$ and $0.95 < X_{HH} \leq 1.6$, with the lower value corresponding to a higher purity since its closer to the SR centre. The $|\Delta\eta_{HH}|$ variable is used to create three further categories: $0 \leq \Delta\eta_{HH} < 0.5$, $0.5 \leq \Delta\eta_{HH} < 1.0$ and $1.0 \leq \Delta\eta_{HH} < 1.5$.



(a) $4b$ **(b)** $4b$

(c) $3b1l$ **(d)** $3b1l$

Figure 7.12.: Non-resonant signal region categories for both $4b$ and $3b1l$. A total of 6 categories are formed by splitting X_{HH} into 2 $X_{HH} \leq 0.95$ and $0.95 < X_{HH} \leq 1.6$ and $|\Delta\eta_{HH}|$ into 3 $0 \leq \Delta\eta_{HH} < 0.5$, $0.5 \leq \Delta\eta_{HH} < 1.0$ and $1.0 \leq \Delta\eta_{HH} < 1.5$. The low X_{HH} and low $|\Delta\eta_{HH}|$ categories create regions with the highest signal purity.

7.10. Analysis cutflows

This chapter has highlighted the main event selections and reconstruction methods for resonant and non-resonant searches. Table 7.3 shows the event yield of each subsequent step used in the non-resonant search. Similarly, Table 7.4 shows the event yield of each subsequent step used in the resonant search.

Selection	**Data**		**SM HH**	**HH $\kappa_\lambda = 10$**
	4b	***2b***	***4b***	***4b***
Initial (unweighted for MC)	1.59×10^{10}	1.59×10^{10}	4.6×10^{6}	4.81×10^{7}
Preselection	5.7×10^{8}	5.7×10^{8}	523	7.27×10^{3}
Trigger	2.81×10^{8}	2.81×10^{8}	471	6.85×10^{3}
Trigger bucket	2.49×10^{8}	2.49×10^{8}	416	6.31×10^{3}
Not VBF channel	2.46×10^{8}	2.46×10^{8}	410	6.22×10^{3}
At least 4 jets	1.81×10^{8}	1.81×10^{8}	351	5.21×10^{3}
Specified number of b-tagged jets	1.89×10^{6}	1.58×10^{8}	96.2	1.16×10^{3}
$\lvert\Delta\eta_{HH}\rvert < 1.5$	1.01×10^{6}	8.27×10^{7}	80.1	986
Signal region	2.25×10^{4}	1.77×10^{6}	35.2	237
Control region 1	4.42×10^{4}	3.49×10^{6}	15.2	129
Control region 2	4.38×10^{4}	3.46×10^{6}	9.56	104

Table 7.3.: Event cutflow for the non-resonant search. The cutflow shows the event yield for each subsequent selection step of the non-resonant search, shown for the $4b$ and $2b$ data as well as the SM HH and $\kappa_\lambda = 10$ signal samples.

Selection	Data		Scalar signal			Graviton signal		
	4b	*2b*	*280 GeV*	*400 GeV*	*600 GeV*	*300 GeV*	*500 GeV*	*700 GeV*
Initial (Unweighted for MC)	15 905 923 744	15 905 923 744	240 566 598	131 193 650	10 847 624	305 000	295 000	155 000
Reweighted to 126 fb^{-1} × gen. x-sec × branching ratio			1.317×10^7	1.437×10^7	2.438×10^6	5.629×10^4	3.807×10^4	8.602×10^3
Pass preselection	356 647 420	356 647 420	1.476×10^6	4.262×10^6	1.408×10^6	1.275×10^4	2.245×10^4	6.666×10^3
Pass trigger	302 389 262	302 839 262	1.457×10^6	4.197×10^6	1.381×10^6	1.231×10^4	2.181×10^4	6.327×10^3
At least four jets with $p_T \geq 40$ GeV, $\lvert\eta\rvert < 2.5$	212 331 447	212 331 447	1.107×10^6	3.280×10^6	1.125×10^6	1.071×10^4	1.787×10^3	5.391×10^3
Four (two for background) b-tagged jets	2 117 122	186 163 666	2.177×10^5	6.758×10^5	2.912×10^5	1.813×10^3	4.628×10^3	1.586×10^3
Trigger bucket cut	1 898 018	162 701 483	2.164×10^5	6.435×10^5	2.647×10^5	1.714×10^3	4.138×10^3	1.399×10^3
$\Delta\eta_{HH} < 1.5$	1 018 025	84 038 338	1.633×10^5	5.366×10^5	2.184×10^5	1.476×10^3	3.870×10^3	1.302×10^3
Top veto	747 013	70 000 679	7.768×10^4	4.170×10^5	1.989×10^5	8.986×10^2	3.384×10^3	1.205×10^3
Signal region	39 196	3 094 095	2.755×10^4	1.996×10^5	1.015×10^5	2.955×10^2	1.627×10^3	6.486×10^2
Validation region	54 481	4 241 770	2.521×10^4	1.105×10^5	5.189×10^4	2.540×10^2	8.611×10^2	2.980×10^2
Control region	93 667	7 176 867	1.573×10^4	5.337×10^4	2.322×10^4	1.505×10^2	3.700×10^2	1.126×10^2

Table 7.4.: Event cutflow for the resonant search. The cutflow shows the event yield for each subsequent selection step of the resonant search. The yields shown are for the $4b$ and $2b$ data and 3 mass points of the scalar and graviton signal samples.

Chapter 8.

Background Estimation

"I have not failed. I've just found 10,000 ways that won't work."
— Thomas A. Edison

The main aspects of the background estimation procedure of the $HH \to b\bar{b}b\bar{b}$ analysis were briefly introduced previously. This chapter is dedicated into delving deeper into the reweighting procedure and the validation efforts.

8.1. Kinematic reweighting model

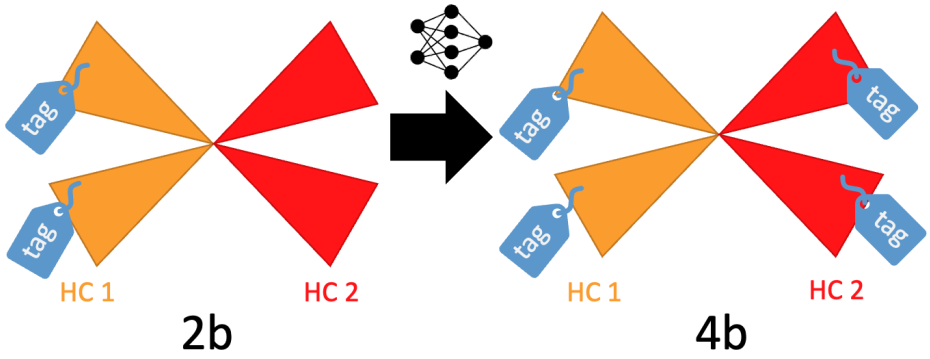


Figure 8.1.: Kinematic Reweighting of $2b$ to $4b$ events by using a neural network to act as a transfer function.

To reweight the $2b$ region into the $4b$ region machine learning binary classification models can be leveraged, such as but not limited to decision trees and neural networks.

This reweighting idea is depicted in Figure 8.1. This is because a classifier will estimate the probability densities of each target, $p_{2b}(\boldsymbol{x})$ and $p_{4b}(\boldsymbol{x})$, for the given input feature space vector $\boldsymbol{x}$. By taking the ratio of these two probability densities, a set of weights, $w(\boldsymbol{x})$, can be derived that is applied for each event, as is shown in Equation 8.1. The use of BDT to perform the reweighting by using the GBReweighter [98] and XGBoost was studied and it was found to perform worse than a neural network model. This is likely due to the fact that a neural network model is able to correctly treat all of the correlations between variables and better model local variations.

$$w(\boldsymbol{x}) = \frac{p_{4b}(\boldsymbol{x})}{p_{2b}(\boldsymbol{x})} \tag{8.1}$$

Reweighting falls into the domain of density ratio estimation, which is a well-understood area. A classifier could be used to solve for $w(\boldsymbol{x})$ by estimating the two density ratios, this however is solving a redundant intermediate problem. The weights can be estimated directly by considering the loss function in Equation 8.2, where $R(\boldsymbol{x})$ is the output of the estimator (the neural network in this case). $\mathbb{E}_{\boldsymbol{x}\sim p_{2b}}$ and $\mathbb{E}_{\boldsymbol{x}\sim p_{4b}}$ are the expectation values of the $2b$ and $4b$ probability densities respectively.

$$\mathcal{L}(R(\boldsymbol{x})) = \mathbb{E}_{\boldsymbol{x}\sim p_{2b}}\left[\sqrt{R(\boldsymbol{x})}\right] + \mathbb{E}_{\boldsymbol{x}\sim p_{4b}}\left[\frac{1}{\sqrt{R(\boldsymbol{x})}}\right] \tag{8.2}$$

The training procedure of a neural network will ensure that the estimator $R(\boldsymbol{x})$ minimizes the loss function, $\mathcal{L}(R(\boldsymbol{x}))$. This procedure is equivalent to estimating the ratio of the two densities directly as shown in Equation 8.3.

$$\arg\min_{R} \mathcal{L}(R(\boldsymbol{x})) = \frac{p_{4b}(\boldsymbol{x})}{p_{2b}(\boldsymbol{x})} \tag{8.3}$$

The substitution $Q(\boldsymbol{x}) \equiv logR(\boldsymbol{x})$ is made in practice to avoid positivity constraints. The substitution changes the loss function to the form shown in Equation 8.4, which is solved by Equation 8.5.

$$\mathcal{L}(Q(\boldsymbol{x})) = \mathbb{E}_{\boldsymbol{x}\sim p_{2b}}\left[\sqrt{e^{Q(\boldsymbol{x})}}\right] + \mathbb{E}_{\boldsymbol{x}\sim p_{4b}}\left[\frac{1}{\sqrt{e^{Q(\boldsymbol{x})}}}\right] \tag{8.4}$$

$$\arg\min_{Q} \mathcal{L}(Q(\boldsymbol{x})) = \log \frac{p_{4b}(\boldsymbol{x})}{p_{2b}(\boldsymbol{x})} \tag{8.5}$$

The optimal reweighting function is then obtained by taking the exponent of the network's output. Thus this loss function allows the density ratio estimation problem to be solved directly, with no positivity constraints. This has shown to yield comparable or improved reweighting relative to the binary categorical cross-entropy approach.

8.1.1. Neural network reweighting

Both the resonant and non-resonant searches use the same neural network based reweighting approach to model the background, with minor differences to the architecture. This section describes the details of the neural network training.

The input features used in the neural network reweighting are described in Table 8.1. The features are almost identical between the non-resonant and resonant search except for two. The first is the top veto that was changed to be limited to using only b-tagged jets in the non-resonant search. The second is that the trigger bucket category was included into the feature space of the non-resonant search as it was shown to improve the modelling in the boundary region between the two buckets. Since the trigger bucket is a categorical feature with no natural order it is processed as a one-hot encoded vector [99]. One-hot encoding transforms a categorical feature of n-categories and creates a new feature for each class. The new features are filled with 0 if they do not fall into that class and 1 if they do belong[1]. Each numerical feature is standardized by subtracting the mean and scaling to unit variance to avoid a single feature inflating the weights. To perform this transformation, the *standard scaler* of *scikit* is used [100]. Some distributions are also transformed by a log function to simply make them smoother. The NN model is built using the *Keras* library [101] with the *TensorFlow* backend [102]. The reweighting is performed on each year of data separately, rather than reweighting all the data together,

[1]Example with 2 categories 'a' and 'b': $[a] \to [1, 0]$ and $[b] \to [0, 1]$

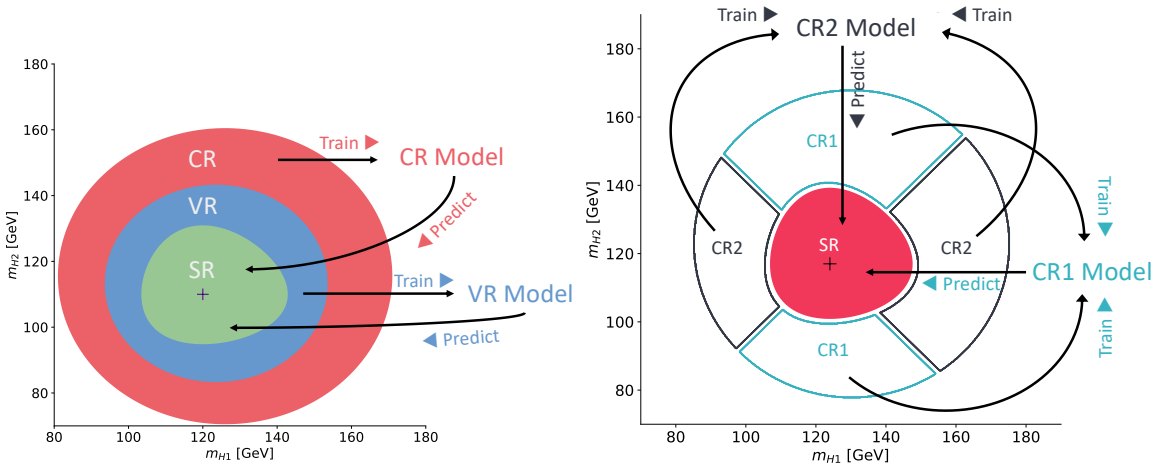
as this was shown to perform better. The improved performance is likely due to the fact that each year brings slightly different run conditions due to the triggers, pile-up and b-tagging performance.

Variable	Description
$\log(p_{T_4})$	$\log(p_T)$ of the 4th leading Higgs candidate jet
$\log(p_{T_2})$	$\log(p_T)$ of the 2nd leading Higgs candidate jet
$\log(\Delta R_{jj}$ Close)	$\log(\Delta R)$ between the closest two Higgs candidate jets
$\log(\Delta R_{jj}$ Not Close)	$\log(\Delta R)$ between the two other Higgs candidate jets
$\langle \lvert HC\eta \rvert \rangle$	average absolute value of the Higgs candidate jet pseudorapidity
$\log(p_{T_{HH}})$	$\log(p_T)$ of the dihiggs system
ΔR_{HH}	ΔR between the two Higgs candidates
$\Delta\phi_{H1}$	$\Delta\phi$ between the jets in the leading Higgs candidate
$\Delta\phi_{H2}$	$\Delta\phi$ between the jets in the subleading Higgs candidate
$log(X_{Wt})$	log of the top veto variable
njets	number of jets in the event
bucket index	trigger bucket category

Table 8.1.: Features of the dihiggs system used in the neural network reweighting model. The bucket index feature is only used in the non-resonant search and is transformed via a one-hot encoding scheme.

A key aspect of the reweighting model is that it must be able to learn the mapping between $2b$ and $4b$ events outside of the signal region, since the $4b$ events in this region are blinded. The reweighting model then needs to be able to extrapolate the mapping into the signal region. To illustrate this the kinematic regions used in the resonant analysis are shown in Figure 8.2a. In this Figure there are two neural network reweighting models trained, one in the control region and one in the validation region. As the control region is the furthest from the signal region and suffers the least from any potential signal contamination, it is used as the nominal background estimate. The validation region is used to provide an alternative estimate that can be used to provide a systematic uncertainty on this extrapolation of the reweighting to the SR. Each of the CR-derived and VR-derived weights, applied to the $2b$ events in the SR, give two different predictions of the $4b$ background model in the SR. The symmetrized difference between the CR and VR models in the signal region is used as the shape systematic uncertainty and is the most important uncertainty of the entire search. As mentioned in Chapter 7, the kinematic regions of the non-resonant search were changed to using quadrants in order

to dilute the signal contamination in the validation region. In this case, the CR1 region is used for the nominal estimate and the CR2 model is used as the alternative model as shown in Figure 8.2b.



(a) resonant reweighting **(b)** non-resonant reweighting

Figure 8.2.: Kinematic regions used to derive the reweighting model in both resonant and non-resonant searches.

To understand how the features impact the reweighting model, the Wasserstein distance metric is used [103]. This metric measures the *cost* required to turn a distribution into another and is also sometimes known as the Earth-Mover's Distance (EMD). This metric is used to compare the level of disagreement between the $2b$ and $4b$ distributions in each of the features used. This is as a proxy for how important each feature might be. Figure 8.3, shows the Wasserstein distance between $2b$ and $4b$ events in each of the variables used for reweighting. The number of jets in the event (njets) seems to be the distribution with the most disagreement to start with. This can then be compared with the distribution of SHAP values shown in Figure 8.4, which shows how the feature impacts the output of the reweighting model [47]. They are also ranked in order of feature importance determined by the mean absolute value of the SHAP values. The njets distribution seems to have the most impact in the model, but other variables with lower Wasserstein distance such as ΔR_{jj} Not Close can still have a significant impact.

8.1.2. Training hyperparameters

The optimal network architecture and hyperparameters were chosen as a result of an optimization study that minimized the area created by the envelope of the shape

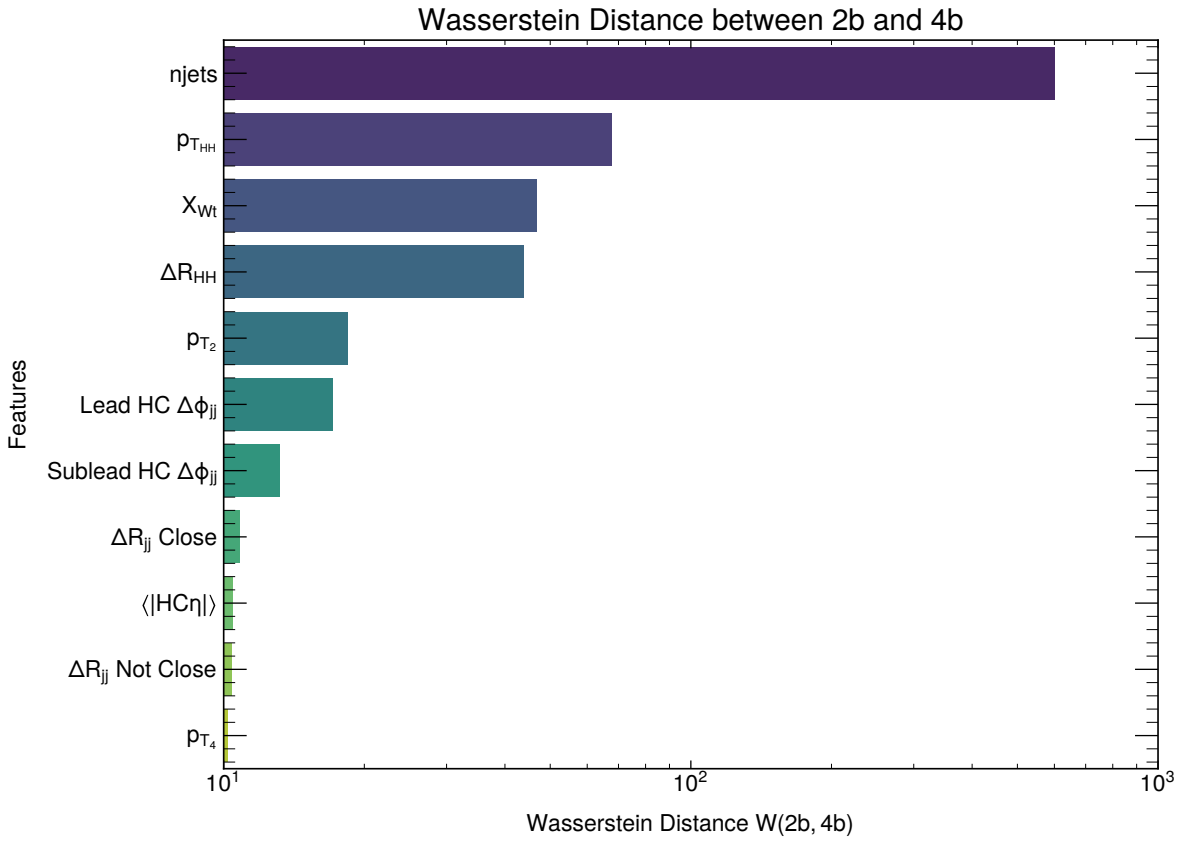


Figure 8.3.: Resonant search Wasserstein distance, $W(2b, 4b)$, between $2b$ and $4b$ events in the control region. Metric indicates the level of discrepancy in each feature of the $2b$ and $4b$ systems.

uncertainty. The shape uncertainty is the extrapolation uncertainty used when the model is inferred in the signal region and is discussed in Section 8.2.3. The network has three hidden layers that are 50 units wide with ReLU activation functions. The output layer has a single unit with a linear activation function. To avoid overfitting during training the dataset is split into two subsets with a 60:40 split. The larger subset is used to train the network while the smaller subset is used to test the network performance on unseen data. The loss function of the test dataset is monitored to stop the training procedure at the optimal point before any overfitting occurs.

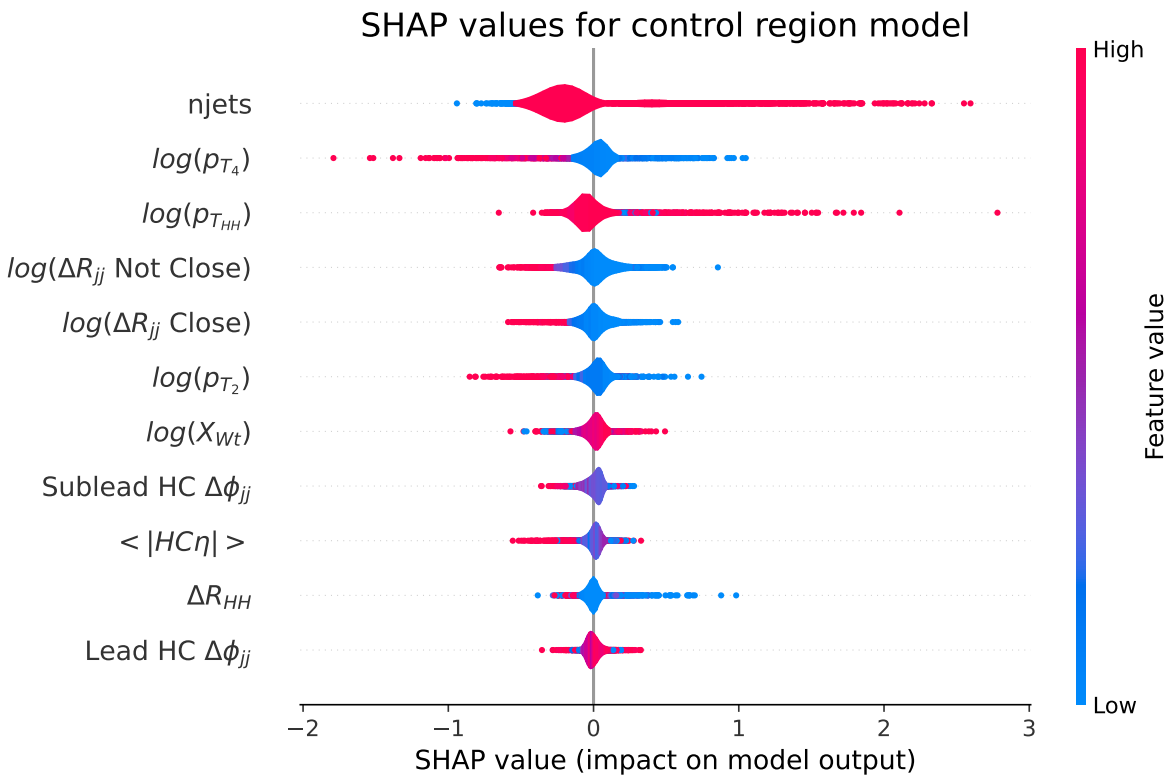


Figure 8.4.: Beeswarm distribution of SHAP values for each feature used in the reweighting model. The features are ranked in descending order of importance. The importance of a feature is measured by the mean absolute value of the SHAP value. These values were calculated using the *DeepExplainer* method of python SHAP package [46, 104].

8.2. Uncertainties

8.2.1. Poisson uncertainties

The standard Poisson-based statistical uncertainties are calculated for both $2b$ and $4b$ distributions. Since $4b$ events do not have weights associated to them their histogram level uncertainties can be determined by the standard $\delta n^i_{4b} = \sqrt{n^i_{4b}}$, where n^i_{4b}, represents the number of $4b$ events in the i-th bin of the histogram. If this approach was applied to the histogram of $2b$ events after reweighting, the statistical uncertainties would be inaccurately estimated. In fact, in this case the uncertainties would be overestimated since there are roughly 100 times more $2b$ events than $4b$ events. To correct for this the weights of the reweighting are taken into account to calculate the Poisson uncertainty of a weighted histogram. This is shown in Equation 8.6, where $n^i_{2b(RW)}$ is the number of $2b$ events after reweighting is applied in the i-th bin of the histogram and $w^j_{2b(RW)}$ are the

weights of events j that fall within the bounds of the i-th histogram bin. If each of these weights was 1, then Equation 8.6 would boil down to be $\delta n^i_{2b(RW)} = \sqrt{n^i_{2b(RW)}}$, which is of the same form as the $4b$ Poisson uncertainty. Since the statistics of $2b$ events is in general very generous, the Poisson uncertainty of the $2b$ (RW) histogram are insignificant except in the tail regions of the distributions.

$$\delta n^i_{2b(RW)} = \sqrt{\sum_{j \in i} \left(w^j_{2b(RW)} \right)^2} \tag{8.6}$$

8.2.2. Repeating the experiment: bootstraps

Previously, the Poisson statistical uncertainty of the $2b$ and $4b$ events was described. This, however, does not account for the statistical uncertainty in the reweighting model due to finite statistics of $4b$ events and the randomized initialization of the NN parameters. In the reweighting model, it is always the $4b$ statistics that are the limiting factor. To estimate this uncertainty, a bootstrap resampling technique is used to create different, but statistically equivalent, *datasets* from the original dataset. These new datasets are created by *sampling with replacement* from the original dataset. Each of these new datasets are then used to retrain the neural network and from all of these neural networks a variance on the output can be established.

The procedure of sampling with replacement is equivalent to drawing from a multinomial distribution. If datasets are large then this procedure can become computationally expensive and it is typically preferable to result to using the *Poisson Bootstrap* technique. This technique effectively weights the training dataset with weights drawn from a Poisson distribution with $\lambda = 1$ to create a new dataset. In the context of the reweighting model this means that all of the $2b$ and $4b$ events are used in the reweighting but each event will receive a different weight for each bootstrap. The reason why this Poisson approximation can be used is because in the limit of a large dataset with size n the Binomal distribution approximates a Poisson distribution as shown in Equation 8.7. This tells us that the procedure of sampling with replacement is approximately equal to weighting the dataset by an integer weight drawn from a Poisson distribution with $\lambda = 1$. Moreover, this bootstrap procedure also coincidentally allows for a more accurate nominal estimate from the ensemble as the reweighting model has a significant variance.

$$\lim_{n\to\infty} = Binomial(n, 1/n) = Poisson(\lambda = 1) \tag{8.7}$$

These bootstrap trainings yield a variation in the reweighting procedure which is used to assign an uncertainty at histogram level. This is then used as an additional uncertainty in the statistical analysis. The reweighting procedure is repeated 100 times to produce 100 different bootstraps. The problem now is to convert these *event-level* variations to a *histogram-level*. Ideally, as each event has 100 different weights associated to it, you would produce 100 different histograms of each variable of interest. Then the nominal histogram can be taken to be the mean count for each bin and the uncertainty can be given by the standard deviation in the counts of each bin. In reality this approach is rather unfeasible as storing the 100 weights for each event is exhaustive of computational resources and since the variables are already histogrammed, this leaves little flexibility for analysis. To overcome these limitations a procedure to derive the nominal estimate and the bootstrap uncertainty is implemented.

The nominal estimate

The first part of deriving the nominal estimate histogram is to find out the normalization factor between $2b$ and $4b$ events. The ratio between $4b$ and $2b$ events, μ_{QCD}, is used as a normalization factor to construct the reweighted $2b$ histogram. Although the definition of this normalization is simple, the usage of Poisson and neural network weights makes it slightly more complicated. The normalization for each bootstrap is defined in Equation 8.8, where $W^{4b}_{p,j}$ and $W^{2b}_{p,i}$ are the Poisson weights for $4b$ and $2b$ events respectively. Similarly, $W^{2b}_{NN,i}$ are the weights the neural network predicted for the $2b$ event. In practice the value of μ_{QCD} is approximately 1, as the neural network is able to learn about the different class proportions and fold this information into the weights. It is also important to note that this normalization factor is always calculated in the kinematic region that the neural network was trained in.

$$\mu_{QCD} = \frac{N_{4b}}{N_{2b}} = \frac{\sum_{j=1}^{N_{4b}} W^{4b}_{p,j}}{\sum_{j=1}^{N_{2b}} W^{2b}_{p,i} \times W^{2b}_{NN,i}} \tag{8.8}$$

Now each event has 100 normalization factors, μ_{QCD}, and 100 weights W_i, which can be combined to derive a median weight $\overline{W}_i$ for each event as shown by Equation 8.9, where the subscript i is used to indicate that this operation is to be performed on each $2b$ event.

$$\overline{W}_i = median\left(\mu^1_{QCD} \cdot W^1_i, ..., \mu^{100}_{QCD} \cdot W^{100}_i\right) \tag{8.9}$$

By using the median $2b$ event weights, $\overline{W}_i$, the median normalization $\overline{\mu}_{QCD}$ can be derived by using Equation 8.8, without the addition of the Poisson weights. The nominal estimate is then obtained from these median weights, $\overline{W}_i$ and the median normalization $\overline{\mu}_{QCD}$.

Bootstrap variation

A variation histogram can then be constructed by using the Interquartile Range (IQR). In this case this is defined as the difference between the 75th and 25th percentiles. The IQR of the weights W^{IQR}_i and normalizations μ^{IQR}_{QCD} are defined in Equations 8.10 and 8.11 respectively.

$$W^{IQR}_i = IQR\left(\mu^1_{QCD} \cdot W^1_i, ..., \mu^{100}_{QCD} \cdot W^{100}_i\right) \tag{8.10}$$

$$\mu^{IQR}_{QCD} = IQR\left(\mu^1_{QCD}, ..., \mu^{100}_{QCD}\right) \tag{8.11}$$

To explain how the final histogram gets constructed consider the two objects $H(weights)$ and $Y(weights)$. The first object is a histogram that has been constructed with the given *weights* and similarly the second object is simply the event yield with the given weights. Note that Y can be determined by summing over the bin contents of a histogram that has been constructed using the same weights. The up varied histogram can be built using $W_{up} = \overline{W} + \frac{1}{2}W^{IQR}$, which then needs to be scaled by the event yield of the nominal histogram. Finally an additional up variation is added to encapsulate the variation of the

normalization. The full construction of the bootstrap variation histogram is summarized in Equation 8.12, and this is symmetrized to obtain the variation band. The usage of the median and IQR operations was needed to be more resilient to outlier values, which would explode the uncertainty band. Ultimately, the histogram level Poisson errors and bootstrap errors are added together in quadrature to represent the total statistical uncertainty for $2b$ events. This approach does not guarantee that the histogram variation yields similar results to creating a histogram for each of the 100 bootstraps. It has however been shown to be an empirically equivalent approximation in the context of the ggF $HH \to b\bar{b}b\bar{b}$ search.

$$H_{up} = H(\overline{W} + \frac{1}{2}W^{IQR}) \cdot \frac{Y(\overline{\mu}_{QCD} \cdot \overline{W})}{Y(\overline{W} + \frac{1}{2}W^{IQR})} + \frac{1}{2}\mu^{IQR}_{QCD} \cdot H(\overline{\mu}_{QCD}\overline{W}) \tag{8.12}$$

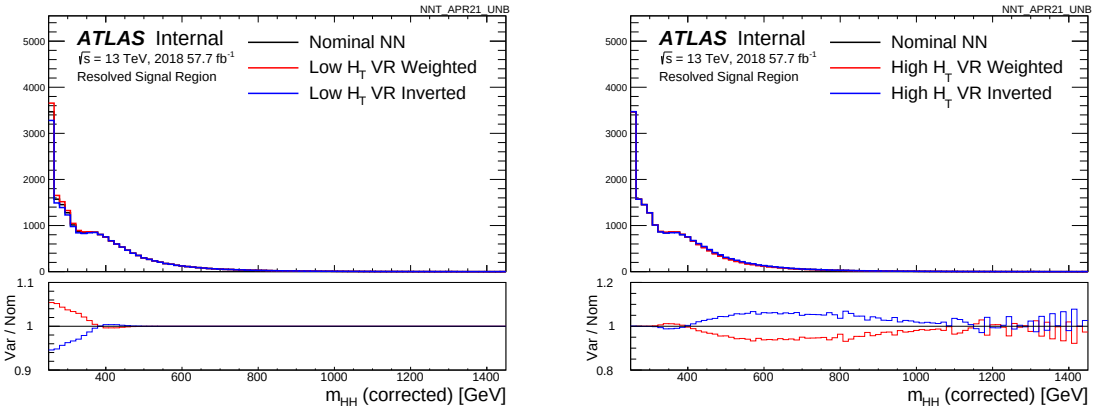
8.2.3. The shape uncertainty

The uncertainties previously discussed are statistical in nature and they do not capture possible biases arising from extrapolating the network into a region of phase space it was not trained on. The *shape systematic* is designed to encapsulate this and is essentially the most challenging part of the entire $4b$ search. This is the reason why an alternative model is trained on the validation region to compare with the nominal model in the signal region. The differences between the alternative and baseline model define the shape uncertainty in the variable of interest, such as m_{HH}. Then the uncertainty band is created by symmetrizing the difference from the baseline.

Resonant search H_T splitting

In the resonant analysis the uncertainty band is split into two independent variations in the m_{HH} spectrum. To do this the variable H_T is used which is determined by the scalar sum of the p_T of jets that make up the Higgs candidates. This variable is used as it is highly correlated with m_{HH} and it allows us to define essentially low mass and high mass uncertainties without creating a sharp discontinuity in the m_{HH} spectrum itself. The decision boundary is set at $H_T = 300$ GeV. Events with $H_T < 300\,\text{GeV}$ this value are put into the low-H_T variation and events above this are put in the high-H_T variation. This split based on H_T is designed to prevent low mass bins constraining the high mass bins and vice-versa and was also used in the Early Run-2 search. Alternative

methods were investigated, but this seemingly simple split yields nearly identical results. An example of the low and high H_T shape uncertainty using data from 2018 are shown in Figure 8.5. In Figure 8.5a, the low-H_T variation provides an uncertainty envelope for corrected m_{HH} values below 400 GeV. Similarly, Figure 8.5b provides an uncertainty band for the region $m_{HH} > 400\,\text{GeV}$.



(a) Low-H_T variation **(b)** High-H_T variation

Figure 8.5.: Resonant low and high H_T shape uncertainty

Non-resonant search quadrants splitting

One of the main differences between the two searches is that the non-resonant search splits the kinematic regions into quadrants, as shown in Figure 7.9b. Using H_T splitting in this case is possible, but it was found that the nuisance parameters are less constrained with this quadrant splitting. This split of the nuisance parameter is shown in Figure 8.6. The shape uncertainty obtained by using these quadrants definitions is shown in Figure 8.7.

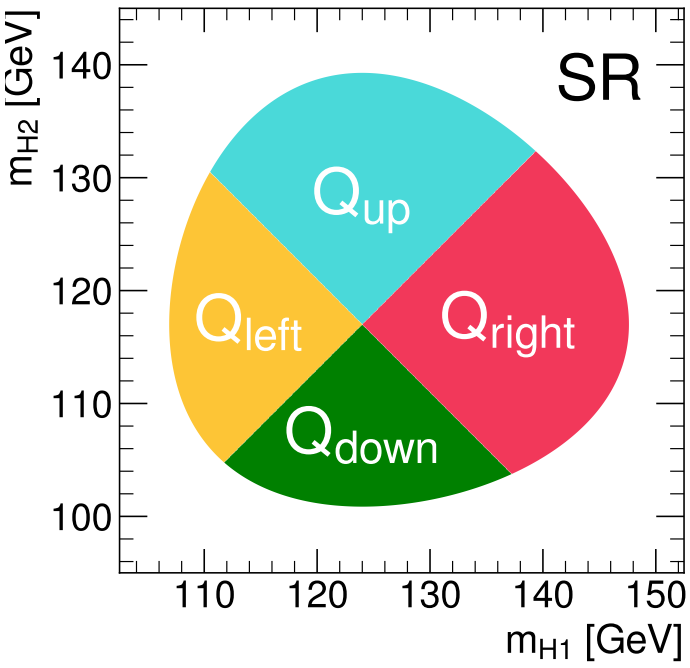
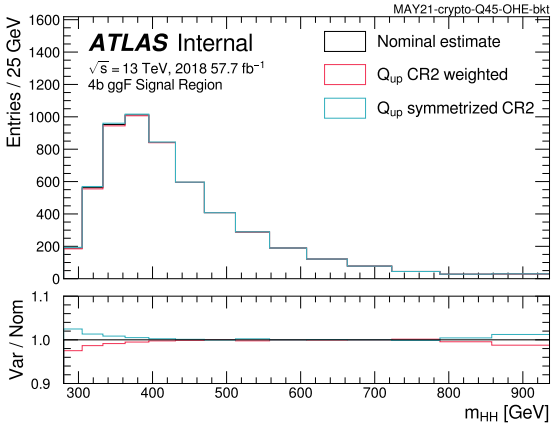
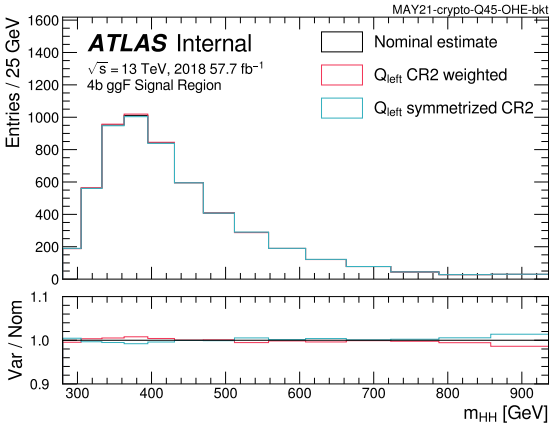


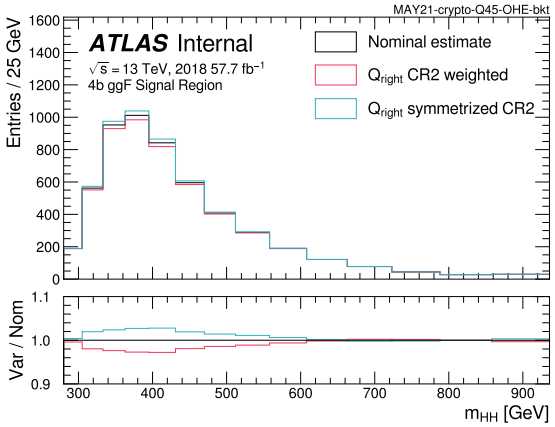
Figure 8.6.: Quadrants nuisance parameter split for non-resonant search.



