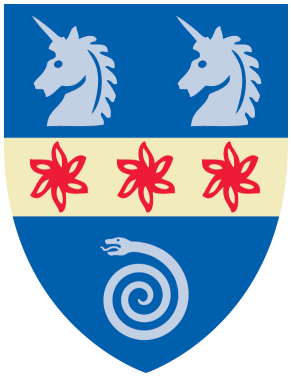
Investigation of Non-Resonant *HH* Production and Couplings in the $b\bar{b}b\bar{b}$ Final State using the ATLAS Detector





James Cameron Grundy,
St Hilda's College,
The University of Oxford

Submitted for the degree of
Doctor of Philosophy
in
Particle Physics

To my parents –

your love and support make this, and everything else, possible.

Acknowledgements

First, I would like to thank my supervisor, Prof. Todd Huffman, for his guidance, support and generosity. Throughout my PhD, I always felt like you had my back, and for that I am very grateful.

I would also like to thank my parents, Brian and Elizabeth Grundy, for their unending love and support. In moments of doubt, stress and worry, you were always there. I would not be anywhere near where I am today without you.

I would like to thank James Frost and Bill Balunas, two fonts of particle physics knowledge, for their invaluable guidance both in and out of our weekly meetings. Thank you to both Beojan Stanislaus and Santiago Paredes Sáenz for their patience in putting up with a new student with many questions and for being role models for me. Thank you to the entire Exotics Group; all of you contributed positively to my PhD experience.

I would like to thank the senior members of the Particle Physics Sub-department for giving me the opportunity to study in Oxford and for running the DPhil programme. Thank you also to the entire Particle Physics Sub-department in Oxford for making me feel at home for the past four years. In particular to Sue Geddes and Kim Proudfoot, without whom the department wouldn't run.

I would like to thank everyone at St Hilda's College for supporting me in my studies. I will miss College dearly, but know I always have a home here.

I would also like to thank everyone on the ATLAS Experiment and at CERN, my scientific home for the past four years. Thank you to the entire $HH \rightarrow b\bar{b}b\bar{b}$

analysis team for their contributions both to my work and the analysis generally. In particular, I'd like to thank analysis conveners Maximilian Swiatlowski and Rafael Teixeira de Lima for spearheading the analysis with a keen collaborative spirit, and my fellow Internal Note editors Nicole Hartman and Rui Zhang for their enthusiasm and tireless work ethic.

I have many great friends, some old and some new, who have contributed so much to my life generally and since arriving in Oxford. To my friends from St Hilda's College – Ailish, Chris, Nick, Jess, David, Jan, Geertje, Rob and Emma. To my friends from the Particle Physics Sub-department – Hayden, Luke and Maria. To my friends in London – Charles, Hamish, Jen, Laura, Lee, Liam, Paula and Will. To the fellas back home – Alex, Connor, Harry, Joe, Michael and Zac. Thank you to you all; I cherish my relationship with every one of you.

I would also like to thank the Science and Technology Facilities Council (STFC) for their financial support.

Finally, I would like to thank the late Dr Barry King. You inspired myself, and many other state school students across Liverpool, to become physicists. You are sorely missed.

Abstract

The full Run 2 ATLAS $HH \to b\bar{b}b\bar{b}$ analysis targeted the simultaneous non-resonant production of two Higgs bosons (HH) in the $b\bar{b}b\bar{b}$ final state [1]. This analysis utilised 126 fb^{-1} of 13 TeV proton-proton collision data from the ATLAS Experiment on CERN's Large Hadron Collider. The increase in data, alongside a program of intensive research and development that I played a significant role in, produced a comprehensive set of important physics results that showed a marked improvement on predecessor ATLAS $HH \to b\bar{b}b\bar{b}$ analyses [2, 3]. An observed (expected) upper limit on HH production via Gluon-Gluon Fusion (ggF) and Vector Boson Fusion (VBF) was set at 5.4 (8.1) times the SM prediction at the 95% confidence level (CL). Constraints on the coupling modifiers κ_λ, κ_V and κ_{2V} (defined as $\kappa = c/c_{SM}$) were set using two different statistical methods. Performing a comparison of the theoretical cross-section to the upper limit on HH production, κ_λ was constrained to $-3.9 < \kappa_\lambda < 11.1$ ($-4.6 < \kappa_\lambda < 10.8$) and κ_{2V} to $-0.03 < \kappa_{2V} < 2.11$ ($-0.05 < \kappa_{2V} < 2.12$) at the 95% CL. Performing a direct measurement of the log likelihood ratio, κ_λ was constrained to $-3.5 < \kappa_\lambda < 11.3$ ($-5.4 < \kappa_\lambda < 11.4$) and κ_{2V} to $0.0 < \kappa_{2V} < 2.1$ ($-0.1 < \kappa_{2V} < 2.1$) at 2σ. Two-dimensional limits were also set on a range of values in κ_{2V}-κ_λ and κ_{2V}-κ_V space at the 95% CL, the former being the first of its kind.

Preface

The focus of this thesis is the ATLAS analysis targeting the simultaneous non-resonant production of two Higgs bosons (the di-Higgs or simply HH) in the $b\bar{b}b\bar{b}$ final state using the full Run 2 ($\mathcal{L}_{INT} = 126\,\text{fb}^{-1}$) dataset [1]. The first three chapters, discussed further below, provide important contextual information regarding particle physics theory and the ATLAS Experiment, whilst the latter chapters consist of a detailed description of all steps in the analysis. The motivation for studying HH production is to improve our understanding of the Higgs sector of the Standard Model (SM). The analysis sets experimental constraints on SM HH production, as well as a number of important interactions. For example, the HHH self-interaction, or *trilinear Higgs self-coupling*, whose associated constant, λ, is present in the SM Higgs potential, a key part of our current understanding of the development of the universe. Also, the HVV and $HHVV$ interactions, which provide insight into the connection between the Higgs boson and the carriers of the weak force, the W and Z bosons. This analysis utilises the largest quantity of data ever recorded by the ATLAS experiment at CERN's Large Hadron Collider (LHC). This, coupled with an intense program of research and development in which I played a significant role, provided a comprehensive set of physics results that showed a substantial improvement on predecessor ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analyses [2, 3].

Chapter 1 provides a theoretical overview of the Standard Model, the Higgs boson and the di-Higgs, followed by a description of HH production and decay at the LHC. Chapter 2 is a review of the design of the ATLAS experiment and the operation of its constituent sub-detectors. Chapter 3 describes the analysis objects and algorithms employed on ATLAS. Chapters 4 and 5 provide a detailed overview of the steps in the full Run 2 ATLAS non-resonant $HH \rightarrow b\bar{b}b\bar{b}$ analysis selection. This includes a description of a specific category optimised for HH production via Vector Boson Fusion (VBF), which I spearheaded and designed. This is a unique feature of this analysis. Chapter 6 provides an account of a study I conducted into including VBF-targeting triggers into the analysis. Chapter 7 describes the procedure for generating the fully data-driven background estimate used in the analysis. Chapter 8 is a description of a study I conducted to redesign the analysis to perform a statistical combination of independent categories targeting the leading HH production modes. Once implemented, this resulted in a significant improvement in analysis sensitivity. Chapter 9 provides a detailed overview of the statistical analysis employed to derive the final results, starting from the basics. Chapter 10 introduces and lists the statistical and systematic sources of uncertainty accounted for in the analysis. Chapter 11 outlines the final results of the analysis, providing a comparison to similar analyses. Finally, Chapter 12 provides a brief summary.

Contents

1 Theory **15**
- 1.1 The Standard Model and The Higgs Boson . . . 15
  - 1.1.1 The Cabibbo–Kobayashi–Maskawa Matrix . . . 21
  - 1.1.2 Deficiencies of the Standard Model . . . 22
- 1.2 H and HH Production at the LHC . . . 23
  - 1.2.1 Gluon-Gluon Fusion HH Production . . . 25
  - 1.2.2 Vector Boson Fusion HH Production . . . 27
  - 1.2.3 HH Decay Modes and Analyses . . . 32

2 The Experiment **35**
- 2.1 CERN and the LHC . . . 35
- 2.2 The ATLAS Detector . . . 38
  - 2.2.1 Inner Detector . . . 40
  - 2.2.2 Calorimeters . . . 45
  - 2.2.3 Muon Spectrometer . . . 49
  - 2.2.4 The Trigger . . . 50

3 Physics Objects and Algorithms **53**
- 3.1 Tracks . . . 53
- 3.2 Vertices . . . 55
- 3.3 Jets . . . 55
  - 3.3.1 The anti-k_t Algorithm . . . 56
  - 3.3.2 Particle Flow . . . 58
  - 3.3.3 JES Calibration . . . 58
  - 3.3.4 JER Calibration . . . 59
- 3.4 Flavour Tagging . . . 60
  - 3.4.1 Flavour Tagging Algorithms . . . 60
  - 3.4.2 Working Points . . . 64
  - 3.4.3 Flavour Tagging Scale Factors . . . 65
- 3.5 Muons . . . 66

3.6 b-Jet Corrections . . . 67

4 Reconstruction and Selection **71**
4.1 Introduction to the $HH \to b\bar{b}b\bar{b}$ Analysis . . . 71
4.2 Data and Monte Carlo Simulation . . . 72
4.3 Triggers . . . 76
4.3.1 Trigger Bucket Strategy . . . 76
4.4 Object Selection . . . 81
4.5 Event Selection . . . 81
4.6 VBF Channel Optimisation . . . 83
4.7 *Two-Tag* and *Four-Tag* Events . . . 84
4.8 *HH* Reconstruction . . . 86
4.8.1 *HH* Pairing Strategy . . . 86
4.9 Background Reduction . . . 89
4.10 Region Definitions . . . 91
4.11 Event Yields . . . 94

5 Categorisation **99**
5.1 ggF Channel Categorisation . . . 101
5.2 VBF Channel Categorisation . . . 102
5.3 m_{HH} Binning . . . 103

6 VBF Trigger Studies **109**
6.1 Motivation . . . 109
6.2 Impact on Signal Yields . . . 111
6.3 Impact on Sensitivity . . . 112
6.4 The VBF Trigger Category . . . 113
6.5 Defining a `2b 2jMjj HT` Trigger Bucket . . . 117
6.6 Summary . . . 120

7 Background Estimation **123**
7.1 The *Two-Tag* Region . . . 124
7.2 Neural Network Reweighting . . . 124
7.3 Background Yields . . . 133
7.4 Background Validation . . . 134
7.4.1 Reversed $|\Delta\eta_{HH}|$ (`rev`$|\Delta\eta_{HH}|$) . . . 135
7.4.2 Shifted Regions . . . 138
7.4.3 `3b1f` . . . 143
7.4.4 QCD Dijet and $t\bar{t}$ Monte Carlo Simulation . . . 144

8 A ggF-VBF Combination for κ_λ-related Results **147**
8.1 *HH* Production Cross-Section and κ_λ . . . 147
8.2 Comparison of Analysis Sensitivity . . . 148
8.3 *HH* Kinematics and κ_λ . . . 151
8.4 Results . . . 152
8.5 Conclusion . . . 154

9 Statistical Analysis **155**

9.1 The Histogram . . . 155
9.2 The Likelihood . . . 155
9.3 Hypothesis Testing . . . 158
9.4 The CL_s Method . . . 160
9.5 Expected Sensitivity . . . 162
9.6 Pulls and Impacts . . . 163

10 Uncertainties **165**
10.1 Statistical Uncertainties . . . 165
10.2 Systematic Uncertainties . . . 167
10.2.1 JES and JER Uncertainties . . . 168
10.2.2 Flavour Tagging Uncertainties . . . 169
10.2.3 Theory Uncertainties . . . 170
10.2.4 Background Estimation Shape Uncertainty . . . 171
10.2.5 `3b1f` Non-Closure Uncertainty . . . 176
10.2.6 Summary of Background Uncertainties . . . 176

11 Results and Discussion **181**
11.1 Checks of Fit Validity . . . 181
11.2 Analysis Results . . . 188
11.2.1 Limit on Standard Model *HH* Production . . . 188
11.2.2 1D Limits on *HH* Coupling Modifiers . . . 188
11.2.3 2D Limits on *HH* Coupling Modifiers . . . 191
11.2.4 Log Likelihood Limits on *HH* Coupling Modifiers . . . 194

12 Summary and Outlook **197**
12.1 Summary . . . 197
12.2 Outlook . . . 198

A Additional Acceptance × Efficiency Plots **201**

B Analysis Selection Cutflows **203**

C Trigger Lists for VBF Trigger Studies **211**

D Additional Background Uncertainty Plots **215**

E Nuisance Parameter Naming Scheme **223**

F Additional Fit Validity Plots **227**

1 Theory

1.1 The Standard Model and The Higgs Boson

The Standard Model (SM) is a mathematical theory that describes the elementary particles that populate our universe and their interactions through three of the four fundamental forces: weak, electromagnetic and strong. It is the foundation on which modern particle physics is built, and has proven to be highly accurate when compared to experimental measurement. Much of the physics program of the Large Hadron Collider (LHC) has been dedicated to investigating its predictions, culminating in the discovery of the Higgs boson in 2012 [4, 5].

The particle content of the SM is shown in Figure 1.1. Particles belong to two groups: *fermions* and *bosons*. Fermions possess $\frac{1}{2}$-integer spin and follow Fermi-Dirac statistics, whilst bosons possess integer spin and follow Bose-Einstein statistics. Fermions can be further separated into two groups: *leptons* and *quarks*. Quarks are the only fermion to interact via the strong force, as they possess colour charge. Bosons can also be further separated into two types: *gauge* and *scalar* bosons. The gauge bosons are the propagators of the fundamental forces: gluon (g) for the strong force, photon (γ) for the electromagnetic force, and W/Z for the weak force. The Higgs boson,

H, is the only fundamental boson in the SM.

The Standard Model is a *gauge quantum field theory*. In quantum field theory, particles are described by quantum states, which are acted on by quantum mechanical operators associated with each point in space-time. The dynamics of the quantum fields are expressed using the Lagrangian formalism. A condensed version of the SM Lagrangian is shown in Equation 1.1.[1]

$$\begin{aligned} \mathcal{L}_{\text{SM}} &= -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ &+ i\bar{\psi}\gamma_\mu D^\mu \psi \\ &+ \psi_i y_{ij} \psi_j \Phi + \text{h.c.} \\ &+ |D_\mu \Phi|^2 - V(\Phi) \end{aligned} \tag{1.1}$$

Here, ψ represents the fermion fields and Φ the Higgs field. $-\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$ is the kinetic term associated with the gauge boson fields, being the contraction of the field strength tensor, $F_{\mu\nu}$. $i\bar{\psi}\gamma_\mu D^\mu \psi$ is the kinematic term for the fermion fields and their interaction with the gauge boson fields (implicitly contained in D_μ). $\psi_i y_{ij} \psi_j \Phi$ describes the interaction between the Higgs and fermion fields, with interaction strength proportional to the Yukawa couplings, y_{ij}. The h.c. simply means the Hermitian conjugate of $\psi_i y_{ij} \psi_j \Phi$. Finally, $|D_\mu \Phi|^2$ describes the kinematics of the Higgs field and its interaction with the gauge boson fields, and $V(\Phi)$ is the Higgs potential, defined in Equation 1.2 and shown in Figure 1.2 for the case when $\mu^2 < 0$. Equation 1.2 is the Landau-Ginzburg version of the Higgs potential, which is the simplest gauge invariant and renormalisable function, but there are other potential candidates.

[1] This version is no doubt familiar to those who have been to CERN and seen it on the souvenir mugs and t-shirts.

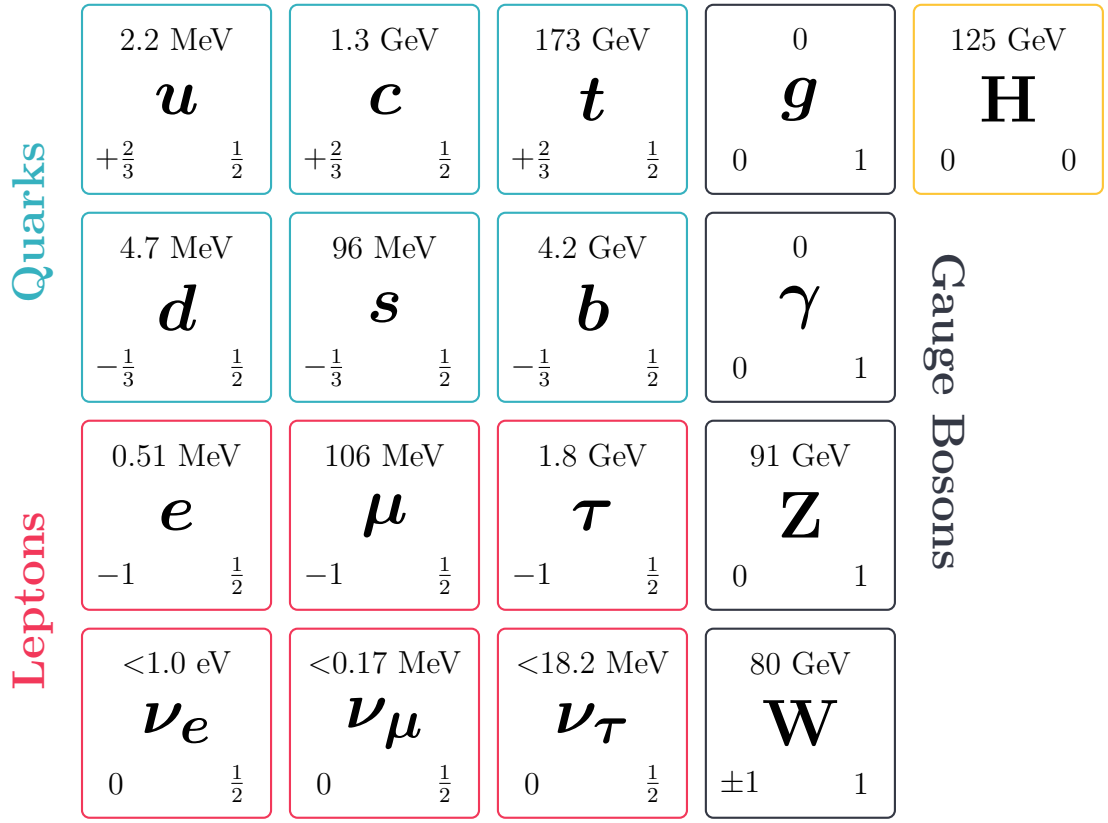


Figure 1.1: The particle content of the Standard Model of Particle Physics. The quarks are labelled in blue, leptons in pink, gauge bosons in navy and Higgs boson in yellow. For each particle, the mass (top), charge (bottom left) and spin (bottom right) are shown.

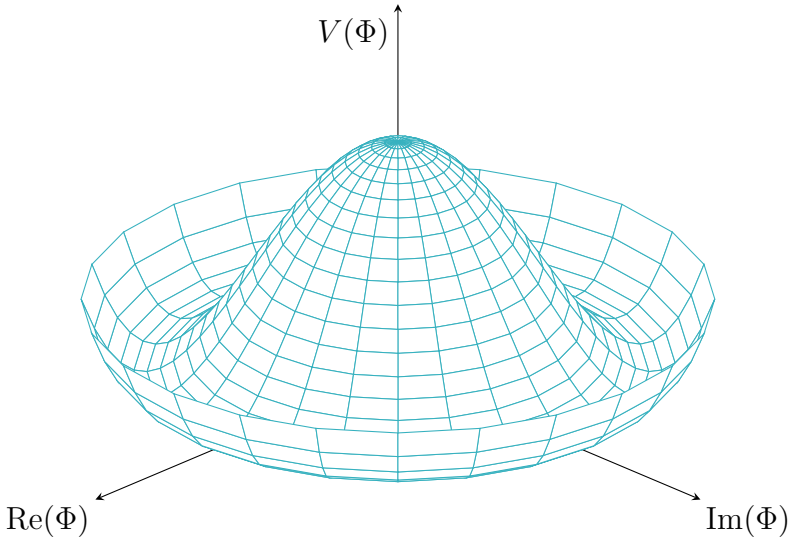


Figure 1.2: Graphical visualisation of the Landau-Ginzburg version of Higgs potential defined in Equation 1.2 for the case where $\mu^2 < 0$. The z-axis shows the value of the potential, $V(\Phi)$, and the x and y-axes show the real and imaginary components of the Higgs field, Φ.

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{1.2}$$

As a gauge theory, the SM Lagrangian is *gauge invariant*, which means it is constructed such that it is unchanged under a group of local transformations (a *gauge group*). To ensure the SM Lagrangian is invariant under a specific gauge group, a number of gauge fields have to be introduced into the Lagrangian. The quanta of each gauge field are the gauge bosons, introduced earlier as g, γ and W/Z. The overall SM gauge group is $\mathrm{U(1)_Y} \times \mathrm{SU(2)_L} \times \mathrm{SU(3)_c}$. $\mathrm{U(1)_Y}$ is associated with Quantum Electrodynamics (QED), and the photon field is created to enforce $\mathrm{U(1)_Y}$ gauge invariance. $\mathrm{SU(2)_L}$ is associated with the electroweak interaction and the W/Z bosons. $\mathrm{SU(3)_c}$ is associated with Quantum Chromodynamics (QCD) and the gluon. By requiring that the SM Lagrangian is gauge invariant, three of the four fundamental forces are naturally integrated into the theory.

A key element missing from Equation 1.1 is a description of the masses of the SM

particles. Mass terms, e.g. $m_Z^2 Z_\mu Z^\mu$, can be introduced into the SM Lagrangian by hand, but would break gauge invariance. The Higgs mechanism is a theoretical framework designed to provide particles with mass [6, 7, 8].[2] It is initiated by a change in state of the universe after passing a critical temperature, T_c, when rapidly cooling post-Big Bang. The value of μ^2 in the Higgs potential changes as the universe passes T_c, causing the spontaneous breaking of the electroweak symmetry, $SU(2)_L \times U(1)_Y$. When $\mu^2 > 0$, the minimum of the Higgs potential, called the vacuum expectation value, $\langle\Phi\rangle$, is 0, which preserves the electroweak symmetry. However, when $\mu^2 < 0$, then potential takes the form shown in Figure 1.2, known commonly as the *Mexican hat* potential. This form has a set of degenerate minima that lie in the brim of the hat, where $\langle\Phi\rangle = \sqrt{-\frac{\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}}$, and all such states are not invariant under $SU(2)_L \times U(1)_Y$.

To provide an idea of how the Higgs mechanism instils particles with mass, the example of the W and Z bosons will be discussed. The electroweak Lagrangian describing only the Higgs boson and its interactions with the gauge bosons is shown in Equation 1.3.[3]

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger D^\mu \Phi + \mu^2 \Phi^\dagger \Phi - \lambda(\Phi^\dagger \Phi)^2 \tag{1.3}$$

In electroweak theory, Φ is defined as an $SU(2)_L$ doublet of positively-charged (ϕ^+) and neutral (ϕ^0) complex scalar Higgs fields, as shown in Equation 1.4.

[2] Also know as the *Englert-Brout-Higgs-Guralnik-Hagen-Kibble* mechanism.

[3] D_μ is the *covariant derivative* and in electroweak theory $= \partial_\mu + i\frac{g}{2}\sigma_j W_\mu^j + i\frac{g'}{2} Y B_\mu$, where g and g' are electroweak coupling constants, Y is the hypercharge, σ_j are the Pauli matrices, and W_μ^j and B_μ can be transformed into $W_\mu^\pm$ and Z_μ.

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.4}$$

When $\mu^2 < 0$, the vacuum state lies somewhere in the brim of the Mexican hat, but the choice is arbitrary. As such, it is possible to make the convenient choice of $\langle\phi_3\rangle = v$ and $\langle\phi_1\rangle = \langle\phi_2\rangle = \langle\phi_4\rangle = 0$. Then, ϕ_3 can be expanded about this minimum as $\phi_3 = H + v$, where H is the neutral scalar Higgs field whose quanta is the Higgs boson. This choice is called the *unitary gauge*, shown in Equation 1.5.

$$\Phi = \begin{pmatrix} 0 \\ H + v \end{pmatrix} \tag{1.5}$$

By inserting the unitary gauge Higgs doublet into the electroweak Lagrangian shown in Equation 1.3 and manipulating the result, Equation 1.6 can be generated.

$$\begin{aligned} \mathcal{L}_\Phi = &\frac{1}{2}\partial_\mu H \partial^\mu H + \mu^2 H^2 + \frac{\lambda}{4}(H^4 + 4vH^3) \\ &+ \frac{1}{4}g^2(H^2 + 2vH + v^2)W_\mu^+ W^{-\mu} \\ &+ \frac{1}{8}(g^2 + g'^2)(H^2 + 2vH + v^2)Z_\mu Z^\mu \end{aligned} \tag{1.6}$$

The first row of Equation 1.6 describes the kinematics of the Higgs field and its mass, $m_H = \sqrt{-2\mu^2}$, alongside the trilinear Higgs self-interaction, $v\lambda H^3$, and quartic self-interaction, $\frac{\lambda}{4}H^4$. The self-interaction terms contain λ, a parameter also present in the Higgs potential. As such, the Higgs self-interactions provide a direct probe of the Higgs potential and the veracity of the Landau-Ginzburg equation. The second and third rows describe the interaction between the Higgs and W/Z bosons alongside newly-introduced W and Z mass terms, where $m_W = \frac{1}{2}gv$ and $m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}$.

The Higgs mechanism described here can be extended to provide fermions with mass, with the exception of neutrinos, through the additional of Yukawa terms to the Lagrangian, which describe interactions between Φ and the fermionic fields, ψ, as shown in the third row of Equation 1.1.

1.1.1 The Cabibbo–Kobayashi–Maskawa Matrix

In the Standard Model, a quark of one flavour can transition into that of another via the weak interaction, e.g. $b \to cW^-$. This is the primary mechanism behind the decay of hadrons, colourless composite particles comprising quarks, which are readily produced at the LHC. The Cabibbo–Kobayashi–Maskawa (CKM) matrix describes the probability of flavour-changing quark transitions. The probability of a quark of flavour j to transition to a quark of flavour i is proportional to the square of a corresponding element of the CKM matrix, $|V_{ij}|^2$. For example, the probability of the $b \to cW^-$ transition is proportional to $|V_{cb}|^2$. The values of the CKM matrix elements are shown in Equation 1.7 [9].[4]

$$V_{CKM} = \begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.974 & 0.225 & 0.00369 \\ 0.225 & 0.973 & 0.0418 \\ 0.00857 & 0.0411 & 0.999 \end{bmatrix} \tag{1.7}$$

An important property of the CKM matrix is unitarity, i.e. $V_{CKM}^{\dagger} V_{CKM} = I$. For this to hold, the sum of the square elements in a row or column must equal one, e.g. $|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = 1$.

The probability of different flavour-changing quark transitions differs greatly due to the significant variation in the CKM matrix. For example, a top quark decays to

[4]Equation 1.7 was generated using a global fit to all available CKM matrix measurements and imposing the SM constraints. The CKM matrix element values have been rounded to three significant figures and uncertainties have been excluded for ease of reading.

a bottom quark with near 100% probability as $|V_{tb}| \sim 1$; however, it decays to a down quark with <1% probability as $|V_{td}| = 0.00857$. This is also important for understanding the likelihood of different decay pathways for hadrons.

1.1.2 Deficiencies of the Standard Model

The Standard Model is an incredibly successful physical theory, describing three of the four fundamental forces and all known quarks, leptons and bosons. Since its inception, its predictions have consistently demonstrated remarkable accuracy when compared to experimental results. However, it does have deficiencies, a number of important examples of which will now be described. The SM provides no description of the gravitational force, with attempts to integrate the most successful theory of gravity, General Relatively, being unsuccessful. Neutrinos are massless in the SM, in direct disagreement with the observation of neutrino flavour mixing, the process of a neutrino of one flavour spontaneously transforming into that of another, which requires the mixing particle to be massive. Cosmological observations have shown two significant deficiencies of the SM. First, that the SM only describes ∼5% of the matter content of the universe, with the rest comprising Dark Matter and Dark Energy. The SM contains no viable Dark Matter candidate. Second, the majority of the particle content of the universe is matter rather than antimatter, which cannot be explained in terms of the SM, with level of *CP*-violation being insufficient to produce the observed levels of matter-antimatter asymmetry.[5] Additionally, there are a number of important theoretical problems with the Standard Model, such as the Higgs mass Hierarchy Problem, which poses the question why the Higgs mass is ∼125 GeV when the expectation is large quantum corrections should make its mass many orders of magnitude larger.

[5] *CP* is a combination of the Charge (C) and Parity (P) symmetries, which correspond to flipping the sign of particle charge, e.g. $- \rightarrow +$, and particle position, e.g. $\{x, y, z\} \rightarrow \{-x, -y, -z\}$. *CP*-violation is the breaking of the *CP* symmetry. Note, the flipping of charge in the C symmetry is equivalent to changing a particle into its antiparticle, which explains its association with matter-antimatter asymmetry.

1.2 *H* and *HH* Production at the LHC

In the summer of 2012, a neutral boson of mass ∼125 GeV, consistent with the predictions for the SM Higgs boson, was discovered by the ATLAS [4] and CMS experiments [5] on the Large Hadron Collider at CERN. Since then, a significant fraction of the research conducted by ATLAS and CMS has focused on studying the properties of this newly-discovered particle. This has yielded a wealth of information about the Higgs boson and its interactions, all of which is in agreement with Standard Model prediction. Ref. [10] is a compendium of the properties of the Higgs boson measured by the ATLAS detector as of 2022 and will be briefly summarised here. The Higgs mass is known to ∼0.1%, with the combined ATLAS and CMS measurement equal to 125.09 GeV. It has been observed that the Higgs boson is spin 0 and *CP* even, i.e. its interactions are unchanged under the *CP* symmetry. The cross-section of the three leading H production processes has been measured to between approximately 7 and 24%. These are, in order of theoretical cross-section, Gluon-Gluon Fusion, Vector Boson Fusion and Higgstrahlung. The Higgs branching fractions to $\gamma\gamma$, ZZ, $W^{\pm}W^{\mp}$, $\tau^{+}\tau^{-}$ and $b\bar{b}$ have been measured to between ∼10 and 15% accuracy. The coupling strength of Higgs to top, bottom, Z, W, τ and μ have been measured and are shown versus their mass in Figure 1.3. In the SM, the Higgs boson couples to particle mass, with coupling strength proportional to mass for fermions and mass squared for bosons. Figure 1.3 neatly agrees with this prediction, showing the linear relationship between coupling strength and fermion mass/boson mass squared.

One exciting avenue of Higgs research is the study of the trilinear Higgs self-interaction, shown in Figure 1.4, as this provides a direct probe of the Higgs potential and the process of electroweak symmetry breaking. Here, the coupling modifier, which is the ratio of the coupling to its SM prediction, $\kappa = c/c_{SM}$, is shown instead of the coupling, c. The use of coupling modifiers provides a framework to parame-

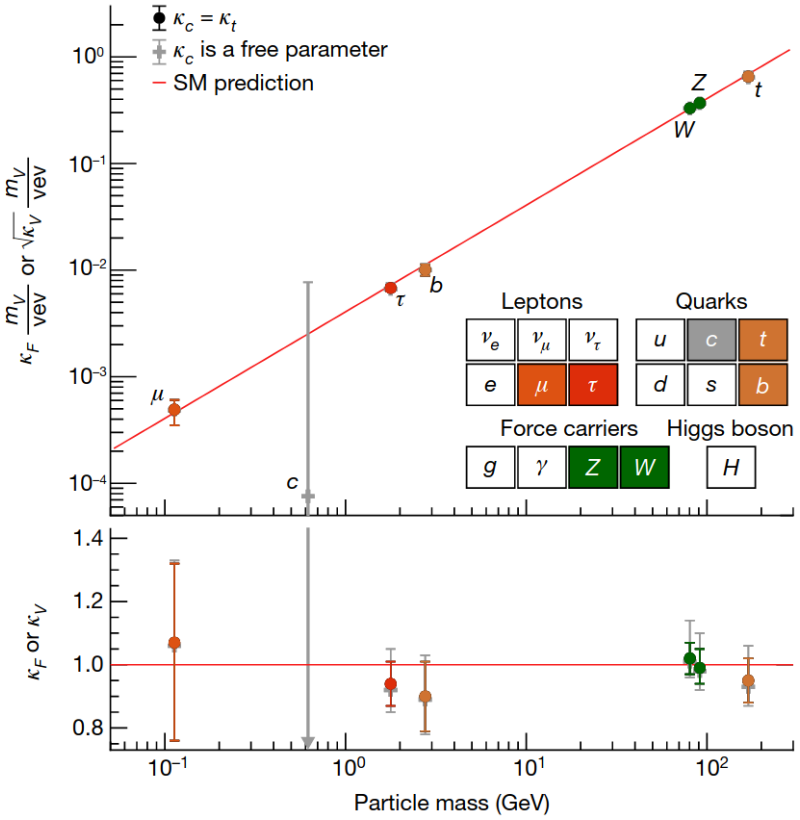


Figure 1.3: Top: plot of particle mass over the vacuum expectation value (VEV) multiplied by the measured coupling modifier for fermions (labelled with F) and the square root of the measured coupling modifier for bosons (labelled with V) versus particle mass. Bottom: coupling modifier value versus particle mass. The coupling modifiers, κ, are defined as the ratio of the measured coupling strength to the Standard Model prediction. A linear red line is included to demonstrate the agreement to the theoretical prediction that Higgs bosons couples to mass for fermions and mass squared for bosons. Two fit scenarios are shown: one where $\kappa_c = \kappa_t$ (coloured circle markers) and one where κ_c is floated in the fit (grey cross markers). The Higgs-charm coupling has not been experimentally measured but a observed (expected) limit at 8.5 (12.4) has been placed on κ_c by ATLAS [11].

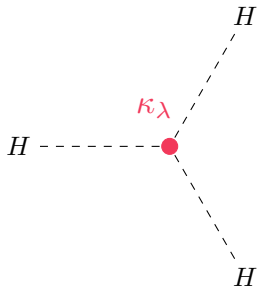


Figure 1.4: Feynman diagram of the interaction of three Higgs bosons, known as the trilinear Higgs self-interaction, with the trilinear Higgs self-coupling modifier, $\kappa_\lambda = \lambda/\lambda_{SM}$, highlighted in pink.

terise deviations from the Standard Model, and its use is standard practice in Higgs boson analyses in ATLAS [12].

1.2.1 Gluon-Gluon Fusion *HH* Production

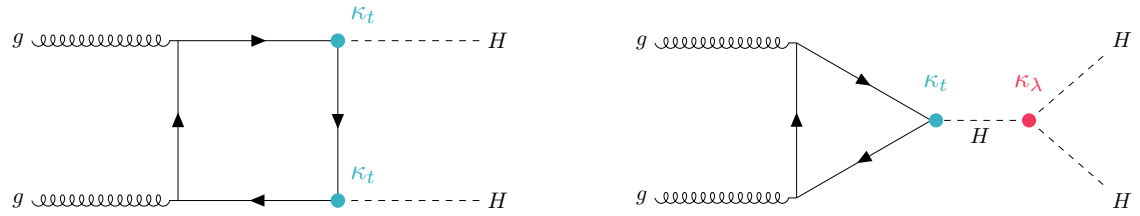
HH production occurs via a number of mechanisms at the LHC. The leading production mode is Gluon-Gluon Fusion (ggF), with a SM cross-section of 31.05 fb [13]. At tree-level, there are two Feynman diagrams that describe ggF, shown in Figure 1.5. These are known as the *box* and *triangle* diagrams, where only the triangle diagram has sensitivity to the trilinear Higgs self-coupling.

The ggF *HH* amplitude can be expressed in terms of the individual contributions from the triangle and box diagrams, as shown in Equation 1.8 [14].

$$\mathcal{A}(\kappa_t, \kappa_\lambda) = \kappa_t^2 \mathcal{A}_{\text{box}} + \kappa_t \kappa_\lambda \mathcal{A}_{\text{tri}} \tag{1.8}$$

Taking the absolute square of Equation 1.8 and setting κ_t to its SM value of one gives:[6]

[6]The coupling modifier κ_t is not normally of interest to *HH* analyses, as analyses targeting production of a single Higgs boson have measured it to high precision. For example, the 80 fb^{-1} combined ATLAS Higgs boson measurement determined $\kappa_t = 1.09^{+0.15}_{-0.14}$ [15].



(a) The *box* diagram.

(b) The *triangle* diagram.

Figure 1.5: The two tree-level Feynman diagrams for di-Higgs production via Gluon-Gluon Fusion. Shown in blue is the top Yukawa coupling modifier, κ_t, and in pink is the trilinear Higgs self-coupling modifier, κ_λ.

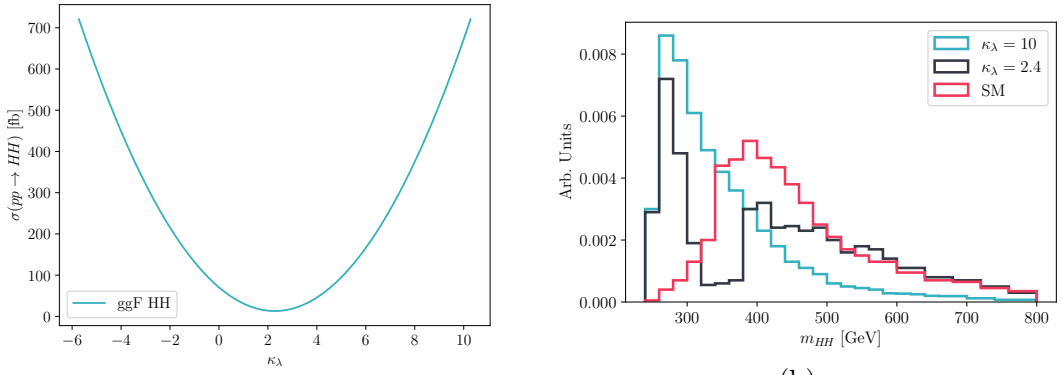
$$\sigma_{ggF}(pp \to HH) \propto |\mathcal{A}_{\text{box}}|^2 + 2\kappa_\lambda \text{Re}(\mathcal{A}^*_{\text{box}}\mathcal{A}_{\text{tri}}) + \kappa_\lambda^2 |\mathcal{A}_{\text{tri}}|^2 \tag{1.9}$$

A number of Beyond the Standard Model (BSM) theories predict deviations from $\kappa_\lambda = 1$ [16, 17, 18, 19]. The term corresponding to the triangle diagram, $\kappa_\lambda^2|\mathcal{A}_{\text{tri}}|^2$, will dominate if κ_λ is equal to a large positive or negative value. On the other hand, if κ_λ is set to a small value, the term corresponding to the box diagram, $|\mathcal{A}_{\text{box}}|^2$, will dominate. The term in the centre, $2\kappa_\lambda \text{Re}(\mathcal{A}^*_{\text{box}}\mathcal{A}_{\text{tri}})$, causes destructive interference, which reduces the SM *HH* cross-section and peaks at $\kappa_\lambda \approx 2.4$.

The ggF *HH* cross-section is shown as a function of κ_λ in Figure 1.6a.[7] The minimum cross-section corresponds to the point of maximum destructive interference, $\kappa_\lambda \approx 2.4$. Deviations away from this minimum point result in significant increases in ggF *HH* cross-section. As such, a large range of potential κ_λ values have been excluded by experimental measurement. The current best limit on κ_λ from an ATLAS analysis is $-1.0 < \kappa_\lambda < 6.6$ at the 95% confidence level [20].

Figure 1.6b shows the ggF *HH* mass distribution for $\kappa_\lambda = 1$, 2.4 and 10, and demonstrates how different values of κ_λ, and the resulting variation in magnitude of the different diagrams, affects the event kinematics.[8] At $\kappa_\lambda = 1$, the box diagram

[7] The cross-section values are equal to those that the $HH \to b\bar{b}b\bar{b}$ analysis Monte Carlo simulation samples are scaled to, as described in Section 4.2.



(a) Cross-section

(b) m_{HH}

Figure 1.6: The value of the cross-section, σ, and distribution of HH mass, m_{HH}, for Gluon-Gluon Fusion HH production at different values of the trilinear Higgs self-coupling modifier, κ_λ, at $\sqrt{s} = 13\,\text{TeV}$ [21]. The m_{HH} distributions are normalised to unity.

dominates, resulting in a relatively hard m_{HH} distribution. At $\kappa_\lambda = 10$, the triangle diagram dominates, changing the kinematics and producing a relatively soft m_{HH} distribution. At $\kappa_\lambda \approx 2.4$, destructive interference is at a maximum, causing the yield to drop significantly at the centre of the m_{HH} distribution and the individual contributions at low and high mass from the triangle and box diagrams respectively to become visible.

1.2.2 Vector Boson Fusion *HH* Production

The second leading HH production mode is Vector Boson Fusion (VBF). VBF is characterised by two high-momentum and well-separated jets present in the final state alongside the HH. The SM cross-section of VBF is over an order of magnitude lower than ggF, at 1.726 fb [22]. The three tree-level Feynman diagrams for VBF HH production are shown in Figure 1.7.

The VBF HH interaction depends on three couplings – κ_V, κ_{2V} and κ_λ. κ_V has been tightly constrained by analyses targeting production of a single Higgs boson; for example, the $80\,\text{fb}^{-1}$ combined ATLAS Higgs boson measurement determined

[8]The m_{HH} distributions are from Monte Carlo simulation samples using the settings described in Section 4.2, with the Herwig7 package performing the hadronisation and showering.

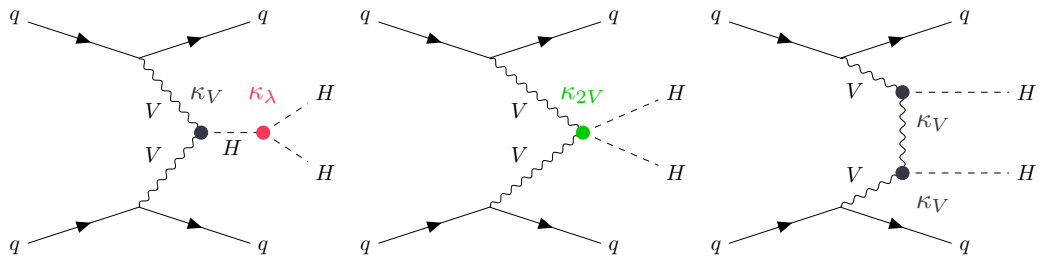


Figure 1.7: The three tree-level Feynman diagrams for di-Higgs production via Vector Boson Fusion. Shown in dark blue is the HVV vertex coupling modifier, κ_V, in pink is the trilinear Higgs self-coupling modifier, κ_λ, and in green is the $HHVV$ vertex coupling modifier, κ_{2V}.

$\kappa_V = 1.05 \pm 0.04$ [15]. Additionally, experimental analyses focusing on κ_λ are normally tailored to the ggF HH production mode, given its superior cross-section. However, VBF HH does provide a unique and sensitive probe of the $HHVV$ vertex, as explained below.

In the Standard Model, the second and third diagrams shown in Figure 1.7 cancel to ensure perturbative unitarity, resulting in the relatively small cross-section. However, in BSM theories where κ_V and κ_{2V} are not equal, the amplitude can be significantly larger [23, 24, 25]. Using the general parameterisation of the couplings of the Higgs boson to vector bosons and fermions in Ref. [26], the dominant contribution to the amplitude is seen to derive from the hard scatter of a pair of longitudinally-polarised vector bosons, as shown in Equation 1.10 [27].[9]

$$\mathcal{A}(V_L V_L \to hh) \simeq \frac{\hat{s}}{v^2}(\kappa_{2V} - \kappa_V^2) \tag{1.10}$$

Here, $\hat{s}$ is the centre of mass energy and v is the vacuum expectation value. If $\kappa_V \neq \kappa_{2V}$, the overall amplitude is significantly increased, and grows linearly with centre of mass energy, resulting in hard HH kinematics.

[9] This is only the dominant contribution when $m_W \ll \hat{s} \equiv m_h \ll \Lambda$, where Λ is the cut-off scale.

Figure 1.8 shows the variation in the VBF *HH* cross-section with κ_λ, alongside normalised distributions of key kinematic variables for three representative values of κ_λ.[10] The minimum VBF *HH* cross-section with respect to κ_λ corresponds to $\kappa_\lambda = 1.56$, close to the SM value. If κ_λ is shifted to large positive or negative values, the VBF *HH* cross-section increases significantly. At the same time, the m_{HH} distribution becomes softer, centring itself at lower m_{HH} values. The kinematics of the initial scatter jets characteristic of the VBF topology are roughly unchanged with κ_λ.

Figure 1.9 shows the variation in the VBF *HH* cross-section with κ_{2V}, alongside normalised distributions of key kinematic variables for three representative values of κ_{2V}.[11] The minimum VBF *HH* cross-section with respect to κ_{2V} corresponds to $\kappa_{2V} = 1.17$, also close to the SM value. If κ_{2V} is shifted to large positive or negative values, the VBF *HH* cross-section increases by several orders of magnitude. At the same time, the m_{HH} distribution becomes harder. As with the variation in κ_λ, the kinematics of the initial scatter jets are roughly unchanged with κ_{2V}.

The tightest experimental constraints on κ_{2V} come from a $HH \to b\bar{b}b\bar{b}$ CMS analysis that targeted highly-Lorentz boosted final states to reject background, and found $0.62 < \kappa_{2V} < 1.41$ at the 95% confidence level [28].

The derivation of Equation 1.10 is shown in Ref. [27], which builds on the framework laid out in Ref. [26].

[10] The cross-section values are equal to those used in the $HH \to b\bar{b}b\bar{b}$ analysis, as described in Section 4.2. The m_{HH} distributions are also from the Monte Carlo simulation samples described in Section 4.2, with a simple preselection requiring at least four jets, and that all jets possess $p_T \geq 25\,\text{GeV}$ and $|\eta| \leq 5.0$.

[11] See Footnote 10.

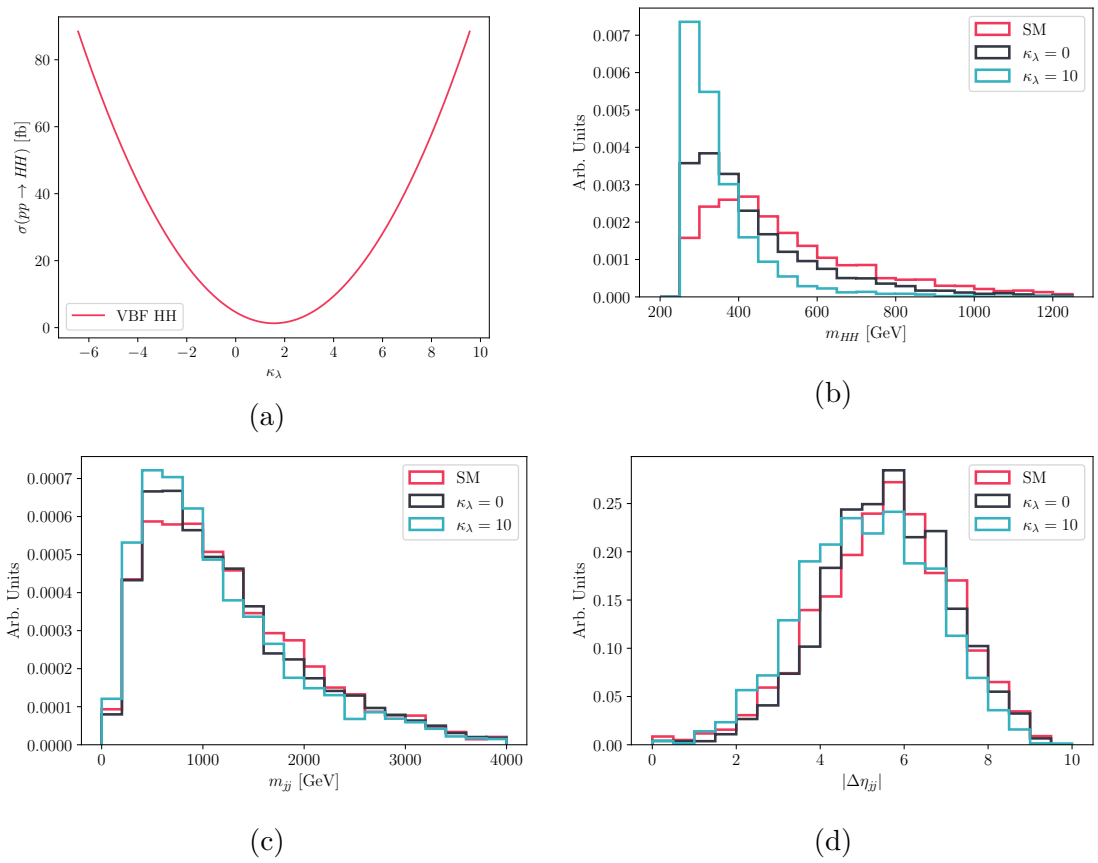


Figure 1.8: Cross-section and kinematics of VBF *HH* events for a selection of κ_λ values. The top left plot shows the VBF *HH* cross-section as a function of κ_λ. The top right, bottom left and bottom right show the *HH* mass alongside the di-jet mass and pseudorapidity gap for the two initial scatter jets for three key values of κ_λ. The m_{HH}, m_{jj} and $|\Delta\eta_{jj}|$ distributions are normalised to unity.

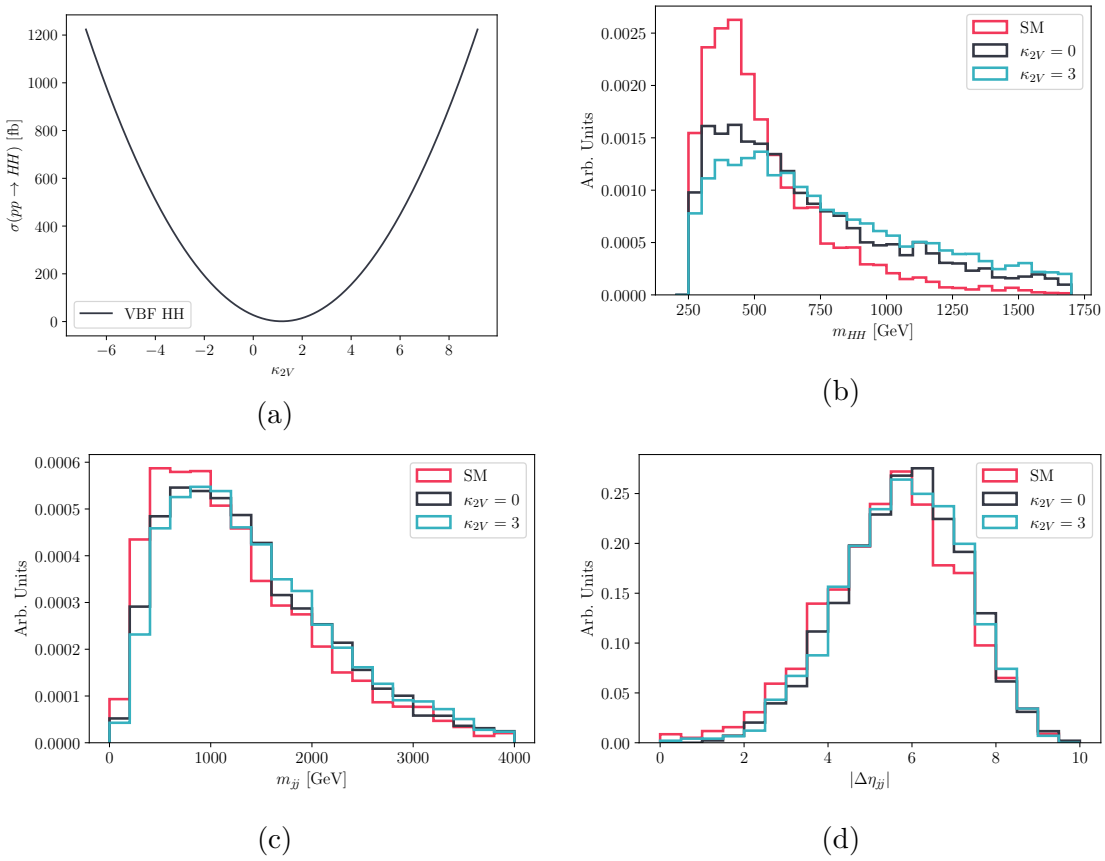


Figure 1.9: Cross-section and kinematics of VBF *HH* events for a selection of κ_{2V} values. The top left plot shows the VBF *HH* cross-section as a function of κ_{2V}. The top right, bottom left and bottom right show the *HH* mass alongside the di-jet mass and pseudorapidity gap for the two initial scatter jets for three key values of κ_{2V}. The m_{HH}, m_{jj} and $|\Delta\eta_{jj}|$ distributions are normalised to unity.

1.2.3 *HH* Decay Modes and Analyses

As introduced previously, the Higgs boson interacts with all massive Standard Model particles, with the coupling strength depending on the mass of the particle the Higgs boson is interacting with. Satisfying the requirement that the mass of the initial state cannot exceed the final state, $H \to b\bar{b}$ is the leading Higgs boson decay mode, with a branching ratio of 58.2% [29]. By extension, the leading *HH* decay mode is $HH \to b\bar{b}b\bar{b}$, with a branching ratio of 33.9%. Figure 1.10 displays the branching ratios for the leading *HH* decay modes and those relevant for current *HH* analyses on ATLAS.

Individual ATLAS analyses generally study *HH* by targeting a single decay mode. The three SM *HH* decay modes that ATLAS analyses have the highest sensitivity to are $HH \to b\bar{b}b\bar{b}$, $HH \to b\bar{b}\tau^-\tau^+$ and $HH \to b\bar{b}\gamma\gamma$. Targeting each has its own advantages and disadvantages, as discussed below.

The LHC is a proton-proton collider, so, as protons are made up of quarks and gluons, the largest background for detectors based on the LHC comes from hadronic activity initiated by the strong interaction. Therefore, even though $HH \to b\bar{b}b\bar{b}$ has the largest branching ratio of all *HH* decay modes, analysis sensitivity is diluted by significant levels of background contamination, as $b\bar{b}b\bar{b}$ is a fully-hadronic final state. $HH \to b\bar{b}\gamma\gamma$, by contrast, has the opposite problem. The two photons in its final state produce a clean and distinct detector signature, but the low $H \to \gamma\gamma$ branching ratio (0.227%) means it has a low event yield. $HH \to b\bar{b}\tau^-\tau^+$ is seen as a middle ground between the two, having a relatively large branching ratio at 7.3% and $H \to \tau^+\tau^-$ being a relatively clean signature.

As it has the largest branching ratio, $HH \to b\bar{b}b\bar{b}$ provides the best sensitivity to *HH* events with larger m_{HH} values, where the other channels can possess a negligible

number of events. As it has a clean signature, $HH \rightarrow b\bar{b}\gamma\gamma$ provides the best sensitivity to HH events with smaller m_{HH} values, where the other channels are swamped by the significant QCD background contamination. This is relevant for the setting of constraints on κ_λ and κ_{2V}. The hardening of the m_{HH} spectrum with varying κ_{2V} (see Figure 1.9b) means the $HH \rightarrow b\bar{b}b\bar{b}$ analysis has the best sensitivity to BSM κ_{2V} HH production. Again, in contrast, the softening of the m_{HH} spectrum with varying κ_λ (see Figure 1.6b) means the $HH \rightarrow b\bar{b}\gamma\gamma$ analysis has the best sensitivity to BSM κ_λ HH production.

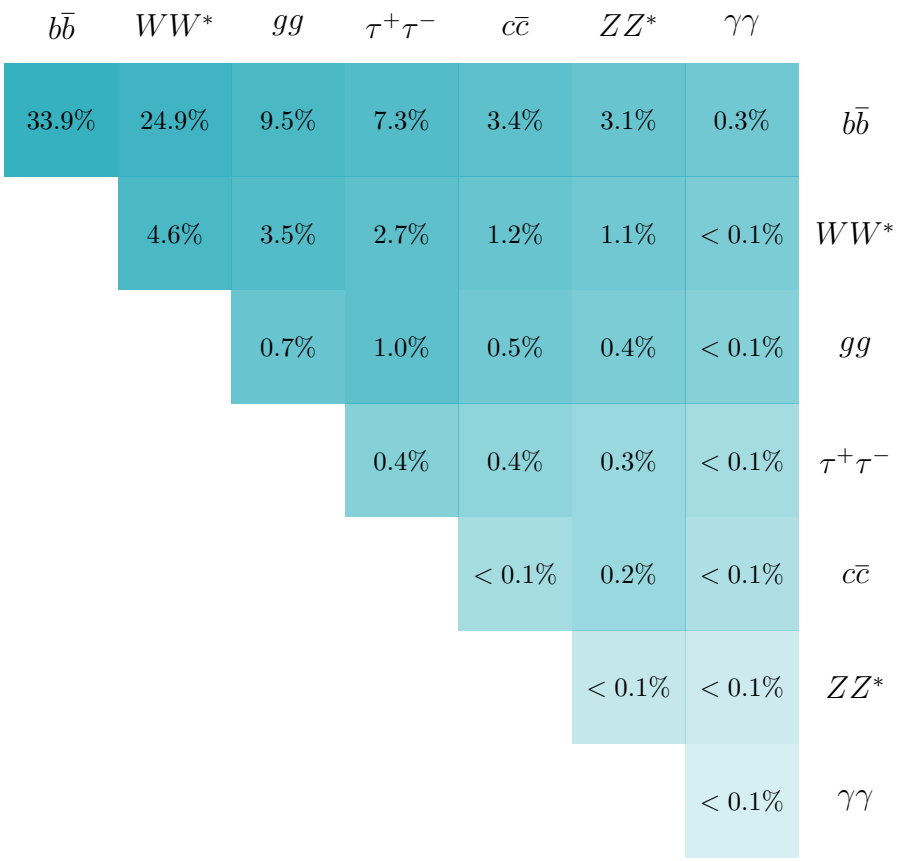


Figure 1.10: The branching ratios of the HH decay modes for Higgs bosons with mass $m_H = 125\,\text{GeV}$ [30]. Lower probability modes exist but are not included for clarity.

2 The Experiment

2.1 CERN and the LHC

The European Organisation for Nuclear Research (CERN) is the world's largest particle physics laboratory, based on the outskirts of Geneva, Switzerland. CERN is home to a vast and diverse particle accelerator complex, which facilitates the study of high energy physics for a number of large particle physics experiments. Figure 2.1 shows the various accelerator systems in the CERN accelerator complex.

The largest accelerator is the Large Hadron Collider (LHC), which is capable of accelerating protons (and heavy ions) to velocities close to the speed of light. At a maximum centre-of-mass energy of $\sqrt{s}$ =13.6 TeV (as of 2022), it is the most powerful particle accelerator in the world. The LHC is a synchrotron, where charged particles are accelerated around a ring of fixed circumference. Acceleration is performed by a series of 16 radiofrequency (RF) cavities, which contain an oscillating electric field synchronised such that the field points in the correct direction to accelerate the protons as they pass through. Hundreds of dipole and quadrupole magnets are stationed incrementally around the LHC ring to bend and focus the proton beam respectively. The magnet systems operate at ultra-high field strengths, with

the dipole magnets reaching 8.3 T for beam energies of 6.5 TeV. The acceleration and magnet systems rely on superconducting technology so are kept at ultra-low temperatures of −271.3 °C using liquid helium.

Protons reach the LHC ring after a multistage acceleration process (see Figure 2.1 for reference). This starts with negative hydrogen ions being accelerated to 160 MeV by LINAC4 and fed into the Proton Synchrotron Booster (BOOSTER), where the ions are further accelerated and stripped of their electrons to leave the hydrogen nuclei, protons. The accelerated protons are then injected in the Proton Synchrotron (PS) for further acceleration before being fed into the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. Finally, the protons circulating the SPS are injected into the LHC, where they are accelerated to the maximum centre-of-mass energy. Protons are grouped into bunches of $\sim 10^{11}$ protons to increase likelihood of interaction. Two sets of bunches transverse the LHC ring in opposite directions in separate ultra-high vacuum beam pipes before intersecting at the collision points. Each bunch is ∼30 cm in length and ∼10 µm in width when maximally focused. 2808 equally-spaced bunches of protons are accelerated in opposite directions around the LHC ring, which is 27 km in circumference, and are collided over the course of several hours at four separate points. At these points are four particle detectors that measure the spray of particles produced from the proton-proton (or ion-ion) collisions. There are two general purpose detectors – ATLAS [31, 32] and CMS [33] – that have a broad particle physics research remit. The ALICE experiment [34] studies quark-gluon plasma through heavy-ion collisions and the LHCb [35] experiment specialises in the physics of the bottom quark. The focus of this chapter will be the ATLAS detector.

The LHC programme is composed of run periods were the LHC is operational and data is collected and shutdown periods were the LHC is not operating and upgrades or fixes are made to the accelerators and detectors. LHC Run 1 lasted from 2009–

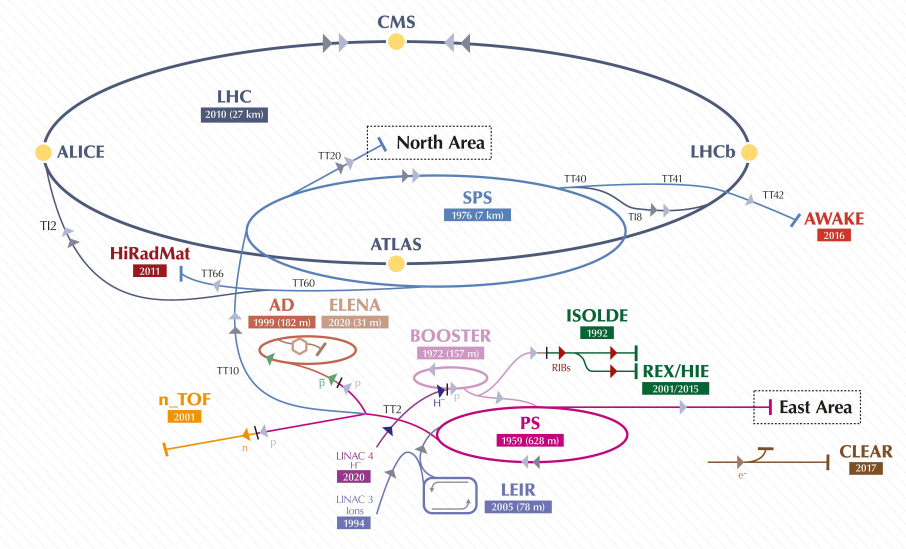


Figure 2.1: Schematic of the CERN accelerator complex. Each individual accelerator system is represented by a different colour. The lines joining the different accelerators highlight that particles are accelerated in a series of stages using increasingly more powerful accelerators. The largest dark blue ring is the Large Hadron Collider (LHC), with the four LHC experiments – CMS, LHCb, ATLAS and ALICE – marked with a yellow dot.