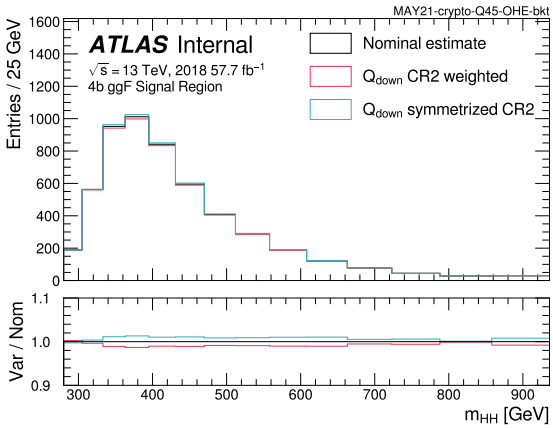
(a) Q_{up} variation



(b) Q_{left} variation



(c) Q_{right} variation

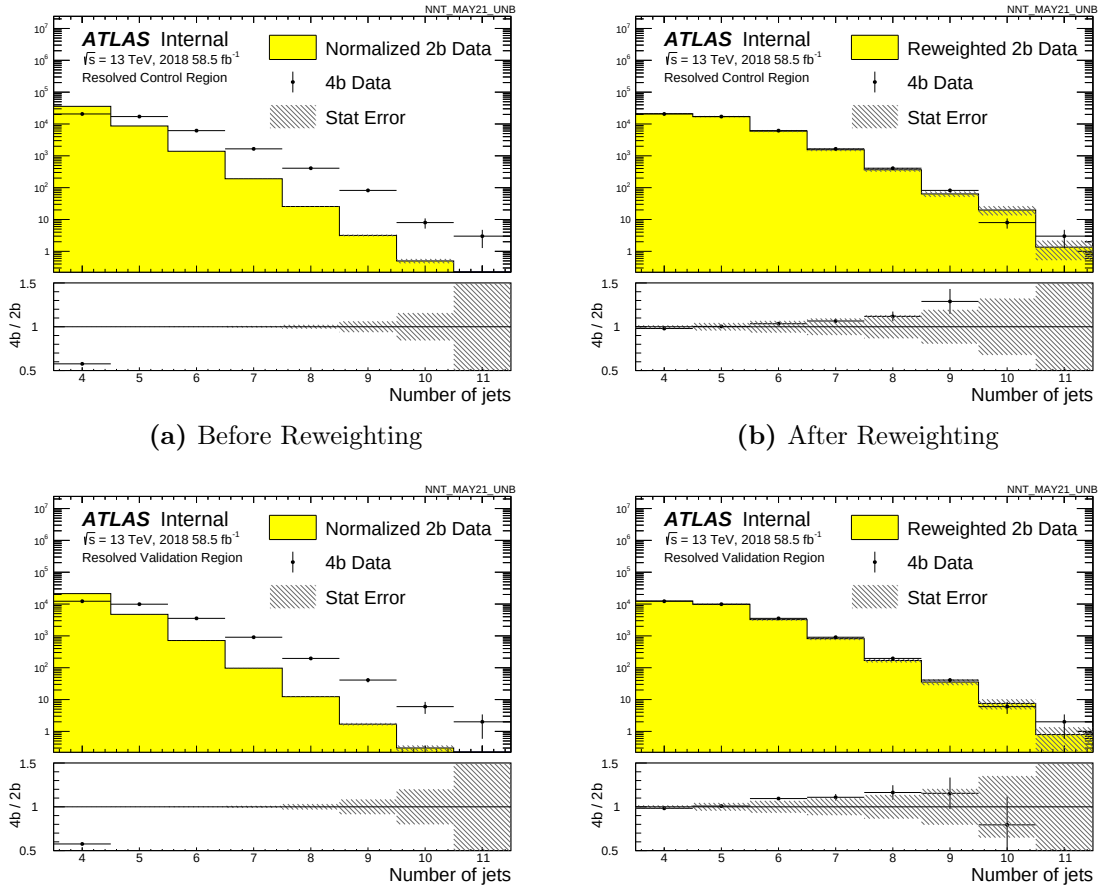


(d) Q_{down} variation

Figure 8.7.: Non-resonant shape uncertainty variations using massplane quadrants splitting.

8.3. Resonant reweighting

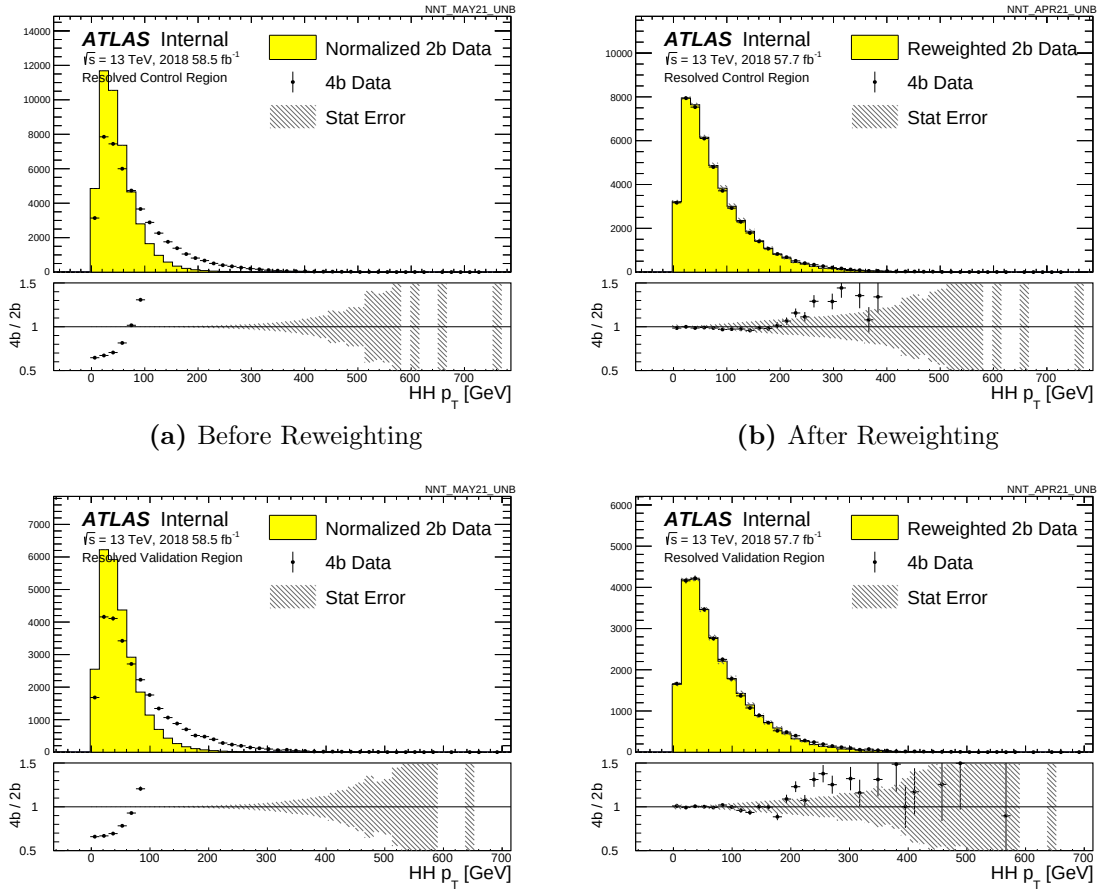
It is important to check how well the reweighting model performs in the region it has been trained on to see if the level of closure is adequate. Figures 8.8-8.12 show how the reweighting maps the $2b$ distributions into $4b$ distributions of a few key variables in both Control and Validation regions. The examples shown here are only for 2018 data as it has the highest integrated luminosity relative to 2016 and 2017. Since this is a closure check the uncertainty is only statistical and is aThe statistical uncertainty on the $2b$ distributions is made up of both the Poisson and bootstrap uncertainties added in quadrature. These figures show that the reweighting model can adequately map the $2b$ region into the $4b$ region within the level of statistical uncertainty. Some distributions, such as $p_{T_{HH}}$ show discrepancies between the reweighted data and the $4b$ target that are larger than the statistical uncertainty. In this case, this occurs in the tail end of the distribution that is very difficult for the network to model since there are only a few $2b$ events to model $4b$ events in this tail region, which is evidenced by Figures 8.9a and 8.9c. Only a handful of the distributions used to train the neural network are shown here for conciseness, the remaining distributions are shown in Appendix A.1.



(a) Before Reweighting **(b)** After Reweighting

(c) Before Reweighting **(d)** After Reweighting

Figure 8.8.: Resonant njets distributions before and after reweighting in both Control and Validation regions.



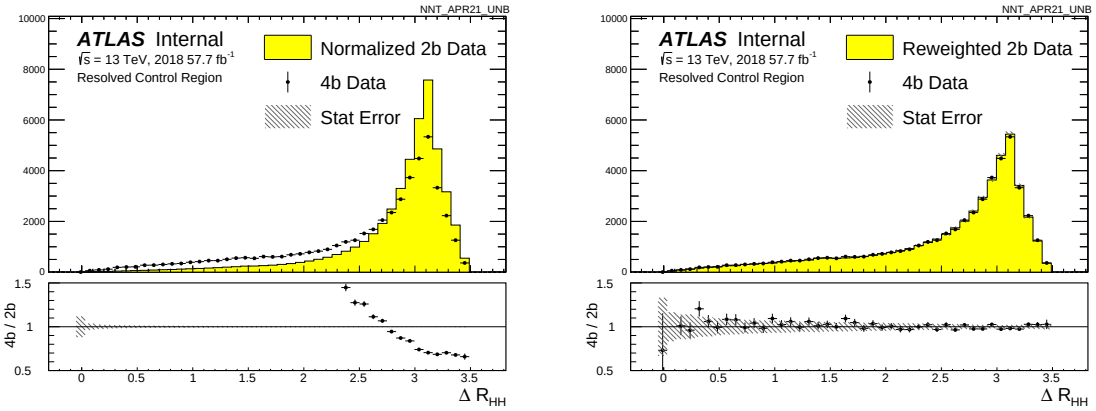
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

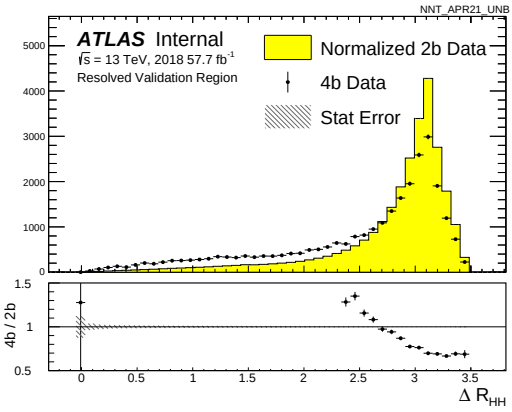
(d) After Reweighting

Figure 8.9.: Resonant $p_{T_{HH}}$ distributions before and after reweighting in both Control and Validation regions.

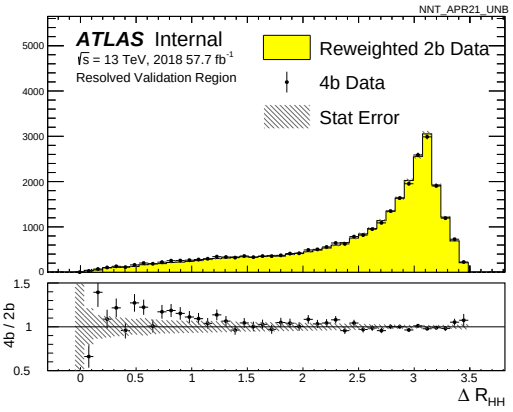


(a) Before Reweighting

(b) After Reweighting

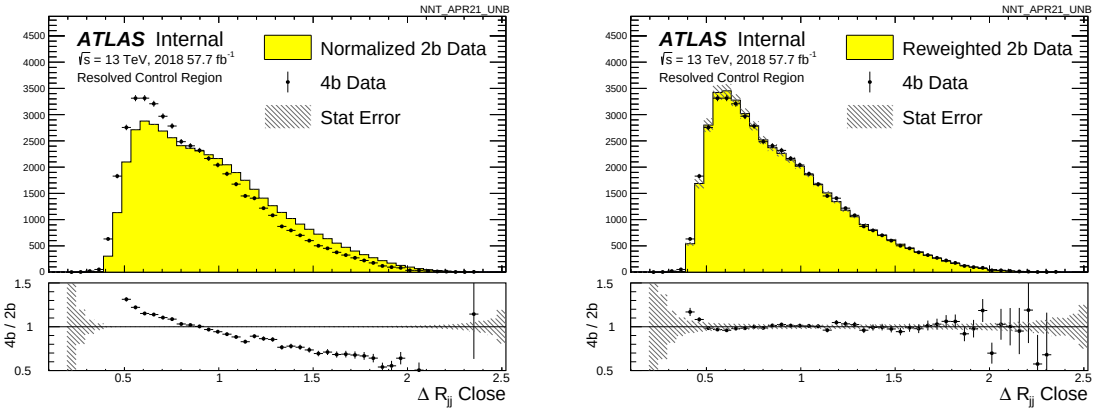


(c) Before Reweighting



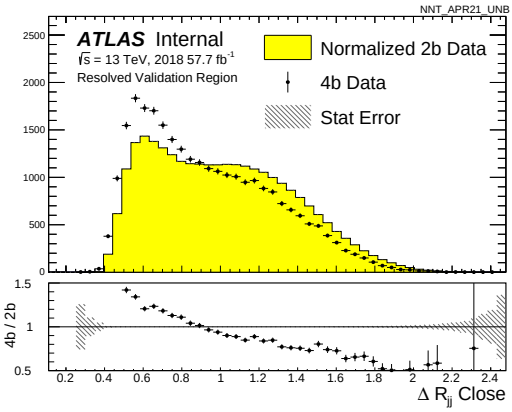
(d) After Reweighting

Figure 8.10.: Resonant ΔR_{HH} distributions before and after reweighting in both Control and Validation regions.

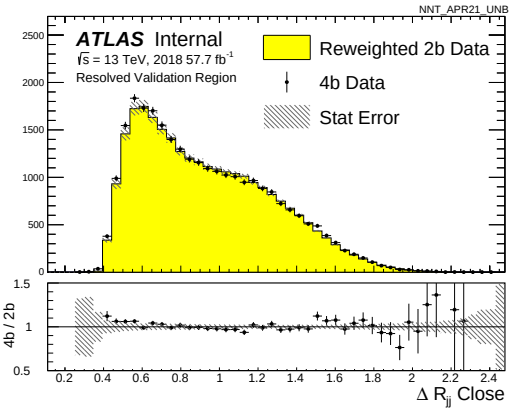


(a) Before Reweighting

(b) After Reweighting

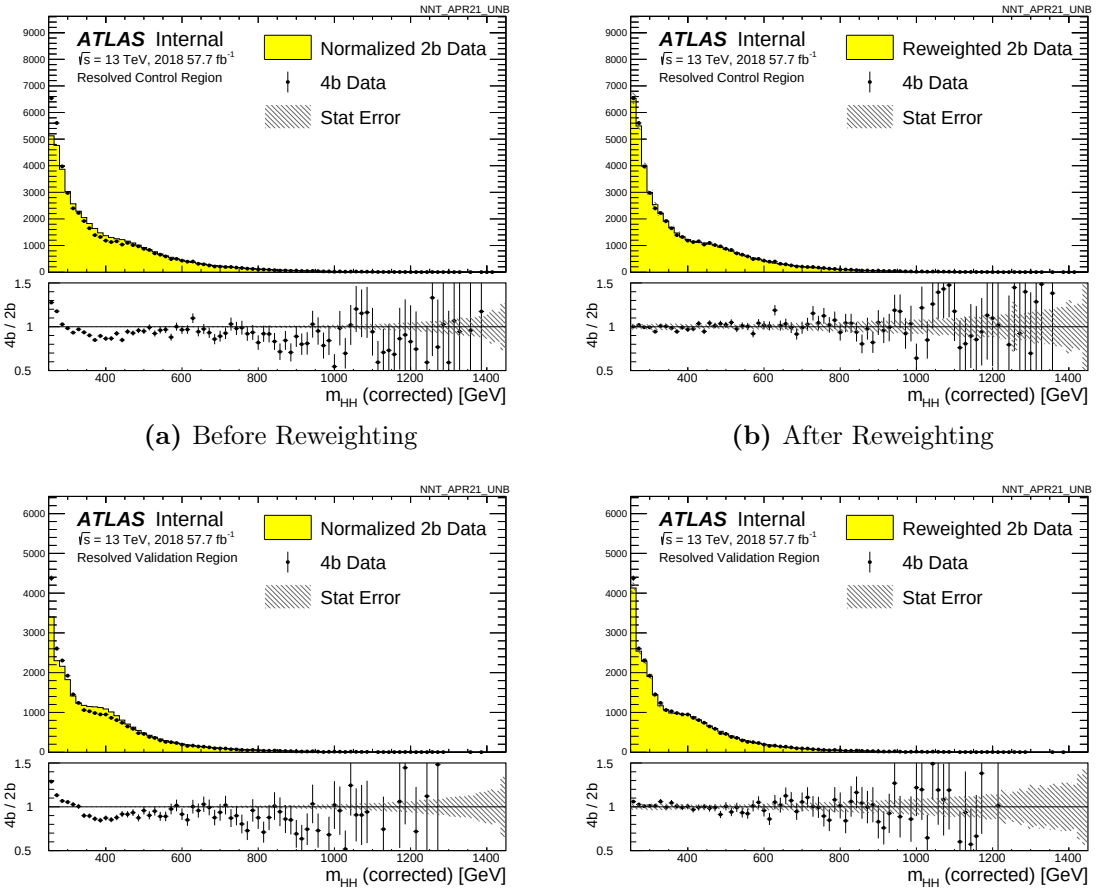


(c) Before Reweighting



(d) After Reweighting

Figure 8.11.: Resonant ΔR_1 distributions before and after reweighting in both Control and Validation regions.



(a) Before Reweighting **(b)** After Reweighting

(c) Before Reweighting **(d)** After Reweighting

Figure 8.12.: Resonant (corrected) m_{HH} distributions before and after reweighting in both Control and Validation regions.

8.3.1. Reversed $|\Delta\eta_{HH}|$ validation

Previously, in Section 8.3, events in the $4b$ category were used to validate the performance of the reweighting model in the control and validation regions. In the signal region we do not have this luxury. The estimated uncertainties would hopefully cover any potential mismodelling of the background, without being so large that the sensitivity of the analysis is diminished. The key issue that remains is that there is no gauge whether the assigned uncertainties sufficiently cover any potential mismodelling. To do this an orthogonal control sample was designed by reversing the $|\Delta\eta_{HH}|$ cut and keeping every other step of the analysis identical. In this space, $4b$ events in the signal region can be safely unblinded since the signal yields are minuscule in comparison to nominal selection. The upper limit, from the reversed region, on the cross-section for all the signal processes considered in this search have already been excluded by the ATLAS collaboration. The $4b$ events in this reversed region are then used to verify the closure of the background estimation machinery. This control sample is referred to as the *reversed* $|\Delta\eta_{HH}|$ region.

As the resonant search uses the corrected m_{HH} variable for the sensitivity limits this variable needs to be correctly modelled by the background estimation procedure. To validate this, the reversed $|\Delta\eta_{HH}|$ region is used in the 2017 data and the results are shown in Figure 8.13. This Figure shows that in general the envelope provided by the shape systematic and the bootstrap uncertainty provides sufficient coverage to the $4b$ events in the signal region. The Figure shows the χ^2/ndf and p-value between the $4b$ distribution and the reweighted $2b$ distributions. Both of these metrics indicate that we cannot reject the hypothesis that the two histograms are sampled from the same distribution. To further support this statement the bottom panel shows the bin-by-bin significance (commonly called *pulls*), which is calculated by using the $2b$ background model and its associated uncertainty along with the $4b$ data and its associated statistical uncertainty. A binned likelihood Gaussian fit on the distribution of these significances is also shown. This particular fit yields a $\mu = -0.063 \pm 0.070$ and $\sigma = 0.960 \pm 0.053$, which is statistically consistent with the ideal distribution of these significances $\sim\mathcal{N}(\mu = 0, \sigma = 1)$.

Although it is encouraging that the Gaussian's fit of the pulls in this reversed region does not show any significant bias in a particular direction, this could be by a random consequence of the feature that was chosen for this test. To make this test more reliable all of the features used in the reweighted model were used to estimate the distribution of pulls. Then the same Gaussian fit was performed and the results are shown in Figure 8.14. This Figure also shows that there is no significant bias in the pulls and we can confidently

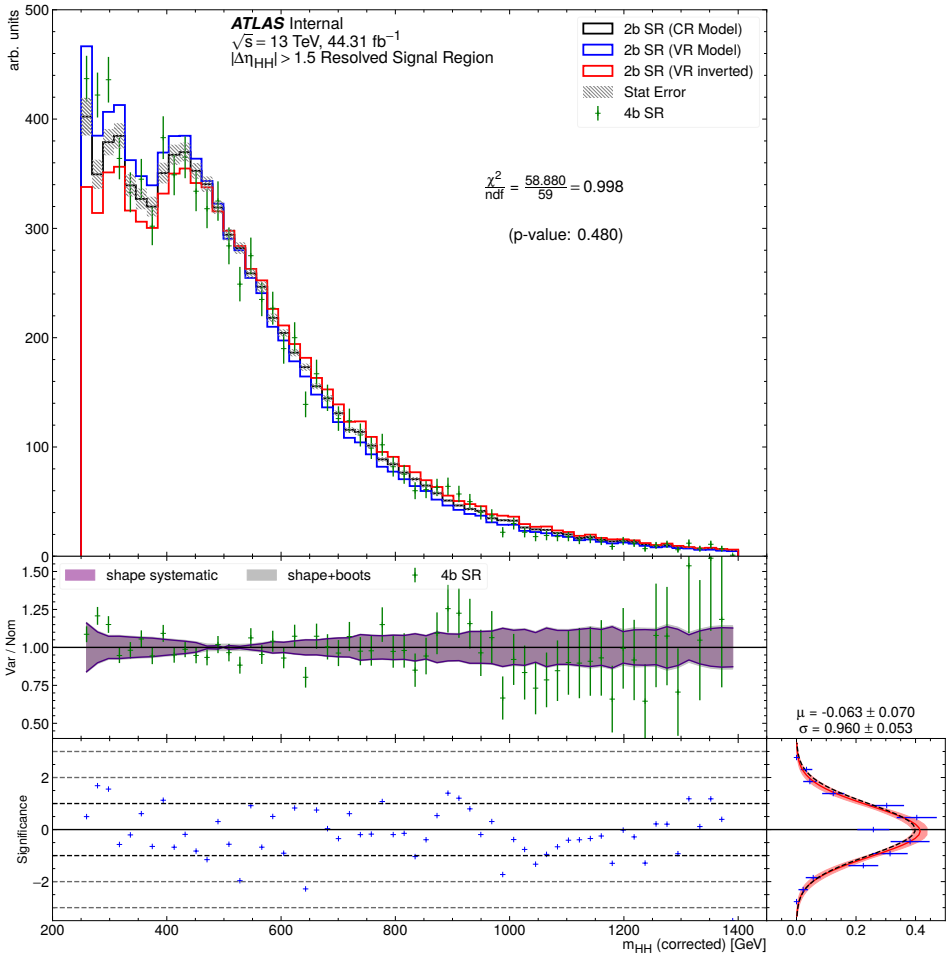


Figure 8.13.: Corrected m_{HH} distribution in the reversed $|\Delta\eta_{HH}|$ validation sample. Reweighted $2b$ events are compared with the $4b$ events in the signal region of this validation sample. The data used for this test are from the 2017 dataset.

say that the uncertainties of the background model provide sufficient coverage for any mismodelling. For completeness, the validation of the remaining kinematic variables used to train the background model are shown in Appendix B.1.

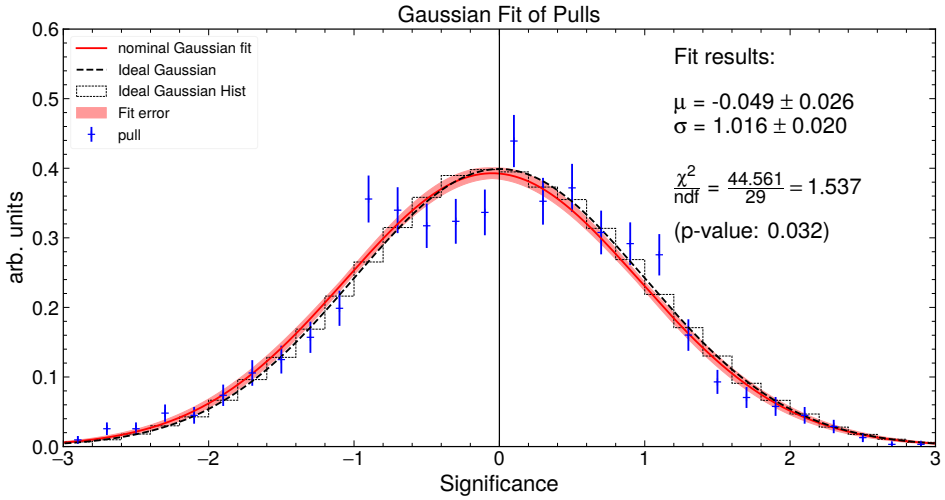
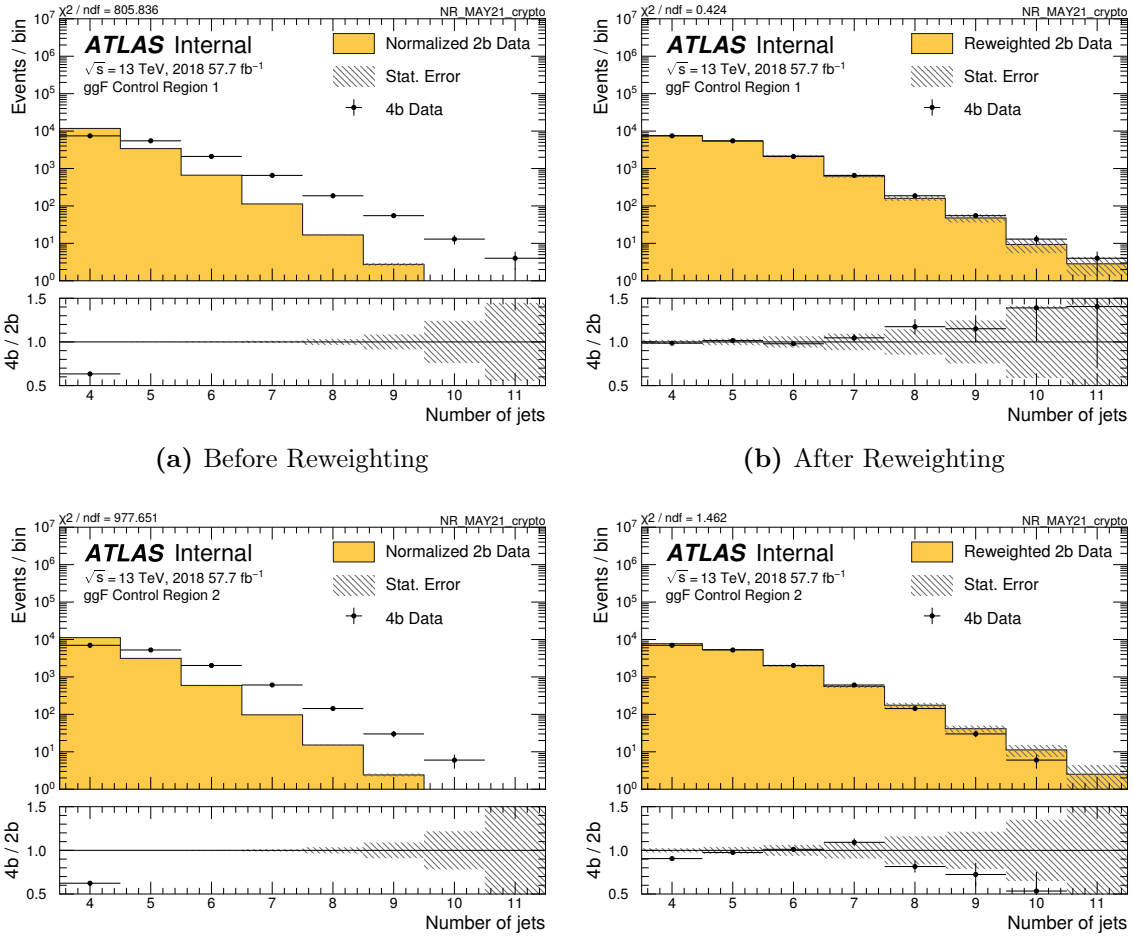


Figure 8.14.: Resonant search binned likelihood Gaussian fit of bin-by-bin significances from the reversed $|\Delta\eta_{HH}|$ validation in the signal region. The significances were obtained by calculating the histogram for all of the distributions using the reweighting mechanism. The fit is compared to an ideal Gaussian distribution with $\mu = 0$ and $\sigma = 1$.

8.4. Non-resonant reweighting

Similar to the resonant reweighting examples, the performance of the non-resonant reweighting is checked for adequate closure. Figures 8.15-8.19, show the effect of the reweighting model for a few key features in both CR1 and CR2 kinematic regions. The examples shown are only using 2018 data for simplicity. Similar to the resonant reweighting, the remaining distributions used to train the reweighting model are shown in Appendix B.2.



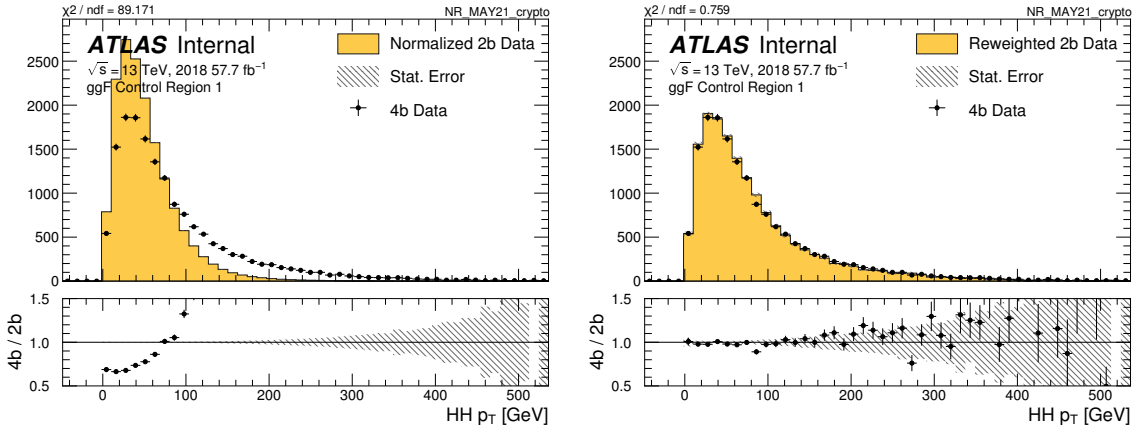
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

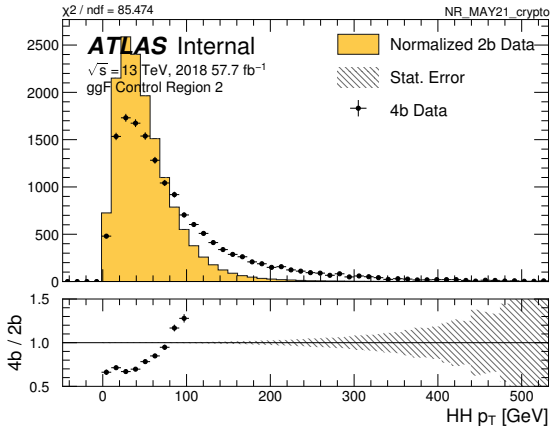
(d) After Reweighting

Figure 8.15.: Non-resonant njets distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

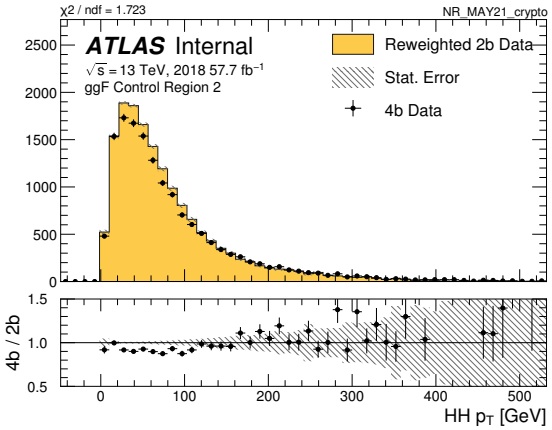


(a) Before Reweighting

(b) After Reweighting

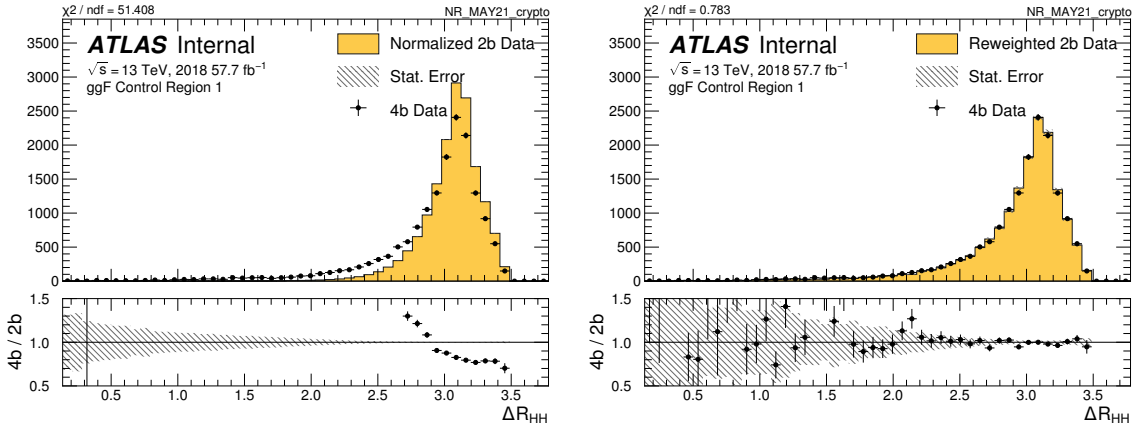


(c) Before Reweighting



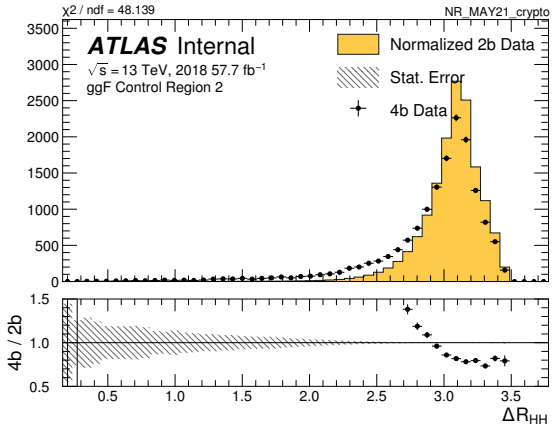
(d) After Reweighting

Figure 8.16.: Non-resonant $p_{T_{HH}}$ distributions before and after reweighting in both CR1 and CR2 kinematic regions

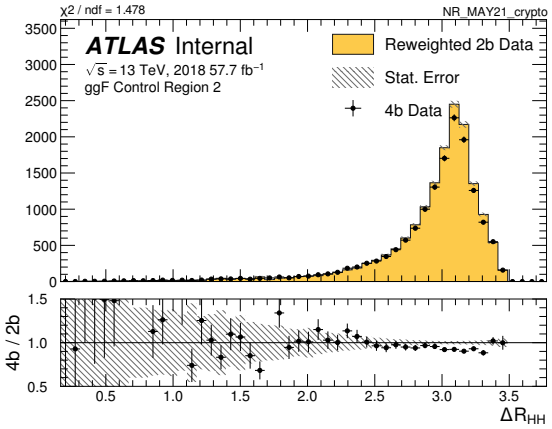


(a) Before Reweighting

(b) After Reweighting

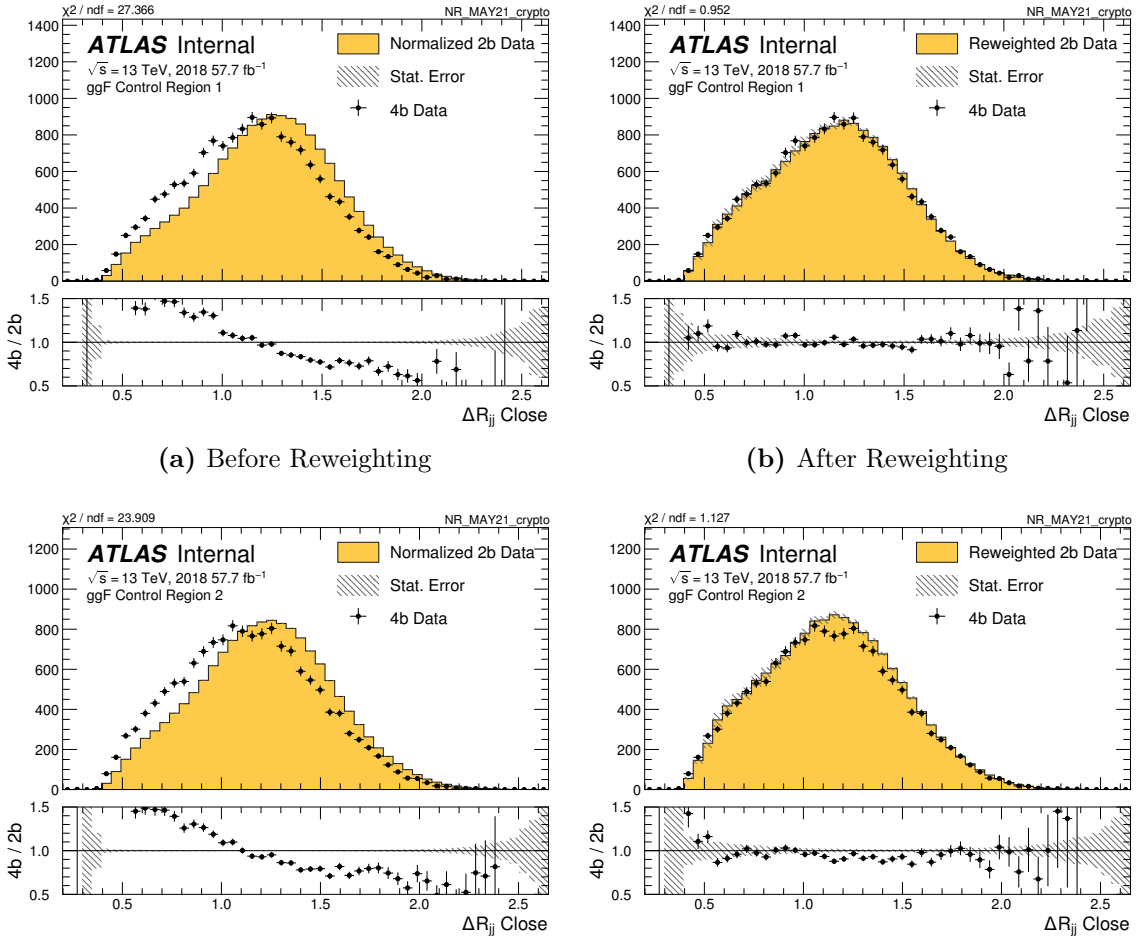


(c) Before Reweighting



(d) After Reweighting

Figure 8.17.: Non-resonant ΔR_{HH} distributions before and after reweighting in both CR1 and CR2 kinematic regions



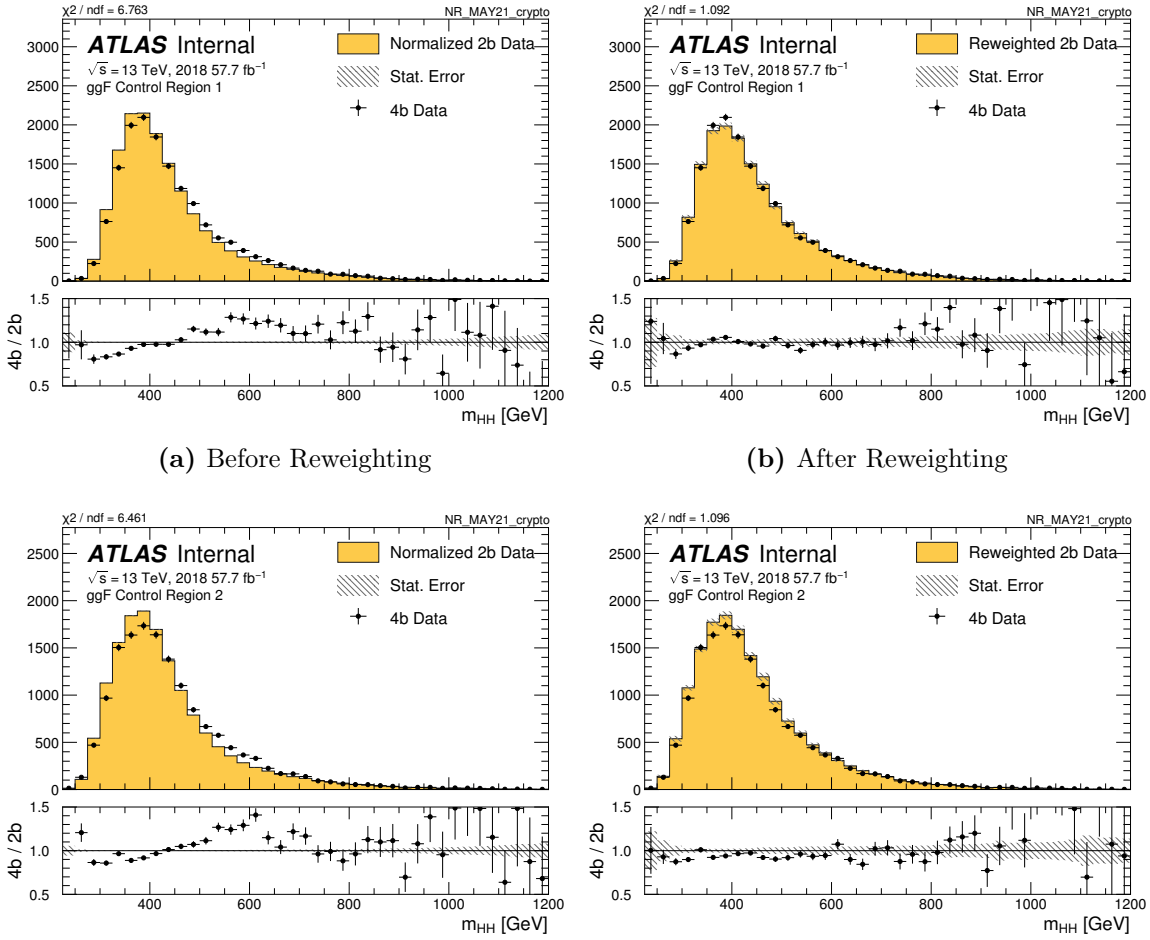
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

(d) After Reweighting

Figure 8.18.: Non-resonant ΔR_1 distributions before and after reweighting in both CR1 and CR2 kinematic regions



(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

(d) After Reweighting

Figure 8.19.: Non-resonant m_{HH} distributions before and after reweighting in both CR1 and CR2 distributions.

From the experience gained in the resonant search the validation of the background modelling was expanded in the non-resonant through the use of additional, orthogonal control samples of events. The next sections describe the validation of the reweighting technique in 3 orthogonal samples.

8.4.1. Non-resonant $3b1f$ validation

The $3b1f$ control sample was the main driving force behind many of the decisions and optimizations for the non-resonant search. Initially it was used to validate the choice of kinematic regions and reweighting setup. To test this setup the agreement of the reweighted $2b$ events and the target $3b1f$ events was evaluated in the signal region. It was found that the current setup yielded adequate and consistent agreement across the whole signal region. Figure 8.20 shows the closure of extrapolating the reweighting model into the signal region of the $3b1f$ validation sample.

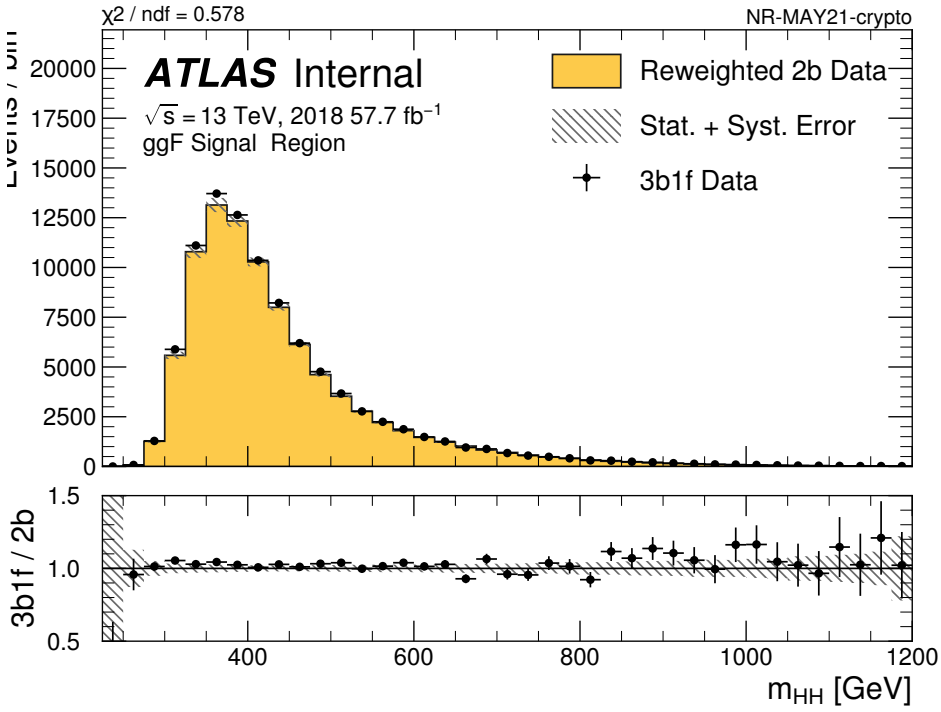


Figure 8.20.: Non-resonant $3b1f$ signal region validation

8.4.2. Non-resonant shifted massplane validation

The background modelling should not be dependent on the location of the massplane, therefore to gain more confidence on the background modelling procedure the mechanism should be tested on different regions of the massplane. Figure 8.21 shows the location of the different kinematic regions across the massplane. In each of these regions, the reweighting procedure is applied in the same way as in the nominal regions of the analysis, and the derived background model can be compared directly with the $4b$ distributions in the corresponding signal regions. In each of the shifted signal region the $4b$ distribution is compared to the distribution of the background prediction along with the relevant uncertainties. From these values a significance can be determined as shown in Equation 8.13, where $N_{background}$ is the yield of the background model, N_{4b} is the $4b$ yield in the shifted signal region and σ_{stat} is the statistical uncertainty. The statistical uncertainty is determined by summing in quadrature the $4b$ Poisson uncertainty, the $2b$ Poisson error, the bootstrap uncertainty and the shape systematic. The distribution of the significances determined by Equation 8.13 is shown in Figure 8.22. In this case, both the $4b$ and $3b1l$ signal regions were used to determine the distribution of these significances. The Gaussian fit shows that there are no significant biases, which provides confidence in the background model and the choice of kinematic regions.

$$\mu_{norm} = \frac{N_{background} - N_{4b}}{\sigma_{stat}} \tag{8.13}$$

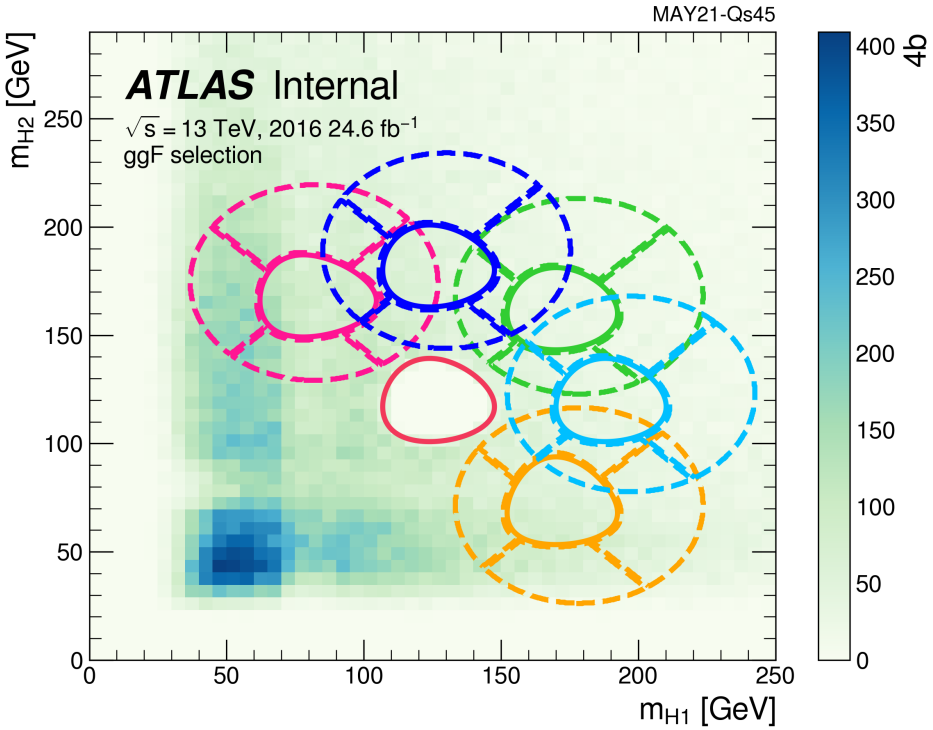


Figure 8.21.: Validation regions across the m_{H1}-m_{H2} massplane. Each of these regions is used to performa closure test for the reweighting model and the target $4b$ events in the signal region.

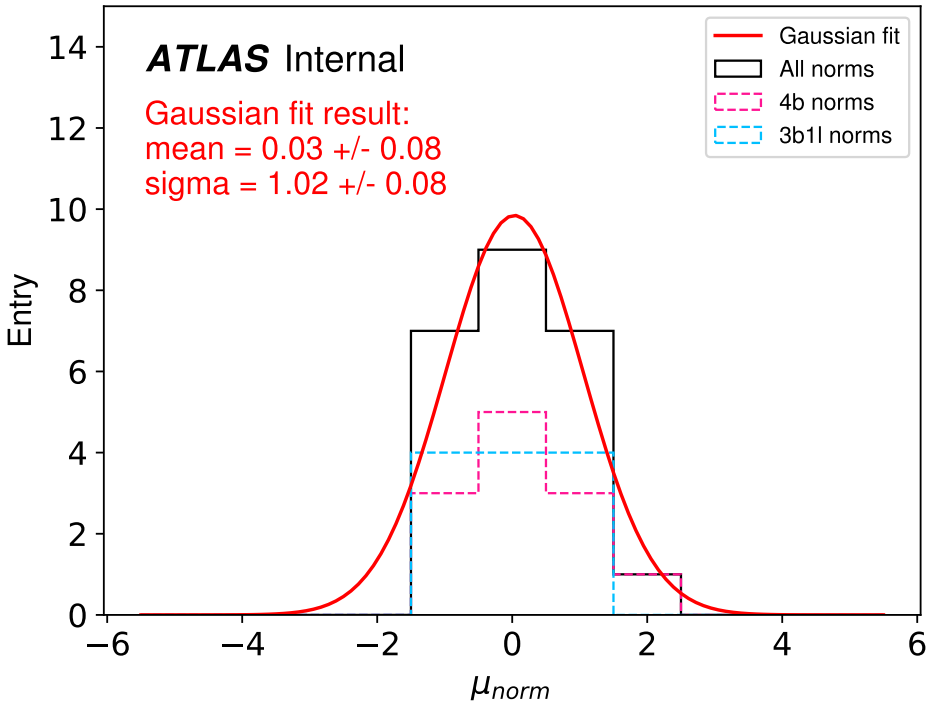


Figure 8.22.: Gaussian fit on shifted regions significances. The differences in both the $4b$ and $3b1l$ signal regions are used to fill the fitted histogram.

8.4.3. Non-resonant reversed $|\Delta\eta_{HH}|$ validation

The same exercise is carried out with the reversed $|\Delta\eta_{HH}|$ sample in the non-resonant search to validate the background model. Figure 8.23 shows the m_{HH} distribution and Figure 8.24 shows the distributions of pulls in this control sample using the non-resonant reconstruction. The χ^2/ndf in m_{HH}, indicates that the $4b$ and reweighted $2b$ events are in agreement with each other. However the bin-by-bin significance shows that there is a consistent over-prediction by the background modelling beyond the coverage provided by the uncertainty bands around 400 GeV. These significances were determined by only including the shape and bootstrap uncertainties while the background normalization uncertainty was omitted. This uncertainty is included in the statistical limit setting framework and hence the pulls seen here are likely to improve with the inclusion of this uncertainty. Moreover, the over-prediction is not seen in the $3b1f$ and shifted massplane validation regions, which may indicate that this over-prediction is an artefact of the reversed $|\Delta\eta_{HH}|$ region only. It is for these reasons that the over-prediction seen in the reversed $|\Delta\eta_{HH}|$ region is not an indication that the background modelling of the nominal $4b$ region is biased. The validation of the remaining kinematic distributions used to train the background model are shown in Appendix section B.2.

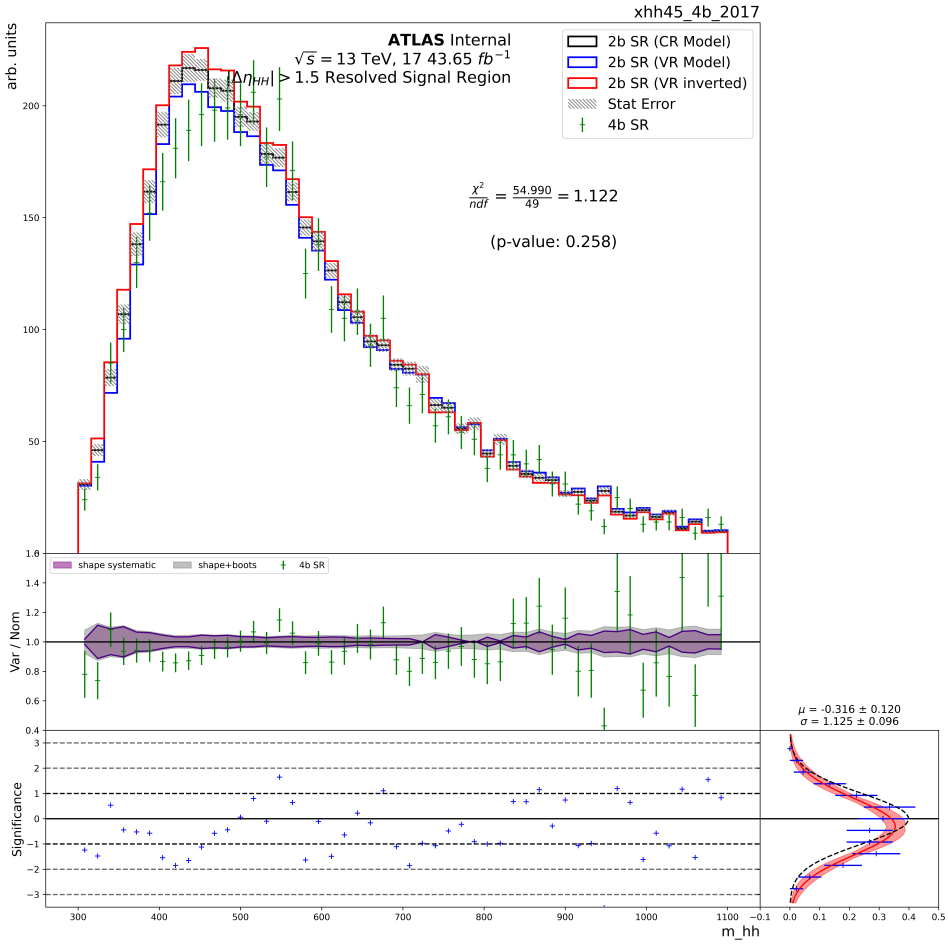


Figure 8.23.: Non-resonant m_{HH} distribution in the reversed $|\Delta\eta_{HH}|$ validation sample. Reweighted $2b$ events are compared with the $4b$ events in the signal region of this validation sample. The data used for this test are from the 2017 dataset.

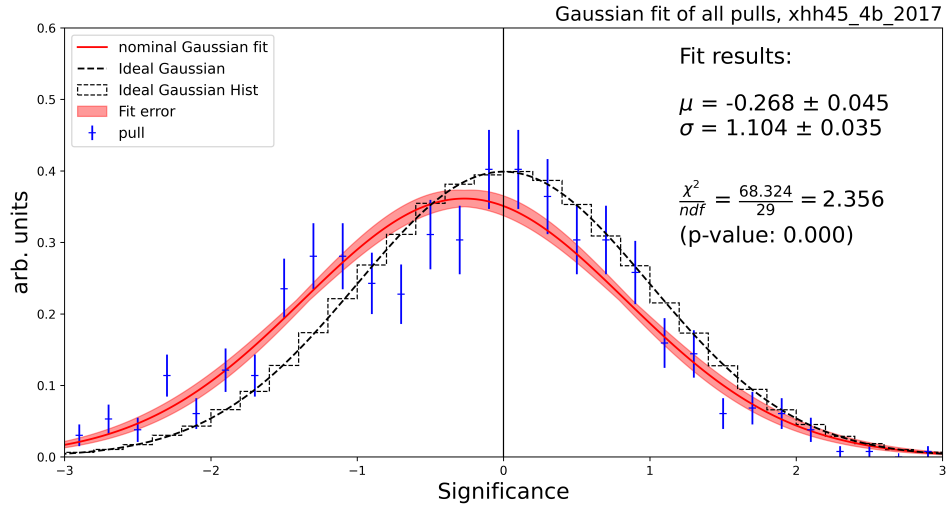


Figure 8.24.: Non-resonant search binned likelihood Gaussian fit of bin-by-bin significances from the reversed $|\Delta\eta_{HH}|$ validation in the signal region. The significances were obtained by calculating the histogram for all of the distributions using the reweighting mechanism. The fit is compared to an ideal Gaussian distribution with $\mu = 0$ and $\sigma = 1$.

Chapter 9.

Trigger Optimization

The early Run-2 analysis used a combination of b-jet triggers to maximize its potential, which is used as a basis for the full Run-2 search. This chapter will discuss the trigger selections, optimizations and design choices for both the resonant and non-resonant analyses.

9.1. Trigger exploration

ATLAS provides a list of b-jet triggers that were used for each year of data taking and have been approved for use in searches. Graviton signal samples within the resolved mass window, 300–1300 GeV, were used for the optimization of the resonant analysis. Similarly, the trigger optimization for the non-resonant analysis was done using the SM HH signal samples. Since the background model requires events with exactly two b-jets, only triggers requiring at least one or two b-jets can be used. Figures 9.1 and 9.2 show the $1b$ and $2b$ trigger efficiency as a function of the resonance mass of the G^*_{KK} sample, respectively. In these Figures the efficiency is defined as the ratio of events that passed the trigger or trigger combination with respect to the number of events that have passed the offline reconstruction requirements. The $1b$ triggers shown in Figure 9.1 are generally more efficient at high graviton masses. This is because $1b$ triggers have a high p_T threshold, in the range of 225–360 GeV, in order to keep the event rate manageable. The $1b + 3j$ triggers also seen in Figure 9.1 have a higher low mass efficiency than the $1b$ triggers, but their overall efficiency is much lower. The difference in their efficiency curve comes from the requirement of three additional non-b-tagged jets, which allows them to use lower p_T thresholds, between 75–85 GeV, than the $1b$ triggers. Additionally,

some triggers have an additional HT[1] constraint. In the case of the $1b + HT$ trigger, this allows it to have a p_T threshold of 55 GeV, which improves the low mass efficiency. Similarly, due to their lower p_T thresholds the $2b$ triggers shown in Figure 9.2 are also most efficient at low masses. At higher masses events become more collimated causing two jets to overlap and be reconstructed as a single jet, resulting in the lower efficiency of $2b$ triggers. Both $1b$ and $2b$ triggers can have tighter or looser b-tagging requirements, which has little effect on the shape of the efficiency curve but can significantly change the overall efficiency. From these two Figures it is evident that a combination of $1b$ and $2b$ triggers can maximize the efficiency across the whole mass range.

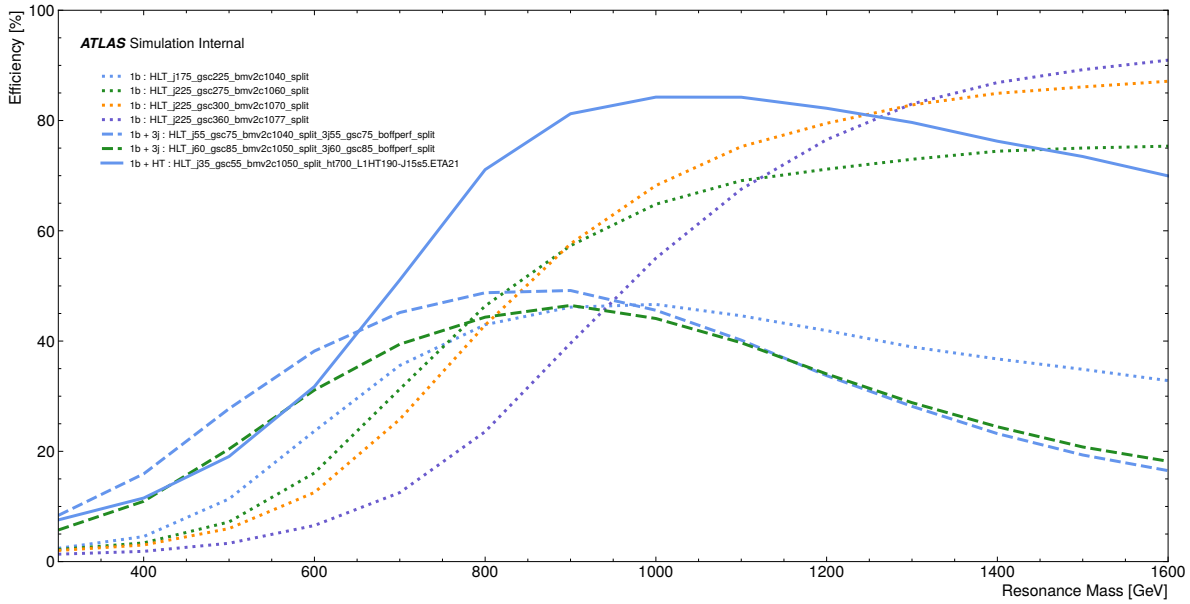


Figure 9.1.: Efficiency of $1b$ triggers as a function of the G^*_{KK} resonance mass.

9.2. Trigger combinations

Initially the resonant and non-resonant search were combined together and the trigger optimization was performed with both of these searches in mind. To find the best combination of triggers to use, a brute force approach that calculates the overall efficiency for each possible trigger combination was used. The search was performed on a dataset made up of both SM HH and G^*_{KK} signal samples. In this dataset the number of SM HH events was equal to the total number G^*_{KK} events that are evenly distributed across the resolved resonance mass window of 300–1300 GeV. For 2015 and 2016 the best trigger

[1] HT is the scalar sum of jet p_T

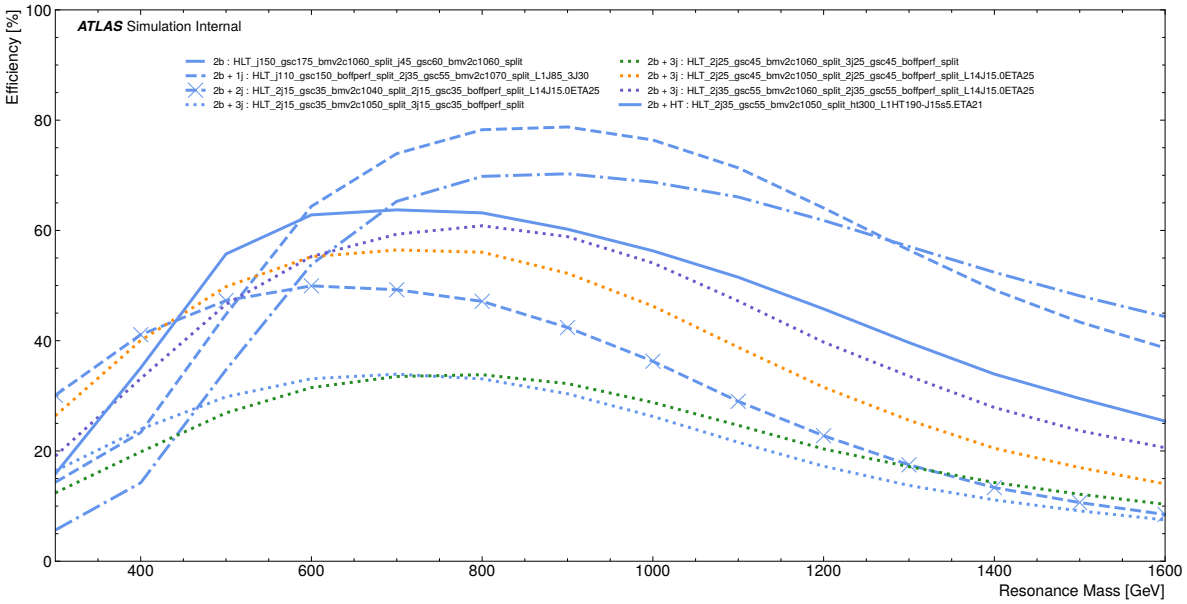


Figure 9.2.: Efficiency of $2b$ triggers as a function of the G^*_{KK} resonance mass.

combinations were found to align with the triggers used in the early Run-2 analysis. Table 9.1, shows the three best triggers to combine in 2017, which are of the same trigger type as what was found in the early Run-2 analysis. To improve the efficiency for low m_{HH} events, which is where the $HH \to b\bar{b}b\bar{b}$ suffers the most, the $2b + HT$ trigger was added to the list of 2017 and 2018 triggers. This trigger is not used in 2016 since the $2b + HT$ trigger in this year uses the same Level-1 seed as the $2b + 2j$ trigger, which results in a negligible added contribution from the $2b + HT$ trigger. Table 9.3 shows the full list of triggers, in both resonant and non-resonant searches, across each year. To complement Table 9.3, Table 9.2 shows the different Level-1 seeds used in all of the selected triggers. Figure 9.3 shows how the combination of all these triggers behave as a function of the G^*_{KK} resonance mass. The combined efficiency shows that an efficiency plateau is reached at around 800 GeV, but it does show that the low mass region is where the b-jet triggers suffer the most.

Trigger Type	Trigger Name
$1b$	HLT_j225_gsc300_bmv2c1070_split
$2b + 2j$	HLT_2j15_gsc35_bmv2c1040_split_2j15_gsc35_boffperf_split_L14J15.0ETA25
$2b + 1j$	HLT_j110_gsc150_boffperf_split_2j35_gsc55_bmv2c1070_split_L1J85_3J30
Yield [%]	78.44

Table 9.1.: Combination of three triggers in 2017 that yield the highest efficiency in the mixed G^*_{KK} and SM HH dataset.

L1 Seed Index	Requirements
1	At least 1 jet $p_T > 100\,\mathrm{GeV}$
2	At least 1 jet $p_T > 75\,\mathrm{GeV}$ and 2 with $p_T > 20\,\mathrm{GeV}$
3	4 jets $p_T > 15\,\mathrm{GeV}$ and $\lvert\eta\rvert < 2.5$
4	At least 1 jet with $p_T > 85\,\mathrm{GeV}$ and 2 with $p_T > 30\,\mathrm{GeV}$
5	At least one jet with $p_T > 15\,\mathrm{GeV}$ as well as $\lvert\eta\rvert < 2.1$. $H_T > 190\,\mathrm{GeV}$.

Table 9.2.: Level 1 seeds used in the triggers of the $HH \to b\bar{b}b\bar{b}$ analysis.

Year	Analysis Use	Trigger Type	Trigger Name	Trigger Description	L1 Seed Index
2016	Resonant	$1b$	*HLT_j**225**_bmv2c20**60**_split*	One b-jet with $p_T > 225$ GeV at 60 % W.P.	1
	both	$2b + 1j$	*HLT_j**100**_2j**55**_bmv2c20**60**_split*	Two b-jets with $p_T > 55$ GeV at 60 % W.P. and one extra jet with $p_T > 100$ GeV	2
	both	$2b + 2j$	*HLT_**2j35**_bmv2c20**60**_split_**2j35**_L14J15.0ETA25*	Two b-jets with $p_T > 35$ GeV at 60 % W.P. and two extra jets with $p_T > 35$ GeV	3
2017	Resonant	$1b$	*HLT_j225_gsc**300**_bmv2c10**70**_split*	One b-jet with $p_T > 300$ GeV at 70 % W.P.	1
	both	$2b + 1j$	*HLT_j110_gsc**150**_boffperf_split_**2**j35_gsc**55**_bmv2c10**70**_split_L1J85_3J30*	Two b-jets with $p_T > 55$ GeV at 70 % W.P. and one extra jet with $p_T > 150$ GeV	4
	both	$2b + 2j$	*HLT_**2**j15_gsc**35**_bmv2c10**40**_split_**2**j15_gsc**35**_boffperf_split_L14J15.0ETA25*	Two b-jets with $p_T > 35$ GeV at 40 % W.P. and two extra jets with $p_T > 35$ GeV	3
	Resonant	$2b + HT$	*HLT_**2**j35_gsc**55**_bmv2c10**50**_split_ht**300**_L1HT190-J15s5.ETA21*	Two b-jets with $p_T > 55$ GeV at 50 % W.P. and HT > 300 GeV	5
2018	Resonant	$1b$	*HLT_j225_gsc**300**_bmv2c10**70**_split*	One b-jet with $p_T > 300$ GeV at 70 % W.P.	1
	both	$2b + 1j$	*HLT_j110_gsc**150**_boffperf_split_**2**j45_gsc**55**_bmv2c10**70**_split_L1J85_3J30*	Two b-jets with $p_T > 55$ GeV at 70 % W.P. and one extra jet with $p_T > 150$ GeV	4
	both	$2b + 2j$	*HLT_**2j35**_bmv2c10**60**_split_**2j35**_L14J15.0ETA25*	Two b-jets with $p_T > 35$ GeV at 60 % W.P. and two extra jets with $p_T > 35$ GeV	3
	Resonant	$2b + HT$	*HLT_**2**j45_gsc**55**_bmv2c10**50**_split_ht**300**_L1HT190-J15s5.ETA21*	Two b-jets with $p_T > 55$ GeV at 50 % W.P and HT > 300 GeV	5

Table 9.3.: Selected triggers for both resonant and non-resonant $HH \to b\bar{b}b\bar{b}$ searches. Each trigger has a unique identifier, which describes the constraints that the trigger requires. The description of each trigger indicates the number of b-tagged jets along with their p_T threshold and b-tagging working point (W.P.). Some triggers have additional constraints beyond the b-tagged jets, such as one or two extra non-b-tagged jets. The non-resonant search only makes use of trigger types $2b + 2j$ and $2b + 1j$, while the resonant search also uses $1b$ and $2b + HT$ triggers. These triggers are formed by software (HLT) triggering system, which uses the seeds provided by the Level-1 firmware trigger. The complete definition of the Level-1 seed used for each trigger can be found by correlating the index in this table with Table 9.2.

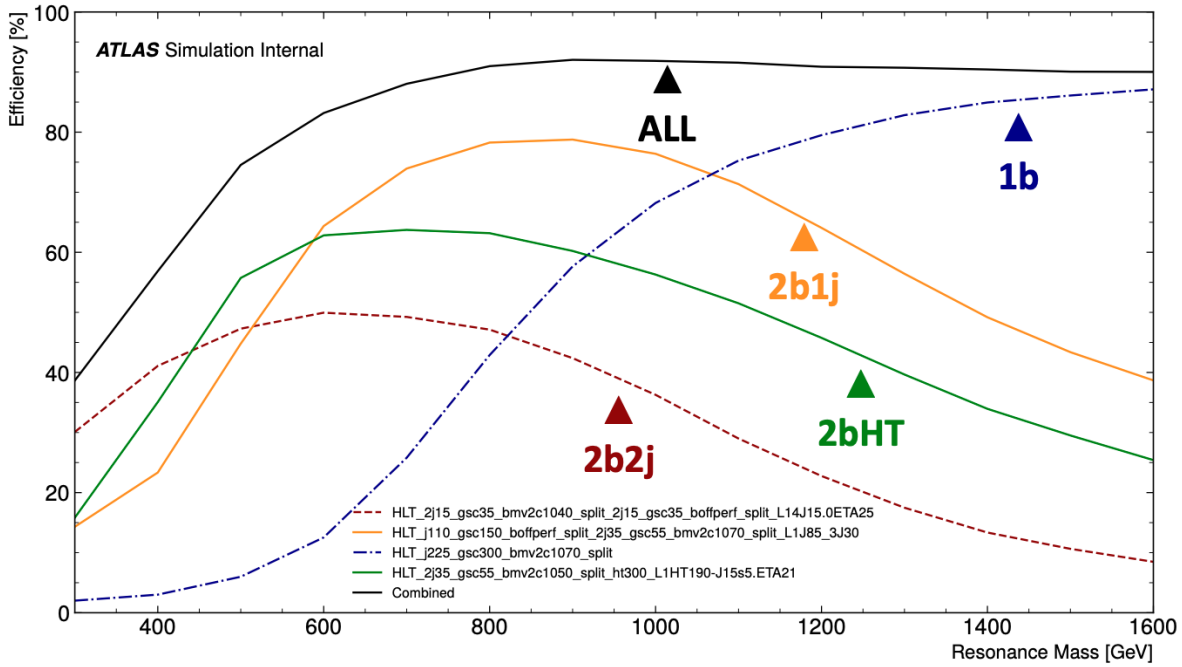


Figure 9.3.: The efficiency of the four selected triggers and their combination across the resonance mass spectrum G^*_{KK} events for 2017. The efficiencies are determined prior to any analysis cuts. The mean signal efficiency for the combination of triggers is approximately 82.4 %.

9.3. Trigger matching

To correctly calibrate our b-jet triggers, only events that satisfy the trigger matching condition must be used. This is due to the fact that trigger efficiencies and scale-factors are measured on a per-jet basis. The trigger matching procedure builds a connection between reconstructed jets and trigger cuts by matching the offline jets to online jets. The exact nature of the trigger matching condition depends on the trigger being used, but in general, an event needs to have a sufficient number of offline jets matched to online jets for the trigger in question. An event can be matched to multiple triggers and if the event is not matched to any trigger it is discarded. The offline jets are first geometrically matched to online jets with a geometrical constraint of $\Delta R < 0.4$. Those online jets that are geometrically matched to b-tagged offline jets are then checked to see if the online b-tagging requirements of the trigger are also met. The matching efficiency ($\varepsilon^{trigger}$) can be defined as a function of the trigger in question and is determined by the ratio of matched events to total events for that trigger. To illustrate this, the 2017 $2b+2j$ trigger can be used. This trigger requires 2 jets to be b-tagged at a 60 % working point, which

for the MV2c10 tagger corresponds to a score of $b_{score}^{2b+2j} = 0.846$. The matching efficiency of this trigger was 99.4 % when tested on 10000 SM HH events. Table 9.4 and 9.5 show the matching efficiency for both data and signal samples, all of which are close to unity. The details of the trigger matching algorithm are shown in algorithm 1.

Algorithm 1: Offline to online jet trigger matching pseudo-code

input : $b_{score}^{trigger}$ jet b-tagging score requirement of trigger
input : $n_{b-tag}^{trigger}$ number of b-tagged jets required by trigger
output : Decision whether the event is matched to triggers

```
begin
    n_M = 0 (number of offline jets matched to online jets)
    for each offline jet do
        for each online jet do
            if ΔR(online jet, offline jet) < 0.4 then
                if b_score^offlinejet > b_score^trigger then
                    n_M = n_M + 1
    if n_M ≥ n_b-tag^trigger then
        Event has been matched to this trigger.
```

9.4. Trigger buckets

Using multiple triggers carries the complication that an event may pass any subset of these triggers. Trigger efficiencies can be calibrated between data and MC, but when

Signal Type	**SM HH**			**G^*_{KK} 400 GeV**	**Scalar 500 GeV**
Campaign	**MC16a**	**MC16d**	**MC16e**	**MC16a**	**MC16e**
Year	**2016**	**2017**	**2018**	**2016**	**2018**
ε^{2b+1j} [%]	99.8	99.7	99.7	99.5	99.9
ε^{2b+2j} [%]	99.5	99.4	99.6	99.7	99.7
ε^{1b} [%]	99.5	99.4	99.4	99.4	100
ε^{2b+HT} [%]	99.7	99.8	99.4	99.3	99.5
Tagger	MV2c20	MV2c10	MV2c10	MV2c20	MV2c10

Table 9.4.: Trigger matching efficiencies of different signal samples.

	Data		
Year	**2016**	**2017**	**2018**
ε^{2b+1j} [%]	98.5	98.9	99.3
ε^{2b+2j} [%]	98.2	98.6	97.3
ε^{1b} [%]	97.7	98.7	98.4
ε^{2b+HT} [%]	98.9	98.1	98.3
Tagger	MV2c20	MV2c10	MV2c10

Table 9.5.: Trigger matching efficiencies of data across the three different years.

multiple triggers are used these calibrations become far more difficult. To illustrate this consider the case of only using 2 triggers for simplicity. In this case, it is possible that an event fails the first trigger but passes the second trigger. For those events, the difference between data and MC needs to be accounted for by calculating this probability, commonly known as an *anti-scale factor*. Ultimately, this would require the conditional efficiency of passing the second trigger given that the first trigger failed and this quantity is not measured by the collaboration. To avoid these complications the *trigger buckets* categorization allows events to be compared to exactly one trigger, but it comes with a penalty.

The trigger buckets orthogonalize events into categories which depend on the events offline kinematics. The categories are designed, using offline kinematic cuts, so that the trigger being used for that category remains efficient. The penalty of this scheme is that if an event does not pass the trigger of the category it has been placed in, then the event is discarded. Therefore this trigger bucket scheme must be designed to minimize the events lost. An example diagram of a trigger bucket scheme using N sets of triggers is shown in Figure 9.4. The following sections will describe the design choices for the trigger bucketing scheme in both the resonant and non-resonant searches.