13, with $5.08\,\text{fb}^{-1}$ of proton-proton collision data collected at $\sqrt{s} = 7\,\text{TeV}$ in 2011 and $21.3\,\text{fb}^{-1}$ of collision data collected at $\sqrt{s} = 8\,\text{TeV}$ in 2012. This period saw the famous discovery of the Higgs boson, a landmark moment for the LHC. Run 2 lasted from 2015–18 and saw a significant increase in energy to $\sqrt{s} = 13\,\text{TeV}$ and a large increase in the amount of data recorded at $139\,\text{fb}^{-1}$. The increased energy and luminosity of Run 2 allowed physicists to build on the work on Run 1; for example, measuring a number of important properties of the Higgs boson, such as its couplings to a number of the heavier particles in the SM. Run 3 started in 2022 and is due to last until 2026 with a small increase in $\sqrt{s}$ to 13.6 TeV and a projected doubling in the amount of data recorded to $\sim 300\,\text{fb}^{-1}$. After Run 3, the LHC will undergo an extensive upgrade to move to ultra-high luminosities, becoming the High Luminosity LHC (HL-LHC). A modest increase in energy to $\sqrt{s} = 14\,\text{TeV}$ is planned alongside a significant increase in projected amounts of data recorded to $\sim 3000\,\text{fb}^{-1}$ by the end of its lifetime. A key physics goal of Run 3 and the HL-LHC is the study of HH production and the trilinear Higgs self-coupling, which are the focus of this thesis.

2.2 The ATLAS Detector

The ATLAS detector, an acronym of **A** **T**oroidal **L**HC **A**pparatu**S**, is one of two general purpose particle detectors based on the LHC, alongside its sister experiment CMS. It is a *hermetic* detector, meaning it is designed to surround as much of the proton-proton interaction point as possible and to measure all known decay products of the particles produced post-collision.[1] The detector measures the trajectory, momentum and energy of particles that travel through it. If these particles are the product of an earlier particle's decay, the earlier particle can be reconstructed. The detector record of each collision of the proton bunches is called an *event*, with approximately 40 million events occurring each second on the LHC. The series of events that are saved, the data, are the starting point of all research on ATLAS. With over 5500 scientists working on ATLAS as of March 2022, the experiment has an extensive particle physics research output. A detailed description of the ATLAS detector can be found in Ref. [36], which informs much of this chapter.

ATLAS comprises three sub-detectors, each layered on top of one another in concentric cylinders around the interaction point: the Inner Detector (ID), the Calorimeters (ECAL and HCAL) and the Muon Spectrometer (MS). There is a 2 T solenoid magnet surrounding the ID and and three toroid magnets within the MS. A general overview of the ATLAS detector is shown in in Figure 2.2. These individual systems will be discussed in detail in individual sections in this chapter.

ATLAS uses a right-handed spherical coordinate system, as shown in Figure 2.3. The origin lies at the interaction point, with the z-axis pointing along the beamline, the x-axis pointing towards the centre of the LHC ring and the y-axis pointing skyward. Typically, the azimuthal angle, θ, is not used to measure the angle of

[1] The ATLAS detector cannot measure neutrinos directly; however, it is able to reconstruct neutrinos as the missing energy in an interaction, as it is able to measure all other particles.

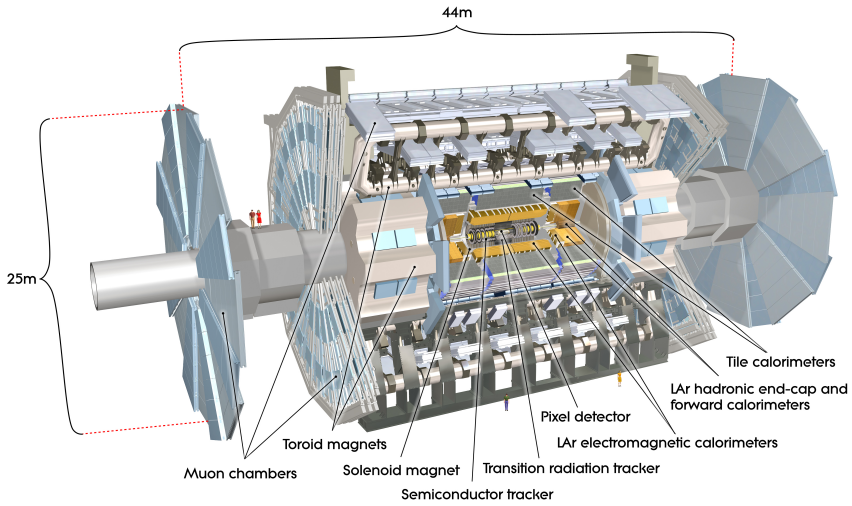


Figure 2.2: Schematic of the ATLAS detector, with the infrastructure and sub-detector constituents labelled. Bunches of protons accelerated by the LHC pass through the horizontal grey pipe bisecting the detector and collide in the centre. The spray of particles produced by the collision pass through the detector and are recorded as signal. Two figures are shown on top of the beam pipe to the left of the central structure to indicate scale [37].

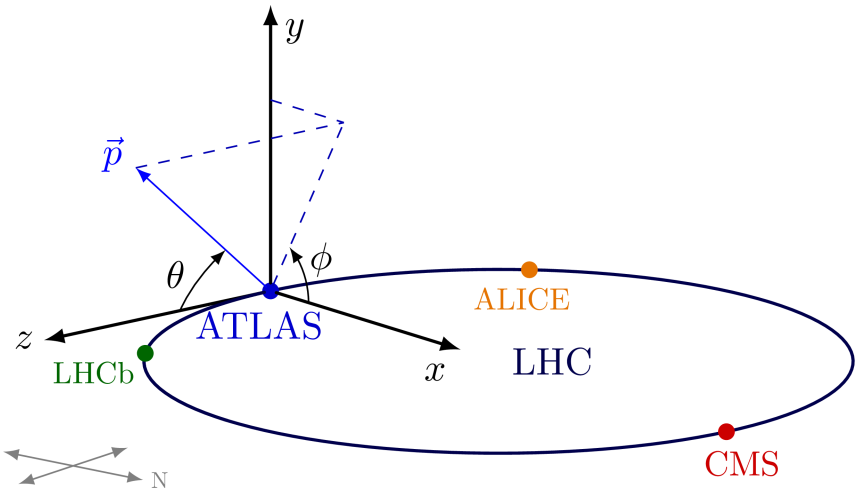


Figure 2.3: Schematic showing the ATLAS right-handed spherical coordinate system. The black ring represents the LHC, with the individual coloured points showing the four experiments. The ATLAS experiment is shown in blue. The azimuthal angle, θ, is measured from the z-axis, which lies along the beam line. The polar angle, ϕ, is measured from the x-axis, which points to the centre of the LHC. The y-axis points skyward [38].

a particle's trajectory from the beam-line, as it is not Lorentz invariant. Instead, pseudorapidity, η, defined in Equation 2.1, is used. η is 0 for a particle whose trajectory is perpendicular to the z-axis and infinity (or negative infinity) when along the positive z-axis (or negative z-axis).

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{2.1}$$

Momentum is typically quoted transverse to the beam line, simply called transverse momentum, $\mathrm{p_T}$, as standard momentum is altered by the initial momentum of the incoming protons and their constituents, whilst $\mathrm{p_T}$ is conserved.

2.2.1 Inner Detector

The Inner Detector (ID) is the closest sub-detector to the interaction point, covering the region 3.325 cm to 108.2 cm in radial distance [39, 40]. Its purpose is to provide precision measurement of charged particle trajectories. This aids identification of interaction vertices, which is otherwise difficult in the high track density environment around the interaction point. In particular, the ID facilitates reconstruction of secondary vertices from the decay of long-lived b-hadrons. This allows jets initiated by b-hadrons to be identified or *b-tagged* (see Section 3.4 for details), which is vitally important for analyses with final states containing multiple b-quarks e.g. the $HH \rightarrow b\bar{b}b\bar{b}$ analysis. Particles that traverse the ID leave *hits* in its layers. A particle's trajectory or *track* can be reconstructed by interpolating between these hits. Extrapolating backwards, the common point where a group of tracks originate can be identified as a distinct interaction vertex, which is required when reconstructing particle decay chains and rejecting background processes.

The ID comprises three sub-systems, which are arranged in concentric cylindrical layers around one another. They are, from closest to the interaction point to furthest

away: the Pixel Detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT). They cover the region $|\eta| \leq 2.5$ and all ϕ. An overview of the ID is shown in Figure 2.4.

The ID is surrounded by a 2 T solenoid magnet, which facilitates the measurement of particle momentum. Charged particles bend in the magnetic field due to the Lorentz force, and their momentum, p, can be measured using their track in the ID via Equation 2.2.

$$p = 0.3 \times rB \tag{2.2}$$

Here, p is in units of GeV/c, r [m] is the radius of the curved track and B [T] is the magnetic field strength. The direction of curvature allows the charge of the particle to be determined, as oppositely charged particles curve in opposite directions in the same magnetic field.

Semiconductor Detectors

The Pixel Detector and Semiconductor Tracker both utilise semiconductor detector technology, the physics behind which will be explained briefly now.

There are two sets of electron energy levels in a solid – the *valence band* and the *conduction band*. Electrons in the conduction band are free to move, allowing the material to conduct electricity. In a semiconductor material, the energy of the conduction band is only slightly higher than the valence band, with the distance called the *band gap*. When an electron moves to the conduction band, it leaves behind a vacant site in the valence band, called a *hole*.

Impure semiconductors are said to be *doped*. There are two types of doped semi-

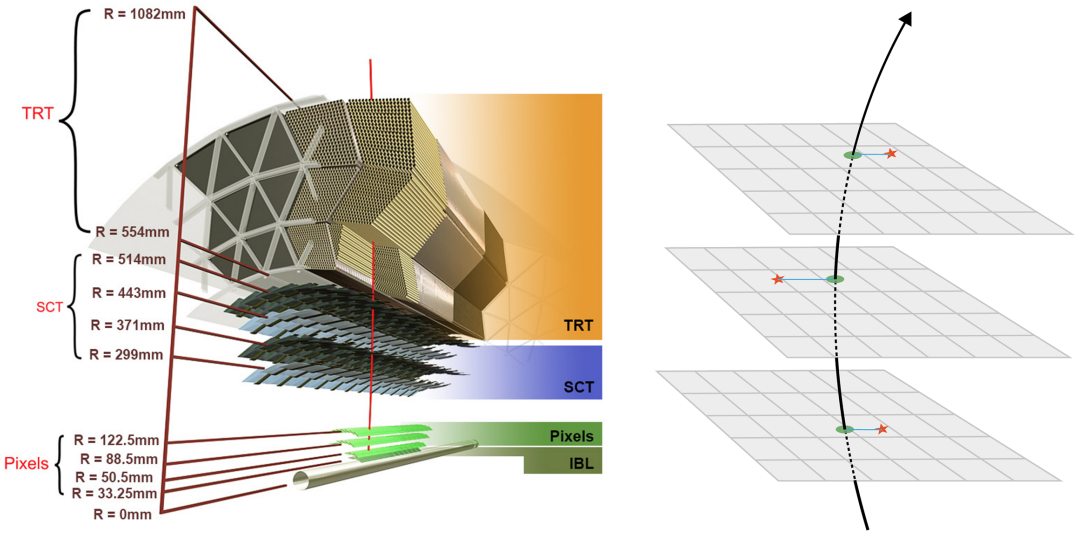


Figure 2.4: Left: schematic of the Inner Detector, with the three sub-detectors - Pixel Detector, Semiconductor Tracker (SCT) and Transition Radiation Tracker (TRT) - from the beam line outward. Here, the Insertable B-Layer (IBL) is shown separately but can be considered part of the Pixel Detector. Right: representation of a track being reconstructed from hits left by a charged particle traversing the layers of the Inner Detector (ID) [41].

conductors – *n*-type and *p*-type. *n*-type semiconductors contain impurities that add electrons, which lie in the *donor level*, close to the conduction band. *p*-type semiconductors contain impurities that add holes, which lie in the *acceptor level*, close to the valence band. A *n-p junction* is formed when an *n*-type semiconductor is interfaced with a *p*-type semiconductor. This creates an electric field across the material, leaving a central region devoid of charge carriers, called the *depletion region*. By placing electrodes across the n-p junction and applying a *reversed bias voltage*, the electric field can be enhanced, increasing the size of the depletion region.

In a semiconductor detector, incident particles traverse the depletion region providing energy to the valence band electrons, transitioning them to the conduction band and creating an electron-hole pair. Due to the electric field, the electrons migrate towards the positive electrode and the holes towards the negative electrode. This produces an electric pulse, which can be detected. By creating a layered array of such detectors, one can reconstruct a particle's trajectory by drawing a line through

the positions of the different detectors it crossed.

Pixel Detector

The Pixel Detector is the first layer of the ID, closest to the interaction point [42]. It has the highest granularity of the ID sub-systems and plays a pivotal role in reconstructing decay vertices. The Pixel Detector is composed of *modules*. Each module is composed of a silicon wafer, the *sensor*, connected to a set of radiation-hard electronics, with minuscule *bump bonds*, i.e. tiny metal connections, joining each pixel to the electronics. Each module is approximately 6 cm by 2 cm and contains 46 080 pixels, each themselves 50 µm by 400 µm. The modules are arranged in four concentric cylindrical layers (*barrels*) around the beam-line, and are bookended by three disks also made of modules (*end-caps*).

After the end of Run 1 (2011-2012) of the LHC, the Insertable B-Layer (IBL) was added as the inner-most layer of the Pixel Detector [43]. The primary motivation was to maintain Run 1 ID performance at the increased instantaneous luminosity levels of Run 2. So, as well as adding an additional hit for track reconstruction, the IBL has slightly smaller pixels at 50 µm by 250 µm and a slightly larger active detector area.

Semiconductor Tracker Detector

The Semiconductor Tracker is the second layer of the ID. It provides precision tracking information like the Pixel Detector, but has slightly lower granularity and a more cost effective design. It comprises sets of semiconductor detector modules made from silicon wafers connected to a set of radiation-hard electronics; however, the module sensors are composed of micro-strips, 80 µm by 12 cm, instead of pixels. Each module has two silicon wafers, aligned at very slightly different angles (a 40 mrad difference), which allow two-dimensional position information to be determined. There are 4088 modules, organised in four concentric cylindrical layers around the ID, bookended by 9 disks on each side.

Transition Radiation Tracker

The Transition Radiation Tracker is located furthest away from the beam line, surrounding the Pixel Detector and SCT. It is a *straw chamber*, composed of an array of tightly packed gas-filled drift tubes (diameter 4 mm) with high tension gold-plated tungsten wires running through each. The drift tubes are filled with a mixture of gases – 70 % Xe, 27 % CO_2 and 3 % O_2. The TRT is composed of a barrel region of 52 544 straws that cover $|\eta| < 1$ and two end caps of 122 880 straws each that cover $1 < |\eta| < 2$. When a particle passes through the drift tube, it ionises the gas atoms. The walls of the straws are held at a high negative potential difference with respect to the central wire, creating an electric field. The freed electrons will then drift towards the wire, discharge and produce an electric signal, which is detected. Particles will cross several drift tubes as they traverse the TRT, allowing for their track to be reconstructed from the signals from the tubes. In addition to the drift tubes, the region between the tubes contains polypropylene fibres and foils. As particles propagate from the polypropylene-filled region into the gas of the drift tube, they emit *transition radiation*, which is absorbed by the Xe atoms, increasing their energy and boosting the output signal. This allows for some particle identification, as the intensity of the transition radiation is greatest for electrons.

Inner Detector Resolution

A comparison of the spatial resolution of the different sub-systems in the ID is shown in Table 2.1. The sub-systems with the highest spatial resolution are placed closest to the interaction point, as this is where track density is highest. The IBL layer of the Pixel Detector has the highest spatial resolution due to its high-granularity pixel-based design. Although the TRT has the lowest $r\phi$ resolution, it measures the largest number of hits, normally $\sim$36 per track, due to its straw-based design. This is far more than the number of layers in the Pixel Detector or SCT.

Table 2.1: Spatial resolution of the Inner Detector sub-systems in terms of the radial distance (r), azimuthal angle (ϕ) and beam line distance (z), as measured from the interaction point. $r\phi$ is used instead of simply ϕ as the angular resolution changes with radial distance [44].

	Resolution (µm)		
Sub-System	$r\phi$	r	z
IBL	8	-	40
Pixels (Barrel)	10	-	115
Pixels (End-cap)	10	115	-
SCT (Barrel)	17	-	580
SCT (End-cap)	17	580	-
TRT	130	130	-

2.2.2 Calorimeters

The ATLAS calorimeters measure the energy of incident particles [45, 46, 47]. They cover $|\eta| < 4.9$ and the full range in ϕ, providing more forward coverage than the ID. They are designed to absorb the total energy of all known particles, except muons and neutrinos, thus stopping them. An overview of the ATLAS calorimetry system is shown in Figure 2.5.

Particles *shower* as they propagate through a calorimeter. There are two types of particle showers, which are normally initiated by two sets of particles. *Electromagnetic showers* are typically initiated by photons and electrons interacting primarily via the electromagnetic force. Hadronic showers are typically initiated by hadrons and are dominated by interactions via the strong force.

In an electromagnetic shower, two physical processes dominate - *bremsstrahlung* and *pair production*. Bremsstrahlung describes the process by which an electron

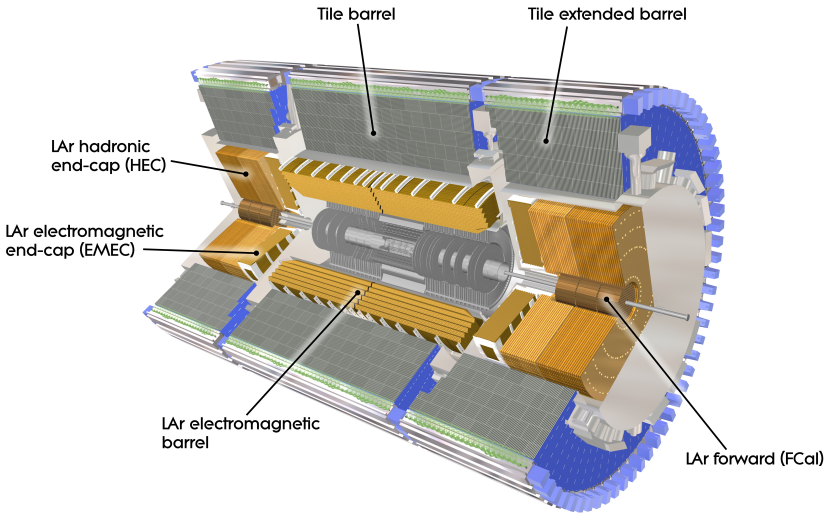


Figure 2.5: Schematic of the ATLAS calorimetry system, which surrounds the interaction point and Inner Detector (ID) [48]. The electromagnetic calorimeter (ECAL) comprises the liquid argon (LAr) barrel and end-caps (EMEC). The hadronic calorimeter (HCAL) comprises the scintillating tile barrel and extended barrel, alongside the end-caps (HEC) and forward detectors (FCal).

entering the electric field of an atomic nucleus will deflect, decelerate and radiate lost kinematic energy in the form of a photon or *bremsstrahlung radiation*. In pair production, a photon with sufficiently high energy will spontaneously transform into an electron-positron pair. In an electromagnetic shower, these two processes happen repeatedly, with an initial particle creating a cascade of lower-energy electrons and photons until all of its initial energy is deposited in the calorimeter.

Hadronic showers undergo a more complex spectrum of physical processes than electromagnetic showers, taking longer to develop and producing a broader shower profile. As the hadron propagates through the calorimeter, it undergoes inelastic scattering and nuclear interactions with the calorimeter nuclei. This creates a series of new hadrons, of which pions, i.e., π^-, π^0 and π^+, are the most common. This process can excite the nuclei, which emit further particles as they de-excite. These new particles interact with the other calorimeter nuclei creating further particles, and the process repeats until all the initial hadron's energy is deposited in the

calorimeter. Additionally, the neutral pion, π^0, decays readily into a photon pair, which creates an additional electromagnetic shower component.

There are two ATLAS calorimeters to target these two scenarios – the *electromagnetic calorimeter* (ECAL) and *hadronic calorimeter* (HCAL). Both the ECAL and HCAL are *sampling calorimeters*. Sampling calorimeters are composed of alternative layers of detection and absorption materials. The detection material measures the energy of the incident particles, whilst the absorption material initiates the particle showering. This is a cost-effective alternative to *homogeneous calorimeters*, which are made from a single material that plays both roles simultaneously.

The ECAL has a liquid argon (LAr) detection material and lead absorption material. LAr was selected because of its radiation tolerance, stability of response over time and the linear relationship between its energy response and temperature. The ECAL is composed of a barrel (covering $|\eta| < 1.475$) and two end-caps (covering $1.375 < |\eta| < 3.2$). Additionally, the sections covering the high-precision ID region ($|\eta| < 2.5$) is split into three radial regions for added resolution. The HCAL has two sections that utilise different calorimeter technologies – the central *tile calorimeter* and the LAr end-caps and forward calorimeters. In the tile calorimeter, the absorption material is steel and the detection material is polystyrene scintillating tile. Scintillation caused by passing particles is detected using a series of photomultiplier tubes. This technology was selected as a cost-effective means of providing the maximum radial depth for capturing the relatively long hadronic showers. The tile calorimeter is composed of a central barrel, covering $|\eta| < 1.0$, and a two extended barrels, covering $0.8 < |\eta| < 1.7$. As it covers the central high-precision region, it is also segmented in three layers radially. The Hadronic End-cap Calorimeter (HEC) relies on LAr technology, like the EM calorimeter, and covers $1.5 < |\eta| < 3.2$. It is segmented radially in four layers. The Forward Calorimeter (FCal), which covers the region $3.1 < |\eta| < 4.9$, relies on LAr technology too. The FCal comprises

three detector modules, with the first made of copper (providing good heat removal) and designed for electromagnetic interactions, and the latter two made of tungsten (providing good shower containment) and designed for hadronic showers.

The length of electromagnetic and hadronic showers are usually measured in terms of *radiation lengths* (X_0) and *nuclear interaction lengths* (λ), respectively.[2] The depth of the ECAL in terms of radiation lengths varies with η but is never less than $\sim 22X_0$. The depth of the HCAL varies similarly but is never less than $\sim 10\lambda$. This depth is sufficient to provide good containment of the full shower length and significantly limits the amount of *punch-through* (high-energy particles that are not stopped in the calorimeter).

Calorimeter Granularity and Resolution

Granularity varies across different sections of the ECAL. Granularity is highest in the initial layers of the central barrel, at 0.025×0.025 in $\Delta\eta \times \Delta\phi$, and is lowest in the furthest forward regions of the end-cap, at 0.1×0.1. The granularity of the HCAL also varies, but is generally less fine. The central scintillator tile calorimeter has the best granularity, at 0.1×0.1 for the first layer of barrel and extended barrel. The LAr HEC also has a granularity of 0.1×0.1 at its most central region, but the majority of the LAr forward calorimeter has less fine granularity.

The targeted energy resolution of the calorimeters is shown as a function of particle energy in Table 2.2. The energy resolution in the ECAL is very precise, due to the fine granularity. This facilitates precision measurements of electrons and photons, complementing the ID. This was important for measuring the Higgs boson decay to the photons, one of the golden channels for the Higgs discovery effort. The resolution of the HCAL is also precise – albeit slightly less so than the ECAL – facilitating the

[2]The *radiation length* is defined as the mean distance a charged particle travels in a material before its energy is reduced by a factor of e through the electromagnetic interaction with the material. The *nuclear interaction length* is the mean distance a particle travels in a material before undergoing an inelastic collision with one of the atoms in the material.

Table 2.2: Energy resolution performance goals for the ATLAS calorimetry system. Units for E are GeV.

Calorimeter	Region	Energy Resolution	η Coverage
Electromagnetic	–	$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$	± 3.2
Hadronic	Barrel/End-caps	$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$	± 3.2
Hadronic	Forward	$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$

accurate reconstruction of hadron-initiated jets.[3]

2.2.3 Muon Spectrometer

The outermost layer of the ATLAS detector is the Muon Spectrometer, due to the high penetrating powers of high-energy muons [49, 50, 51]. The MS is designed to measure the momentum of charged particles that exit the calorimeter system. Other particles besides muons can exit the calorimeters if they have sufficient energy. As mentioned, this is referred to as punch-through, and is a background process in muon identification. The MS covers $|\eta| < 2.7$ and all ϕ. An overview of the MS is shown in Figure 2.6.

The MS comprises a central barrel region covering $|\eta| < 1.2$ and two end-caps covering $1.0 < |\eta| < 2.7$. Both the barrel and end-caps are surrounded by superconducting toroid magnets, designed to bend the trajectory of charge particles to facilitate measurement of their momentum and identification of their charge, as in the ID [53, 54]. There are three concentric cylindrical layers of muon detection chambers in the barrel region at approximately 5 m, 7.5 m and 10 m. In the end-cap region, there are four large wheels composed of chambers, aligned perpendicularly to the z-axis at $|z| = 7.4$ m, 10.8 m, 14 m and 21.5 m.

There are four types of detector in the MS – Monitored Drift Tubes (MDT), Cathode

[3] See Section 3.3 for details on *jets*.

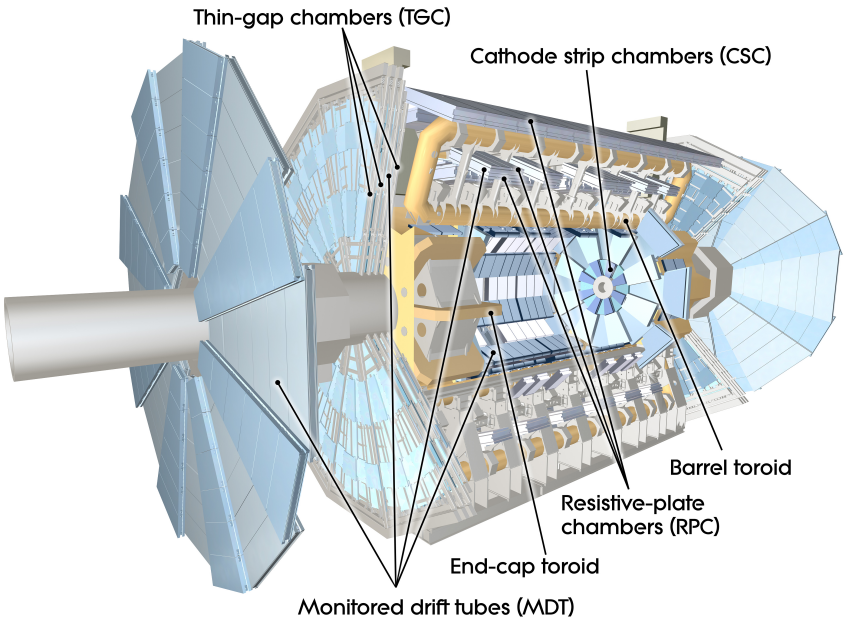


Figure 2.6: Schematic of the Muon Spectrometer (MS), which is the outermost layer of the ATLAS detector, including the two large *wheels* at either end of the detector [52]. There are four types of muon-detection chamber labelled – the Thin Gap Chambers, the Cathode Strip Chambers, the Monitored Drift Tubes and the Resistive Plate Chambers. Additionally, the three magnet systems, shown in yellow, are labelled – the barrel toroid magnets and the two end-cap toroid magnets.

Strip Chambers (CSC), Thin Gap Chambers (TGP) and Resistive Plate Chambers (RPC). The MDTs provide precise muon momentum measurement in the region $|\eta| < 2.7$. They possess a straw-based design similar to the TRT and possess an average resolution of ∼35 µm per chamber. The CSCs are used in place of the MDTs in the first layer of the MS in the region $2 < |\eta| < 2.7$, as particle flux and muon-track density is high in this region and the CSCs possess faster response time and lower sensitivity to neutrons. The Thin Gap Chambers and Resistive Plate Chambers are used to provide fast muon track information for triggering (see Section 2.2.4).

2.2.4 The Trigger

The instantaneous luminosity of the LHC peaked in 2018 at $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, which is too high to record every event in permanent storage [55]. Instead, the ATLAS Trigger Data Acquisition (TDAQ) system makes the decision whether to record an event or not, lowering the flow of event data to a manageable level [56, 57, 58]. The

TDAQ is a two-stage trigger system starting with the Level-1 (L1) trigger followed by the High Level Trigger (HLT). The L1 trigger uses basic, coarse granularity detector information to inform the trigger decision. After L1, the event rate drops from ∼40 MHz to ∼100 kHz. Events that pass the L1 trigger proceed to the HLT, where more precise detector information and advanced calculations are used to inform the trigger decision. After the HLT, the event rate drops to ∼1 kHz, which is sufficiently low that events can be recorded to permanent storage. ATLAS employs a suite of different triggers in their *Trigger Menu*, which specify different requirements on, for example, the number of a certain type of particle or the p_T of a particle. One example trigger is `HLT_j420`, which requires an event contains at least one jet with p_T of 420 GeV as measured by the HLT, dropping all events that do not meet this requirement. A list of the triggers used in the $HH \to b\bar{b}b\bar{b}$ analysis is included in Section 4.3.

The L1 trigger is hardware-based, using custom electronics. It comprises the L1 Calo and L1 Muon sub-systems, which provide input to L1 Topo. The final decision is implemented by the Central Trigger Processor (CTP). For a given event, the hits from the Muon Spectrometer are sent to the L1 Muon sub-system, which performs a basic calculation of the p_T of any potential muons. Events are typically kept if the p_T of the muon is sufficiently high or if there is a sufficient number of them. Simultaneously, the L1 Calo sub-system is provided with signals from the calorimetry system. Basic digitisation and calibration of the signal is performed, relevant particles reconstructed and measurements of the total energy and missing energy are made. Information from the L1 Muon and Calo are provided to L1 Topo, which uses it to perform some basic calculations, e.g. angular distance between particles or the combined invariant mass. Events will be kept if they pass the threshold on such topological variables. The locations of interesting detector signatures, called *Regions of Interest* (ROI), are passed onto the HLT, which may only utilise the information from the ROIs to reduce processing time. The HLT is software-based. An initial

decision is made by the quick but more rudimentary *fast trigger*. Then, particles are reconstructed, any pertinent kinematic variables calculated and cut thresholds applied. The algorithms used are very similar to those applied in offline analyses.

3 Physics Objects and Algorithms

In the first step of an ATLAS analysis, the final state of interest must be accurately reconstructed from a mixture of signals in the various ATLAS sub-detectors. This is a complex process, relying on a suite of algorithms, calibrations and correction factors. The methods utilised in the $HH \rightarrow b\bar{b}b\bar{b}$ analysis will be described in detail in this chapter.

3.1 Tracks

A *track* is a representation of the trajectory of a charged particle that traversed the ATLAS detector, which is built using a combination of measurements in the different sub-systems of the Inner Detector. Tracks are vital to ATLAS analyses, allowing the common origin of particles to be identified (*vertices*) and aiding in the identification of quark flavour (a process called *flavour-tagging*, which is discussed in Section 3.4). Refs. [59, 60] provides a detailed description of primary track reconstruction in ATLAS, which will be summarised here.

First, clusters are created by grouping neighbouring pixels or strips in the Pixel and SCT detectors with charge measured to be a pre-defined level over noise. These clusters are positioned at a *space-point*, a three-dimensional position in ATLAS that

represents the point an incident particle crossed that part of the detector. After the collection of clusters are built, the track reconstruction algorithm is applied. Track *seeds* are built from groups of three space-points. Requirements are placed on the transverse momentum and impact parameter of the track seeds in order to reject seeds that do not originate from a real particle crossing.[1] A search for an additional space-point that fits with the track seed is also performed. During Run 2, there could be as many as 20 000 track seeds produced in a single proton bunch collision. A combinatorial Kalman filter [61] is then used to build track *candidates* from the track seeds by incorporating other space-points that fit with the pattern of a particle crossing. In the next stage, the *ambiguity solver* is applied to resolve overlaps between track candidates, e.g. when two track candidates contain the same cluster, and reject *fake tracks*, i.e. those not originating from a true particle crossing. Here, a *track score* is calculated for each track candidate, which is based on a number of track quality criteria, such as the number of space-points that make up the track and the fit of the space-points to the track candidate trajectory using the χ^2 goodness-of-fit metric. Only track candidates with sufficiently high track scores are kept. The ambiguity solver requires clusters to only be associated with a maximum of two tracks (so-called *shared* clusters) with the decision for which tracks keep the cluster being based on track score. Also, tracks are required to be associated with a maximum of two shared clusters. A Neural Network (NN) is used to identify single clusters originating from two particles that happened to cross the same point in the detector (so-called *merged* clusters). A penalty is applied to a track candidate's score if it contains a shared cluster that is not identified as a merged cluster by the NN. A high-resolution global-χ^2 track fit is performed using all available information, including additional information from the TRT, producing the final collection of reconstructed tracks.

[1] Generally, the impact parameter is the distance of closest approach of a track to a vertex. Here, it is with respect to the centre of the interaction region and tracks are assumed to have a perfect helical trajectory.

3.2 Vertices

Primary vertices are the precise location in the ATLAS detector where a proton-proton interaction occurred. Multiple proton-proton interactions occur in a single proton bunch collision, but the vast majority are soft-QCD interactions, which are not of interest. The products of such interactions are referred to as *pile-up* and correct assignment of particles to vertices is a vital step in rejecting pile-up.

Vertices are located using charge particle tracks in the ID. Ref. [62] provides a detailed description of primary vertex reconstruction, which will be summarised briefly here. Vertex seeds are built using the x and y coordinates of the beam spot and the mode of the z coordinates of the collection of tracks at their point of closest approach to the centre of the beam spot. The collection of reconstruction tracks are fit to vertex seeds using an iterative vertex fitting algorithm based on the χ^2 goodness-of-fit metric. Only tracks with a pre-defined level of compatibility with the vertex seed are assigned to that seed. Vertex seeds with two or more associated tracks with a combined $\mathrm{p_T}$ above 500 MeV are added to the collection of reconstructed vertices. The vertex with the largest sum of squared track momenta ($\sum p_{\mathrm{T}}^2$) is taken as the hard-scatter primary vertex.

3.3 Jets

In proton-proton collisions, quarks and gluons (collectively called *partons*) are created, fragment and hadronise. Fragmentation is the process of a single parton gradually generating several different partons via the QCD interaction. The spray of quarks and gluons formed is referred to as a *parton shower*. In hadronisation, the quarks in the parton shower bind together to form hadrons due to *colour confinement*, the fundamental prohibition of free coloured states. So, what starts as a single initial parton, ends up as a varied topology of high complexity.

In ATLAS, the initial parton's four-momentum is reconstructed using a *jet*. The spray of hadronic activity from a parton acts as inputs to the jet, which is defined by an algorithm and set of parameters. The algorithm decides how to group the inputs to reconstruct the parton, and the parameters control the algorithm. The inputs come from either real data or Monte Carlo (MC) simulation and input type can vary. For example, inputs can be calibrated calorimeter clusters or particle information in MC simulation. A key jet parameter is jet radius, R, which controls the maximum distance a particle can be from the centre of the jet. In ATLAS, the two standard jet radii are 0.4 (small-R), typically used for a single parton, and 1.0 (large-R), typically used for high-momentum collimated partons from the same parent particle, e.g. a Lorentz-boosted $H \to b\bar{b}$ system.

In the $HH \to b\bar{b}b\bar{b}$ analysis, all hadronic activity is reconstructed using small-R jets, defined using the anti-k_t algorithm with Particle Flow (PFlow) objects as input. The anti-k_t algorithm and PFlow inputs will now be explained.

3.3.1 The anti-k_t Algorithm

The anti-k_t algorithm is applied to an entire event, building a collection of jets. The algorithm is a *sequential recombination* algorithm [63]. These algorithms view fragmentation as a series of $1 \to 2$ QCD interactions, e.g. $g \to q\overline{q}$, and reconstruct the initial parton by iteratively combining pairs of inputs starting from the final state. The algorithm relies on two key quantities, d_{ij} and d_{iB}, defined in Equations 3.1 and 3.3 respectively. The choice of how to pair inputs depends on the distance metric, d_{ij}.

$$d_{ij} = \min\left(\frac{1}{\mathrm{p}_{\mathrm{T},i}^2}, \frac{1}{\mathrm{p}_{\mathrm{T},j}^2}\right)\frac{\Delta R_{ij}^2}{R} \tag{3.1}$$

Here, ΔR_{ij} is the angular distance between two inputs, i and j, and is defined in

Equation 3.2. R is the dimensionless jet radius parameter. It controls the angular reach of the jet algorithm, being the maximum angular separation two inputs can have and still be part of the same jet.

$$\Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \tag{3.2}$$

The choice of whether to stop pairing inputs and declare that a jet has been identified depends on d_{iB}, the inverse of the squared $\mathrm{p_T}$ for i.

$$d_{i\mathrm{B}} = \frac{1}{\mathrm{p}_{\mathrm{T},i}^2} \tag{3.3}$$

In the first iteration of the algorithm, the values of d_{ij} and d_{iB} are calculated for all inputs. These values are collated and the lowest selected. If a d_{ij} is the lowest value, the i and j inputs are combined to form k. i and j are then replaced in the list of inputs with k. If a d_{iB} is the lowest value, i is identified as a jet and removed from the list of inputs. The next iteration then proceeds with the updated list of inputs, and further iterations proceed until all jets are identified.

d_{ij} is smaller for inputs that are closer in proximity, so the anti-k_t algorithm has a tendency to group inputs that are near to one another. d_{ij} is also small when the $\mathrm{p_T}$ of i or j is large, but if the inputs are not close together then d_{iB} is typically smaller than d_{ij}, and one of the inputs will be labelled as a jet instead of being combined.

The inverse $\mathrm{p_T}$ factor in the definitions of d_{ij} and d_{iB} ensure that the highest $\mathrm{p_T}$ inputs are combined first. As a result, jets will be constructed around a high $\mathrm{p_T}$ core, with softer inputs having a small impact on overall jet shape.

3.3.2 Particle Flow

Particle Flow (PFlow) jets are the standard for most ATLAS analyses [64]. PFlow jets are built from objects formed from a combination of TopoClusters from the calorimeters and tracks from the Inner Detector. TopoClusters, a shortening of Topological Clusters, are built from groups (or *clusters*) of adjacent cells in the calorimeter with energy over a defined level above noise [65]. By combining TopoCluster with high resolution tracking information, PFlow jets show improved momentum resolution, improved angular resolution and reduced contamination from pile-up when compared with the previous standard, EMTopo jets, which are built from TopoClusters alone.

The primary challenge when building PFlow jets is preventing the double-counting of a particle's energy, as the jet may contain both a particle's track and TopoCluster. To avoid this, the algorithm tries to match tracks to TopoClusters. If it succeeds, it calculates the expected energy deposited in the calorimeter by the track, and removes the corresponding number of cells from the TopoClusters. Then, the tracks, matched or unmatched, and the TopoClusters, altered or unaltered, form the group of PFlow objects, which are provided as input to the jet algorithm.

3.3.3 JES Calibration

The measured properties of a jet can be biased by a number of systematic effects. To name a few: the measured energy can be inflated by pile-up; energy can be lost in the detection process, e.g. deposited in passive material in the detector; and energy can be lost in jet reconstruction, e.g. parton decay products might not be added into the jet (*out-of-cone effects*). To counteract this, the Jet Energy Scale (JES) calibration is applied to the jet four-momenta [66].

The JES calibration comprises a number of factors, each targeting a different sys-

tematic effect. First, the *pile-up correction* removes any contribution to the jet energy that originates from pile-up. Second, the *absolute JES* corrects the jet four-momentum to match equivalent jets reconstructed at particle-level, to remove biases from reconstruction-related effects.[2] Then, the *Global Sequential Calibration (GSC)* corrects for variability in simulated detector response to equivalent-energy jets of differing characteristics, improving jet p_T resolution and corresponding uncertainty. Finally, the *in situ correction* corrects the four-momentum of jets reconstructed in data such that they match equivalent jets reconstructed in MC simulation, to remove biases from detector-related effects.

The systematic uncertainties associated with each step in the measurement of the JES calibration are discussed in Section 10.2.1.

3.3.4 JER Calibration

The Jet Energy Resolution (JER) quantifies the variation in the measured energy of jets due to detection and reconstruction effects [66]. Several sources are accounted for when measuring the JER. For example, electronic noise in the detection equipment, stochasticity in the calorimeter energy deposits, and variation in energy losses in passive material.

The JER is measured using $2 \rightarrow 2$ dijet events, and is then combined with a factor independently derived to account for electronic noise. Measuring the JER relies on the *dijet balance* method. In a theoretical dijet event, the momentum of the two final state jets should be equal. So, any difference between the two, called *asymmetry*, can be ascribed to resolution effects. Using a large dijet dataset, the distribution for the asymmetry is generated, a Gaussian fit performed, and the JER extracted as the standard deviation.

[2] *Particle-level* means the jets are reconstructed using the particle information recorded in Monte Carlo simulation.

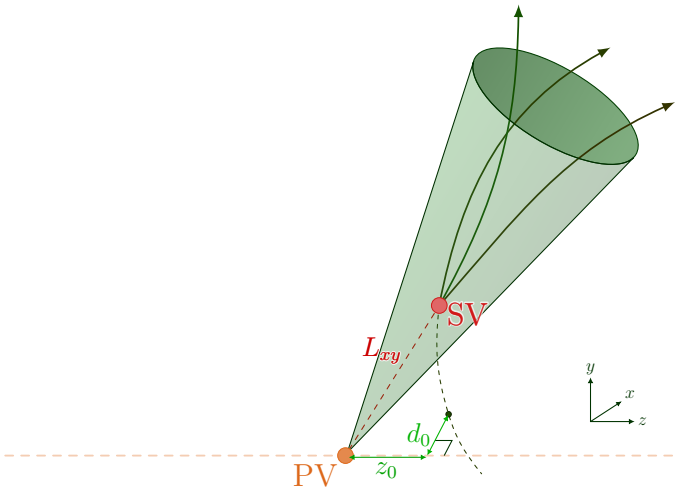


Figure 3.1: Diagram of the decay of a b-hadron in a jet, with a number of key characteristics of the b-hadron that are leveraged in b-tagging labelled, such as the Primary Vertex (PV), the flight length of the b-hadron (L_{xy}), the Secondary Vertex (SV) and the impact parameters, z_0 and d_0. (Credit to Izaak Neutelings and Nazar Bartosik.)

If the jet energy resolution in data is greater than MC simulation, the latter is *smeared* until it matches. Smearing is performed by weighting the energy of jets, drawing the weights from a Gaussian with standard deviation derived from the nominal JER and the JER uncertainty (discussed in Section 10.2.1).[3]

3.4 Flavour Tagging

Flavour tagging is the process of identifying the type of parton that initiated a jet. One category of flavour tagging is b-tagging, the process of identifying whether a jet was initiated by a b-quark. The $HH \rightarrow b\bar{b}b\bar{b}$ analysis targets a final state containing four b-jets, and, as such, b-tagging is a vital step in the analysis.

3.4.1 Flavour Tagging Algorithms

A suite of algorithms that identify the flavour of a jet are available to ATLAS analyses. The majority leverage the distinctive features of the decay of a b-hadron, using variables that show discrimination for *light* (uds), c and b-jets. Figure 3.1 shows a diagram highlighting these features. b-hadrons are emitted from the collision

[3] $\sigma^2_{\text{smear}} = (\sigma_{\text{nom}} + |\sigma_{\text{syst}}|)^2 - \sigma^2_{\text{nom}}$

point at velocities close to the speed of light, receiving a Lorentz boost. The lifetime, τ, of a typical b-hadron is $\sim$1.5 ps. The b-hadron lifetime is slightly longer than expected given their relatively large mass ($\sim$5 GeV). This is because the CKM matrix elements that govern the weak decay of a bottom quark are relatively small: $|V_{ub}| = 0.00369$ and $|V_{cb}| = 0.0418$. Therefore, the decay of b-hadron is slightly suppressed, extending its lifetime.[4] Assuming that a b-hadron is travelling at the speed of light, c, after it is produced, it will travel at least $\tau \times c = 0.45$ mm before it decays, but can travel much further due to the Lorentz boost. In the ATLAS detector, this distance, called the flight length, L_{xy}, is large enough that the point of decay of the b-hadron can be distinguished from the primary vertex (PV) using tracks in the Inner Detector. Here, a track is constructed from the hits left by a particle as it crosses the layers of the Inner Detector, and, thus, represents the particle's trajectory. The *primary vertex* (PV) is the point at which the initial proton-proton interaction occurred. The point of decay of the b-hadron is called the *secondary vertex* (SV). The large flight length of the b-hadron is a key characteristic exploited when b-tagging jets.

An important quantity used in b-tagging is the distance of closest approach of a track to the PV. This point is called the *perigee*, a term adopted from astronomy. The variable z_0 is the projection of this distance along the beam-line (the z-axis), and the variable d_0 is the projection in the x-y plane. The variables used in b-tagging are the *impact parameters* (IP) [67]. The longitudinal IP is $\text{IP}_{r\phi} = |d_0|$ and the transverse IP is $\text{IP}_z = |z_0 \sin(\theta)|$. Roughly speaking, the larger the IP, the more unlikely a jet is a *light*-jet, given a *light*-jet is unlikely to have decay products far from the PV.

The flavour tagging algorithm applied in the $HH \to b\bar{b}b\bar{b}$ analysis is `DL1r`, the current

[4] See Section 1.1.1 for an explanation of the CKM matrix and the values of its elements. Also, note that b-quarks will preferentially decay to c-quarks as $|V_{cb}|$ is approximately an order of magnitude larger than $|V_{ub}|$ and a $b \to t$ transition will not occur as $m_t > m_b$.

standard on ATLAS [68]. `DL1r` is a *high-level* tagger, defined as such because its inputs are the outputs of specialised *low-level* taggers. The details of these low-level taggers and `DL1r` will now be discussed.

`IP2D` and `IP3D` are two low-level taggers that exploit the differences between the impact parameter information for *light*, *c* and *b*-jets [68]. Tracks are associated to a jet using the angular separation, ΔR, between the track and the jet axis.[5] The impact parameters of all associated tracks are then used to quantify the probability of a jet being a particular flavour. `IP2D` uses $\mathrm{IP}_{r\phi}$, whilst `IP3D` uses both IP_z and $\mathrm{IP}_{r\phi}$, leveraging the correlations between the two. The final discriminants are three log-likelihood ratios comparing the probabilities of a jet being different flavours, which are provided as input to `DL1r`.

`SV1` is a low-level tagger that exploits the distinct secondary vertex in *b*-jets [69]. `SV1` looks to reconstruct a single SV in a jet by performing a χ^2-fit of tracks associated with the jet to all two-track vertices. If a SV is found, a set of eight discriminating variables associated with the SV, e.g. number of associated tracks, invariant mass, and energy fraction, are provided as input to `DL1r`.

JetFitter is a low-level tagger that looks to exploit the distinct topological vertex structure in *b*-jets [70]. The concept behind JetFitter is that if a *b*-hadron was created at a PV, decayed to a *c*-hadron at a SV, and then the *c*-hadron decayed itself, one would expected to see an approximately flush line of vertices and associated tracks, given the initial momentum of the *b*-hadron. JetFitter uses advanced statistical techniques to reconstruct this decay topology in the jet. As output, it provides eight discriminating variables that describe this chain of tracks and vertices to `DL1r`.

[5] The jet axis points in the direction of the vector sum of the four-momenta of the clusters corresponding to the jet.

The final low-level b-tagging algorithm that provides input to DL1r is RNNIP [68]. RNNIP also uses tracks and associated IPs, but leverages the correlations between the tracks' spatial and kinematic variables by using a Recurrent Neural Network, which is designed to work with sequences of inputs. This is an advantage over traditional IP-based algorithms, like IP2D/3D, which consider tracks as independent objects. Additionally, RNNIP can use a greater number of input variables, given the ease of adding more variables to Neural Networks. Like IP2D/3D, the final discriminants are log-likelihood ratios of the probabilities for different flavours, which are provided as input to DL1r.

DL1r is a high-level b-tagging algorithm, which utilises a Deep Neural Network, whose inputs, as discussed, are the outputs of IP2D/3D, SV1, JetFitter, and RNNIP, as well as jet $\mathrm{p_T}$ and η. DL1r stands for Deep Learning 1 with RNNIP. RNNIP is mentioned specifically because DL1r is based on the precursor DL1 algorithm, which did not use RNNIP. Details of the Neural Network architecture and hyperparameters can be found in Table 3 of Ref. [71]. DL1r is trained and validated with a mixture of $t\bar{t}$ and Z′ Monte Carlo simulation events, $t\bar{t}$ being used at low-$\mathrm{p_T}$ and Z′ at high. In this sample, the branching fractions are set to one third for $b\bar{b}$, $c\bar{c}$ and *light-light* quark pairs, so there is an equal distribution of flavours. The output of DL1r is a probability score that a given jet is initiated by a light (p_u), charm (p_c), or bottom (p_b) quark. These probabilities are combined to give the final DL1r discriminant, shown in Equation 3.4, which is a log-likelihood ratio of the hypotheses that a jet is a b-jet compared to a *light* or c-jet.

$$D_{\mathrm{DL1r}} = \log\left(\frac{p_b}{f_c \times p_c + (1 - f_c) \times p_u}\right) \tag{3.4}$$

Here, f_c is the effective c-jet fraction in the training sample. The value of f_c was

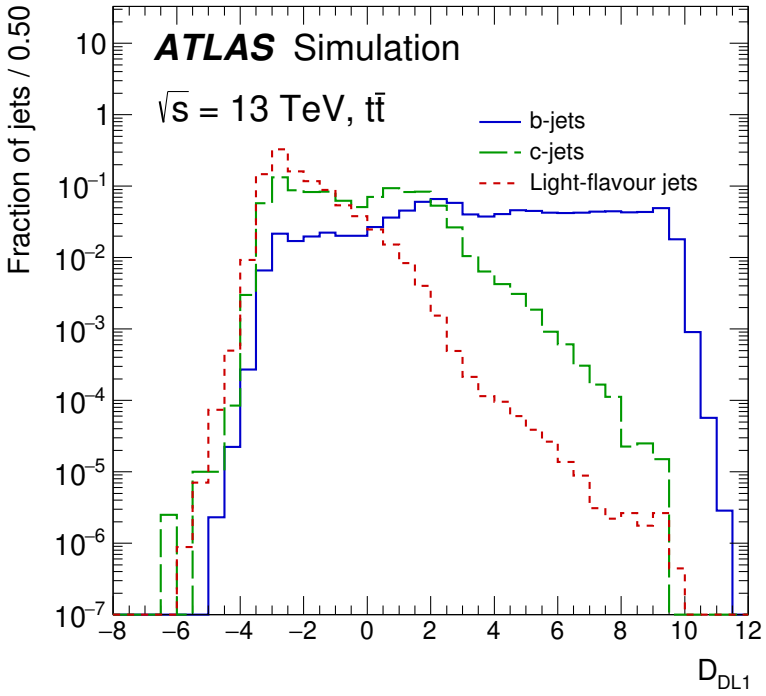


Figure 3.2: Output discriminant distribution of the `DL1` b-tagging algorithm for b (solid blue), c (dashed green) and *light*-jets (finely-dashed red) in a reference $t\bar{t}$ Monte Carlo simulation sample.

selected as 0.018 after training in order to optimise tagging performance of the algorithm.

3.4.2 Working Points

The `DL1r` discriminant is a measure of the probability a jet is a b-jet. Figure 3.2 shows an example of the distribution of the discriminant for the `DL1` tagger [71].[6] In order to decide whether a jet is a b-jet, a cut must be applied to the value of the discriminant. If the `DL1r` discriminant value is above the cut value then the jet is b-tagged. Where to place this cut is a question of balance between efficiency and purity. A tight cut provides a pure but small b-jet sample. On the other hand, a loose cut provides a large sample, but with contamination from *light* and c-jets. In ATLAS, the set of standard cut values are referred to as *working points* (WP). There are four calibrated WPs, each defined such that the rate at which b-jets are accepted (the *efficiency*, ϵ_b) is a fixed value in a reference $t\bar{t}$ sample. The working points are

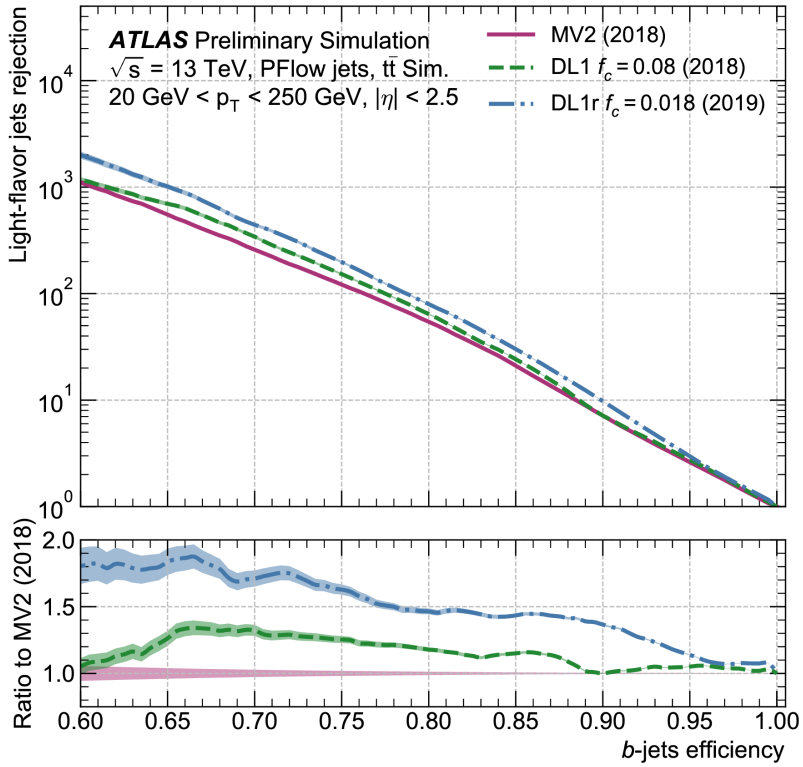


Figure 3.3: Top: the b-jet efficiency (ϵ_b) versus *light*-jet rejection ($1/\epsilon_l$) for three b-tagging algorithms – `MV2` (pink), `DL1` (green) and `DL1r` (blue) – using $R = 0.4$ Particle Flow jets in a reference $t\bar{t}$ Monte Carlo simulation sample. Jets are selected in the range 20 GeV $< \mathrm{p_T} <$ 250 GeV. Bottom: ratio of the b-tagging efficiencies of the three b-jet algorithms to the `MV2` algorithm.

60%, 70%, 77% and 85%. The performance of `DL1r` compared to two other b-tagging algorithms, `MV2` and `DL1`, in terms of b-jet efficiency (ϵ_b) and *light*-jet rejection ($1/\epsilon_l$) is shown in Figure 3.3 [72]. In the $HH \rightarrow b\bar{b}b\bar{b}$ analysis, the 77% WP was used. This was part of an agreement with other HH and single H analyses aiming to be involved in a future combination, as using a single WP facilitates combination of results.

3.4.3 Flavour Tagging Scale Factors

In general, the b-tagging efficiency, ϵ_b, in Monte Carlo (MC) simulation does not agree with that in data. This is due to imperfect modelling in detector response and the physics processes in MC simulation. To correct this, scale factors, SF($\mathrm{p_T}$), are applied to the events in MC simulation to ensure agreement with data. For b-tagged

[6]Note, this is `DL1` not `DL1r`, as the equivalent plot for `DL1r` was not available from the Flavour Tagging CP group. This provides an idea of what the output distribution for `DL1r` would look like. The two methods are the similar, but `DL1` does not use the output from `RNNIP`.

jets, they are defined simply as the ratio of the b-tagging efficiency in data and MC simulation, as shown in Equation 3.5.

$$\mathrm{SF}(p_{\mathrm{T}}) = \frac{\epsilon_b^{\mathrm{data}}}{\epsilon_b^{\mathrm{MC}}} \tag{3.5}$$

Scale factors are measured for each WP and in a range of $\mathrm{p_T}$ bins. The efficiency in both data and simulation is measured in $t\bar{t}$ events, as there is a high probability they contain b-jets due to the high likelihood of top decays via $t \to Wb$. The b-tagging efficiency in MC simulation is extracted by identifying true b-jets using truth particle information to check whether a b-hadron is in close proximity in ΔR to a given jet.[7] For data, a $t\bar{t}$ enriched data sample is produced by applying a rigorous event selection, and the b-tagging efficiency is measured using a likelihood-based method, as described in Ref. [71]. Once the scale factors are measured, a smoothing and extrapolation procedure is applied. The final scale factors are a smooth function of jet $\mathrm{p_T}$ and provide coverage even at high $\mathrm{p_T}$ values. The b-tagging scale factors for anti-k_t $R = 0.4$ PFlow Jets tagged by `DL1r` at the 70% WP are shown in Figure 3.4 [73].

3.5 Muons

A b-hadron has a ∼10% probability of decaying via $b \to W(\to \mu\nu_\mu)c$ [74]. As a result, a significant number of b-jets contain muons. In the $HH \to b\bar{b}b\bar{b}$ analysis, muons are reconstructed using the *combined* muon algorithm. Combined muons are built using a global refit of tracks from both the Inner Detector (ID) and Muon Spectrometer (MS) [75]. There are three standard muon reconstruction working points (WP) – `loose`, `medium` and `tight`. The tighter the WP the higher the muon reconstruction purity but the lower the efficiency. The lower the WP the higher

[7] $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$

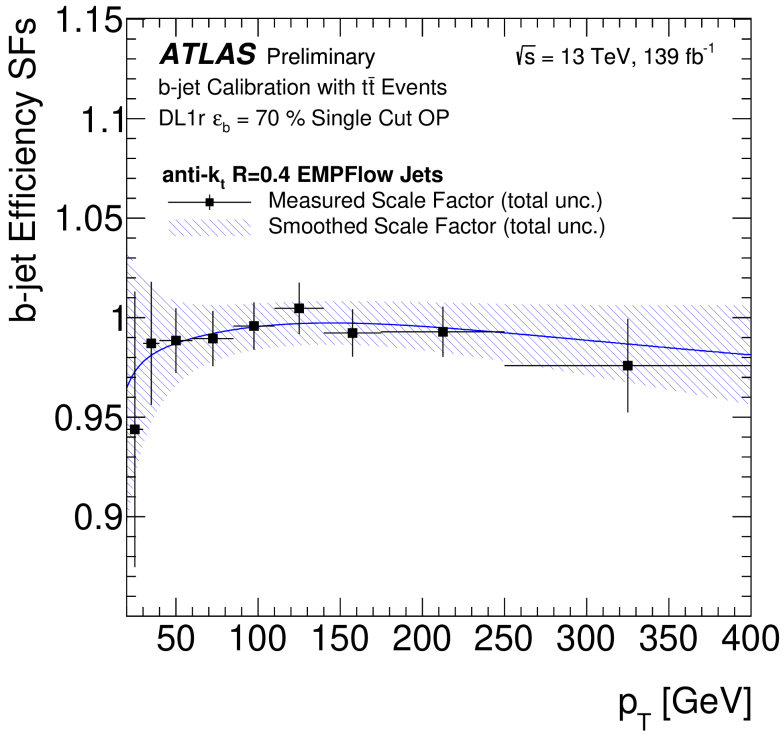


Figure 3.4: Measured (black) and smoothed (blue) b-tagging efficiency scale factors versus p_T for the b-tagging algorithm `DL1r` at the 70% working point (WP) with $R = 0.4$ Particle Flow jets in a reference $t\bar{t}$ Monte Carlo simulation sample.

the efficiency but the greater the degree of contamination from *fakes* (particles that are not muons but are reconstructed as such). The `medium` WP was used in the $HH \rightarrow b\bar{b}b\bar{b}$ analysis as the middle ground and given high precision muons were not required. After all the muons are reconstructed in the analysis, any found in close proximity to a b-jet are added into the jet, as will be described in Section 3.6.

3.6 *b*-Jet Corrections

In a semi-leptonic decay of a b-hadron, the neutrino produced would pass through ATLAS undetected, and the muon would only deposit a fraction of its energy in the calorimeter. Additionally, the parton shower of the fragmenting b-hadron can also extend beyond the jet boundary, reducing the measured energy. This is called an *out-of-cone effect*. In scenarios such as these, there would be a discrepancy between the measured energy of the jet and the energy of the initiating b-hadron.

In the $HH \rightarrow b\bar{b}b\bar{b}$ analysis, corrections are applied to counteract the energy losses due to semi-leptonic b-hadron decays and out-of-cone effects. The first is the *μ-in-jet* correction. A search for a muon is performed in a variable radius cone around the jet axis, as defined in Equation 3.6. The cone radius decreases with muon $\mathrm{p_T}$ to account for the collimation of decay products of high energy jets. If a muon is found within the cone, its four-vector is added to that of the jet. This ensures the full energy of the muon is accounted for in the b-jet. The expected energy the muon deposited in the calorimeter is then subtracted from the jet energy to prevent double-counting.

$$\Delta R(\mu, \mathrm{jet}) < \min\left(0.4, 0.04 + 10/\mathrm{p}_{\mathrm{T}}^{\mu}\ \mathrm{GeV}\right) \tag{3.6}$$

The second correction is to account for energy losses due to the missing neutrino and out-of-cone effects. The correction is a $\mathrm{p_T}$-dependent scale factor applied to b-jets reconstructed in the detector. The scale factor ensures their $\mathrm{p_T}$ is in agreement with equivalent b-jets reconstructed from truth information in Monte Carlo simulation, where all missing particles and energy is included. The correction is derived in $t\bar{t}$ events. The scale factors for jets with and without a muon are derived separately as the two have a consistent difference in energy loss.

These corrections significantly improve mass resolution in Monte Carlo simulation for the individual Higgs bosons and HH, as shown in Figure 3.5.

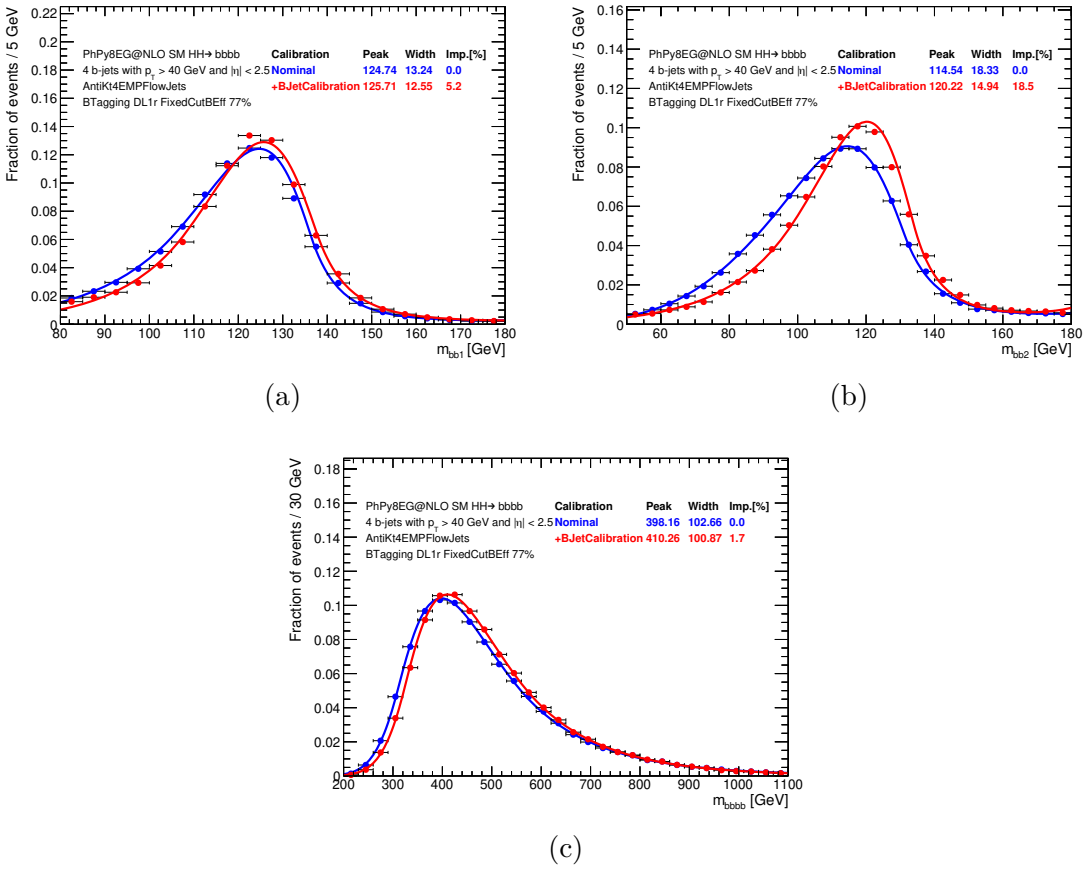


Figure 3.5: Comparison of the m_{H1} (top left), m_{H2} (top right) and m_{HH} (bottom) distributions before (blue) and after (red) the application of the b-jet energy corrections. The legend in the upper right of each plot outlines the improvement in the mass resolution.