9.4.1. Resonant trigger buckets

The resonant analysis is where the usage of multiple b-jet triggers is truly essential. This is because the varying resonance mass requires the triggers to be efficient across a wide m_{HH} spectrum. For example, the $2b + 2j$ trigger would be more efficient at capturing low mass gravitons of around 300 GeV, while the $1b$ trigger is more useful for high mass gravitons above 1000 GeV. This means that the range of resonance masses needs to be

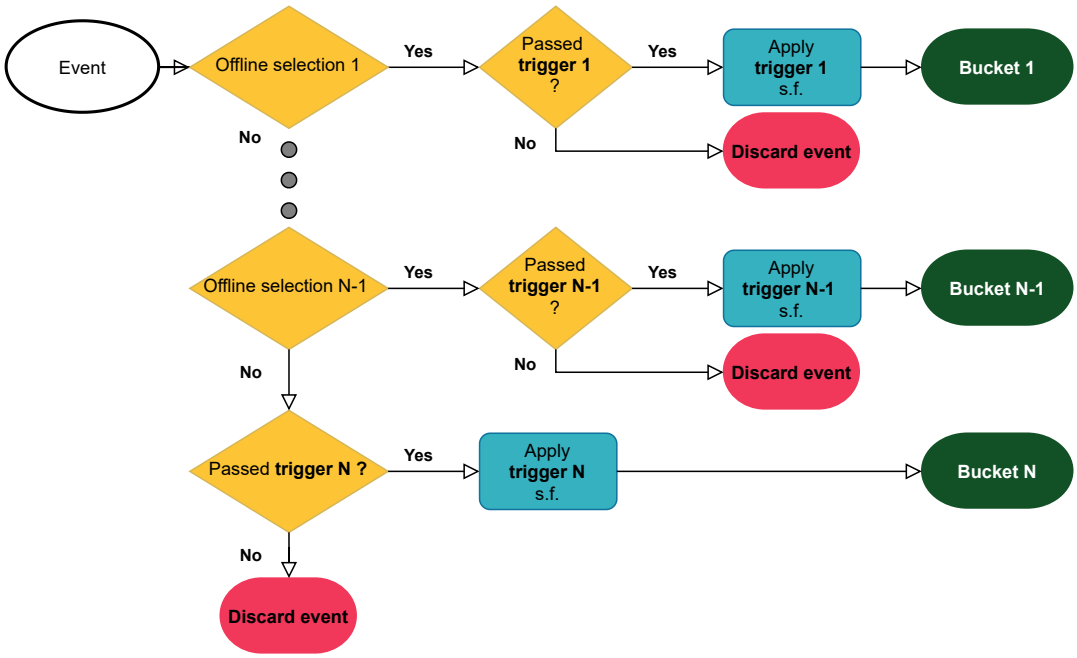


Figure 9.4.: Example diagram of the trigger buckets scheme using N triggers. Events are categorized by offline selections and then kept if they passed the trigger in question. If an event does not pass the trigger then it is discarded. If an event does not pass the offline selection, then it will be checked for the offline selection of the next bucket and so forth. The final bucket does not have an offline selection and the event is only required to pass the final trigger to assign it to the last bucket. In this case, triggers are also required to be matched in order for them to pass.

	Trigger	**Offline Selection Variables**
Bucket 1	$1b$	lead jet $p_T > 325$ GeV and is b-tagged
Bucket 2	$2b + 1j$	lead jet $p_T > 168.75$ GeV and is not b-tagged
Bucket 3	$2b + HT$	$HT > 900$ GeV
Bucket 4	$2b + 2j$	-

Table 9.6.: Resonant trigger buckets, optimal offline selections and trigger order.

	Signal (%)	**Data (%)**	**Offline Cut [GeV]**
Bucket 1	6.5	0.6	325
Bucket 2	24.2	9.7	168.75
Bucket 3	1.8	0.1	900
Bucket 4	49.1	77.3	N/A
Discarded	18.5	12.3	N/A

Table 9.7.: Cutflow of the optimal resonant trigger buckets. Percentages indicate the fraction of events from the total number of events from the G^*_{KK} signal samples within a resonance mass window of 300-1200 GeV. Only 2017 MC16d G^*_{KK} events were used for these results.

considered when designing the trigger buckets scheme. The efficiencies of the individual triggers drive the decision for the order of the trigger buckets. For example the $1b$ trigger is relatively inefficient so it is placed with the top priority in order to allow for enough statistics. Conversely, the $2b + 2j$ trigger is the most efficient so it is given the lowest priority in the trigger bucket order. The order of the triggers and offline selections chosen for the resonant buckets are shown in table 9.6. The p_T of the leading jet is determined by looking at all jets with $p_T > 40$ GeV and $|\eta| < 2.5$. Similarly, the HT is determined by the scalar p_T sum of all jets with $p_T > 25$ GeV and $|\eta| < 2.5$. The offline variables used were chosen to operate on the highest efficiency plateau of each trigger. The cuts used for each bucket were obtained through a grid-search that maximizes the area under the curve formed by significance as a function of the G^*_{KK} mass. The full logic of the resonant trigger buckets is shown in Figure 9.5.

The resonant trigger bucket scheme results in about 18.5 % of the signal to be discarded, which is far from negligible. At the same time it also results in 12.3 % of the data to be discarded. This results in comparable level of significance prior to the trigger bucketing. To fully understand the impact of the resonant trigger bucket scheme,

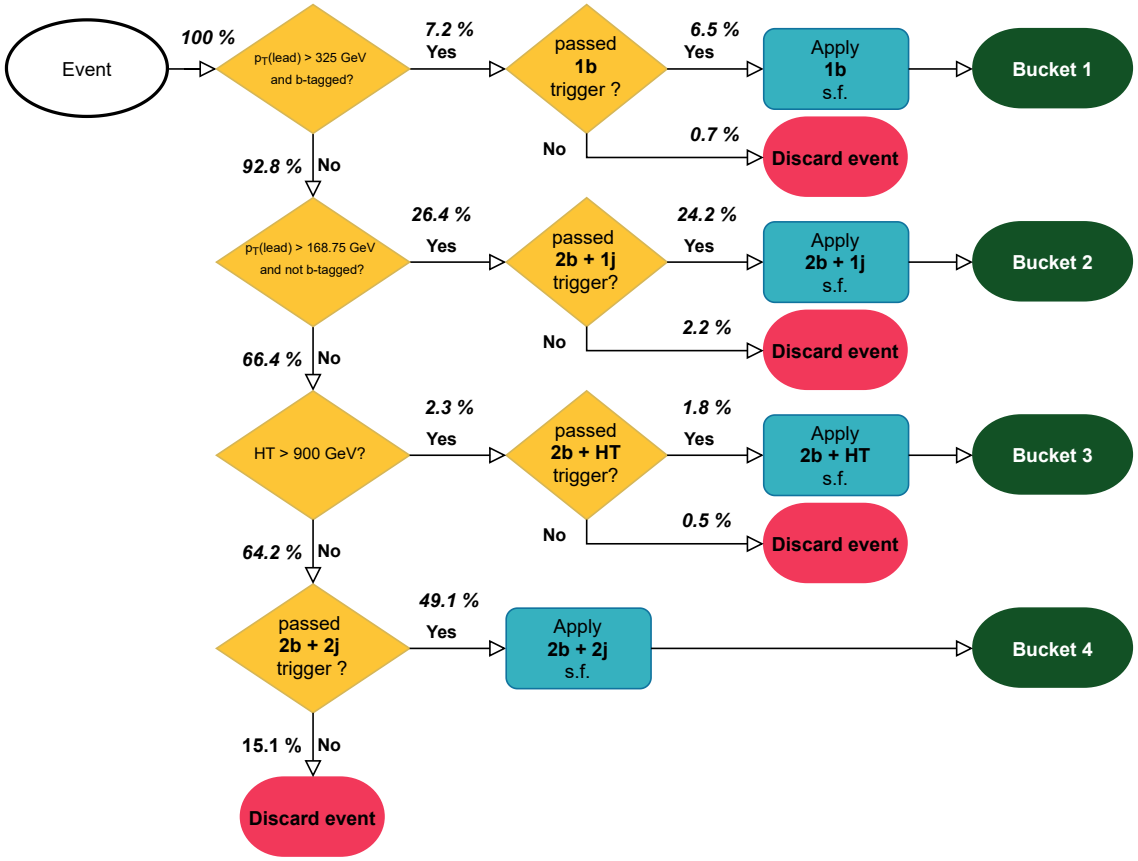
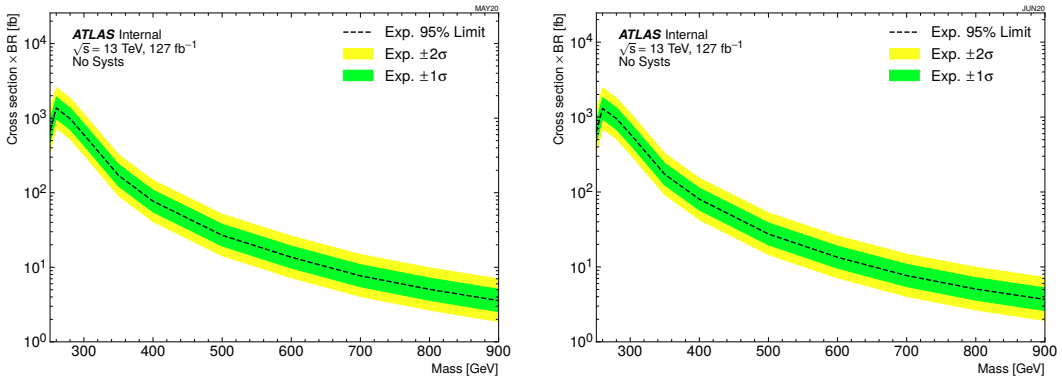


Figure 9.5.: Resonant trigger buckets event categorization flow diagram. Percentages indicate the ratio of events with respect to the total after each decision step. The percentages were determined by using G^*_{KK} 2017 samples with masses ranging from 300 to 1300 GeV. A total of 18.5 % of events are discarded, with the majority coming from the last $2b + 2j$ trigger.

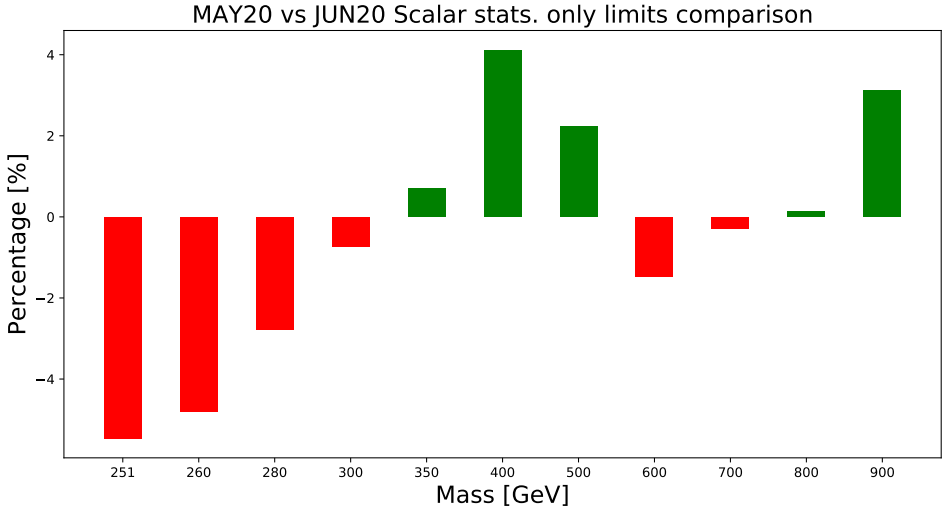
the expected 95 % confidence level limits on the cross section for the scalar sample was produced. The limits are compared with and without the trigger buckets selection in Figure 9.6. Figure 9.6a shows the limits prior to trigger bucketing, matching and it also includes 2015 data. Figure 9.6b, shows the limits, with data 2015 excluded, after the trigger bucketing and matching has been applied. The percentage difference as a function of the resonance mass is shown in Figure 9.6c. This shows that the use of the trigger buckets scheme for the resonant analysis is reasonable as the percentage difference is minimal.

9.4.2. Non-resonant trigger buckets

The non-resonant signal allows for a much simpler trigger buckets scheme. The $1b$ trigger is only about 5 % efficient in the $4b$ SR and can be safely dropped from consideration. Similarly the $2b + HT$ trigger has a significant overlap with the $2b + 2j$ trigger and designing an efficient trigger bucket is not possible with the $2b + HT$ trigger. Therefore only the $2b + 1j$ and $2b + 2j$ triggers are considered for the non-resonant trigger buckets scheme. To do this a grid-search was performed to evaluate the $4b$ SR yields on the non-resonant signal samples. The search showed that the first bucket can be designed by requiring that the leading jet $p_T > 170$ GeV and the third leading jet $p_T > 70$ GeV, while requiring that the event passes the $2b + 1j$ trigger. Then the $2b + 2j$ trigger is used for the second bucket. The simplified trigger buckets scheme for the non-resonant analysis is shown in Figure 9.7. This strategy was compared to a strategy using only the $2b + 2j$ without any bucketing and it was found to improve S/B yields by 9 %, thus the benefit of the trigger buckets scheme is significant over the additional complexity of the trigger buckets. Figure 9.8 shows the m_{HH} distribution of $4b$ events in the SR for both bucket 1 and bucket 2. Bucket 1 as it uses the $2b + 2j$ trigger captures the high mass region while the $2b + 2j$ trigger captures the low mass region.



(a) Before trigger buckets and matching **(b)** After trigger buckets and matching



(c)

Figure 9.6.: Expected confidence level limits on cross-section of $pp \to X \to b\bar{b}b\bar{b}$ with only statistical uncertainties included. These limits ignore the application of scale factors to correct the signal prediction. a) Shows the expected limits before trigger bucketing and matching. b) Shows the expected limits after both trigger bucketing and matching have been applied. This also excludes data taken in 2015. The offline selections of the trigger buckets used in the resonant search are shown in Table 9.6. c) Shows the percentage difference between the limits shown in a) and b).

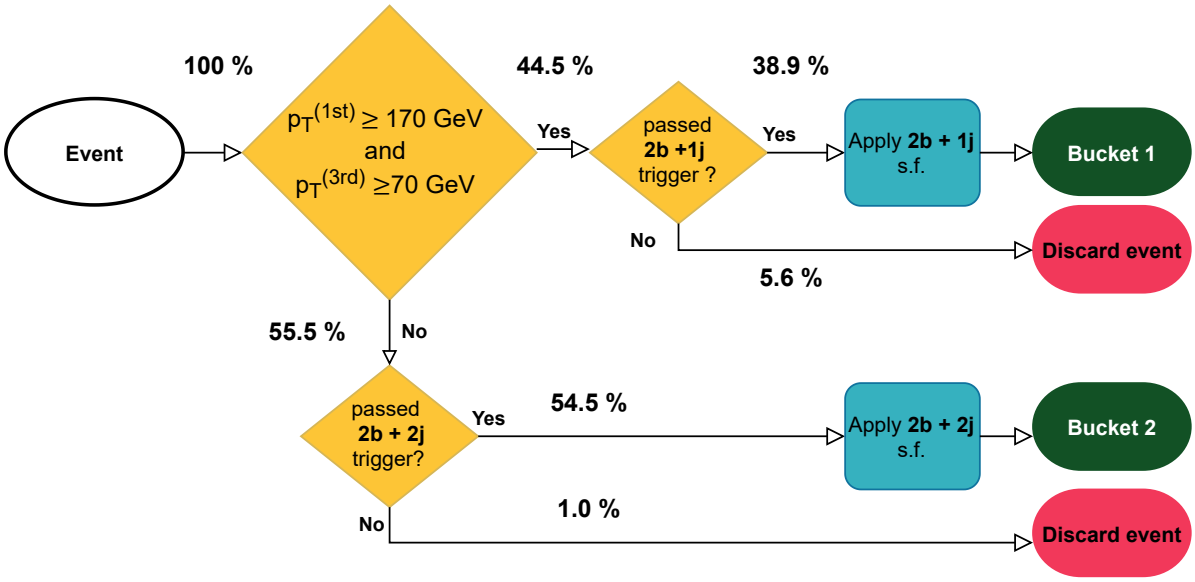
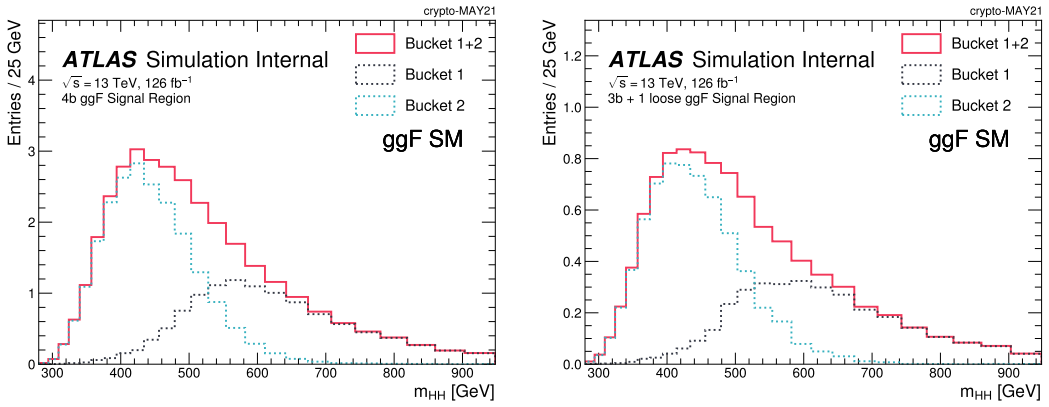


Figure 9.7.: Non-resonant trigger buckets diagram. Percentages indicate the fraction of $4b$ SM HH events in the signal region that pass or fail each stage of the trigger buckets. A total of 6.6 % of these signal events are discarded through this scheme.



(a) 4b **(b)** 3b1l

Figure 9.8.: SM HH signal region m_{HH} distribution of non-resonant buckets for both $4b$ and $3b1l$ categories.

Chapter 10.

Unblinded Results

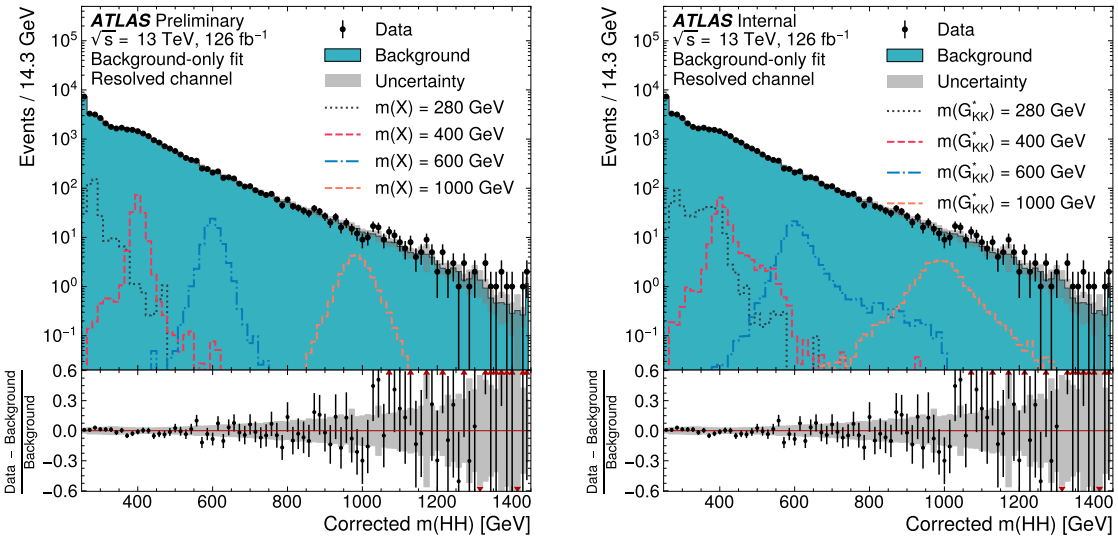
All of the necessary elements of the $HH \to b\bar{b}b\bar{b}$ search have been outlined and the unblinded results of both the resonant and non-resonant searches are presented in this chapter.

10.1. Resonant results

10.1.1. Distributions

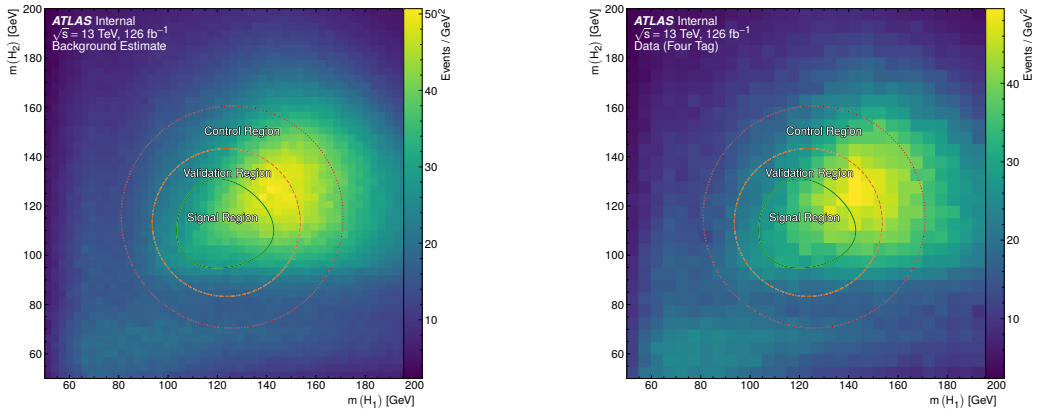
The corrected invariant HH mass, m_{HH}^{cor}, is used as the final discriminant of the resonant analysis. Figure 10.1 primarily shows the unblinded distributions of 4b data in the signal region, which is then compared with the background estimate and the signal samples for both scalars and gravitons. Additionally, a background-only likelihood fit is performed on the signal region distributions. In general the background modelling shows adequate agreement with the 4b data, except for the high mass tail regions. To complement these distributions the $m_{H1} - m_{H2}$ massplane for both the background estimate and the unblinded 4b events are shown in Figure 10.2. These massplanes indicate that the background estimate is able to replicate the structure of the 4b events. Tables 10.1 and 10.2, show the event yields for the 4b data, background estimate and a selection of signal hypothesis. The yields are integrated over a window in m_{HH}^{cor} that contains 90 % of each signal hypothesis. The ranges are typically larger for the spin-2 G_{KK}^{*} samples as these benchmark signals have a broader distribution than the spin-0 scalar samples. At low mass, near the kinematic threshold of 250 GeV, the wide nature of the spin-2 signals

result in extremely high mass tails as a consequence of the restricted phase space. This effect causes the double peaked structure of the 280 GeV spin-2 signal sample.



(a) spin-0 **(b)** spin-2

Figure 10.1.: Resonant search m_{HH}^{cor} signal region distributions. The background estimate is compared to the observed data after a background-only profile likelihood fit, which are in agreement with each other except for the high mass tail region. a) Shows the data along with the benchmark spin-0 signal samples for a few resonance masses. b) Shows the same resonance mass points but of the spin-2 benchmark signal instead.



(a) Reweighted 2b **(b)** 4b

Figure 10.2.: Massplanes of both the background estimate and the unblinded 4b category.

m(X) [GeV]	$\boldsymbol{m_{HH}^{cor}}$ **[GeV]**	**Data**	**Background model**	**spin-0 signal model**
260	[250, 321]	18 554	18 300 ± 110	503 ± 43
500	[464, 536]	2 827	2 866 ± 22	105.40 ± 5.70
800	[750, 850]	358	366.2 ± 7.3	37.70 ± 1.70
1200	[1079, 1250]	68	52.6 ± 1.7	11.71 ± 0.62

Table 10.1.: Resonant $4b$ signal region yields of data, background model and benchmark spin-0 signal hypothesis. The yields are determined by choosing an m_{HH}^{cor} window that contains 90 % of the spin-0 benchmark signal. The signal yield is normalized to the expected cross-section and the uncertainties are determined by summing the individual components in quadrature. A background-only fit to the data is used to evaluate the background yields and uncertainties [16].

m(G^*_{KK}) [GeV]	$\boldsymbol{m_{HH}^{cor}}$ **[GeV]**	**Data**	**Background model**	**spin-2 signal model**
260	[250, 393]	26 775	26 650 ± 130	368 ± 25
500	[464, 636]	4 655	4 719 ± 37	138.6 ± 5.70
800	[707, 950]	795	811 ± 13	52.1 ± 1.9
1200	[993, 1279]	146	120.6 ± 2.8	14.45 ± 0.67

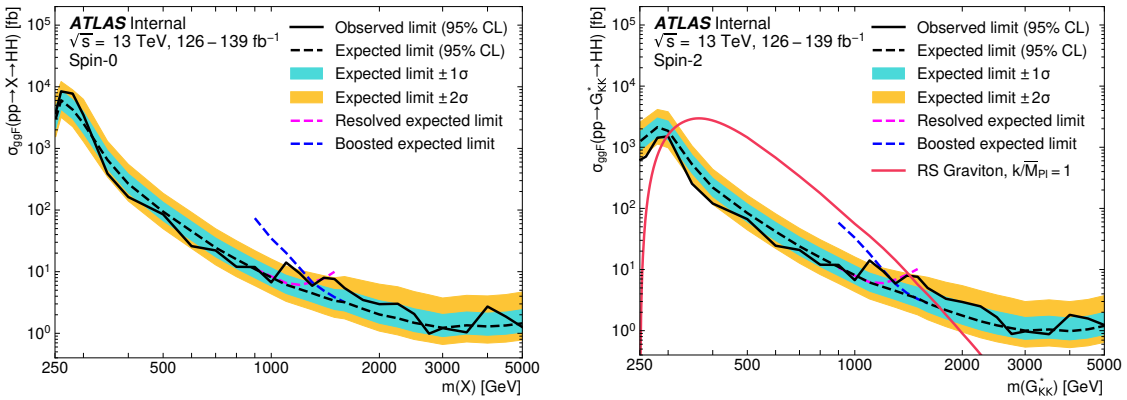
Table 10.2.: Resonant $4b$ signal region yields of data, background model and benchmark spin-2 signal hypothesis. The yields are determined by choosing an m_{HH}^{cor} window that contains 90 % of the spin-2 benchmark signal. The signal yield is normalized to the expected cross-section and the uncertainties are determined by summing the individual components in quadrature. A background-only fit to the data is used to evaluate the background yields and uncertainties [16].

Uncertainty Category	Relative Impact (%)			
	280 [GeV]	600 [GeV]	1600 [GeV]	4000 [GeV]
background m_{HH}^{cor} shape	12	8.7	1.3	1.9
Jet momentum/mass scale	0.6	0.1	1.5	1.5
Jet momentum/mass resolution	2.1	1.5	7.4	7.3
b-tagging calibration	0.7	0.4	1.8	7.6
Theory (signal)	0.6	0.6	1.6	1.0
Theory ($t\bar{t}$ background)	N/A	N/A	0.7	0.4
All systematic uncertainties	16	11	13	17

Table 10.3.: Impact of uncertainties on the expected 95 % CL upper limit for the spin-0 cross-section. These are defined by the decrease in the expected limit when the nuisance parameter is fixed to its best-fit value. The uncertainties at low masses of 280 GeV and 600 GeV, which correspond to the resolved analysis, are dominated by the background estimation uncertainty. Masses of 1600 GeV and 4000 GeV correspond to the boosted topology, in which the background estimate does not significantly contribute to the systematic uncertainties relative to the jet resolution and b-tagging calibrations.

10.1.2. Resonant search limits

Due to the various masses of the signal hypothesis in the resonant search, the 95 % upper limits on the cross-section are parametrized as a function of the mass of the benchmark signal. Local p-values are determined by the CL_S method using the background only hypothesis as the null hypothesis as outlined Equation 4.9 in Chapter 4. The fits are obtained by using 84 equally spaced bins from 250 GeV to 1450 GeV in m_{HH}^{cor} and each year is fitted as a separate channel. The results of these fits for both spin-0 and spin-2 benchmark signals are shown in Figure 10.3. To cover the full spectrum of resonance mass, the resolved channel and boosted channels are combined to produce the final limit. Previously an excess with a global significance of 2.5σ was observed in the early Run-2 spin-0 scalar limits at 280 GeV. In the full Run-2 limits, this excess vanishes as the global significance lowers down to 0.2σ. A new excess is now present at the 1100 GeV mass point for both the spin-0 and spin-2 benchmark signals. These have a global (local) significance of 1.0 (2.6) and 1.2 (2.7) for the spin-0 and spin-2 signals respectively. Table 10.3 shows how each source of uncertainty contributes to the expected limit of the spin-0 signal.



(a) spin-0 **(b)** spin-2

Figure 10.3.: Expected and observed 95 % CL upper limits on the cross-section of HH production via the benchmark signal samples a) spin-0 scalar and b) spin-2 G^*_{KK}. The coloured bands indicate the $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals on the expected limits. These results are obtained by the combination of resolved and boosted channels. The contribution of each channel is shown as coloured dashed lines. The resolved channel covers a mass range of 251 to 1500 GeV while the boosted channel starts at 900 GeV. Events are given priority to the resolved channel if they are selected by both analyses. The theoretical cross-section as a function of $m(G^*_{\mathrm{KK}})$ for the bulk RS model with $k/\overline{M}_{Pl} = 1$ is shown by the solid red line.

10.1.3. Comparison with other channels

To comprehend how the $b\bar{b}b\bar{b}$ search performs, the full Run-2 limit is compared with the other final states in Figure 10.4. The comparison shows the full Run-2 limits for the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$ (resolved and boosted) and $b\bar{b}\gamma\gamma$ final states. Additionally, the previous combination using the early Run-2 searches is overlaid for comparison. In the low $m(X)$ range the $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$ outperform the $b\bar{b}b\bar{b}$. This is unsurprising since the background yield is very high at low masses the $b\bar{b}b\bar{b}$ channel relative to $b\bar{b}\gamma\gamma$ and $b\bar{b}\tau^+\tau^-$ channels. Moreover, the low m_{HH} region is particularly difficult to model and leads to large shape uncertainties. These aspects diminish the sensitivity in the low m_{HH} region. However in the mass region above 600 GeV, the $b\bar{b}b\bar{b}$ channel begins to outperform the other channels. This is where the $b\bar{b}b\bar{b}$ channel will have the strongest contribution to the full Run-2 combined limits.

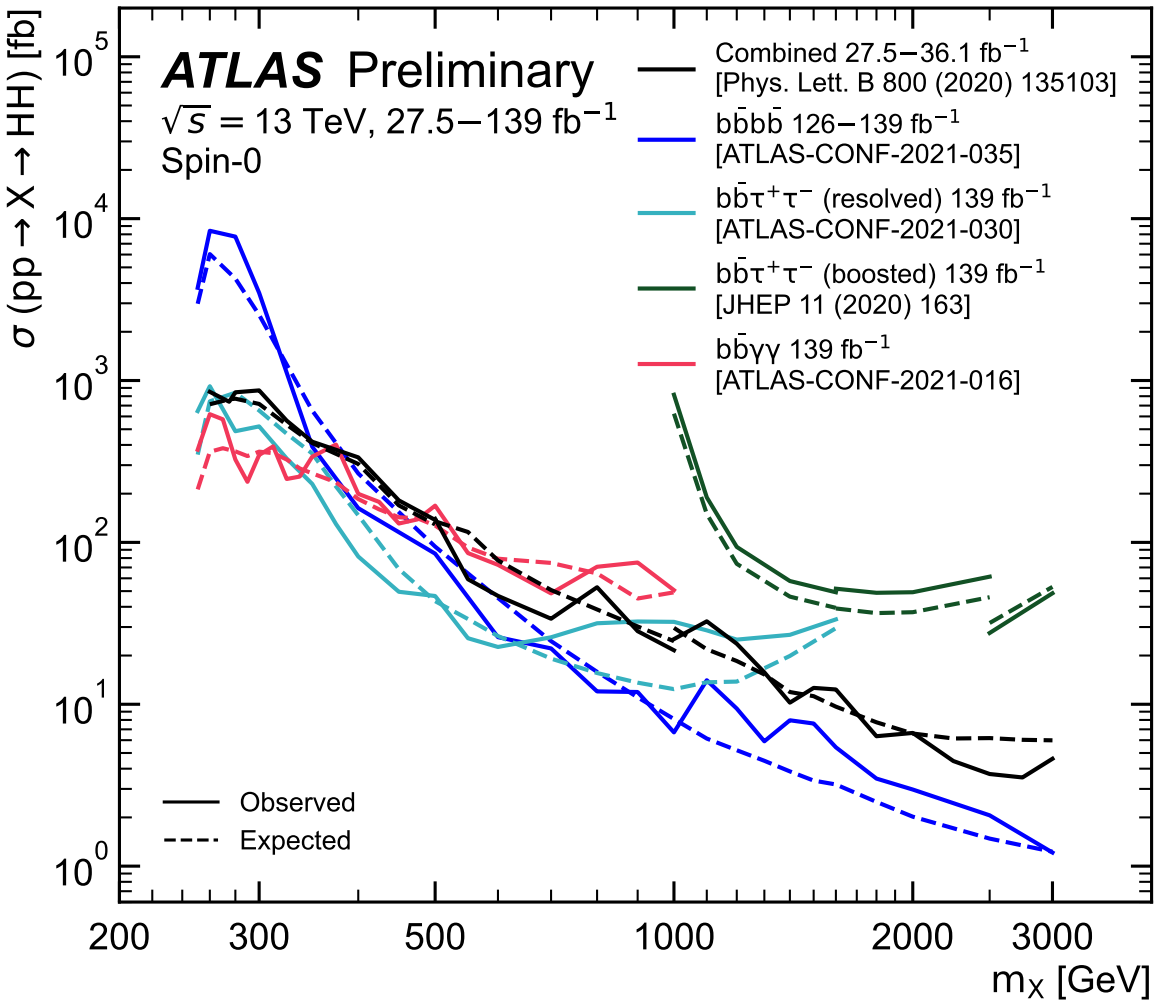
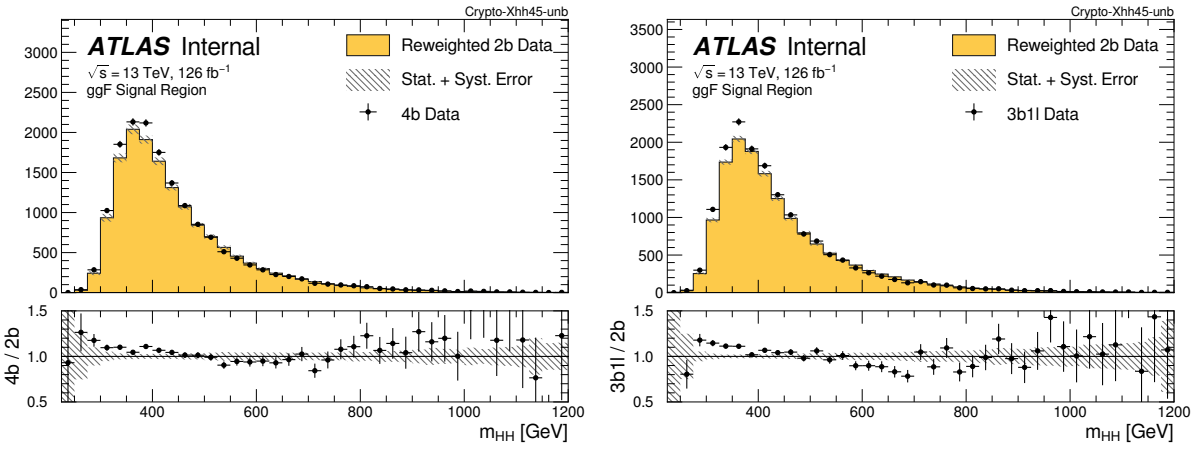


Figure 10.4.: Upper limits at 95 % CL on resonant HH production cross-section as function of the spin-0 scalar resonance mass. The combined early Run-2 results are overlaid for comparison. [105]

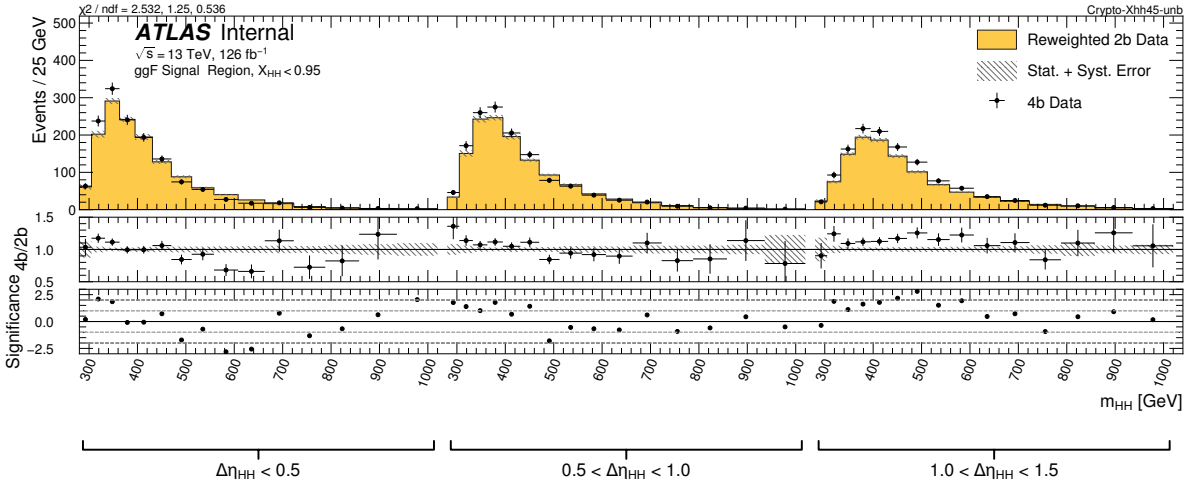
10.2. Non-resonant results

The invariant HH mass, m_{HH}, is used as the final discriminant in the non-resonant search. The unblinded signal region m_{HH} distributions of both $4b$ and $3b1l$ data is shown in Figure 10.5, which also has the background estimate overlaid. In both $4b$ and $3b1l$ regions, there appears to be a consistent underestimation of the level of background compared with what we observe in the data in the peak of the distribution that is not covered by the uncertainties. Furthermore there is a deficit in the 700 GeV region in $3b1l$. The two signal regions, $4b$ and $3b1l$, are split into 6 categories by using 2 X_{HH} bins and 3 $|\Delta\eta_{HH}|$ bins. Figures 10.6 and 10.7 the same unblinded distributions as shown previously, but in each of the categories used instead. These distributions indicate that the discrepancies seen in Figure 10.5 are consistent throughout all the categories. The massplanes of both the background estimate and the $4b$ region are shown in Figure 10.8

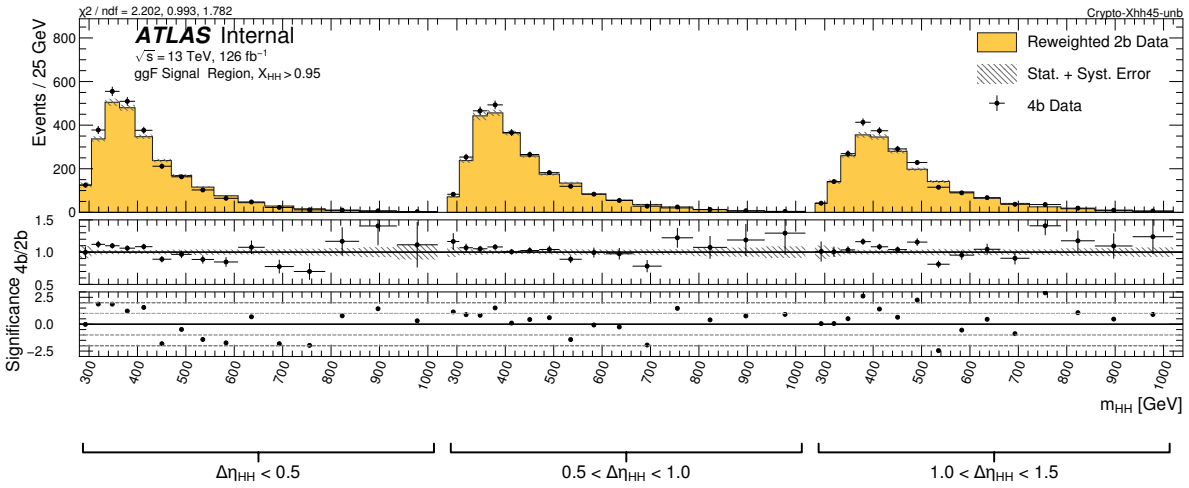


(a) $4b$ **(b)** $3b1l$

Figure 10.5.: Inclusive signal region m_{HH} distribution for both $4b$ and $3b1l$ categories. Distributions are obtained by combining all years together.