4 Reconstruction and Selection

4.1 Introduction to the $HH \rightarrow b\bar{b}b\bar{b}$ Analysis

The following chapters of this thesis describe the activities and results of the non-resonant full Run 2 ($126\,\text{fb}^{-1}$) ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analysis, which I have contributed towards through the majority of my PhD [1].

The focus of the analysis is the simultaneous non-resonant production of two Higgs bosons, known as the di-Higgs (HH), in the $b\bar{b}b\bar{b}$ final state. The high branching ratio of $H \rightarrow b\bar{b}$ results in a relatively large cross-section compared to other HH analyses. However, the primary background source is QCD, and generating sufficient complex QCD Monte Carlo simulation is challenging. Instead, a unique data-driven approach is used to model the background.

This analysis has two separate channels for the two leading HH production processes, ggF and VBF, which are statistically combined when generating results. I spearheaded the VBF analysis, designing and implementing the VBF channel, as described in this chapter. Additionally, I proposed the statistical combination with the ggF channel after conducting a series of studies, as described in Chapter 8.

This analysis is the first *HH* analysis on ATLAS to be designed with two optimised channels for ggF and VBF, providing an improved and broader set of physics results.

Current sensitivity levels are insufficient to observe *HH* production, so this analysis sets statistical constraints on a number of important associated quantities. These include the SM *HH* cross-section (σ_{HH}), the *HHVV* vertex coupling modifier (κ_{2V}), and the trilinear Higgs self-coupling modifier (κ_{λ}).

In Chapter 4, details of the data and simulation samples are described alongside the event selection used to isolate *HH* events and reject background. Chapter 5 describes a number of analysis categories introduced to boost sensitivity. How the prominent background processes are modelled using data-driven techniques is described in Chapter 7. The statistical analysis used to generate the limits on cross-section and coupling modifiers is examined in Chapter 9. In Chapter 10, influential sources of uncertainty are listed and how they were accounted for when deriving results explained. Finally, the results of the analysis are presented and discussed in Chapter 11.

4.2 Data and Monte Carlo Simulation

The $HH \to b\bar{b}b\bar{b}$ analysis used data recorded by the ATLAS detector and taken during the years 2016–18 of Run 2 of the Large Hadron Collider (LHC) at CERN. In total, this amounted to an integrated luminosity of $126\,\mathrm{fb}^{-1}$. The year-by-year breakdown is shown in Table 4.1.

Table 4.1: Per-year integrated luminosity of data used by the ATLAS $HH \to b\bar{b}b\bar{b}$ analysis.

	2016	**2017**	**2018**
$\boldsymbol{\mathcal{L}_{INT}}$	$24.6\,\mathrm{fb}^{-1}$	$43.7\,\mathrm{fb}^{-1}$	$57.7\,\mathrm{fb}^{-1}$

The 3.2 fb^{-1} of data taken during 2015 is not included as the required calibrations are not available.[1] The integrated luminosity for 2016 is below the total recorded value of 32.9 fb^{-1} due to a misconfiguration of one of the b-jet triggers used in 2016, resulting in a loss of usable data for this analysis.

In addition to data, a number of samples containing Monte Carlo (MC) simulated ggF and VBF *HH* events are used in the analysis. For Run 2 of the LHC, an official ATLAS MC production campaign, called *MC16*, was launched. As data-taking conditions vary each year, a new sub-campaign was launched each year. For 2016 the sub-campaign is *MC16a*; for 2017, *MC16d*; and for 2018, *MC16e*. MC simulation samples are labelled using their sub-campaign.

The ggF *HH* samples are simulated using the Powheg Box (V2) event generator [76, 77, 78] at Next-to-Leading Order (NLO) with the Finite-Top-Mass scheme and the Higgs mass, m_H, set to 125 GeV. Pythia8 [79] is used to perform the parton showering and hadronisation. The Parton Distribution Functions (PDFs) are supplied by PDF4LHC [80]. The samples are normalised to 31.05 fb, the SM ggF *HH* cross-section calculated at Next-to-Next-to-Leading Order (NNLO) with the Approximate Finite-Top-Mass scheme [13].

The VBF *HH* samples are simulated using the MadGraph [81] event generator at Leading Order (LO), and using Pythia8 for performing the parton showering and hadronisation. The PDFs are supplied by `NNPDF2.3lo`, with the A14 set of tuned parameters for the underlying-event simulation [82, 83]. These samples are normalised to 1.726 fb, the VBF *HH* cross-section calculated at Next-to-Next-to-Next Leading order (N^{3}LO) [22].

[1] ATLAS calibrations are calculated by dedicated task forces (working under the umbrella term *Combined Performance* (CP)) to meet the needs of analysers. The b-jet trigger CP group has decided not to calculate calibrations for 2015 data due to its relatively small integrated luminosity.

Table 4.2: The range of the nominal *HH* Monte Carlo simulation samples available to the $HH \to b\bar{b}b\bar{b}$ analysis, listed by the variations of their coupling modifiers, κ. All other coupling modifiers are set to the Standard Model value of one unless stated in the table. The Leading Order (LO) cross-section is also included.

***HH* production mode**	**κ value(s)**	**LO σ [fb]**
ggF	SM	27.884
ggF	$\kappa_\lambda = 10$	583.9
VBF	SM	1.1836
VBF	$\kappa_{2V} = 0$	18.593
VBF	$\kappa_{2V} = 0.5$	6.6177
VBF	$\kappa_{2V} = 1.5$	2.3032
VBF	$\kappa_{2V} = 2$	9.9726
VBF	$\kappa_{2V} = 3$	44.944
VBF	$\kappa_\lambda = 0$	3.1659
VBF	$\kappa_\lambda = 2$	1.068
VBF	$\kappa_\lambda = 10$	67.377
VBF	$\kappa_V = 0.5$	7.5289
VBF	$\kappa_V = 1.5$	45.412
VBF	$\kappa_\lambda = 0$, $\kappa_{2V} = 0$	25.4
VBF	$\kappa_\lambda = -5$, $\kappa_V = 0.5$	4.4474

Along with the Standard Model ggF and VBF *HH* samples, a set of samples with non-SM values of the coupling modifiers κ_λ, κ_{2V} and κ_V were also generated (with the same production settings). The *HH* Monte Carlo simulation samples available to the $HH \to b\bar{b}b\bar{b}$ analysis are summarised in Table 4.2. A subset of the non-SM samples were linearly combined as function of κ_λ, and κ_{2V} and κ_V in the case of VBF, to generate the kinematic distributions for any coupling value [84, 85]. The remaining non-SM samples were used to validate the linear combination method. The *HH* mass, m_{HH}, distributions generated directly from the samples and using the linear combination method agreed within the statistical uncertainty.

A set of *alternative* ggF and VBF *HH* samples were generated to assist in estimating the uncertainty due to selecting a specific event generator when other equivalent generators exist (see Section 10.2.3 for details). These samples possess the same con-

figuration as the nominal samples bar using Herwig7 to perform the parton showering and hadronisation [86].

A set of $t\bar{t}$ and dijet QCD samples were generated to help validate the methodology for generating the data-driven background estimate. The $t\bar{t}$ samples are split by decay mode. There are a set of *all hadronic* samples, where the $t\bar{t}$ system decays to a fully hadronic final state, and *non-all hadronic* samples, where the $t\bar{t}$ system decays to a semi-leptonic final state. The $t\bar{t}$ samples are simulated using the POWHEG BOX (V2) event generator and the parton showering and hadronisation is performed using PYTHIA8. The PDFs are supplied by `NNPDF2.3lo`, with the A14 set of tuned parameters for the underlying-event simulation. The resummation damping factor, h_{damp}, which regulates high energy radiation, is tuned to $1.5m_{top}$ [87]. The QCD samples are generated using the same software packages as $t\bar{t}$ apart from using solely PYTHIA8 to perform the parton showering and hadronisation. A requirement for at least four jets with $\mathrm{p_T} > 15\,\mathrm{GeV}$ and two b-jets is stipulated.

The ATLAS detector and the interactions of particles traversing the infrastructure and detector systems are simulated using the GEANT4 toolkit [88].

After their production, the non-SM κ_λ ggF *HH* samples are scaled to their NNLO value [89].[2] The non-SM VBF *HH* samples are scaled to their approximate N^{3}LO value by applying a scale factor equal to the ratio of the SM VBF *HH* N^{3}LO cross-section to the SM VBF *HH* LO cross-section. These samples are not scaled to their exact N^{3}LO value because their cross-sections are only known to LO.

Figure 4.1 shows the variation of the ggF *HH* cross-section with κ_λ, displaying a sharp increase in cross-section as κ_λ is moved away from one. Figures 4.2a and 4.2b

[2] $\sigma_{NNLO} = 70.3874 - 50.4111 \times \kappa_\lambda + 11.0595 \times \kappa_\lambda^2$. To be exact, this is NNLO in the Heavy-Top-Limit with full NLO.

shows the variation in the VBF HH cross-section with κ_λ and κ_{2V} respectively. Both show a steep increase in the cross-section as the coupling modifiers are moved from one, but the increase is far greater when altering κ_{2V}.

κ_V is not discussed at length because the $HH \rightarrow b\bar{b}b\bar{b}$ analysis did not aim to set constraints solely on this modifier, the existing experimental constraints from single H analyses already being to relatively high precision ($\kappa_V = 1.05 \pm 0.04$) [15].

4.3 Triggers

In the $HH \rightarrow b\bar{b}b\bar{b}$ analysis, the $b\bar{b}b\bar{b}$ final state is reconstructed as four b-jets. As such, a suite of b-jet triggers was selected for the triggering stage. The Trigger Menu (the list of all available ATLAS triggers) differs each year. The list of triggers used for each of the different years by the $HH \rightarrow b\bar{b}b\bar{b}$ analysis is summarised in Table 4.3. Each trigger falls into one of two broad categories – `2b2j` or `2b1j`– where the first two characters indicate the number of b-jets required and the latter two indicate the number of additional jets required.

4.3.1 Trigger Bucket Strategy

There is often a discrepancy between the trigger efficiency (the rate at which events pass a given trigger) in data and Monte Carlo (MC) simulation. Therefore, *trigger scale-factors* (SFs) are applied to MC simulation to correct this difference. Deriving and applying these SFs can be a challenge, and the complexity increases with the number of triggers used. In order to simplify the handling of the trigger SFs in this analysis, a strategy was devised based on *trigger buckets*.

Fundamentally, a trigger bucket is just a label applied to an event that states which trigger SF should be applied. The trigger bucket strategy is applied to data from each year individually, since the data and triggers are separated by year. For each year, there are two categories of trigger – `2b2j` and `2b1j` – and two corresponding

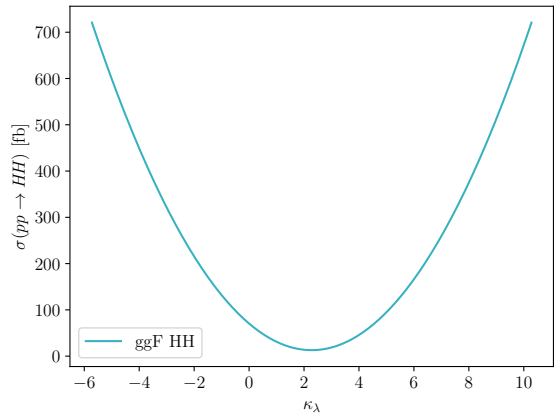
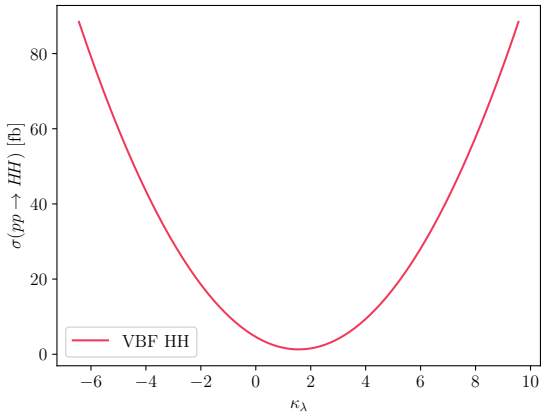
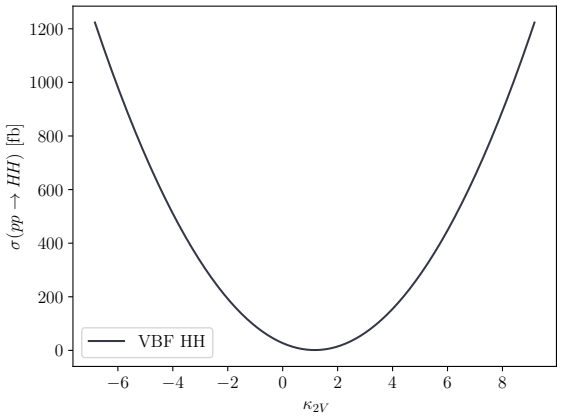


Figure 4.1: Variation in the ggF *HH* cross-section with the trilinear Higgs self-coupling modifier, κ_λ. The minimum point lies at $\kappa_\lambda = 2.28$.



(a)



(b)

Figure 4.2: Variation in the VBF *HH* cross-section with the trilinear Higgs self-coupling modifier, κ_λ, (top) and the *HHVV* vertex coupling modifier, κ_{2V}, (bottom). The minimum points lie at $\kappa_\lambda = 1.56$ and $\kappa_{2V} = 1.17$.

Table 4.3: Selection of b-jet triggers, listed by year and including a brief description, used by the $HH \rightarrow b\bar{b}b\bar{b}$ analysis. WP stands for *working point*, the efficiency at which b-jets are correctly identified by the b-tagging algorithm.

Year	Trigger	Category	Description
2016	HLT_j100_2j55_bmv2c2060_split	2b1j	Two b-jets (60% WP) $p_T > 55\,GeV$ *and* one jet $p_T > 100\,GeV$
	HLT_2j35_bmv2c2060_split_2j35_L14J15.0ETA25	2b2j	Two b-jets (60% WP) $p_T > 35\,GeV$ *and* two jets $p_T > 35\,GeV$
2017	HLT_j110_gsc150_boffperf_split_2j35_gsc55_bmv2c1070_split_L1J85_3J30	2b1j	Two b-jets (70% WP) $p_T > 55\,GeV$ *and* one jet $p_T > 150\,GeV$
	HLT_2j15_gsc35_bmv2c1040_split_2j15_gsc35_boffperf_split_L14J15.0ETA25	2b2j	Two b-jets (40% WP) $p_T > 35\,GeV$ *and* two jets $p_T > 35\,GeV$
2018	HLT_j110_gsc150_boffperf_split_2j45_gsc55_bmv2c1070_split_L1J85_3J30	2b1j	Twp b-jets (70% WP) $p_T > 55\,GeV$ *and* one jet $p_T > 150\,GeV$
	HLT_2j35_bmv2c1060_split_2j35_L14J15.0ETA25	2b2j	Two b-jets (60% WP) $p_T > 35\,GeV$ *and* two jets $p_T > 35\,GeV$

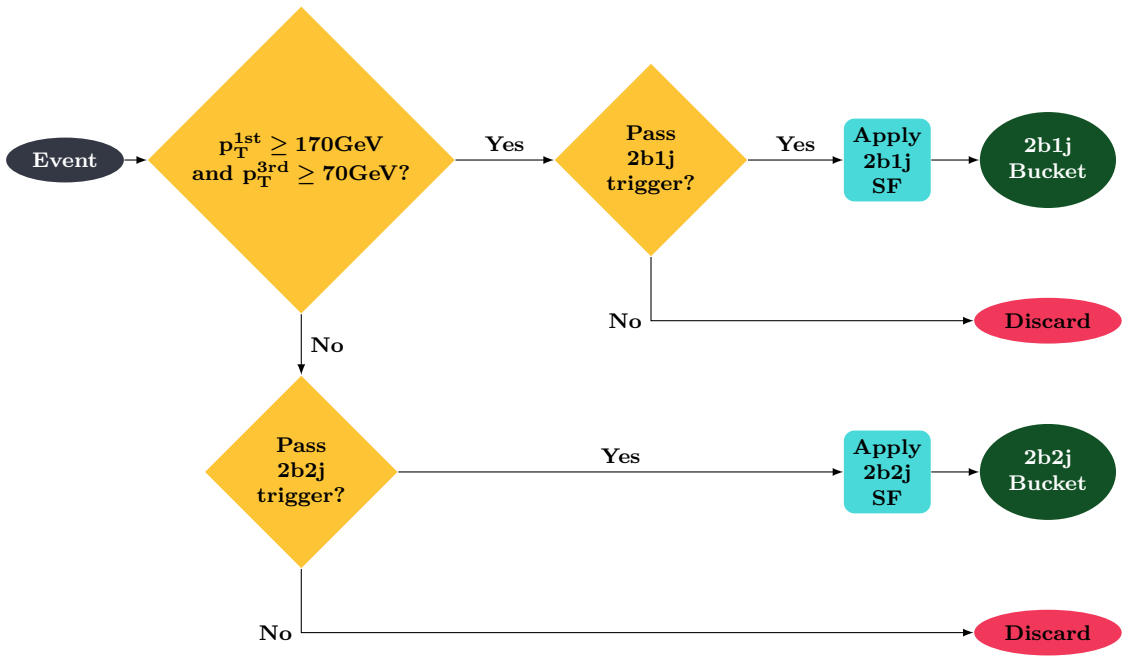
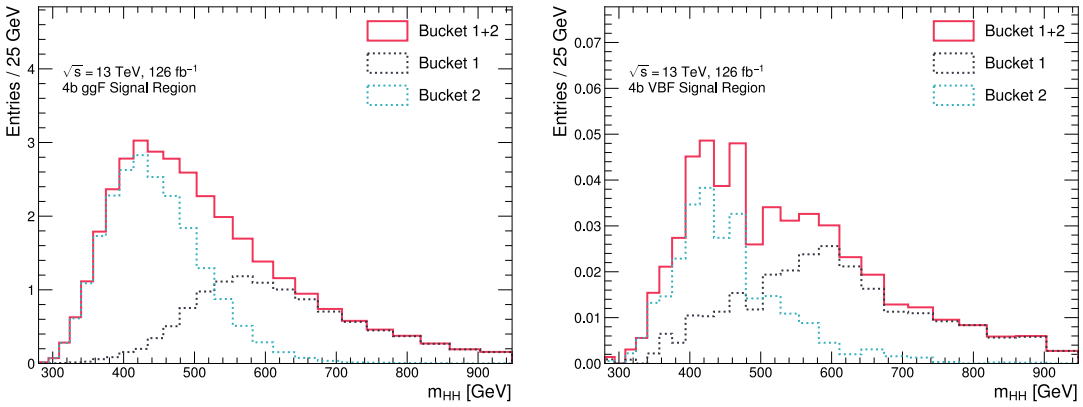


Figure 4.3: Schematic of the *trigger buckets* strategy in the $HH \rightarrow b\bar{b}b\bar{b}$ analysis. (Design credit: Lucas Borgna.)

buckets.

A flowchart outlining how an event is placed in a specific bucket is shown in Figure 4.3 and described below. An event is taken and the assessment for the `2b1j` bucket is conducted. First, a selection that mirrors the cuts of the `2b1j` trigger is applied. If the event passes, then it is checked whether the event passes the `2b1j` trigger. If it does, it is placed in the `2b1j` bucket and the `2b1j` SF is applied. If it does not, it is discarded from the analysis. If the event fails the selection meant to mirror the `2b1j` cuts, it is then checked if it passes the `2b2j` trigger. If it does, then it is placed in the `2b2j` bucket and the `2b2j` SF is applied. If not, the event is discarded.

This strategy simplifies the analysis by removing the ambiguity over which SF to apply to an event that passes both `2b1j` and `2b2j` triggers. Additionally, it removes the need for deriving and applying anti-trigger SFs. Anti-trigger SFs are applied to MC simulation to ensure the rate at which events *fail* a given trigger is the same in data and MC simulation. Instead, the trigger bucket strategy discards events



(a) SM ggF HH (b) $\kappa_{2V} = 0$ VBF HH

Figure 4.4: Breakdown of the contribution of the two trigger buckets to the m_{HH} distribution of SM ggF (left) and $\kappa_{2V} = 0$ VBF HH MC simulation. Bucket 1 (dark blue) corresponds to the `2b1j` trigger and Bucket 2 (light blue) corresponds to the `2b2j` trigger.

that have been observed to fail a trigger. The rationale behind applying a selection that mirrors the trigger cuts before it is checked whether the event passed a trigger is to minimise the number of events discarded. This ensures that there is a high probability an event will have passed the trigger before it is checked if this is the case. An analysis pre-selection mirrors the `2b2j` trigger, so an equivalent cut in the trigger bucketing strategy is not required for `2b2j`.

Figure 4.4 shows the breakdown of the contribution of the two trigger buckets to the m_{HH} distribution for ggF and VBF HH MC simulation. The lower threshold `2b2j` trigger contributes the most events at low m_{HH} and the higher threshold `2b1j` trigger contributes the most events at high m_{HH}.

Analysis studies found that the loss in events due to the trigger bucket strategy had only a small impact on sensitivity, with the loss in data and HH MC simulation being approximately equal at ∼12% [90].

4.4 Object Selection

Hadronic activity, such as the four b-quarks from the *HH* decay and the two additional jets in the VBF topology, are reconstructed as small-radius ($R = 0.4$) anti-k_t Particle Flow (PFlow) jets. Jets reconstructed in the analysis fall into one of two categories, depending on their position in the ATLAS detector. A cut is also applied to the jet p_T to remove low-p_T background jets.

***Central* jets:** $p_T > 40\,\text{GeV}$ and $|\eta| < 2.5$

***Forward* jets:** $p_T > 30\,\text{GeV}$ and $|\eta| \geq 2.5$

Jets initiated by a *b*-quark are identified using the `DL1r` *b*-tagging algorithm with a chosen WP of 77%. By definition, forward jets cannot be *b*-tagged as `DL1r` relies on information from the Inner Detector, which only has coverage up to $|\eta| = 2.5$.

Each *H* is reconstructed from a pair of central jets, and the forward jets are reconstructed to help identify VBF *HH* events, a pair of high-p_T forward jets being characteristic of the VBF topology.

4.5 Event Selection

An overview of the analysis selection is shown in Figure 4.5 and is outlined below.

This analysis has an orthogonal channel for each of the two leading *HH* production modes, ggF and VBF. The cuts described in Items (4), (5) and (6) in Figure 4.5 describe how events are separated into these two channels. These cuts are designed to identify whether an event possesses the two high-p_T and well-separated jets characteristic of the VBF topology, referred to as the *VBF jets* from now on. An event

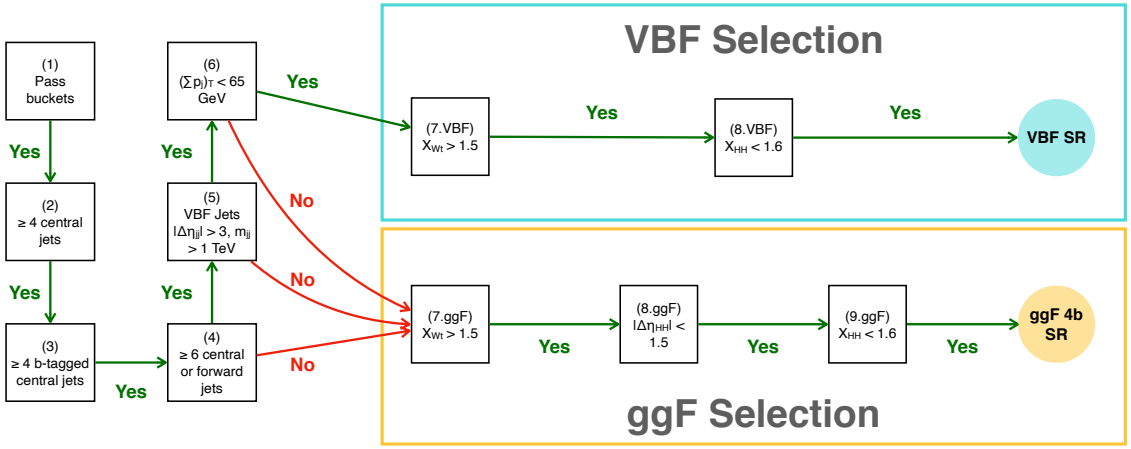


Figure 4.5: Flow chart of the steps in the $HH \to b\bar{b}b\bar{b}$ analysis selection. After Item (6), events are separated in the VBF (blue) and ggF (yellow) channels, which have their own selection steps. (Design credit: Rafael Teixera De Lima.)

must pass all three to be classified as VBF. If it fails any one step, it is classified as ggF, and skips straight to Item (7 ggF).

(1) Events are placed in a trigger bucket.

(2) Events are required to have at least four central jets.

(3) At least four of these central jets must be b-tagged.

(4) Events are required to have at least six of any combination of central and forward jets.

(5) From all non-b-tagged central and forward jets, the pair of with the highest combined invariant mass, m_{jj}, are selected as the VBF jets. Cuts are then applied to check whether the VBF jets have pseudorapidity separation, $|\Delta\eta_{jj}|$, > 3 and a combined invariant mass, m_{jj}, $> 1000\,\text{GeV}$.

(6) Events are required to have the transverse component of the momentum of the

vector sum of the *HH* jets and the VBF jets, $[\sum_j \vec{\mathrm{p}}_j]_{\mathrm{T}}$, $< 65\,\mathrm{GeV}$.

(7 VBF)/(7 ggF) Events are required to have the $t\bar{t}$-veto variable, X_{Wt}, > 1.5 (as described in Section 4.9).

(8 ggF) Events are required to have a pseudorapidity separation between the reconstructed Higgs bosons, $|\Delta\eta_{HH}|$, < 1.5.

(8 VBF)/(9 VBF) Events are required to existed in the *Signal Region*, $X_{HH} < 1.6$ (as described in Section 4.10).

4.6 VBF Channel Optimisation

I designed and implemented the VBF channel in the $HH \to b\bar{b}b\bar{b}$ analysis. There were two significant changes to the selection compared with the previous iteration of the VBF $HH \to b\bar{b}b\bar{b}$ analysis [91]. First, the VBF jets could be reconstructed from *all* jets rather than just forward jets (defined as $|\eta| > 2.0$ in the previous iteration). Second, as the $|\eta|$ requirement was changed, the cuts on m_{jj}, $|\Delta\eta_{jj}|$, and $[\sum_j \vec{\mathrm{p}}_j]_{\mathrm{T}}$ were re-optimised.

The VBF jets were reconstructed from both central and forward jets in order to increase event yields. Increasing the signal yield will improve sensitivity, but the aim was to increase yields generally too, as a large number of events are required for the multi-variate analysis techniques used to derive the data-driven background to be functional and robust.

In order to assess the impact of reconstructing the VBF jets with and without central jets, the significance metric $S/\sqrt{B}$ was calculated and compared for these two scenarios, using the SM VBF *HH* MC simulation signal and background yields

Table 4.4: The cut values tested on VBF-related variables m_{jj}, $|\Delta\eta_{jj}|$ and $[\sum_j \vec{\mathrm{p}}_j]_\mathrm{T}$ when optimising the event selection for the VBF channel.

Variables	**Cuts Tested**
m_{jj}	600, 800, 1000, 1200 GeV
$\|\Delta\eta_{jj}\|$	2.5, 3.0, 4.0, 5.0
$[\sum_j \vec{\mathrm{p}}_j]_\mathrm{T}$	60, 65, 70, 75 GeV

from 2016. The result was a 6% increase in $S/\sqrt{B}$ when using central jets.

With this change, the cuts on m_{jj}, $|\Delta\eta_{jj}|$, and $[\sum_j \vec{\mathrm{p}}_j]_\mathrm{T}$ were re-optimised using $S/\sqrt{B}$ and the same samples. $S/\sqrt{B}$ values were calculated for all permutations of the cuts listed in Table 4.4. The range of cuts were selected starting with the cuts from the previous iteration of the analysis ($m_{jj} > 1000\,\mathrm{GeV}$, $|\Delta\eta_{jj}| > 5.0$ and $[\sum_j \vec{\mathrm{p}}_j]_\mathrm{T} < 60\,\mathrm{GeV}$) and balancing covering as wide a range of cuts as finely as possible with the computational resources and time to run all the calculations.

Figures 4.6a – f display the $S/\sqrt{B}$ values for all cut permutations. The cuts that maximised $S/\sqrt{B}$ were $m_{jj} > 1000\,\mathrm{GeV}$, $|\Delta\eta_{jj}| > 3$ and $[\sum_j \vec{\mathrm{p}}_j]_\mathrm{T} < 65\,\mathrm{GeV}$. The figures show that the optimal cut value was within the range of cuts tested, i.e. there is a peak in $S/\sqrt{B}$ in each plot, and that the value selected for the cuts was close to the true optimal, i.e. close to the peak in the distribution.

4.7 *Two-Tag* and *Four-Tag* Events

As part of the analysis selection, events are labelled as either *two-tag* (or simply *2b*) or *four-tag* (*4b*). This label describes the number of *b*-jets in a given event, with *two-tag* corresponding to two *b*-jets and *four-tag* corresponding to four *or more*. Accurately reconstructed *HH* signal events will be *four-tag* events. *Two-tag* events are kept in the analysis to assist in the building of the data-driven background estimate. Events with zero, one or three *b*-jets are discarded. *Two-tag* and *four-tag* events are

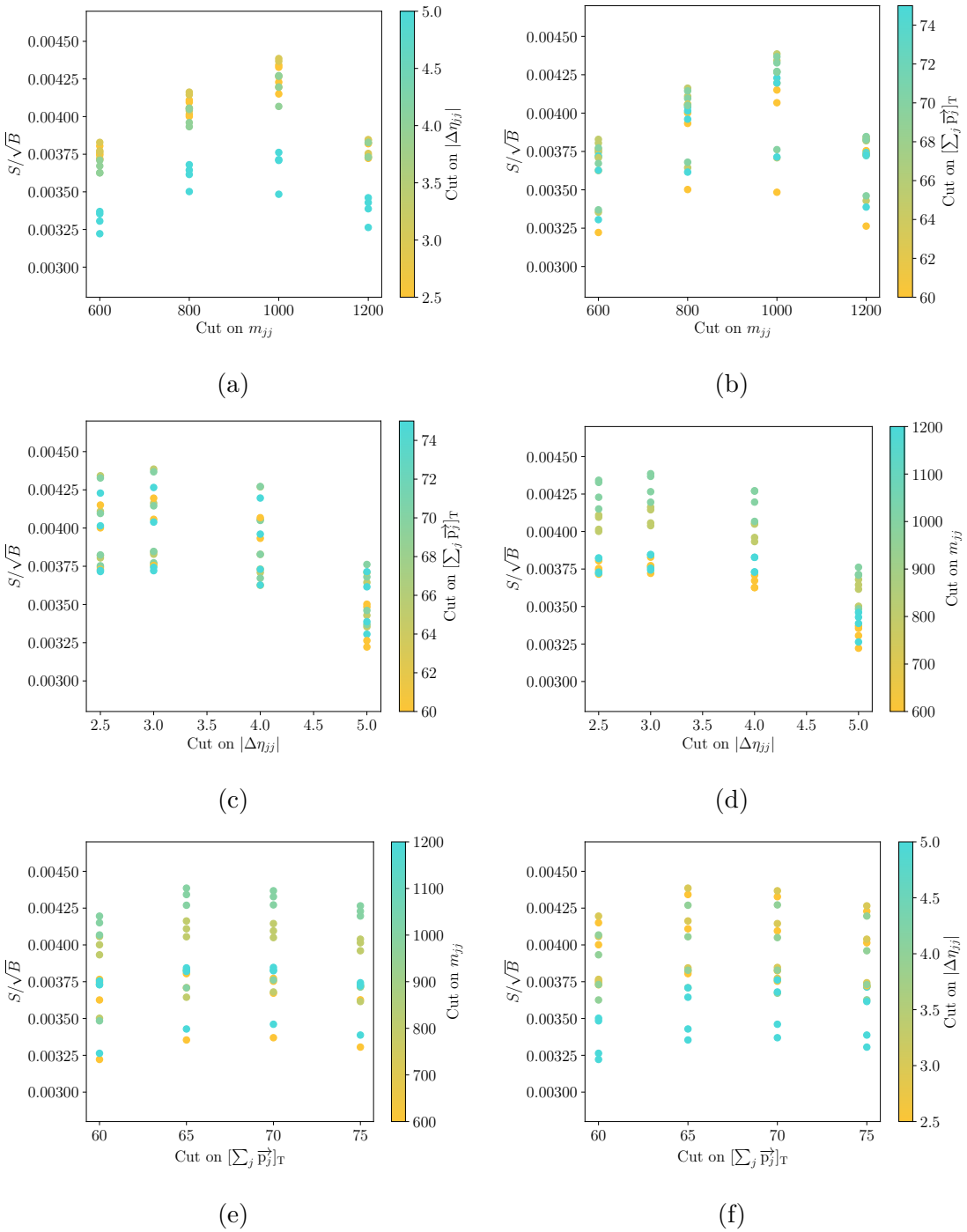


Figure 4.6: Variation in $S/\sqrt{B}$ for SM VBF HH MC simulation and background for 2016 with various cuts on m_{jj}, $|\Delta\eta_{jj}|$, and $[\sum_j \vec{p}_j]_T$. Each plot shows the cut on one variable on the x-axis, and the colour of the points indicates a cut on a second variable, as shown on the colour-bar.

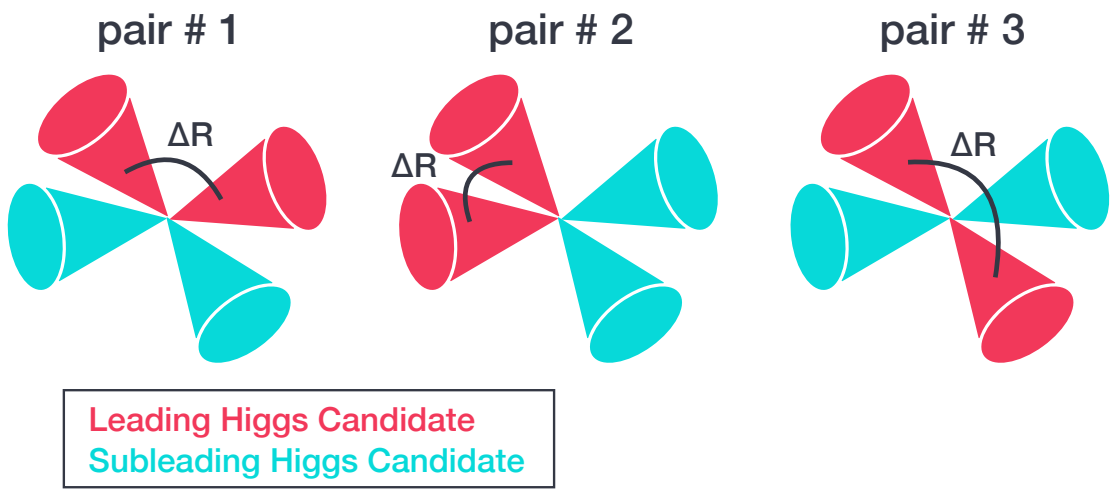


Figure 4.7: The three possible H pairing arrangements of the four b-jets of the HH. The leading Higgs boson candidate is shown in pink, with a representation of the ΔR distance between the two constituent jets included. The sub-leading Higgs is shown in blue. Here, the $HH \rightarrow b\bar{b}b\bar{b}$ analysis pairing strategy would select pair arrangement #2, as the ΔR between the constituent jets of the leading Higgs is smallest.

treated identically in the analysis selection, including in HH reconstruction, even though the *two-tag* events are not placed in the analysis signal region.

4.8 *HH* Reconstruction

The HH system is reconstructed from two pairs of central jets, where each pair is used to reconstruct a single Higgs boson.

4.8.1 *HH* Pairing Strategy

In a given event, there can be several b-jets, so it must be discerned which four are the decay products of the HH. For *four-tag* events with more than four b-jets, the four with the highest p_T are identified as those from the HH decay. For *two-tag* events, the two b-jets are identified as being from the HH decay, in addition to the two highest p_T non-b-tagged jets.

When the four HH jets are identified, it must be decided which should be paired up to reconstruct the two Higgs bosons. There are three possible HH pairing arrangements, as shown in Figure 4.7. The pairing method used in this analysis relies on the principle that in a two-body decay the angle between the two decay products will decrease as the momentum of the parent particle increases. Applying this logic, it is expected that a high-momentum H that decays via $H \to b\bar{b}$ will be observed as two collimated b-jets in the ATLAS detector. For each arrangement, the higher p_T (pink) and lower p_T (blue) Higgs bosons are identified. For the leading H in each arrangement, the angle, ΔR, between the two constituent jets is calculated.[3] The arrangement with smallest ΔR is selected as the correctly paired-up arrangement.

Figure 4.8 shows the pairing accuracy as a function of truth m_{HH} for SM ggF and $\kappa_{2V} = 0$ VBF HH MC in their respective channels. A clear trend is visible – harder events have a higher pairing accuracy. The higher the energy of the initial H the more collimated the two b-jets, making them easier to identify and pair. Figure 4.9 shows the pairing accuracy as a function of κ_λ and κ_{2V} for ggF and VBF HH MC simulation in their respective channels. The same trend is visible, given non-SM κ_{2V} signals possess a relatively hard m_{HH} distribution and non-SM κ_λ signals possess a relatively soft m_{HH} distribution. Alternative pairing methods were explored to increase the pairing accuracy for soft signal. This included a Boosted Decision Tree-based method, which increased the soft signal pairing accuracy substantially [92]. However, this method sculpted the background distribution, resulting in a large background peak in the centre of the analysis signal region, which reduced analysis sensitivity. As such, this method was not used.

[3] *Leading* means the highest in p_T and *sub-leading* means the second highest in p_T.

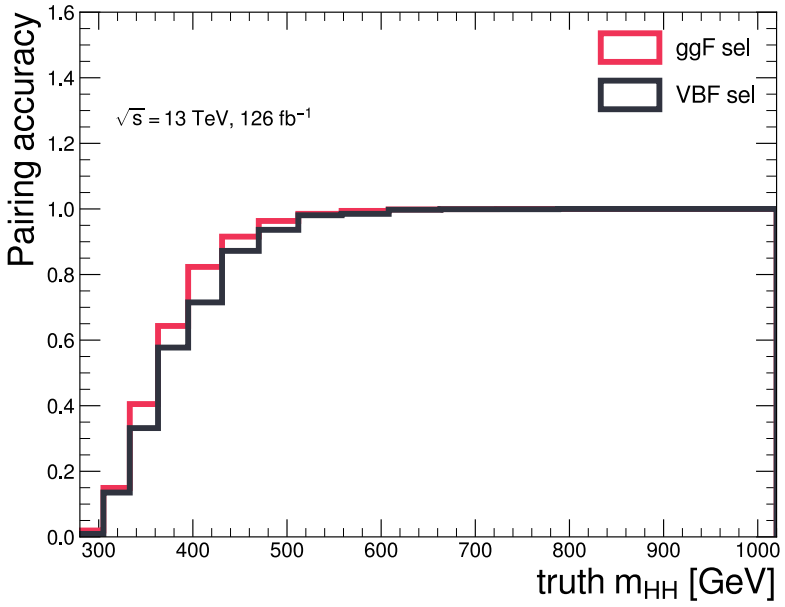
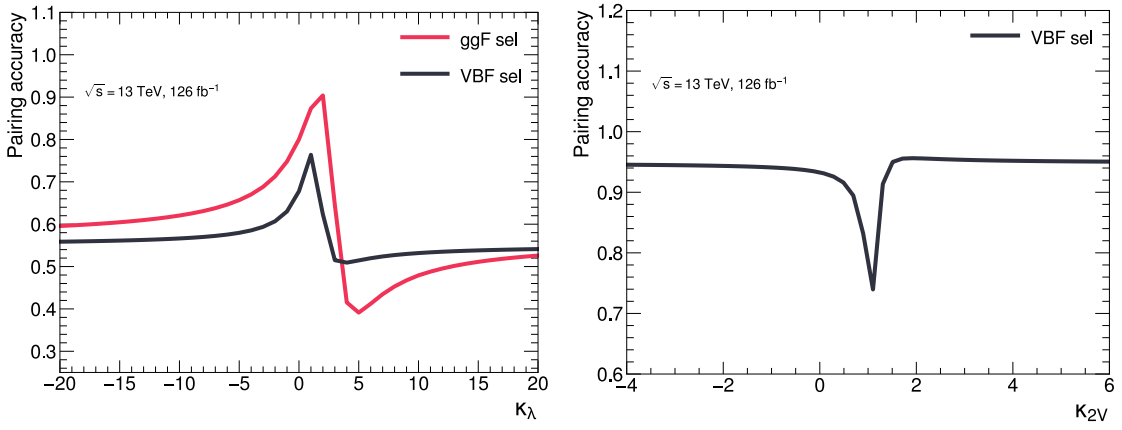


Figure 4.8: Percentage of events where the four jets from the decay of the *HH* are correctly paired into those from individual Higgs decays as a function of the true value of the *HH* mass derived from Monte Carlo (MC) simulation for ggF and VBF *HH* MC signal samples in their respective channels.



(a) κ_λ (b) κ_{2V}

Figure 4.9: Percentage of events where the four jets from the decay of the *HH* are correctly paired into those from individual Higgs decays as a function of κ_λ (left) and κ_{2V} (right) for ggF and VBF *HH* Monte Carlo simulation signal samples in their respective channels.

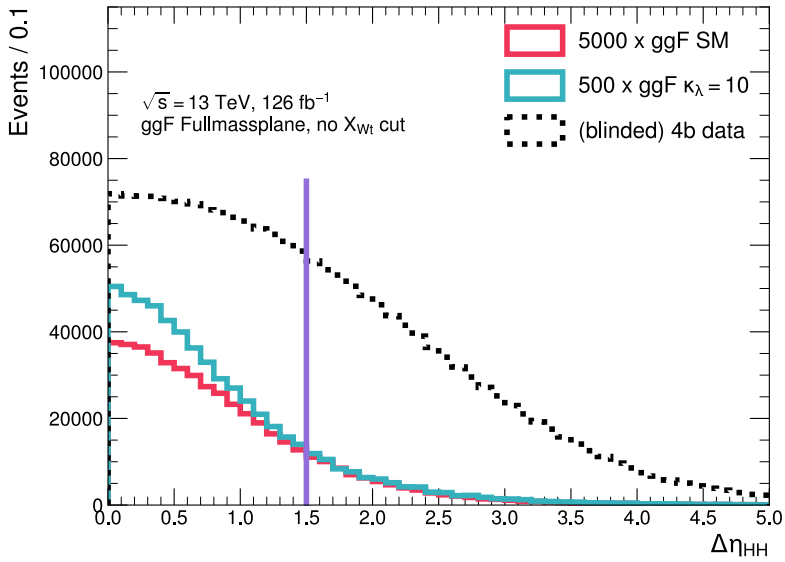


Figure 4.10: The $|\Delta\eta_{HH}|$ distribution for SM (pink) and $\kappa_\lambda = 10$ (light blue) ggF HH Monte Carlo simulation and blinded *four-tag* data (dark blue). The $|\Delta\eta_{HH}| < 1.5$ cut is represented by the purple vertical line.

4.9 Background Reduction

The primary source of background in this analysis is from QCD, e.g. $g \to b\bar{b}$ being misidentified as $H \to b\bar{b}$. The angular distribution of background jets should be random; whereas, for true ggF HH events, there is a high probability the Higgs bosons will be back-to-back. To reduce background, a maximum cut on the pseudorapidity separation of the two Higgs bosons, $|\Delta\eta_{HH}| < 1.5$, is applied to events in the ggF channel. The same cut is not applied to events in the VBF channel as it reduces sensitivity to SM VBF HH signal. The $|\Delta\eta_{HH}|$ distribution for ggF HH signal for two important values of κ_λ and for blinded *four-tag* data (used to represent the background) is shown in Figure 4.10.[4]

The second-leading source of background comes from top quark pair ($t\bar{t}$) production. Top quarks readily decay via the weak interaction into a bottom quark and W boson. This W boson can either decay to a quark or lepton pair; if the former, the overall

[4] *Blinded* means that the data does not contain events from the analysis signal region. The events only come from the control regions.

top decay is called *all-hadronic*, and, if the latter, it's called *semi-leptonic*.

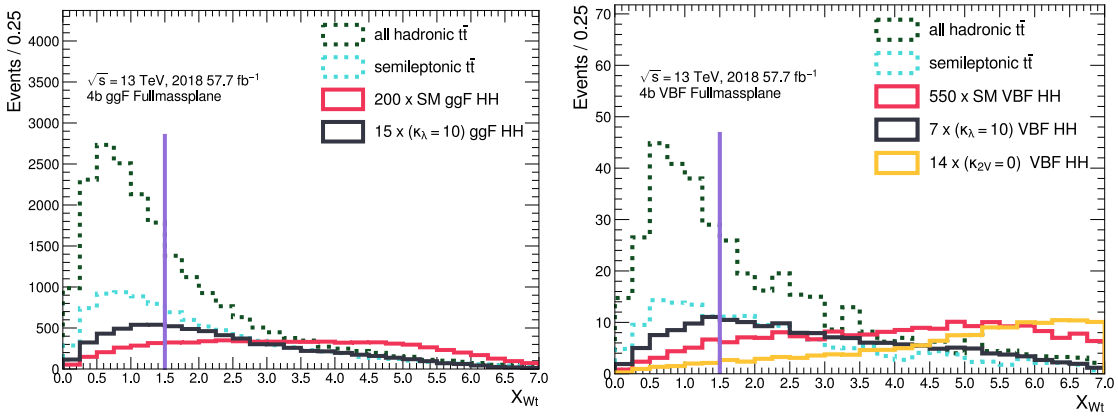
In order to reduce the background from top quark pairs, a cut is placed on X_{Wt} defined in Equation 4.1. X_{Wt} is designed to quantify the likelihood an event contains a hadronically decaying t i.e. $t \to bW(\to q\bar{q})$. The smaller the value of X_{Wt}, the higher the likelihood.

$$X_{Wt} = \sqrt{\left(\frac{m_W - 80.4\text{GeV}}{0.1 \times m_W}\right)^2 + \left(\frac{m_t - 172.5\text{GeV}}{0.1 \times m_t}\right)^2} \tag{4.1}$$

In order to calculate X_{Wt}, potential W bosons and top quarks must be reconstructed. For a given event, a single value of X_{Wt} is not calculated. Rather, every possible t and W combination is reconstructed, a corresponding series of X_{Wt} values calculated, and the minimum value of X_{Wt} is cut on.

To explain the X_{Wt} cut in more detail, first, a b-tagged central jet is identified and removed from the pool of all central jets (excluding any VBF jets). This b-jet is meant to originate from $t \to bW$. From the remaining jets, all possible jets pairings are used to reconstruct all W candidates. Then, all the W candidates are combined with the singled-out b-jet to reconstruct all possible t candidates. For each arrangement of t and W candidate, a value of X_{Wt} is calculated and added to a list. Then, another b-tagged central jet is identified and the process begins again, adding more X_{Wt} values to the existing list. The process continues until all b-tagged central jets have been used. Once all the X_{Wt} values have been calculated, the minimum from the list is taken, as it is the most likely to correspond to true $t \to bW(\to qq)$. A cut, $X_{Wt} < 1.5$, is then applied to this minimum value of X_{Wt}.

Distributions of minimum X_{Wt} are shown in Figure 4.11 for important coupling



(a) ggF Channel

(b) VBF Channel

Figure 4.11: X_{Wt} distributions for ggF and VBF *HH* Monte Carlo (MC) simulation in their respective channels. In both cases, the distributions for all-hadronic and semi-leptonic $t\bar{t}$ MC simulation is also shown. The cut at $X_{Wt} < 1.5$ is shown as the vertical purple line.

points for ggF, VBF and $t\bar{t}$ MC simulation in both the ggF and VBF channels. From these distributions, it can be seen that the $X_{Wt} < 1.5$ cut removes a substantial number of the top quark-initiated events, but has a far smaller effect on ggF and VBF *HH*.

4.10 Region Definitions

The last step in the selection is to place events in the analysis regions. There is the Signal Region (SR) and two control regions: Control Region 1 (CR1) and Control Region 2 (CR2). The Signal Region is designed to capture the majority of the ggF and VBF *HH* events. Control Region 1 is used to help build the data-driven background estimate, and Control Region 2 is used to help validate it. All regions are designated using X_{HH}, defined in Equation 4.2.

$$X_{HH} = \sqrt{\left(\frac{m_{H1} - 124\text{GeV}}{0.1 \times m_{H1}}\right)^2 + \left(\frac{m_{H2} - 117\text{GeV}}{0.1 \times m_{H2}}\right)^2} \tag{4.2}$$

There are two event variables in X_{HH}: the masses of the Higgs bosons, m_{H1} and m_{H2}.

The individual distributions of m_{H1} and m_{H2} should be centred around 125 GeV (the Higgs boson mass), with a small degree of spread due to detector resolution. Therefore, on a 2D m_{H1} vs. m_{H2} plot, the distribution for true *HH* events should peak at (125, 125). For background events, the distribution should be a smoothly decreasing slope from low Higgs masses to high, given low energy background events are more probable than high.

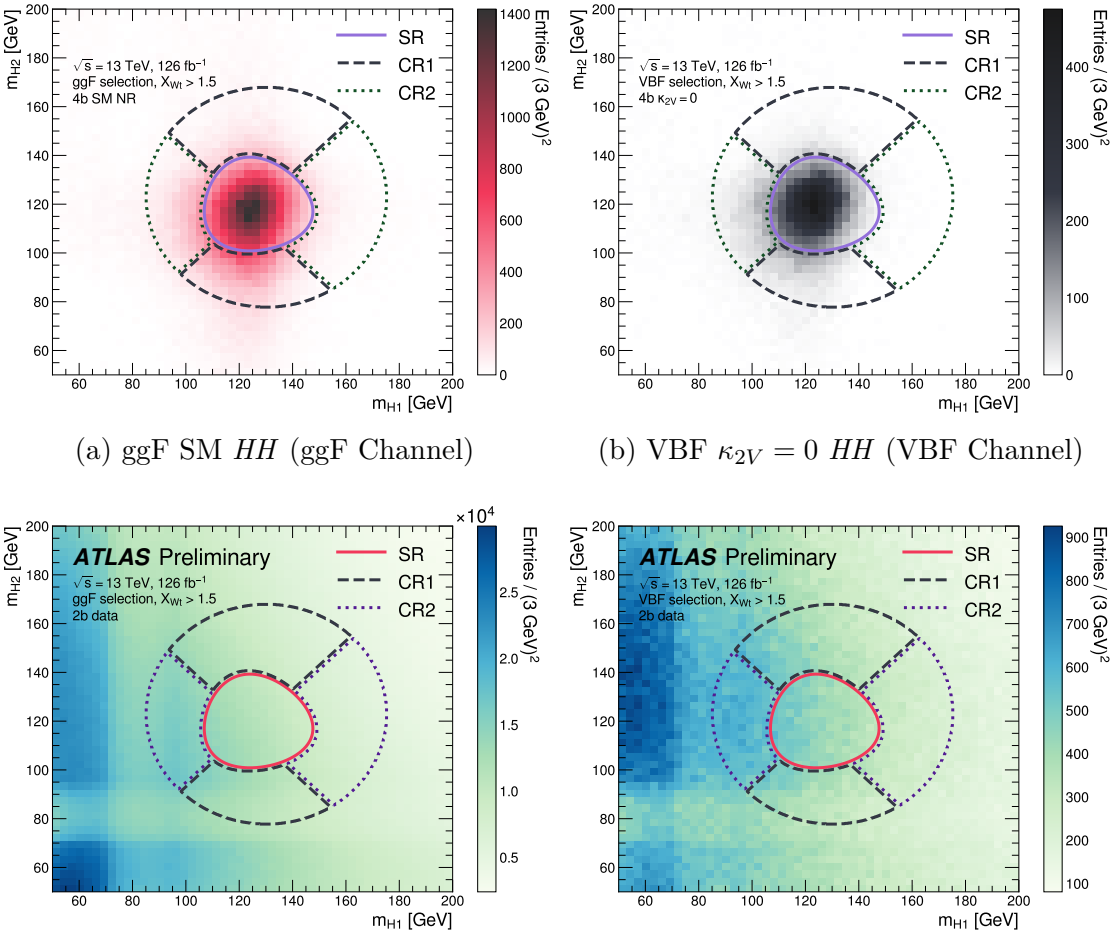
X_{HH} quantifies the proximity of a given event to the central (125, 125) peak in the theoretical *HH* m_{H1} vs. m_{H2} distribution. The closer to the peak, the smaller the value of X_{HH}. In practice, the peak is not centred at exactly (125, 125), due to leptonic decays of *b*-hadrons resulting in lost energy in the form of undetected neutrinos. The centre was found to be closer to (124, 117), m_{H2} being defined as the mass of the sub-leading *H*, so systematically possessing a smaller value (thus 117).

Figure 4.12 shows the distributions in the m_{H1} vs. m_{H2} plane for two important signal samples and *two-tag* data in the ggF and VBF channels. This provides a visual representation of the SR, CR1 and CR2. The definition of the SR is given in Equations 4.3. The cut value of 1.6 and shape of the SR were optimised for this analysis. CR1 and CR2 are made up of two oppositely-facing quadrants within the band with edges defined in Equations 4.4 and 4.5.

$$\text{SR: } X_{HH} < 1.6 \tag{4.3}$$

$$\text{CRs inner edge: } X_{HH} = 1.6 \tag{4.4}$$

$$\text{CRs outer edge: } \sqrt{\left(m_{H1} - 1.05 \cdot 124\text{GeV}\right)^2 + \left(m_{H2} - 1.05 \cdot 117\text{GeV}\right)^2} = 45\text{GeV} \tag{4.5}$$



(a) ggF SM HH (ggF Channel)

(b) VBF $\kappa_{2V} = 0$ HH (VBF Channel)

(c) *Two-tag* data (ggF Channel)

(d) *Two-tag* data (VBF Channel)

Figure 4.12: Distributions in m_{H1} vs. m_{H2} plane for HH MC simulation samples and data in the ggF and VBF Channels. The Signal Region (SR) is the area within the solid line, Control Region 1 (CR1) is the area within the wide-dashed lines, and Control Region 2 (CR2) is the area within the fine-dashed lines.

In Figures 4.12a and 4.12b, the signal peaks in the SR, with the majority being captured within it. This is the case for all signal samples used in the $HH \rightarrow b\bar{b}b\bar{b}$ analysis. In Figures 4.12c and 4.12d, the majority of *two-tag* events lie outside of the SR, peaking around low m_{H1} and decreasing as m_{H1} and m_{H2} increase. The lightly-shaded bands that start at the bottom left corner of these plots and extend vertically and horizontally are the result of the X_{Wt} cut (they are centred around 80.4 GeV, the W boson mass).

The control regions are designed to be in close proximity to the SR to ensure the background events within both regions are as similar as possible. This is because the events in the control regions are used to help build the data-driven background estimate for the SR. The orientation of the control region quadrants within the band was optimised to ensure the similarity of CR1 and CR2 events. The final choice of orientation is such that the boundaries between CR1 and CR2 are at, counting clockwise, 45°, 135°, 225° and 315° with respect an imaginary vertical line that passes through the centre of the SR. This ensures that the drop in events due to the X_{Wt} cuts impacts the CR1 and CR2 equally (the lightly-shaded bands extend into CR1 and CR2 by approximately the same amount in Figures 4.12c and 4.12d). See Chapter 7 for details on the background estimation.

4.11 Event Yields

Figure 4.13 shows the values of acceptance times efficiency (A × E) for each step in the analysis event selection for κ_λ variations of ggF HH MC simulation in the ggF channel, κ_λ variations of VBF HH MC simulation in the VBF channel and κ_{2V} variations of VBF HH MC simulation in the VBF channel. The A × E for BSM ggF HH κ_λ points tends to be lower than for the SM point. This is due to BSM κ_λ signal being relatively soft and thus the acceptance being lower. The reverse is true for BSM VBF HH κ_{2V} signal, which is relatively hard. Additional A × E plots are

included in Appendix A.

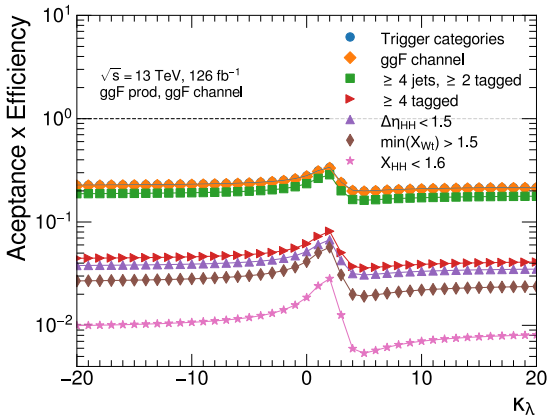
Tables 4.5 and 4.6 show the *four-tag* SR yields for ggF and VBF *HH* MC simulation in the ggF and VBF channels.[5] Yield means the number of events, often expressed as a decimal because MC is weighted. Approximately 1–2% of the overall ggF yield for the various coupling points shown falls in the VBF channel, demonstrating that the addition of the VBF channel does very little to dilute the sensitivity to ggF. A far larger percentage of the overall VBF yield falls in the ggF channel, but the definition of the VBF channel was optimised to maximise significance, resulting in tight cuts that reject far more background than signal. Background yields are shown in Tables 7.2 and 7.3 in Chapter 7.

Figures 4.14 and 4.15 show the variation in the *four-tag* SR yields and significance ($S/\sqrt{B}$) versus κ_λ and κ_{2V} respectively.[6] Even though there is more VBF signal in the ggF channel than the VBF channel for the majority of modifier values, the significance is far larger in the VBF channel. As ggF *HH* possesses no sensitivity to κ_{2V} it is treated as a background process when setting constraints on κ_{2V}.

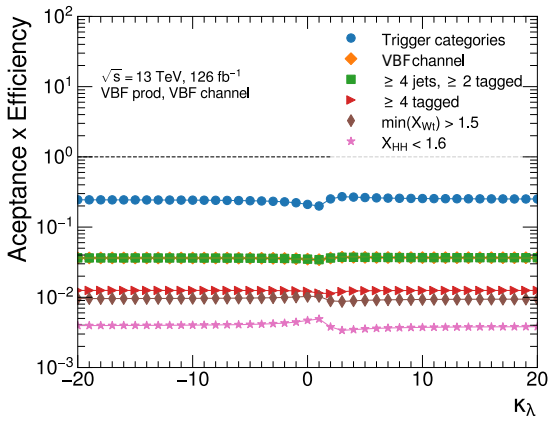
The full list of event selection cutflows for data, SM and $\kappa_\lambda = 10$ ggF *HH* MC simulation and SM, $\kappa_\lambda = 10$ and $\kappa_{2V} = 0$ VBF *HH* MC simulation are shown in Appendix B.

[5]The yields shown in Tables 4.5 and 4.6 and Figures 4.14 and 4.15 are extracted from the final m_{HH} histograms. The m_{HH} binning is described in Section 5.3, and starts at 400 GeV in the VBF channel.

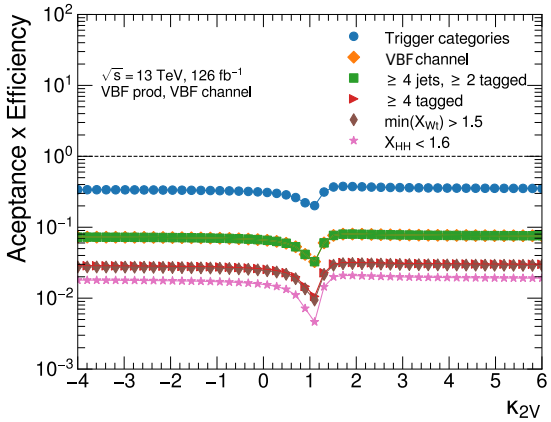
[6]See Footnote 5.



(a) κ_λ ggF *HH*, ggF channel



(b) κ_λ VBF *HH*, VBF channel



(c) κ_{2V} VBF *HH*, VBF channel

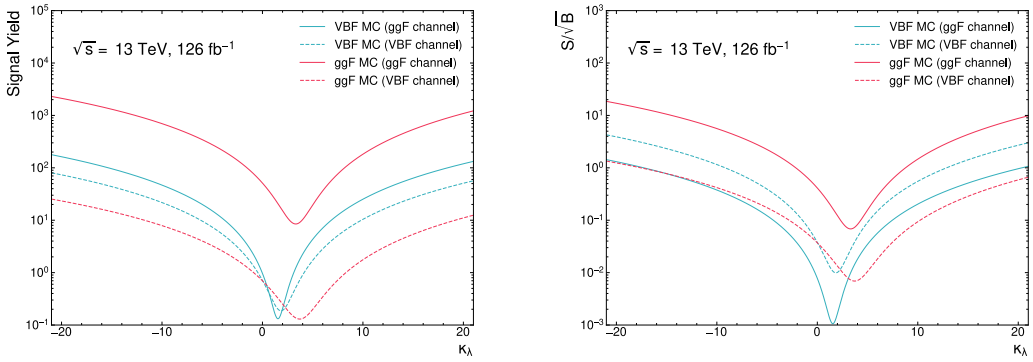
Figure 4.13: The acceptance times efficiency for each step in the analysis event selection for κ_λ variations of ggF *HH* Monte Carlo (MC) simulation in the ggF channel (top), κ_λ variations of VBF *HH* MC simulation in the VBF channel (bottom left) and κ_{2V} variations of VBF *HH* MC simulation in the VBF channel (bottom right). Each step in the selection is represented by a different coloured line, with the legend in the top right corner of the plot showing the steps in order.

Table 4.5: The event yields in the *four-tag* Signal Region for two coupling points of interest for ggF *HH* Monte Carlo simulation normalised to $126\,\text{fb}^{-1}$.

Coupling Point	**ggF Channel Yield**	**VBF Channel Yield**
SM	29.1	0.430
$\kappa_\lambda = 10$	182	1.77

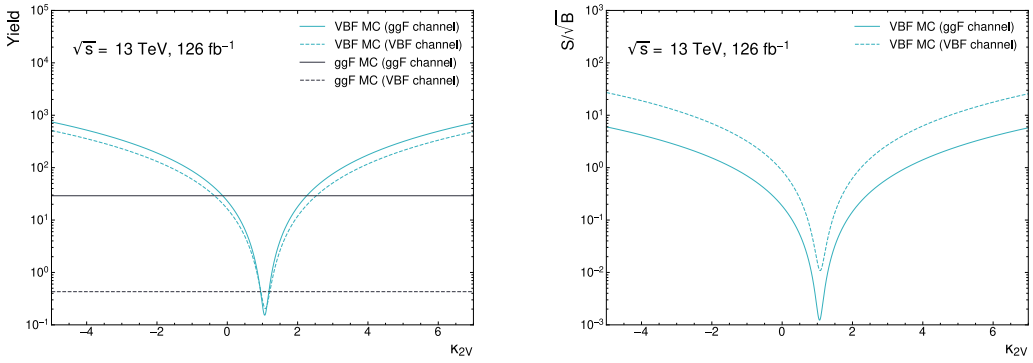
Table 4.6: The event yields in the *four-tag* Signal Region for three coupling points of interest for VBF *HH* Monte Carlo simulation normalised to $126\,\text{fb}^{-1}$.

Coupling Point	**ggF Channel Yield**	**VBF Channel Yield**
SM	0.235	0.297
$\kappa_\lambda = 10$	25.2	10.4
$\kappa_{2V} = 0$	23.1	16.4



(a) (b)

Figure 4.14: *four-tag* Signal Region yield (left) and statistical significance (right) of the VBF and ggF *HH* Monte Carlo simulation in the VBF and ggF channels versus κ_λ.



(a) (b)

Figure 4.15: Left: VBF *HH* signal and ggF *HH* background yields in the *four-tag* Signal Region in the VBF and ggF channels versus κ_{2V} using VBF and ggF *HH* Monte Carlo simulation. Right: statistical significance of VBF *HH* signal in the VBF and ggF channels versus κ_{2V}. The statistical significance is with respect to the full background estimate not solely the ggF *HH* background.

5 Categorisation

In the $HH \to b\bar{b}b\bar{b}$ analysis, the data is stored in histograms of m_{HH} prior to being used in the statistical analysis that sets limits on HH production and couplings. m_{HH} displays a good degree of separation between signal and background, and data was binned in m_{HH} in the previous iteration of the $HH \to b\bar{b}b\bar{b}$ analysis [93]. Binning in m_{HH} improves analysis sensitivity, as it allows the contrasting shapes of the signal and background m_{HH} distributions to be accounted for in the statistical analysis.

Two other variables that showed good discrimination between HH signal and background were $|\Delta\eta_{HH}|$ and X_{HH}.[1] The discriminating power of these variables was leveraged by introducing *categories*. Categories are simply large analysis regions, events being placed in a particular region depending on their value of $|\Delta\eta_{HH}|$ and X_{HH}. A schematic overview of the $|\Delta\eta_{HH}|$ and X_{HH} categories, including the bin boundaries, is shown in Figure 5.1.

The categorisation differs between the ggF and VBF channels. This is primarily due to the ggF channel containing more events, so it can be divided into a greater number of categories without the increased statistical uncertainty becoming a limiting factor.

[1] Defined in Equation 4.2.

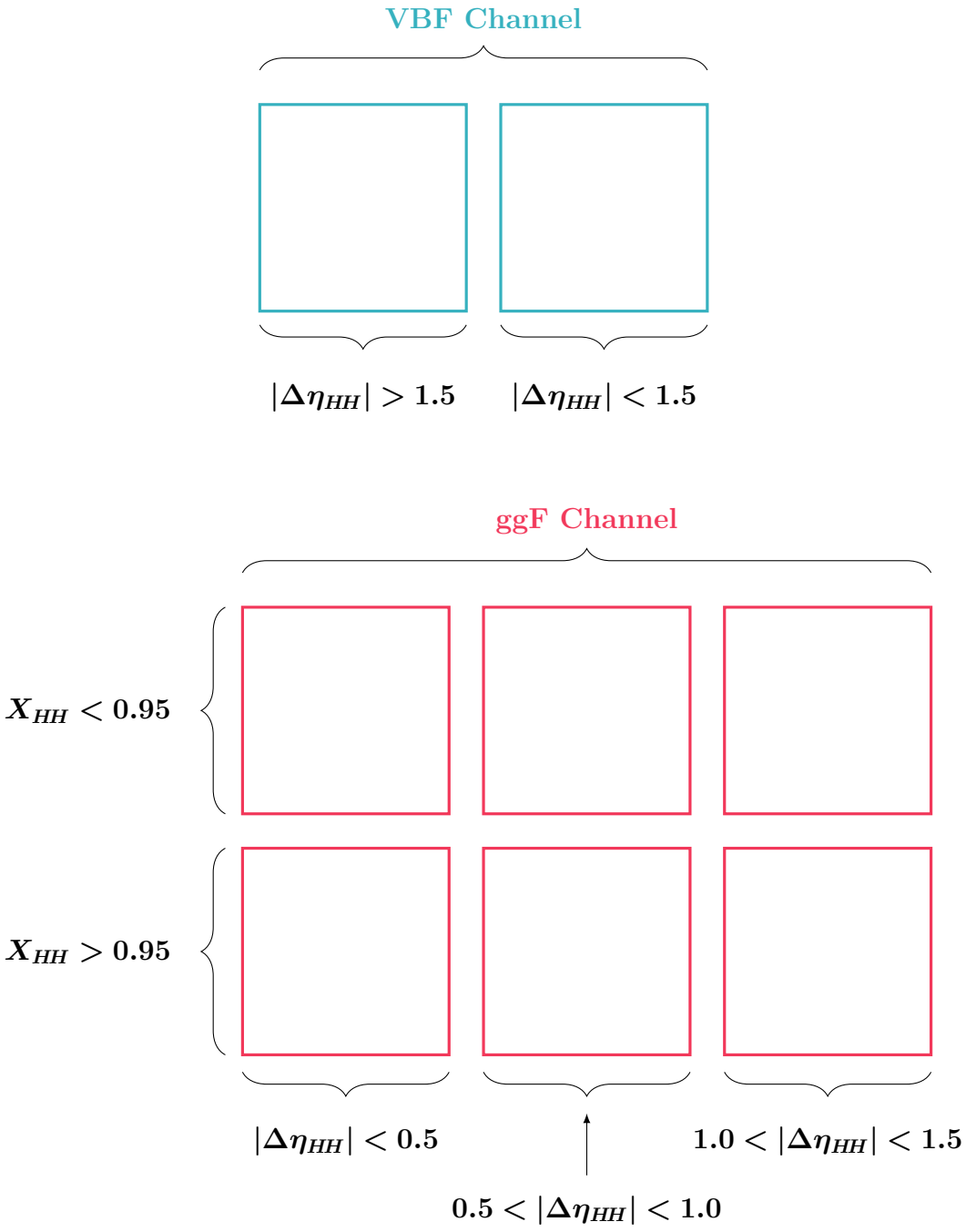
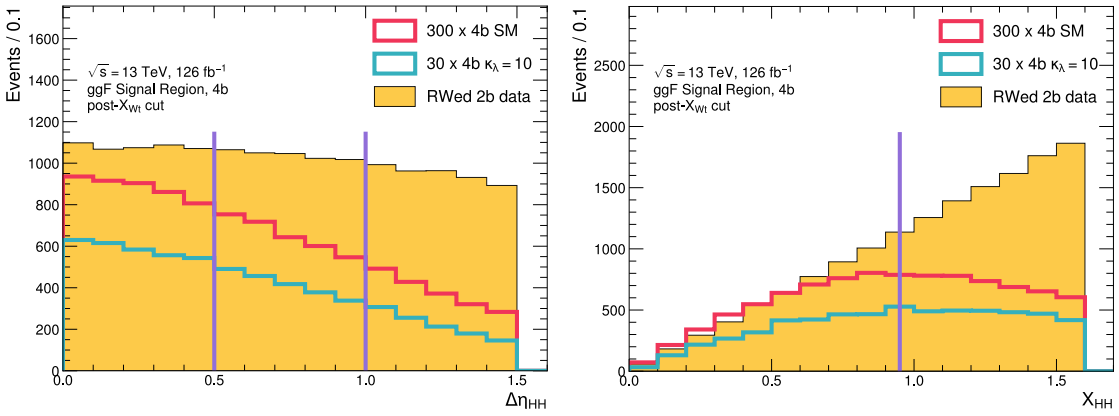


Figure 5.1: Schematic overview of the $|\Delta\eta_{HH}|$ and X_{HH} categories in the $HH \to b\bar{b}b\bar{b}$ analysis, each category being represented by a coloured square, with the variables and cuts used to define each category positioned on the axes.



(a) $|\Delta\eta_{HH}|$ (b) X_{HH}

Figure 5.2: Distribution of the categorisation variables, $|\Delta\eta_{HH}|$ and X_{HH}, for SM (pink) and $\kappa_\lambda = 10$ (blue) ggF HH signal and background (yellow, *RWed 2b data*) in the *four-tag* Signal Region of the ggF channel. The vertical purple lines show the category boundaries. Signal distributions are scaled as to be visible next to the more dominant background.

These differences will be covered in the following sections.

5.1 ggF Channel Categorisation

The ggF channel is divided into six categories, shown in Figure 5.1. As part of the analysis selection, a cut is placed on both X_{HH} and $|\Delta\eta_{HH}|$: $X_{HH} < 1.6$ and $|\Delta\eta_{HH}| < 1.5$. Figure 5.2 shows the distributions of ggF HH signal and background in terms of X_{HH} and $|\Delta\eta_{HH}|$.[2]

For simplicity, the position of the $|\Delta\eta_{HH}|$ categorisation boundaries were equally-spaced within $|\Delta\eta_{HH}| < 1.5$. The position of the boundary in X_{HH} was selected as it maximised the significance metric, $S/\sqrt{B}$, for SM ggF HH signal. It was observed that sensitivity increases with the fineness of the binning in X_{HH} and $|\Delta\eta_{HH}|$; however, this also elevates the complexity of the analysis, and the amount of work to validate the background estimate, which must be checked in every category. The choice of 3 $|\Delta\eta_{HH}|$ × 2 X_{HH} categories was selected as a compromise between

[2]The legend in the figures in this chapter use the shorthand "RWed 2b data" for reweighted *two-tag* data.

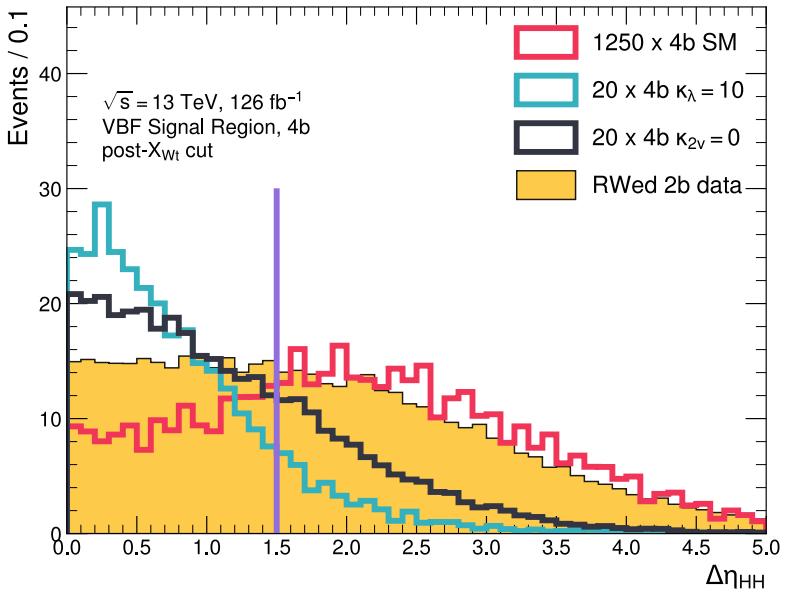


Figure 5.3: $|\Delta\eta_{HH}|$ distributions for SM (pink), $\kappa_\lambda = 10$ (light blue) and $\kappa_{2V} = 0$ (dark blue) VBF *HH* Monte Carlo simulation and background (yellow, *RWed 2b data*) in the *four-tag* Signal Region of the VBF channel. The vertical purple line shows the boundary of the $|\Delta\eta_{HH}|$ categorisation. Signal distributions are scaled as to be visible next to the more dominant background.

these two competing factors. It was found that by separating events into X_{HH} and $|\Delta\eta_{HH}|$ categories, the upper limit on ggF *HH* production was reduced by between ∼28–35% across the κ_λ spectrum.

5.2 VBF Channel Categorisation

The VBF channel has two categories: $|\Delta\eta_{HH}| > 1.5$ and $|\Delta\eta_{HH}| < 1.5$. Figure 5.3 shows the $|\Delta\eta_{HH}|$ distributions for VBF *HH* signal and background. In the previous iteration of the VBF $HH \to b\bar{b}b\bar{b}$ analysis, a $|\Delta\eta_{HH}| < 1.5$ cut was applied in the selection [3]. As can be seen from Figure 5.3, this cut is effective at accepting non-SM VBF *HH* signal and rejecting background; however, the signal acceptance for SM VBF *HH* is poor. I observed this first, and brought it to the attention of the analysis group. In response, the $|\Delta\eta_{HH}|$ cut was changed to a categorisation, with the cut value turned into the category boundary. With this change, the SM VBF *HH* signal was kept in the analysis, and the sensitivity to non-SM *HH* signal was

maintained.

Categorisation in X_{HH} and finer categorisation in $|\Delta\eta_{HH}|$ was not implemented in the VBF channel due to the relatively low event yields. The larger the number of categories, the fewer events in each category. For a channel with relatively low statistics, this would mean each category would have high statistical uncertainties and poses challenges when validating the background estimate in each category.

5.3 m_{HH} Binning

Once the analysis categories were defined, the m_{HH} binning was optimised. The binning for m_{HH} was assigned using a logarithmic binning scheme, which works as follows. After the lowest bin edge is stipulated, the second bin edge is placed at a distance from the first equal to X% of the first bin edge. X is a user-defined percentage. Then, the third bin edge is placed at a distance away from the second equal to X% of the second bin edge, and so on. For example, if the user specifies the first bin edge is at 250 GeV and X = 10%, the bin edges would be:

$$[\mathbf{250}, \quad 250 + (10\% \text{ of } 250) = \mathbf{275}, \quad 275 + (10\% \text{ of } 275) = \mathbf{302.5}, \quad ...].$$

The logarithmic scheme provides fine bins at small bin edge values that gradually increase in width as the bin edge value rises. Generally, the m_{HH} distributions peaks around low values of m_{HH}, and then steady decreases as m_{HH} increases, with a long tail stretching out to higher values of m_{HH}. The logarithmic scheme ensures there are fine bins in the high-statistics low-m_{HH} region and wide bins in the low-statistics high-m_{HH} region. So, the fine details of the low m_{HH} region can be exploited, whilst minimising the impact of statistical fluctuations in the high m_{HH} tail.

The position of the lowest and highest m_{HH} bin edges, as well as the value of X,

Table 5.1: List of parameter settings associated with the logarithmic m_{HH} binning scheme in each channel and category. *Min* is the lowest bin edge, *Max* is the highest bin edge, X is the bin size parameter and *rounding* is the rounding applied to each bin edge.

Channel	Category	Min [GeV]	Max [GeV]	X [%]	Rounding
ggF	All	280	950	9	to nearest 1
VBF	$\lvert\Delta\eta_{HH}\rvert < 1.5$	400	965	10	to nearest 5
VBF	$\lvert\Delta\eta_{HH}\rvert > 1.5$	400	1490	9	to nearest 5

were optimised independently for the ggF and VBF channels. I performed the optimisation for the VBF channel, designing separate m_{HH} binnings for the two $|\Delta\eta_{HH}|$ categories, due to a difference in the length of their high-m_{HH} tails. In all instances the aim was to maximise the fineness of the bins, whilst ensuring the statistical uncertainty in any bin did not rise above 30%. This analysis uses an asymptotic approximation for the test statistic when deriving results (described in Chapter 9). As a rule of thumb, the threshold for this approximation to falter is when there are less than ∼10 events. The statistical uncertainty on 10 events as a percentage is $\sqrt{10}/10 = 0.316 = 31.6\%$. This was rounded to the 30% value.

Table 5.1 describes the final m_{HH} binning in each channel and category. For the VBF channel, mismodeling of the background estimate was observed at low values of m_{HH} in the `3b1f` background validation region.[3] To counteract this, the lowest bin edge was moved to 400 GeV, with events below this edge being discarded.[4] This was found to have negligible impact on analysis sensitivity, the low m_{HH} region being very high in background. All bin edges above 400 GeV were left unchanged.

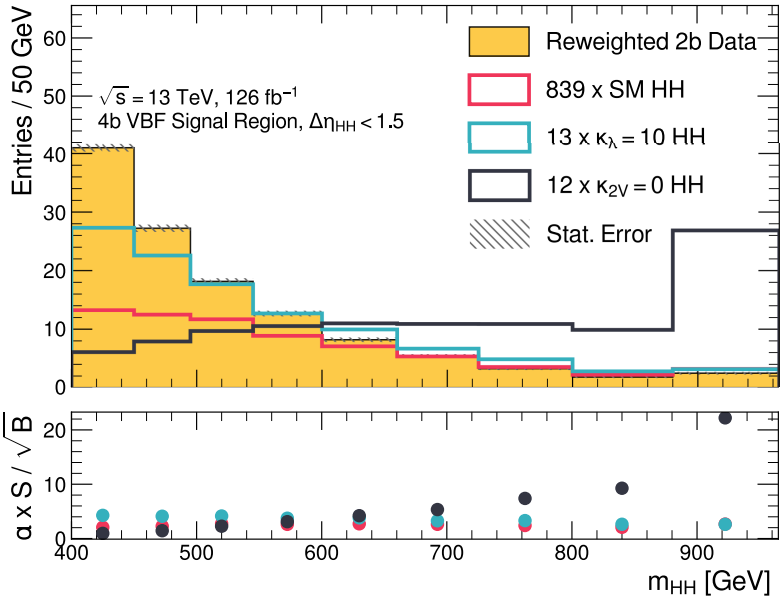
[3]The derivation and validation of the background estimate is described in Chapter 7, with the definition of `3b1f` included in Section 7.4.3.

[4]Prior to this change the lowest bin edge was 280 for the $|\Delta\eta_{HH}| < 1.5$ category and 290 for $|\Delta\eta_{HH}| > 1.5$.

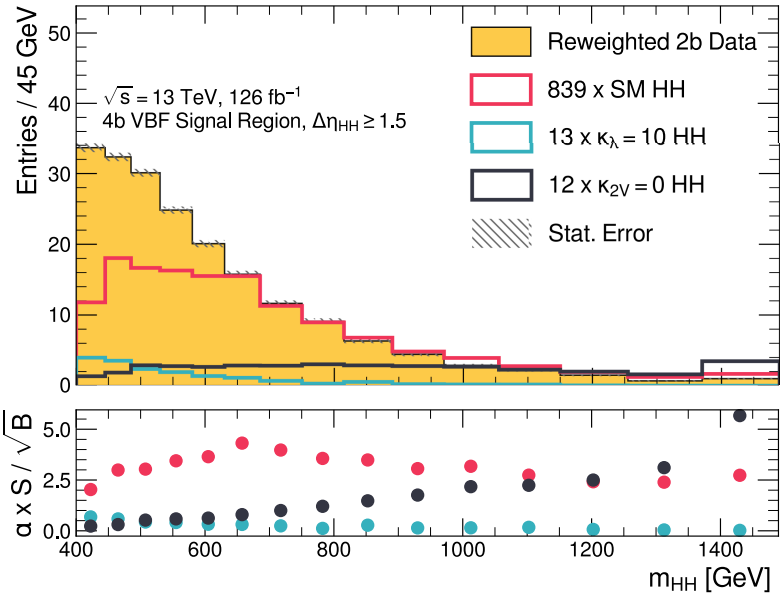
The m_{HH} distributions for each year vary due to the different trigger requirements. As a result, the background estimate in the ggF channel is derived exclusively for each year, so ggF channel events are further separated by data-taking year. The effect of the year-by-year differences is small in the VBF channel due to the larger statistical uncertainty, and the background estimate is run inclusive of year to maximise training statistics. So, in contrast, VBF channel events are not separated by year.

Figure 5.4 shows the m_{HH} distributions for VBF HH signal and background in the $|\Delta\eta_{HH}|$ categories in the VBF channel for all data-taking years. Sub-plots of $S/\sqrt{B}$ are included, and show how the $|\Delta\eta_{HH}| < 1.5$ category drives sensitivity to the non-SM VBF HH signal, whilst the $|\Delta\eta_{HH}| > 1.5$ category drives sensitivity to SM VBF HH signal.

Figure 5.5 shows the m_{HH} distributions for ggF HH signal and background in the X_{HH} and $|\Delta\eta_{HH}|$ categories for each data-taking year. They show that the sensitivity increases with decreasing X_{HH} and $|\Delta\eta_{HH}|$, with the $X_{HH} < 0.95 \times |\Delta\eta_{HH}| < 0.5$ category possessing the highest $S/\sqrt{B}$. This could be inferred from Figure 5.2. This is expected, given that, theoretically, HH signal should minimise X_{HH} and $|\Delta\eta_{HH}|$.

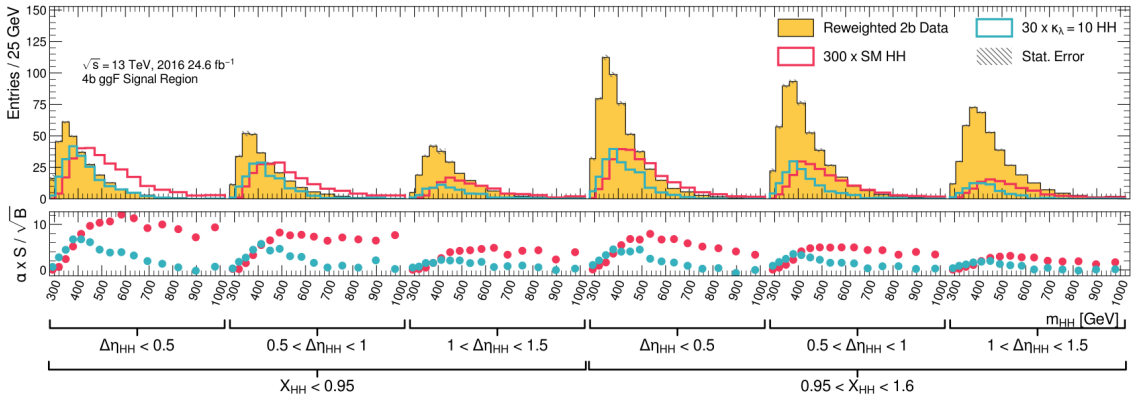


(a) $|\Delta\eta_{HH}|< 1.5$

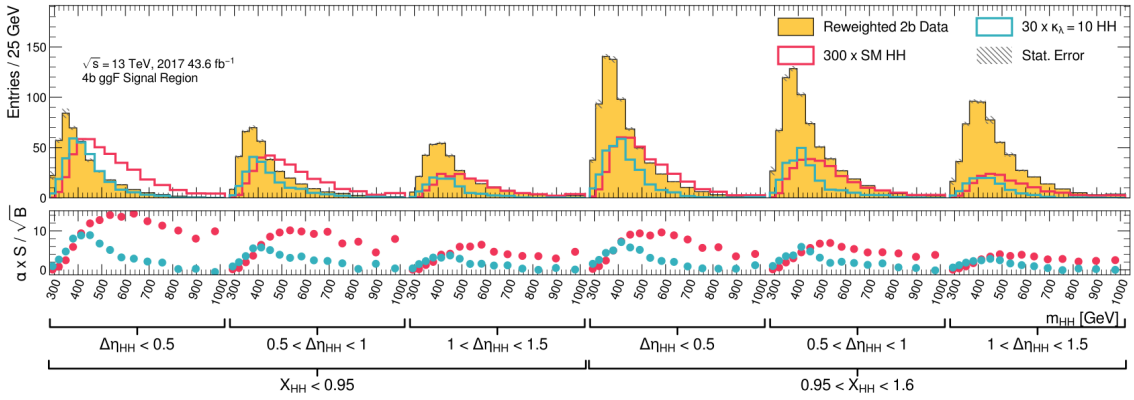


(b) $|\Delta\eta_{HH}|> 1.5$

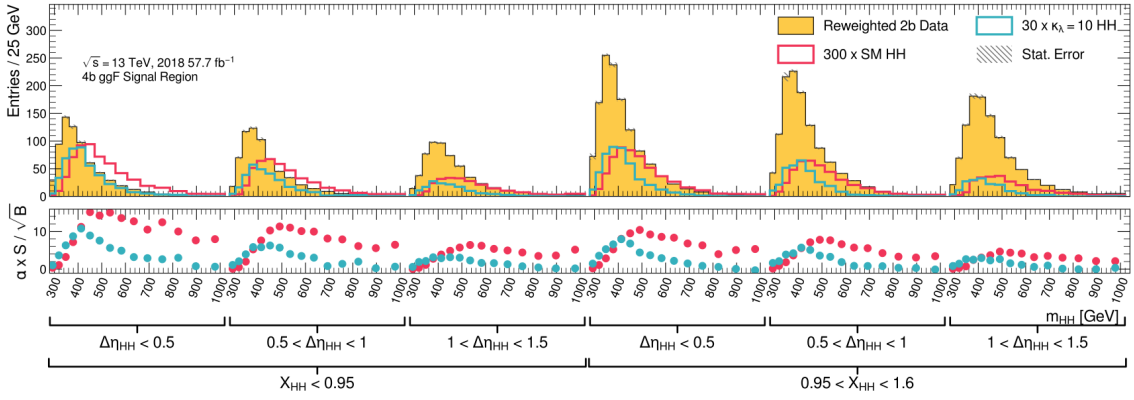
Figure 5.4: m_{HH} distributions for SM (pink), $\kappa_\lambda = 10$ (light blue) and $\kappa_{2V} = 0$ (dark blue) VBF HH signal and background (yellow, *RWed 2b data*) in the two $|\Delta\eta_{HH}|$ categories of the VBF channel. Sub-plots of $S/\sqrt{B}$ are included to show the relative sensitivity of each category to different VBF HH coupling points. The signals are scaled for visual purposes.



(a) MC16a/2016



(b) MC16d/2017



(c) MC16e/2018

Figure 5.5: Per-year m_{HH} distributions for SM (pink) and $\kappa_\lambda = 10$ (blue) ggF HH signal and background (yellow, *RWed 2b data*) in the X_{HH} and $|\Delta\eta_{HH}|$ categories in the ggF channel. Each plot simultaneously shows the six m_{HH} distributions for the six $|\Delta\eta_{HH}|$–X_{HH} ggF categories, with the labels below the x-axis indicating the category a particular m_{HH} distribution corresponds to. Sub-plots of $S/\sqrt{B}$ are included to show the relative sensitivity of each category to different ggF HH coupling points. The signals are scaled for visual purposes.

6 VBF Trigger Studies

6.1 Motivation

Historically, the $HH \rightarrow b\bar{b}b\bar{b}$ analysis targeted HH production via ggF, as the HH production process with the highest cross-section. The VBF channel was added later due to the unique sensitivity of VBF HH to κ_{2V} and was built on top of the ggF-optimised analysis. As such, it was expected that a number of the steps in the ggF-optimised selection would be non-optimal for VBF HH signal. For example, I observed a significant gap between the signal acceptance for SM ggF and VBF HH at the triggering stage – 73.8% compared to 47.7% respectively.[1] A number of triggers designed to target the VBF topology were introduced to the ATLAS Trigger Menu during Run 2 [94, 95, 96]. This chapter describes the studies I conducted to investigate whether analysis sensitivity to VBF HH signal could be improved by introducing these VBF-targeting triggers.

The first of these studies was conducted prior to the implementation and optimisation of the trigger bucketing strategy, which reduced the number of b-jet triggers used in the analysis to just `2b2j` and `2b1j` (see Section 4.3.1). At the time, the

[1] This was observed for the version of the analysis that existed in October 2019.

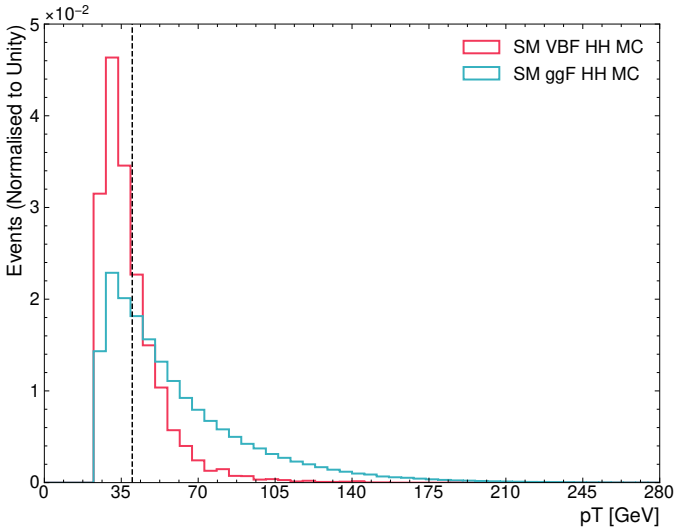


Figure 6.1: Pre-trigger p_T distributions for the lowest p_T Higgs candidate jet for SM ggF and VBF *HH* MC simulation (MC16a). The dashed black line outlines the jet $p_T > 40\,\text{GeV}$ cut applied as part of the analysis selection.

analysis used a suite of *b*-jet triggers, which keep events with a certain number of *b*-jets above a designated p_T threshold (see Table C.2 in Appendix C for the full list).

To better understand why the triggering step had a relatively low signal acceptance for SM VBF *HH*, the p_T distributions of the lowest p_T Higgs candidate (HC) jet in ggF and VBF *HH* MC simulation before the application of the trigger were compared. The results are shown in Figure 6.1. The distribution is far softer, i.e. of lower energy, for VBF *HH* than for ggF *HH*, meaning the VBF *HH* events are less likely to pass the jet p_T thresholds required by the *b*-jet triggers. This explains the difference in the signal acceptance for this step.

The analysis selection also requires that an event has four central jets with $p_T > 40\,\text{GeV}$ (Item (2) in Section 4.5). So, even if these softer events did pass the *b*-jet triggers, they would likely be removed by this cut. Any change to the triggering must be accompanied by a lowering of the p_T threshold for this cut to increase the

signal acceptance for VBF *HH*. This cut intentionally mirrors the thresholds of the `2b2j` trigger to ensure the analysis lies on the plateau of its turn-on curve.[2]

The choice of exploring the VBF triggers was well-motivated. Typically, VBF triggers reduce the event rate by targeting events that contain the two high-momentum well-separated jets characteristic of the VBF topology. Practically, this means they apply cuts to related variables, e.g. m_{jj}, and look for jets in a wide η range to allow for forward jets. As such, the requirements on the HC jet p_T is often less stringent than for *b*-jet triggers.

6.2 Impact on Signal Yields

The suite of VBF triggers investigated is shown in Table C.1 in Appendix C. These are all the VBF triggers in the Trigger Menu that do not have requirements designed for specific analyses, e.g. the trigger that required two taus for the $H \to \tau\tau$ analysis was excluded. There are different VBF triggers available for each year, with more triggers being available in later years, as the use of VBF triggers became more established.

In order to gauge how much additional signal the VBF triggers could provide, the percentage increase in the VBF *HH* signal yield when including the suite of tested VBF triggers on top of the nominal *b*-jet triggers was measured. This is summarised in Table 6.1. Only the analysis pre-selection, which requires four jets including two *b*-jets with $p_T > 25\,\text{GeV}$, is applied to these events. For 2016, the VBF triggers add little additional signal. This is to be expected, given there were only two VBF triggers used in the 2016 and both required the existence of a photon in the event.

[2] A *trigger turn-on curve* is a plot of the trigger efficiency versus the variable the trigger cuts on. Theoretically, this should be a step function, starting at zero efficiency and switching to 100% efficiency, i.e. *turning-on*, at the exact point the cut on the variable is passed. In practice, resolution effects broaden the turn-on, meaning it's a high-gradient slope rather than a vertical line.

Table 6.1: Percentage increase in pre-selection event yields for VBF *HH* samples when including the suite of tested VBF triggers.

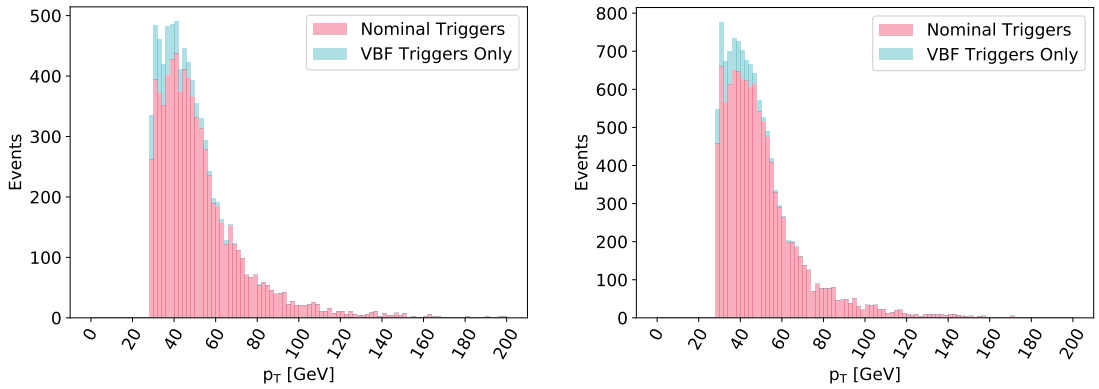
κ-value	Year	Increase in Yield [%]
SM	2016	1.07
	2017	26.0
	2018	20.7
$\kappa_{2V} = 0$	2016	0.54
	2017	14.1
	2018	12.2
$\kappa_{\lambda} = 10$	2016	0.97
	2017	23.3
	2018	18.2

For 2017 and 2018, the results are more promising, with the post-preselection VBF *HH* signal yield increasing by 12.2–26.0%, depending on the coupling value and year.

Figure 6.2 shows the distribution of the lowest p_T HC jet, with the additional signal from the VBF triggers highlighted in blue. This plot shows that the additional signal provided by the VBF triggers lies at low momentum. This indicates that hard VBF *HH* events are likely to trigger both the VBF and nominal *b*-jet triggers. Additionally, a large fraction of this additional signal would be removed by the analysis selection's aforementioned 40 GeV jet p_T cut.

6.3 Impact on Sensitivity

In order to assess how including the VBF triggers impacts analysis sensitivity, values of $S/\sqrt{B}$ for SM VBF *HH* signal after the full analysis selection were calculated for the 2017 data-taking year. Lower jet p_T cuts of 30 and 35 GeV were tested to increase the acceptance of the soft signal provided by the VBF triggers. When including the VBF triggers with the nominal p_T cut, $S/\sqrt{B}$ increases by 4.4%. However, lowering the p_T cut to 35 GeV reduced the increase to 2.0% and lowering to 30 GeV decreased



(a) MC16d (b) MC16e

Figure 6.2: Distributions of the lowest p_T Higgs candidate jet in SM VBF *HH* MC simulation, where events that pass the nominal set of *b*-jet triggers are highlighted in pink, alongside additional events that only passed the tested VBF triggers stacked on top in blue.

$S/\sqrt{B}$ by 4.4%. This is due to the introduction of additional background. Overall, the VBF triggers do introduce a significant amount of additional soft VBF *HH* signal over the *b*-jet triggers at the pre-selection stage. Unfortunately, the impact on sensitivity is minimal, a large fraction of the additional signal being removed by the analysis jet p_T cut, and simply reducing this cut threshold introduces a nullifying amount of background.

6.4 The VBF Trigger Category

The idea of introducing trigger categories into the analysis was explored as a strategy for keeping the additional signal from the VBF triggers whilst minimising the impact from the additional background. The proposal was to include two categories – one supplied by the nominal *b*-jet triggers and with the standard 40 GeV jet p_T cut, and the other supplied by VBF trigger(s) and with a reduced 25 GeV jet p_T cut.

Due to the complexity of calculating and applying trigger scale factors, it was decided that only a single trigger would be used for the VBF trigger category.

Table 6.2: Per-trigger yields and significance for $\kappa_{2V} = 0$ VBF HH signal (2017).

Trigger Short	S Yield	B Yield	$S/\sqrt{B}$
1b 2fwrdj	0.29	6.44	0.11
2b 1fwrdj	2.28	99.1	0.23
2b 2jMjj HT	3.08	62.9	0.39

Table 6.3: Per-trigger yields and significance for $\kappa_{\lambda} = 10$ VBF HH signal (2017).

Trigger Short	S Yield	B Yield	$S/\sqrt{B}$
1b 2fwrdj	0.39	6.44	0.16
2b 1fwrdj	3.19	99.1	0.32
2b 2jMjj HT	3.46	62.9	0.44

Table 6.4: Per-trigger yields and significance for $\kappa_{2V} = 0$ VBF HH signal (2018).

Trigger Short	S Yield	B Yield	$S/\sqrt{B}$
1b 2frwdj	0.493	13.2	0.14
2b 1frwdj	3.06	123	0.28
2b 2jMjj HT	4.00	83.9	0.44
2j 2jMjj 2b	2.75	79.9	0.31
2b 1frwdj 1j 1cntlj	3.06	138	0.26
2j Mjj+dPhijj+dEtajj	1.13	30.3	0.21

Table 6.5: Per-trigger yields and significance for $\kappa_{\lambda} = 10$ VBF HH signal (2018).

Trigger Short	S Yield	B Yield	$S/\sqrt{B}$
1b 2frwdj	0.477	13.2	0.13
2b 1frwdj	4.42	123	0.40
2b 2jMjj HT	5.36	83.9	0.59
2j 2jMjj 2b	3.81	79.9	0.43
2b 1frwdj 1j 1cntlj	4.54	138	0.39
2j Mjj+dPhijj+dEtajj	1.34	30.3	0.24

Table 6.6: Per-year comparison of the significance of VBF *HH* signal when using the nominal *b*-jet trigger category and when combining it with the proposed VBF trigger category. Z is a measure of statistical significance, defined in Equation 6.1.

κ-value	Year	Nominal Z	Combined Z	Gain in Z [%]
$\kappa_{2V} = 0$	2017	1.31	1.38	5%
	2018	1.05	1.32	26%
$\kappa_{\lambda} = 10$	2017	0.46	0.58	27%
	2018	0.62	0.79	27%

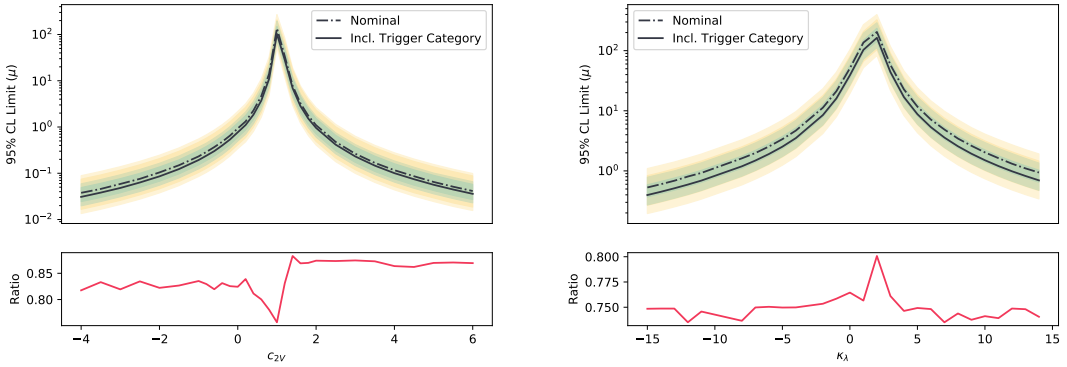
A comparison of the $S/\sqrt{B}$ for all VBF triggers tested is shown in Tables 6.2 – 6.5 for 2017/2018 and the $\kappa_{\lambda} = 10/\kappa_{2V} = 0$ coupling points.[3] The non-SM points were used as they are relevant for the analysis aim of setting limits on κ_{λ} and κ_{2V}. The trigger that maximised $S/\sqrt{B}$ in all instances was `2b 2jMjj HT`.[4] This trigger was used to supply the VBF trigger category. The key for the shortened trigger names is shown in Table C.3 in Appendix C.

Table 6.6 shows the gain in significance when combining the *b*-jet and VBF trigger categories. Events which pass both triggers are placed in the VBF trigger category, as it was found this produced a greater increase in significance than vice versa.

A new estimate of the significance, Z, defined in Equation 6.1, was used instead of $S/\sqrt{B}$ in Table 6.6 onward. This was following the recommendation of Ref. [97], where the authors compared a range of formulae for estimating the significance, including Equation 6.1 and $S/\sqrt{B}$, and found Equation 6.1 to be the best at approximating the measured significance in a toy model. This is not unexpected – $S/\sqrt{B}$ is derived by approximating the Poisson distribution of the number of events as a Gaussian; whereas, Z is more complex metric, derived by applying statistical methods from hypothesis testing to a Poisson-counting likelihood.

[3] HT is defined as the scalar sum of the transverse momentum, p_T, of a collection of physics objects.

[4] Full chain for 2017(2018): HLT_ht300_2j40_0eta490_invm700_L1HT150-J20s5.ETA31... _MJJ-400-CF_AND_2j25(35)_gsc45_bmv2c1070_split



(a) κ_{2V} (b) κ_λ

Figure 6.3: Comparison of the expected upper limits on signal strength, μ (= σ/σ_{theory}), for VBF *HH* signal when including the proposed VBF trigger category (solid line) and when not (dashed line). The one and two standard deviation uncertainty bands on the expected limits are shown in green and yellow respectively. The pink line on the sub-plots shows the ratio of the upper limit when including the VBF trigger category to when not including it.

$$\mathrm{Z} = \sqrt{2\left(n \ln\left[\frac{n}{b}\right] - (n-b)\right)} \tag{6.1}$$

Here, n is the number of combined signal and background events and b is the number of background events.

Additionally, as the relative shapes of the m_{HH} distributions of signal and background are important, Z was calculated for individual bins in m_{HH} and summed in quadrature.[5] A 26–27% gain in significance is observed for $\kappa_\lambda = 10$ VBF *HH* signal in both 2017 and 2018, and for $\kappa_{2V} = 0$ in 2018.

The expected upper limits on the signal strength, μ (= σ/σ_{theory}), at the 95% confidence limit were calculated when including the trigger categories and when not for VBF *HH* signal versus κ_λ and κ_{2V}. The results are displayed in Figure 6.3, and show ∼12.5–17.5% improvement for non-SM values of κ_{2V}, ∼20% for SM and ∼25%

[5] In this context, m_{HH} bins = [250, 300, 350, 400, 450, 500, 600, 750, 1500].

for non-SM κ_λ when including the trigger categories.

6.5 Defining a 2b 2jMjj HT Trigger Bucket

After the trigger bucket strategy was implemented in the $HH \rightarrow b\bar{b}b\bar{b}$ analysis, a trigger bucket for 2b 2jMjj HT had to be designed.[6] To build a trigger bucket, offline cuts mirroring the trigger requirements must be determined. Below is the full chain for 2b 2jMjj HT, with a description of its requirements:

Full chain: HLT_ht300_2j40_0eta490_invm700_L1HT150-J20s5.ETA31_... MJJ-400-CF_AND_2j35_gsc45_bmv2c1070_split

Req. 1: HT cut of 300 GeV at HLT and 150 GeV at L1.[7]

Req. 2: *Two jets with $p_T > 40$ GeV.*

Req. 3: M_{jj} cut of 700 GeV at HLT and 400 GeV at L1.

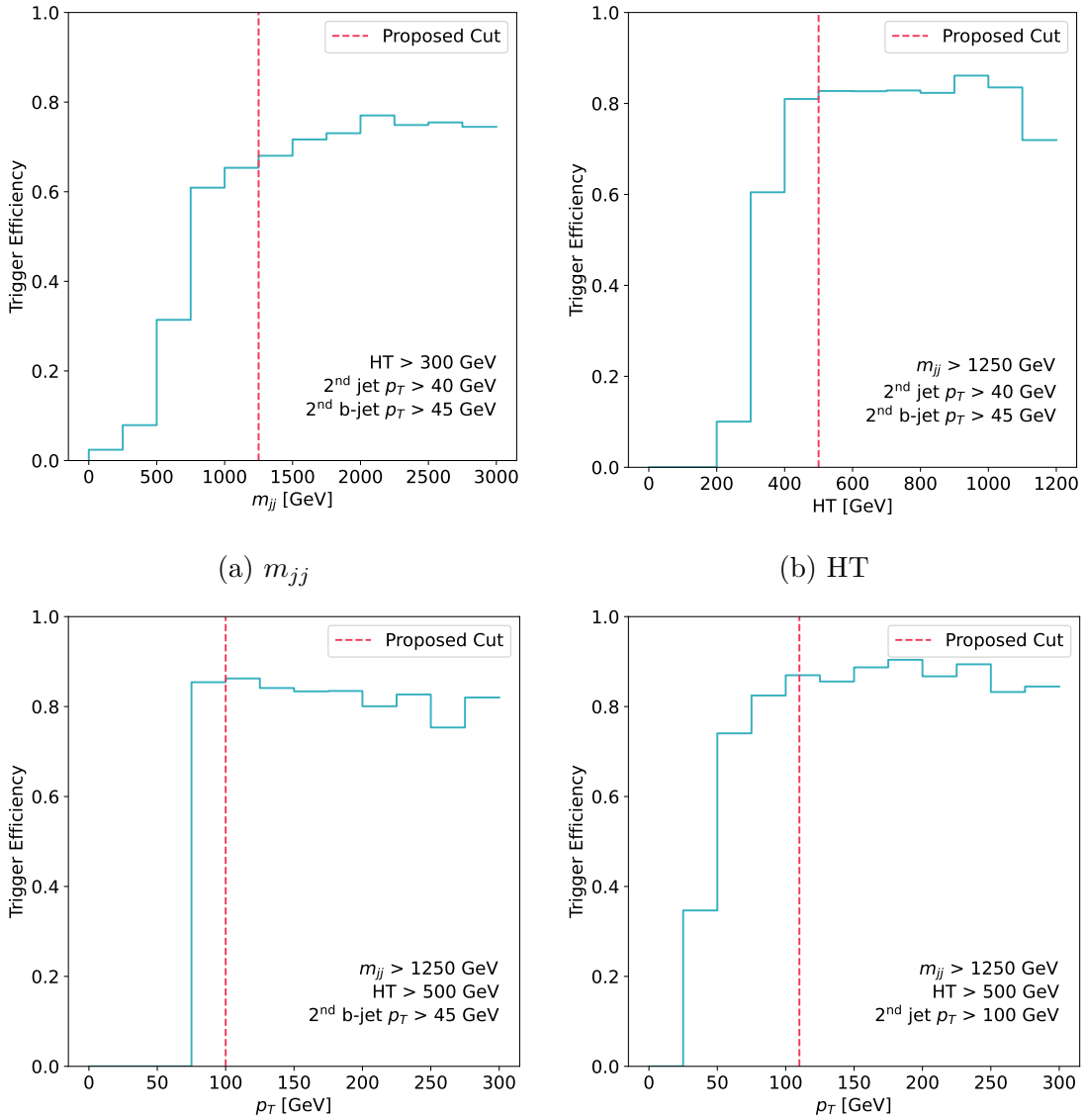
Req. 4: Two *b*-jets with $p_T > 45$ GeV.

For the 2b 2jMjj HT trigger bucket, these four requirements were applied as offline cuts on m_{jj}, HT, second-leading jet p_T and second-leading b-jet p_T. In order to determine the threshold for these offline cuts, 2b 2jMjj HT trigger efficiency curves in terms of these variables were generated. The aim was to place the cuts such that the analysis lies on the trigger turn-on plateau, which makes the derivation of trigger scale factors simpler. The trigger efficiency curves are shown in Figure 6.4.

Figure 6.4a shows the first trigger efficiency curve, which is in terms of m_{jj}. When calculating the trigger efficiency in terms of m_{jj}, cuts matching the HLT were placed on the other bucket variables. Once the bucket cut for m_{jj} was determined (1250 GeV), this, not the HLT value, was used when determining the bucket cut for

[6] See Section 4.3.1 for a description of the trigger bucket strategy.

[7] *L1* stands for Level-1, the first stage of the ATLAS trigger system, which prioritises speed over accuracy to reduce the high event rate. *HLT* stands for High Level Trigger, the second stage of the ATLAS trigger, which performs more complex, and therefore accurate, calculations.



(a) m_{jj}

(b) HT

(c) Second-leading jet p_T

(d) Second-leading b-jet p_T

Figure 6.4: Efficiency of the `2b 2jMjj HT` trigger with respect to the variables used for the offline cuts in the associated trigger bucket. The pink dashed line shows the position of the bucket offline cut.

Table 6.7: The variables used in the definition of the `2b 2jMjj HT` trigger bucket alongside the value of the cut on these variables applied by the HLT and by the bucket.

Variable	HLT Cut Value [GeV]	Bucket Cut Value [GeV]
m_{jj}	700	1250
HT	300	500
2nd jet p_T	40	100
2nd b-jet p_T	45	110

HT. As each new bucket cut was determined, it was applied when determining the bucket cut on the next variable. This can be seen when going from Figure 6.4a to 6.4d and looking at the cut values in bottom right corner. This ensures that events that, for example, fail the m_{jj} bucket cut are not used when determining the bucket cut on HT. Table 6.7 displays the HLT cuts stipulated in the `2b 2jMjj HT` trigger alongside their corresponding offline cuts for the `2b 2jMjj HT` bucket, and shows that the offline cuts required are significantly higher than the original HLT cuts.

With the `2b 2jMjj HT` trigger bucket implemented, the impact on sensitivity of using the *b*-jet and VBF trigger categories was reassessed. Table 6.8 shows the change in the VBF *HH* yield in the *four-tag* Signal Region when including the `2b 2jMjj HT` trigger bucket for the 2018 data-taking year. The increase in yield is negligible, in contrast to what was observed previously. It was observed that the majority of events in the `2b 2jMjj HT` trigger bucket lie at high m_{HH} values, not at low m_{HH}, where the additional signal provided by the VBF triggers is expected.

Although it is unfortunate that the promising increase in signal yield observed in previous studies was lost when implementing the trigger buckets, it is not surprising. In order to place the analysis on the plateau of the `2b 2jMjj HT` trigger turn-on, stringent cuts on m_{jj} HT and jet $\mathrm{p_T}$ were required. These cuts will favour harder events, removing the additional soft signal the VBF triggers provided.

Table 6.9 shows the change in significance, Z, when combining the *b*-jet and VBF trigger categories, with the full trigger bucketing strategy implemented, using 2018 MC simulation and data. Here, finer m_{HH} binning was used to better encode the shape information into Z.[8] The gain in significance varies from -3–10%, depending on the coupling point. Combining the *b*-jet and VBF trigger categories can lower

[8]In this context, m_{HH} bins = [250, 312.5, 375, 437.5, 500, 562.5, 625, 687.5, 750, 812.5, 875, 937.5, 1000, 1062.5, 1125, 1187.5, 1250, 1312.5, 1375, 1437.5, 1500]

the value of Z because events that pass both sets of triggers are placed in the VBF trigger category, and there can be events that fail the VBF bucket cut when they would pass the b-jet bucket cut.

Balancing the work required to determine the `2b 2jMjj HT` trigger scale factors and the modest increase in significance provided by the `2b 2jMjj HT` trigger category when included in the trigger bucket strategy, the proposal for trigger categories was not pursued for this iteration of the $HH \rightarrow b\bar{b}b\bar{b}$ analysis.

6.6 Summary

In the $HH \rightarrow b\bar{b}b\bar{b}$ analysis, the signal acceptance for the trigger selection step is significantly higher for SM ggF HH than for SM VBF HH. I pursued a series of studies to assess whether the inclusion of VBF triggers would increase VBF HH signal acceptance and general analysis sensitivity. A suite of VBF triggers were tested and found to provide an increase in the pre-selection signal yield of between 12.2–26.0% for a range of κ_λ and κ_{2V} coupling points in 2017 and 2018. The increase for 2016 was negligible. Studies of Higgs candidate jet p_T distributions showed that the majority of this additional signal was soft, i.e. low-energy, so would be removed by the analysis selection. Loosening the analysis selection provided no increase in sensitivity, as an equivalent number of new background events entered the analysis. The VBF trigger `2b 2jMjj HT` provided the maximum significance of all the VBF triggers tested, so was investigated further. A trigger categorisation was devised and assessed, with events being placed in either a category supplied by the `2b 2jMjj HT` trigger or the b-jet triggers nominally used by the analysis. Although early results showed large improvements to analysis sensitivity, the implementation of the trigger bucket strategy placed stringent cuts on events which passed the `2b 2jMjj HT` trigger, removing the additional soft signal it supplied. As a result, implementing VBF triggers, and a trigger categorisation, was not pursued in this analysis.

Table 6.8: Change in *four-tag* SR event yield for SM, $\kappa_\lambda = 10$ and $\kappa_{2V} = 0$ VBF *HH* signal points (MC16e) when including the `2b 2jMjj HT` trigger bucket and not.

κ-value	Incl. VBF Trigger Bucket	Signal Yield
SM	Yes	0.112
	No	0.114
$\kappa_{2V} = 0$	Yes	6.59
	No	6.54
$\kappa_\lambda = 10$	Yes	5.87
	No	5.89

Table 6.9: Comparison of the significance of VBF *HH* signal for the 2018 data-taking year when using the nominal *b*-jet trigger category and when combining it with the proposed VBF trigger category, with the implementation of trigger buckets. Z is a measure of statistical significance, defined in Equation 6.1. *Nominal Z* is the significance when just using the *b*-jet triggers, *Combination Z* is when using a combination of the *b*-jet and VBF trigger categories, and *Gain in Z* is the increase in significance when using the combination instead of the nominal.

κ-value	Year	Nominal Z	Combination Z	Gain in Z [%]
SM	2018	0.0108	0.0119	10%
$\kappa_{2V} = 0$	2018	2.13	2.33	9%
$\kappa_\lambda = 10$	2018	0.878	0.852	-3%

7

Background Estimation

The dominant source of background in the $HH \to b\bar{b}b\bar{b}$ analysis is multi-jet events initiated by underlying QCD, contributing ∼90% of the background. The remaining ∼10% is from $t\bar{t}$, which readily decay via $t \to bW$. Complex QCD processes are challenging to model accurately, and it is highly computationally expensive to generate sufficient QCD Monte Carlo simulation. Instead, this analysis builds the entirety of the background estimate, including the sub-dominant $t\bar{t}$ contribution, from data, weighting events from Control Region 1 (CR1) to generate the expected distributions of the background processes in the Signal Region (SR).[1]

There is one exception to the data-driven approach. When setting constraints on κ_{2V}, ggF HH is treated as a background process, and is modelled using ggF HH Monte Carlo simulation; however, ggF HH background was found to have negligible impact on analysis sensitivity, being many orders of magnitude smaller than the QCD background.

[1] See Section 4.10 for region definitions.

7.1 The *Two-Tag* Region

In the *four-tag* SR, QCD multijet processes, e.g. $2\times(g \to b\bar{b})$, are the dominant background to $HH \to b\bar{b}b\bar{b}$. In the *two-tag* SR, similar QCD multijet processes, e.g. $(g \to b\bar{b})(g \to q\bar{q})$, also dominate. However, the predicted signal contribution in the *two-tag* SR is negligible as the background yield is approximately two orders of magnitude higher in the *two-tag* SR than the *four-tag* SR. Due to the similarities between *two-tag* data and *four-tag* background, the *two-tag* SR provides the data for the data-driven background estimate.

7.2 Neural Network Reweighting

In spite of the similarity between *two-tag* and *four-tag* events, the differing b-jet requirements do result in differences in their kinematics, as shown later in Figures 7.2, 7.3 and 7.4. Weights are applied to the *two-tag* SR events to correct for any disparities. The challenge is to find a reweighting function, $w(\vec{x})$, of a set of kinematics variables, $\vec{x}$, that maps the *two-tag* probability density function (PDF), $p_{2b}(\vec{x})$, to the *four-tag* PDF, $p_{4b}(\vec{x})$, as shown in Equation 7.1.

$$p_{4b}(\vec{x}) = w(\vec{x}) \cdot p_{2b}(\vec{x}) \tag{7.1}$$

Equation 7.1 can be trivially rearranged to show $w(\vec{x})$ is the ratio of the two PDFs:

$$w(\vec{x}) = \frac{p_{4b}(\vec{x})}{p_{2b}(\vec{x})} \tag{7.2}$$

$w(\vec{x})$ can be approximated by training a Neural Network using a machine learning

technique known as *density ratio estimation* [98]. Following this method, the NN loss function is defined as shown in Equation 7.3.

$$\mathcal{L}(R(\vec{x})) = \mathbb{E}_{x\sim p_{2b}}\left[\sqrt{R(\vec{x})}\right] + \mathbb{E}_{x\sim p_{4b}}\left[\frac{1}{\sqrt{R(\vec{x})}}\right] \tag{7.3}$$

Here, $\mathbb{E}_{x\sim p_{2b}}[...]$ and $\mathbb{E}_{x\sim p_{4b}}[...]$ represent the expectation value from the *two-tag* and *four-tag* PDFs, and $R(\vec{x})$ is an estimator. As training proceeds, the NN will find the form of $R(\vec{x})$ that minimises the loss function. This definition of the loss function is chosen precisely because this minimum corresponds to an $R(\vec{x})$ that is the ratio of the PDFs:

$$\arg\min_{R} \mathcal{L}(R(\vec{x})) = \frac{p_{4b}(\vec{x})}{p_{2b}(\vec{x})} \tag{7.4}$$

Once trained, the NN will have learned $w(\vec{x})$ and will output an event-level weight when supplied with an event expressed in terms of $\vec{x}$. Technically, $\log R(\vec{x})$ is used to impose positivity, with the exponent taken at the end. The NN loss function is discussed in detail in Section 7.2.1 in Ref. [90], with the derivation of loss function minimiser shown in Equation 7.4 outlined in Appendix D in the same reference.

The NN derives the reweighting function using data from CR1. The procedure follows the *ABCD* method, i.e. deriving weights that map A $\rightarrow$ B and applying those weights to C such that C $\rightarrow$ D. A schematic for the application of this strategy in this context is shown in Figure 7.1.

CR1 is designed to provide the *two*-to-*four-tag* weights. The control regions are placed as close as possible to the SR in the m_{H1} vs. m_{H2} plane to ensure events in

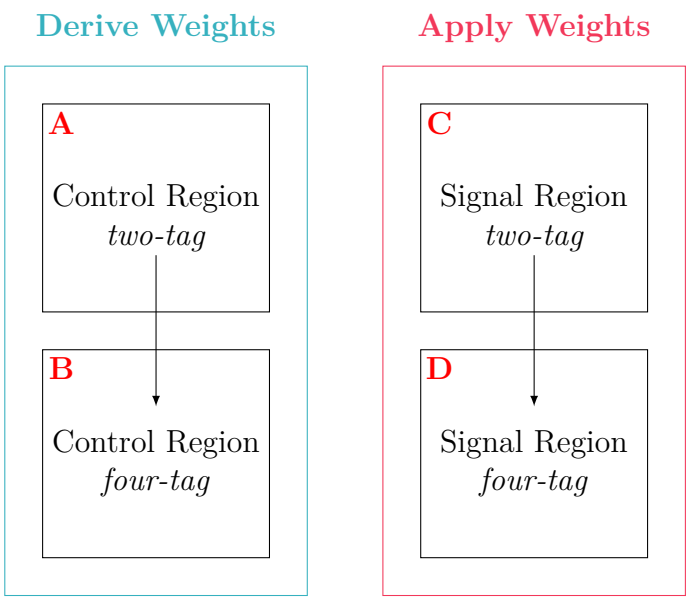


Figure 7.1: Schematic of the derivation and application of the *two*-to-*four-tag* background estimate weights using the ABCD method. The weights are extracted from the mapping of A to B and applied to C to map C to D.

these regions are kinematically similar. Additionally, there is minimal signal content in the control region, being away from the theoretical *HH* signal peak. As a result, the differences between the *two-tag* and *four-tag* events in CR1 are similar to those for background in the SR.

Being distinct topologies, the background estimate is derived separately for the ggF and VBF channels. Additionally, separate background estimates for each data-taking year are generated in the ggF channel. This is because the triggers vary for each year, and, as a result, so may the m_{HH} distribution. The background estimate for the VBF channel is derived using a single data set containing all the 2016–18 data, the trigger-induced differences being less impactful in this statistics-limited channel, and a large number of events being required for training the NN. The year is included as a one-hot encoded variable in the inputs to the NN.[2]

[2]One-hot encoding is method of packaging category variables for input to a Neural Network. For n category variables, each variable is represented by a different basis vector in an n-dimensional space. For example, the three years 2016, 2017 and 2018 are input to the Neural Network as [1, 0, 0], [0, 1, 0] and [0, 0, 1]. This prevents the Neural Network from using the ordered relationship between category variables, e.g. $2016 < 2017 < 2018$, which is not pertinent information.

Due to the yield in the ggF channel being far larger than the VBF channel, the architecture of the NN used in each differs. In the ggF channel, the NN is composed of three layers of 50 nodes, each with a Rectified Linear Unit (ReLU) activation function. For the VBF channel, the number of nodes per layer was reduced to 20, a smaller network being required to prevent under-training. The inputs are the values of relevant variables associated with an event, and the output is a weight, which is then placed on the *two-tag* SR event when building the data-driven background estimate. The list of input variables in the ggF and VBF channel also differs. These are shown in Table 7.1. The nominal NN showed good performance in modelling both m_{HH} and other event variables when compared with Neural Networks with different architectures but similar size.

The NN input variables for the VBF channel were optimised after mismodeling of the m_{HH} distribution was observed when using the ggF channel set. Variables were first assessed in terms of their correlation with m_{HH}, then in terms of their correlation with each other. Those with high correlation to m_{HH} but low correlation to one another were chosen. m_{HH} is the final discriminating variable, so it is important to include variables with significant influence on m_{HH}. Additionally, it is important model different aspects of the event to better model different influences on m_{HH}, and picking minimally correlated variables facilitates this. This optimised list of input variables improved the m_{HH} modelling in the VBF channel.

Two prominent sources of uncertainty were identified in the training of the NN. First, the statistics of the training dataset, and, second, the initial values of the node-to-node weights in the NN.[3] To account for these uncertainties, the *Bootstrapping* methodology is applied [99, 100].

[3]Neural Networks are composed of layers of nodes, where the nodes in consecutive layers are connected to one another. Along this connection is a weight. The process of training a Neural Network is passing data through the Network and updating these weights steadily. Initially, these weights are given a random value. The initial values of these weights can influence the training of the Network.

Table 7.1: List of input variables to the Neural Network when deriving the *two*-to-*four*-tag weights for the ggF and VBF channels. *HC* stands for Higgs candidate, and X_{Wt} is defined in Eq. 4.1.