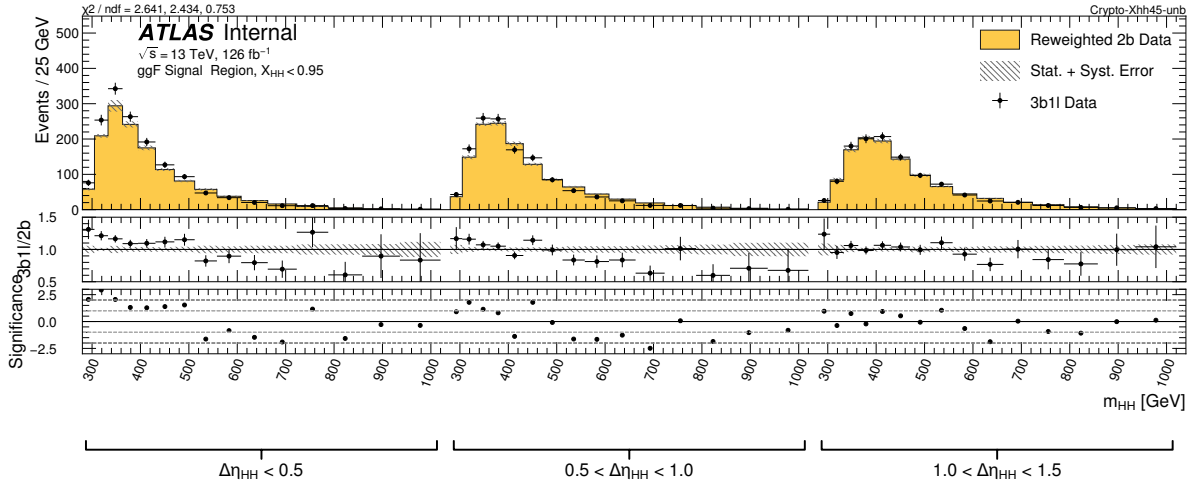


(a) $X_{HH} < 0.95$

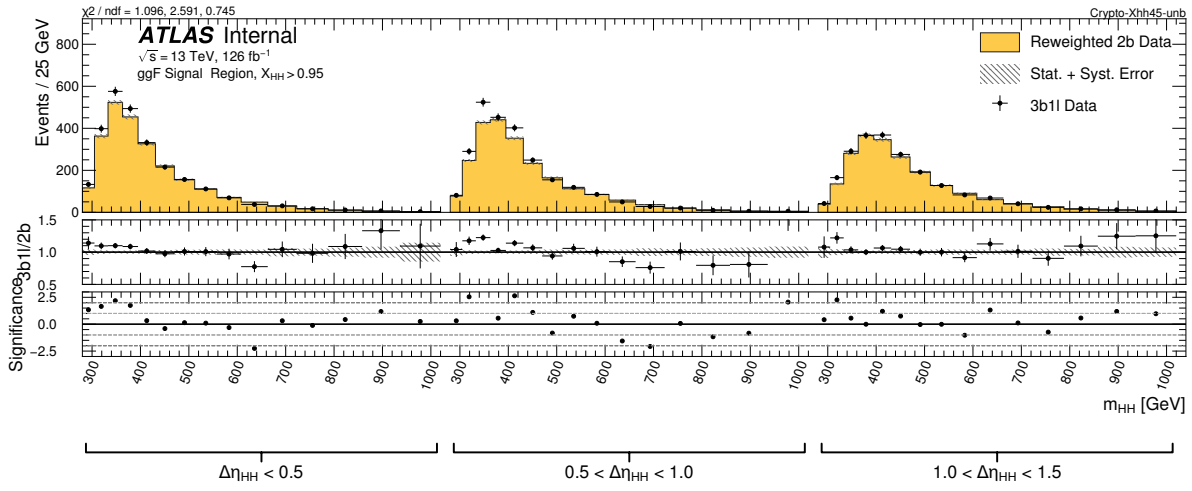


(b) $X_{HH} > 0.95$

Figure 10.6.: Signal region m_{HH} distributions for $4b$ data and the background estimate, split into the different categories. a) Shows the three $|\Delta\eta_{HH}|$ categories for the inner signal region of $X_{HH} < 0.95$, while b) shows the $|\Delta\eta_{HH}|$ categories for the outer signal region $X_{HH} > 0.95$.



(a) $X_{HH} < 0.95$



(b) $X_{HH} > 0.95$

Figure 10.7.: Signal region m_{HH} distributions for $3b1l$ data and the background estimate, split into the different categories. a) Shows the three $|\Delta\eta_{HH}|$ categories for the inner signal region of $X_{HH} < 0.95$, while b) shows the $|\Delta\eta_{HH}|$ categories for the outer signal region $X_{HH} > 0.95$.

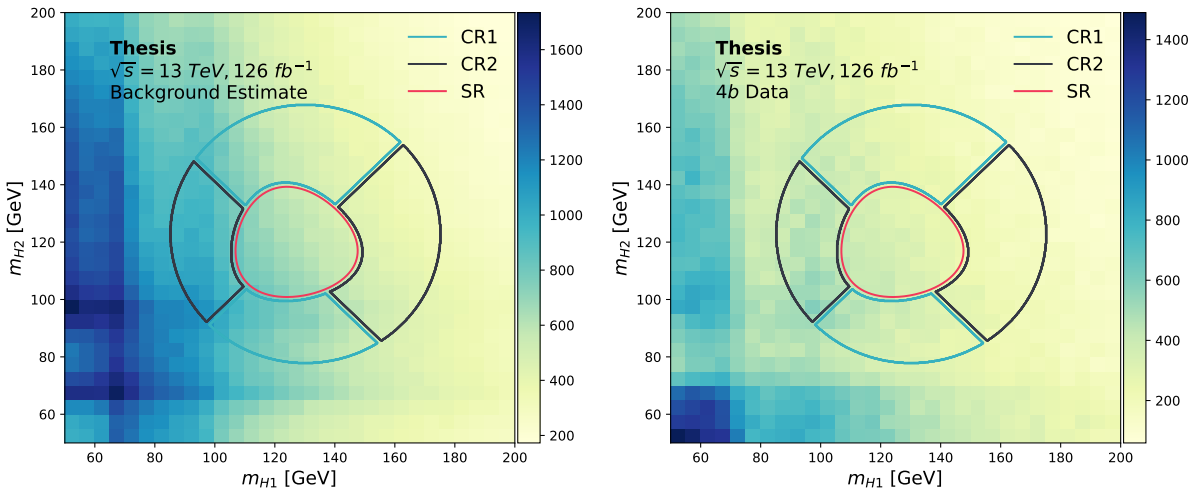


Figure 10.8.: Non-resonant search: Massplanes for the a) background estimate and b) $4b$ category.

10.2.1. Non-resonant search limits

The limits obtained on the Standard Model cross-section is shown in Table 10.4. The limits were obtained by only including the background uncertainties, since the uncertainties on the signal process are expected to have a minimal effect. The observed and expected limits on the parameter of interest, μ, are 5.1 and 5.8 respectively. The observed limit is well within the uncertainty bands of the expected limit. The early Run-2 $b\bar{b}b\bar{b}$ search obtained an observed and expected limits of 12.9 and 21 respectively, while only using 27.5 fb^{-1} of luminosity. Through luminosity scaling, the expected limits should lower down to 9.8 for the full 126.7 fb^{-1} Run-2 dataset. The optimizations made in the $b\bar{b}b\bar{b}$ search have significantly improved the results for the Standard Model point where the expected limit improved by 40 %.

Similarly, the limits as a function of the Higgs coupling parameter, κ_λ, are shown in Figure 10.9. The κ_λ scan in Figure 10.9, shows a significant deviation between the expected and observed limits for values above $\kappa_\lambda = 5$. The $\kappa_\lambda > 5$ signal variations correspond to soft m_{HH} spectrum which is where the analysis has poor sensitivity as well difficulties in modelling the background as seen in Figure 10.5. As a result, the observed constraint on the Higgs coupling for the $b\bar{b}b\bar{b}$ channel is $\kappa_\lambda^{observed} \in [-6.0, 15.0]$. The Standard Model prediction of the trilinear self-coupling, $\lambda_{SM}^{HHH} = 0.13$, which means that the constraints on λ^{HHH} are [-0.78, 1.95]. The constraints obtained from the full Run-2 search show a significant improvement over the early Run-2 $b\bar{b}b\bar{b}$ search which constrained this parameter within [-10.9, 20.1].

	Observed	**-2 σ**	**-1 σ**	**Expected**	**+1 σ**	**+2 σ**
$\boldsymbol{\mu}$	5.1	3.1	4.2	5.8	8.1	10.9
$\boldsymbol{\sigma}$ **[fb]**	158	96	127	180	252	338

Table 10.4.: Non-resonant search, upper limits on the Standard Model, ggF, $HH \to b\bar{b}b\bar{b}$ production. The limits are shown for the parameter of interest, μ, which is defined as the ratio σ/σ_{SM}, where $\sigma_{SM} = 31.05\,\text{fb}$. The limits are also shown for the cross-section directly. These results do not include uncertainties on the signal process.

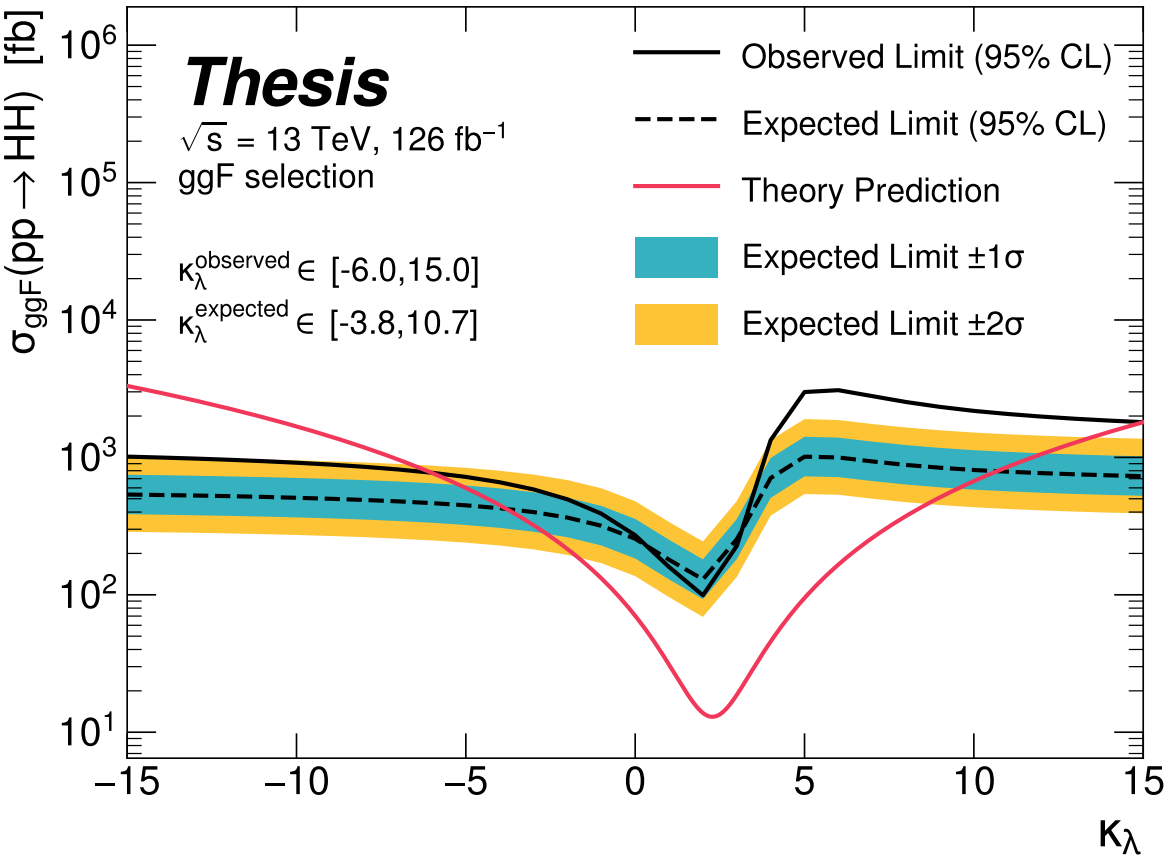


Figure 10.9.: Non-resonant search 95 % CL upper limits on the HH ggF cross-section as a function of the Higgs self-coupling parameter, κ_λ. The dashed black line represents the expected limit, while the solid black line shows the observed limit. The coloured bands represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties on the expected limit. The expected constraint on the κ_λ parameter is [-3.8, 10.7], while the observed constraint is [-6.0, 15.0]. The observed (expected) cross-section limit for the SM process, $\kappa_\lambda = 1$, is 158 fb (180 fb).

10.2.2. Comparison with other channels

The results obtained for the non-resonant search can be compared with the full Run-2 results of other HH channels, such as, $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$. The results of the SM HH process are shown in Figure 10.10. There are some slight differences with the results presented in Figure 10.10 and those presented in this thesis. The first major difference is that the $b\bar{b}b\bar{b}$ analysis is only able to use 126.7 fb^{-1} of data rather than the full 139 fb^{-1} Run-2 dataset, which is due to the usage of b-jet triggers. The loss in luminosity corresponds to a 5 % higher expected limit in the $b\bar{b}b\bar{b}$ search relative to $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$. Secondly, $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$ normalized their results to the combined cross-section of ggF and VBF production modes, $\sigma_{ggF+VBF}$. Lastly, due to technical difficulties, the $b\bar{b}b\bar{b}$ results do not include systematic uncertainties on the signals. These signal uncertainties were seen to have a small effect in the resonant search relative to the large background uncertainty and hence it may be ignored in the non-resonant search.

Additionally, the $b\bar{b}\gamma\gamma$ channel has produced a κ_λ parameter scan which is shown in Figure 10.11. The $b\bar{b}\gamma\gamma$ channel has a significant sensitivity in the low m_{HH} range which helps to constrain the κ_λ parameter. The observed κ_λ constraint from the $b\bar{b}\gamma\gamma$ channel is [-1.5, 6.7], which is a significant improvement over the results obtained from $b\bar{b}b\bar{b}$ in this thesis, [-6.0, 15.0]. Furthermore the deviations between expected and observed limits in the $b\bar{b}b\bar{b}$ results presented in Figure 10.9 are not seen in the $b\bar{b}\gamma\gamma$ channel. The results from the $b\bar{b}\gamma\gamma$ channel reinforces the assertion that the deviations seen in Figure 10.9 originate from the poor modelling of the low m_{HH} region, which gives unreliable estimates of the expected limit.

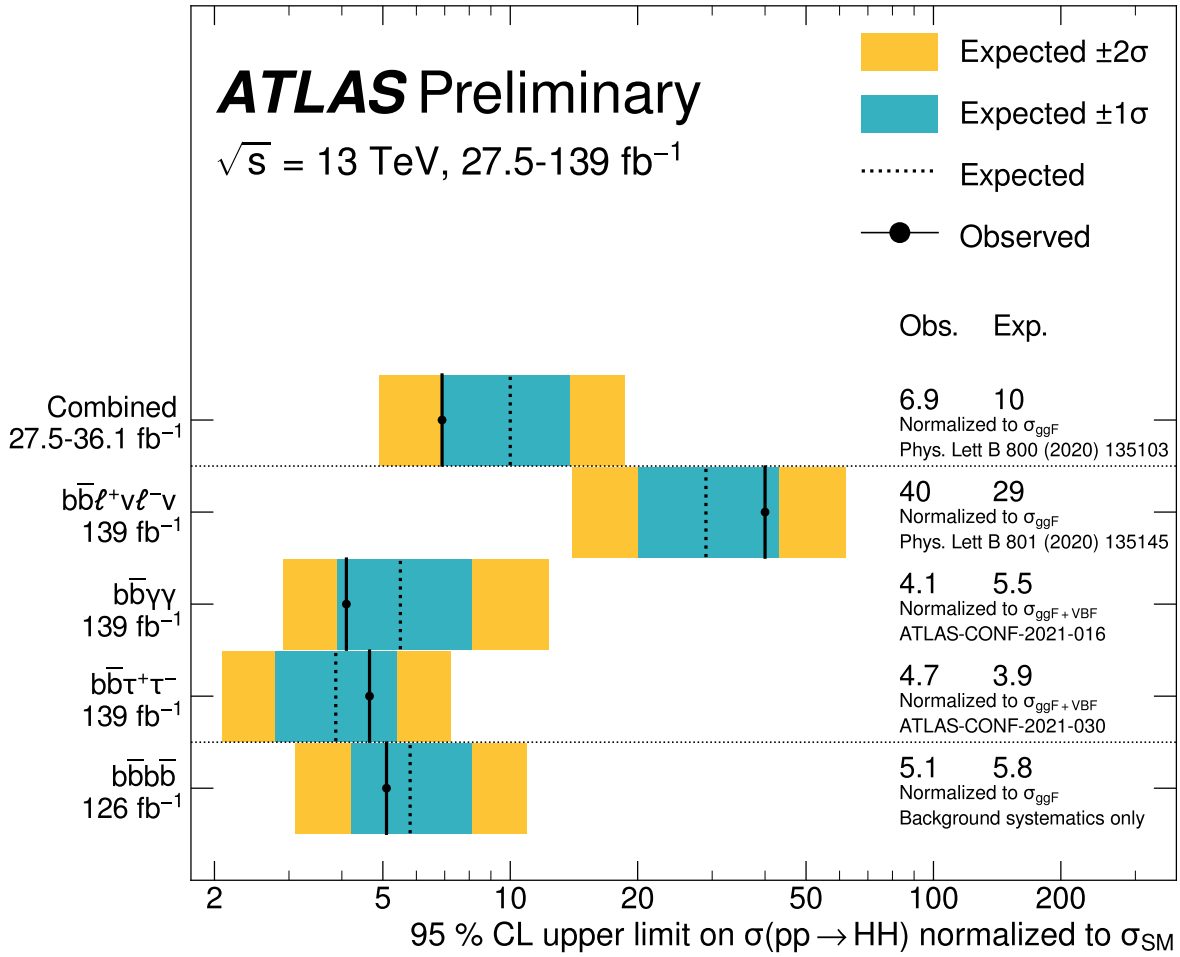


Figure 10.10.: Upper limits at 95 % CL on the non-resonant HH production cross-section relative to the SM prediction. The early Run-2 combined results are shown for comparison. $b\bar{b}\gamma\gamma$ and $b\bar{b}\tau^{+}\tau^{-}$ are normalized to the cross section of both ggF and VBF production mode. Figure edited from [105].

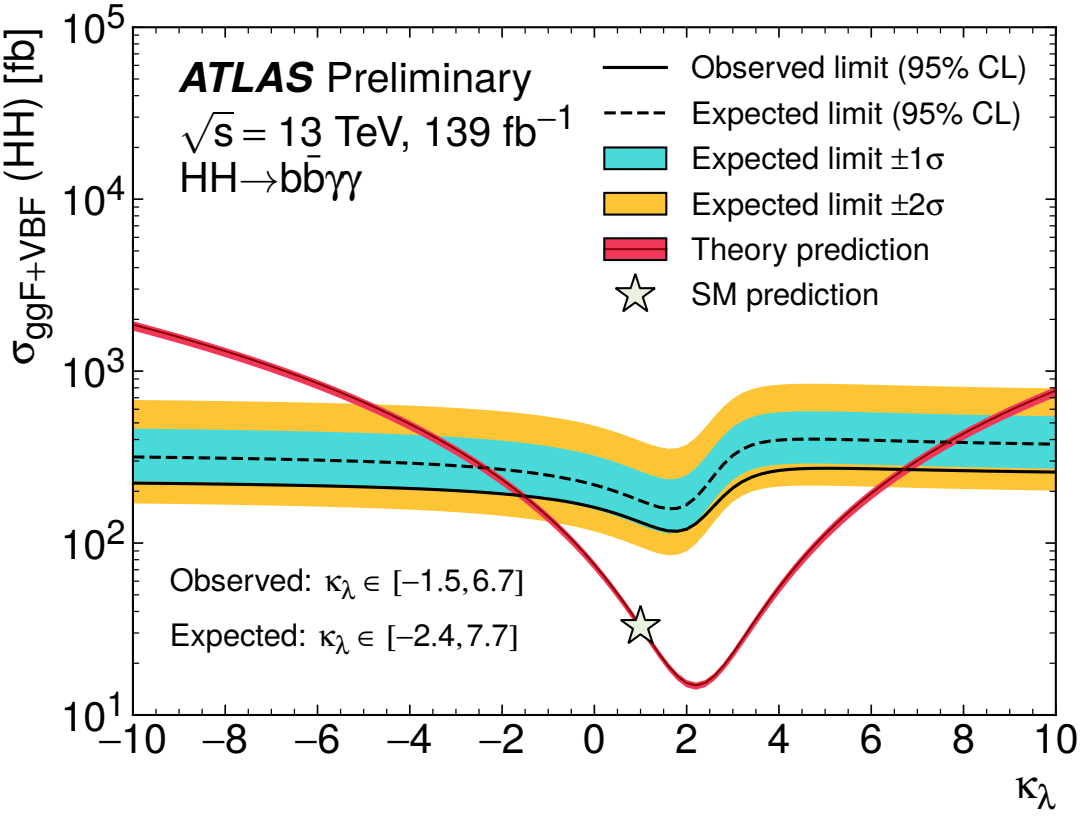


Figure 10.11.: Observed and expected limits at 95 % CL for the non-resonant cross-section of HH production as a function of κ_λ obtained by the $b\bar{b}\gamma\gamma$ search using the full Run-2 data. The cross-section includes both ggF and VBF production modes [30].

Chapter 11.

Conclusion

In summary, the Standard Model of particle physics, while hugely successful, is imperfect and extensions are needed to address some of the fundamental questions of nature. The pair-production of Higgs Bosons, HH, can be used to probe deeper into the structure of the theory. This thesis covers the search of the HH decaying to four bottom quarks as this has the largest branching ratio. Unfortunately, the $b\bar{b}b\bar{b}$ final state has experimental signatures that are difficult to deal with and the background estimate presents a major challenge to the search. In this thesis, the full Run-2 data recorded by the ATLAS detector at the Large Hadron Collider was used for the $HH \to b\bar{b}b\bar{b}$ search. The total integrated luminosity of this dataset is 126 fb^{-1}.

The resonant search was used to look for BSM physics by searching for heavy resonances. No significant excess has been observed in this search. Remarkably, the excess at 280 GeV seen in the early Run-2 search has disappeared. The largest excess had a global significance of 1.0 and 1.2 for the spin-0 and spin-2 signals respectively. Further investigation, perhaps by combining with the other HH channels, is needed to see if this excess increases. The non-resonant search was used to constrain the values the Higgs trilinear self-coupling could take. The constraints on the κ_λ parameter were observed to be [-6.0, 15.0]. The strategy of this search was optimized for the Standard Model, $\kappa_\lambda = 1$, signal as this is where the $b\bar{b}b\bar{b}$ search can have competitive limits relative to the other HH channels. The upper limit on the observed (expected) signal strength of the ggF, SM HH, process was 5.1 (5.8) times the theoretical Standard Model cross-section. This corresponds to an upper limit on the observed (expected) cross-section of 158 fb (180 fb).

The non-resonant search was optimized for the Standard Model signal, which resulted in a 40 % improvement in the expected results over the early Run-2 search after luminosity scaling. These optimizations were needed to make the $b\bar{b}b\bar{b}$ search competitive with respect

to $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$. The lack of sensitivity and poor modelling in the low m_{HH} region negatively impacts the constraints on the κ_λ coupling parameter, which is considerably lower than the results seen from the $b\bar{b}\gamma\gamma$ channel. Furthermore, the observed results show deviations from the expected results when the signal hypothesis has a value of $\kappa_\lambda > 5$. These deviations are the result of the underestimation of the background yields and uncertainties in the low m_{HH} region. These deviations are not seen in the κ_λ scan obtained from the $b\bar{b}\gamma\gamma$ channel, which has a stronger sensitivity in this region.

The $HH \to b\bar{b}b\bar{b}$ can be primarily improved by reducing the extrapolation uncertainty from the background estimate, which is largest in the low m_{HH} region. Secondly, an improvement to the low m_{HH} yields of b-jet triggers will also improve the analysis. These improvements will in turn likely improve the constraints on the Higgs self coupling. The research from this thesis has provided valuable insight into the $HH \to b\bar{b}b\bar{b}$ search. This insight is being used to design future strategy of the analysis and event reconstruction algorithms. These developments are necessary to maximize the potential of HL-LHC data and potentially surpass the current projections of HH searches.

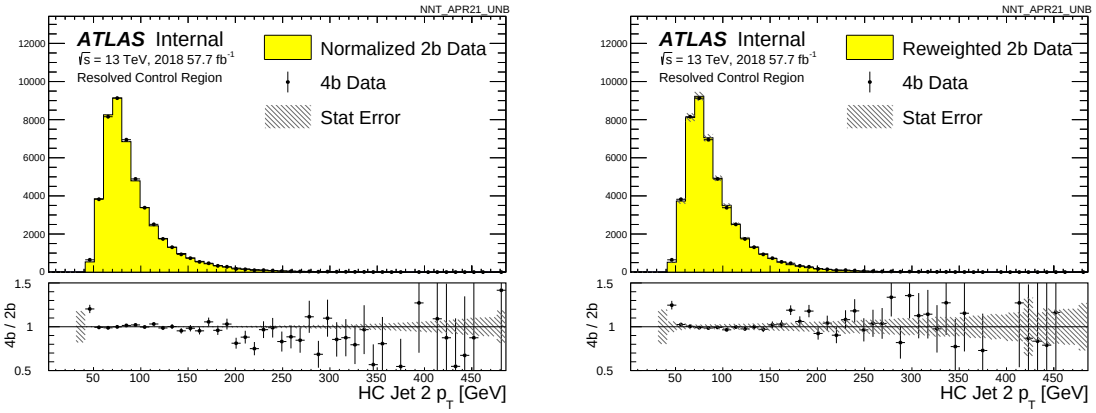
Appendix A.

Background estimate reweighting

This section shows the key distributions, that were not shown in the main body, and are used to derive the background model in both the resonant and non resonant search. Section A.1 shows the distributions used in the resonant search. Sections A.2 and A.3 show the $4b$ and $3b1l$ distributions of non-resonant search respectively. In general the performance of the reweighting model is adequate across these distributions.

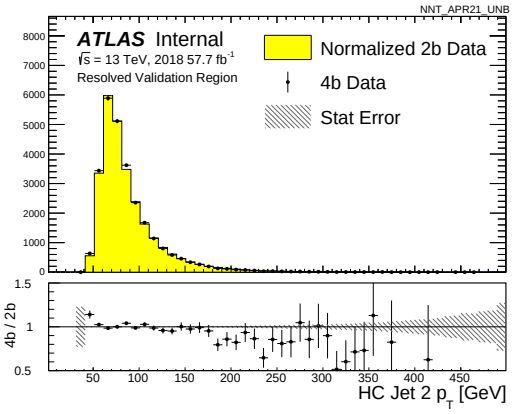
A.1. Resonant reweighting

Figures A.1-A.7 show performance of the reweighting model in both control and validation regions of the resonant search.

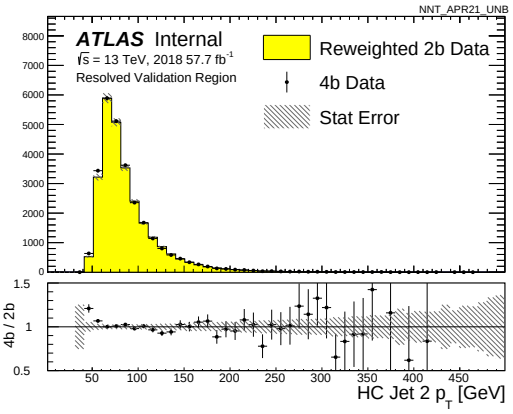


(a) Before Reweighting

(b) After Reweighting

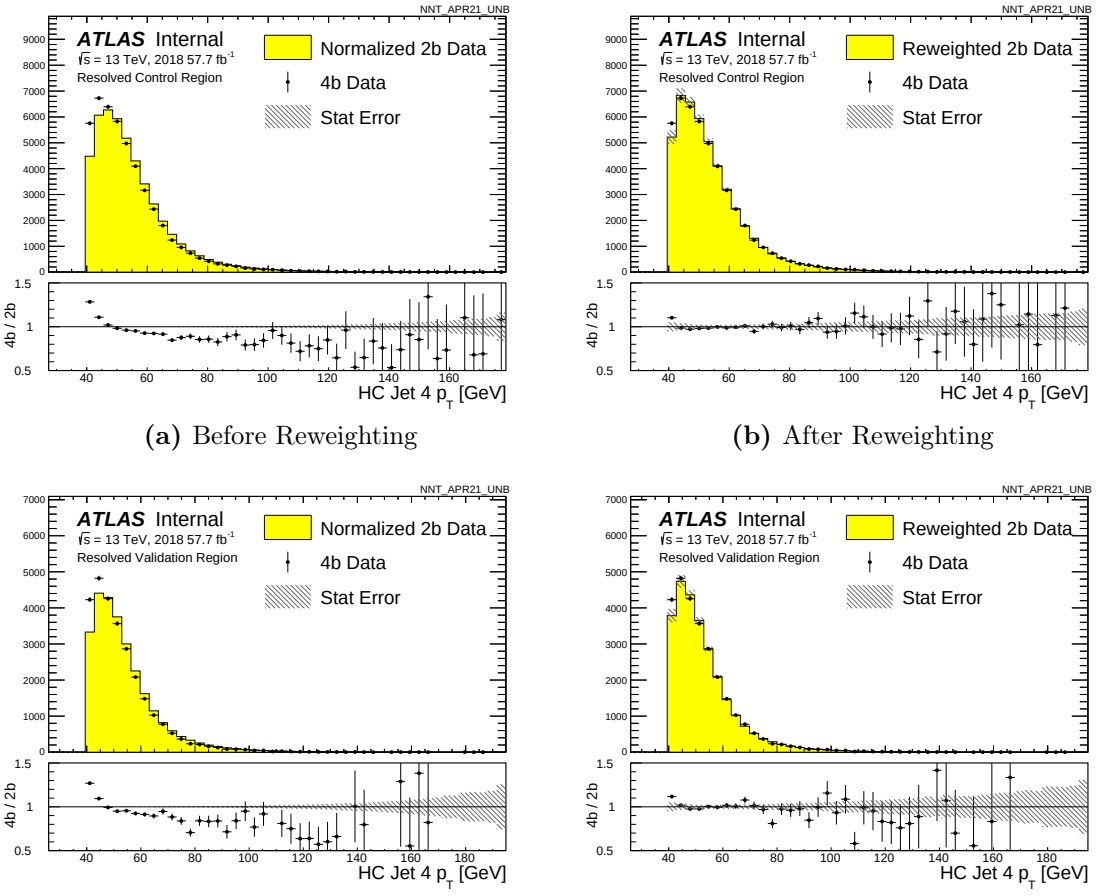


(c) Before Reweighting



(d) After Reweighting

Figure A.1.: Resonant p_{T_2} distributions before and after reweighting in both Control and Validation regions.



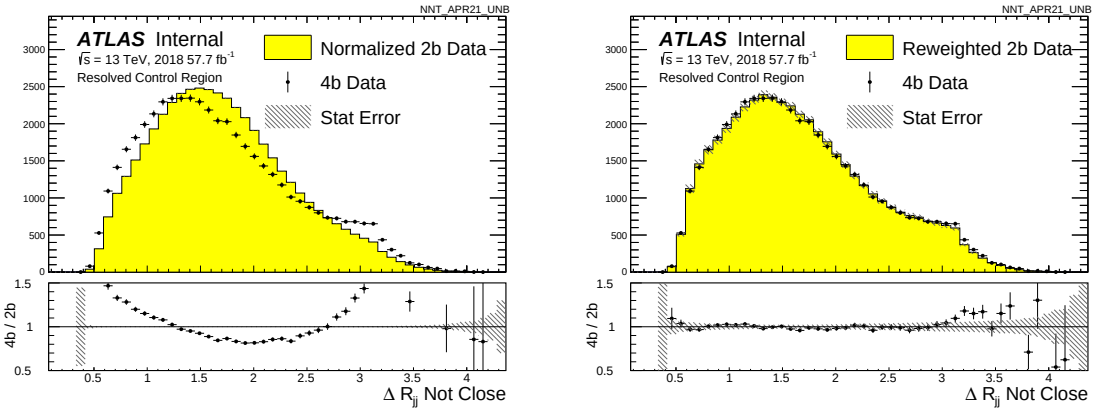
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

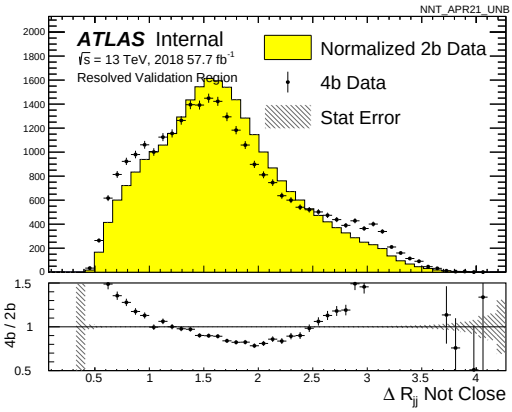
(d) After Reweighting

Figure A.2.: Resonant p_{T_4} distributions before and after reweighting in both Control and Validation regions.

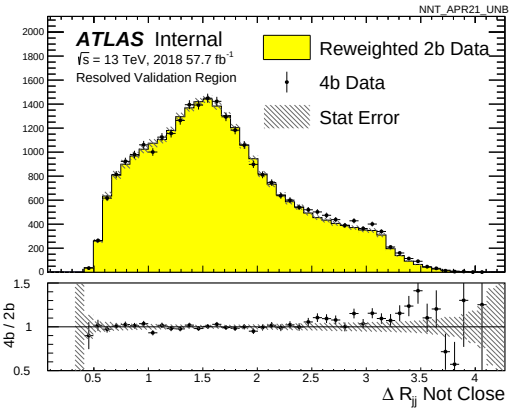


(a) Before Reweighting

(b) After Reweighting

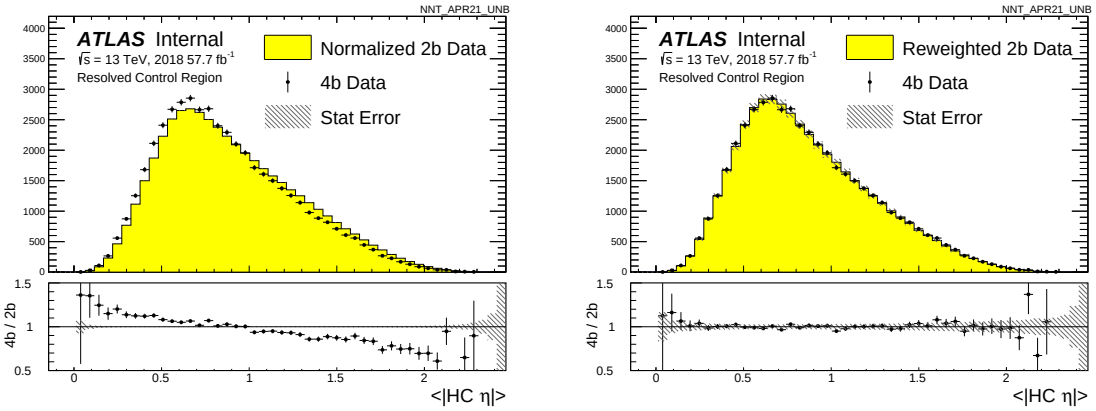


(c) Before Reweighting



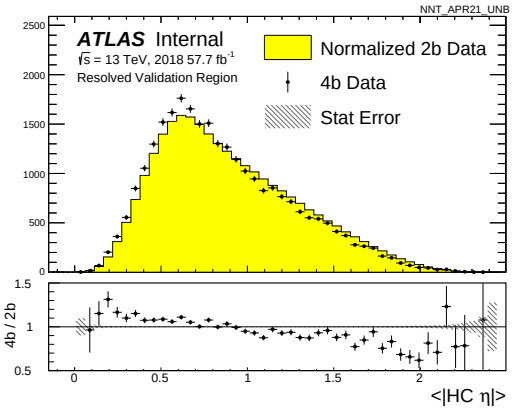
(d) After Reweighting

Figure A.3.: Resonant ΔR_{jj} Not Close distributions before and after reweighting in both Control and Validation regions.