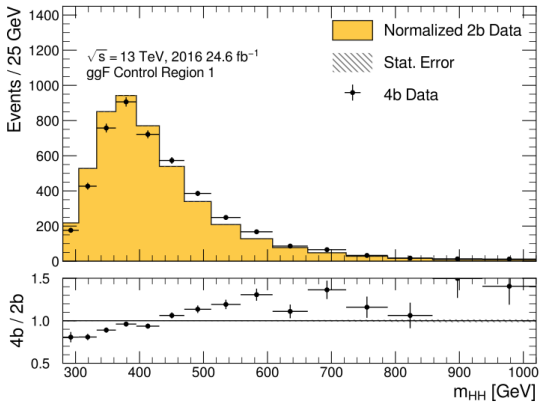
ggF channel	**VBF channel**
1. $\log(\mathrm{p_T})$ of the 4th leading HC jet.	1. Maximum m_{jj} out of the four HC jets.
2. $\log(\mathrm{p_T})$ of the 2nd leading HC jet.	2. Minimum m_{jj} out of the four HC jets.
3. $\log(\Delta R)$ between the closest two HC jets.	3. Energy of the leading HC.
4. $\log(\Delta R)$ between the other two HC jets.	4. Energy of the sub-leading HC.
5. Absolute value of the average HC jet η.	5. 2nd smallest ΔR in minΔR *HH* pairing method (see Sub-sec. 4.8.1).
6. $\log(\mathrm{p_T})$ of the *HH* system.	6. $\log(X_{Wt})$.
7. ΔR between the two HCs.	7. Absolute value of the average HC jet η.
8. $\Delta\phi$ between the jets in the leading HC.	8. Trigger bucket as one-hot encoded variable.
9. $\Delta\phi$ between the jets in the subleading HC.	9. Year as a one-hot encoded variable.
10. $\log(X_{Wt})$.	
11. Number of jets in the event.	
12. Trigger bucket as a one-hot encoded variable.	

Bootstrapping is used to approximate measurements of accuracy, e.g. variance, of an estimator. The basis of bootstrapping is the *bootstrap sample* (or simply *bootstrap*). For a dataset $\vec{x}$ with n entries, a bootstrap sample $\vec{x}^*$ is constructed by randomly sampling n times from $\vec{x}$ with replacement . Therefore, $\vec{x}^*$ is the same size as $\vec{x}$, but contains each element of $\vec{x}$ a random number of times (including zero). Example bootstrap samples of $\vec{x} = [1, 2, 3]$ are $\vec{x}^*_1 = [1, 3, 3], \vec{x}^*_2 = [2, 2, 2], \vec{x}^*_3 = [2, 1, 3]$, etc. To approximate the variance of an estimator $\hat{\theta}(\vec{x})$, m bootstrap samples $\{\vec{x}^*_1 ... \ \vec{x}^*_m\}$ can be generated and, from each, a series of equivalent estimators $\{\hat{\theta}(\vec{x}^*_1) ... \ \hat{\theta}(\vec{x}^*_m)\}$ calculated. Then, the estimate of the variance of the nominal estimator, $\hat{\theta}(\vec{x})$, can then be taken as the variance of the distribution of bootstrap sample estimators, $\{\hat{\theta}(\vec{x}^*_1) ... \hat{\theta}(\vec{x}^*_m)\}$. It can be shown that as the number of bootstrap samples approaches infinity, i.e. $m \to \infty$, the variance of the bootstrap sample estimators, $\{\hat{\theta}(\vec{x}^*_1) ... \hat{\theta}(\vec{x}^*_m)\}$, approaches the population variance of the nominal estimator, $\hat{\theta}(\vec{x})$.[4] An infinite number of bootstrap samples cannot be generated; however, accurate approximations of the variance of estimators are normally achievable when using between 50 to 200 bootstrap samples [100].

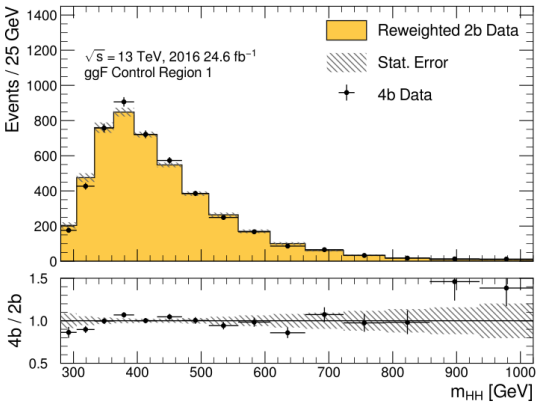
The exact bootstrapping procedure used to estimate the uncertainty due to the statistics of the training dataset and the initial values of the node-to-node weights in the NN will now be described. 100 Neural Networks are trained, each with a randomised set of initial weights, and using a different bootstrap of the data from the CR1. From the 100 trainings, 100 histograms of the background estimate are constructed. Therefore, there are 100 potential values for the yield in each bin for the background estimate. The nominal background estimate is built from the mean of the bin yields, whilst the Inter-Quantile-Range (68%) in each bin is taken as the uncertainty, referred to as the *bootstrap* uncertainty. Further systematic uncertainties associated with the background estimate are discussed in Section 10.2.4.

[4]Here, the "population" is the original dataset, $\vec{x}$.

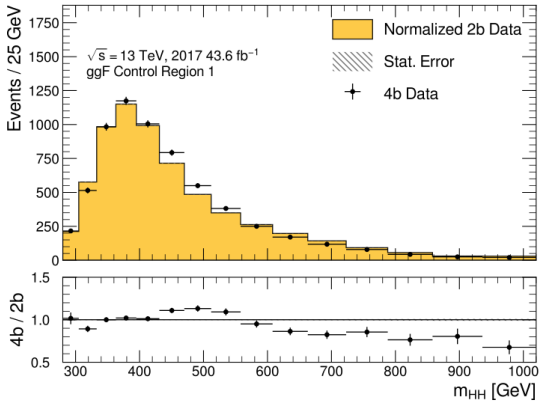
Figure 7.2 shows the m_{HH} distributions for *two-tag* data before and after application of the *two*-to-*four-tag* weights, alongside the *four-tag* data, in CR1 in the ggF channel. There is visible improvement for all three years, with good closure observed between reweighted-*two-tag* and *four-tag* data in all cases. Figure 7.3 shows the equivalent m_{HH} plots in the VBF channel. The impact of the reweighting is smaller in the VBF channel; however, a good level of closure between reweighted *two-tag* and *four-tag* data is observed. The corrective impact of the reweighting in the VBF channel is more visible in other analysis variables beside from m_{HH}. For example, the reweighted *two-tag* X_{Wt} distribution shows a significantly higher degree of closure with the *four-tag* distribution after reweighting, as seen in Figure 7.4. Overall, the agreement between the *two-tag* and *four-tag* distributions, as measured using the χ^2 goodness-of-fit metric, consistently improved across multiple variables in both channels after the application of the *two*-to-*four*-tag weights.



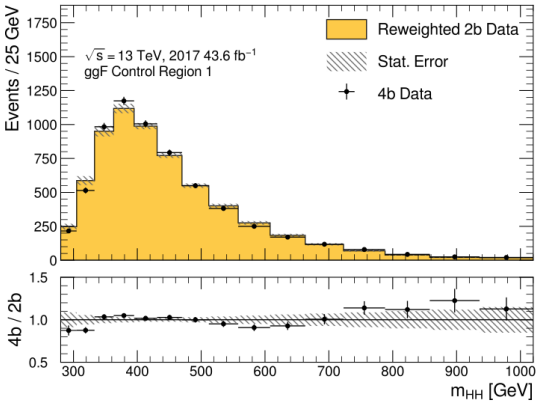
(a) Pre-reweighting, 2016 data.



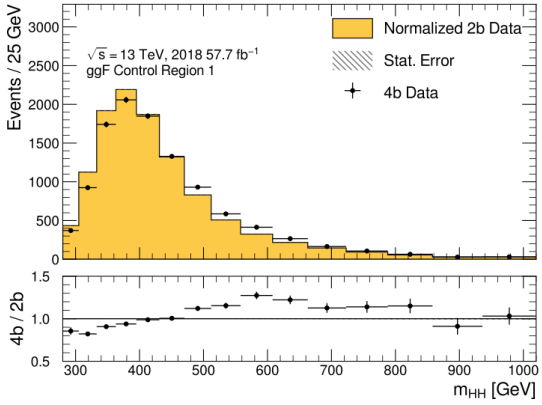
(b) Post-reweighting, 2016 data.



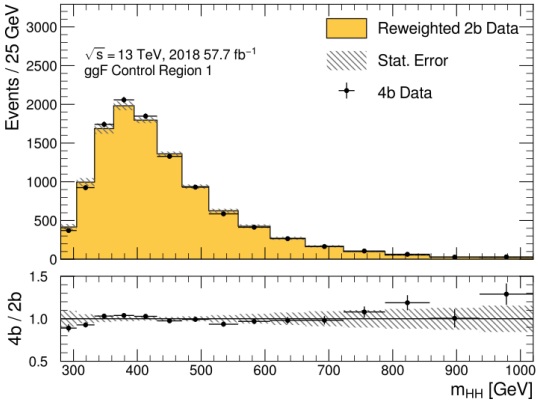
(c) Pre-reweighting, 2017 data.



(d) Post-reweighting, 2017 data.

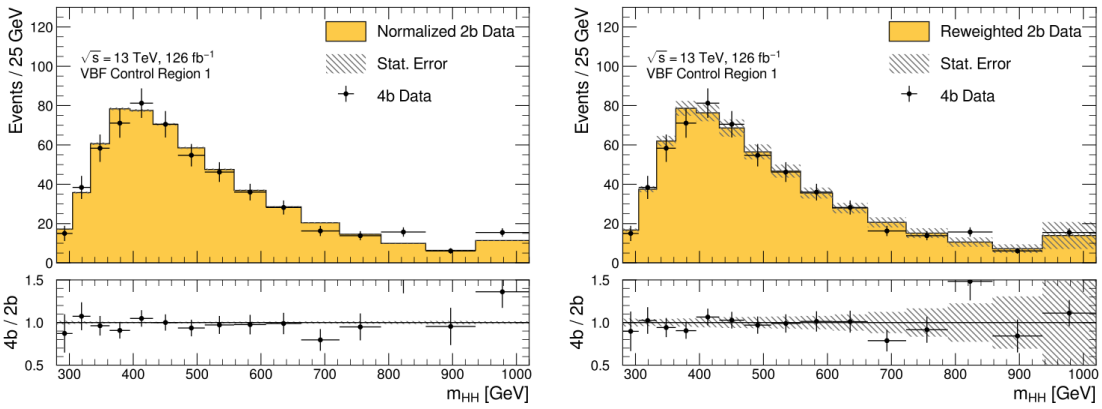


(e) Pre-reweighting, 2018 data.



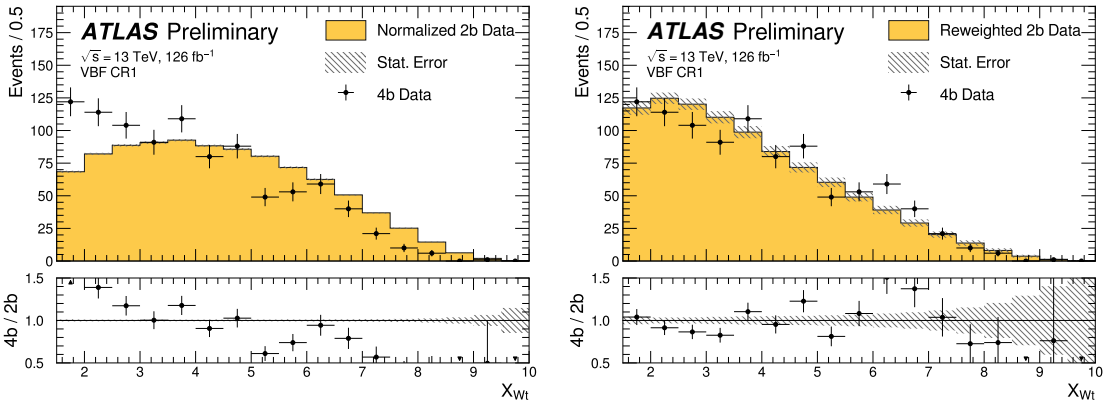
(f) Post-reweighting, 2018 data.

Figure 7.2: Per-year m_{HH} distributions for *two-tag* data before and after application of the *two*-to-*four-tag* weights, alongside the *four-tag* data, in CR1 in the ggF channel. The statistical uncertainty band in the post-reweighting plot also includes the bootstrap uncertainty.



(a) Pre-reweighting (b) Post-reweighting

Figure 7.3: m_{HH} distributions for *two-tag* data before and after application of the *two*-to-*four-tag* weights, alongside the *four-tag* data, in CR1 in the VBF channel for all data-taking years. The statistical uncertainty band in the post-reweighting plot also includes the bootstrap uncertainty.



(a) Pre-reweighting (b) Post-reweighting

Figure 7.4: X_{Wt} distributions for *two-tag* data before and after application of the *two*-to-*four-tag* weights, alongside the *four-tag* data, in CR1 in the VBF channel for all data-taking years. The statistical uncertainty band in the post-reweighting plot also includes the bootstrap uncertainty.

7.3 Background Yields

Tables 7.2 and 7.3 show the background estimate yields in the *four-tag* SR of the ggF and VBF channels respectively. The yield is far smaller in the VBF channel than the ggF channel due to the tight cuts used to target the VBF topology.

Table 7.2: The background estimate event yields in the *four-tag* Signal Region of the ggF channel for 2016–18 data-taking years.

Year	Background Yield
2016	3211
2017	4493
2018	7720

Table 7.3: The background estimate event yields in the *four-tag* Signal Region of the VBF channel for all data-taking years.

Year	Background Yield
2016–18	354.6

7.4 Background Validation

Although similar, there are differences between events in CR1 and the SR. As such, it is important to validate the method of applying weights in the SR that were derived in CR1. The background estimation strategy is validated using three validation regions. These are altered versions of the analysis, such that the final analysis regions are similar but orthogonal to the nominal choice. The three validation regions are defined in Table 7.4. Additionally, $t\bar{t}$ and QCD dijet Monte Carlo (MC) simulation samples were run through the nominal background estimation methodology as a further validation measure.

All validation regions possess the same benefit: analysers can look in the *four-tag* SR. As a result, it can be confirmed whether the reweighted *two-tag* SR distributions are in agreement with those in the *four-tag* SR. Analysers cannot look in the nominal *four-tag* SR because this region is blinded initially and could contain a significant amount of *HH* signal.

Table 7.4: Names and descriptions of the validation regions. All other steps in the analysis selection are kept the same unless stated otherwise.

Name	Description
`3b1f`	Four b-jet (*four-tag*) region replaced with `3b1f`. `3b1f` requires three b-jets tagged at the 77% WP and none at any looser WP.
`rev`$\lvert\Delta\eta_{HH}\rvert$	$\lvert\Delta\eta_{HH}\rvert > 1.5$ required instead of $\lvert\Delta\eta_{HH}\rvert < 1.5$. (Not used for VBF.)
Shifted Regions	Centre of SR/CR1/CR2 changed to be elsewhere in the m_{H1} vs. m_{H2} plane.

7.4.1 Reversed $|\Delta\eta_{HH}|$ (rev$|\Delta\eta_{HH}|$)

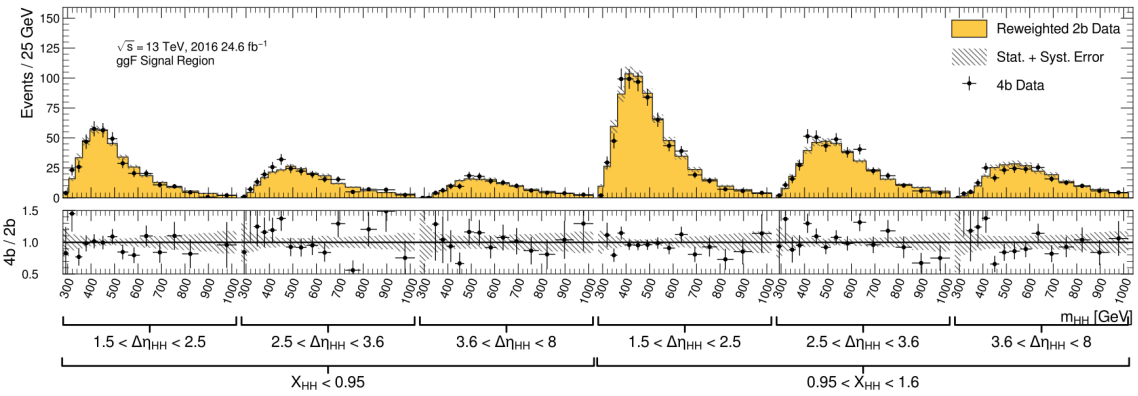
The $|\Delta\eta_{HH}| < 1.5$ cut in the nominal analysis selection is highly efficient at isolating *HH* signal and rejecting QCD background. The rev$|\Delta\eta_{HH}|$ validation region reverses the cut, i.e. requiring $|\Delta\eta_{HH}| > 1.5$, and, as a result, is high in QCD background and low in *HH* signal. All other analysis steps are kept identical, so the kinematics in the rev$|\Delta\eta_{HH}|$ region are similar to those in the nominal set. The VBF background estimate cannot be validated in the rev$|\Delta\eta_{HH}|$ region, as the $|\Delta\eta_{HH}| < 1.5$ cut is not applied in this channel. Only the ggF channel background estimate is validated in rev$|\Delta\eta_{HH}|$.

Events in the rev$|\Delta\eta_{HH}|$ region are divided into $|\Delta\eta_{HH}|$–X_{HH} categories to allow the background modelling in each category to be assessed. The nominal set of $|\Delta\eta_{HH}|$ boundaries cannot be used due to the inversion of the $|\Delta\eta_{HH}|$ cut. Instead, a set of new boundaries, which separate events into bins of approximately equal statistics, is used. This strategy matches that used to set the nominal $|\Delta\eta_{HH}|$ boundaries. The rev$|\Delta\eta_{HH}|$ category bins are:

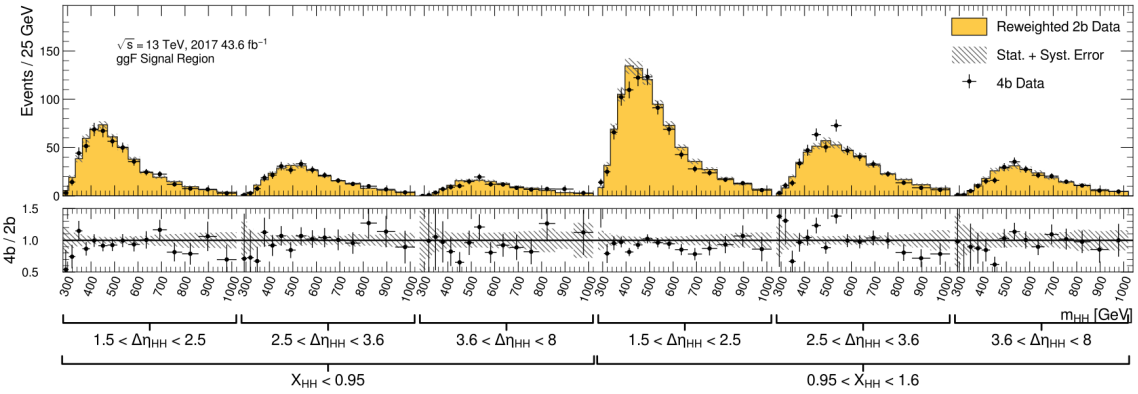
$$1.5 < |\Delta\eta_{HH}| < 2.5, \quad 2.5 < |\Delta\eta_{HH}| < 3.6, \quad 3.6 < |\Delta\eta_{HH}| < 8.0.$$

The nominal X_{HH} categorisation ($0.0 < X_{HH} < 0.95, 0.95 < X_{HH} < 1.6$) is kept.

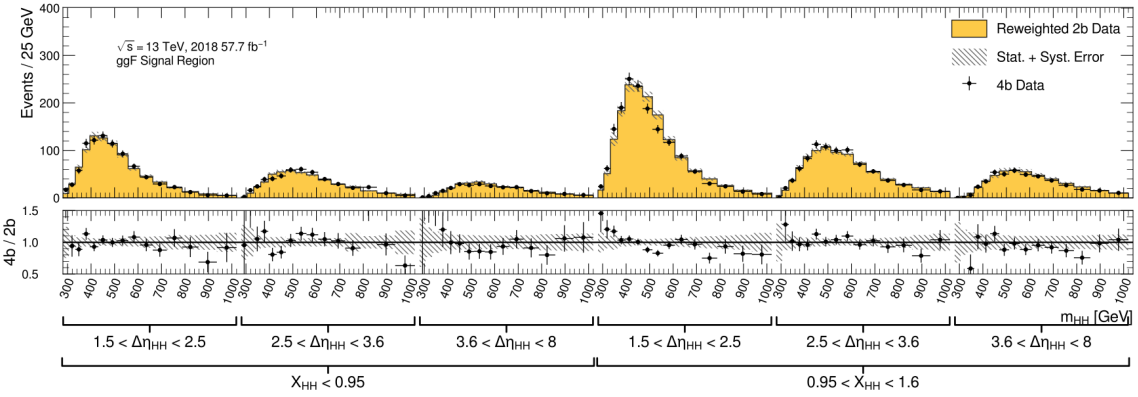
The background estimation procedure was performed in the ggF channel of the rev$|\Delta\eta_{HH}|$ region. Figure 7.5 shows a comparison of the m_{HH} distributions for reweighted-*two-tag* and *four-tag* data in the rev$|\Delta\eta_{HH}|$ Signal Region. The reweighted-*two-tag* distributions show good agreement with *four-tag* across categories and years. The vast majority of points lie within each others' uncertainty bands. This result provides confidence in the accuracy of the background estimation strategy.



(a) 2016



(b) 2017



(c) 2018

Figure 7.5: The per-year m_{HH} distributions for *four-tag* (black points) and reweighted *two-tag* (yellow) data in the Signal Region of the reversed $|\Delta\eta_{HH}|$ alternative region in the ggF channel. Each plot simultaneously shows the six m_{HH} distributions for the six $|\Delta\eta_{HH}|$-X_{HH} ggF categories, with the labels below the x-axis indicating the category a particular m_{HH} distribution corresponds to. The smaller distribution at the bottom of each plot shows the ratio of *four-tag* to reweighted *two-tag* data.

Additionally, a background-only maximum likelihood estimator (MLE) fit of the likelihood model to data was performed in the *four-tag* SR of the `rev`$|\Delta\eta_{HH}|$ region. The pulls of all nuisance parameters (NP) were within the acceptable bounds and were not observed to be overconstrained.[5]

[5]Chapter 9 has a description of the statistical procedure and Section 9.6 has an explanation of pulls.

7.4.2 Shifted Regions

The *Shifted Regions* are a collection of five validation regions, where the position of the SR/CR1/CR2 are located at a different point in the m_{H1} vs. m_{H2} plane, close in proximity to the nominal SR, but not overlapping it. All other steps in the analysis are kept the same. The Shifted Regions are shown in Figure 7.6, with the region centres listed in Table 7.5. No region is placed in the lower left of the m_{H1} vs. m_{H2} plane, as this is an area where the background distribution is distinct.

The SR of each Shifted Region is defined using Equation 7.5, which is the same as the nominal definition, X_{HH} (Equation 4.2), but with the central values, 124 and 117 GeV, and the resolutions, 0.1 and 0.1, replaced by input variables $m_{H1,\text{ Centre}}$ and $m_{H2,\text{ Centre}}$, and $\sigma_{m_{H1}}$ and $\sigma_{m_{H2}}$, respectively. The definitions of the resolutions are shown in Equation 7.6. Similarly, the control regions are defined using the same equations as the nominal definition (Equations 4.4 and 4.5), but with the central values, 124 and 117 GeV, also replaced by $m_{H1,\text{ Centre}}$ and $m_{H2,\text{ Centre}}$. By altering the region centre and resolution, the Shifted Regions possess a similar size and shape to the nominal regions, but are located at a different point.

$$X_{HH,\text{ Shifted}} = \sqrt{\left(\frac{m_{H1} - m_{H1,\text{ Centre}}}{\sigma_{m_{H1}} \times m_{H1}}\right)^2 + \left(\frac{m_{H2} - m_{H2,\text{ Centre}}}{\sigma_{m_{H2}} \times m_{H2}}\right)^2} \tag{7.5}$$

$$\sigma_{m_{H1}} = 0.1 \times \frac{124}{m_{H1,\text{ Centre}}}, \quad \sigma_{m_{H2}} = 0.1 \times \frac{117}{m_{H2,\text{ Centre}}} \tag{7.6}$$

The background estimation procedure was performed in the each of the five Shifted Regions for both the ggF and VBF channels, and the agreement between the reweighted-*two-tag* and *four-tag* data in the SR compared. The reweighted-*two-*

tag and *four-tag* data yields were observed to be in good agreement across all years, all being within two standard deviations of one another, and a large majority being within one standard deviation. Additionally, the m_{HH} distributions are also in good agreement, with the vast majority of *four-tag* data points lying within the uncertainty band of the reweighted *two-tag* distribution. The distributions for 2018 data in the ggF channel are shown as an example in Figure 7.7, and the distributions for all data-taking years in the VBF channel are shown in Figure 7.8. These results are in agreement with the `rev`$|\Delta\eta_{HH}|$ validation studies, providing additional confidence in the background estimation strategy.

A background-only maximum likelihood estimator (MLE) fit of the likelihood model to data was performed in the *four-tag* SR of the ggF channel in the Shifted Regions. The pulls of all nuisance parameters (NP) were within the acceptable bounds and were not observed to be overconstrained.[6]

[6]See footnote 5.

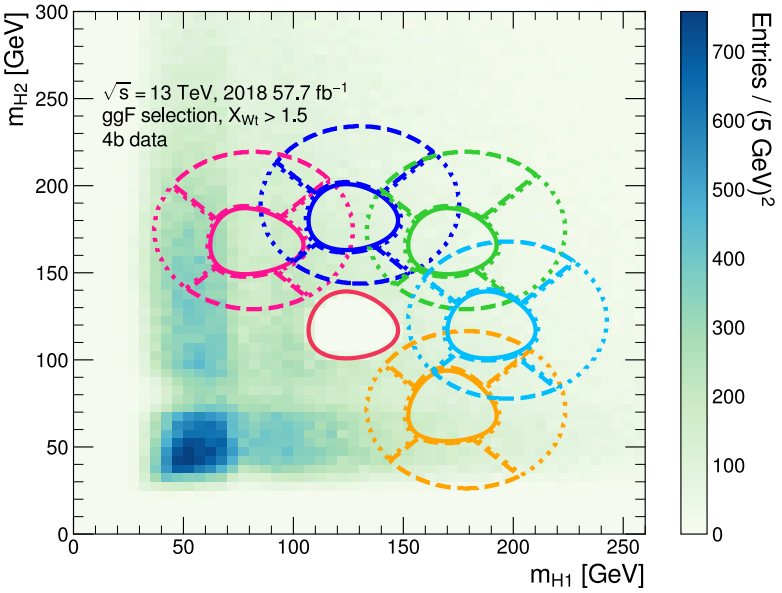
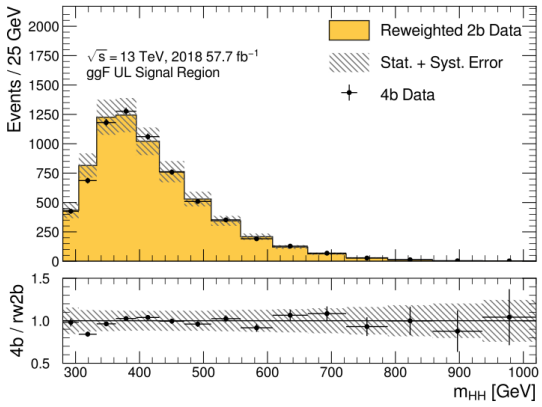


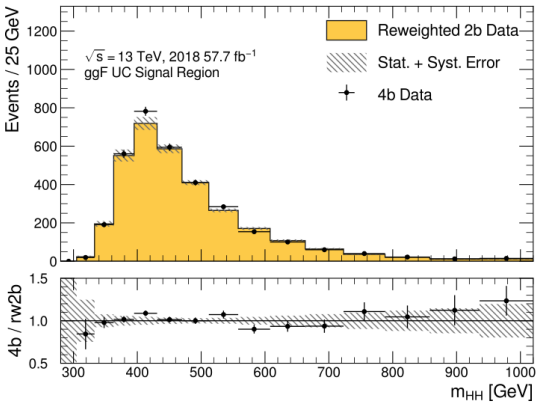
Figure 7.6: Position of the five Shifted Regions in the m_{H1} vs. m_{H2} plane, each shown in a different colour, with the true Signal Region shown with the solid pink line and white interior.

Table 7.5: Centres of the five Shifted Regions in the m_{H1} vs. m_{H2} plane. Each Shifted Region is named after its relative position in the m_{H1} vs. m_{H2} plane

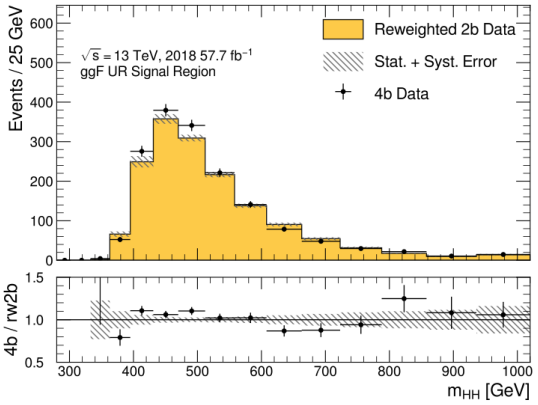
Shifted region name	**Region centre**	
	$\boldsymbol{m_{H1}}$ **[GeV]**	$\boldsymbol{m_{H2}}$ **[GeV]**
Upper Left	78	166
Upper Centre	124	180
Upper Right	170	160
Centre Right	188	117
Lower Right	170	68



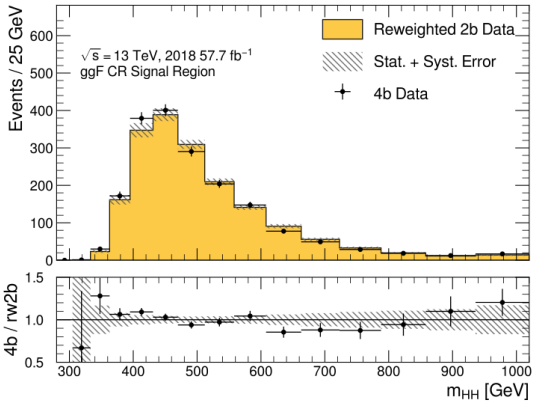
(a) Upper Left.



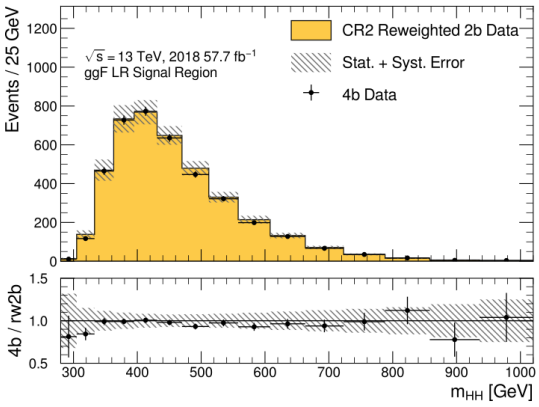
(b) Upper Centre.



(c) Upper Right.

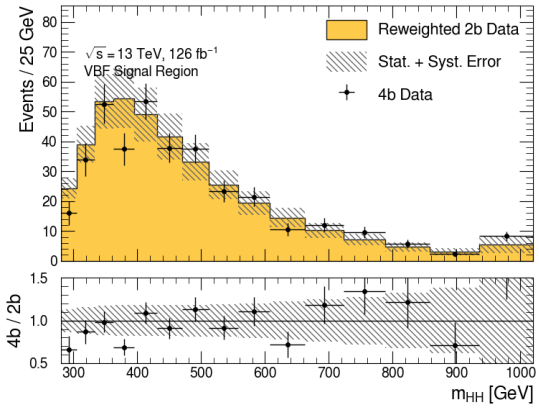


(d) Centre Right.

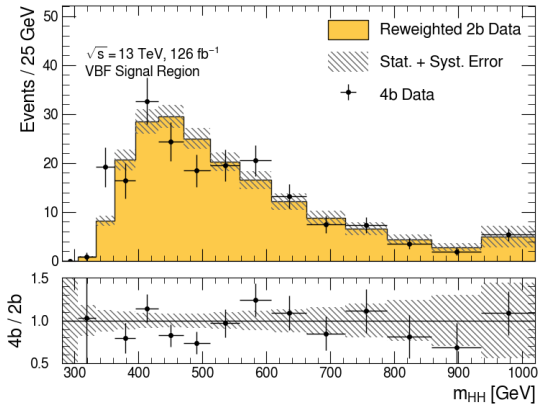


(e) Lower Right.

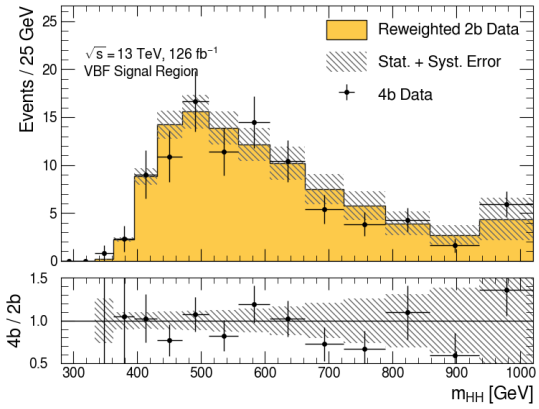
Figure 7.7: m_{HH} distributions for reweighted-*two-tag* and *four-tag* data in the SR of the ggF channel of the five Shifted Regions for 2018 data.



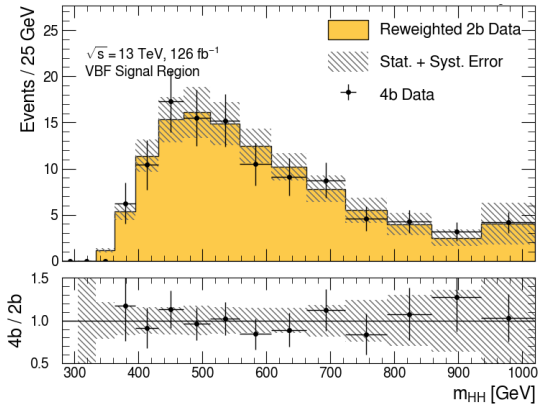
(a) Upper Left.



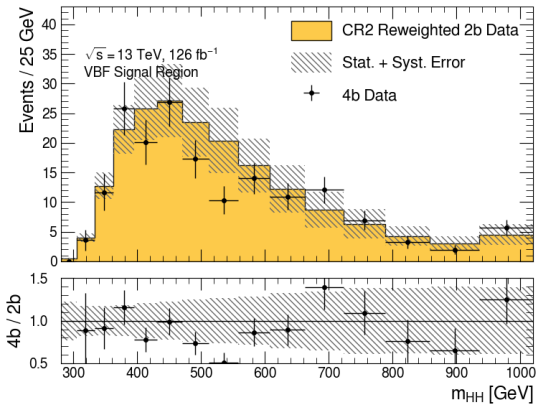
(b) Upper Centre.



(c) Upper Right.



(d) Centre Right.



(e) Lower Right.

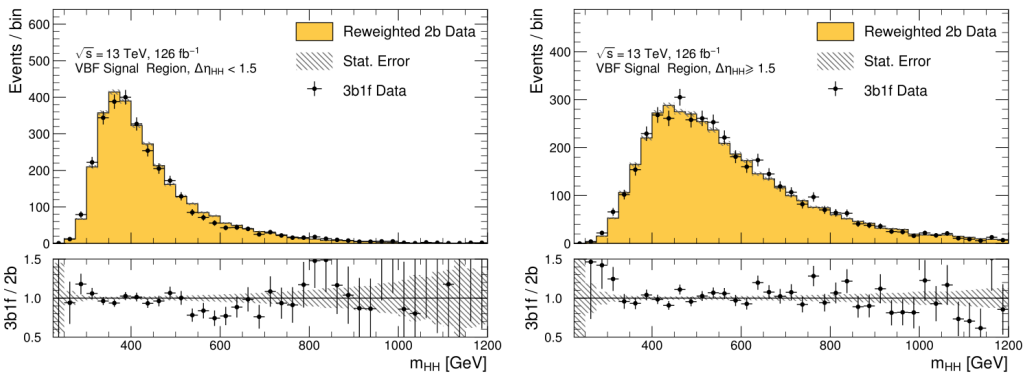
Figure 7.8: m_{HH} distributions for reweighted-*two-tag* and *four-tag* data in the SR of the VBF channel of the five Shifted Regions for all data-taking years.

7.4.3 `3b1f`

In the `3b1f` validation region, the *four-tag* requirement is replaced with the provision that events possess only three b-jets at the nominal 77% WP and none at any looser WP. These events are normally discarded by the nominal analysis selection. All other parts of the analysis set-up are kept the same.

The background estimation procedure was performed, replacing the *four-tag* requirement with `3b1f`, and the reweighted-*two-tag* and `3b1f` data compared. In the VBF channel, the events were further split into the nominal $|\Delta\eta_{HH}|$ categories, to assess the modelling of the background estimate in each category. Figure 7.9 shows the results. Good agreement between the reweighted-*two-tag* and `3b1f` data is observed across both VBF categories, with the majority of `3b1f` data points lying within the uncertainty band of reweighted-*two-tag*. As with the results in `rev`$|\Delta\eta_{HH}|$ and the Shifted Regions, the results in `3b1f` provide further confidence in the accuracy of the background estimation strategy.

In the ggF channel, a small non-closure was observed in the `3b1f` region. This discrepancy was accounted for with an additional systematic uncertainty, and is discussed in Section 10.2.5.

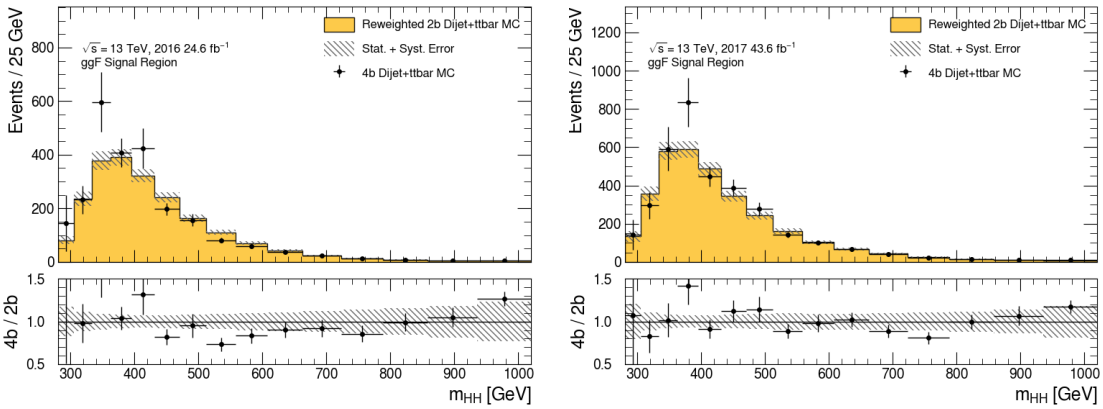


(a) $|\Delta\eta_{HH}| < 1.5$ (b) $|\Delta\eta_{HH}| > 1.5$

Figure 7.9: m_{HH} distributions for reweighted-*two-tag* and `3b1f` data in the two $|\Delta\eta_{HH}|$ categories of the SR of the VBF channel for all data-taking years.

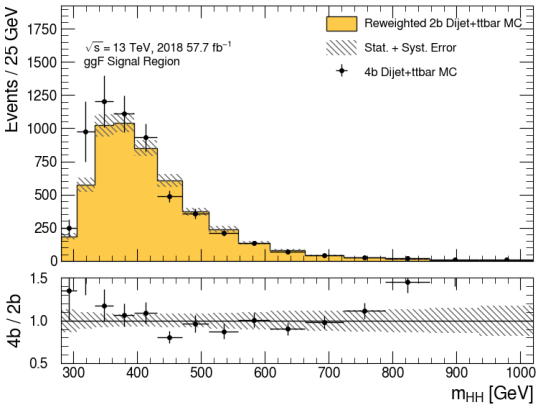
7.4.4 QCD Dijet and $t\bar{t}$ Monte Carlo Simulation

The two leading contributions to the background come from QCD multijet and $t\bar{t}$ events. Dijet QCD and $t\bar{t}$ MC simulation samples were generated to help validate the background estimate (see Section 4.2 for sample details). These samples were run through the procedure for generating the background estimate. From this, it was estimated that 10.2% of the background distribution across all years in the ggF channel is composed of $t\bar{t}$, and 8.1% in the VBF channel. Additionally, for the ggF channel, the reweighted MC *two-tag* QCD and $t\bar{t}$ yields and m_{HH} distributions were compared to those in the *four-tag* SR, as shown in Figure 7.10. No significant disagreement was observed in yields. Good agreement was also seen in the m_{HH} distributions, with the majority of *four-tag* data points lying within the uncertainty band of the reweighted-*two-tag* distribution. This provided further confidence in the validity of the nominal background estimate.



(a) 2016

(b) 2017



(c) 2018

Figure 7.10: Reweighted *two-tag* (yellow) and *four-tag* (black points) m_{HH} distributions from QCD dijet and $t\bar{t}$ Monte Carlo simulation in the Signal Region of the ggF channel for each data-taking year.

8

A ggF-VBF Combination for κ_λ-related Results

When probing *HH* production, ATLAS analyses are normally optimised to target the Gluon-Gluon Fusion (ggF) production process [93, 101, 102]. This is because ggF cross-section is higher than the second-leading *HH* production processes in the Standard Model (SM) by approximately an order of magnitude. There are ATLAS *HH* analyses that have been optimised for the second-leading *HH* production process, Vector Boson Fusion (VBF), but they focused on its unique sensitivity to the *HHVV* vertex, setting constraints on the associated coupling modifier κ_{2V} [91]. Theoretically, the trilinear Higgs self-coupling modifier, κ_λ, can be probed using VBF (see Figure 1.7a), but no ATLAS analysis aiming to set constraints on κ_λ has yet to included a channel optimised for VBF. In this chapter, I describe studies I performed to investigate whether adding a VBF-optimised channel to the ggF-optimised ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analysis could improve the sensitivity to *HH* production at SM and Beyond the Standard Model (BSM) κ_λ values.

8.1 *HH* Production Cross-Section and κ_λ

In the Standard Model, the cross-section for *HH* production via VBF is ∼5% the size of that via ggF. To better understand how the amplitude of these two processes varied with κ_λ, I compared the variation of their respective cross-sections with this

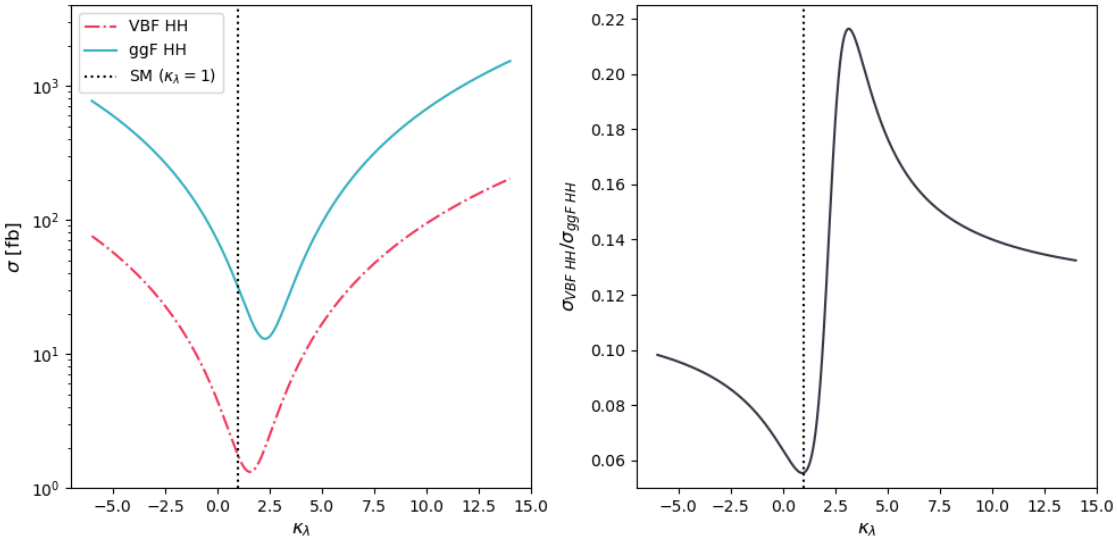


Figure 8.1: Left: comparison of the cross-section of *HH* production via Vector Boson Fusion (VBF) and Gluon-Gluon Fusion (ggF) for different values of the trilinear Higgs self-coupling modifier, κ_λ ($= c_\lambda / c_\lambda^{SM}$). Right: ratio of the VBF to ggF cross-sections shown in the left-hand plot. The black dotted line shows the Standard Model value ($\kappa_\lambda = 1$).

modifier. The results are shown in Figure 8.1. The right-hand plot shows the ratio between σ_{VBF} and σ_{ggF}. At the SM value, $\kappa_\lambda = 1$, the difference is close to its maximum, as the two VBF Feynman diagrams shown in Figures 1.7b and 1.7c destructively interfere and cancel [27]. The difference is vastly reduced for values of κ_λ below or above 1, in particular above 1. At $\kappa_\lambda = \sim 3$, σ_{VBF} is approximately a factor of four closer to σ_{ggF} than at the SM value. This is close to the point of maximal destructive interference between the *box* and *triangle* ggF diagrams (Figure 1.5). The relatively low value of σ_{VBF} that made VBF an undesirable production process to target compared to ggF is many times less significant at BSM values of κ_λ.

8.2 Comparison of Analysis Sensitivity

To establish how an additional VBF-optimised channel would impact the $HH \to b\bar{b}b\bar{b}$ analysis' constraints on κ_λ, I ran Monte Carlo simulation samples for ggF *HH* signal, VBF *HH* signal and data (used for the data-driven background estimate) through

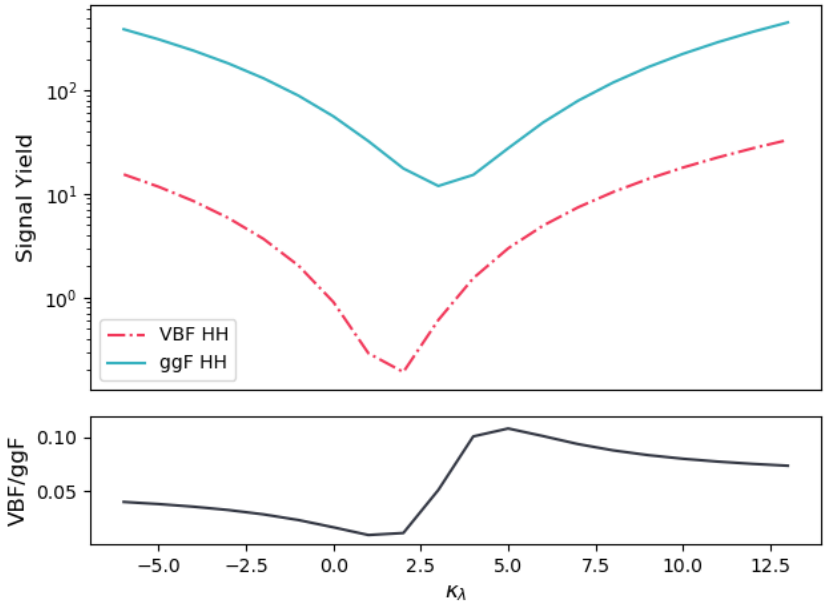


Figure 8.2: Top: comparison of the signal yield in the *four-tag* Signal Region at a range of κ_λ points using ggF and VBF *HH* Monte Carlo simulation samples for all data-taking years. Bottom: ratio of the VBF to ggF signal yields.

the analysis outlined in Chapter 4, using the definition of the VBF channel described there, and set upper limits on the rate of ggF and VBF *HH* production at values of κ_λ in our range of interest.[1]

Figure 8.2 shows a comparison of the post-selection signal yield for the ggF and VBF *HH* MC simulation samples in the *four-tag* Signal Region for different κ_λ values. The VBF yield is significantly larger for the BSM κ_λ values than for $\kappa_\lambda = 1$. This indicates that the increase in signal cross-section shown in Figure 8.1 translates into an increase in signal yield in the analysis. These events could have been removed by the selection, as the event kinematics do change with κ_λ.

The ratio of the VBF to ggF signal yield is smaller than that shown in Figure 8.1. However, this is due the VBF events having been through the additional selection

[1]The analysis described in this Chapter 4 has evolved since this study was conducted, so, although there are no major structural differences, the analysis is not identical. However, the limits at the end of this Chapter, used to show the final impact of the additional VBF-optimised category, are produced using the most up-to-date version of the analysis.

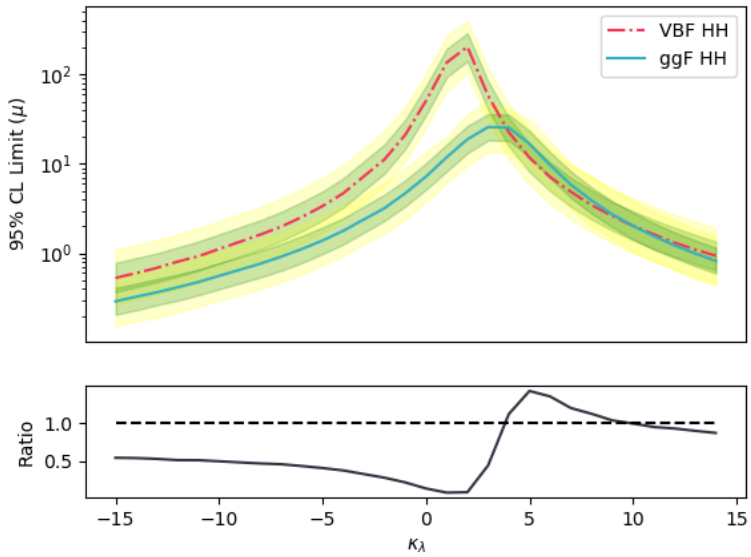
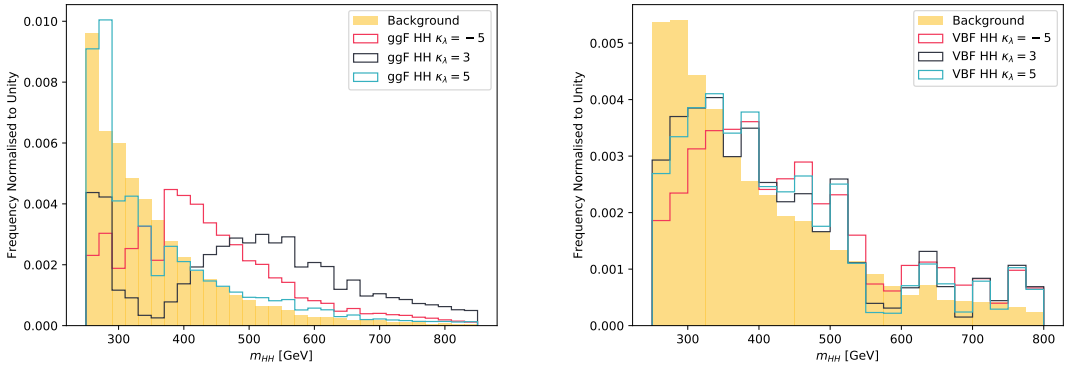


Figure 8.3: Top: comparison of the expected upper limits (95% CL) on signal strength, μ ($= \sigma/\sigma_{theory}$), for VBF *HH* Monte Carlo (MC) simulation in the VBF channel (dashed pink) and ggF *HH* MC in the ggF channel (blue) for 2016–18 years of data-taking. Bottom: ratio of the VBF limit to the ggF limit. The one and two standard deviation uncertainty bands on the expected limits are shown in green and yellow respectively.

required to be placed in the VBF channel, which also reduces the background yield.

Figure 8.3 shows a comparison of the expected upper limit (95%) on signal strength, μ ($= \sigma/\sigma_{theory}$), of the ggF and VBF *HH* processes in their respective channels. At $\kappa_\lambda = 1$, the limit on μ_{ggF} is far tighter than on μ_{VBF}. Given the large gap in the cross-section shown in Figure 8.1, this is expected, and indicates the inclusion of a VBF-optimised channel would likely not improve sensitivity to SM *HH* production. Away from $\kappa_\lambda = 1$, and particularly at high positive κ_λ, the limit on μ_{VBF} is more competitive. For example, at $\kappa_\lambda = 5$, the limit on μ_{VBF} and μ_{ggF} lie within each others' one standard deviation uncertainty band.

The results shown in Figure 8.3 indicate that if the ggF and VBF channels were statistically combined, the overall limit would be an improvement on using a single ggF-optimised channel. Furthermore, the improvement would be localised in the



(a) ggF channel (b) VBF channel

Figure 8.4: The distributions of the HH mass, m_{HH}, for VBF and ggF HH Monte Carlo simulation for three representative κ_λ values and for the data-driven background estimates in the ggF and VBF channels. All distributions are normalised to unity in order to be compared.

regions where the thresholds of the current-best exclusion limit on κ_λ ($-1.0 < \kappa_\lambda < 6.6$) lie [20]. Therefore, by introducing a VBF-optimised channel to the current $HH \to b\bar{b}b\bar{b}$ analysis, improved constraints on κ_λ can be expected.

The uncertainties accounted for when calculating the limits shown in Figure 8.3 were statistical and the systematic uncertainty on the background estimate.

8.3 *HH* Kinematics and κ_λ

Varying the value of κ_λ alters the prominence of diagrams including κ_λ with respect to other ggF/VBF diagrams. This changes the kinematics of ggF and VBF HH events. To better understand how the change in the kinematics impacts analysis sensitivity, I studied the post-selection distributions of m_{HH} (the variable of interest) in VBF and ggF HH MC simulation and in background in the VBF and ggF channels for a range of κ_λ values. The results are shown Figure 8.4.

The variation in the shape of the m_{HH} distributions with κ_λ is small for VBF HH MC, but is far greater for ggF. Two reasons explain this: a). the interference

between the *box* and *triangle* ggF diagrams (Figure 1.5) result in high variability in ggF kinematics with κ_λ, and b). the tight requirements of the VBF-selection result in the events in the VBF channel being kinematically similar.

For $\kappa_\lambda = 5$, the ggF m_{HH} distribution is soft, centred at low values of m_{HH}, where the background distribution is most prominent. For $\kappa_\lambda = -5$, the distribution is harder, centred at higher values of m_{HH}, where there is less background. This difference is expressed in the upper limits on ggF and VBF *HH* cross-section: VBF is close to ggF at $\kappa_\lambda = 5$, where the ggF distribution lies in a high background region; however, it is further away at $\kappa_\lambda = -5$, where the ggF distribution lies in a lower background region.

8.4 Results

In response to the studies outlined in the previous sections of this chapter, the ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analysis adopted a proposal to include a VBF-optimised channel and pursue a ggF-VBF combination when deriving results. Figure 8.5 shows the expected upper limit (95% CL) on the combined ggF and VBF *HH* cross-section for a range of κ_λ values using the VBF channel, ggF channel and a statistical combination of the two. Table 8.1 summarises these results, showing the exclusion limits set on κ_λ for each channel and their combination. With the addition of the VBF channel, the expected constraints on κ_λ show good improvement, decreasing from $-4.8 < \kappa_\lambda < 11.5$ to $-4.6 < \kappa_\lambda < 10.8$.

Whilst there is good improvement provided by implementing the two channel strategy, it didn't fully reflect the expected improvement from the studies described in this chapter. Since those studies were performed, the ggF channel, being the primary channel, has undergone further optimisation. Therefore, the VBF channel offers less improvement now than when these studies were performed. However, the future is

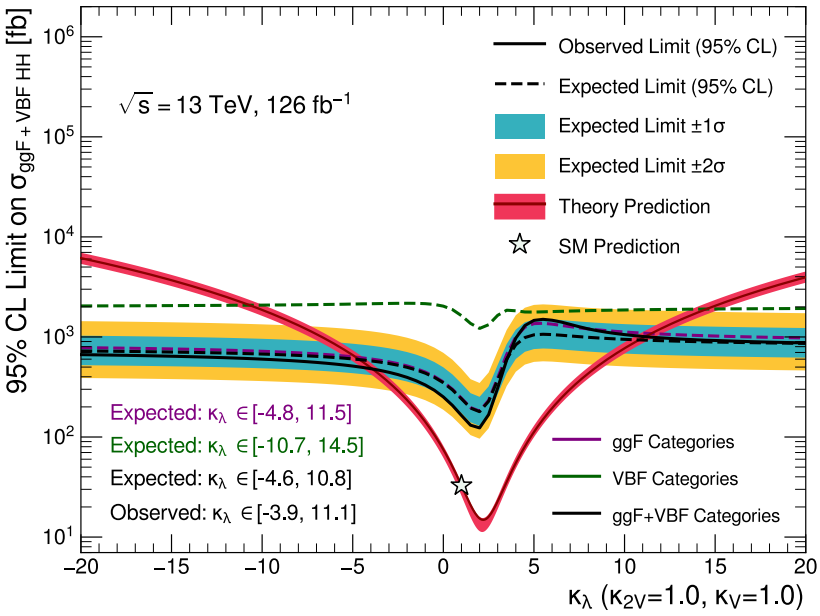


Figure 8.5: The expected upper limit (95% CL) on the combined ggF and VBF *HH* production cross-sections at a range of κ_λ values, with the individual limits from the ggF (purple) and VBF (orange) channels overlaid. The observed upper limit for the combined ggF and VBF channels is also shown (solid black).

Table 8.1: The expected constraints (95% CL) on the Higgs trilinear coupling modifier, κ_λ, for the ggF channel, the VBF channel and the combination of the two.

Channel	**Expected Constraints on κ_λ**
ggF channel	-4.8 – 11.5
VBF channel	-10.7 – 14.5
Combination	-4.6 – 10.8

promising. Being a new addition, there is space for further optimisation of the VBF channel with respect to κ_λ, which could increase sensitivity further. Additionally, as constraints on κ_λ get tighter, the potential improvements offered by the VBF channel with respect to positive κ_λ increase (down to $\kappa_\lambda \approx 4.$), as shown in Figure 8.5.

8.5 Conclusion

ATLAS analyses targeting *HH* production are normally optimised for the leading *HH* production process, Gluon-Gluon Fusion (ggF), due to its relatively large SM cross-section. The study of the second leading *HH* production process, Vector Boson Fusion (VBF), is primarily motivated by its unique sensitivity to the *HHVV* vertex coupling, κ_{2V}. I studied the prospect of improving the $HH \rightarrow b\bar{b}b\bar{b}$ analysis' exclusion limits on κ_λ by incorporating an additional channel optimised for VBF. The outcome of these studies indicated that this two-channel strategy had the potential to improve the analysis constraints on κ_λ, and, so, this strategy was adopted by the analysis. This is the first ATLAS *HH* analysis to have an optimised channel for each of the two leading *HH* production processes and the first to combine them for their final results. Expected results indicate that the inclusion of the two-channel strategy improved the constraints on κ_λ from $-4.8 < \kappa_\lambda < 11.5$ to $-4.6 < \kappa_\lambda < 10.8$ and the improvement should increase as the constraints get tighter. The improvement is due to the increase in the VBF cross-section away from the SM value of κ_λ (one) and the m_{HH} distribution for ggF tending to sit at low values of m_{HH} when $\kappa_\lambda > 1$, where background is high.

9 Statistical Analysis

The goal of the $HH \to b\bar{b}b\bar{b}$ analysis is to make statements about the rate of HH production and the values of HH couplings, κ_λ, κ_V and κ_{2V}, using the available data. In order to make a meaningful statement, the analysis utilises statistical analysis. The details of the statistical analysis are outlined in this chapter.

9.1 The Histogram

After the analysis selection, the data is organised into histograms in terms of the variable of interest, the HH mass, m_{HH}. m_{HH} was the variable of interest in the previous iteration of the $HH \to b\bar{b}b\bar{b}$ analysis, and displays a good level of separation between signal and background. Data in each analysis channel and category is organised into an individual histogram. Additionally, the data in the ggF channel is further separated by year. This follows the scheme described in Chapter 5. The binning of the m_{HH} histograms is described in Section 5.3.

9.2 The Likelihood

Searches for unobserved processes in ATLAS seek to, first, state whether the process exists (a *discovery*), then, if not, rule out a section of the phase space where the process could have existed (a *confidence limit*).

In the case of the $HH \rightarrow b\bar{b}b\bar{b}$ analysis, extrapolations to $\mathcal{L}_{INT} = 126\,\text{fb}^{-1}$ using the previous iteration of the analysis demonstrated that sensitivity levels are insufficient for a discovery. As such, this iteration of the analysis set confidence limits (CL) on HH production cross-section and HH couplings: κ_λ, κ_V and κ_{2V}.

Equation 9.1 describes the number of events in an analysis region in terms of the number of signal (s) and background (b) events. Associated with s is a scale-factor, μ, called the signal strength. μ is taken as the *parameter of interest* (POI) when performing the statistical analysis, and, as such, confidence limits are set directly on this parameter.

$$n = \mu s + b \tag{9.1}$$

The primary technique used in the statistical analysis is the *hypothesis test*. In this context, a hypothesis is a statement regarding which processes contribute to the data e.g. the background processes only ($\mu = 0$), the expected signal and background ($\mu = 1$) or some scaling of the signal and background ($\mu = \mu'$). A hypothesis is manifested using *probability density functions* (PDFs). The agreement between a hypothesis and data is quantified using a *likelihood*, referred to as the *model*, which is the product of the PDFs for each point in the data.

In the $HH \rightarrow b\bar{b}b\bar{b}$ analysis, the likelihood is the product of the Poisson distributions describing the event count in each bin in every histogram and category and a set of terms associated with the systematic uncertainties. The likelihood is shown in Equation 9.2, with the key terms explained afterwards.[1]

[1]Note, $\lambda_{c,b}(\hat{\mu}, \hat{\boldsymbol{\theta}})$ is the expected rate parameter in the Poisson distribution and not the likelihood ratio, $\tilde{\lambda}(\mu)$.

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{c \in \text{cats}} \prod_{b \in \text{bins}} \text{Pois}(n_{c,b} \mid \lambda_{c,b}(\hat{\mu}, \hat{\boldsymbol{\theta}})) \prod_{s \in \text{systs}} \text{Gaus}(\hat{\theta}_s \mid \theta_s). \tag{9.2}$$

Here, $\text{Pois}(n_{c,b} \mid \lambda_{c,b}(\hat{\mu}, \hat{\boldsymbol{\theta}}))$ is a Poisson PDF describing the number of events, $n_{c,b}$, for an expected rate of $\lambda_{c,b}(\hat{\mu}, \hat{\boldsymbol{\theta}})$ (the $\hat{}$ and θ symbols will be explained shortly). The subscript c and b references the product over the categories and bins. The symbol $\hat{}$, visible in Equation 9.2, represents the *maximum likelihood estimator* (MLE) of a parameter (the *post-fit* value). The MLE is an estimate of a parameter derived from data in the analysis, and corresponds to the value of the parameter that maximised the likelihood. The MLE plays an important role in hypothesis testing, discussed in the next section.

The last product in Equation 9.2 integrates the analysis's systematic uncertainties into the model. Systematic uncertainties are introduced into the likelihood using additional parameters called *nuisance parameters* (NP or θ). In Equation 9.2, the NPs are stored in a vector, $\boldsymbol{\theta}$. NPs increase the flexibility of the model, allowing it to cover deviations due to systematic uncertainties in exchange for a reduction in sensitivity to the POI. The initial value of an NP (the *pre-fit* value) is typically measured in an auxiliary measurement. This can be as part of a calibration exercise; for example, in the measurement of the Jet Energy Spectrum. A distribution describing the NP, which is centred at the auxiliary measurement, is incorporated into the likelihood. This distribution accounts for the uncertainty associated with the auxiliary measurement and is referred to as the NP *constraint term*. The most common option is to approximate the NP distribution as a Gaussian, thus the Gaussian PDF term, $\text{Gaus}(\hat{\theta}_s \mid \theta_s)$, at the end of Equation 9.2, the subscript s referencing the product over all systematic uncertainties [103]. This is a valid approximation in the large sample limit, as the distribution of an MLE approaches a Gaussian in this limit [104]. Two other common choices of constraint terms are *Poisson* (also referred

to as *Gamma*) and *Log-Normal*. The Poisson constraint term describes the NP as a Poisson distribution, which is valid when the auxiliary measurement is based on a counting experiment. As such, NPs associated with statistical uncertainties are encoded into the likelihood using Poisson constraints. The Log-Normal describes a distribution whose logarithm is a Gaussian. The Log-Normal is used an alternative to the Gaussian because it is strictly non-negative, matching NPs which can only be positive, e.g. luminosity measurements.

9.3 Hypothesis Testing

Hypothesis testing is a method of statistical inference used to compare a *null* hypothesis (H_0) to an *alternative* hypothesis (H_1). It relies on the use of a *test statistic*, which quantifies the agreement between two hypotheses.

According to the Neyman-Pearson Lemma, the most powerful test statistic is the *likelihood ratio*.[2] The likelihood ratio, $\tilde{\lambda}(\mu)$, used in the $HH\rightarrow b\bar{b}b\bar{b}$ analysis is shown in Equation 9.3. It is used to quantify the level of agreement between the model with a hypothesised value of μ and the best fit of the model to data. μ and $\boldsymbol{\theta}$ are simultaneously fit to the data using the MLE method and inserted into the model to obtain the best fit of the model to data, $\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})$.

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} & \hat{\mu} \geq 0, \\ \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(0, \hat{\boldsymbol{\theta}}(0))} & \hat{\mu} < 0. \end{cases} \tag{9.3}$$

The symbol $\hat{\ }$, visible in Equation 9.3, represents the *profiled* MLE. This is the MLE of a parameter when the value of a parameter it is dependent on is fixed. Typically, the NPs in a likelihood are profiled when testing a specific μ hypothesis. The MLE

[2]The *power* of a hypothesis test is the probability to reject H_0 and accept H_1 given H_1 is true.

of $\boldsymbol{\theta}(\mu)$ when stipulating $\mu = \mu'$ is $\hat{\hat{\boldsymbol{\theta}}}(\mu')$.

When $\hat{\mu} < 0$, $\hat{\mu}$ is replaced with zero in $\tilde{\lambda}(\mu)$ to prevent the use of a negative signal contribution, which is unphysical.

The final test statistic is taken as $\tilde{q}_\mu$, which is the natural logarithm of $\tilde{\lambda}(\mu)$ multiplied by -2, as shown in Equation 9.4.[3] The $HH \to b\bar{b}b\bar{b}$ analysis seeks to set upper limits on μ, so only μ above $\hat{\mu}$ is considered, with $\tilde{q}_\mu$ set to zero otherwise.

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} = \begin{cases} -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(0,\hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{\mu} \leq 0, \\ -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\boldsymbol{\theta}})} & 0 \leq \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \tag{9.4}$$

The *p-value* is a measure of the agreement between an observation and a hypothesis. It is defined as the probability, given a hypothesis, of observing a measurement more extreme than the one obtained. The equation for the p-value for a hypothesised value of μ, p_μ, is shown in Equation 9.5.

$$p_\mu = \int_{\tilde{q}_{\mu,\text{obs}}}^{\infty} f(\tilde{q}_\mu \mid \mu, \boldsymbol{\theta}(\mu))\, d\tilde{q}_\mu \tag{9.5}$$

Here, $f(\tilde{q}_\mu \mid \mu, \boldsymbol{\theta}(\mu))$ is the PDF of the test statistic, $\tilde{q}_\mu$, assuming a hypothesised value of μ, and $\tilde{q}_{\mu,obs}$ is the observed value of $\tilde{q}_\mu$.

A μ hypothesis is rejected if p_μ is less than a user-specified threshold, α. In ATLAS,

[3] Taking the natural logarithm multiplied by -2 ensures that when the null hypothesis is true the LLR converges to being χ^2-distributed in the asymptotic limit.

the standard for deriving confidence limits is $\alpha = 0.05$, which is the 95% CL. The boundaries of the 95% CL on μ are derived by solving $p_\mu = 0.05$ for μ. In the $HH \to b\bar{b}b\bar{b}$ analysis, this is done numerically.

The test statistic PDF, $f(\tilde{q}_\mu \mid \mu, \boldsymbol{\theta}(\mu))$, can be generated using a number of independent Monte Carlo simulation samples; however, this is very computationally expensive. Instead, $f(\tilde{q}_\mu \mid \mu, \boldsymbol{\theta}(\mu))$ is approximated using a $\frac{1}{2}$-χ^2 distribution, which is the asymptotic limit of the true PDF under the assumption of a large number of events [105].[4]

9.4 The CL_s Method

In the $HH \to b\bar{b}b\bar{b}$ analysis, the background yield is significantly larger than the signal yield. In such a scenario, an increase in the event yield due to signal processes could be nullified by a downward fluctuation in the background, and valid signal hypotheses could be rejected as a result. To counteract this, the p-value is modified using the CL_s method, as shown in Equation 9.6.

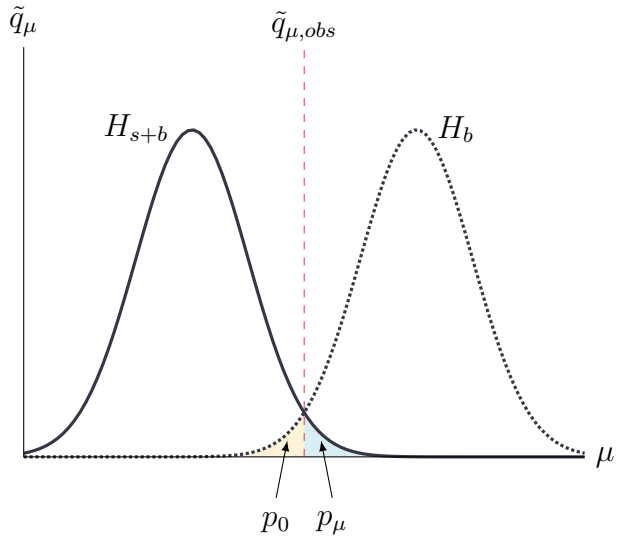
$$p'_\mu = \frac{p_\mu}{1 - p_0} \tag{9.6}$$

Here, p_0 is the p-value under the background-only hypothesis ($\mu = 0$), as defined in Equation 9.7.[5]

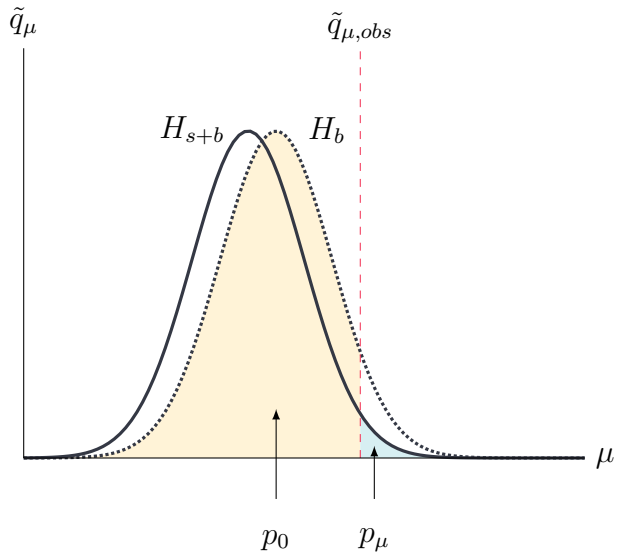
$$p_0 = 1 - \int_{\tilde{q}_{\mu,\text{obs}}}^{\infty} f(\tilde{q}_\mu \mid 0, \boldsymbol{\theta}(0))\, d\tilde{q}_\mu \tag{9.7}$$

[4] This is the sum of a delta function, $\delta(\tilde{q}_\mu)$, and a χ^2 PDF, each given a weight equal to $\frac{1}{2}$.

[5] It may appear contradictory that $p_0 = 1 - integral$ whereas $p_\mu = integral$. Figure 9.1a provides a clear explanation. The p-value is measured from opposite sides of the observed test statistic for H_b and H_{s+b}, explaining the need to take away the integral from one for p_0.



(a) H_b and H_{s+b} are dissimilar.



(b) $H_b \sim H_{s+b}$.

Figure 9.1: The distributions of the test statistic under the background-only (H_b, thinly-dashed navy line) and signal & background (H_{s+b}, filled navy line) hypotheses, alongside the observed value of the test statistic, $\tilde{q}_{\mu,obs}$ (dashed pink line). The p-value under H_b is labelled as p_0 and under H_{s+b} for a given μ as p_μ. Top is when the test statistic distributions under the two hypothesis are dissimilar and bottom is when they are similar.

Figure 9.1 shows a schematic of the test statistic PDFs under the background-only (H_b) and signal & background (H_{s+b}) hypotheses, with p_0 and p_μ labelled. In a scenario where the signal contribution is small relative to the background, the PDFs of the test statistic under H_b and H_{s+b} are very similar, as seen in Figure 9.1b. As such, one might reject H_{s+b} even though the probability that H_b is correct, $(1-p_0)$, is small. To counteract this, p_μ is weighted by the inverse of $(1-p_0)$. As $(1-p_0)$ decreases towards p_μ, p'_μ increases away from p_μ, preventing the incorrect rejection of H_{s+b}.

9.5 Expected Sensitivity

Alongside the observed upper limit on μ, the *expected* upper limit is also calculated. This is a vital optimisation metric prior to the unblinding of the Signal Region, and a reference point for deviations between expectation and observation.

The expected sensitivity is derived using the *Asimov dataset* [105]. This is defined as the binned dataset where the bin content is equal to the expected bin content.

Under the assumption of the asymptotic approximation of the test statistic PDF described earlier, the expected $\mathrm{CL_s}$ upper limit on μ and the $\pm N$ standard deviations on this limit are given by Equations 9.8 and 9.9 [105].

$$\mu_{up+N} = \sigma(\Phi^{-1}(1-\alpha\Phi(N)) + N) \tag{9.8}$$

$$\sigma = \frac{\mu}{\sqrt{q_{\mu,A}}} \tag{9.9}$$

Here, Φ is the *cumulative distribution function* (CDF) for a standard Gaussian,

$\Phi(N)$ is the value of the CDF for N standard deviations and $q_{\mu,A}$ is the test statistic for a hypothesised value of μ derived from the Asimov dataset, thus the subscript A.

9.6 Pulls and Impacts

A series of diagnostic techniques are used to check the validity of the fit of the nuisance parameters performed as part of the calculating the likelihood ratio. The pull, defined in Equation 9.10, is a measure of the difference between the MLE value of a NP and its pre-fit value.

$$\alpha = \frac{\hat{\theta} - \theta_0}{\sigma_\theta} \tag{9.10}$$

It is expected that the *post-fit* NPs will be pulled to some degree; however, significant pulls, particularly to values outside of the $\pm 1\sigma$ NP uncertainty bounds, can be a problem, as it indicates the original uncertainty estimate is inaccurate. Additionally, a significant post-fit reduction of the NP uncertainty bounds, referred to as an *over-constraint*, can potentially be a problem for the same reason.

The *impact* associated with a particular NP is defined as the change in the parameter of interest, e.g. μ, when the NP is shifted to its upper and lower uncertainty bounds. A large impact indicates that the associated uncertainty is highly influential on the measurement of the POI. Often, the pull values associated with the series of NPs used in a physics measurement are displayed in a *ranking plot*, where those with the largest impact are positioned at the top and those with the smallest at the bottom.

When a large-impact NP is found to have a significant pull or to be over-constrained, a discussion, and potential investigation, follows. If the investigation finds the NP

parameterisation poorly approximates the underlying uncertainty, this must be corrected.

Additionally, the values of the correlation coefficients between all NPs (the correlation matrix) should be assessed. NPs representing independent sources of uncertainty should be uncorrelated. There can be legitimate physical reasons why NPs are highly correlated or anti-correlated, but these should be understood if observed. This is because large (anti-)correlations can be a sign that the algorithm performing the MLE fit has not converged to extremum of the log likelihood, as the correlation matrix is extracted directly from the fit. If the fit does not fully converge, the final results will be inaccurate.

The checks of the validity of the fit in the $HH \rightarrow b\bar{b}b\bar{b}$ analysis are described in Section 11.1.

10 Uncertainties

Uncertainties can be separated into two broad categories: *statistical* and *systematic*. Statistical uncertainties cover random fluctuations. If an experiment was repeated identically several times, the results would differ simply by chance. Systematic uncertainties cover biases, which can arise in all steps in the analysis, from data collection to result generation. Sources of systematic uncertainty include: mis-modelling in the data-driven background estimate, and inaccuracy in the jet energy and b-tagging efficiency corrections. In this chapter, the origin of the uncertainties that impact the $HH \to b\bar{b}b\bar{b}$ analysis will be discussed alongside how these uncertainties were quantified and accounted for.

10.1 Statistical Uncertainties

In the $HH \to b\bar{b}b\bar{b}$ statistical analysis, data is stored in histograms of m_{HH}. The number of counts in a histogram bin follows a Poisson distribution (Equation 10.1), as this describes the number of events that occurred in a fixed interval.

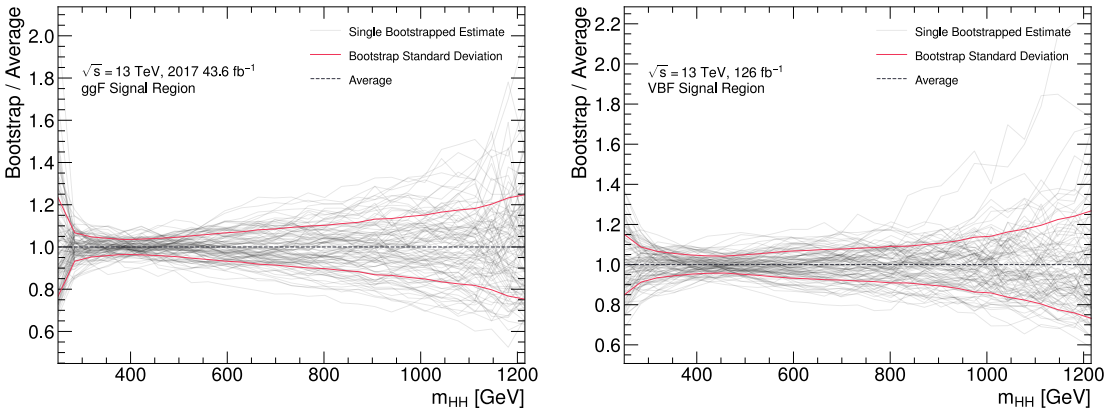
$$f(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \tag{10.1}$$

Here, λ is the expectation value of the number of events, which also happens to equal the variance. As such, the standard deviation, σ, is equal to $\sqrt{\lambda}$. The statistical uncertainty on the number of events in a histogram bin, n, is taken as $\pm\sigma$, but, as λ is generally not known, $\sqrt{n}$ is used as an approximation.

Using $\sqrt{n}$ is not appropriate for quantifying the statistical uncertainty for a weighted histogram, which is the case for both Monte Carlo simulation and the background estimate. In this scenario, the statistical uncertainty is taken as the square root of the sum of the squared event weights, $\sqrt{\sum_i w_i^2}$.[1]

In Section 7.2, the bootstrap uncertainty was described. As a reminder, this is an uncertainty on the background estimate, and covers the effect of the random initialisation of the Neural Network and the finite size of the training dataset. As this uncertainty aims to cover statistical fluctuations, it is treated as a statistical uncertainty in the fit. An individual nuisance parameter is associated with each m_{HH} bin in the model. There are three primary motivations behind treating the bootstrap NPs as uncorrelated. First, it is attempting to account for a uncertainty that is statistical in origin, so it follows that it should be implemented similarly to the standard Poisson statistical uncertainty. Second, it is a more conservative choice than the alternative of correlating across m_{HH} bins. Third, a bootstrap shape uncertainty would be highly related to the existing shape uncertainty associated with the background estimate (see Section 10.2.4). Figure 10.1 shows the distribution of the bootstraps around the nominal background estimate alongside the bootstrap uncertainty for 2017 data in the ggF channel and 2016–18 data in the VBF channel. This is representative of the bootstrap uncertainty in all categories and for all years. The same plots for 2016 and 2018 data in the ggF channel are shown in Figures D.1a and D.1b in Appendix D.

[1] $\mathrm{V}[\sum_i w_i] = \sum_i \mathrm{V}[w_i] = \sum_i w_i^2$, given $\mathrm{V}[w_i] = \mathrm{V}[w_i \times 1 \text{ event}] = w_i^2 \times \mathrm{V}[1 \text{ event}] = w_i^2$.



(a) ggF channel (2017)

(b) VBF channel

Figure 10.1: Ratio of the mean background estimate histogram to the ninety-nine other bootstrap histograms (grey), where the bootstrap uncertainty (red) is the standard deviation of the distribution of bootstraps. Left: 2017 data in the ggF channel. Right: all-years (2016–18) data in the VBF channel.

10.2 Systematic Uncertainties

As introduced in Chapter 9, nuisance parameters, $\boldsymbol{\theta}$, add flexibility to the model (Equation 9.2) to allow for coverage of sources of systematic uncertainty. The procedure to ensure a model responds appropriately to a pull in a NP is complex. Take the Jet Energy Spectrum (JES) calibration as an example. First, the JES calibration value and uncertainty is measured externally by the `Jet/Etmiss` CP group in a *subsidiary measurement*. Then, the entire analysis chain is run three times: once with the nominal JES calibration, once with the JES calibration reduced to its lower uncertainty bound (a *down variation*), and once with it increased to its upper bound (an *up variation*).[2] This generates three histograms of the variable of interest, here m_{HH}. A technique known as *template morphing* is then used to combine the down and up variation histograms with the nominal by interpolating across equivalent bins in each variation using a function of the JES NP. Thus, as the JES NP value is varied up or down, the model approaches the result when the JES calibration was fluctuated upwards or downwards. The same procedure is used to encode all

systematic uncertainties into the model used in the $HH \to b\bar{b}b\bar{b}$ analysis.

10.2.1 JES and JER Uncertainties

As discussed in Section 3.3.3, the JES calibration corrects the energy of jets reconstructed in the detector to match that of equivalent jets reconstructed using particle information from MC simulation. This is not automatically true due to systematic biases in detection and reconstruction. The JES calibration is measured in a series of stages, using a range of event topologies in both MC simulation and data. As such, there is a number of systematic uncertainties associated with the measurement of the JES. This includes uncertainties inherited from the calibration of physics objects; the choice of analysis cuts applied when performing the measurement; and the event generator used for the MC simulation. In total, there are 125 individual nuisance parameters associated with the JES calibration (for the full list, see Table 2 in Ref. [66]), but a smaller set is used in the $HH \to b\bar{b}b\bar{b}$ analysis, as will be discussed.

As discussed in Section 3.3.4, the Jet Energy Resolution (JER) quantifies the variation in the measured energy of jets due to detection and reconstruction effects. Biases associated with the dijet balance method are the dominant component of the JER uncertainty. At low $\mathrm{p_T}$, this originates from the uncertainty inherited from the JES calibration, the JES being applied to all jets prior to the JER measurement. At high $\mathrm{p_T}$, it is the *non-closure* uncertainty, which accounts for the difference in the measured resolution for jets reconstructed in the detector and jets built using particle information in MC simulation.

The inclusion of the full set of jet uncertainty nuisance parameters is excessive for analyses not performing precision jet measurements, which includes the $HH \to b\bar{b}b\bar{b}$ analysis. The `Jet/EtMiss` group, who measure the JES and JER and supply the calibration uncertainties, also provide a reduced set of NPs. The $HH \to b\bar{b}b\bar{b}$ analysis

[2]This is assuming that there is a single JES NP for simplicity. If there are several, the same procedure is repeated for each.

utilised the *category reduction* set, which has 30 JES and 13 JER NPs, named as such because the *in situ* NPs are combined if they fall into similar categories: *detector*, *statistical*, *modelling* or *mixed*. This set is recommended for all analyses that will be part of a combination, so was chosen for the the $HH \rightarrow b\bar{b}b\bar{b}$ analysis, as it will be combined with its sister HH analyses and the H analyses in the future.

10.2.2 Flavour Tagging Uncertainties

As discussed in Section 3.4.3, the b-tagging efficiency in data and MC simulation are not always in agreement. Therefore, scale factors, defined in Equation 3.5 as the ratio of b-tagging efficiency in data and MC simulation, are applied. These scale factors rely on the independent measurement of the b-tagging efficiency in data and MC simulation in a high purity $t\bar{t}$ sample. Primary sources of systematic uncertainty include: the inherited JES uncertainty; modelling of top quark processes in MC simulation, e.g. the choice of parton shower and hadronisation model, ISR and FSR rates, and PDF model; and the sizes of the various samples, which have an event selection applied to ensure high levels of $t\bar{t}$ purity.[3]

As with the JES/JER nuisance parameters, a reduced scheme is also used for the flavour tagging NPs. For the $HH \rightarrow b\bar{b}b\bar{b}$ analysis, the *medium* reduced scheme is used. In flavour tagging, there is a NP for every histogram bin in the scale factor measurement. In the reduced scheme, these NPs are combined into an envelope. For the medium scheme, there are nine NPs associated with the b-jet calibration, four with the c-jet and four with the *light*-jet. Additionally, two NPs are included to cover any inaccuracy in the extrapolation of the scale factors to high $\mathrm{p_T}$.

[3] *ISR* and *FSR* stand for to initial and final state radiation, i.e. QCD radiation from quarks involved in the interaction before and after proton-scattering. *PDF* stands for parton distribution function, the probability density for quarks and gluons within the proton.

10.2.3 Theory Uncertainties

Theory uncertainties cover biases associated with the choice of quantities and modelling procedures based on particle physics theory. The theory uncertainties accounted for in the $HH \to b\bar{b}b\bar{b}$ analysis are described below.

The uncertainty on the $H \to b\bar{b}$ branching ratio was taken as $^{+3.42\%}_{-3.48\%}$ using the uncertainty values listed in Ref. [106] and the combination procedure in Ref. [107].

The uncertainty due to the choice of event generator was quantified by comparing the m_{HH} distributions in MC simulation generated using either PYTHIA8 or Herwig7 for the same HH production process.[4] The uncertainty measurement was repeated for a number of ggF and VBF HH samples and within each category and, for ggF, year. The uncertainty was measured by calculating the weighted average of the difference in bin yields between the nominal (PYTHIA8) and alternative (Herwig7) m_{HH} histograms. In the weighted average, the weight applied is the inverse of the statistical error in a bin, so those bins with large statistical errors contribute less to the average. The uncertainties were measured for the SM and $\kappa_\lambda = 10$ ggF samples and for the SM, $\kappa_{2V} = 0$ and $\kappa_\lambda = 10$ VBF samples. For any other coupling point, the greatest uncertainty out of these two (ggF) or three (VBF) options is used. The uncertainties on the VBF HH samples were ∼10%, with a small number of coupling points in a small number of categories reaching ∼22%. The uncertainties on the ggF HH samples were typically below ∼20%, with a small number of coupling points in a small number of categories between ∼20–35%, and one reaching ∼60%.

Values of the HH cross-section for various coupling points are used to convert upper limits on signal strength, $\mu = \sigma/\sigma_{theory}$, to cross-section, σ. Uncertainties on these values comprise of two factors, themselves a combination of two other factors: PDF+

[4] See Section 4.2 for sample details.

α_S is the combination of the uncertainties on the PDF and strong coupling constant, α_S. scale + mtop is the combination of the scale and top mass scheme uncertainties, and only affects ggF *HH* significantly. The PDF + α_S and scale + mtop uncertainties for a Higgs boson of mass 125 GeV were ±3% and $^{+6\%}_{-23\%}$ for ggF *HH* respectively and ±2.1% and $^{+0.03\%}_{-0.04\%}$ for VBF *HH*.

The uncertainty on the *HH* MC simulation samples due to choice of PDF set and α_S, and the renormalisation and factorisation scale uncertainties due to missing higher orders were evaluated following Ref. [108]. This created an uncertainty envelope on the m_{HH} distributions that was generally $< 1\%$ for the PDF + α_S uncertainty and ∼2% for scale uncertainty, with a small number of bins reaching up to 5% for the former and 3% for the latter. The plots of the uncertainty envelope on the m_{HH} signal distributions are available in Appendix AK of Ref. [90].

The measured value of the integrated luminosity is used to scale MC simulation to data. The 1.7% uncertainty on this value is accounted for in the analysis [109].

10.2.4 Background Estimation Shape Uncertainty

The background estimate used in the analysis is built by applying weights derived in Control Region 1 (CR1) to events in the Signal Region (SR).[5] This introduces uncertainty to the background estimate, as CR1 covers a different region in m_{H1} vs. m_{H2} space to the SR.[6] This uncertainty is estimated using an alternative background estimate, generated using a set of weights derived in Control Region 2 (CR2). CR2 is similar to CR1 in terms of size and proximity to the SR, but also covers a different region in m_{H1} vs. m_{H2} space. As a result, the difference between the nominal and alternative background estimates provides a measure of the variation due to the weights being derived in different regions of m_{H1} vs. m_{H2} space. This uncertainty

[5] See Sections 4.10 and 7.2 for details.

[6] CR1 is positioned in close proximity to the SR in the m_{H1} vs. m_{H2} plane to ensure they are as similar as possible.

is referred to as the *background estimation shape uncertainty.*

Simply taking the difference between the CR1 and CR2-weighted background estimates resulted in the associated nuisance parameter (NP) being overconstrained.[7] This can be explained by the complexity of the data-driven background estimate. It is likely there are multiple sources contributing to this uncertainty, which are not well parameterised using a single NP. As a result, a more complex strategy inspired by the previous iteration of the $HH \to b\bar{b}b\bar{b}$ analysis was used [2]. In that analysis, two NPs allowed variation in different regions of HT space (HT $>$ 300 GeV and HT $<$ 300 GeV).[8] HT was used as a proxy for m_{HH}, the idea being that different background processes contribute more in different m_{HH} regions. Thus, separating the NPs in this way better describes the individual contributions to the uncertainty. In this iteration of the analysis, it was observed that the two HT-split NPs were overconstrained. So, a study was conducted to find a new NP decomposition scheme. A range of different high-level variables (e.g. X_{Wt}, HT and m_{H1}) were tested using different numbers of NPs. It was observed that separating in m_{H1} and m_{H2} produced the smallest overconstraints. This inspired the final strategy, discussed next.

The background shape uncertainty is parameterised using four NPs, each expressing the contribution from a different part of the SR (and therefore different parts of m_{H1} vs. m_{H2} space). To do this, the SR is divided into four quadrants, as shown in Figure 10.2. These quadrants are of roughly equal area and are orientated such that the dividing boundaries are aligned with those of the control regions. They are named for their position in the m_{H1} vs. m_{H2} plane: $\mathrm{Q_N}$ for North, $\mathrm{Q_E}$ for East, $\mathrm{Q_S}$ for South, and $\mathrm{Q_W}$ for West. Alternative background estimates are generated by taking SR events, applying the nominal CR1-derived weights to three of the SR quadrants and applying the alternative CR2-derived weights to the remaining SR quadrant. There

[7] In a maximum likelihood estimator fit to the Asimov dataset.

[8] HT is defined as the scalar sum of the $\mathrm{p_T}$ of a set of objects. Here, HT is calculated using the four Higgs candidate jets.

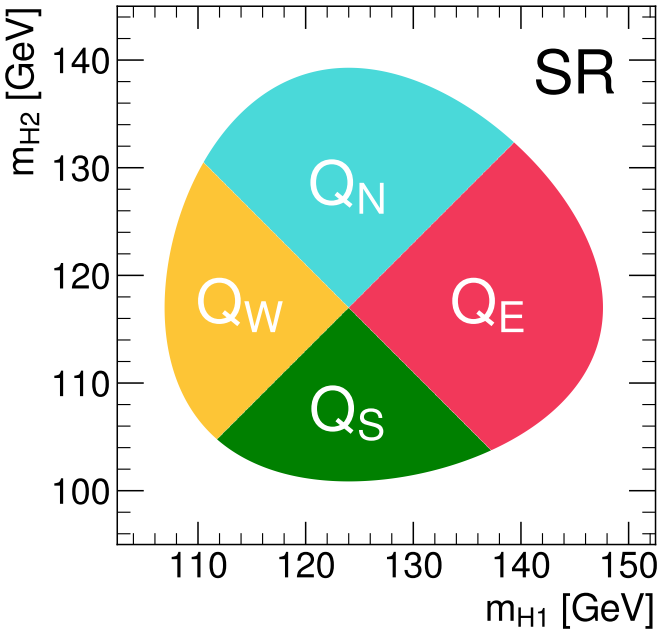
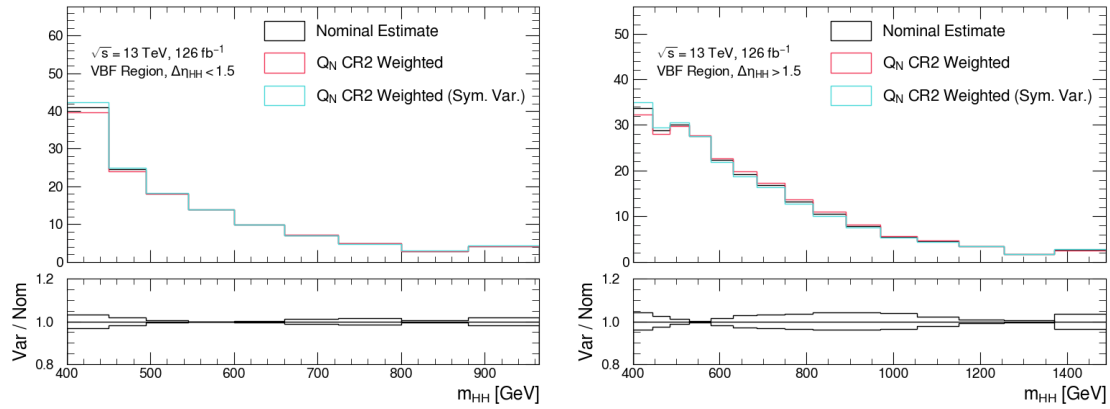


Figure 10.2: Visualisation of the Signal Region quadrants used in the derivation of the background estimation shape uncertainty. The four quadrants span the entirety of the Signal Region. Each quadrant is labelled with a subscript that represents its relative position – N (North), E (East), S (South) and W (West).

are four permutations and therefore four alternative background estimates and four associated NPs. So, roughly three quarters of the events in each alternative background estimate have the same weights applied as those in the nominal background estimate; it is only those that fall in a specific SR quadrant that have the alternative weights applied. The idea behind this choice of NP decomposition scheme is that different background processes will contribute more in different parts of the m_{H1} vs. m_{H2} plane, so separating the NPs like this parameterises these individual contributions. No overconstraints were observed on the background shape uncertainty NPs in a fit to the Asimov dataset when using this decomposition scheme.

The background shape uncertainty is *symmetrised* around the nominal background estimate. This means that the up variation is taken as nominal + (alternative − nominal) and the down variation is taken as nominal − (alternative − nominal).

Figures 10.3 and 10.4 show the symmetrised background shape systematic uncer-

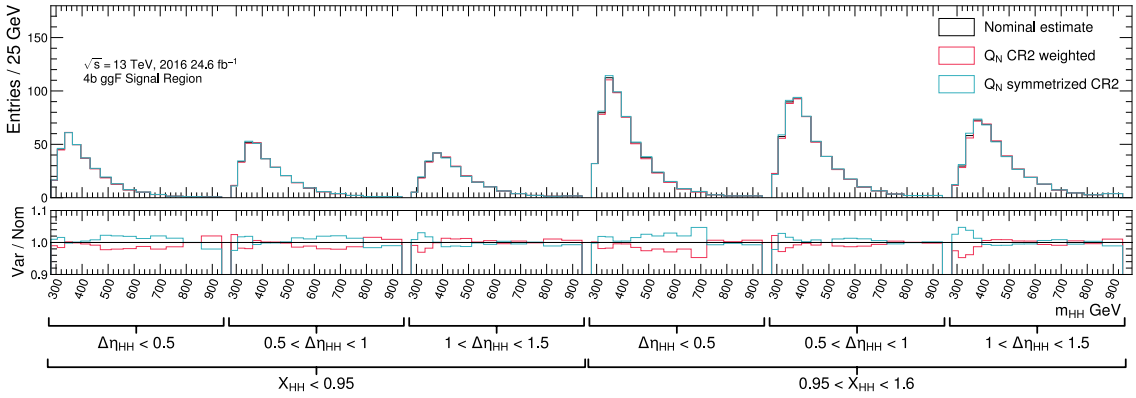


(a) $|\Delta\eta_{HH}| < 1.5$ (b) $|\Delta\eta_{HH}| > 1.5$

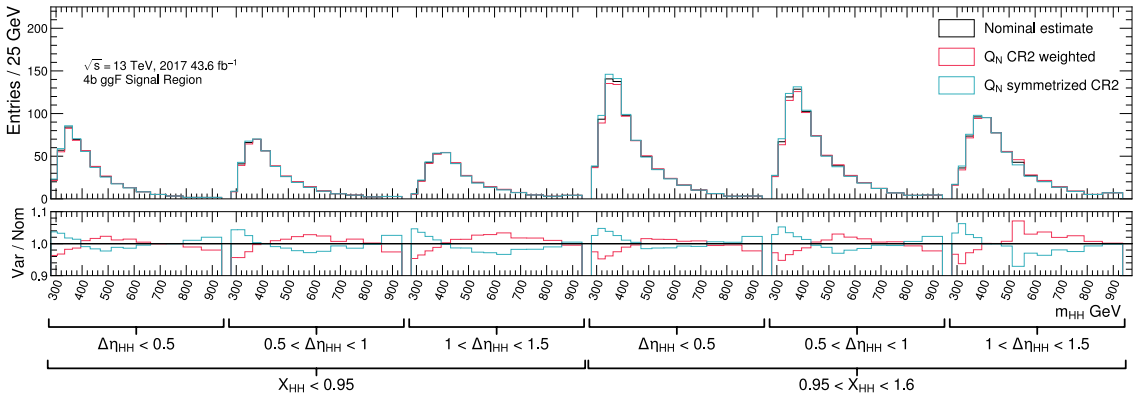
Figure 10.3: Symmetrised shape uncertainty on VBF channel background estimate, where CR2-derived weights are applied to Q_N and CR1-derived weights applied to the other Signal Region quadrants.

tainty in the different categories of the VBF and ggF channels respectively for the case when the CR2-derived weights are applied to Q_N and the CR1-derived weights are applied to the remaining SR quadrants. Histograms for the other quadrant permutations are shown in Appendix D.

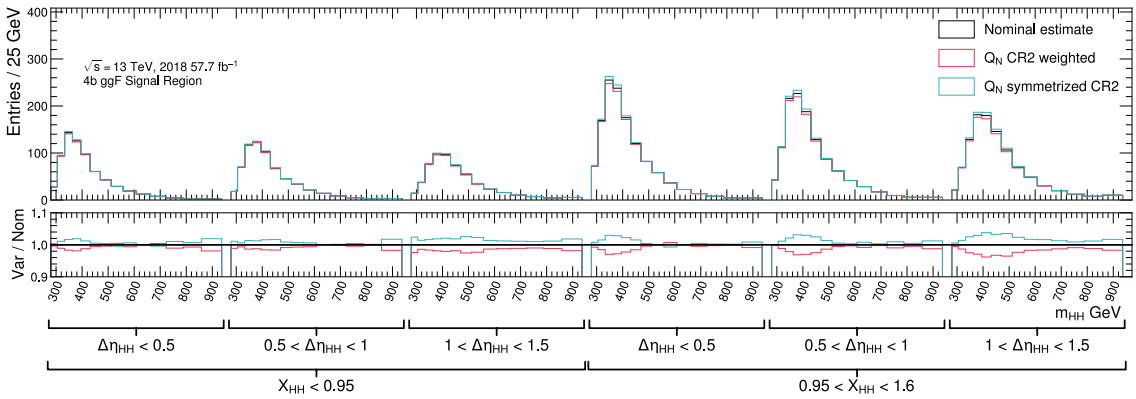
In the ggF channel, there are three years, six categories and four signal region quadrants. In the VBF channel, there are two categories and four quadrants, and the analysis is year-inclusive. Therefore, there are 72 background shape NPs associated with the ggF channel and 8 with the VBF channel. For each channel, the decision was made to correlate the background shape NPs across categories, as the Neural Network-based background estimation strategy is inclusive of category. Therefore, mismodeling in each category is likely to originate from the same source. Additionally, it was observed that the NP variation of the background m_{HH} histogram was similar across categories for specific years and quadrants, suggesting that any variation in the model will tend to cause coherent changes in the distributions.



(a) 2016



(b) 2017



(c) 2018

Figure 10.4: Symmetrised shape uncertainty on ggF channel background estimate, where CR2-derived weights are applied to Q_N and CR1-derived weights applied to the other Signal Region quadrants. Each plot simultaneously shows the six symmetrised shape uncertainty bands for the six $|\Delta\eta_{HH}|$–X_{HH} ggF categories, with the labels below the x-axis indicating the category a particular band corresponds to.

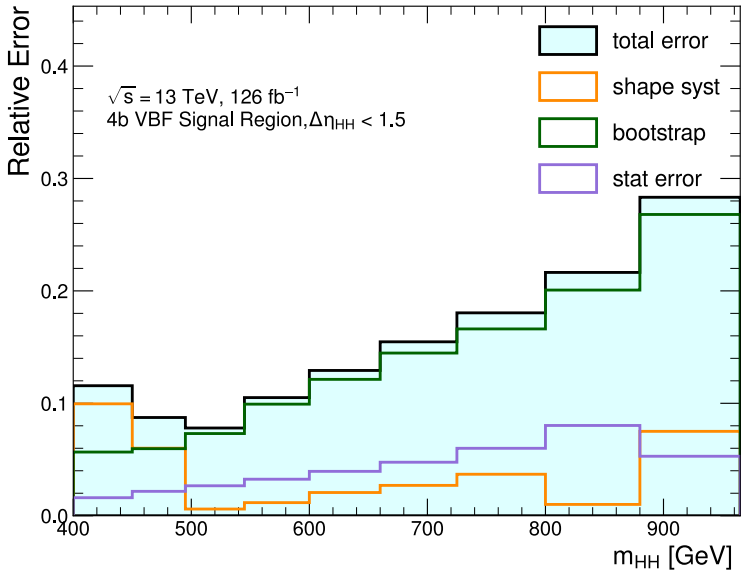
10.2.5 `3b1f` Non-Closure Uncertainty

In the `3b1f` validation region, the *four-tag* requirement is replaced with the stipulation that events must contain three b-jets at the 77% WP and none at any other WP. `3b1f` is discussed in Section 7.4.3. A small normalisation difference, referred to as a *non-closure*, was found between the reweighted-*two-tag* SR m_{HH} distribution in the ggF channel and the `3b1f` distribution it was attempting to model. It was decided this was the result of a systematic bias, which needed to be covered by an uncertainty on the background estimate. The `3b1f` non-closure uncertainty is derived by taking the difference between unity and the statistical uncertainty bound on the ratio of the reweighted-*two-tag* and `3b1f` data. This forms the `3b1f` uncertainty band, which is then smoothed. The smoothing procedure replaces the uncertainty on a bin with the average of the uncertainties on the three bins around that point.[9]

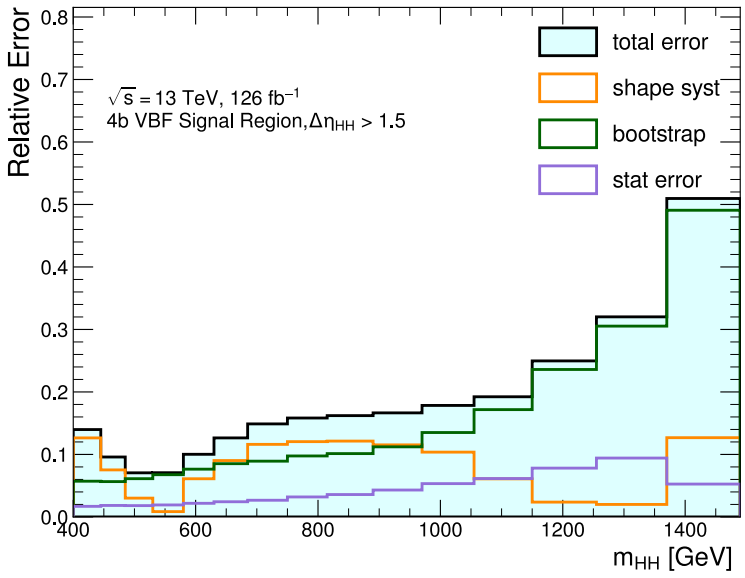
10.2.6 Summary of Background Uncertainties

A summary of different components of the uncertainty on the m_{HH} background estimate distribution is shown in Figures 10.5 and 10.6. The bootstrap uncertainty dominates at high m_{HH}, as statistics are low in the m_{HH} tail. At low m_{HH}, no single background uncertainty dominates. Additionally, a summary of the magnitude of the different components of the uncertainty on the background estimate is shown in Tables 10.1 – 10.2. Overall, the bootstrap uncertainty dominates in both the VBF and ggF channels.

[9]For example, if the uncertainty on the histogram was $[0.1, 0.05, 0.15, 0.2, 0.1]$, the smoothing procedure would replace the first element with $(0 + 0.1 + 0.05)/3 = 0.05$, the second element with $(0.1 + 0.05 + 0.15)/3 = 0.1$, and so on.

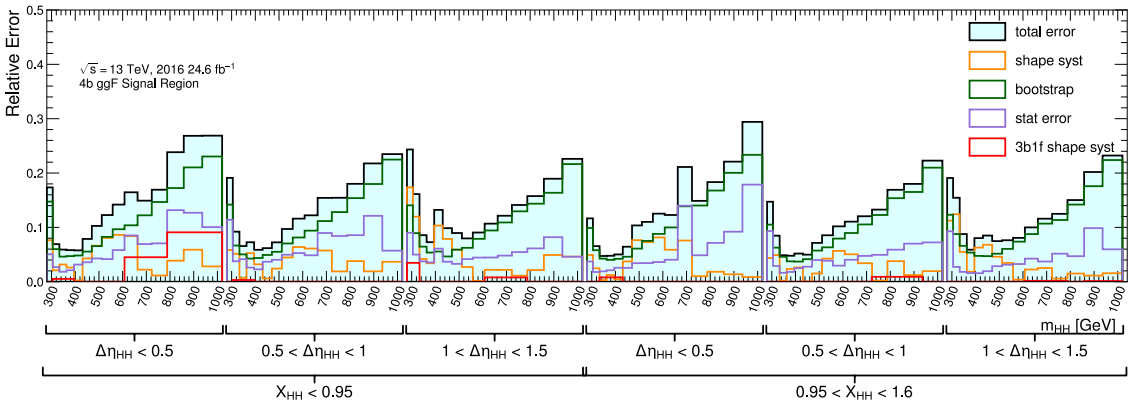


(a) $|\Delta\eta_{HH}|< 1.5$

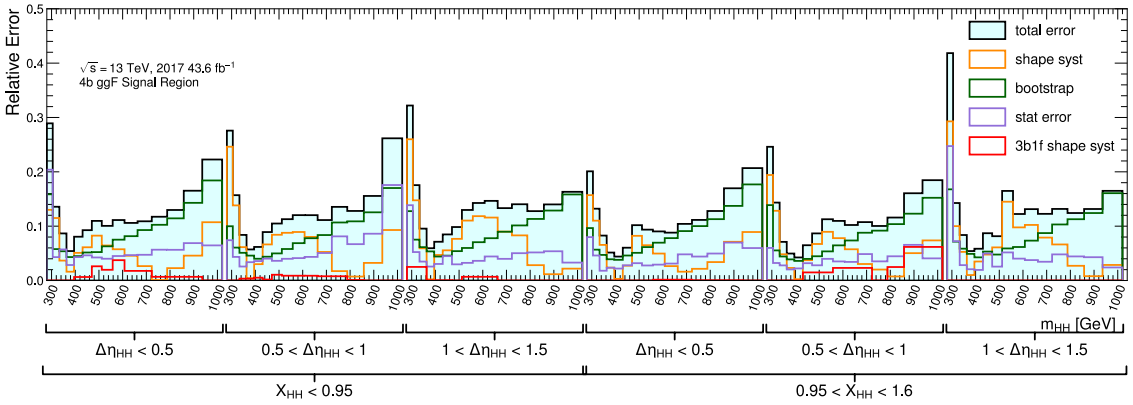


(b) $|\Delta\eta_{HH}|> 1.5$

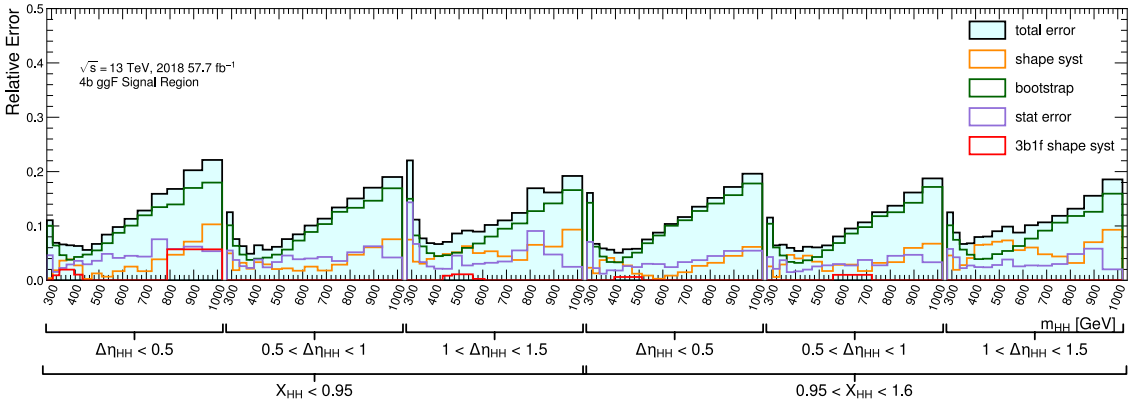
Figure 10.5: Components of the uncertainty on the m_{HH} background estimate distribution in the *four-tag* Signal Region and two $|\Delta\eta_{HH}|$ categories of the VBF channel.



(a) 2016



(b) 2017



(c) 2018

Figure 10.6: Components of the uncertainty on the m_{HH} background estimate distribution in the *four-tag* Signal Region of the ggF channel. Each plot simultaneously shows the uncertainty bands for the six $|\Delta\eta_{HH}|$–X_{HH} ggF categories, with the labels below the x-axis indicating the category a particular set of bands corresponds to.

Table 10.1: Magnitude of the components of the uncertainty on the background estimate in the *four-tag* Signal Region of the VBF channel. Total statistical uncertainty is the quadrature sum of the bootstrap and *two-tag* Poisson uncertainties.

SR 4b (Bkgd Yield = 354.6)	**Counts**	**Relative (%)**
Total Statistical Uncertainty	32.3	9.1
Shape Q_N	−1.2	−0.35
Shape Q_W	−0.8	−0.23
Shape Q_S	−0.5	−0.15
Shape Q_E	−0.3	−0.095
Total Systematic Uncertainty	1.6	0.45
Total Uncertainty	32.3	9.1

Table 10.2: Magnitude of the components of the uncertainty on the background estimate in the *four-tag* Signal Region of the ggF channel for 2016 data. Total statistical uncertainty is the quadrature sum of the bootstrap and *two-tag* Poisson uncertainties.

SR 16 4b (Bkgd Yield = 3211.1)	**Counts**	**Relative (%)**
Total Statistical Uncertainty	225.8	7.0
Shape Q_N	−25.3	−0.79
Shape Q_W	−12.1	−0.38
Shape Q_S	16.2	0.50
Shape Q_E	7.0	0.22
3b1f NC	6.8	0.21
Total Systematic Uncertainty	33.9	1.1
Total Uncertainty	228.3	7.1

Table 10.3: Magnitude of the components of the uncertainty on the background estimate in the *four-tag* Signal Region of the ggF channel for 2017 data. Total statistical uncertainty is the quadrature sum of the bootstrap and *two-tag* Poisson uncertainties.

SR 17 4b (Bkgd Yield = 4492.7)	**Counts**	**Relative (%)**
Total Statistical Uncertainty	308.9	6.9
Shape Q_N	−24.7	−0.55
Shape Q_W	−4.8	−0.11
Shape Q_S	43.0	0.96
Shape Q_E	14.8	0.33
3b1f NC	17.6	0.39
Total Systematic Uncertainty	54.9	1.2
Total Uncertainty	313.7	7.0

Table 10.4: Magnitude of the components of the uncertainty on the background estimate in the *four-tag* Signal Region of the ggF channel for 2018 data. Total statistical uncertainty is the quadrature sum of the bootstrap and *two-tag* Poisson uncertainties.

SR 18 4b (Bkgd Yield = 7720.0)	**Counts**	**Relative (%)**
Total Statistical Uncertainty	483.0	6.3
Shape Q_N	−141.3	−1.8
Shape Q_W	−55.8	−0.72
Shape Q_S	7.0	0.091
Shape Q_E	−87.1	−1.1
3b1f NC	42.6	0.55
Total Systematic Uncertainty	180.4	2.3
Total Uncertainty	515.6	6.7

11

Results and Discussion

The full Run 2 ($126\,\text{fb}^{-1}$) ATLAS $HH \to b\bar{b}b\bar{b}$ analysis produced a number of important and novel limits on HH production and Standard Model couplings [1]. These constraints help elucidate the Higgs and electroweak sectors of the Standard Model, key components of our understanding of the workings of the universe. In this chapter, the results of the analysis are shown, described and discussed.

11.1 Checks of Fit Validity

As described in Chapter 9, uncertainties are encoded into the likelihood model using nuisance parameters (NP). These nuisance parameters, as well as the parameter of interest, are fit to data as part of the process of deriving limits. It is important to check the validity of the fit to ensure the post-fit nuisance parameters accurately describe the uncertainty they represent and that there is no pathological behaviour.

The m_{HH} distributions of the background-only model after the fit to 2016 data in the Signal Region of the ggF channel are shown in Figure 11.1. Overall, agreement is good, with no significant deviations visible. This is also true for the 2017 and 2018 data-taking years, the plots for which are shown in Figures F.1 and F.2 in Appendix F. The equivalent plots for all data-taking years in the VBF channel are shown in

Figure 11.2. These also display good levels of agreement.

The pull and impact of the 20 highest impact nuisance parameters are shown, ordered by impact, in Figure 11.3. This is for the combined ggF and VBF channels. Appendix E provides an overview of the naming scheme used for the NPs shown on the y-axis. To summarise, those NPs including `NC_3b1f` in their name correspond to the background non-closure uncertainty derived in the `3b1f` region; `CR12_shape` corresponds to the shape uncertainty on the background; `JET_JER` is associated with the Jet Energy Resolution; `THEO` is associated with the theoretical uncertainties; and `FT_EFF` is associated with the calibration of the flavour-tagging algorithms. The highest impact NP `alpha_THEO_XS_SCALEMTop_ggf` is associated with the uncertainty on the ggF cross-section due to the choice of scale and top mass schemes. `alpha_THEO_ACC_PS_ggf` and `alpha_THEO_ACC_SCALE_ggf` are the uncertainties on the signal acceptance due to the choice of parton shower model and scale respectively. The salient points to draw from Figure 11.3 are: firstly, the highest impact NPs are associated with the theoretical ggF cross-section value and the `3b1f` and shape uncertainties on the background. This indicates these uncertainties likely caused the largest reduction in sensitivity. Secondly, the pulls of the vast majority of highest impact NPs lie within the $\pm 1\sigma$ bounds, and none are pulled beyond $\pm 1\sigma$ when accounting for their uncertainty bands. Thirdly, no significant overconstraining of NPs is observed. Mild pulls and constraining of the background estimate NPs is observed; however, this can be explained by the fact the background weights are derived in a control region and new information could exist in the signal region that could constrain the NPs. The findings of second and third points provide confidence that the NPs are accurately accounting for the uncertainty they represent. The pulls and impacts of lower ranked NPs are shown in Figure F.3 in Appendix F.[1]

[1]Two of the NPs associated with the shape uncertainties on the 2016 background estimate – `alpha_CR12_shape_E_ggf_16` and `alpha_CR12_shape_S_ggf_16` – are pulled by $\sim 2\sigma$ in Figure F.3. It was found that these NPs have a large influence over the m_{HH} peak, a very challenging part of the distribution to model, resulting in their large pulls. This was not deemed a problem as the origin was well understood.

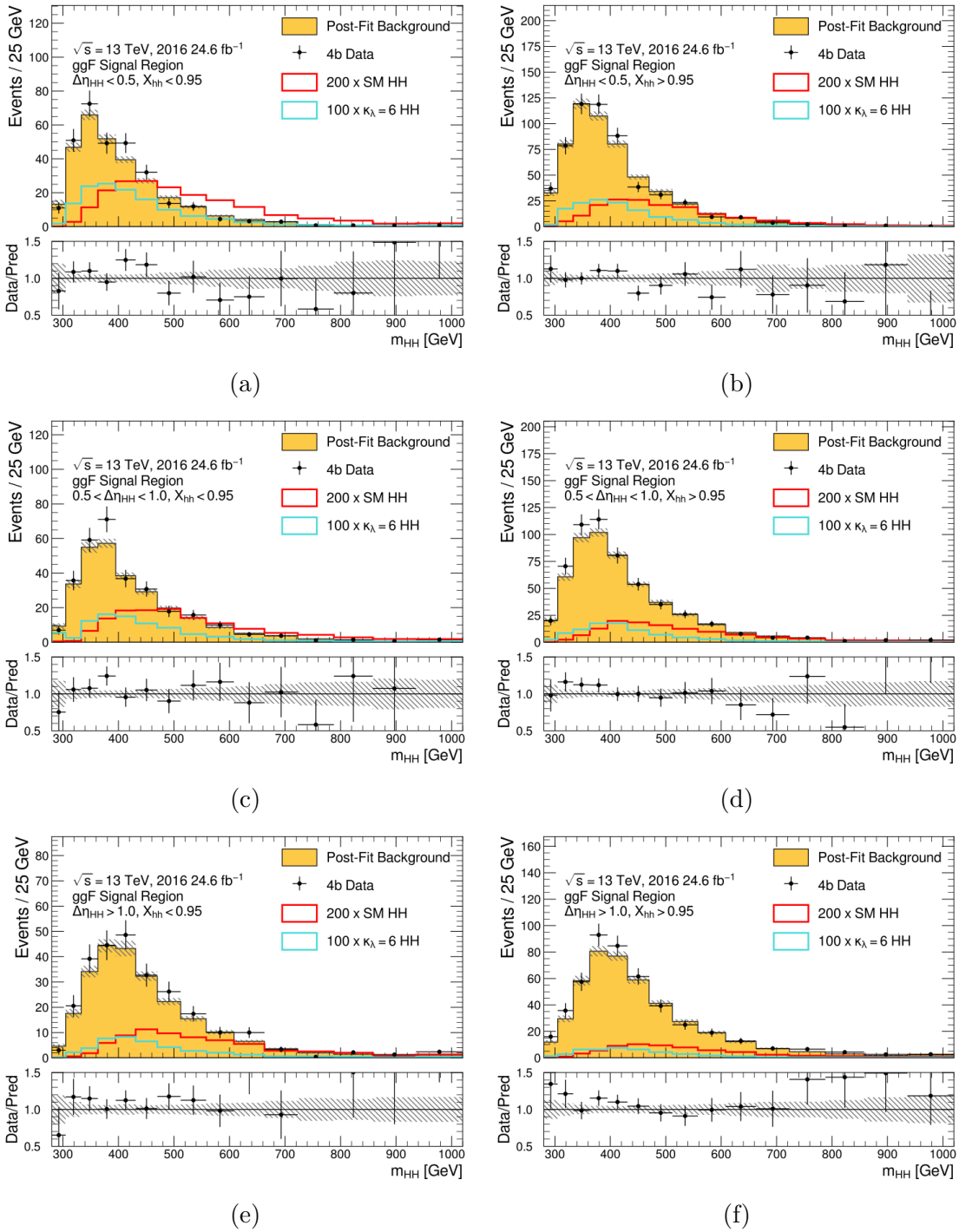
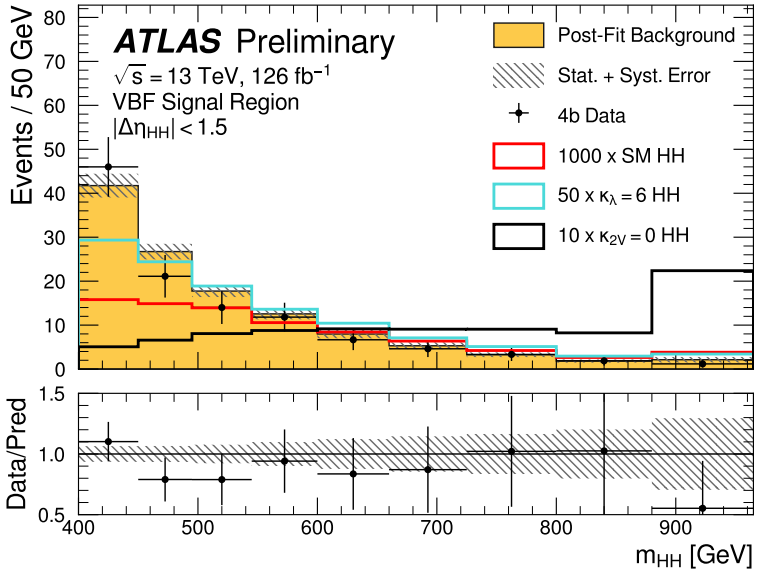
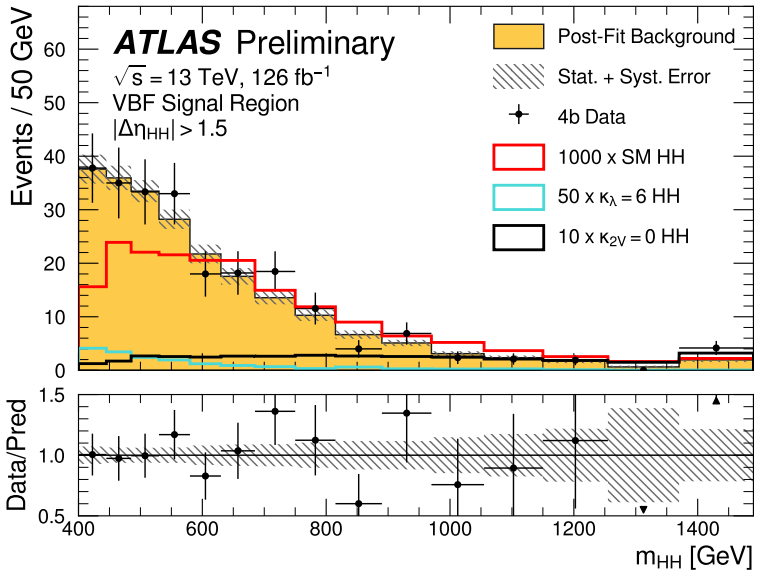


Figure 11.1: Plots of the m_{HH} distributions of the background-only model (yellow) after the maximum likelihood estimator fit to data (black) in the $|\Delta\eta_{HH}|$–X_{HH} categories in the ggF Channel for the 2016 data-taking year. The distributions for two important HH production signal models – SM and $\kappa_\lambda = 10$ – are shown for reference. The signal distributions are scaled as to be visible next to the far more dominant background.



(a) $|\Delta\eta_{HH}| < 1.5$



(b) $|\Delta\eta_{HH}| > 1.5$

Figure 11.2: Plots of the m_{HH} distributions of the background-only model (yellow) after the maximum likelihood estimator fit to data (black) in the $|\Delta\eta_{HH}|$ categories in the VBF Channel for all data-taking years. The distributions for three important HH production signal models – SM, $\kappa_{2V} = 0$ and $\kappa_\lambda = 10$ – are shown for reference. The signal distributions are scaled as to be visible next to the far more dominant background.