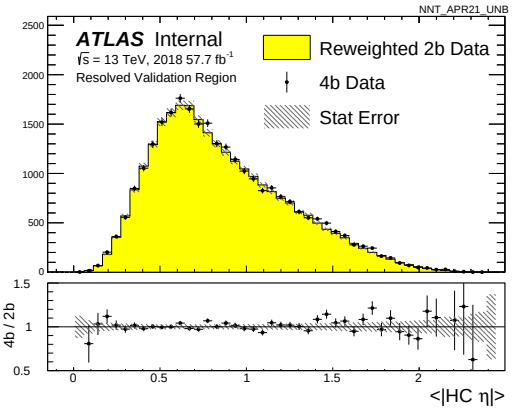


(a) Before Reweighting

(b) After Reweighting

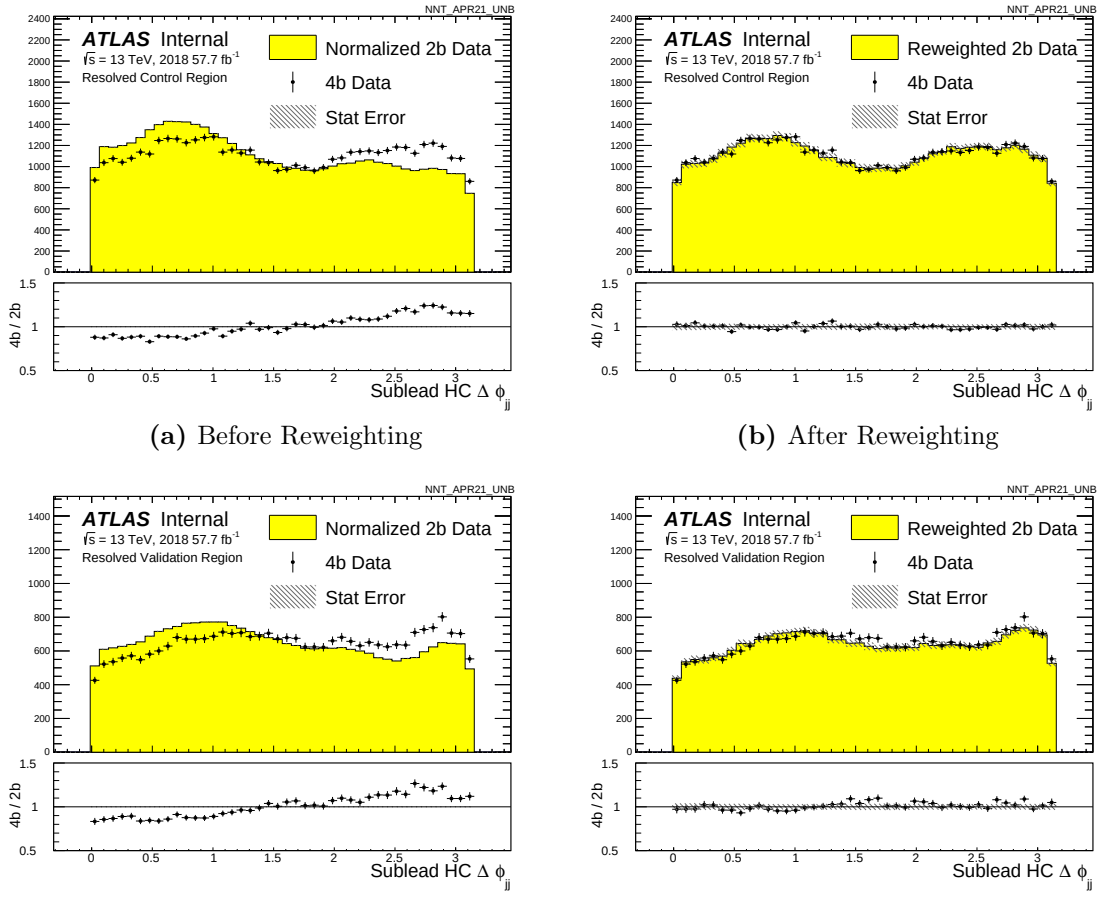


(c) Before Reweighting



(d) After Reweighting

Figure A.4.: Resonant $\langle |HC\eta| \rangle$ distributions before and after reweighting in both Control and Validation regions.



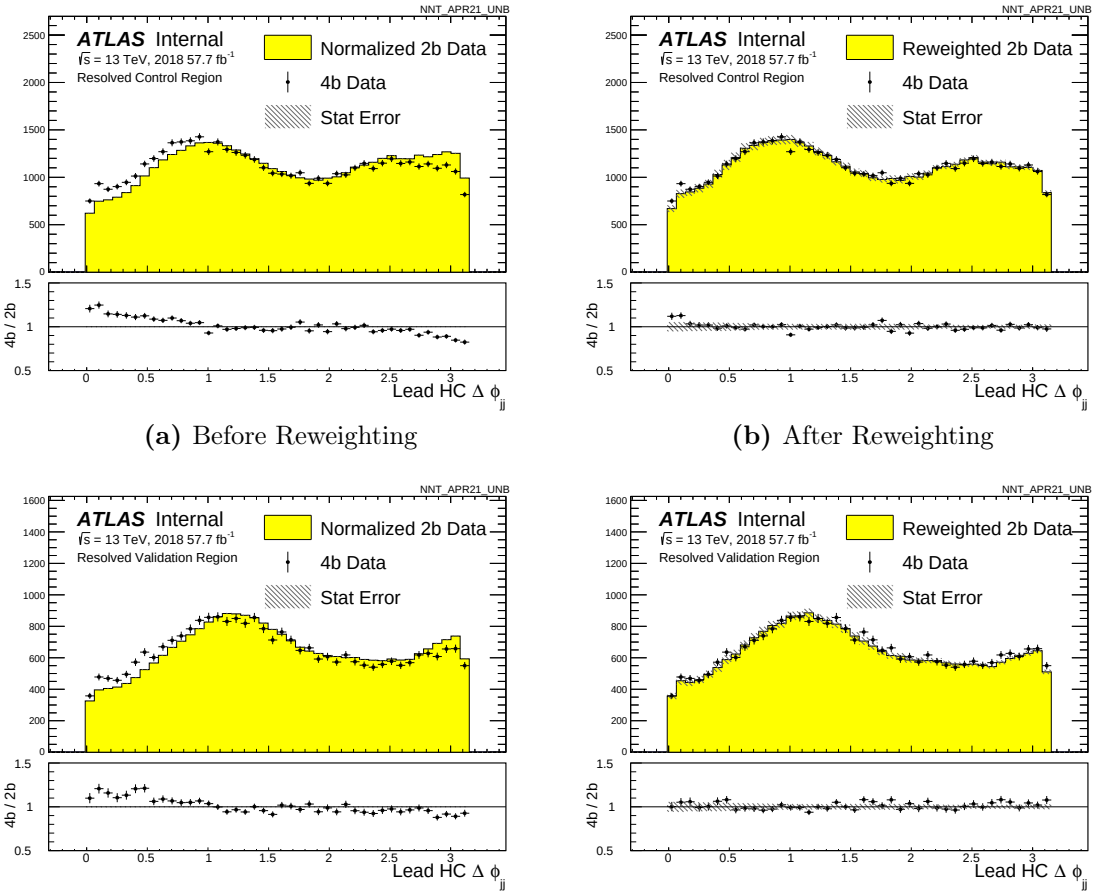
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

(d) After Reweighting

Figure A.5.: Resonant Sublead HC $\Delta\phi_{jj}$ distributions before and after reweighting in both Control and Validation regions.



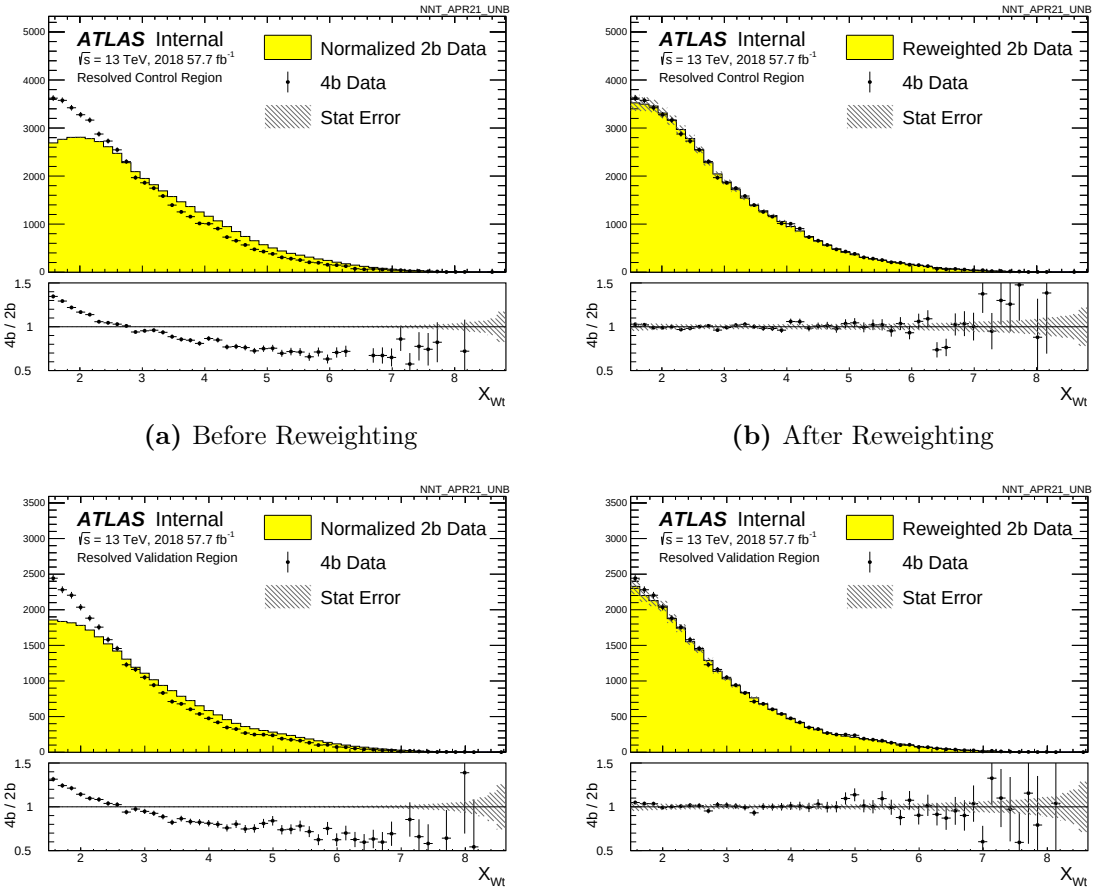
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

(d) After Reweighting

Figure A.6.: Resonant Lead HC $\Delta\phi_{jj}$ distributions before and after reweighting in both Control and Validation regions.



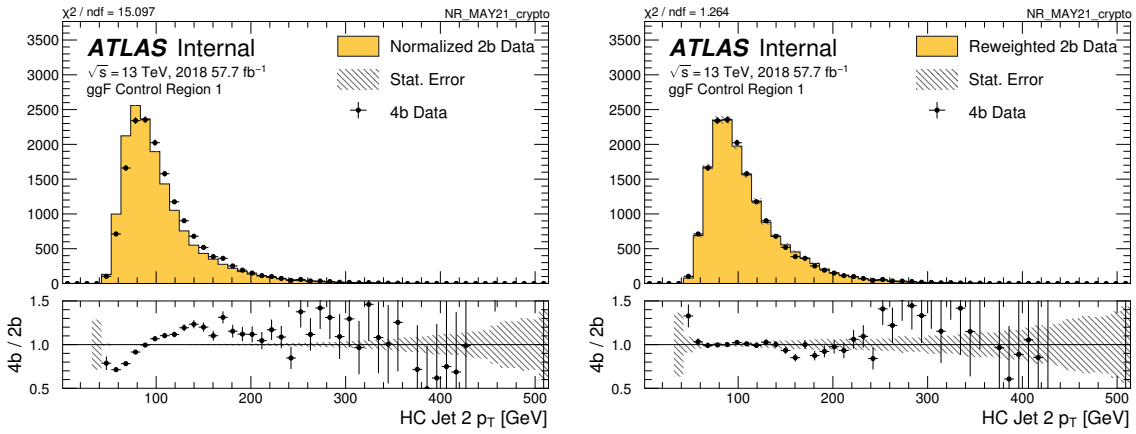
(a) Before Reweighting **(b)** After Reweighting

(c) Before Reweighting **(d)** After Reweighting

Figure A.7.: Resonant X_{Wt} distributions before and after reweighting in the $4b$ in both Control and Validation regions.

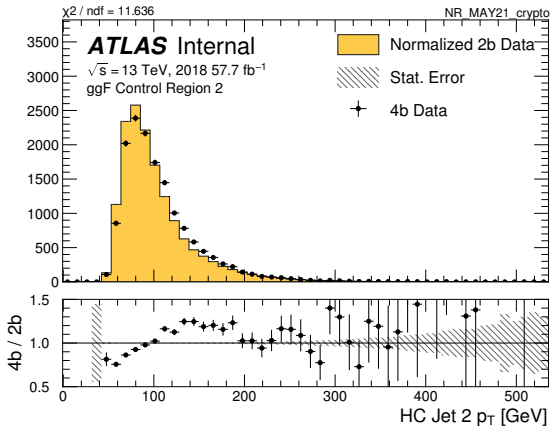
A.2. Non-resonant reweighting 4*b*

Figures A.8-A.14 show performance of the reweighting model of the 4*b* target in both control and validation regions of the non-resonant search.

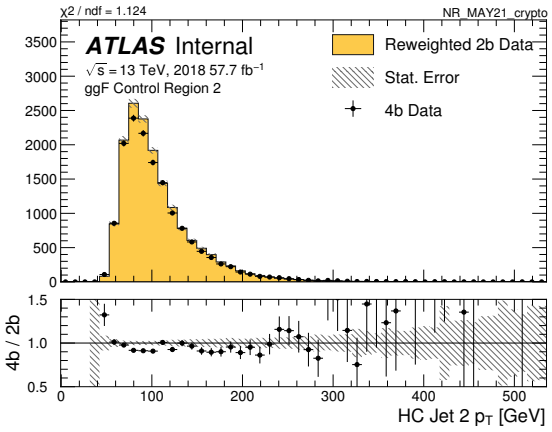


(a) Before Reweighting

(b) After Reweighting

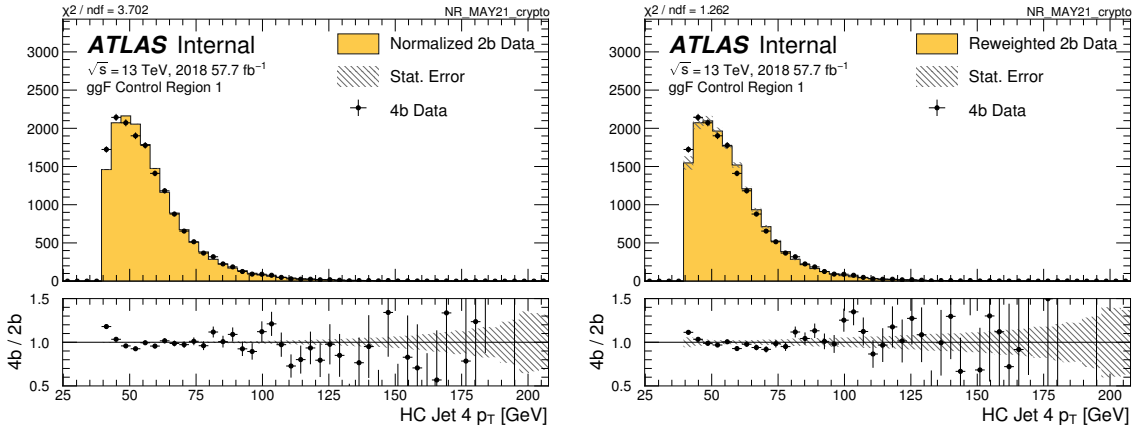


(c) Before Reweighting

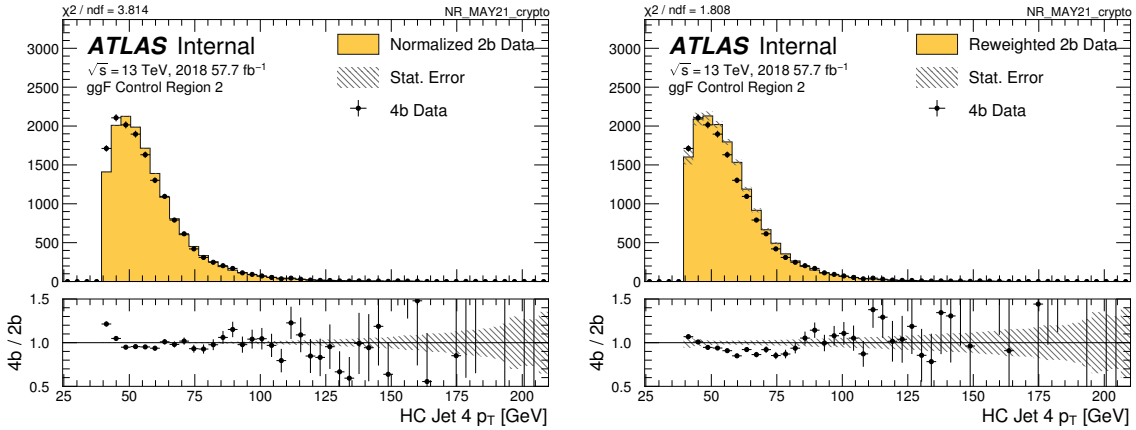


(d) After Reweighting

Figure A.8.: Non-resonant 4*b* p_{T_2} distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

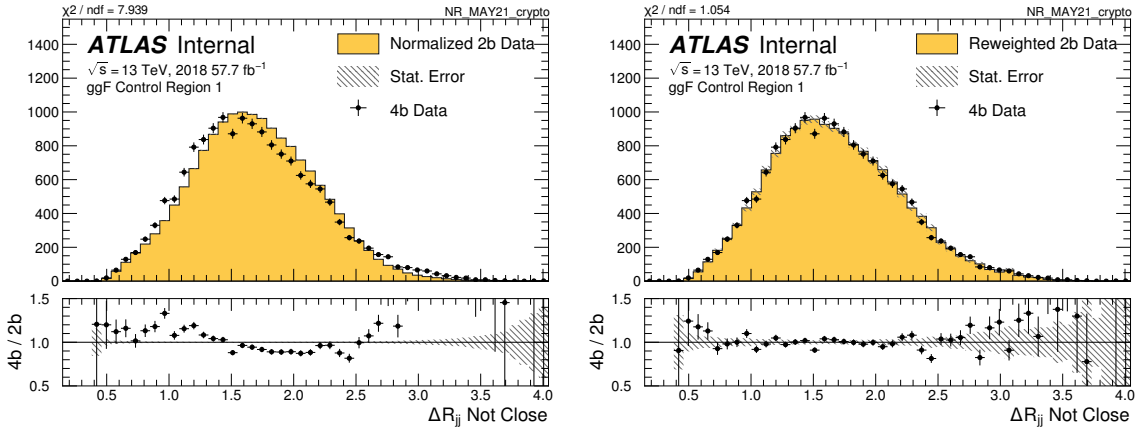


(a) Before Reweighting **(b)** After Reweighting



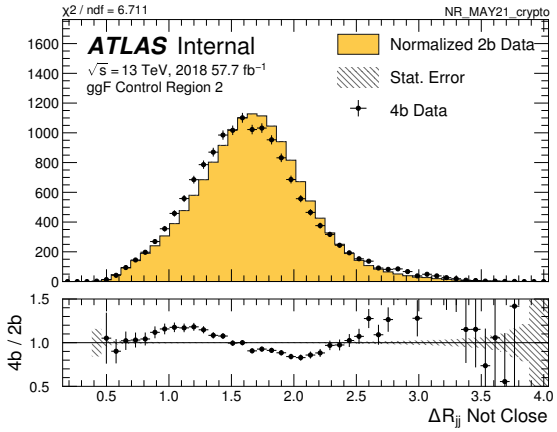
(c) Before Reweighting **(d)** After Reweighting

Figure A.9.: Non-resonant $4b$ p_{T_4} distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

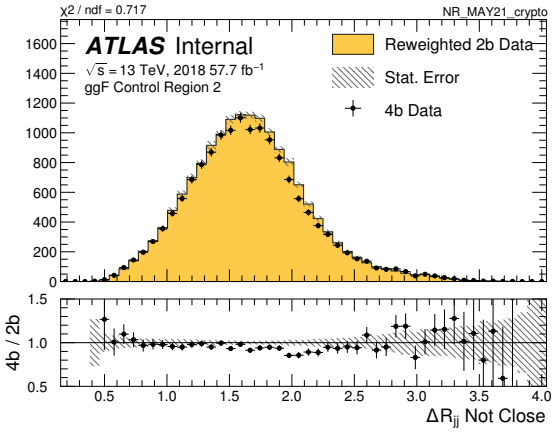


(a) Before Reweighting

(b) After Reweighting

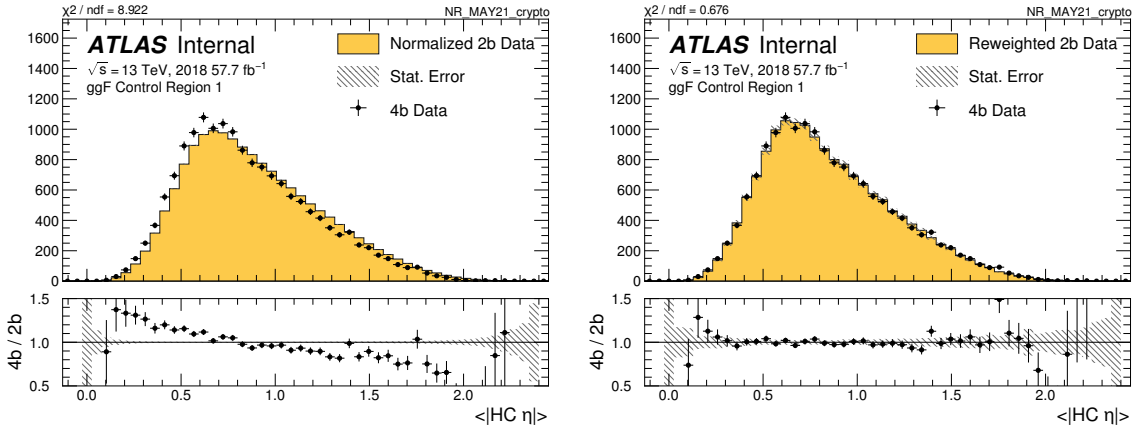


(c) Before Reweighting



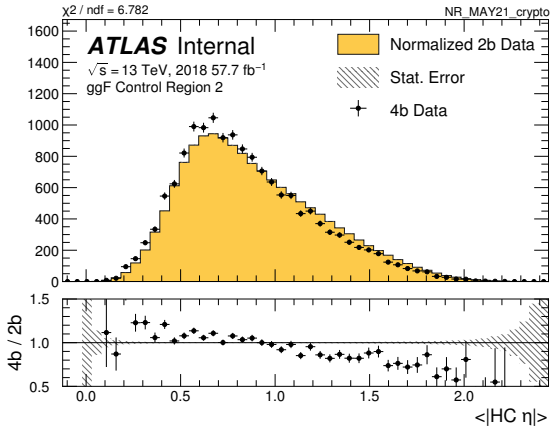
(d) After Reweighting

Figure A.10.: Non-resonant $4b$ ΔR_{jj} Not Close distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

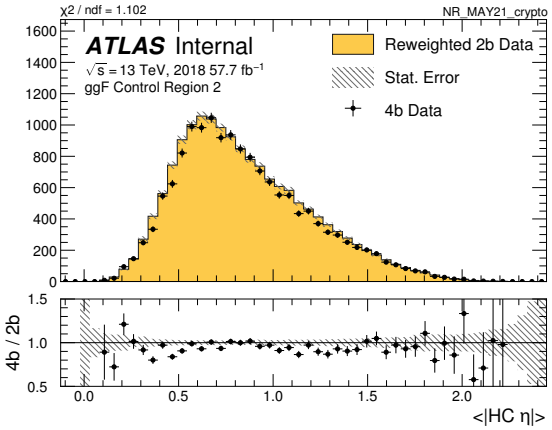


(a) Before Reweighting

(b) After Reweighting

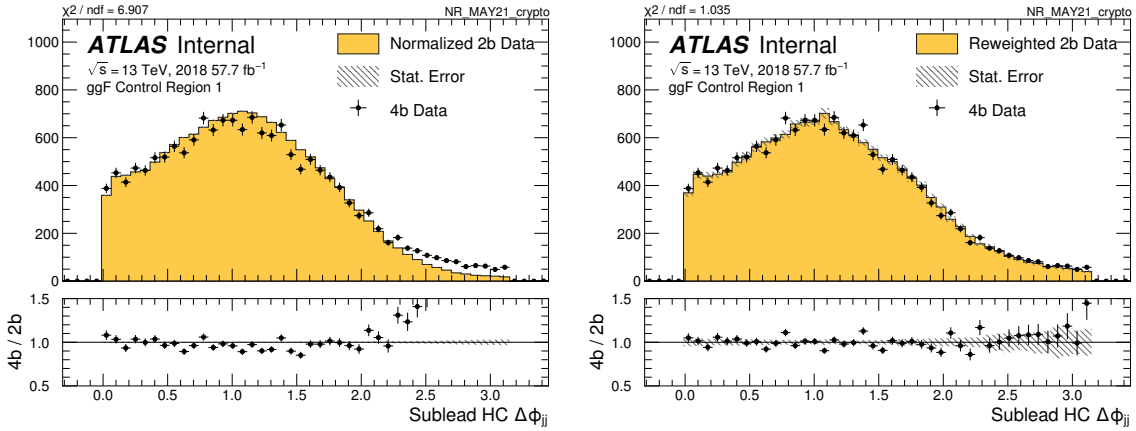


(c) Before Reweighting



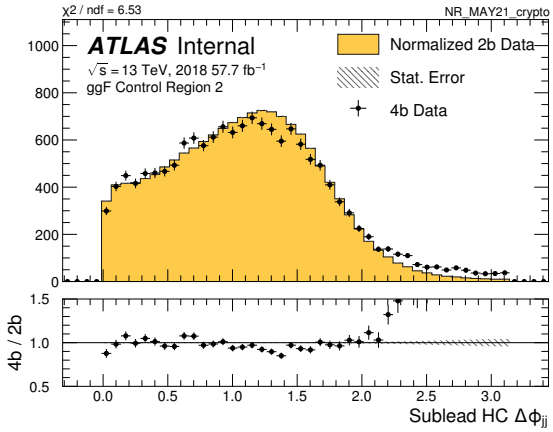
(d) After Reweighting

Figure A.11.: Non-resonant $4b$ $\langle|HC\eta|\rangle$ distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

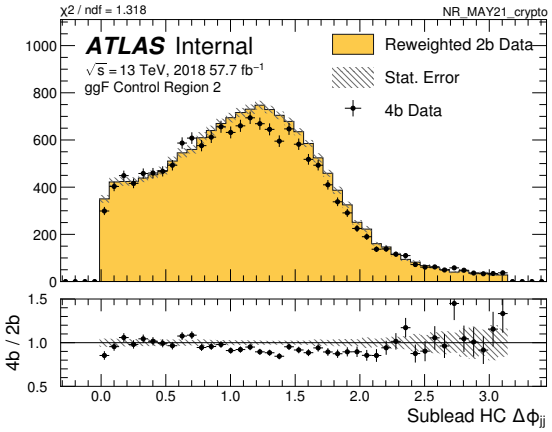


(a) Before Reweighting

(b) After Reweighting

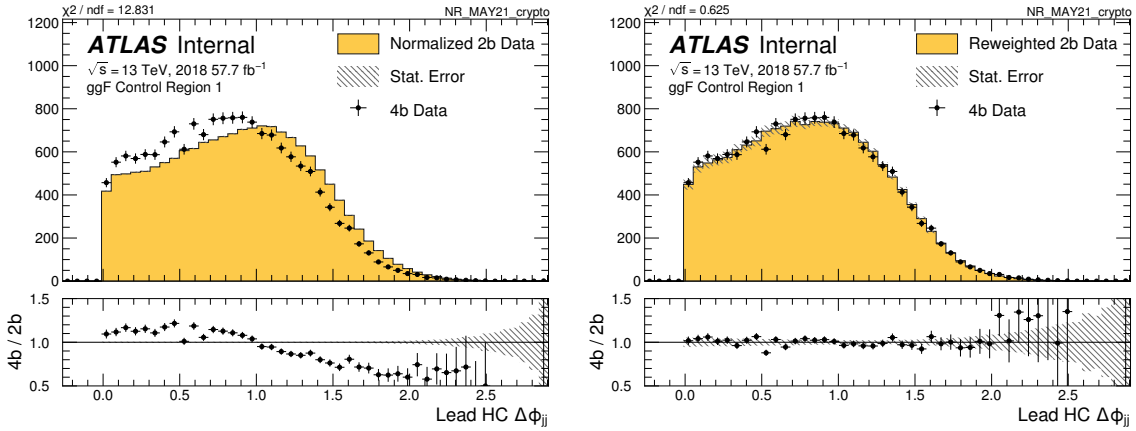


(c) Before Reweighting



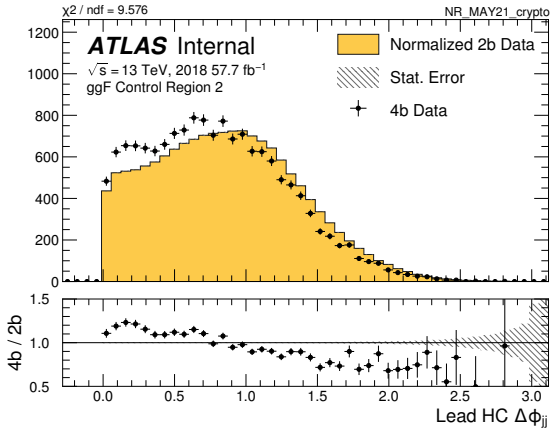
(d) After Reweighting

Figure A.12.: Non-resonant 4*b* Sublead HC $\Delta\phi_{jj}$ distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

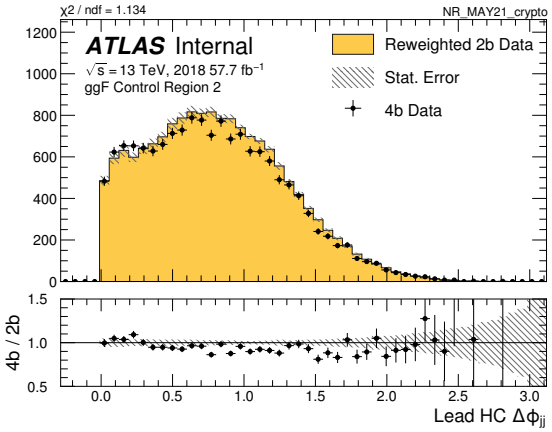


(a) Before Reweighting

(b) After Reweighting

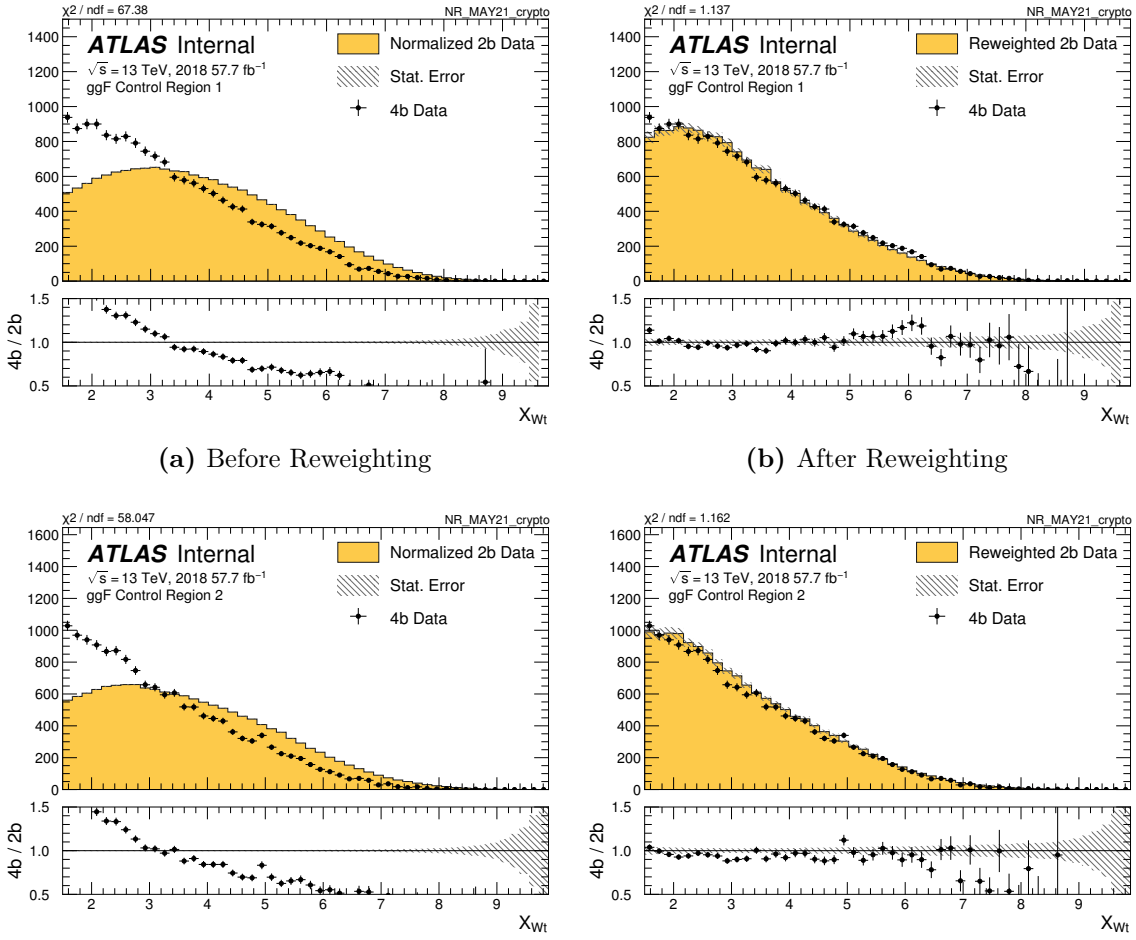


(c) Before Reweighting



(d) After Reweighting

Figure A.13.: Non-resonant 4b Lead HC $\Delta\phi_{jj}$ distributions before and after reweighting, in both CR1 and CR2 kinematic regions.



(a) Before Reweighting

(b) After Reweighting

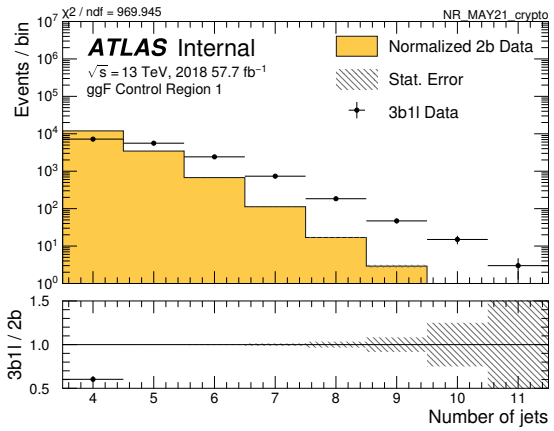
(c) Before Reweighting

(d) After Reweighting

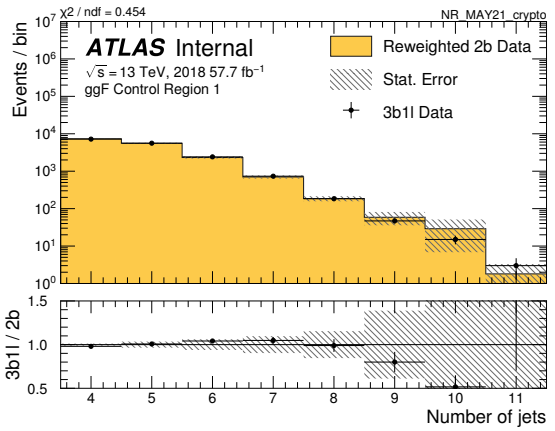
Figure A.14.: Non-resonant $4b$ X_{Wt} distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

A.3. Non-resonant reweighting 3*b*1*l*

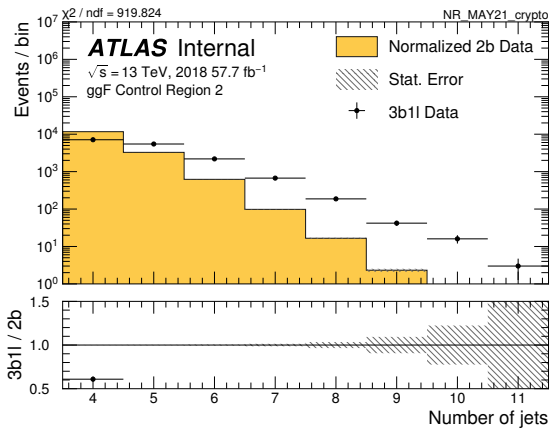
Figures A.15-A.26 show performance of the reweighting model of the 3*b*1*l* target in both control and validation regions of the non-resonant search.



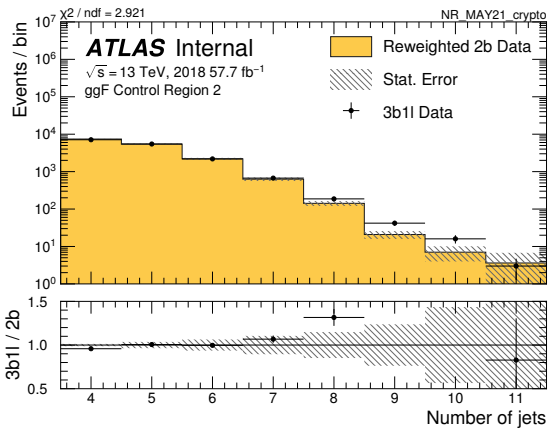
(a) Before Reweighting



(b) After Reweighting

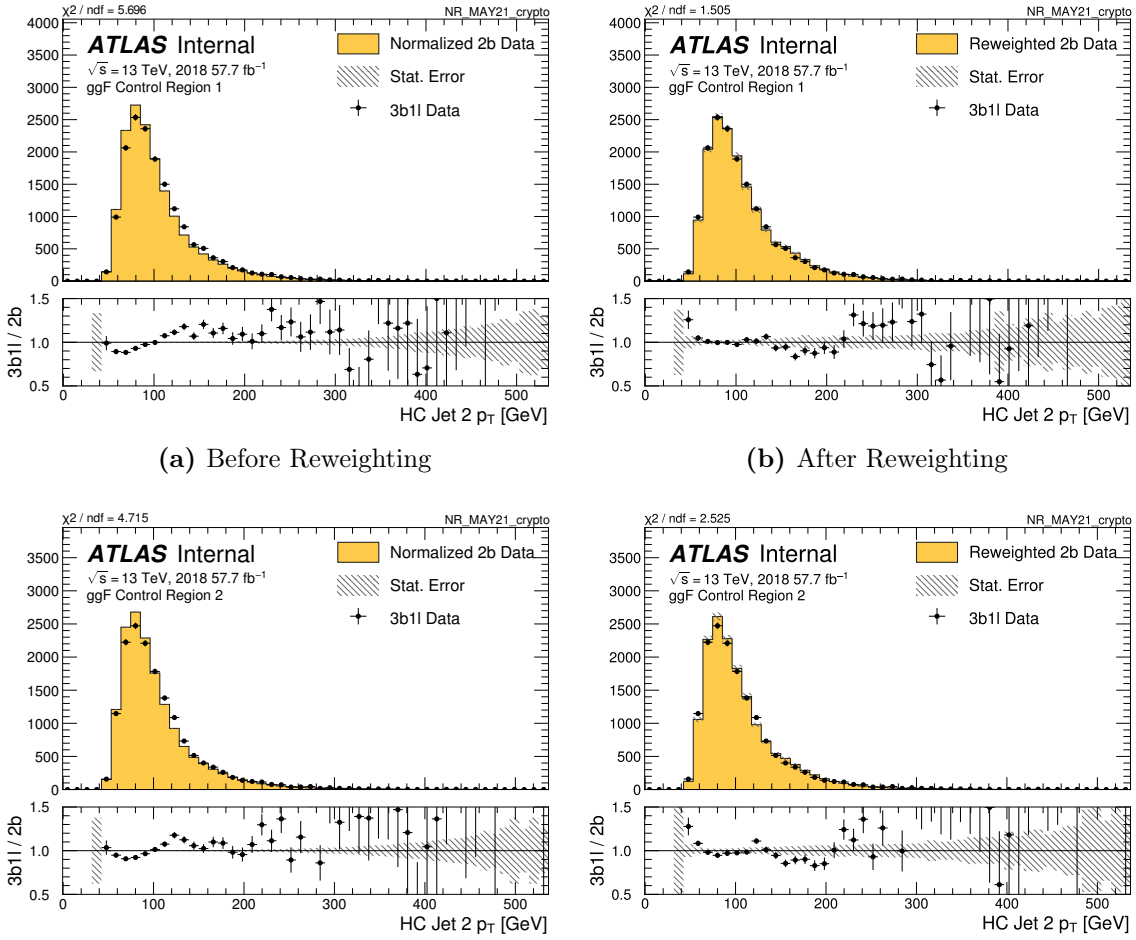


(c) Before Reweighting



(d) After Reweighting

Figure A.15.: Non-resonant 3*b*1*l* number of jets (njets) distributions before and after reweighting, in both CR1 and CR2 kinematic regions.