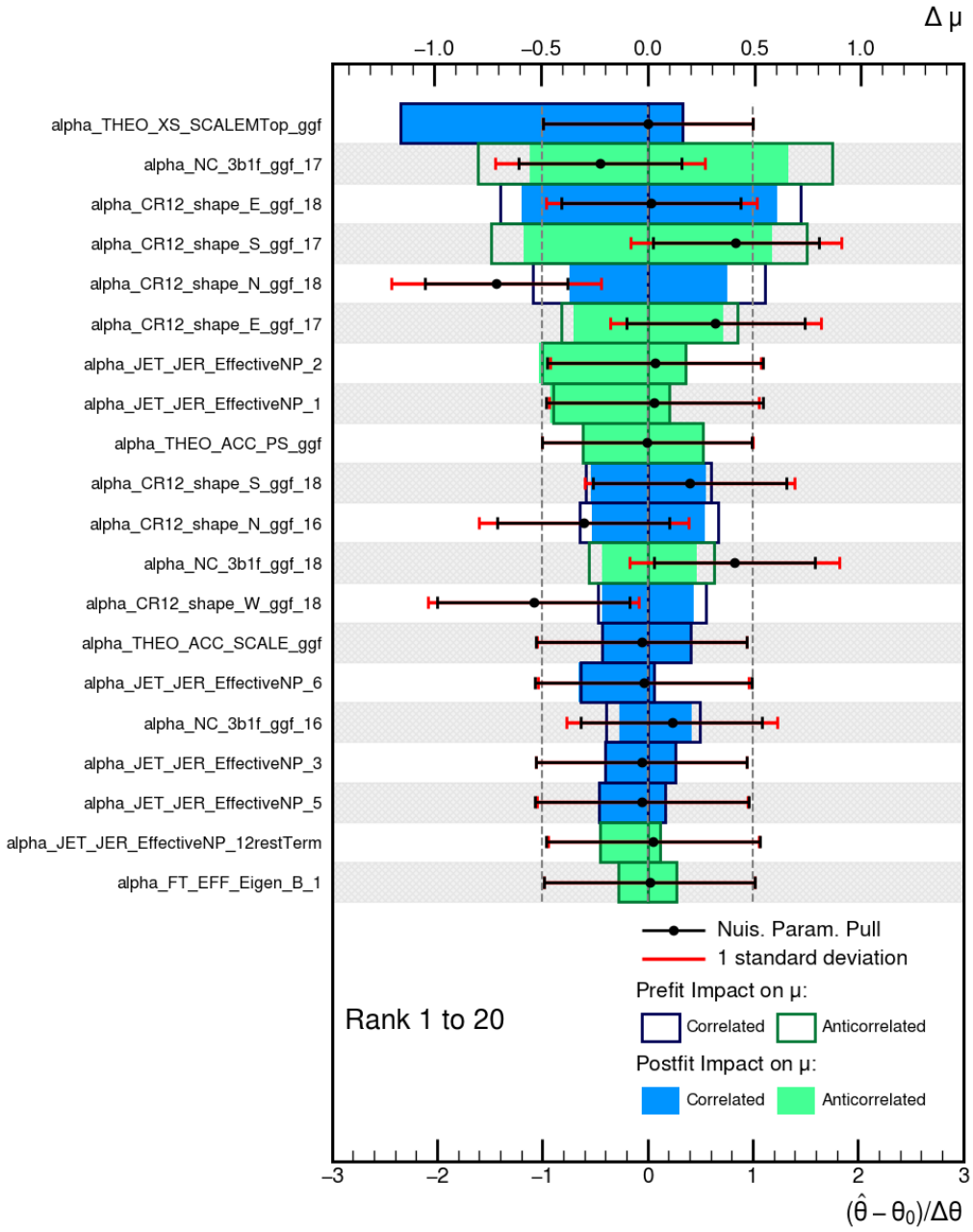


Figure 11.3: Two-in-one plot showing the pulls and impacts of the analysis nuisance parameters (see Section 9.6 for definitions). This plot has two x-axes: impact ($\Delta\mu$) is measured on the upper x-axis and pull ($(\hat{\theta} - \theta_0)/(\Delta\theta)$) is measured on the lower x-axis. The y-axis shows the 20 highest impact NPs in the analysis in descending order, with highest impact at the top. The green and blue bars corresponds to the impact, so should be read using the upper x-axis. If the impact of the NP is correlated with μ, then the bar is coloured blue. If it is anti-correlated, then the bar is coloured green. Additionally, the pre-maximum likelihood estimator fit impact value is shown as an unfilled bar, whilst the post-fit impact is shown as a filled bar. The black and red points with error bars correspond to the pull, so should be read using the lower x-axis. The black point shows the value of the pull, with the black error bar indicating the post-fit error on the NP. The red error bar shows the one standard deviation bounds, and provides a yardstick with which to judge the constraining of the NP.

The correlations between the nuisance parameters describing the shape uncertainty on the background estimate and the `3b1f` non-closure uncertainty in the combined VBF and ggF channels for a SM HH signal and background fit to data are shown in Figure 11.4. No significant correlations were observed except for background shape NPs associated with the same data-taking year. This is expected given NPs associated with the same year are adjusted simultaneously to fit the model to data for that year. The correlations between the other nuisance parameters in the model were negligible. The full correlation matrix is shown in Figure F.4 in Appendix F.

No pathologies were observed in the post-fit model's m_{HH} distributions or the pulls and correlations of its nuisance parameters. This provides confidence in the validity of the fit.

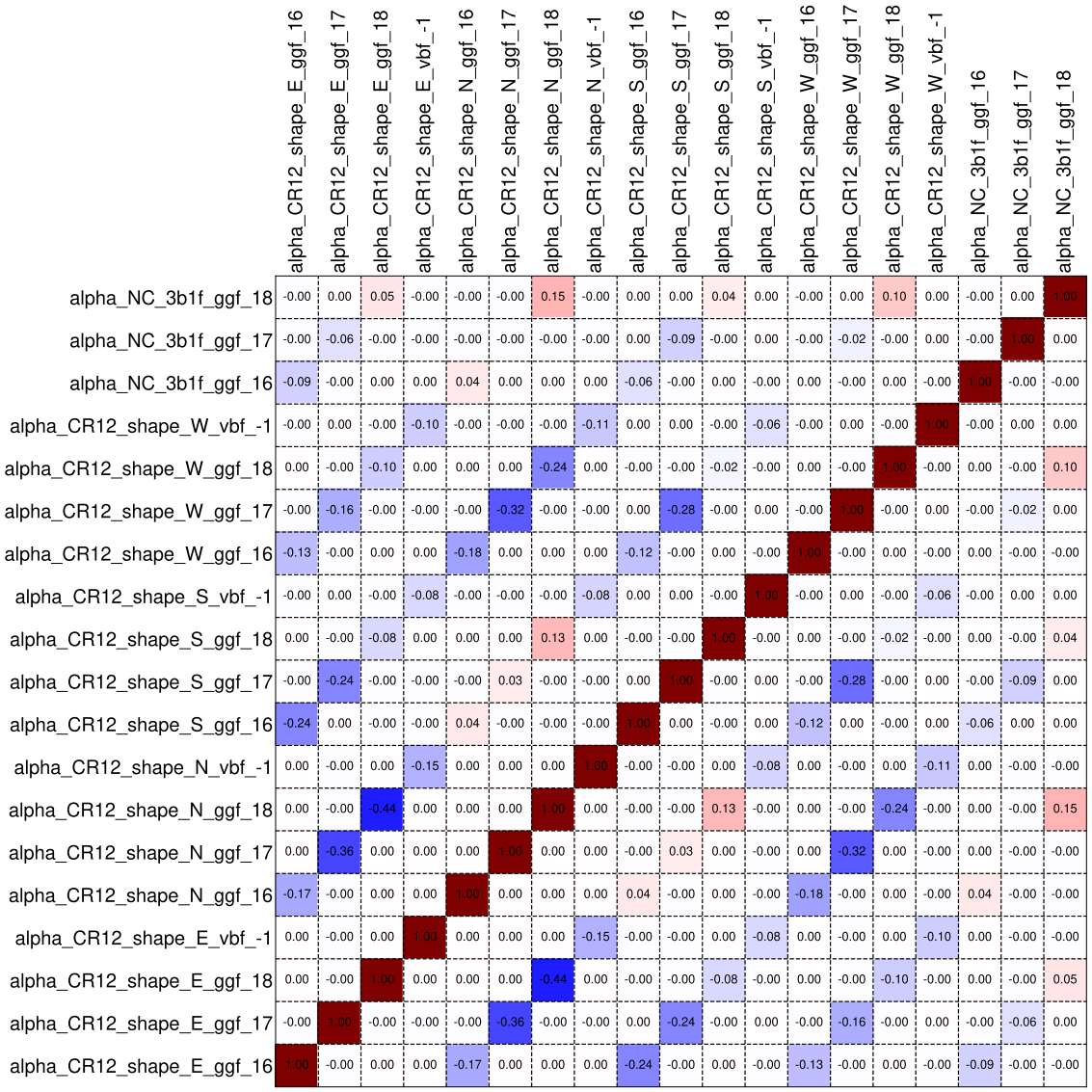


Figure 11.4: Correlations between the nuisance parameters describing the shape uncertainties on the background estimate and the `3b1f` non-closure uncertainty for a Standard Model *HH* signal and background fit to data for the combined ggF and VBF channel.

11.2 Analysis Results

11.2.1 Limit on Standard Model *HH* Production

The observed (expected) limit on *HH* production when considering the ggF and VBF *HH* production modes was set at $5.4 \times \sigma_{SM}$ ($8.1 \times \sigma_{SM}$) at the 95% confidence level (CL), as shown in Table 11.1. The observed value was just over one standard deviation away from the expected value, which is explained by the slight downward fluctuation in the data in the tails of m_{HH} distribution for some categories, as seen in Figure 11.1a. Table 11.1 shows the individual contributions from the ggF and VBF modes. The sensitivity is driven by ggF, which is expected given it has a far higher cross-section in the SM. The new limit represents an approximately factor of 2.5 improvement on the older $\mathcal{L}_{INT} = 36.1\,\text{fb}^{-1}$ ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analysis limit of $12.9 \times \sigma_{SM}$ ($20.7 \times \sigma_{SM}$) [2]. It is also one of the tightest constraints on *HH* production in the $b\bar{b}b\bar{b}$ channel in the world, alongside the CMS full Run 2 result, which was set at $3.9 \times \sigma_{SM}$ ($7.8 \times \sigma_{SM}$) [110].

The two largest sources of uncertainty originate from the background estimation procedure and the theoretical predictions. Accounting only for the statistical uncertainty on the *HH* MC samples and *four-tag* and reweighted *two-tag* data, the limit was $6.0 \times \sigma_{SM}$. By including the uncertainties associated with the background estimation procedure, the limit raised to $7.1 \times \sigma_{SM}$. Also including the uncertainties associated with the theoretical predictions was the primary driver of the raising the limit to its final value of $8.1 \times \sigma_{SM}$.

11.2.2 1D Limits on *HH* Coupling Modifiers

The observed (expected) constraints on the trilinear Higgs self-coupling modifier, κ_λ, were set at $-3.9 < \kappa_\lambda < 11.1$ ($-4.6 < \kappa_\lambda < 10.8$) at the 95% CL. The SM hypothesis, $\kappa_\lambda = 1$, was not excluded at any significant level. The upper limit set

Table 11.1: The observed and expected upper limits on Standard Model *HH* production signal strength ($\mu = \sigma/\sigma_{SM}$) set by the ATLAS full Run 2 $HH \to b\bar{b}b\bar{b}$ analysis. The expected limit is shown with its one and two standard deviation error bounds ($\pm 1\sigma$ and $\pm 2\sigma$). Limits accounting for just the individual ggF and VBF *HH* production modes are shown alongside the full ggF-VBF combination. All limits are set at the 95% confidence level.

***HH* Mode(s)**	**Obs**	$\mathbf{-2\sigma}$	$\mathbf{-\sigma}$	**Exp**	$\mathbf{+\sigma}$	$\mathbf{+2\sigma}$
ggF	5.5	4.4	5.9	8.2	12.4	19.6
VBF	130.5	71.6	96.1	133.4	192.9	279.3
ggF, VBF	**5.4**	4.3	5.8	**8.1**	12.2	19.1

on the ggF and VBF *HH* production cross-section versus κ_λ is shown in Figure 11.5. The upper limit for non-SM κ_λ signals is higher than for the SM point because the events are softer and so the acceptance is lower. The individual contributions from the ggF and VBF channels are shown, demonstrating that a significant improvement on the limit, and, therefore, the constraints on κ_λ, comes from the VBF channel. This is a unique analysis feature, which I pioneered.[2]

The κ_λ constraints are the first ATLAS limits on this coupling in the $b\bar{b}b\bar{b}$ channel. It is also one of the world's tightest κ_λ constraints in this channel, alongside the full Run 2 CMS $HH \to b\bar{b}b\bar{b}$ analysis result: $-2.3 < \kappa_\lambda < 9.4$ ($-5.0 < \kappa_\lambda < 12.0$) [110].

The observed (expected) constraints on the *HHVV* vertex coupling modifier, κ_{2V}, was set at $-0.03 < \kappa_{2V} < 2.11$ ($-0.05 < \kappa_{2V} < 2.12$) at the 95% confidence level. Figure 11.6 shows the upper limit set on ggF and VBF cross-section versus κ_{2V} at the 95% CL. The upper limit for non-SM κ_{2V} signals is lower than for the SM point because the events are harder and so the acceptance is higher. The vast majority of sensitivity comes from the VBF channel, given it was optimised to maximise sensitivity to VBF *HH* production and only VBF has sensitivity to κ_{2V}.

[2] As described in Chapter 8.

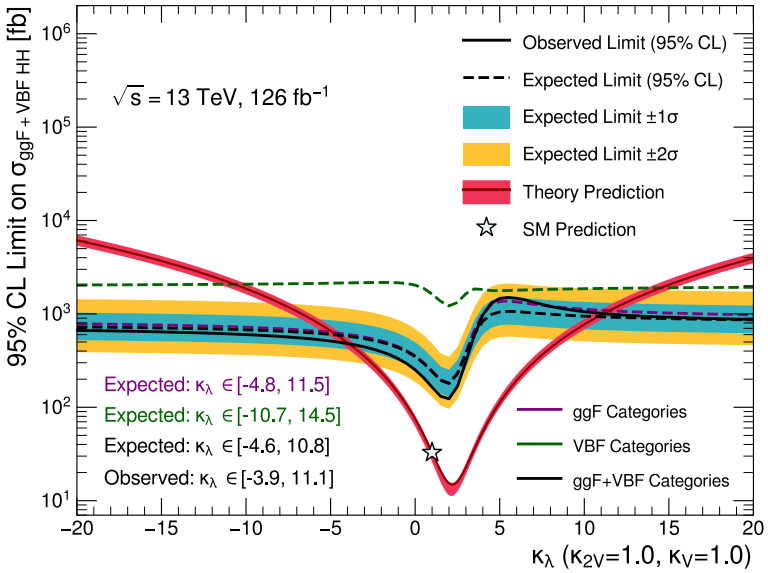


Figure 11.5: Observed (solid) and expected (dashed) 95% CL upper limit on ggF and VBF *HH* production cross-section for various values of the trilinear Higgs self-coupling modifier, κ_λ, with the individual contributions from the ggF channel (purple), VBF channel (green) and their combination (black) shown alongside the theoretical values (red). The corresponding constraints on κ_λ are shown in the bottom left corner. All other coupling modifiers are set to their SM value of one.

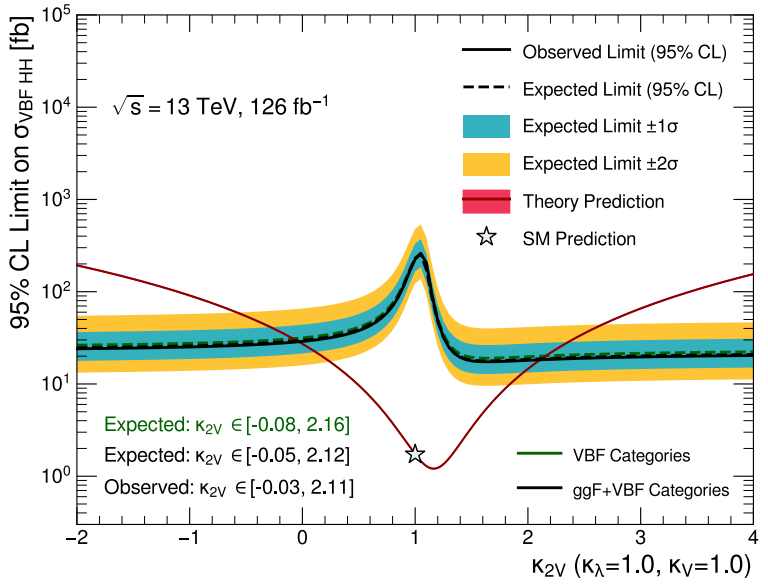


Figure 11.6: Expected 95% CL upper limit on ggF and VBF *HH* production cross-section for various values of the *HHVV* coupling modifier, κ_{2V}, with the results from the VBF channel (green) and the ggF-VBF channel combination (black) shown alongside the theoretical value (red). The corresponding constraints on κ_{2V} are shown in the bottom left corner. All other coupling modifiers are set to their SM value of one.

This κ_{2V} constraint is a significant improvement on the previous VBF $HH \to b\bar{b}b\bar{b}$ analysis, which set a constraint of $-0.43 < \kappa_{2V} < 2.56$ ($-0.55 < \kappa_{2V} < 2.72$). It is also the world's strongest constraint on κ_{2V} in the *resolved* $b\bar{b}b\bar{b}$ channel [3]. In general, the world's best κ_{2V} constraints were set in the *boosted* $b\bar{b}b\bar{b}$ channel of the full Run 2 CMS $HH \to b\bar{b}b\bar{b}$ analysis, at $0.62 < \kappa_{2V} < 1.41$ ($0.66 < \kappa_{2V} < 1.37$) [111].[3]

11.2.3 2D Limits on *HH* Coupling Modifiers

Alongside one-dimensional constraints on κ_λ and κ_{2V}, two-dimensional constraints on κ_{2V}–κ_λ and κ_{2V}–κ_V space were also produced. These results were pursued because many Beyond the Standard Model theories predict differences in multiple coupling values. The Standard Model hypothesis, $\kappa_V = \kappa_{2V} = \kappa_\lambda = 1$, was found to be within the boundaries of the limits set.

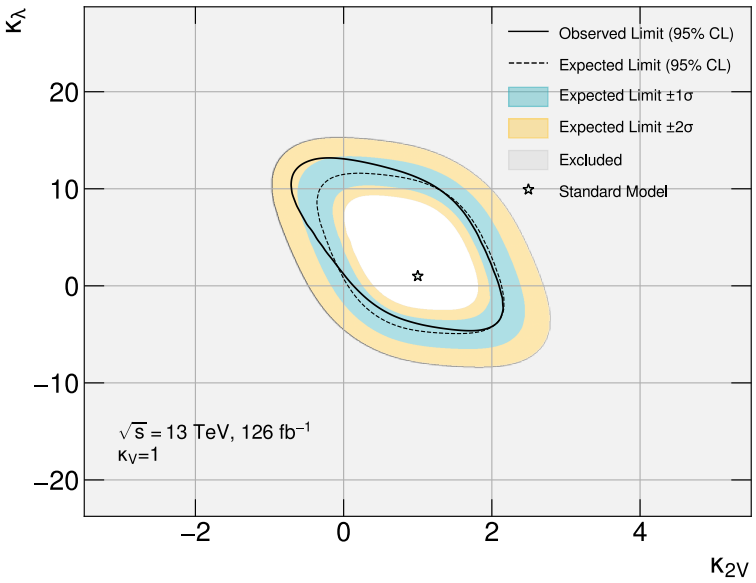
Figure 11.7a shows a two-dimensional limit on ggF and VBF *HH* production for various points in κ_λ–κ_{2V} space. This result benefits from the ggF-VBF channel combination, as sensitivity to κ_{2V} is driven by the VBF channel and sensitivity to κ_λ is driven by the ggF channel. This is one of the world's first joint constraints on κ_λ and κ_{2V}, alongside the result from the aforementioned equivalent CMS analysis [110].

Figure 11.7b shows the same plot for various points in κ_{2V}–κ_V space. [4] The unique shape can be understood by looking back at Equation 1.10 and noting the amplitude increases when $\kappa_{2V} \neq \kappa_V^2$. As a result, points around the curve $\kappa_{2V} = \kappa_V^2$ tend not to be excluded as the amplitude is not large enough. As κ_{2V} and κ_V get larger, the precision required for the cancellation is greater so the effect is smaller. The κ_{2V}–κ_V limit provides a unique and comprehensive probe of the interaction of Higgs bosons

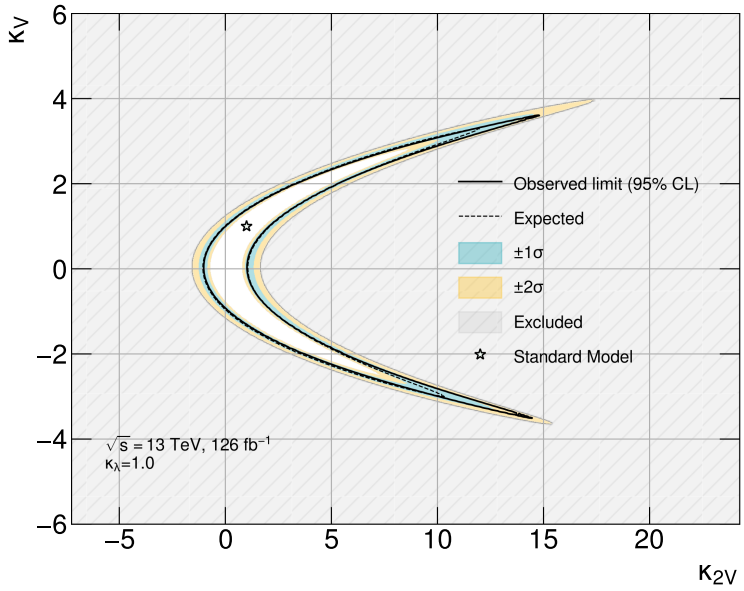
[3] The CMS analysis benefited substantially from a double-b-jet tagger; the ATLAS equivalent is currently in development.

[4] The $H \to b\bar{b}$ branching ratio changes by a non-negligible amount with κ_V; however, this was not accounted for when deriving limits on κ_V due to an error.

and the carriers of the weak force. It is the first joint constraint on these couplings in the world.



(a) κ_{2V} vs. κ_λ



(b) κ_{2V} vs. κ_V

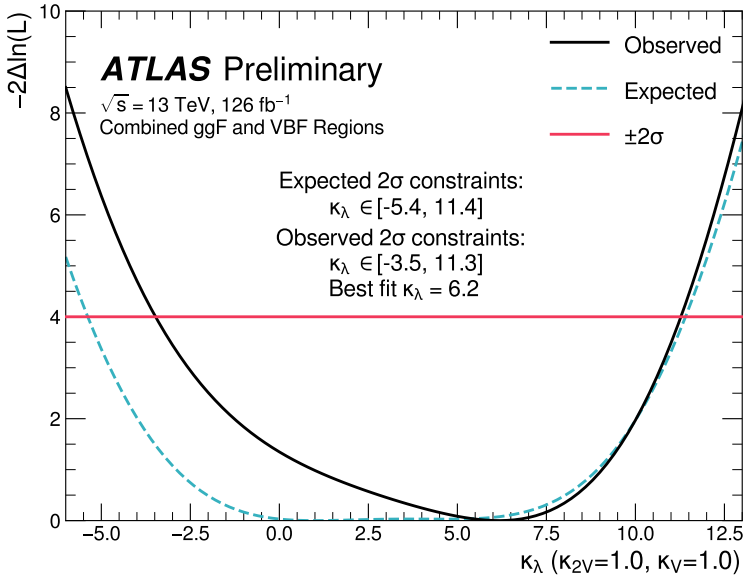
Figure 11.7: The observed (solid black) and expected (dashed black) constraints on ggF and VBF *HH* production at the 95% confidence level (CL) for key points in κ_{2V}–κ_λ (top) and κ_{2V}–κ_V (bottom) space. The one and two standard deviation ($\pm1\sigma$ and $\pm2\sigma$) error bands on the expected limit are shown in blue and yellow respectively. The Standard Model (SM) point, $\kappa_V = \kappa_{2V} = \kappa_\lambda = 1$, is shown as a star. The region excluded at the 95% CL is shaded in grey. All other coupling modifiers are set to their SM value of one.

11.2.4 Log Likelihood Limits on *HH* Coupling Modifiers

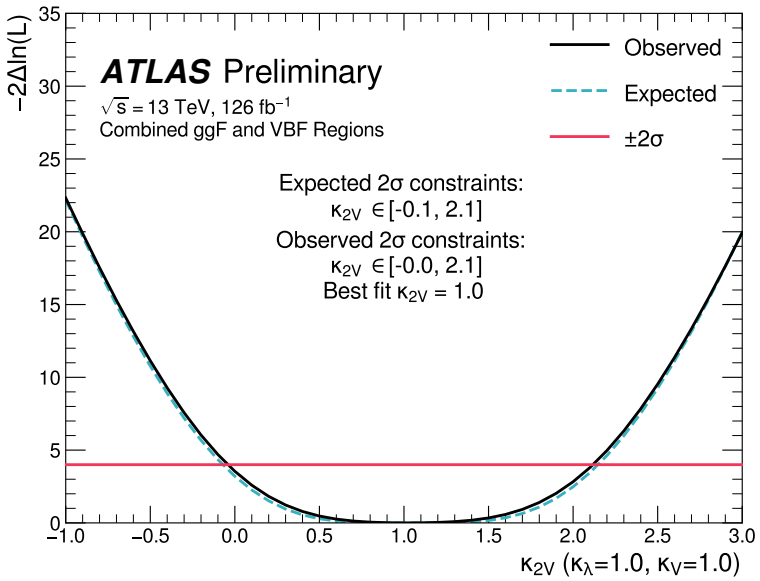
To complement the aforementioned results, equivalent constraints were extracted using the log likelihood ratio (LLR) directly. The LLR of a specific κ_λ hypothesis to the best fit hypothesis is plotted for a range of κ_λ values around the SM prediction, $\kappa_\lambda = 1$. The best fit κ_λ value is the one that minimises the LLR. The value of the LLR increases rapidly as κ_λ is moved away from the best fit value and the hypotheses become more incompatible. The κ_λ values that correspond to an LLR two standard deviations (2σ) or more away from the minimum are excluded. This is approximately equivalent to the 95% CL.

The observed (expected) 2σ constraints placed on κ_λ using the LLR directly are $-3.5 < \kappa_\lambda < 11.3$ ($-5.4 < \kappa_\lambda < 11.4$). The equivalent constraints on κ_{2V} are $0.0 < \kappa_{2V} < 2.1$ ($-0.1 < \kappa_{2V} < 2.1$). Figure 11.8 shows the value of the LLR for various κ_λ and κ_{2V} hypotheses. The difference between the expected and observed constraints in Figure 11.8a stems from a slight excess at low m_{HH}, close to where the relatively soft large positive κ_λ signal peaks, as seen in Figure 11.1a. As a result, the fit finds the best fit value of κ_λ to be 6.2, disfavouring the harder SM and large negative κ_λ hypotheses. The observed (expected) 1σ and 2σ constraints placed on κ_{2V}–κ_λ and κ_{2V}–κ_V space are shown in Figures 11.9a – 11.9c.[5] The SM hypothesis of $\kappa_{2V} = \kappa_\lambda = 1$ is disfavoured at 1σ, a relatively low level of significance. These results are different to those derived using the upper limit on *HH* cross-section, which is expected given the statistical methods differ. Together they offer a comprehensive description of the statistical statements the analysis can make.

[5] See Footnote 4.

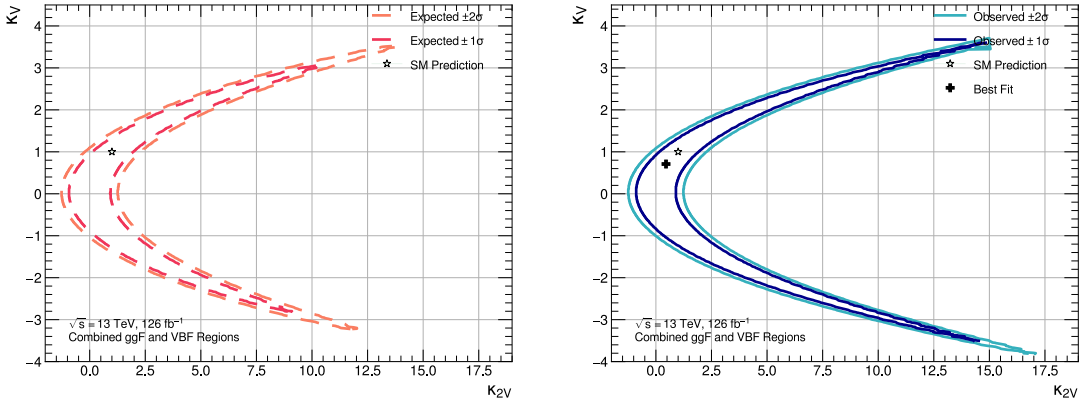


(a) κ_λ



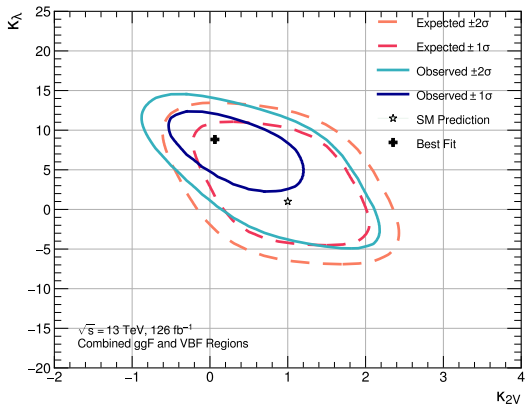
(b) κ_{2V}

Figure 11.8: Observed (solid dark blue) and expected (dashed light blue) values of the log likelihood ratio (LLR) of different κ_λ (top) and κ_{2V} (bottom) hypothesis compared to the maximum likelihood estimator hypothesis for the combined ggF and VBF channels. The two standard deviation (2σ) LLR is shown by the red horizontal line. The best fit value of the coupling modifier along with the 2σ constraints are shown in the centre.



(a) κ_{2V}–κ_V (expected)

(b) κ_{2V}–κ_V (observed)



(c) κ_{2V}–κ_λ

Figure 11.9: Observed (solid blue) and expected (dashed red) one (light) and two (dark) standard deviation exclusion regions in κ_{2V}–κ_V and κ_{2V}–κ_λ space. These are derived using the log likelihood ratio (LLR) of specific coupling modifier hypotheses to the maximum likelihood estimator (MLE) hypothesis for the combined ggF and VBF channels. The Standard Model prediction, $\kappa_V = \kappa_{2V} = \kappa_\lambda = 1$, is shown as a star and the MLE value as a cross.

12 Summary and Outlook

12.1 Summary

This thesis focuses on the full Run 2 ATLAS $HH \to b\bar{b}b\bar{b}$ analysis, which I contributed towards throughout my PhD. This analysis probed the simultaneous production of two Higgs bosons (HH) in the $b\bar{b}b\bar{b}$ final state using $\mathcal{L} = 126\,\text{fb}^{-1}$ of 13 TeV proton-proton collision data from the ATLAS experiment on the Large Hadron Collider (LHC). The aim was to investigate the Higgs sector of the Standard Model (SM) by setting constraints on SM HH production, the trilinear Higgs self-coupling, λ, and the HVV and $HHVV$ vertex couplings, c_V and c_{2V}.

An observed (expected) upper limit on HH production via Gluon-Gluon Fusion (ggF) and Vector Boson Fusion (VBF) was set at 5.4 (8.1) times the SM prediction at the 95% confidence level (CL). Two different statistical methods were used to set constraints on the coupling modifiers κ_λ, κ_V and κ_{2V}.[1] Using a comparison of the theoretical cross-section to the upper limit on HH production for different modifier values, κ_λ was constrained to $-3.9 < \kappa_\lambda < 11.1$ ($-4.6 < \kappa_\lambda < 10.8$) and κ_{2V} to $-0.03 < \kappa_{2V} < 2.11$ ($-0.05 < \kappa_{2V} < 2.12$) at the 95% CL. Measuring the log likelihood ratio directly, κ_λ was constrained to $-3.5 < \kappa_\lambda < 11.3$ ($-5.4 < \kappa_\lambda < 11.4$)

and κ_{2V} to $0.0 < \kappa_{2V} < 2.1$ ($-0.1 < \kappa_{2V} < 2.1$) at 2σ. Limits were also placed on a range of values in κ_{2V}-κ_λ and κ_{2V}-κ_V space at the 95% CL.

12.2 Outlook

The results of this analysis are a significant improvement on predecessor ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analyses [2, 3] and are also competitive with the equivalent CMS $HH \rightarrow b\bar{b}b\bar{b}$ analysis [110]. This is thanks to the increased dataset size, improved ATLAS algorithms and innovative optimisation by the $HH \rightarrow b\bar{b}b\bar{b}$ analysis team. Looking to the future, there are a number of lessons that can be learnt from the full Run 2 ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analysis that will inform the next generation of HH analyses. A number of new techniques were pioneered; for example, the trigger bucketing strategy can help solve ambiguity issues over the application of trigger scale-factors, and the Neural Network-based background estimation can help build background estimates when Monte Carlo simulation is not available. To focus on one innovation in particular, the introduction of a channel optimised for Vector Boson Fusion (VBF), on top of the nominal gluon-gluon Fusion (ggF) channel, offered numerous benefits. It allowed the analysis to extract a broader range of physics result, i.e., competitive limits on both κ_λ and κ_{2V}, as well as two-dimensional limits in κ_λ vs. κ_{2V} space. Additionally, it was discovered that VBF provides a significant level of additional sensitivity to the trilinear Higgs self-coupling. Thus, combining independent ggF and VBF-optimised channels provides tighter constraints on κ_λ than when optimising for ggF alone, which was the previous strategy. There is scope for improvement too, as the VBF channel was not optimised for Beyond the Standard Model (BSM) VBF κ_λ-variations. A study into the benefits of using VBF triggers in the analysis was also conducted, finding they offered little to no additional sensitivity. These studies provide important information to future $HH \rightarrow b\bar{b}b\bar{b}$ analyses that may explore the use of VBF triggers and provides feedback to researchers designing future

VBF triggers with *HH* production in mind.

After the end of Run 3 in 2026, the LHC will be upgraded to the High Luminosity LHC (HL-LHC). This upgrade, due to be completed by 2029, aims to increase the peak instantaneous luminosity to as high as $\sim 7.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. For comparison, the peak instantaneous luminosity during Run 2 was $\sim 2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. As a result, the dataset available to physicists should increase in size by a factor of $\sim$10, to $\sim 3000\,\mathrm{fb}^{-1}$. One of the primary physics goals of the HL-LHC is to probe the properties of the Higgs boson as a means to understanding its link to the process of Electroweak Symmetry Breaking. As such, the study of *HH* production, with its dependence on the trilinear Higgs self-coupling, λ, is central to the HL-LHC programme. Ref. [112] provides a comprehensive overview of the *HH* results that can be expected to be obtained at the HL-LHC based on projections from the partial ($\mathcal{L}_{INT} = \sim 36\,\mathrm{fb}^{-1}$) Run 2 ATLAS and CMS analyses. These projections predict that Standard Model (SM) *HH* production should be observed at a significance level of 3.5σ with ATLAS and 2.6σ with CMS. This gives a combined significance of 4.5σ, close to the level required to confirm a discovery. Projections for the trilinear Higgs self-coupling modifier, κ_λ, predict constraints of $-0.4 < \kappa_\lambda < 7.3$ with ATLAS and $-0.2 < \kappa_\lambda < 3.6$ with CMS at the 95% CL. This gives a combined constraint of $0.1 < \kappa_\lambda < 2.3$. These results are based on a combination of the following *HH* decay channels: $b\bar{b}b\bar{b}$, $b\bar{b}\tau^-\tau^+$, $b\bar{b}\gamma\gamma$, $b\bar{b}VV^*$ and $b\bar{b}ZZ(4l)$. Only projected statistical uncertainties are accounted for. All other assumptions are outlined in Section *8.2 Double Higgs Boson Production Measurements* of Ref. [112]. It is worth noting that the recent full Run 2 *HH* combination from ATLAS set constraints on κ_λ of $-1.0 < \kappa_\lambda < 6.6$ at the 95% CL [20]. This indicates that the aforementioned HL-LHC κ_λ projections for ATLAS are very likely be significantly improved on.

[1] Coupling modifiers are the ratio of the coupling to its Standard Model prediction, i.e., $\kappa_X = c_X/c_{SM}$.

The HL-LHC programme will end by mid-to-late-2030, marking the finishing point of the LHC era at CERN. Looking to the future, plans for a new collider are being drawn up, with proposals broadly falling in one of two categories. Either a hadron-hadron circular collider will be built in order to explore the energy frontier, or an electron-positron linear collider will be built in order to perform ultra-high precision measurements. In the former camp, the leading proposal is the hadron-hadron Future Circular Collider (FCC-hh), a 100 km synchrotron projected to reach energies as high as $\sqrt{s} = 100\,\text{TeV}$ [113, 114].[2] In the latter camp, two leading proposals are the International Linear Collider (ILC) [116] and the Compact Linear Collider (CLIC) [117]. Ref. [112] provides a comprehensive overview of the projected *HH* results obtainable at a number of proposed future colliders. These will be briefly summarised. Starting with the FCC-hh, Standard Model (SM) *HH* production is predicted to be observed at a significance level of over 5σ, the standard required for discovery. In turn, the Higgs self-coupling is projected to be measured at a precision of 5%.[3] Additionally, the *HHVV* vertex coupling, c_{2V}, is projected to be measured to a precision of better than 1%. Regarding the linear colliders, SM *HH* production is projected to be observed with a significance greater than 5σ with both the ILC and CLIC. The observation is projected to be via the Higgstrahlung production process ($e^+e^- \to ZHH$). The Higgs self-coupling is projected to be measured at a precision of 27% for the ILC at 500 GeV and at 10% at 1 TeV. These are the energies projected for the first and second phases of the accelerator. Similarly, the precision is 36% for CLIC at 1.5 TeV and 9% at 3 TeV. Overall, future colliders will allow researchers to probe the *HH* and the Higgs self-coupling at a vastly improved level of precision, thereby elucidating our understanding of the Higgs boson and its connection to the process of Electroweak Symmetry Breaking.

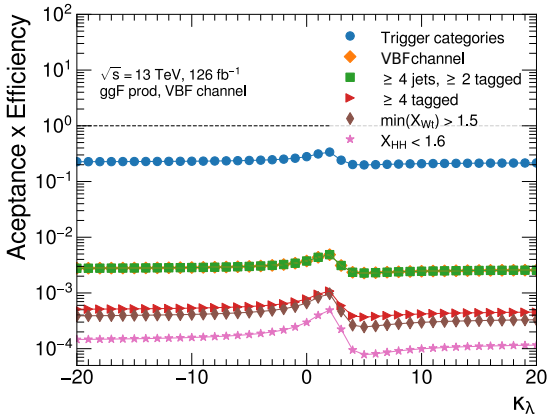
[2] An electron-positron version of the FCC (FCC-ee) [113, 115] has also been proposed, but is projected to have a centre-of-mass energy below the *HH* mass threshold.

[3] All quoted levels of precision are at the 68% Confidence Level.

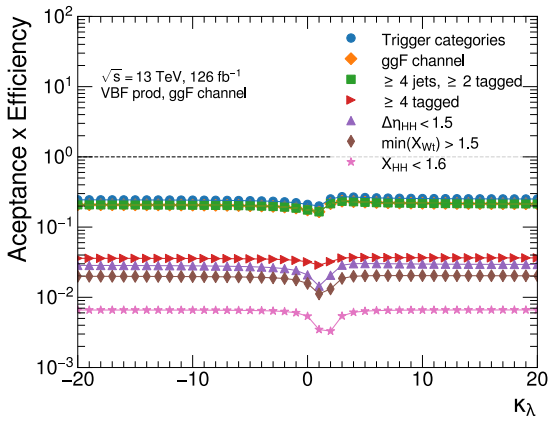
A

Additional Acceptance × Efficiency Plots

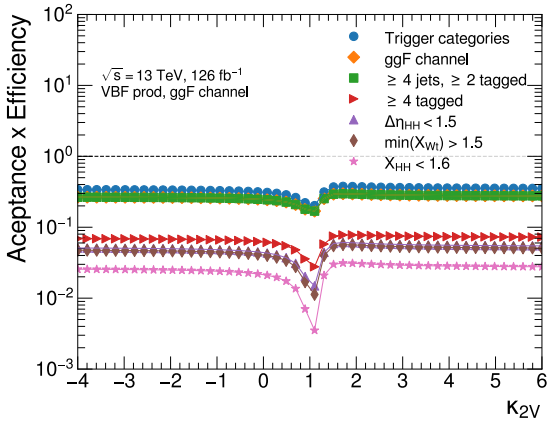
Figure A.1 shows the values of acceptance times efficiency (A × E) for each step in the full Run 2 ATLAS $HH \to b\bar{b}b\bar{b}$ analysis event selection for κ_λ variations of ggF HH Monte Carlo (MC) simulation in the VBF channel, κ_λ variations of VBF HH MC simulation in the ggF channel and κ_{2V} variations of VBF HH MC simulation in the ggF channel. The selection is described in Chapter 4.



(a) κ_λ ggF HH, VBF channel



(b) κ_λ VBF HH, ggF channel



(c) κ_{2V} VBF HH, ggF channel

Figure A.1: Acceptance times efficiency for each step in the analysis event selection for κ_λ variations of ggF HH Monte Carlo (MC) simulation in the VBF channel (top), κ_λ variations of VBF HH MC simulation in the ggF channel (bottom left) and κ_{2V} variations of VBF HH MC simulation in the ggF channel (bottom right). Each step in the selection is represented by a different coloured line, with the legend in the top right corner of the plot showing the steps in order.

B

Analysis Selection Cutflows

This appendix contains tables showing the event yield at each stage in the event selection, otherwise known as *cutflows*, for the full Run 2 ATLAS $HH \to b\bar{b}b\bar{b}$ analysis. The selection is described in Chapter 4. Cutflows are included for data and ggF and VBF *HH* Monte Carlo (MC) simulation in the ggF and VBF channels. The list of cutflows is outlined below.

Data (2016–18, $126\,\text{fb}^{-1}$):

- Table B.1a: *four-tag* events in the ggF channel.
- Table B.1b: *two-tag* events in the ggF channel.
- Table B.2a: *four-tag* events in the VBF channel.
- Table B.2b: *two-tag* events in the VBF channel.

ggF *HH* MC simulation (normalised to $126\,\text{fb}^{-1}$):

- Tables B.3a: *four-tag* events in the ggF channel for SM ggF *HH* signal.

- Tables B.3b: *four-tag* events in the ggF channel for $\kappa_\lambda = 10$ ggF *HH* signal.

- Tables B.4a: *four-tag* events in the VBF channel for SM ggF *HH* signal.

- Tables B.4b: *four-tag* events in the VBF channel for $\kappa_\lambda = 10$ ggF *HH* signal.

VBF *HH* MC simulation (normalised to $126\,\text{fb}^{-1}$):

- Tables B.5a: *four-tag* events in the ggF channel for SM VBF *HH* signal.

- Tables B.5b: *four-tag* events in the ggF channel for $\kappa_\lambda = 10$ VBF *HH* signal.

- Tables B.5c: *four-tag* events in the ggF channel for $\kappa_{2V} = 0$ VBF *HH* signal.

- Tables B.6a: *four-tag* events in the VBF channel for SM VBF *HH* signal.

- Tables B.6b: *four-tag* events in the VBF channel for $\kappa_\lambda = 10$ VBF *HH* signal.

- Tables B.6c: *four-tag* events in the VBF channel for $\kappa_{2V} = 0$ VBF *HH* signal.

Table B.1: 2016–18 data event yields at each step in the analysis event selection for *two-tag* and *four-tag* events in the ggF channel, alongside the ratio of each yield to the yield for the previous cut.

(a) *four-tag* data (ggF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Initial	1.59e+10	-
Pass NTuple Preselection	5.697e+08	-
Trigger	2.807e+08	0.4927
Trigger Buckets	2.49e+08	0.8873
ggF channel	2.457e+08	0.9868
≥ 4 central jets, ≥ 2 b-tags	1.806e+08	0.7349
≥ 4 b-tags	1.886e+06	0.01045
$\lvert\Delta\eta_{HH}\rvert < 1.5$	1.032e+06	0.5469
Top Veto	7.506e+05	0.7276
Signal Region	1.617e+04	0.02154
Control Region 1	3.204e+04	0.04268
Control Region 2	3.067e+04	0.04085

(b) *two-tag* data (ggF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Initial	1.59e+10	-
Pass NTuple Preselection	5.697e+08	-
Trigger	2.807e+08	0.4927
Trigger Buckets	2.49e+08	0.8873
ggF channel	2.457e+08	0.9868
≥ 4 central jets, ≥ 2 b-tags	1.806e+08	0.7349
2 b-tags	1.579e+08	0.8744
$\lvert\Delta\eta_{HH}\rvert < 1.5$	8.27e+07	0.5238
Top Veto	7.22e+07	0.873
Signal Region	1.553e+06	0.02151
Control Region 1	2.983e+06	0.04132
Control Region 2	2.913e+06	0.04035

Table B.2: 2016–18 data event yields at each step in the analysis event selection for *two-tag* and *four-tag* events in the VBF channel, alongside the ratio of each yield to the yield for the previous cut.

(a) *four-tag* data (VBF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Initial	1.59e+10	-
Pass NTuple Preselection	5.697e+08	-
Trigger	2.807e+08	0.4927
Trigger Buckets	2.49e+08	0.8873
VBF channel	3.295e+06	0.01323
$\geq$ 4 central jets, $\geq$ 2 b-tags	3.157e+06	0.9583
$\geq$ 4 b-tags	2.711e+04	0.008586
$\lvert\Delta\eta_{HH}\rvert < 1.5$	2.711e+04	1
Top Veto	2.175e+04	0.8024
Signal Region	502	0.02308
Control Region 1	947	0.04354
Control Region 2	906	0.04165

(b) *two-tag* data (VBF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Initial	1.59e+10	-
Pass NTuple Preselection	5.697e+08	-
Trigger	2.807e+08	0.4927
Trigger Buckets	2.49e+08	0.8873
VBF channel	3.295e+06	0.01323
$\geq$ 4 central jets, $\geq$ 2 b-tags	3.157e+06	0.9583
2 b-tags	2.758e+06	0.8736
$\lvert\Delta\eta_{HH}\rvert < 1.5$	2.758e+06	1
Top Veto	2.469e+06	0.8951
Signal Region	5.873e+04	0.02379
Control Region 1	1.108e+05	0.04489
Control Region 2	1.1e+05	0.04454

Table B.3: ggF *HH* MC simulation event yields at each step in the analysis event selection for *four-tag* events in the ggF channel normalized to $126\,\text{fb}^{-1}$, alongside the ratio of each yield to the yield for the previous cut.

(a) *four-tag* SM ggF *HH* MC simulation (ggF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Pass NTuple Preselection	526.6	-
Trigger	475	0.9019
Trigger Buckets	419	0.8821
Apply FTag, Trigger & JVT SFs	381.8	0.9112
ggF channel	376.6	0.9864
≥ 4 central jets, ≥ 2 b-tags	322.4	0.8561
≥ 4 b-tags	86	0.2668
$\lvert\Delta\eta_{HH}\rvert < 1.5$	71.85	0.8355
Top Veto	60.4	0.8406
Signal Region	29.1	0.4817
Control Region 1	11.41	0.1889
Control Region 2	7.137	0.1182

(b) *four-tag* $\kappa_\lambda = 10$ ggF *HH* MC simulation (ggF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Pass NTuple Preselection	7338	-
Trigger	6917	0.9427
Trigger Buckets	6378	0.922
Apply FTag, Trigger & JVT SFs	5279	0.8278
ggF channel	5198	0.9846
≥ 4 central jets, ≥ 2 b-tags	4314	0.83
≥ 4 b-tags	1002	0.2322
$\lvert\Delta\eta_{HH}\rvert < 1.5$	850.6	0.8492
Top Veto	569	0.6689
Signal Region	182.7	0.3211
Control Region 1	86.18	0.1515
Control Region 2	66.06	0.1161

Table B.4: ggF *HH* MC simulation event yields at each step in the analysis event selection for *four-tag* events in the VBF channel normalized to $126\,\text{fb}^{-1}$, alongside the ratio of each yield to the yield for the previous cut.

(a) *four-tag* SM ggF *HH* MC simulation (VBF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Pass NTuple Preselection	526.6	-
Trigger	475	0.9019
Trigger Buckets	419	0.8821
Apply FTag, Trigger & JVT SFs	381.8	0.9112
VBF channel	5.208	0.01364
≥ 4 central jets, ≥ 2 b-tags	5.155	0.9898
≥ 4 b-tags	1.14	0.2212
$\lvert\Delta\eta_{HH}\rvert < 1.5$	1.14	1
Top Veto	1.008	0.8838
Signal Region	0.4833	0.4796
Control Region 1	0.1703	0.169
Control Region 2	0.123	0.122

(b) *four-tag* $\kappa_\lambda = 10$ ggF *HH* MC simulation (VBF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Pass NTuple Preselection	7338	-
Trigger	6917	0.9427
Trigger Buckets	6378	0.922
Apply FTag, Trigger & JVT SFs	5279	0.8278
VBF channel	81.14	0.01537
≥ 4 central jets, ≥ 2 b-tags	80.22	0.9888
≥ 4 b-tags	15.29	0.1906
$\lvert\Delta\eta_{HH}\rvert < 1.5$	15.29	1
Top Veto	11.15	0.729
Signal Region	3.099	0.278
Control Region 1	1.851	0.166
Control Region 2	1.449	0.13

Table B.5: VBF *HH* MC simulation event yields at each step in the analysis event selection for *four-tag* events in the ggF channel normalized to 126 fb^{-1}, alongside the ratio of each yield to the yield for the previous cut.

(a) *four-tag* SM VBF *HH* MC simulation (ggF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Pass NTuple Preselection	22.26	-
Trigger	20.67	0.9287
Trigger Buckets	18.39	0.8895
Apply FTag, Trigger & JVT SFs	16.14	0.8778
ggF channel	13.9	0.8613
≥ 4 central jets, ≥ 2 b-tags	10.75	0.7731
≥ 4 b-tags	1.87	0.174
$\lvert\Delta\eta_{HH}\rvert < 1.5$	0.9419	0.5036
Top Veto	0.736	0.7814
Signal Region	0.2352	0.3195
Control Region 1	0.1094	0.1486
Control Region 2	0.07797	0.1059

(b) *four-tag* $\kappa_\lambda = 10$ VBF *HH* MC simulation (ggF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Pass NTuple Preselection	1573	-
Trigger	1465	0.9317
Trigger Buckets	1303	0.8896
Apply FTag, Trigger & JVT SFs	1092	0.8374
ggF channel	955.5	0.8754
≥ 4 central jets, ≥ 2 b-tags	756.5	0.7917
≥ 4 b-tags	134.5	0.1778
$\lvert\Delta\eta_{HH}\rvert < 1.5$	109.4	0.8132
Top Veto	76.41	0.6988
Signal Region	25.19	0.3297
Control Region 1	11.42	0.1495
Control Region 2	8.892	0.1164

(c) *four-tag* $\kappa_{2V} = 0$ VBF *HH* MC simulation (ggF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Pass NTuple Preselection	626.1	-
Trigger	513.5	0.8201
Trigger Buckets	444.6	0.8658
Apply FTag, Trigger & JVT SFs	405.2	0.9114
ggF channel	334.4	0.8254
≥ 4 central jets, ≥ 2 b-tags	260.5	0.7788
≥ 4 b-tags	65.23	0.2504
$\lvert\Delta\eta_{HH}\rvert < 1.5$	46.37	0.7108
Top Veto	43.1	0.9296
Signal Region	22.97	0.533
Control Region 1	8.26	0.1916
Control Region 2	4.854	0.1126

Table B.6: VBF *HH* MC simulation event yields at each step in the analysis event selection for *four-tag* events in the VBF channel normalized to $126\,\text{fb}^{-1}$, alongside the ratio of each yield to the yield for the previous cut.

(a) *four-tag* SM VBF *HH* MC simulation (VBF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Pass NTuple Preselection	22.26	-
Trigger	20.67	0.9287
Trigger Buckets	18.39	0.8895
Apply FTag, Trigger & JVT SFs	16.14	0.8778
VBF channel	2.238	0.1387
≥ 4 central jets, ≥ 2 b-tags	2.223	0.9931
≥ 4 b-tags	0.7449	0.3351
$\lvert\Delta\eta_{HH}\rvert < 1.5$	0.7449	1
Top Veto	0.6722	0.9024
Signal Region	0.3265	0.4857
Control Region 1	0.1149	0.171
Control Region 2	0.09092	0.1352

(b) *four-tag* $\kappa_\lambda = 10$ VBF *HH* MC simulation (VBF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Pass NTuple Preselection	1573	-
Trigger	1465	0.9317
Trigger Buckets	1303	0.8896
Apply FTag, Trigger & JVT SFs	1092	0.8374
VBF channel	136.1	0.1246
≥ 4 central jets, ≥ 2 b-tags	135.1	0.9929
≥ 4 b-tags	46.14	0.3416
$\lvert\Delta\eta_{HH}\rvert < 1.5$	46.14	1
Top Veto	34.84	0.7551
Signal Region	14.24	0.4087
Control Region 1	5.03	0.1443
Control Region 2	4.23	0.1214

(c) *four-tag* $\kappa_{2V} = 0$ VBF *HH* MC simulation (VBF channel)

Selection Step	**Yield**	**Yield/Prior Cut**
Pass NTuple Preselection	626.1	-
Trigger	513.5	0.8201
Trigger Buckets	444.6	0.8658
Apply FTag, Trigger & JVT SFs	405.2	0.9114
VBF channel	70.74	0.1746
≥ 4 central jets, ≥ 2 b-tags	70.3	0.9939
≥ 4 b-tags	27.63	0.393
$\lvert\Delta\eta_{HH}\rvert < 1.5$	27.63	1
Top Veto	26.52	0.9601
Signal Region	17.29	0.6517
Control Region 1	3.805	0.1435
Control Region 2	3.074	0.1159

C

Trigger Lists for VBF Trigger Studies

Table C.1 shows the list of VBF trigger chains operational in the period 2016–18 that were investigated as part of the study into utilising VBF triggers in the full Run 2 ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analysis, as described in Chapter 6.

Table C.2 shows the list of b-jet trigger chains used by the $HH \rightarrow b\bar{b}b\bar{b}$ analysis prior to the implementation of the trigger buckets strategy.

Table C.3 shows a key for the shortened versions of the VBF trigger chain names used throughout Chapter 6.

Table C.1: The list of VBF triggers investigated as part of studies into whether their inclusion could improve analysis sensitivity to VBF *HH* signal. The trigger underlined was intended to be included in the VBF trigger list; however, due to a typing error when it was input, it was excluded.

Year	**Trigger Chain**
2016	HLT_g25_medium_L1EM22VHI_j35_0eta490_bmv2c2077_split_3j35_0eta490_invm700
2016	HLT_g25_medium_L1EM22VHI_2j35_0eta490_bmv2c2077_split_2j35_0eta490
2017	HLT_j35_gsc55_bmv2c1070_split_2j45_320eta490_L1J25.0ETA23_2J15.31ETA49
2017	HLT_ht300_2j40_0eta490_invm700_L1HT150-J20s5.ETA31_MJJ-400-CF_AND_2j25_gsc45_bmv2c1070_split
2017/2018	HLT_j55_gsc80_bmv2c1070_split_j45_gsc60_bmv2c1085_split_j45_320eta490
2018	HLT_j45_gsc55_bmv2c1070_split_2j45_320eta490_L1J25.0ETA23_2J15.31ETA49
2018	HLT_j80_0eta240_j60_j45_320eta490_AND_2j35_gsc45_bmv2c1070_split
2018	HLT_ht300_2j40_0eta490_invm700_L1HT150-J20s5.ETA31_MJJ-400-CF_AND_2j35_gsc45_bmv2c1070_split
2018	HLT_j70_j50_0eta490_invm1100j70_dphi20_deta40_L1MJJ-500-NFF
2018	HLT_j70_0eta490_j50_0eta490_2j35_0eta490_invm1000j50_L1MJJ-500-NFF_AND_2j35_bmv2c1070_split
2018	HLT_g35_medium_j70_j50_0eta490_invm900j50_L1MJJ-500-NFF

Table C.2: List of b-jet trigger used in the $HH \rightarrow b\bar{b}b\bar{b}$ prior to the implementation and optimisation of the trigger bucketing strategy.

Year	**Trigger Chain**
2015	HLT_j225_bloose
2015	HLT_j100_2j55_bmedium
2015	HLT_2j35_btight_2j35_L14J15.0ETA25
2016	HLT_j225_bmv2c2060_split
2016	HLT_j100_2j55_bmv2c2060_split
2016	HLT_2j35_bmv2c2060_split_2j35_L14J15.0ETA25
2017	HLT_j110_gsc150_boffperf_split_2j35_gsc55_bmv2c1070_split_L1J85_3J30
2017	HLT_2j15_gsc35_bmv2c1040_split_2j15_gsc35_boffperf_split_L14J15.0ETA25
2017	HLT_2j35_gsc55_bmv2c1050_split_ht300_L1HT190-J15s5.ETA21
2017	HLT_j225_gsc300_bmv2c1070_split
2018	HLT_j110_gsc150_boffperf_split_2j45_gsc55_bmv2c1070_split_L1J85_3J30
2018	HLT_2j35_bmv2c1060_split_2j35_L14J15.0ETA25
2018	HLT_2j45_gsc55_bmv2c1050_split_ht300_L1HT190-J15s5.ETA21

Table C.3: Key for shorthand VBF trigger names used in Chapter 6.

Trigger Short	Trigger Chain
1b 2frwdj	(2017) HLT_j35_gsc55_bmv2c1070_split_2j45_320eta490_L1J25.0ETA23_2J15.31ETA49
	(2018) HLT_j45_gsc55_bmv2c1070_split_2j45_320eta490_L1J25.0ETA23_2J15.31ETA49
2b 1frwdj	HLT_j55_gsc80_bmv2c1070_split_j45_gsc60_bmv2c1085_split_j45_320eta490
2b 2jMjj HT	(2017) HLT_ht300_2j40_0eta490_invm700_L1HT150-J20s5.ETA31_MJJ-400-CF_AND_2j25_gsc45_bmv2c1070_split
	(2018) HLT_ht300_2j40_0eta490_invm700_L1HT150-J20s5.ETA31_MJJ-400-CF_AND_2j35_gsc45_bmv2c1070_split
2b 1frwdj 1j 1cntlj	HLT_j80_0eta240_j60_j45_320eta490_AND_2j35_gsc45_bmv2c1070_split
2j Mjj+dPhijj+dEtajj	HLT_j70_j50_0eta490_invm1100j70_dphi20_deta40_L1MJJ-500-NFF
2j 2jMjj 2b	HLT_j70_0eta490_j50_0eta490_2j35_0eta490_invm1000j50_L1MJJ-500-NFF_AND_2j35_bmv2c1070_split

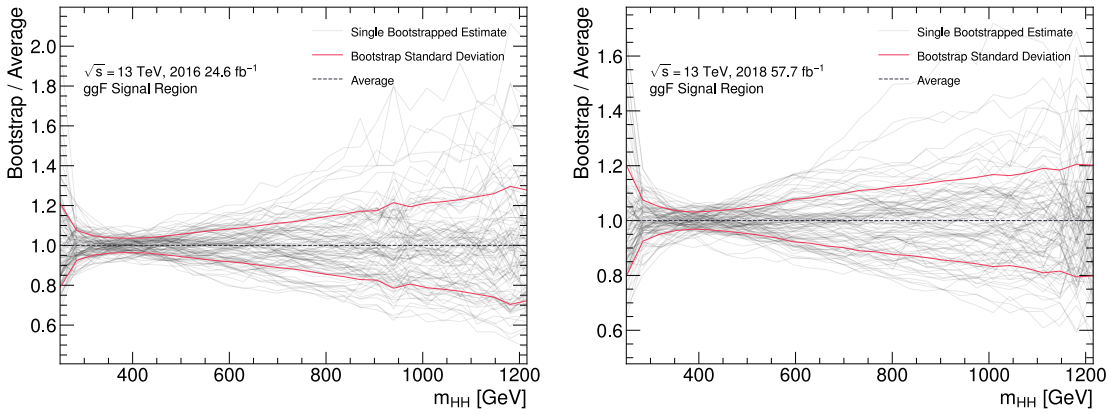
D

Additional Background Uncertainty Plots

The fully data-driven background estimate used in the full Run 2 ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analysis is described in Chapter 7. The uncertainties accounted for in the analysis, including those associated with the background estimate, are described in Chapter 10. This appendix shows additional plots describing the uncertainties associated with the background estimate.

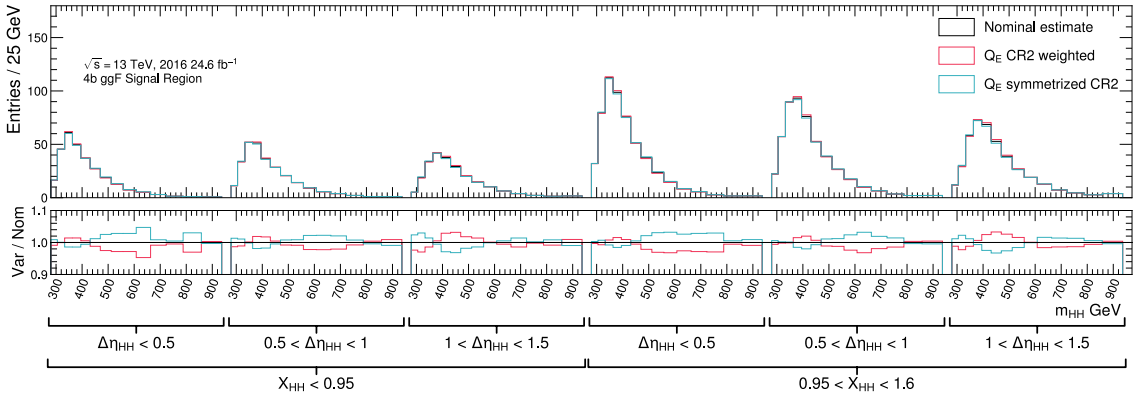
Figure D.1 shows the ratio of the mean background estimate histogram to the ninety-nine other bootstrap histograms, with the bootstrap uncertainty highlighted, for 2016 and 2018 data-taking years in the ggF channel. The plot for 2017 is shown in Figure 10.1a in Section 10.1.

Figures D.2 – D.4 and D.5 – D.7 show the symmetrised background shape systematic uncertainty in the different categories of the ggF and VBF channels respectively for the case when the CR2-derived weights are applied to Q_E, Q_S and Q_W and the CR1-derived weights are applied to the remaining SR quadrants. The plots for Q_N are shown in Figures 10.3 and 10.4 in Section 10.2.4 for the VBF and ggF channels respectively.

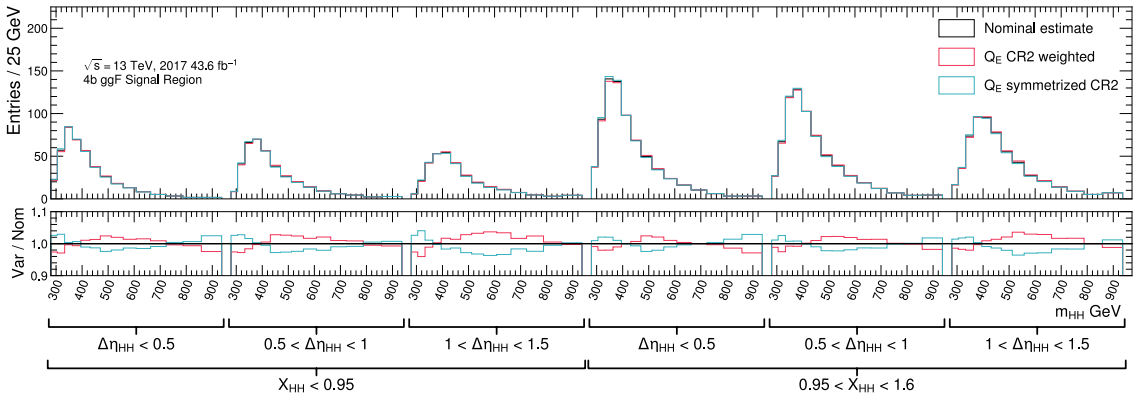


(a) 2016 (b) 2018

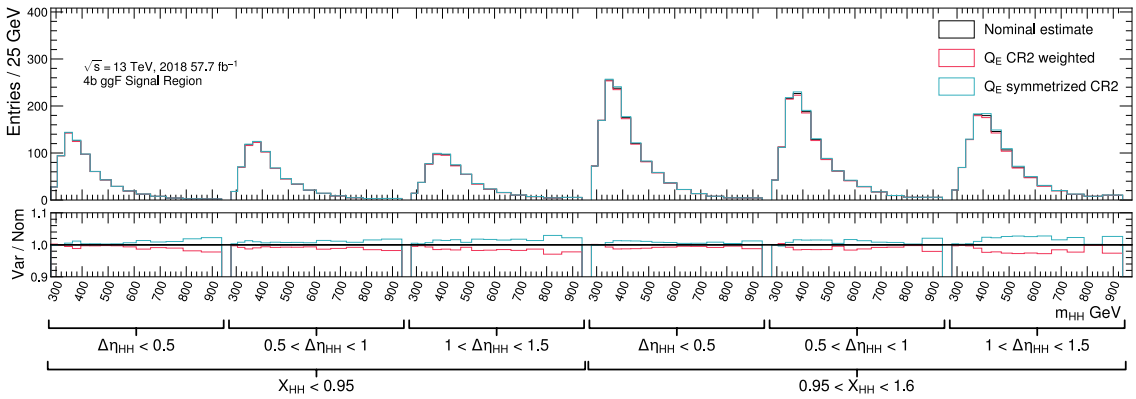
Figure D.1: Ratio of the mean background estimate histogram to the ninety-nine other bootstrap histograms (grey), where the bootstrap uncertainty (red) is the standard deviation of the distribution of bootstraps, for 2016 (left) and 2018 (right) data-taking years in the ggF channel.