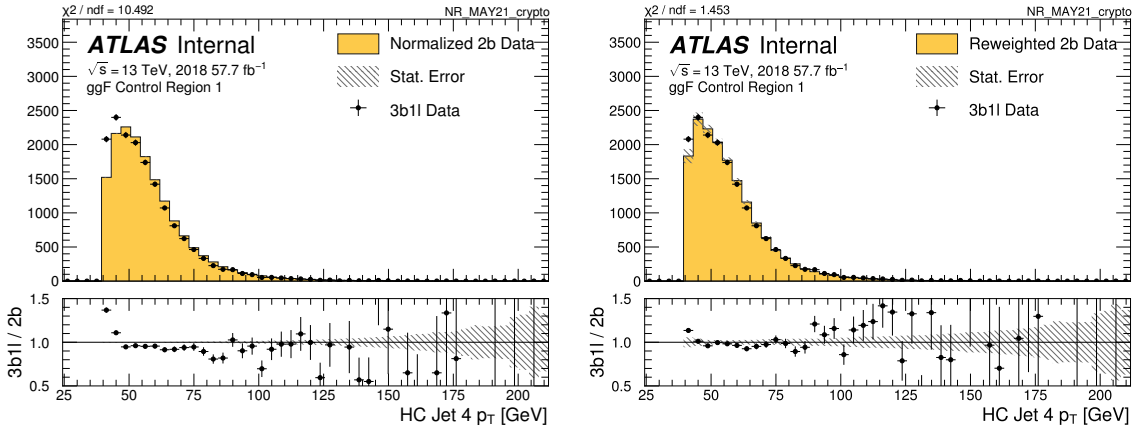
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

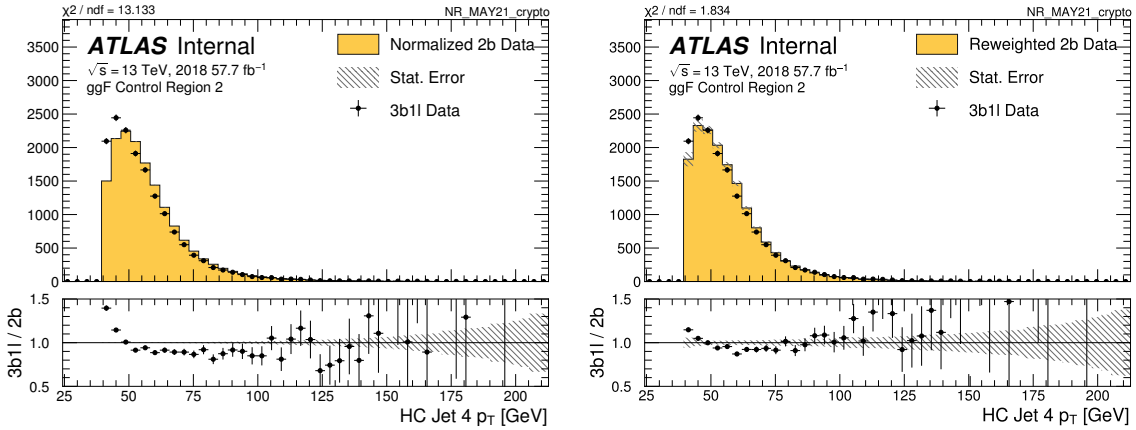
(d) After Reweighting

Figure A.16.: Non-resonant $3b1l$ p_{T_2} distributions before and after reweighting, in both CR1 and CR2 kinematic regions.



(a) Before Reweighting

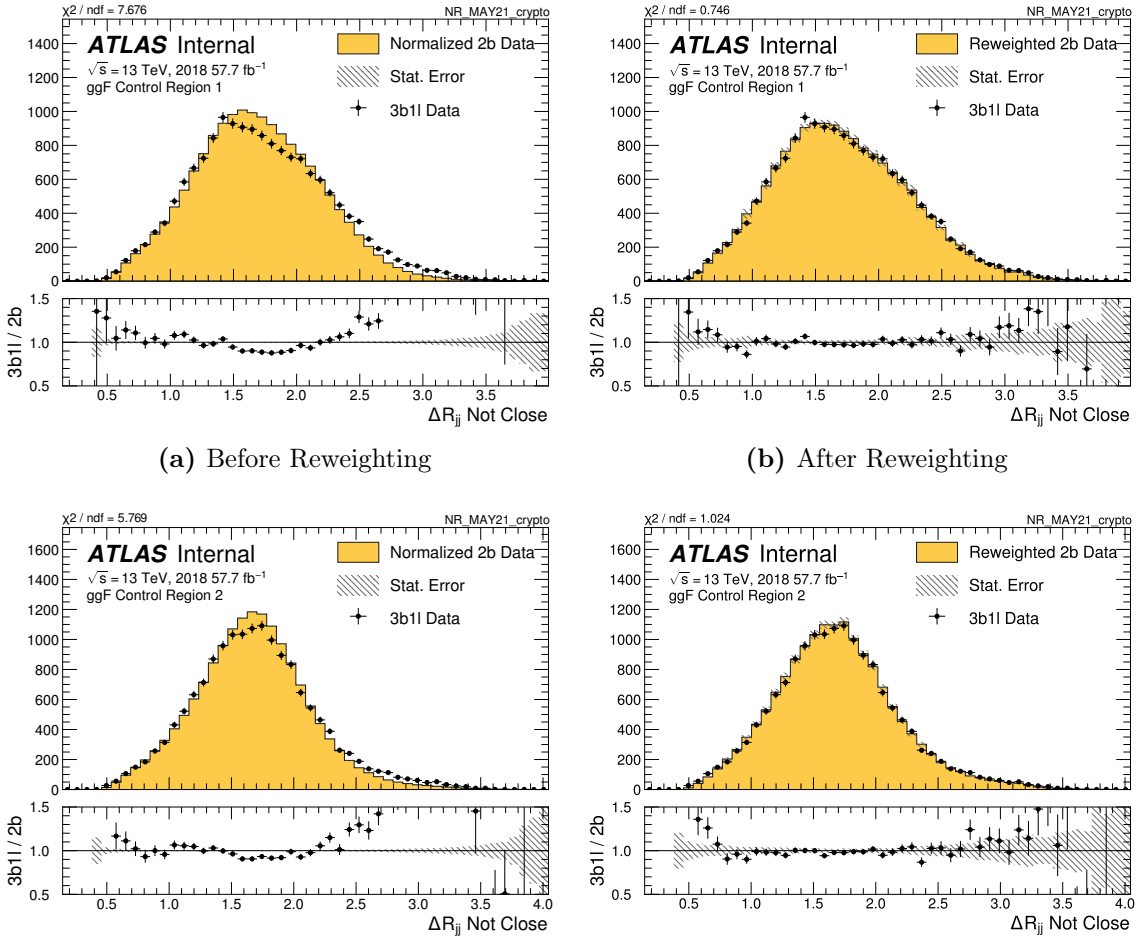
(b) After Reweighting



(c) Before Reweighting

(d) After Reweighting

Figure A.17.: Non-resonant $3b1l$ p_{T_4} distributions before and after reweighting, in both CR1 and CR2 kinematic regions.



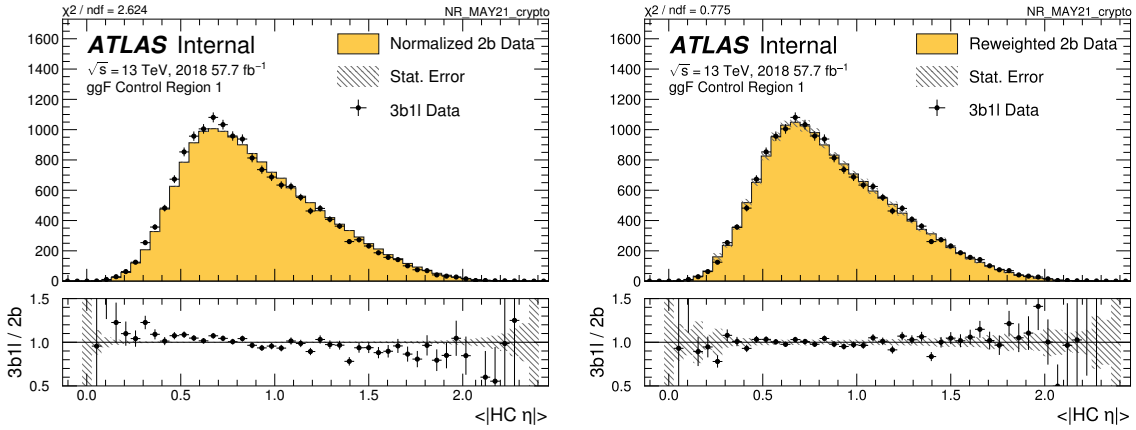
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

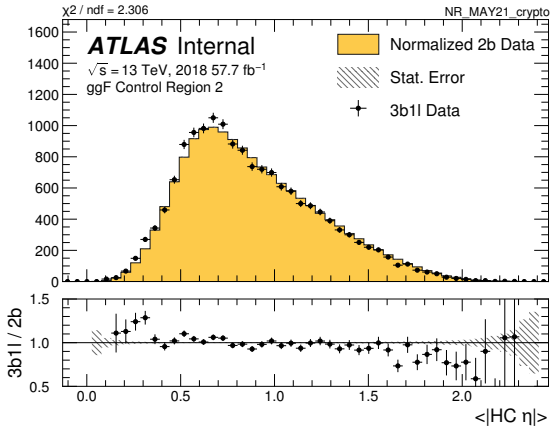
(d) After Reweighting

Figure A.18.: Non-resonant $3b1l$ ΔR_{jj} Not Close distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

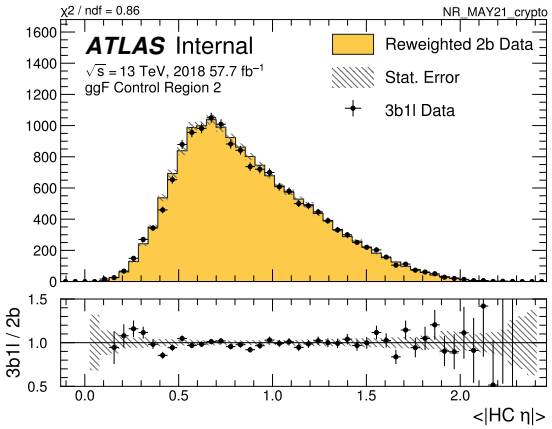


(a) Before Reweighting

(b) After Reweighting

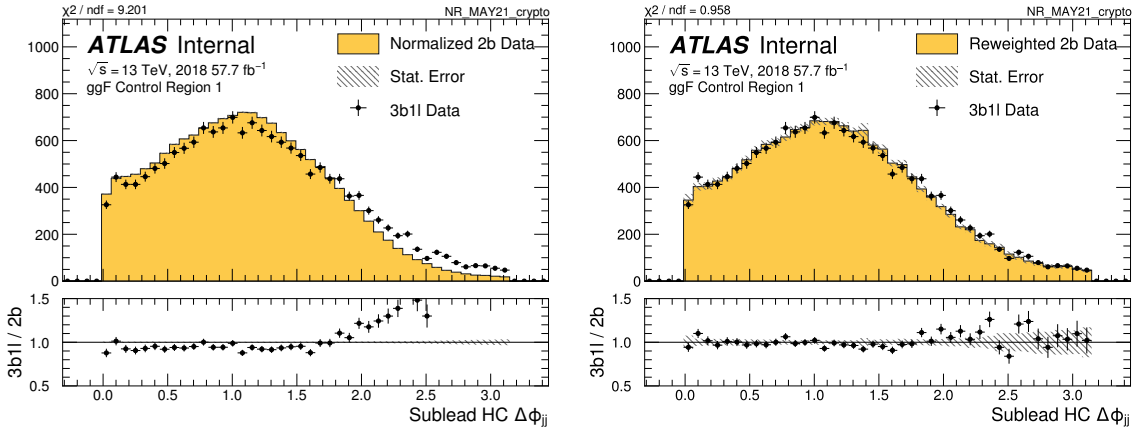


(c) Before Reweighting



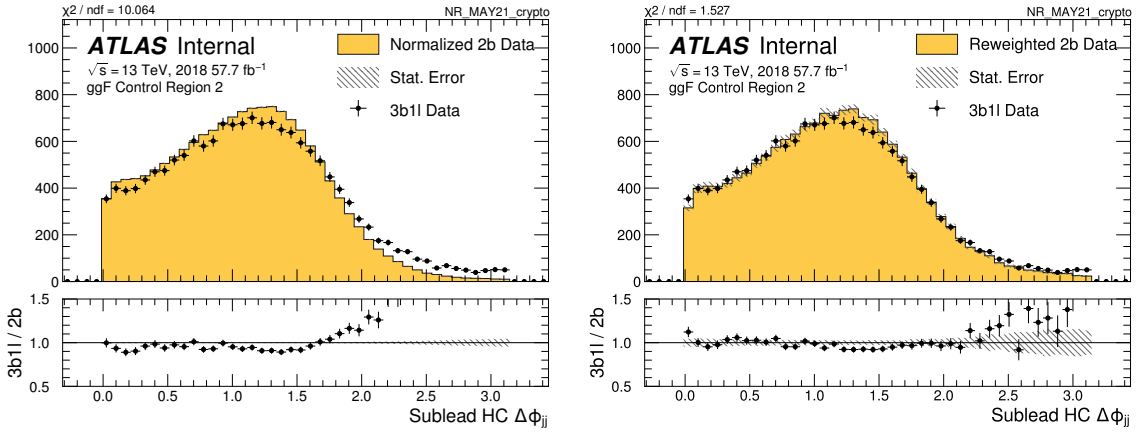
(d) After Reweighting

Figure A.19.: Non-resonant $3b1l$ $\langle|HC\eta|\rangle$ distributions before and after reweighting, in both CR1 and CR2 kinematic regions.



(a) Before Reweighting

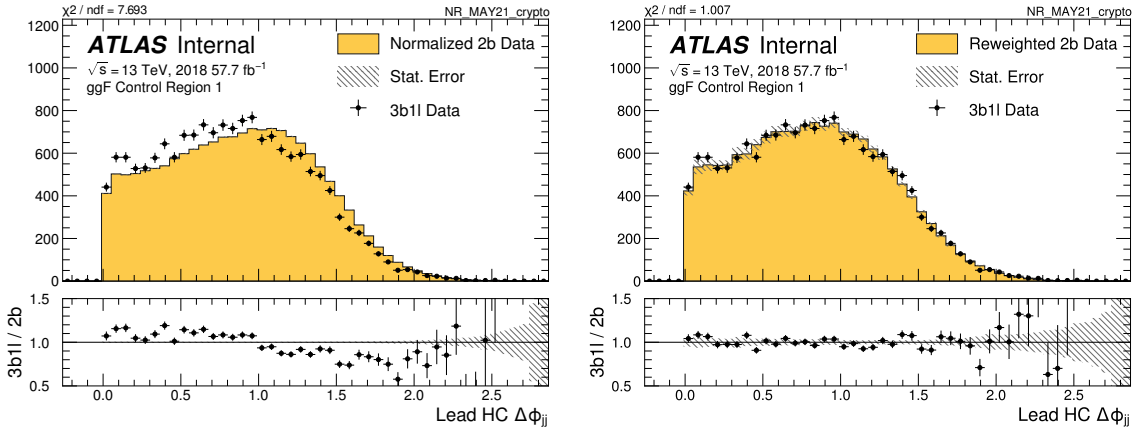
(b) After Reweighting



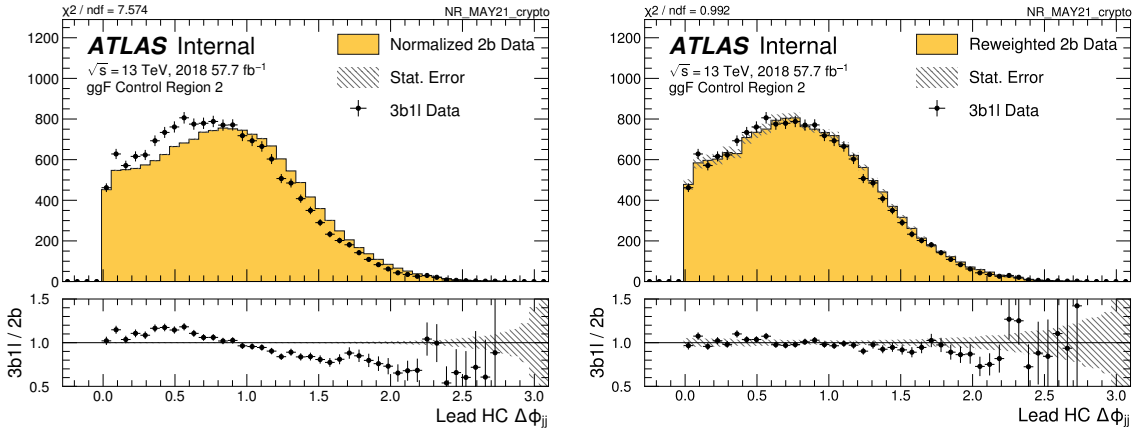
(c) Before Reweighting

(d) After Reweighting

Figure A.20.: Non-resonant $3b1l$ Sublead HC $\Delta\phi_{jj}$ distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

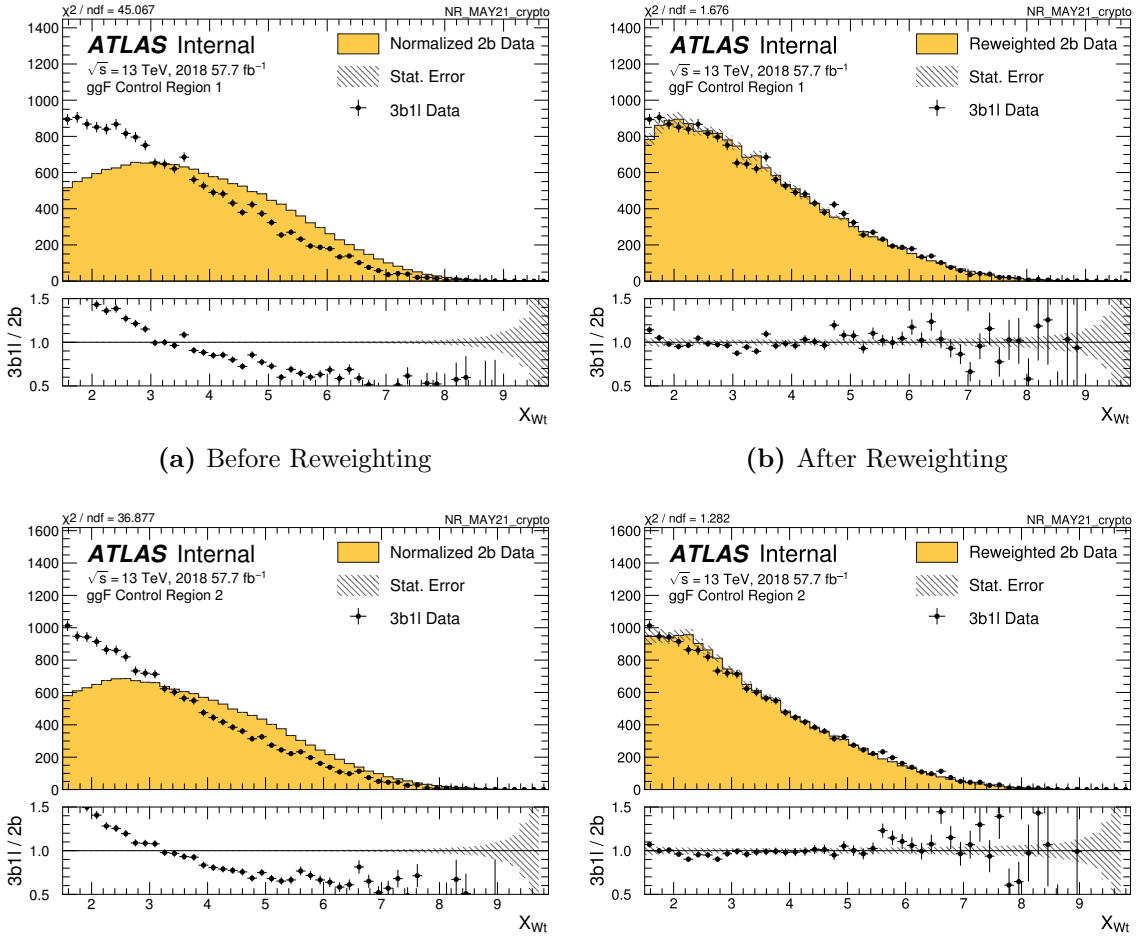


(a) Before Reweighting **(b)** After Reweighting



(c) Before Reweighting **(d)** After Reweighting

Figure A.21.: Non-resonant $3b1l$ Lead HC $\Delta\phi_{jj}$ distributions before and after reweighting, in both CR1 and CR2 kinematic regions.



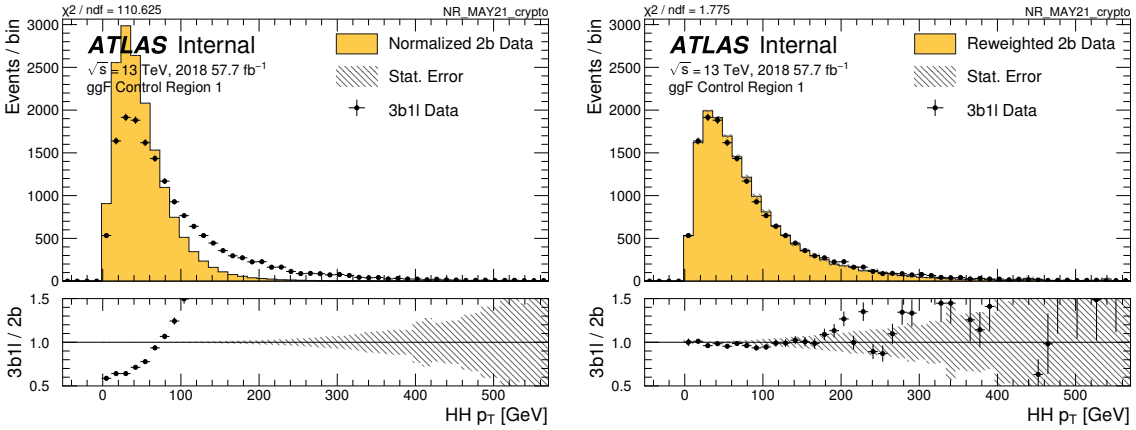
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

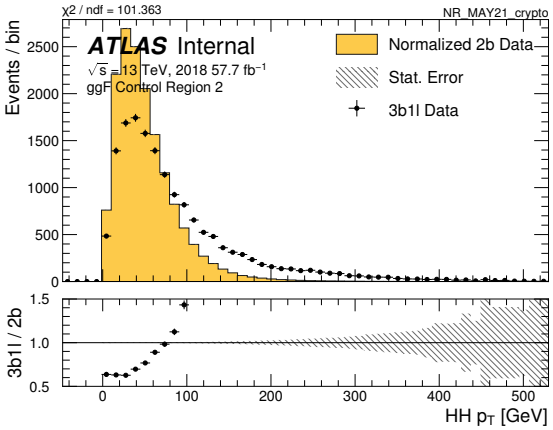
(d) After Reweighting

Figure A.22.: Non-resonant $3b1l$ X_{Wt} distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

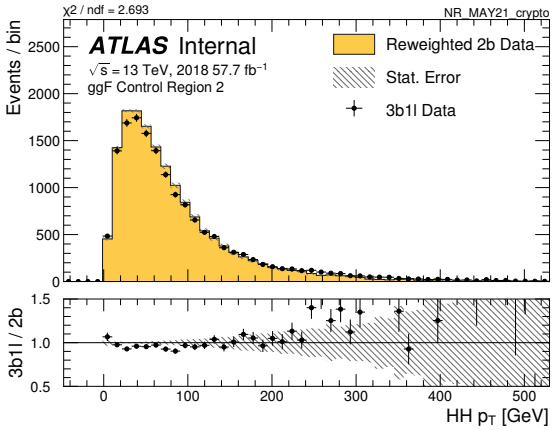


(a) Before Reweighting

(b) After Reweighting

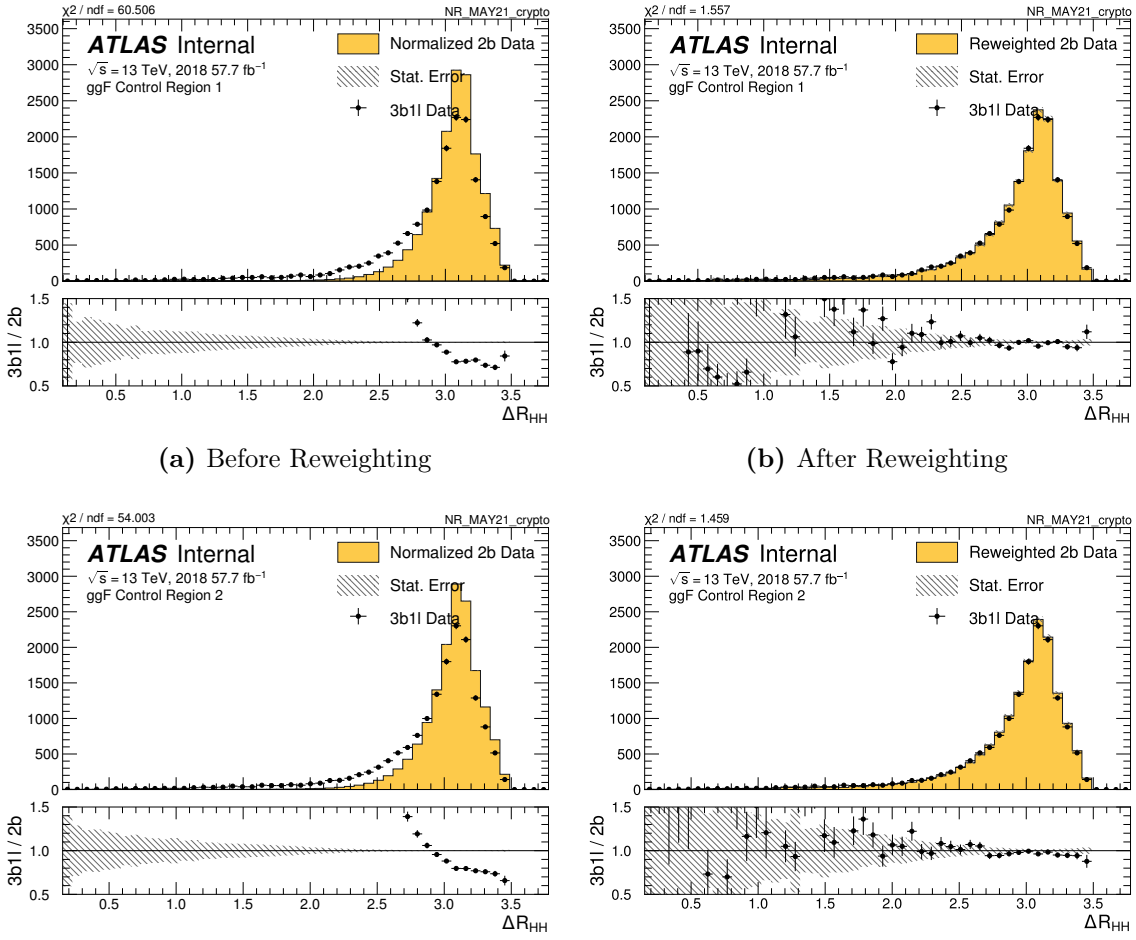


(c) Before Reweighting



(d) After Reweighting

Figure A.23.: Non-resonant $3b1l$ $p_{T_{HH}}$ distributions before and after reweighting, in both CR1 and CR2 kinematic regions.



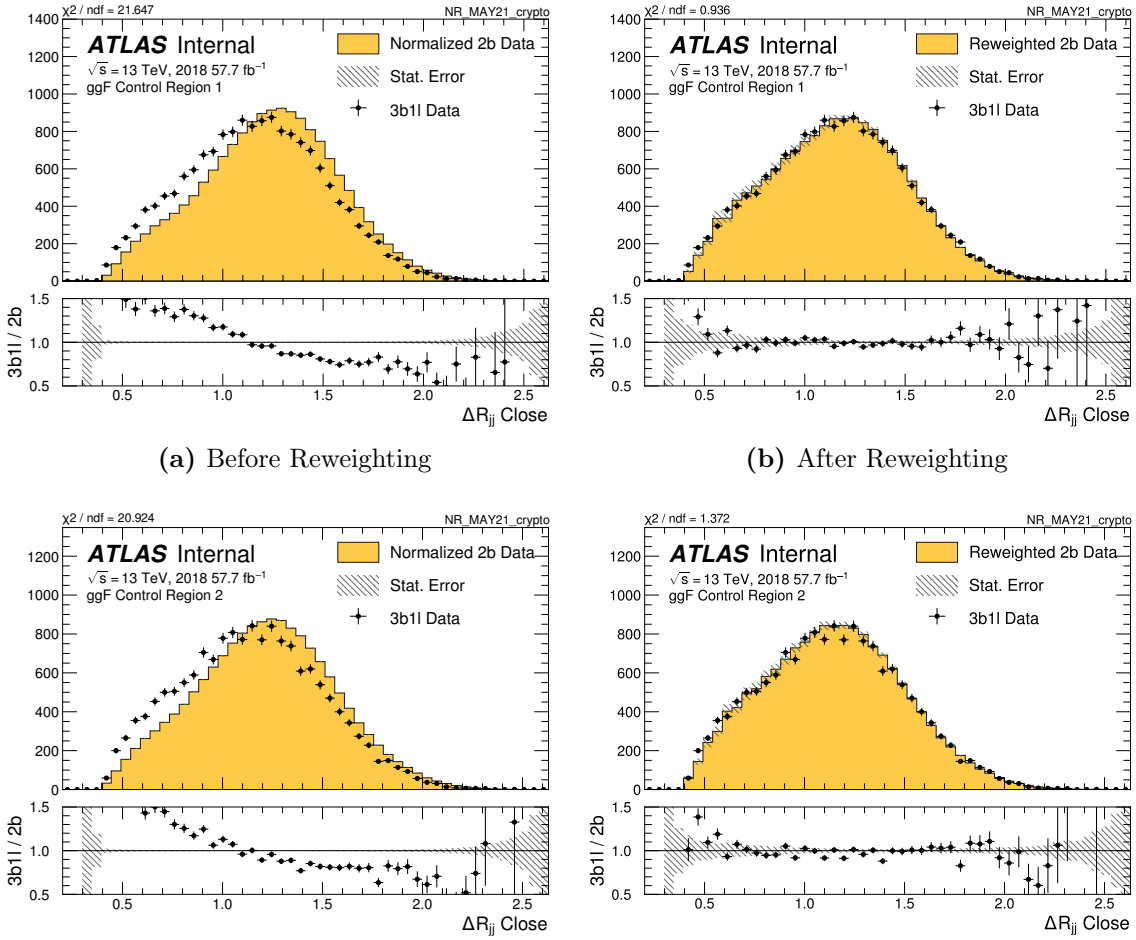
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

(d) After Reweighting

Figure A.24.: Non-resonant $3b1l$ ΔR_{HH} distributions before and after reweighting, in both CR1 and CR2 kinematic regions.



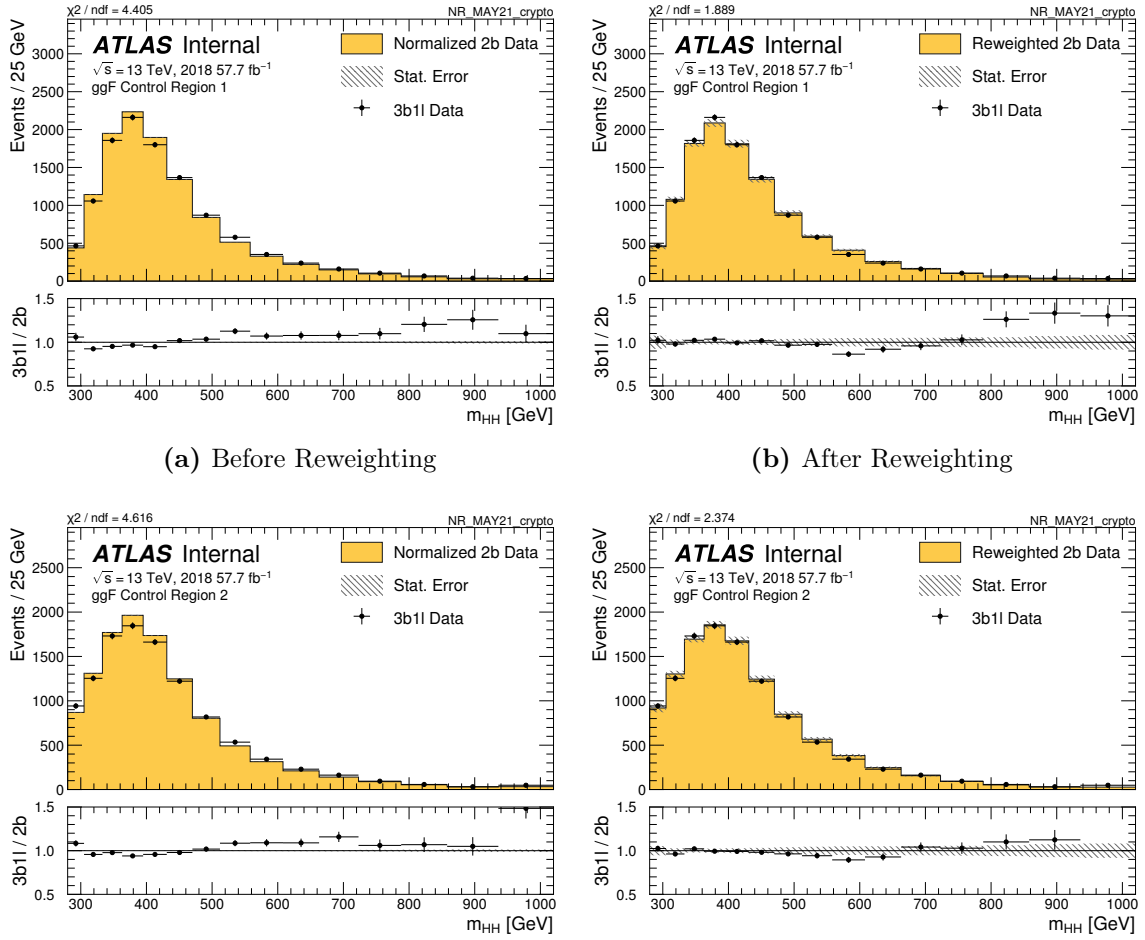
(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

(d) After Reweighting

Figure A.25.: Non-resonant $3b1l$ ΔR_{jj} Close distributions before and after reweighting, in both CR1 and CR2 kinematic regions.



(a) Before Reweighting

(b) After Reweighting

(c) Before Reweighting

(d) After Reweighting

Figure A.26.: Non-resonant $3b1l$ m_{HH} Close distributions before and after reweighting, in both CR1 and CR2 kinematic regions.

Appendix B.

Background Validation

B.1. Resonant search reversed $|\Delta\eta_{HH}|$ validation

The validation of the reversed $|\Delta\eta_{HH}|$ region for the resonant search was also examined for other key kinematic variables used to train the reweightin model. Figures B.1, B.2 and B.3, show the agreement of the background model and the $4b$ data in the reversed $|\Delta\eta_{HH}|$ region. To a large extent these distributions provide further validity to the background model. Only the pT_{HH} distribution shows some discrepancy, but this variable is known to be difficult to model in the nominal CR and VR regions.

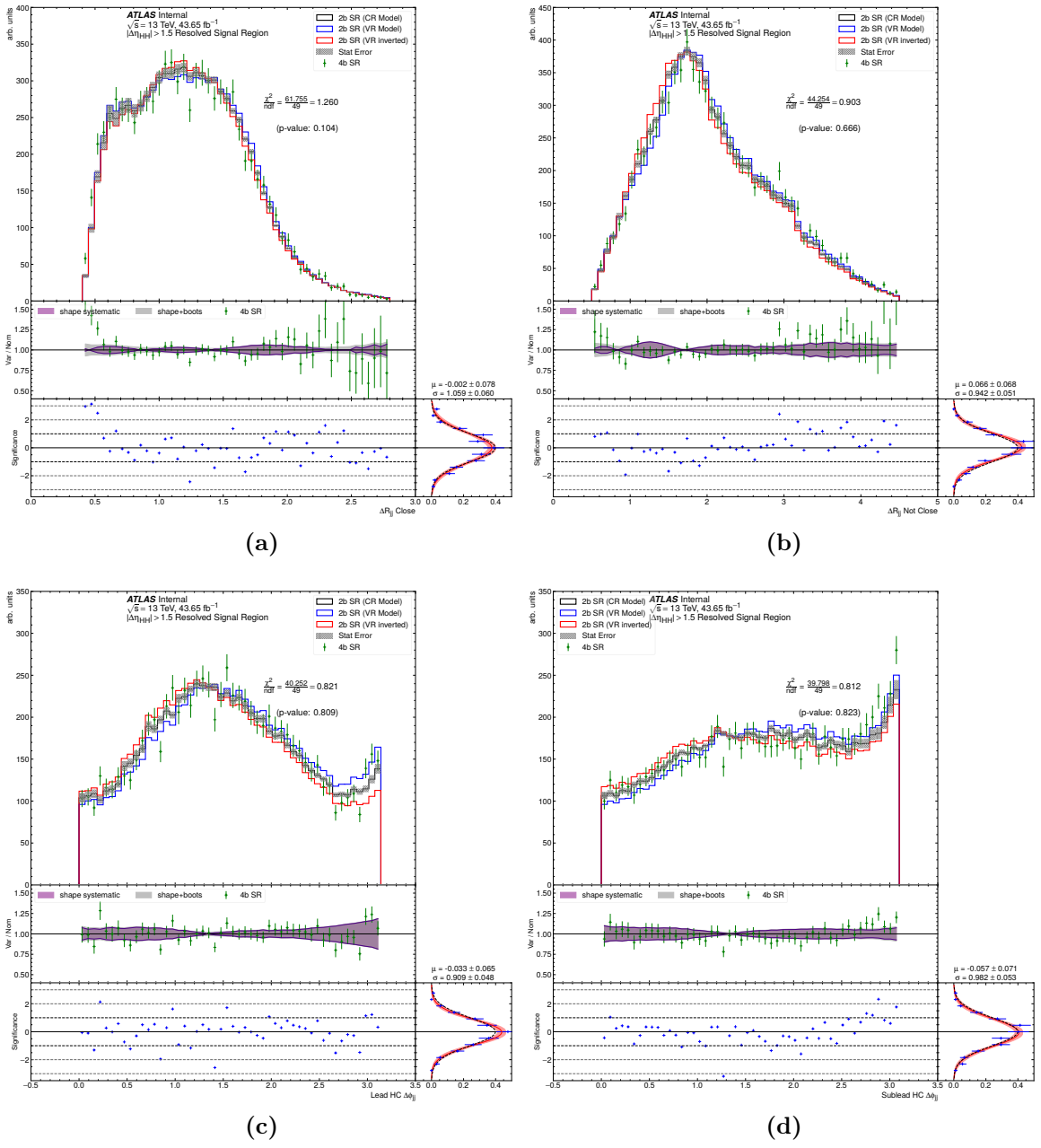


Figure B.1.: Resonant search reversed $|\Delta\eta_{HH}|$ validation in the signal region. ΔR_{jj} Close, ΔR_{jj} Not Close, Lead HC $\Delta\phi_{jj}$ and Sublead HC $\Delta\phi_{jj}$ kinematic distributions.

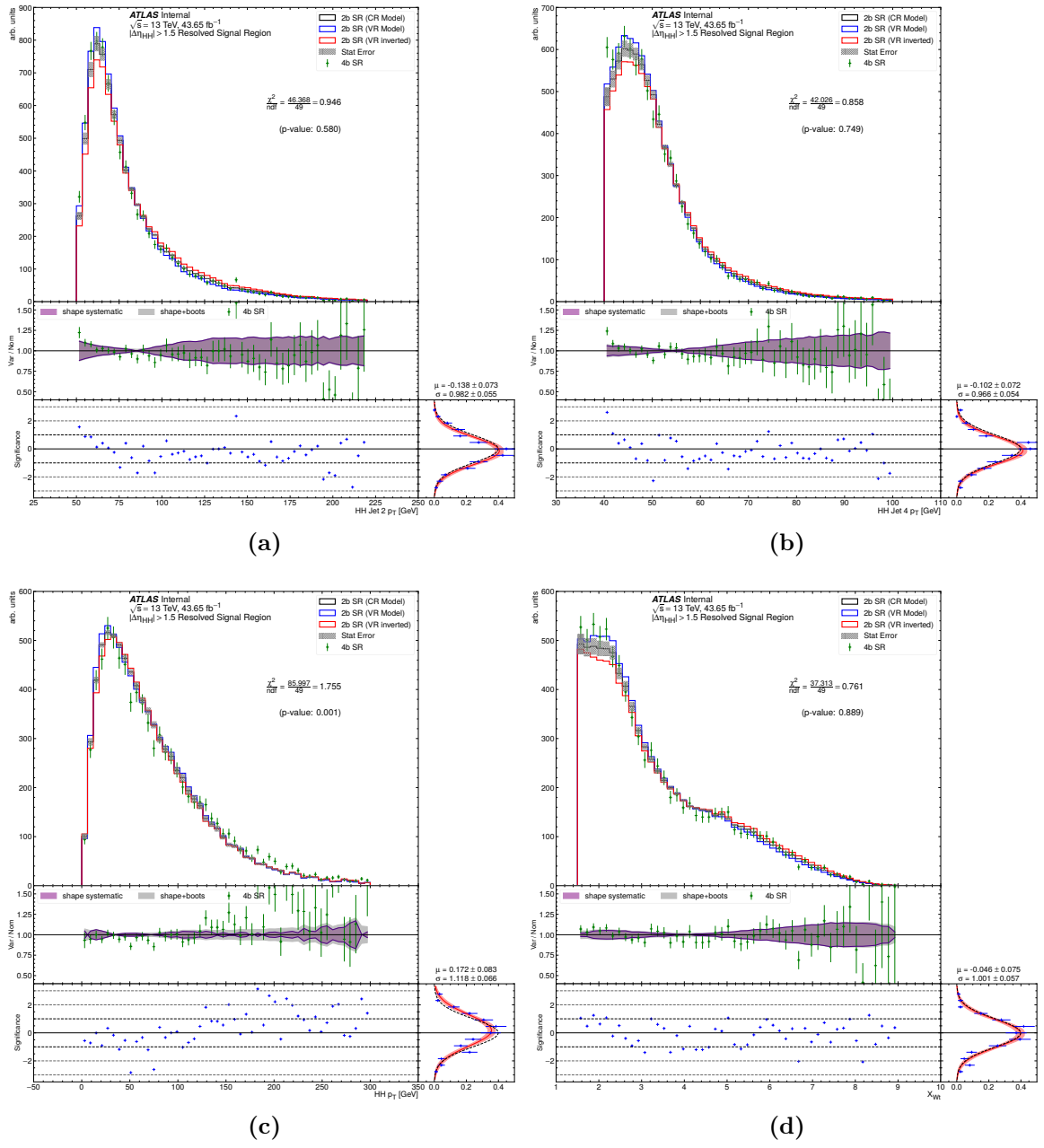
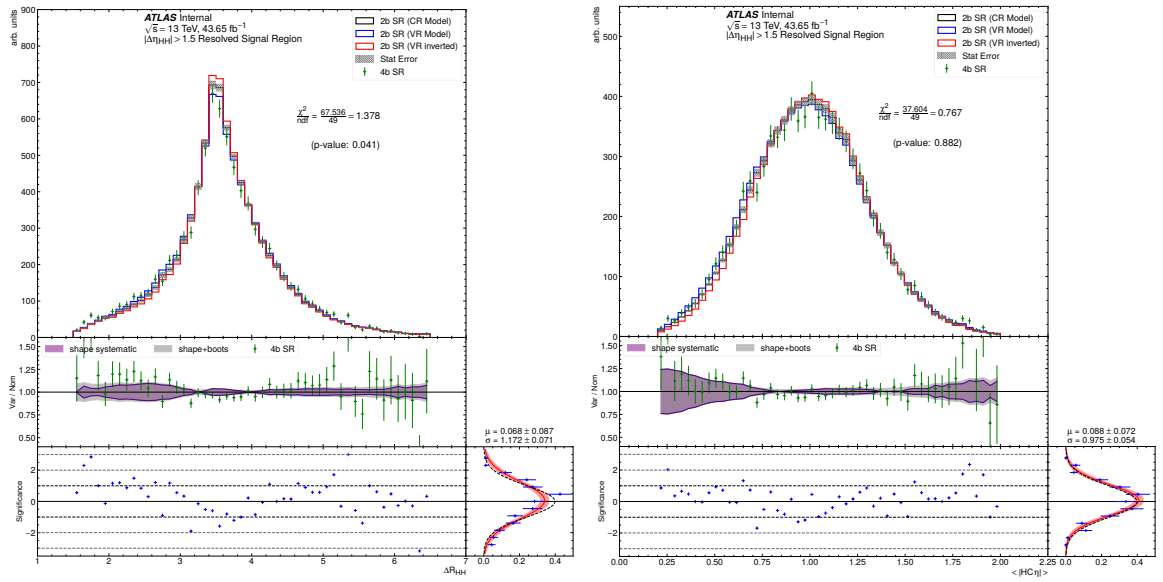


Figure B.2.: Resonant search reversed $|\Delta\eta_{HH}|$ validation in the signal region. HC Jet 2 p_T, HC Jet 4p_T, HH p_T and X_{Wt} kinematic distributions.



(a) **(b)**

Figure B.3.: Resonant search reversed $|\Delta\eta_{HH}|$ validation in the signal region. ΔR_{HH} and η_i kinematic distributions.

B.2. Non-resonant search reversed $|\Delta\eta_{HH}|$ validation

The validation of the reversed $|\Delta\eta_{HH}|$ region for the non-resonant search was also examined for other key kinematic variables used to derive the reweighting model. Figures B.4, B.5 and B.6, show the agreement of the background model and the $4b$ data in the reversed $|\Delta\eta_{HH}|$ region. In general, the same issues seen previously in the m_{HH} distribution in Figure 8.23 are present in the rest of the kinematic distributions.

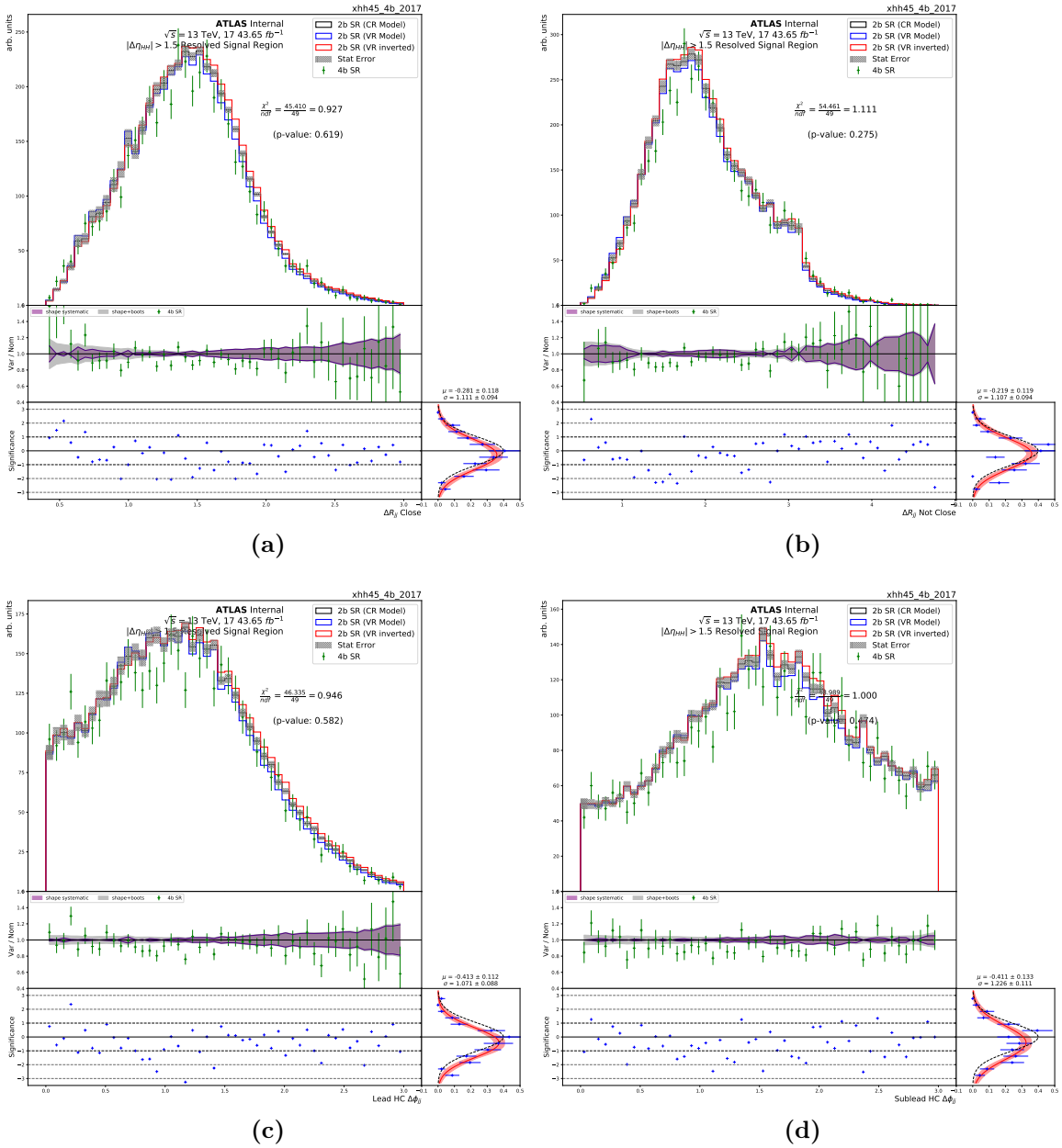


Figure B.4.: Non-resonant search reversed $|\Delta\eta_{HH}|$ validation in the signal region. ΔR_{jj} Close, ΔR_{jj} Not Close, Lead HC $\Delta\phi_{jj}$ and Sublead HC $\Delta\phi_{jj}$ kinematic distributions.

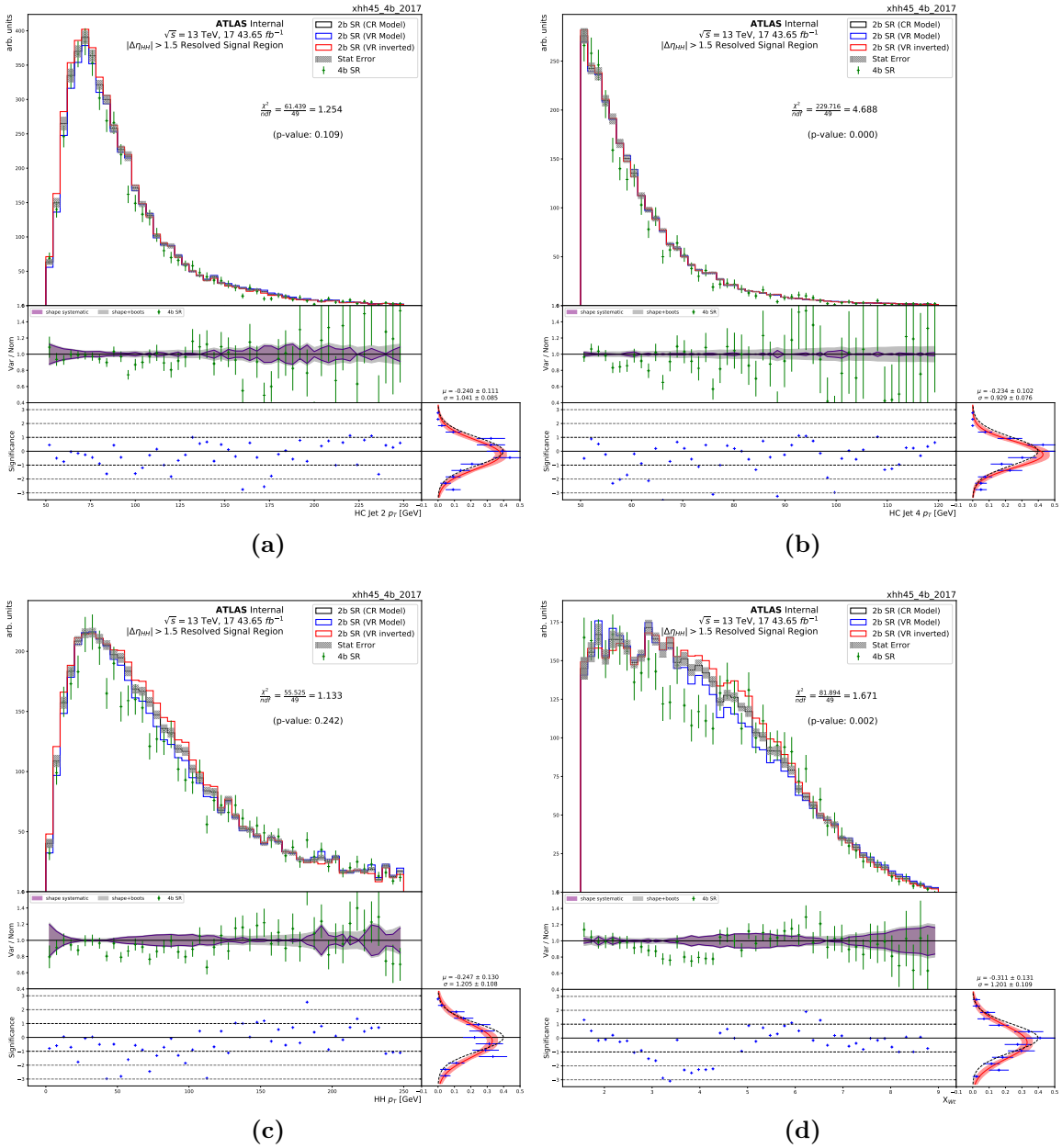
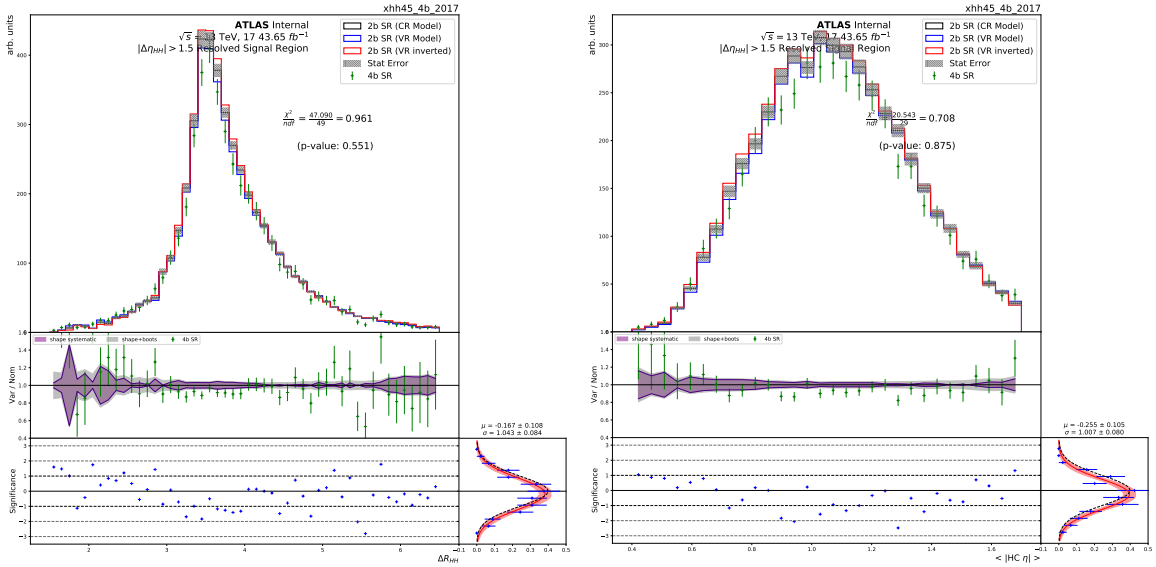


Figure B.5.: Non-resonant search reversed $|\Delta\eta_{HH}|$ validation in the signal region. HC Jet 2 p_T, HC Jet 4 p_T, HH p_T and X_{Wt} kinematic distributions.



(a) **(b)**

Figure B.6.: Non-resonant search reversed $|\Delta\eta_{HH}|$ validation in the signal region. ΔR_{HH} and $\langle|HC\eta|\rangle$ kinematic distributions.

Bibliography

[1] Carsten Burgard. *Example: Standard model of physics*. 2016. URL: https://texample.net/tikz/examples/model-physics/ (visited on 09/21/2021).

[2] Mark Thomson. *Modern particle physics*. New York: Cambridge University Press, 2013. ISBN: 978-1-107-03426-6.

[3] ATLAS Collaboration and CMS Collaboration. "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". In: *Phys. Rev. Lett.* 114 (19 May 2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. URL: https://link.aps.org/doi/10.1103/PhysRevLett.114.191803.

[4] MINOS Collaboration. "Observation of Muon Neutrino Disappearance with the MINOS Detectors in the NuMI Neutrino Beam". In: *Phys. Rev. Lett.* 97 (19 Nov. 2006), p. 191801. DOI: 10.1103/PhysRevLett.97.191801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.97.191801.

[5] The Super-Kamiokande Collaboration. "Evidence for an Oscillatory Signature in Atmospheric Neutrino Oscillations". In: *Phys. Rev. Lett.* 93 (10 Sept. 2004), p. 101801. DOI: 10.1103/PhysRevLett.93.101801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.101801.

[6] CERN. *Dark matter*. 2021. URL: https://home.cern/science/physics/dark-matter (visited on 09/22/2021).

[7] Shinya Kanemura et al. "Can WIMP dark matter overcome the nightmare scenario?" In: *Physical Review D* 82.5 (Sept. 2010). ISSN: 1550-2368. DOI: 10.1103/physrevd.82.055026. URL: http://dx.doi.org/10.1103/PhysRevD.82.055026.

[8] LHCb collaboration. *Test of lepton universality in beauty-quark decays*. 2021. arXiv: 2103.11769 [hep-ex].

[9] Muon $g-2$ Collaboration. "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm". In: *Phys. Rev. Lett.* 126 (14 Apr. 2021), p. 141801. DOI: 10.1103/PhysRevLett.126.141801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.126.141801.

[10] Kaustubh Agashe et al. "Warped gravitons at the CERN LHC and beyond". In: *Phys. Rev. D* 76 (3 Aug. 2007), p. 036006. DOI: 10.1103/PhysRevD.76.036006. URL: https://link.aps.org/doi/10.1103/PhysRevD.76.036006.

[11] Alexandra Carvalho. *Gravity particles from Warped Extra Dimensions, predictions for LHC*. 2018. arXiv: 1404.0102 [hep-ph].

[12] G.C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Physics Reports* 516.1-2 (July 2012), pp. 1–102. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2012.02.002. URL: http://dx.doi.org/10.1016/j.physrep.2012.02.002.

[13] Neil Turok and John Zadrozny. "Electroweak baryogenesis in the two-doublet model". In: *Nuclear Physics B* 358.2 (1991), pp. 471–493. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(91)90356-3. URL: https://www.sciencedirect.com/science/article/pii/0550321391903563.

[14] Laura Lopez Honorez et al. "The inert doublet model: an archetype for dark matter". In: *Journal of Cosmology and Astroparticle Physics* 2007.02 (Feb. 2007), pp. 028–028. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2007/02/028. URL: http://dx.doi.org/10.1088/1475-7516/2007/02/028.

[15] CERN. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. Geneva: CERN, Oct. 2016. DOI: 10.23731/CYRM-2017-002. URL: https://cds.cern.ch/record/2227475.

[16] *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*. Tech. rep. Geneva: CERN, Aug. 2021. URL: https://cds.cern.ch/record/2777861.

[17] M. Grazzini et al. "Higgs boson pair production at NNLO with top quark mass effects". In: *Journal of High Energy Physics* 2018.5 (May 2018). ISSN: 1029-8479. DOI: 10.1007/jhep05(2018)059. URL: http://dx.doi.org/10.1007/JHEP05(2018)059.

[18] G. Aad et al. "Combination of searches for Higgs boson pairs in pp collisions at s=13TeV with the ATLAS detector". In: *Physics Letters B* 800 (Jan. 2020), p. 135103. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2019.135103. URL: http://dx.doi.org/10.1016/j.physletb.2019.135103.

[19] J. Baglio et al. "$gg \to HH$: Combined uncertainties". In: *Physical Review D* 103.5 (Mar. 2021). ISSN: 2470-0029. DOI: 10.1103/physrevd.103.056002. URL: http://dx.doi.org/10.1103/PhysRevD.103.056002.

[20] ATLAS Collaboration. "Search for Resonant and Nonresonant Higgs Boson Pair Production in the $b\bar{b}\tau^+\tau^-$ Decay Channel in pp Collisions at $\sqrt{s} = 13$TeV with the ATLAS Detector". In: *Phys. Rev. Lett.* 121 (19 Nov. 2018), p. 191801. DOI: 10.1103/PhysRevLett.121.191801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.191801.

[21] M. Aaboud 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". In: *Journal of High Energy Physics* 2019.1 (Jan. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep01(2019)030. URL: http://dx.doi.org/10.1007/JHEP01(2019)030.

[22] M. Aaboud et al. "Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state with 13 TeV pp collision data collected by the ATLAS experiment". In: *Journal of High Energy Physics* 2018.11 (Nov. 2018). ISSN: 1029-8479. DOI: 10.1007/jhep11(2018)040. URL: http://dx.doi.org/10.1007/JHEP11(2018)040.

[23] ATLAS Collaboration. "Search for Higgs boson pair production in the WW^*WW^* decay channel using ATLAS data recorded at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2019 (May 2019). DOI: 10.1007/JHEP05(2019)124.

[24] M. Aaboud et al. "Search for Higgs boson pair production in the $\gamma\gamma WW^*$ channel using pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *The European Physical Journal C* 78.12 (Dec. 2018). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-6457-x. URL: http://dx.doi.org/10.1140/epjc/s10052-018-6457-x.

[25] M. Aaboud et al. "Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2019.4 (Apr. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep04(2019)092. URL: http://dx.doi.org/10.1007/JHEP04(2019)092.

[26] A.M. Sirunyan et al. "Combination of Searches for Higgs Boson Pair Production in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV". In: *Physical Review Letters* 122.12 (Mar. 2019). ISSN: 1079-7114. DOI: 10.1103/physrevlett.122.121803. URL: http://dx.doi.org/10.1103/PhysRevLett.122.121803.

[27] Georges Aad et al. "Reconstruction and identification of boosted di-τ systems in a search for Higgs boson pairs using 13 TeV proton-proton collision data in ATLAS". In: *JHEP* 11 (2020), p. 163. DOI: 10.1007/JHEP11(2020)163. arXiv: 2007.14811 [hep-ex].

[28] G. Aad et al. "Erratum to: Search for the HH $\to b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2021.1 (Jan. 2021). ISSN: 1029-8479. DOI: 10.1007/jhep01(2021)145. URL: http://dx.doi.org/10.1007/JHEP01(2021)145.

[29] G. Aad et al. "Search for non-resonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at s=13 TeV". In: *Physics Letters B* 801 (Feb. 2020), p. 135145. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2019.135145. URL: http://dx.doi.org/10.1016/j.physletb.2019.135145.

[30] *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. Geneva: CERN, Mar. 2021. URL: https://cds.cern.ch/record/2759683.

[31] *Search for Higgs boson pair production in the four b quark final state*. Tech. rep. Geneva: CERN, 2021. URL: https://cds.cern.ch/record/2771912.

[32] Falko Dulat, Achilleas Lazopoulos, and Bernhard Mistlberger. "iHixs 2 — Inclusive Higgs cross sections". In: *Computer Physics Communications* 233 (2018), pp. 243–260. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2018.06.025. URL: https://www.sciencedirect.com/science/article/pii/S0010465518302479.

[33] M. et al Cepeda. "Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC". In: *CERN Yellow Rep. Monogr.* 7 (Dec. 2018), 221–584. 364 p. DOI: 10.23731/CYRM-2019-007.221. arXiv: 1902.00134. URL: https://cds.cern.ch/record/2650162.

[34] Albert M Sirunyan et al. "A Deep Neural Network for Simultaneous Estimation of b Jet Energy and Resolution". In: *Comput. Softw. Big Sci.* 4.1 (2020), p. 10. DOI: 10.1007/s41781-020-00041-z. arXiv: 1912.06046 [hep-ex].

[35] Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. *Mathematics for Machine Learning*. Cambridge University Press, 2020.

[36] Andrew Stuart Bell. "b-Tagging and Evidence for the Standard Model $H \to b\bar{b}$ Decay with the ATLAS Experiment". Presented 20 Dec 2018. Nov. 2017. URL: https://cds.cern.ch/record/2644018.

[37] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis". In: (Mar. 2007). arXiv: physics/0703039.

[38] Jerome H. Friedman. "Greedy function approximation: A gradient boosting machine." In: *The Annals of Statistics* 29.5 (2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451. URL: https://doi.org/10.1214/aos/1013203451.

[39] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. URL: http://doi.acm.org/10.1145/2939672.2939785.

[40] Guolin Ke et al. "Lightgbm: A highly efficient gradient boosting decision tree". In: *Advances in neural information processing systems* 30 (2017), pp. 3146–3154.

[41] Anna Veronika Dorogush, Vasily Ershov, and Andrey Gulin. *CatBoost: gradient boosting with categorical features support*. 2018. arXiv: 1810.11363 [cs.LG].

[42] Srishti Bhasin. "Study of the decay $B^0 \to D^0 D^0 K^+ \pi^-$ with the LHCb experiment". Presented 28 Jul 2021. Apr. 2021. URL: https://cds.cern.ch/record/2781221.

[43] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. "Deep Sparse Rectifier Neural Networks". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Geoffrey Gordon, David Dunson, and Miroslav Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, Apr. 2011, pp. 315–323. URL: https://proceedings.mlr.press/v15/glorot11a.html.

[44] Sebastian Ruder. *An overview of gradient descent optimization algorithms*. 2017. arXiv: 1609.04747 [cs.LG].

[45] Lloyd S. Shapley. *A Value for N-Person Games*. Santa Monica, CA: RAND Corporation, 1952. DOI: 10.7249/P0295.

[46] Scott M Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[47] Christoph Molnar. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. 2019.

[48] Olaf Behnke et al. *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*. 1st. Wiley-VCH, 2013. ISBN: 3527410589.

[49] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York: New York U., Jan. 2012. URL: https://cds.cern.ch/record/1456844.

[50] Lukas Heinrich et al. "pyhf: pure-Python implementation of HistFactory statistical models". In: *Journal of Open Source Software* 6.58 (2021), p. 2823. DOI: 10.21105/joss.02823. URL: https://doi.org/10.21105/joss.02823.

[51] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[52] A L Read. "Presentation of search results: the CLs technique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (Sept. 2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313. URL: https://doi.org/10.1088/0954-3899/28/10/313.

[53] Eilam Gross. "LHC Statistics for Pedestrians". In: (2008). DOI: 10.5170/CERN-2008-001.205. URL: https://cds.cern.ch/record/1099994.

[54] Lukas Heinrich, Matthew Feickert, and Giordon Stark. *pyhf: v0.6.3*. Version 0.6.3. https://github.com/scikit-hep/pyhf/releases/tag/v0.6.3. DOI: 10.5281/zenodo.1169739. URL: https://doi.org/10.5281/zenodo.1169739.

[55] "High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1". In: 4/2017 (2017). Ed. by G. Apollinari et al. DOI: 10.23731/CYRM-2017-004.

[56] ATLAS Collaboration. *Luminosity Public Results Run 2*. 2021. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (visited on 10/07/2021).

[57] Joao Pequenao. *Computer generated image of the whole ATLAS detector*. Mar. 2008. URL: https://cds.cern.ch/record/1095924.

[58] ATLAS Collaboration. *Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment*. General Photo. July 2020. URL: https://cds.cern.ch/record/2723878.

[59] Joao Pequenao. *Computer generated image of the ATLAS inner detector*. Mar. 2008. URL: https://cds.cern.ch/record/1095926.

[60] T. Flick. "The phase II ATLAS Pixel upgrade: the Inner Tracker (ITk)". In: *Journal of Instrumentation* 12.01 (Jan. 2017), pp. C01098–C01098. DOI: 10.1088/1748-0221/12/01/c01098. URL: https://doi.org/10.1088/1748-0221/12/01/c01098.

[61] ATLAS Collaboration. *ATLAS Insertable B-Layer Technical Design Report*. ATLAS-TDR-19. 2010. URL: https://cds.cern.ch/record/1291633. Addendum: ATLAS-TDR-19-ADD-1. 2012. URL: https://cds.cern.ch/record/1451888.

[62] Joao Pequenao. *Computer Generated image of the ATLAS calorimeter*. Mar. 2008. URL: https://cds.cern.ch/record/1095927.

[63] P.A. Zyla et al. "Review of Particle Physics". In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[64] J Bremer. "The Cryogenic System for the ATLAS Liquid Argon Detector". In: (June 2000), 5 p. URL: https://cds.cern.ch/record/449276.

[65] *1998 European School on High-Energy Physics: St. Andrews, UK*. CERN. Geneva: CERN, 1999. DOI: 10.5170/CERN-1999-004. URL: https://cds.cern.ch/record/342530.

[66] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[67] ATLAS Collaboration. "Drift Time Measurement in the ATLAS Liquid Argon Electromagnetic Calorimeter using Cosmic Muons". In: *Eur. Phys. J. C* 70 (2010), p. 755. DOI: 10.1140/epjc/s10052-010-1403-6. arXiv: 1002.4189 [hep-ex].

[68] ATLAS Collaboration. *ATLAS Tile Calorimeter: Technical Design Report*. ATLAS-TDR-3. 1996. URL: https://cds.cern.ch/record/331062.

[69] Bernd Stelzer. "The New Small Wheel Upgrade Project of the ATLAS Experiment". In: *Nuclear and Particle Physics Proceedings* 273-275 (2016). 37th International Conference on High Energy Physics (ICHEP), pp. 1160–1165. ISSN: 2405-6014. DOI: https://doi.org/10.1016/j.nuclphysbps.2015.09.182. URL: https://www.sciencedirect.com/science/article/pii/S2405601415006719.

[70] Markus Elsing et al. "The ATLAS Tier-0: Overview and operational experience". In: *Journal of Physics: Conference Series* 219.7 (Apr. 2010), p. 072011. DOI: 10.1088/1742-6596/219/7/072011. URL: https://doi.org/10.1088/1742-6596/219/7/072011.

[71] ATLAS Collaboration. *Performance of the ATLAS Level-1 topological trigger in Run 2*. Tech. rep. 36 pages in total, author list starting page 20, 10 figures, 3 tables, submitted to EPJC. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2019-02/. Geneva: CERN, May 2021. arXiv: 2105.01416. URL: https://cds.cern.ch/record/2765768.

[72] Duarte Magano et al. *Investigating Quantum Speedup for Track Reconstruction: Classical and Quantum Computational Complexity Analysis*. 2021. arXiv: 2104.11583 [quant-ph].

[73] M Shochet et al. *Fast TracKer (FTK) Technical Design Report*. Tech. rep. ATLAS Fast Tracker Technical Design Report. June 2013. URL: https://cds.cern.ch/record/1552953.

[74] Collaboration, ATLAS. *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*. Tech. rep. CERN-LHCC-2017-020. ATLAS-TDR-029. Geneva: CERN, Sept. 2017. URL: http://cds.cern.ch/record/2285584.

[75] D Emeliyanov and J Howard. "GPU-Based Tracking Algorithms for the ATLAS High-Level Trigger". In: *Journal of Physics: Conference Series* 396.1 (Dec. 2012), p. 012018. DOI: 10.1088/1742-6596/396/1/012018. URL: https://doi.org/10.1088/1742-6596/396/1/012018.

[76] ATLAS Collaboration. "Operation of the ATLAS trigger system in Run 2". In: *JINST* 15 (July 2020). 60 pages in total, author list starting page 44, 19 figures, 3 tables, submitted to JINST, All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2019-04/, P10004. 60 p. DOI: 10.1088/1748-0221/15/10/P10004. arXiv: 2007.12539. URL: https://cds.cern.ch/record/2725146.

[77] I. J. R. Aitchison and A. J. G. Hey. *Gauge theories in particle physics: A practical introduction. Vol. 2: Non-Abelian gauge theories: QCD and the electroweak theory*. CRC Press, 2012. ISBN: 978-1-4665-1307-5.

[78] Kenneth G. Wilson. "Confinement of quarks". In: *Phys. Rev. D* 10 (8 Oct. 1974), pp. 2445–2459. DOI: 10.1103/PhysRevD.10.2445. URL: https://link.aps.org/doi/10.1103/PhysRevD.10.2445.

[79] Gavin P. Salam. *A Practical Seedless Infrared Safe Cone Algorithm*. 2007. arXiv: 0705.2696 [hep-ph].

[80] Yu.L Dokshitzer et al. "Better jet clustering algorithms". In: *Journal of High Energy Physics* 1997.08 (Aug. 1997), pp. 001–001. ISSN: 1029-8479. DOI: 10.1088/1126-6708/1997/08/001. URL: http://dx.doi.org/10.1088/1126-6708/1997/08/001.

[81] M. Wobisch and T. Wengler. "Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering". In: (July 1999). arXiv: 9907280 [hep-ph]. URL: https://arxiv.org/abs/hep-ph/9907280v1.

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

[83] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet user manual". In: *The European Physical Journal C* 72.3 (Mar. 2012). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-012-1896-2. URL: http://dx.doi.org/10.1140/epjc/s10052-012-1896-2.

[84] ATLAS Collaboration. "Jet reconstruction and performance using particle flow with the ATLAS Detector". In: *Eur. Phys. J. C* 77 (2017), p. 466. DOI: 10.1140/epjc/s10052-017-5031-2. arXiv: 1703.10485 [hep-ex].

[85] ATLAS Collaboration. *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. 2020. arXiv: 2007.02645 [hep-ex].

[86] ATLAS Collaboration. *Configuration and performance of the ATLAS b-jet triggers in Run 2*. 2021. arXiv: 2106.03584 [hep-ex].

[87] Michela Paganini and. "Machine Learning Algorithms for b-Jet Tagging at the ATLAS Experiment". In: *Journal of Physics: Conference Series* 1085 (Sept. 2018), p. 042031. DOI: 10.1088/1742-6596/1085/4/042031. URL: https://doi.org/10.1088/1742-6596/1085/4/042031.

[88] Nicole Michelle Hartman, Michael Kagan, and Rafael Teixeira De Lima. *Deep Sets for Flavor Tagging on the ATLAS Experiment*. Tech. rep. Geneva: CERN, June 2020. URL: https://cds.cern.ch/record/2721094.

[89] Andrea Sciandra. *Development of a new Soft Muon Tagger for the identification of b-jets in ATLAS*. Tech. rep. Geneva: CERN, Oct. 2017. DOI: 10.22323/1.314.0768. URL: https://cds.cern.ch/record/2287545.

[90] ATLAS Collaboration. "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13\,\text{TeV}$". In: (2019). arXiv: 1907.05120 [hep-ex].

[91] Tony Tong. "Search for pair production of Higgs bosons in the four b quark final state with the ATLAS detector". Presented 27 Apr 2018. May 2018. URL: https://cds.cern.ch/record/2634867.

[92] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (July 2014). DOI: 10.1007/jhep07(2014)079. URL: https://doi.org/10.1007/jhep07(2014)079.

[93] Manuel Bähr et al. "Herwig++ physics and manual". In: *The European Physical Journal C* 58.4 (Nov. 2008), pp. 639–707. DOI: 10.1140/epjc/s10052-008-0798-9. URL: https://doi.org/10.1140/epjc/s10052-008-0798-9.

[94] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *The European Physical Journal C* 76.4 (Apr. 2016). DOI: 10.1140/epjc/s10052-016-4018-8. URL: https://doi.org/10.1140/epjc/s10052-016-4018-8.

[95] David J. Lange. "The EvtGen particle decay simulation package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1-2 (Apr. 2001), pp. 152–155. DOI: 10.1016/s0168-9002(01)00089-4. URL: https://doi.org/10.1016/s0168-9002(01)00089-4.

[96] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2015.01.024. URL: https://www.sciencedirect.com/science/article/pii/S0010465515000442.

[97] Beojan Stanislaus. "Searching for Beyond the Standard Model Resonances in the $HH \to b\bar{b}b\bar{b}$ Final State Using the ATLAS Detector". Presented 30 Sep 2020. 2020. URL: https://cds.cern.ch/record/2743178.

[98] Alex Rogozhnikov. "Reweighting with Boosted Decision Trees". In: *Journal of Physics: Conference Series* 762 (Oct. 2016), p. 012036. ISSN: 1742-6596. DOI: 10.1088/1742-6596/762/1/012036. URL: http://dx.doi.org/10.1088/1742-6596/762/1/012036.

[99] John Hancock and Taghi Khoshgoftaar. "Survey on categorical data for neural networks". In: *Journal of Big Data* 7 (Apr. 2020). DOI: 10.1186/s40537-020-00305-w.

[100] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[101] François Chollet et al. *Keras*. https://keras.io. 2015.

[102] Martin Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[103] Aaditya Ramdas, Nicolas Garcia, and Marco Cuturi. *On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests*. 2015. arXiv: 1509.02237 [math.ST].

[104] Scott M. Lundberg et al. "From local explanations to global understanding with explainable AI for trees". In: *Nature Machine Intelligence* 2.1 (2020), pp. 2522–5839.

[105] *Summary of non-resonant and resonant Higgs boson pair searches from the ATLAS experiment*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-031. Geneva: CERN, July 2021. URL: https://cds.cern.ch/record/2777013.