(a) 2016

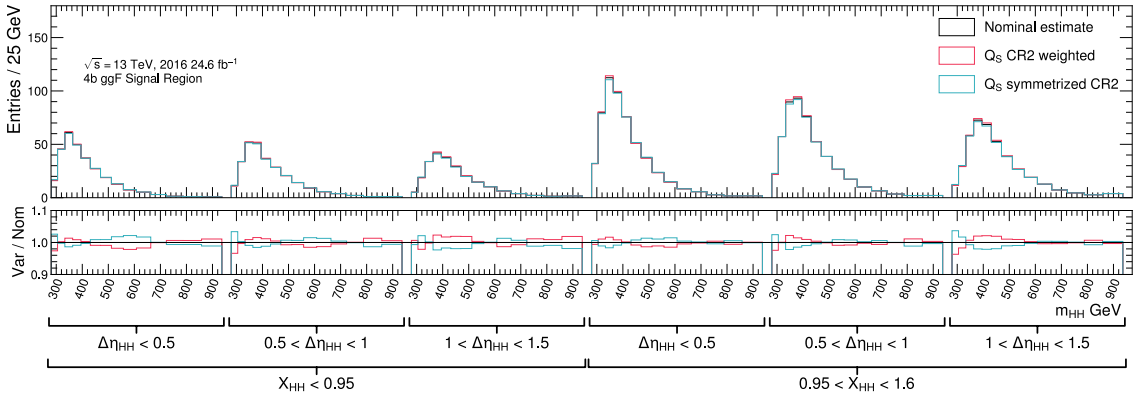


(b) 2017

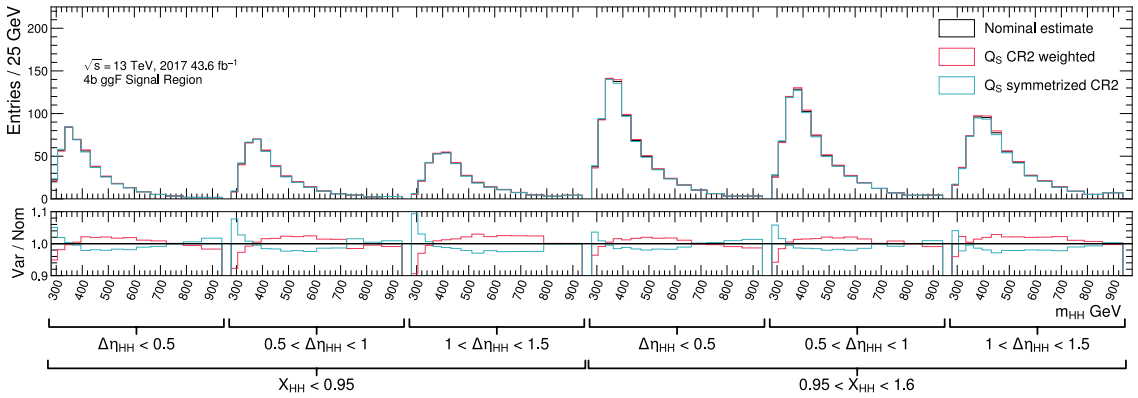


(c) 2018

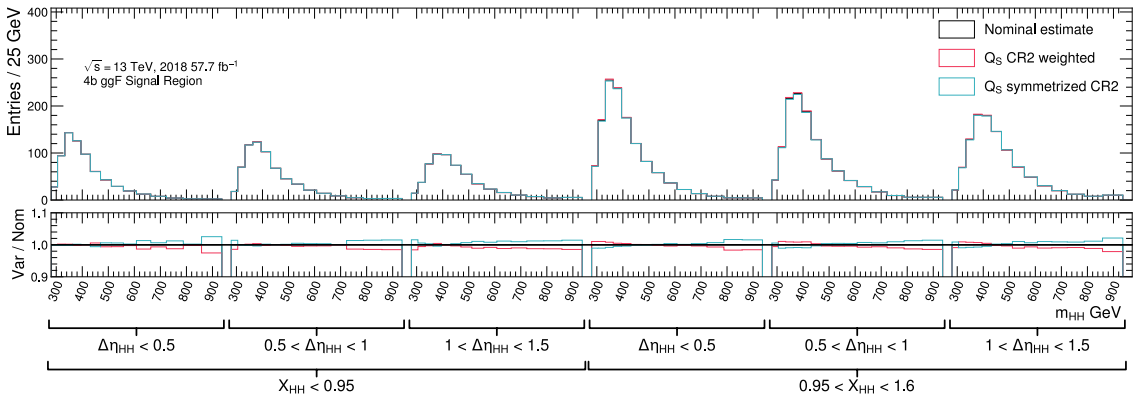
Figure D.2: Symmetrised shape uncertainty on ggF channel background estimate, where CR2-derived weights are applied to Q_E and CR1-derived weights applied to the other Signal Region quadrants. Each plot simultaneously shows the six symmetrised shape uncertainty bands for the six $|\Delta\eta_{HH}|$–X_{HH} ggF categories, with the labels below the x-axis indicating the category a particular band corresponds to.



(a) 2016

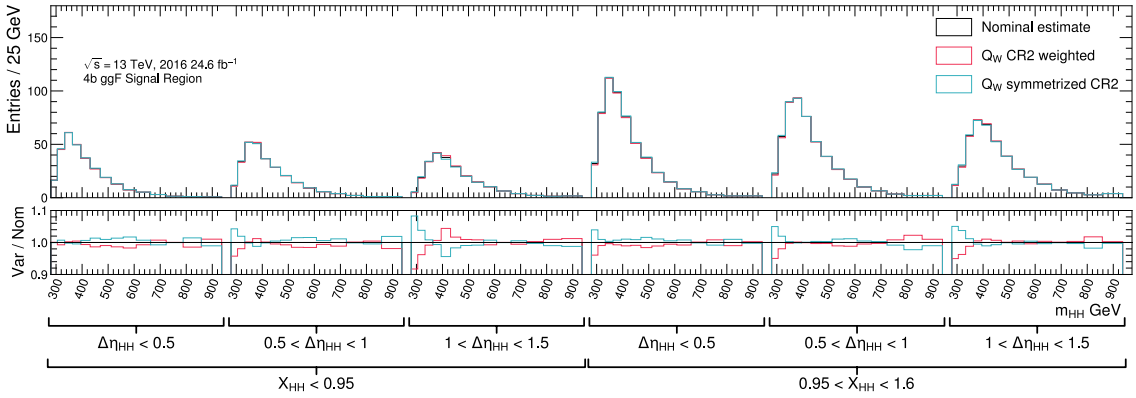


(b) 2017

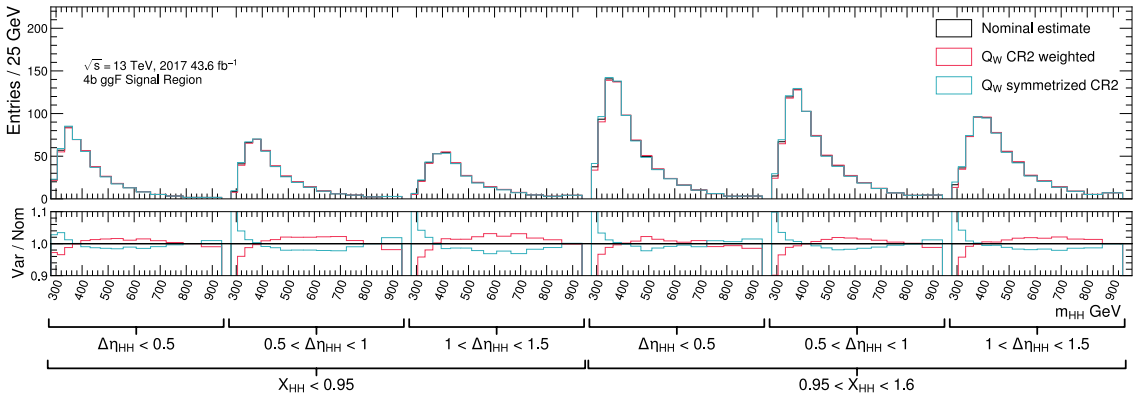


(c) 2018

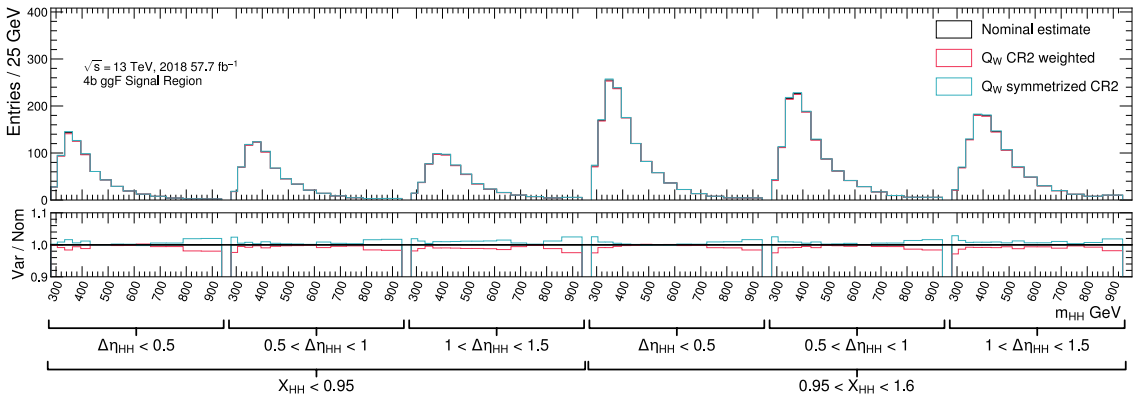
Figure D.3: Symmetrised shape uncertainty on ggF channel background estimate, where CR2-derived weights are applied to Q_S and CR1-derived weights applied to the other Signal Region quadrants. Each plot simultaneously shows the six symmetrised shape uncertainty bands for the six $|\Delta\eta_{HH}|$–X_{HH} ggF categories, with the labels below the x-axis indicating the category a particular band corresponds to.



(a) 2016

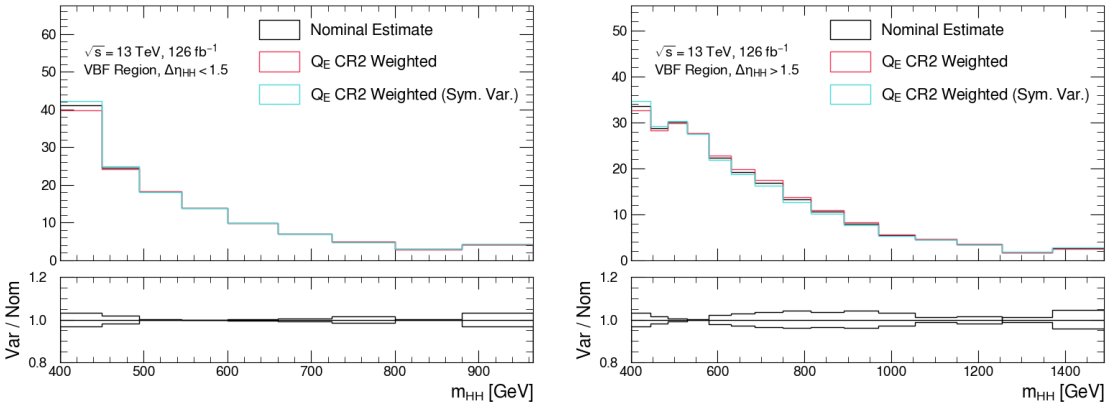


(b) 2017



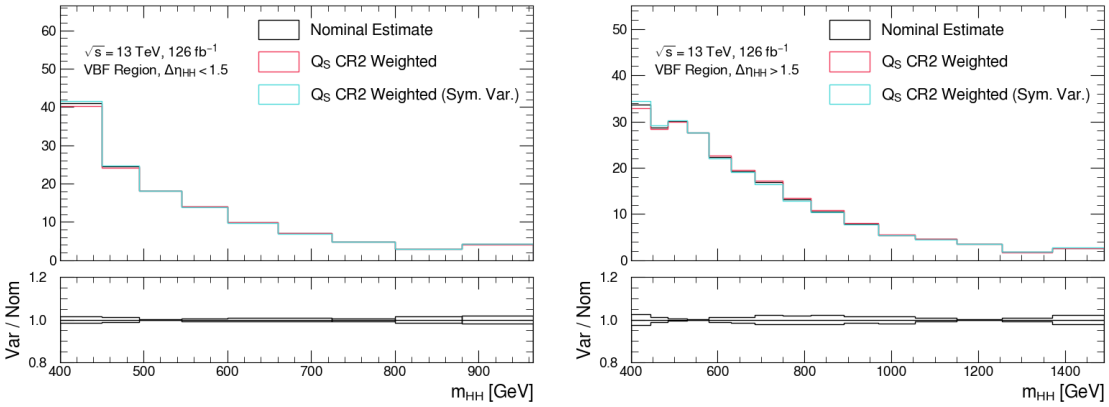
(c) 2018

Figure D.4: Symmetrised shape uncertainty on ggF channel background estimate, where CR2-derived weights are applied to Q_W and CR1-derived weights applied to the other Signal Region quadrants. Each plot simultaneously shows the six symmetrised shape uncertainty bands for the six $|\Delta\eta_{HH}|$–X_{HH} ggF categories, with the labels below the x-axis indicating the category a particular band corresponds to.



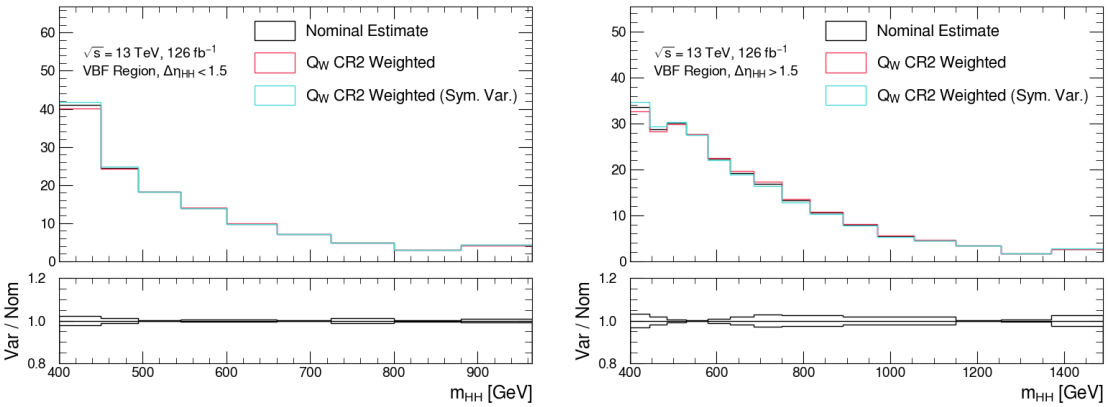
(a) $|\Delta\eta_{HH}| < 1.5$ (b) $|\Delta\eta_{HH}| > 1.5$

Figure D.5: Symmetrised shape uncertainty on VBF channel background estimate in the two $|\Delta\eta_{HH}|$ categories, where CR2-derived weights are applied to Q_E and CR1-derived weights applied to the other Signal Region quadrants.



(a) $|\Delta\eta_{HH}| < 1.5$ (b) $|\Delta\eta_{HH}| > 1.5$

Figure D.6: Symmetrised shape uncertainty on VBF channel background estimate in the two $|\Delta\eta_{HH}|$ categories, where CR2-derived weights are applied to Q_S and CR1-derived weights applied to the other Signal Region quadrants.



(a) $|\Delta\eta_{HH}| < 1.5$ (b) $|\Delta\eta_{HH}| > 1.5$

Figure D.7: Symmetrised shape uncertainty on VBF channel background estimate in the two $|\Delta\eta_{HH}|$ categories, where CR2-derived weights are applied to Q_W and CR1-derived weights applied to the other Signal Region quadrants.

E

Nuisance Parameter Naming Scheme

The uncertainties accounted for in the full Run 2 ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analysis are described in Chapter 10. Uncertainties are integrated into the model using nuisance parameters (NP). This appendix describes the naming scheme used for the NPs in the fit, as shown in Figures 11.3 and F.3. Below "{`year`}" represents the data-taking year and can have a value of `16`, `17`, `18` or `-1`, where `-1` corresponds to all years combined; "{`channel`}" can either be `VBF` or `ggF`; and "{...}" represents any longer chain of descriptors.

Background Estimate Shape Systematics:

NP naming scheme: `alpha_CR12_shape_{quadrant}_{channel}_{year}`

NPs associated with the shape uncertainty on the background estimate (see Section 10.2.4). "{`quadrant`}" indicates the SR quadrant that the CR2-derived weights are applied to to derive the alternative background estimate, with the CR1-derived weights being applied to the remaining SR quadrants. "{`quadrant`}" can have values of `N`, `E`, `S` and `W`.

Background Estimate Non-Closure Systematic:

NP naming scheme: `alpha_NC_3b1f_ggf_{year}`

Non-closure systematic uncertainty on the background estimate derived in the `3b1f` region of the ggF channel (see Section 10.2.5).

Jet Energy Spectrum and Jet Energy Resolution Uncertainties:

JES NP naming scheme: `alpha_JET_{...}`

JER NP naming scheme: `alpha_JET_JER_{...}`

Uncertainties associated with the derivation of the Jet Energy Spectrum (JES) and Jet Energy Resolution (JER) applied to reconstructed jets (see Section 10.2.1). Use the *category reduction* NP scheme for the JES and *FullJER* NP scheme for the JER [118].

Flavour Tagging Uncertainties:

NP naming scheme: `alpha_FT_EFF_{...}`

Uncertainties associated with the calibration of the flavour tagging algorithms used to b-tag jets (see Section 10.2.2) [119]. The number of NPs is reduced by grouping variations using principle component analysis. This produces a set of so-called *eigenvector variations*, thus the `Eigen` in the NP name [120].

Theory Uncertainties:

NP naming scheme: `alpha_THEO_{...}`

Uncertainties associated with the choice of quantities and modelling procedures based on particle physics theory (see Section 10.2.3).

F

Additional Fit Validity Plots

The full Run 2 ATLAS $HH \rightarrow b\bar{b}b\bar{b}$ analysis generates results using the statistical procedure outlined in Chapter 9. A series of checks of the maximum likelihood estimator fit performed as part of this procedure are described in Section 11.1. Additional plots produced as part of these checks are included in this appendix.

The m_{HH} distributions of the background-only model after the fit to 2017 and 2018 data in the Signal Region of the ggF channel are shown in Figures F.1 and F.2 respectively. The plot for 2016 is shown in Figure 11.1 of Section 11.1.

The post-fit pull on the nuisance parameters of the model are shown, ordered by impact, in Figure F.3. The nuisance parameters shown are ranked 21st to 91st. This is for the combined ggF and VBF channels. All behaviour was found to be within acceptable bounds. The pulls and impacts of 20 highest ranked nuisance parameters are shown in Figure 11.3 of Section 11.1.

The correlations between all model nuisance parameters in the combined VBF and ggF channels for a SM HH signal and background fit to data are shown in Figure F.4. The correlation matrix for nuisance parameters associated with the systematic

uncertainties on the background estimate is shown in Figure 11.4 in Section 11.1.

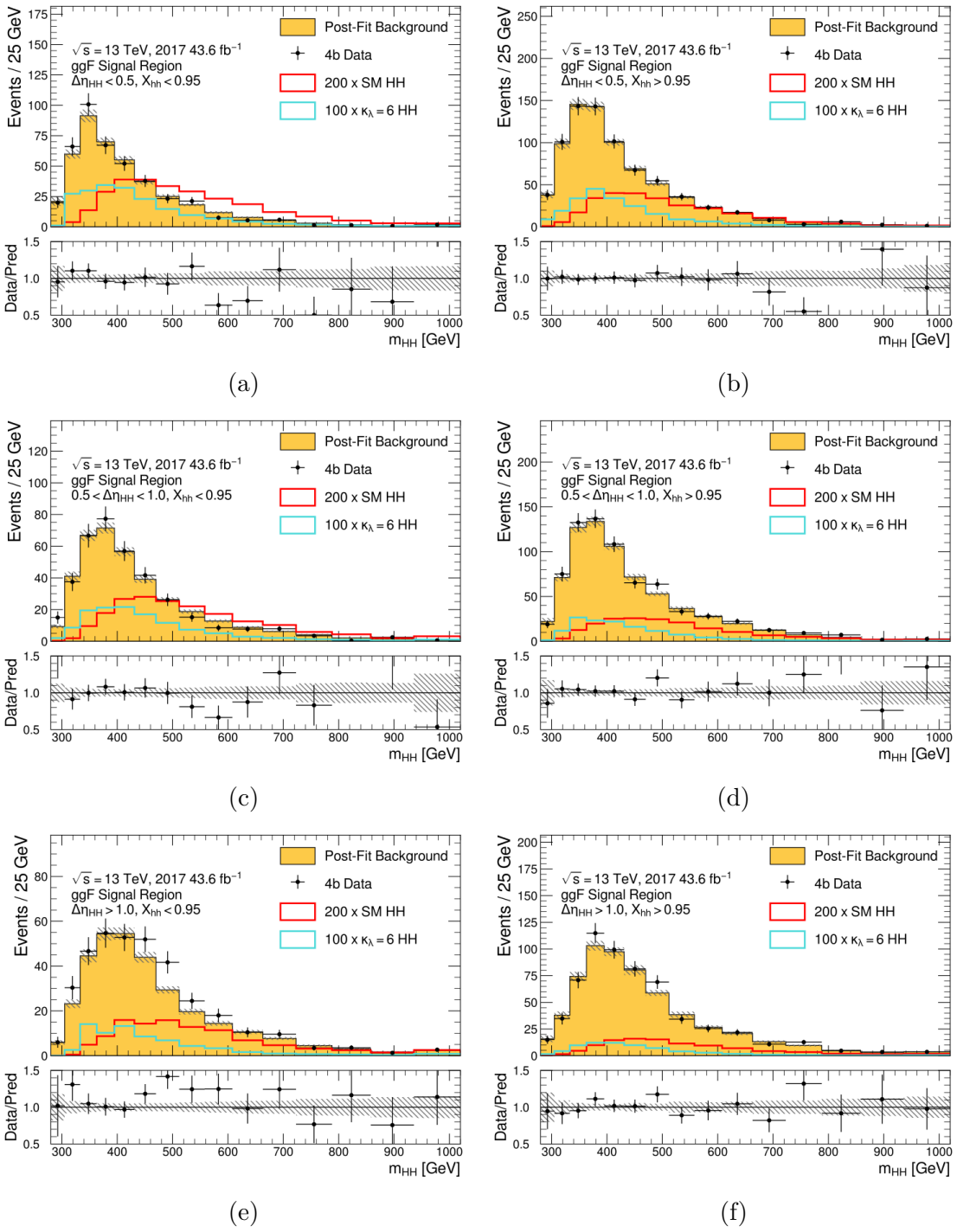


Figure F.1: Plots of the m_{HH} distributions of the background-only model (yellow) after the maximum likelihood estimator fit to data (black) in the $|\Delta\eta_{HH}|$–X_{HH} categories in the ggF Channel for the 2017 data-taking year. The distributions for two important HH production signal models – SM and $\kappa_\lambda = 10$ – are shown for reference. The signal distributions are scaled as to be visible next to the far more dominant background.

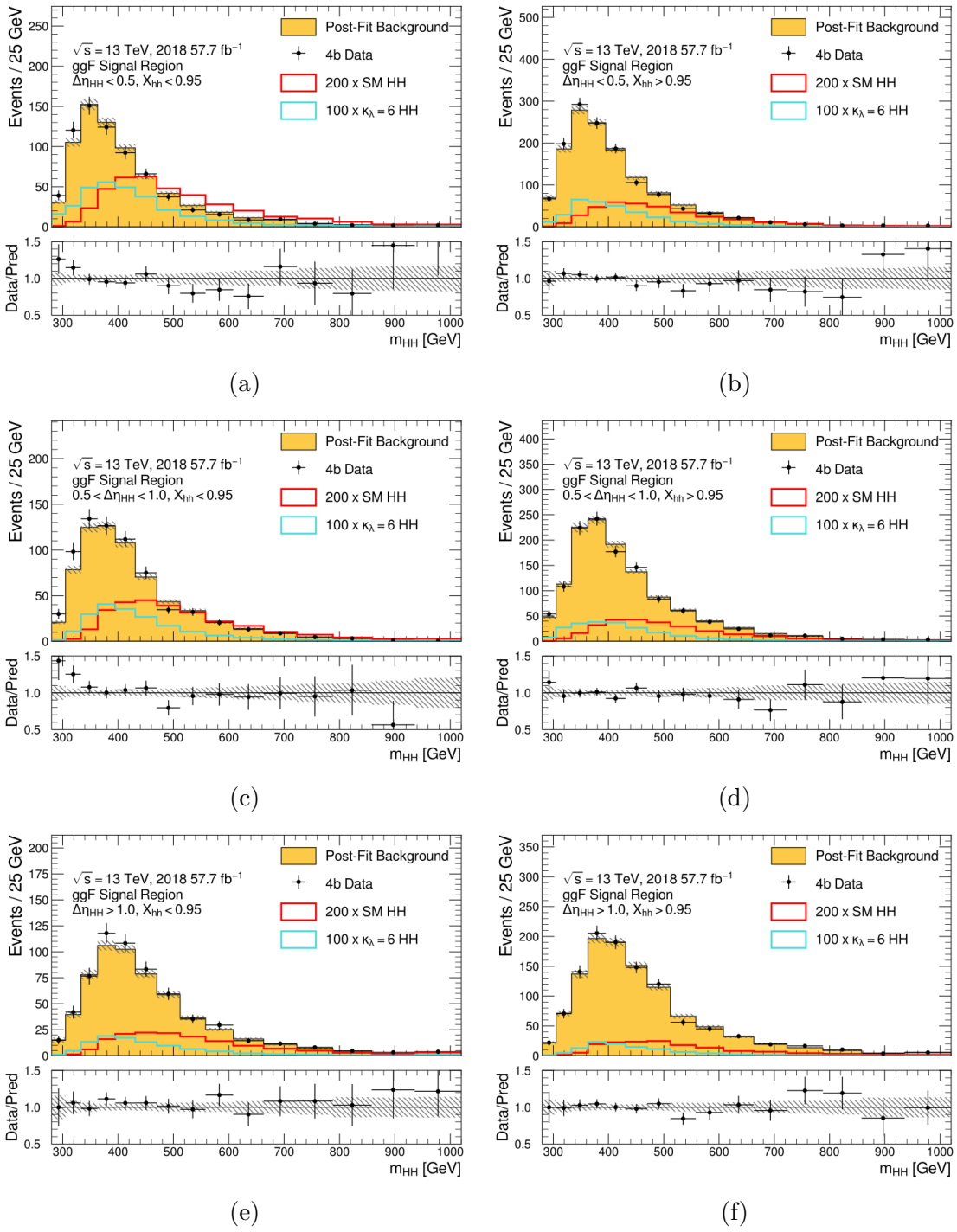
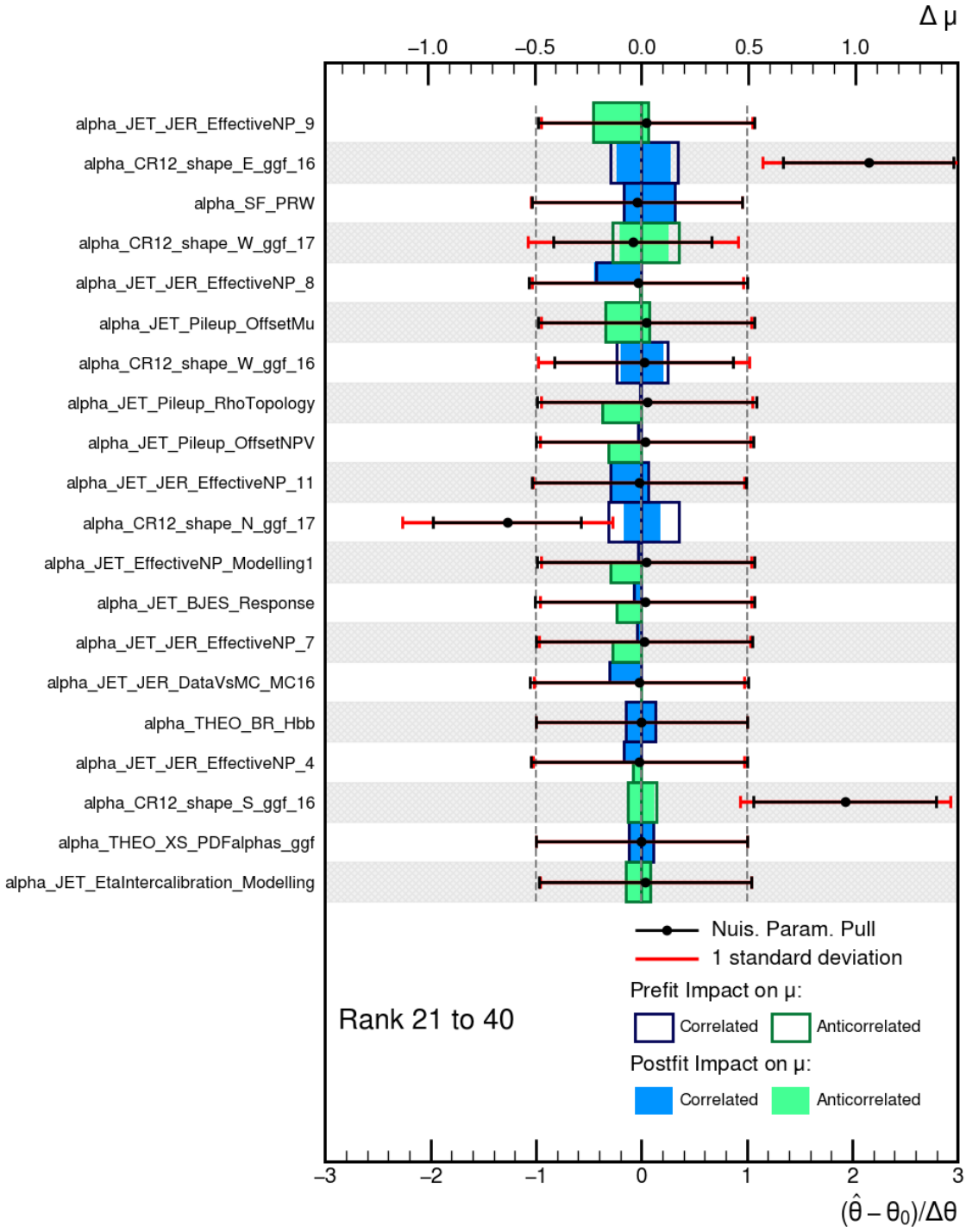
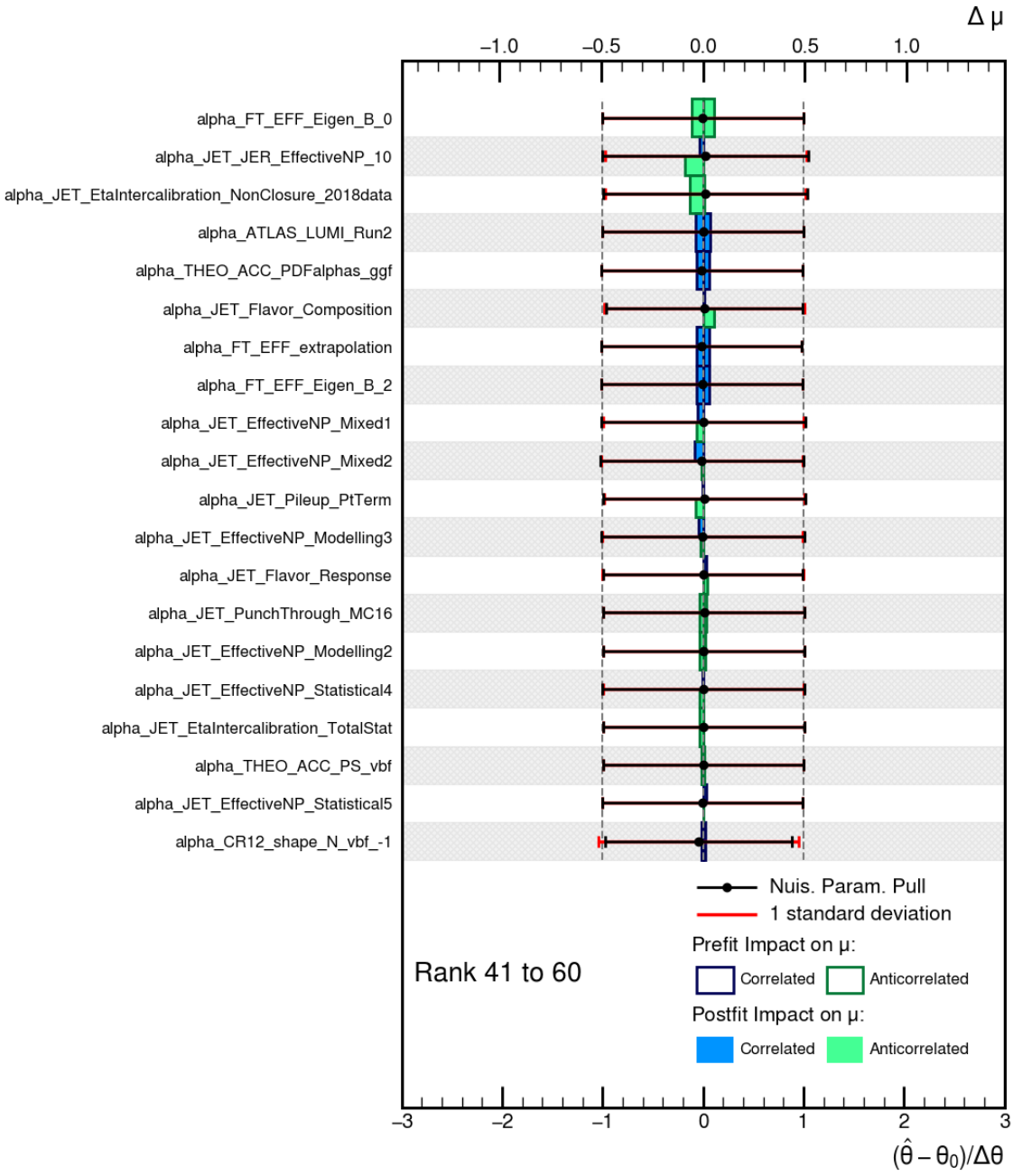


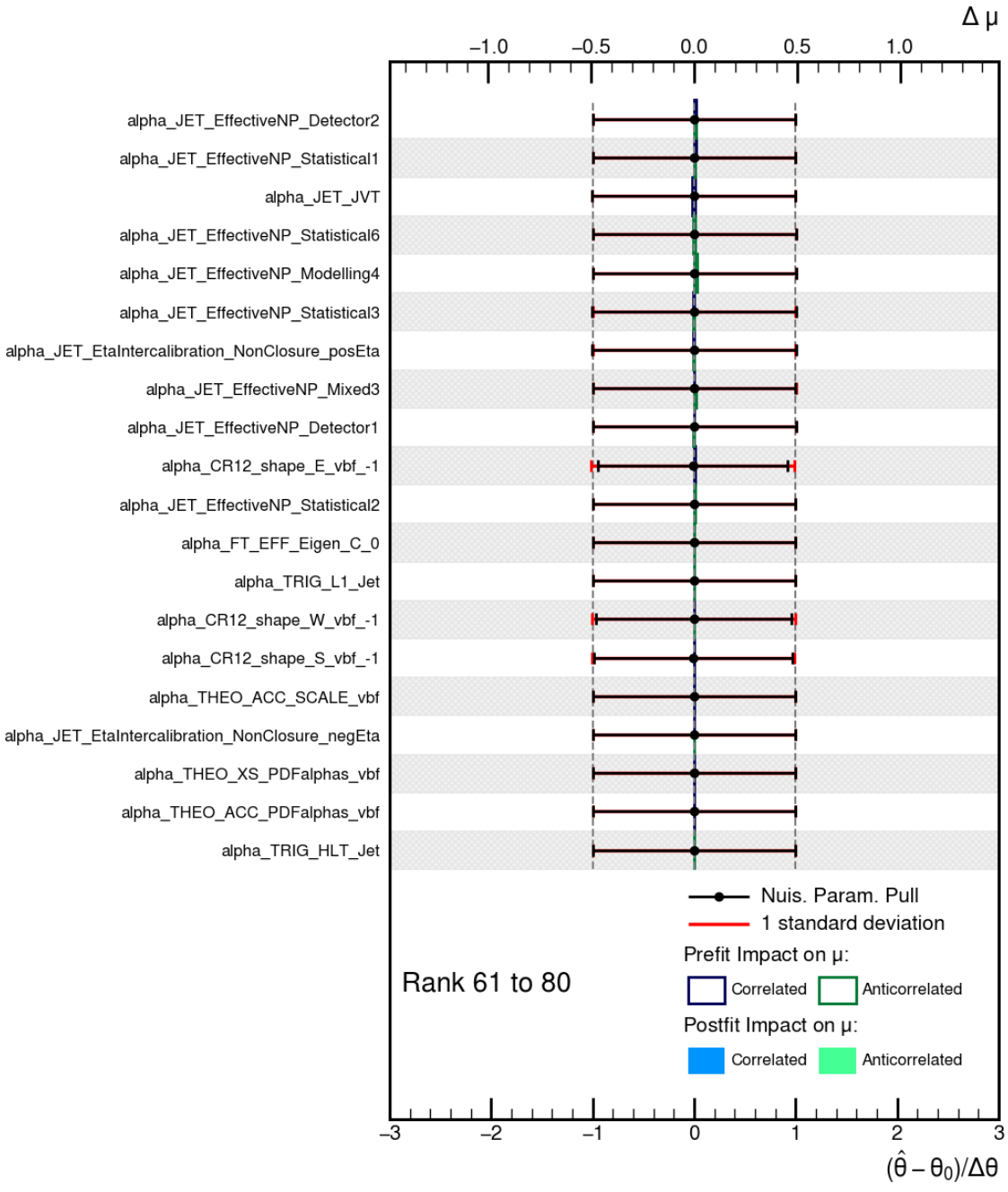
Figure F.2: Plots of the m_{HH} distributions of the background-only model (yellow) after the maximum likelihood estimator fit to data (black) in the $|\Delta\eta_{HH}|$–X_{HH} categories in the ggF Channel for the 2018 data-taking year. The distributions for two important HH production signal models – SM and $\kappa_\lambda = 10$ – are shown for reference. The signal distributions are scaled as to be visible next to the far more dominant background.



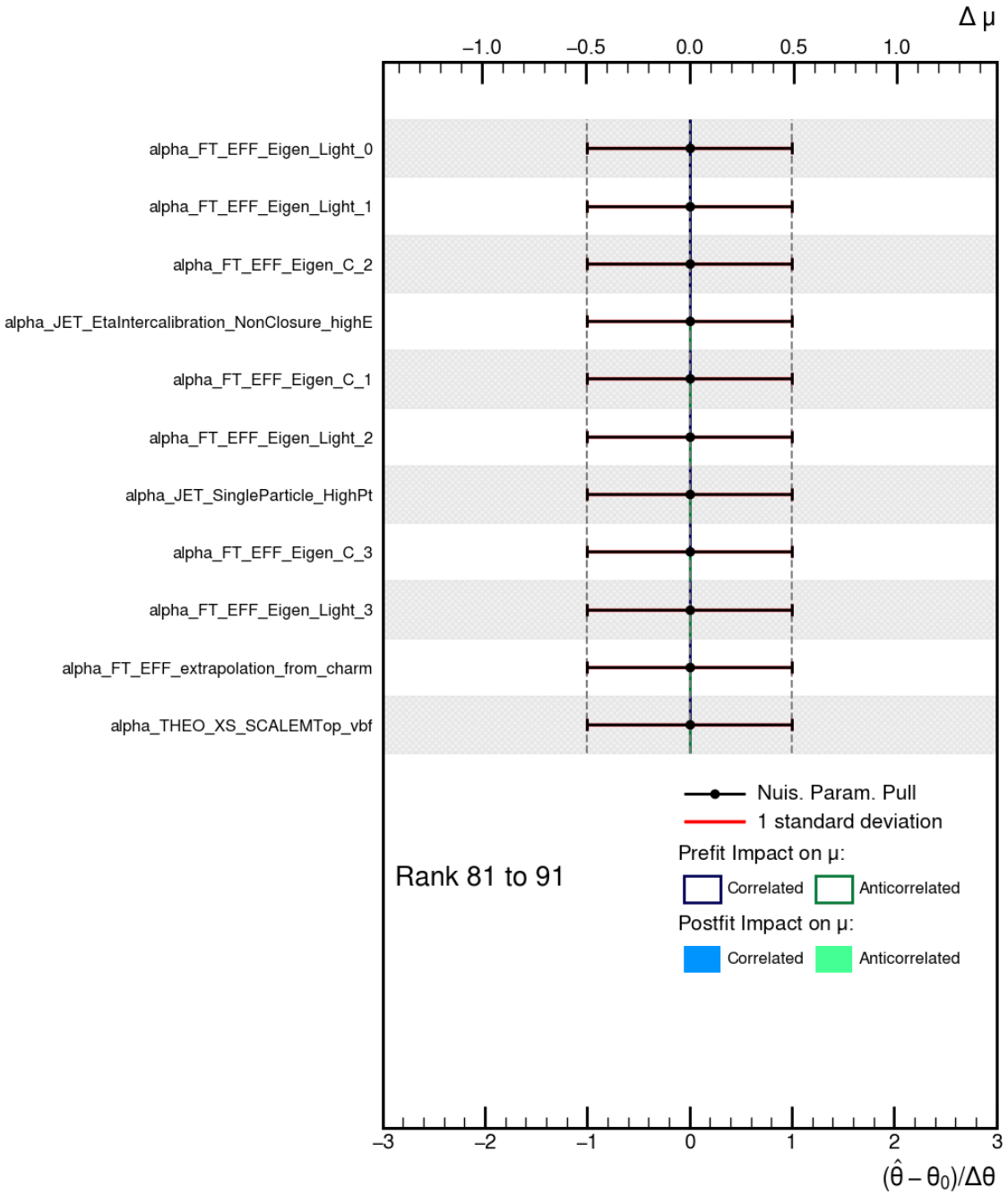
(a) Ranked 21 – 40.



(b) Ranked 41 – 60.



(c) Ranked 61 – 80.



(d) Ranked 81 – 91.

Figure F.3: Combined plot showing the pull and uncertainty on the nuisance parameters (red and black) alongside the impact on the parameter of interest, μ (blue and green). Nuisance parameters are ranked top to bottom with the highest impact at the top. The standard deviation on a nuisance parameter is shown in red as a yardstick to judge the constraint, shown in black. Pulls are shown post the maximum likelihood estimator (MLE) fit of the background to data.

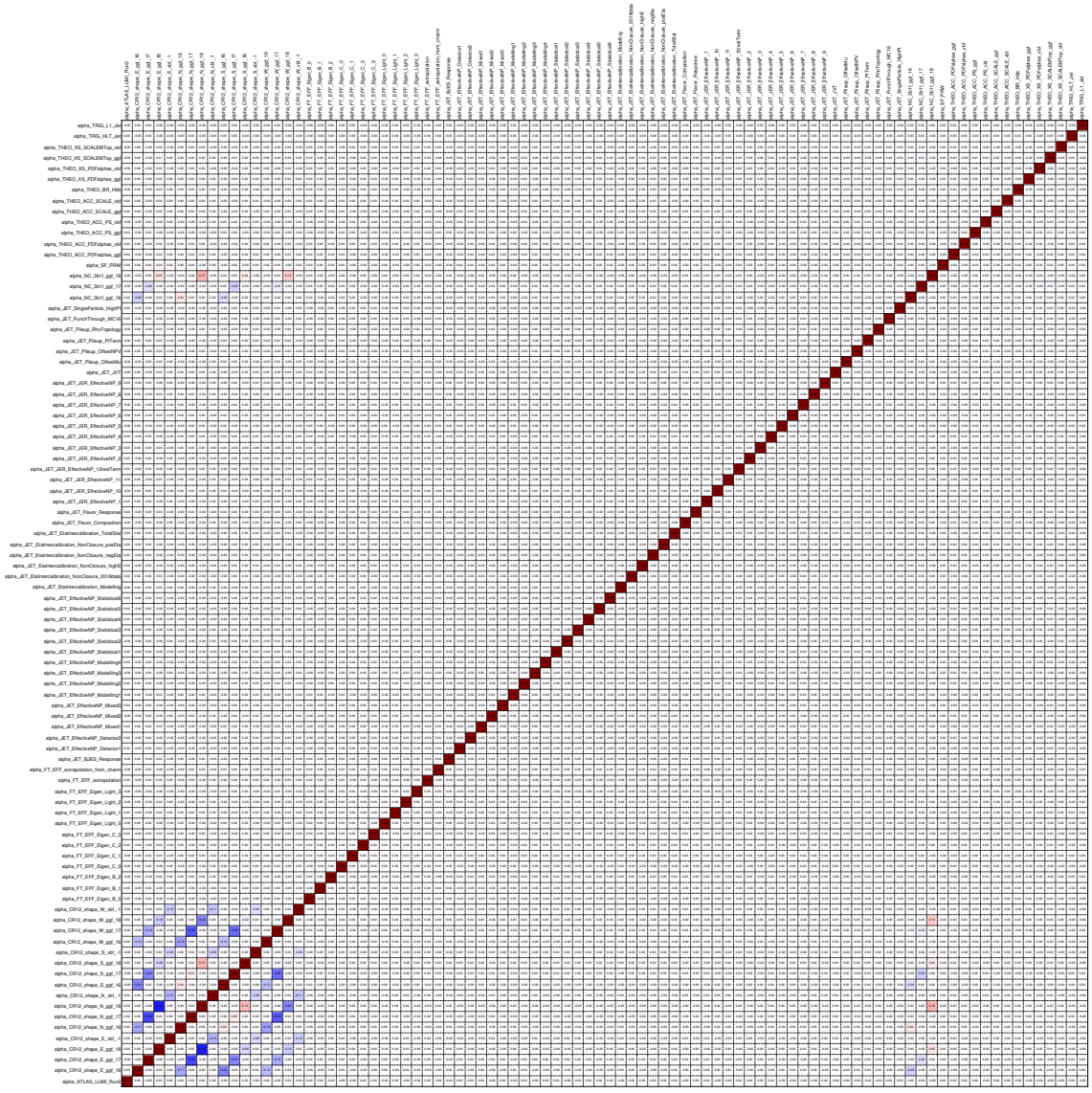


Figure F.4: Correlations between all the nuisance parameter in a SM *HH* signal and background model fit to data in the combined ggF and VBF channels displayed in the form of a correlation matrix.

Bibliography

[1] ATLAS Collaboration. *Search for non-resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. Geneva: CERN, June 2022. URL: http://cds.cern.ch/record/2811390 (cit. on pp. 7, 9, 71, 181).

[2] ATLAS Collaboration. "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Rev. D* 94 (2016), p. 052002. DOI: 10.1103/PhysRevD.94.052002. arXiv: 1606.04782 [hep-ex] (cit. on pp. 7, 9, 172, 188, 198).

[3] ATLAS Collaboration. "Search for the $HH \to b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 07 (2020), p. 108. DOI: 10.1007/JHEP07(2020)108. arXiv: 2001.05178 [hep-ex]. [Erratum: JHEP **01** (2021) 145] (cit. on pp. 7, 9, 102, 191, 198).

[4] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), p. 1. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (cit. on pp. 15, 23).

[5] CMS Collaboration. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV". In: *JHEP* 06 (2013), p. 081. DOI: 10.1007/JHEP06(2013)081. arXiv: 1303.4571 [hep-ex] (cit. on pp. 15, 23).

[6] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321 (cit. on p. 19).

[7] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508 (cit. on p. 19).

[8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585 (cit. on p. 19).

[9] Particle Data Group. *CKM Quark-Mixing Matrix*. 2022. DOI: 10.1093/ptep/ptac097 (cit. on p. 21).

[10] "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery". In: *Nature* 607.7917 (July 2022), pp. 52–59. DOI: 10.1038/s41586-022-04893-w. URL: https://doi.org/10.1038%2Fs41586-022-04893-w (cit. on p. 23).

[11] ATLAS Collaboration. "Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector". In: *Eur. Phys. J. C* 82 (2022), p. 717. DOI: 10.1140/epjc/s10052-022-10588-3. arXiv: 2201.11428 [hep-ex] (cit. on p. 24).

[12] LHC Higgs Cross Section Working Group et al. "LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle". In: (2012). DOI: 10.48550/ARXIV.1209.0040. URL: https://arxiv.org/abs/1209.0040 (cit. on p. 25).

[13] Massimiliano Grazzini et al. "Higgs boson pair production at NNLO with top quark mass effects". In: *JHEP* 05 (2018), p. 059. DOI: 10.1007/JHEP05(2018)059. arXiv: 1803.02463 [hep-ph] (cit. on pp. 25, 73).

[14] ATLAS Collaboration. *Constraints on the Higgs boson self-coupling from the combination of single-Higgs and double-Higgs production analyses performed with the ATLAS experiment.* ATLAS-CONF-2019-049. 2019. URL: https://cds.cern.ch/record/2693958 (cit. on p. 25).

[15] ATLAS Collaboration. "Combined measurements of Higgs boson production and decay using up to $80\,\text{fb}^{-1}$ of proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$ collected with the ATLAS experiment". In: *Phys. Rev. D* 101 (2020), p. 012002. DOI: 10.1103/PhysRevD.101.012002. arXiv: 1909.02845 [hep-ex] (cit. on pp. 25, 28, 76).

[16] Marcela Carena et al. "Alignment limit of the NMSSM Higgs sector". In: *Physical Review D* 93.3 (Feb. 2016). DOI: 10.1103/physrevd.93.035013. URL: https://doi.org/10.1103%2Fphysrevd.93.035013 (cit. on p. 26).

[17] Vernon Barger et al. "Effects of genuine dimension-six Higgs operators". In: *Physical Review D* 67.11 (June 2003). DOI: 10.1103/physrevd.67.115001. URL: https://doi.org/10.1103%2Fphysrevd.67.115001 (cit. on p. 26).

[18] Aaron Pierce, Jesse Thaler, and Lian-Tao Wang. "Disentangling dimension six operators through di-Higgs boson production". In: *Journal of High Energy Physics* 2007.05 (May 2007), pp. 070–070. DOI: 10.1088/1126-6708/2007/05/070. URL: https://doi.org/10.1088/1126-6708/2007/05/070 (cit. on p. 26).

[19] T. D. Lee. "A Theory of Spontaneous T Violation". In: *Phys. Rev. D* 8 (4 Aug. 1973), pp. 1226–1239. DOI: 10.1103/PhysRevD.8.1226. URL: https://link.aps.org/doi/10.1103/PhysRevD.8.1226 (cit. on p. 26).

[20] ATLAS Collaboration. *Combination of searches for non-resonant and resonant Higgs boson pair production in the $b\bar{b}\gamma\gamma$, $b\bar{b}\tau^+\tau^-$ and $b\bar{b}b\bar{b}$ decay channels using pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*. Tech. rep. Geneva: CERN, Oct. 2021. URL: https://cds.cern.ch/record/2786865 (cit. on pp. 26, 151, 199).

[21] Xiaohu Sun et al. *HH non-resonant ggF production: MC validation and request*. URL: https://indico.cern.ch/event/866189/contributions/3649797/attachments/1953958/3244956/HH-XiaohuSUN-2019-11-29-HHMC.pdf (cit. on p. 27).

[22] Frédéric A. Dreyer and Alexander Karlberg. "Vector-boson fusion Higgs pair production at N3LO". In: *Physical Review D* 98.11 (Dec. 2018). ISSN: 2470-0029. DOI: 10.1103/physrevd.98.114016. URL: http://dx.doi.org/10.1103/PhysRevD.98.114016 (cit. on pp. 27, 73).

[23] David B. Kaplan and Howard Georgi. "SU(2) × U(1) breaking by vacuum misalignment". In: *Physics Letters B* 136.3 (1984), pp. 183–186. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(84)91177-8. URL: https://www.sciencedirect.com/science/article/pii/0370269384911778 (cit. on p. 28).

[24] Kaustubh Agashe, Roberto Contino, and Alex Pomarol. "The minimal composite Higgs model". In: *Nuclear Physics B* 719.1 (2005), pp. 165–187. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2005.04.035. URL: https://www.sciencedirect.com/science/article/pii/S0550321305003445 (cit. on p. 28).

[25] Roberto Contino, Leandro Da Rold, and Alex Pomarol. "Light custodians in natural composite Higgs models". In: *Phys. Rev. D* 75 (5 Mar. 2007),

p. 055014. DOI: 10.1103/PhysRevD.75.055014. URL: https://link.aps.org/doi/10.1103/PhysRevD.75.055014 (cit. on p. 28).

[26] Roberto Contino et al. "Strong double higgs production at the LHC". In: *Journal of High Energy Physics* 2010.5 (May 2010). DOI: 10.1007/jhep05(2010)089. URL: https://doi.org/10.1007%2Fjhep05%282010%29089 (cit. on pp. 28, 29).

[27] Fady Bishara, Roberto Contino, and Juan Rojo. "Higgs pair production in vector-boson fusion at the LHC and beyond". In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5037-9. URL: http://dx.doi.org/10.1140/epjc/s10052-017-5037-9 (cit. on pp. 28, 29, 148).

[28] CMS Collaboration. *Search for Higgs boson pair production via vector boson fusion with highly Lorentz-boosted Higgs bosons in the four b quark final state at* $\sqrt{s} = 13$ *TeV*. Tech. rep. Geneva: CERN, 2021. URL: https://cds.cern.ch/record/2776802 (cit. on p. 29).

[29] Particle Data Group. *Review of Particle Physics: 11. Status of Higgs Boson Physics*. 2022. DOI: 10.1093/ptep/ptac097 (cit. on p. 32).

[30] LHC Higgs Cross Section Working Group. *Vol. 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*. DOI: https://doi.org/10.23731/CYRM-2017-002 (cit. on p. 34).

[31] ATLAS Collaboration. *ATLAS Detector and Physics Performance: Technical Design Report, Volume 1*. ATLAS-TDR-14; CERN-LHCC-99-014. 1999. URL: https://cds.cern.ch/record/391176 (cit. on p. 36).

[32] ATLAS Collaboration. *ATLAS Detector and Physics Performance: Technical Design Report, Volume 2*. ATLAS-TDR-15; CERN-LHCC-99-015. 1999. URL: https://cds.cern.ch/record/391177 (cit. on p. 36).

[33] CMS Collaboration. "The CMS experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004 (cit. on p. 36).

[34] ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08002 (cit. on p. 36).

[35] LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08005 (cit. on p. 36).

[36] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003 (cit. on p. 38).

[37] Joao Pequenao. *Computer generated image of the ATLAS inner detector.* Mar. 2008. URL: https://cds.cern.ch/record/1095926 (cit. on p. 39).

[38] Izaak Neutelings. *ATLAS Coordinate System Diagram.* June 2017. URL: https://tikz.net/axis3d_cms/ (cit. on p. 39).

[39] ATLAS Collaboration. *ATLAS Inner Detector: Technical Design Report, Volume 1.* ATLAS-TDR-4; CERN-LHCC-97-016. 1997. URL: https://cds.cern.ch/record/331063 (cit. on p. 40).

[40] ATLAS Collaboration. *ATLAS Inner Detector: Technical Design Report, Volume 2.* ATLAS-TDR-5, CERN-LHCC-97-017. 1997. URL: https://cds.cern.ch/record/331064 (cit. on p. 40).

[41] ATLAS Collaboration. *Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment.* July 2020. URL: https://cds.cern.ch/record/2723878 (cit. on p. 42).

[42] ATLAS Collaboration. *ATLAS Pixel Detector: Technical Design Report*. ATLAS-TDR-11; CERN-LHCC-98-013. 1998. URL: https://cds.cern.ch/record/381263 (cit. on p. 43).

[43] ATLAS Collaboration. *ATLAS Insertable B-Layer Technical Design Report*. ATLAS-TDR-19; CERN-LHCC-2010-013. 2010. URL: https://cds.cern.ch/record/1291633 (cit. on p. 43).

[44] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008). Table 4.1. DOI: 10.1088/1748-0221/3/08/s08003. URL: https://doi.org/10.1088/1748-0221/3/08/s08003 (cit. on p. 45).

[45] ATLAS Collaboration. *ATLAS Liquid Argon Calorimeter: Technical Design Report*. ATLAS-TDR-2; CERN-LHCC-96-041. 1996. URL: https://cds.cern.ch/record/331061 (cit. on p. 45).

[46] ATLAS Collaboration. *ATLAS Tile Calorimeter: Technical Design Report*. ATLAS-TDR-3; CERN-LHCC-96-042. 1996. URL: https://cds.cern.ch/record/331062 (cit. on p. 45).

[47] ATLAS Collaboration. *ATLAS Liquid Argon Calorimeter Phase-I Upgrade: Technical Design Report*. ATLAS-TDR-022; CERN-LHCC-2013-017. 2013. URL: https://cds.cern.ch/record/1602230 (cit. on p. 45).

[48] Joao Pequenao. *Computer Generated image of the ATLAS calorimeter*. Mar. 2008. URL: https://cds.cern.ch/record/1095927 (cit. on p. 46).

[49] ATLAS Collaboration. *ATLAS Muon Spectrometer: Technical Design Report*. ATLAS-TDR-10; CERN-LHCC-97-022. CERN, 1997. URL: https://cds.cern.ch/record/331068 (cit. on p. 49).

[50] ATLAS Collaboration. *ATLAS Muon Spectrometer Phase-II Upgrade: Technical Design Report*. ATLAS-TDR-026; CERN-LHCC-2017-017. 2017. URL: https://cds.cern.ch/record/2285580 (cit. on p. 49).

[51] ATLAS Collaboration. *ATLAS New Small Wheel: Technical Design Report*. ATLAS-TDR-020; CERN-LHCC-2013-006. 2013. URL: https://cds.cern.ch/record/1552862 (cit. on p. 49).

[52] Joao Pequenao. *Computer generated image of the ATLAS Muons subsystem*. Mar. 2008. URL: https://cds.cern.ch/record/1095929 (cit. on p. 50).

[53] ATLAS Collaboration. *ATLAS End-Cap Toroids: Magnet Project Technical Design Report, Volume 3*. ATLAS-TDR-8; CERN-LHCC-97-020. CERN, 1997. URL: https://cds.cern.ch/record/331066 (cit. on p. 49).

[54] ATLAS Collaboration. *ATLAS Barrel Toroid: Magnet Project Technical Design Report, Volume 2*. ATLAS-TDR-7; CERN-LHCC-97-019. 1997. URL: https://cds.cern.ch/record/331065 (cit. on p. 49).

[55] ATLAS Collaboration. "Operation of the ATLAS trigger system in Run 2". In: *JINST* 15 (2020), P10004. DOI: 10.1088/1748-0221/15/10/P10004. arXiv: 2007.12539 [physics.ins-det] (cit. on p. 50).

[56] ATLAS Collaboration. *ATLAS Level-1 Trigger: Technical Design Report*. ATLAS-TDR-12; CERN-LHCC-98-014. 1998. URL: https://cds.cern.ch/record/381429 (cit. on p. 50).

[57] ATLAS Collaboration. *ATLAS High-Level Trigger, Data Acquisition and Controls: Technical Design Report*. ATLAS-TDR-16; CERN-LHCC-2003-022. 2003. URL: https://cds.cern.ch/record/616089 (cit. on p. 50).

[58] ATLAS Collaboration. *ATLAS TDAQ System Phase-I Upgrade: Technical Design Report*. ATLAS-TDR-023; CERN-LHCC-2013-018. 2013. URL: https://cds.cern.ch/record/1602235 (cit. on p. 50).

[59] T Cornelissen et al. "The new ATLAS track reconstruction (NEWT)". In: *Journal of Physics: Conference Series* 119.3 (July 2008), p. 032014. DOI: 10.1088/1742-6596/119/3/032014. URL: https://dx.doi.org/10.1088/1742-6596/119/3/032014 (cit. on p. 53).

[60] ATLAS Collaboration. "Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2". In: *Eur. Phys. J. C* 77 (2017), p. 673. DOI: 10.1140/epjc/s10052-017-5225-7. arXiv: 1704.07983 [hep-ex] (cit. on p. 53).

[61] R. Frühwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(87)90887-4. URL: https://www.sciencedirect.com/science/article/pii/0168900287908874 (cit. on p. 54).

[62] ATLAS Collaboration. "Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC". In: *Eur. Phys. J. C* 77 (2017), p. 332. DOI: 10.1140/epjc/s10052-017-4887-5. arXiv: 1611.10235 [hep-ex] (cit. on p. 55).

[63] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph] (cit. on p. 56).

[64] 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] (cit. on p. 58).

[65] ATLAS Collaboration. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *Eur. Phys. J. C* 77 (2017), p. 490. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex] (cit. on p. 58).

[66] ATLAS Collaboration. "Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: (2020). arXiv: 2007.02645 [hep-ex] (cit. on pp. 58, 59, 168).

[67] P. Billoir and S. Qian. "Fast vertex fitting with a local parametrization of tracks". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 311.1 (1992), pp. 139–150. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(92)90859-3. URL: https://www.sciencedirect.com/science/article/pii/0168900292908593 (cit. on p. 61).

[68] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. ATL-PHYS-PUB-2017-013. 2017. URL: https://cds.cern.ch/record/2273281 (cit. on pp. 62, 63).

[69] ATLAS Collaboration. *Secondary vertex finding for jet flavour identification with the ATLAS detector*. ATL-PHYS-PUB-2017-011. 2017. URL: https://cds.cern.ch/record/2270366 (cit. on p. 62).

[70] ATLAS Collaboration. *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*. ATL-PHYS-PUB-2018-025. 2018. URL: https://cds.cern.ch/record/2645405 (cit. on p. 62).

[71] 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: *Eur. Phys. J. C* 79 (2019), p. 970. DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120 [hep-ex] (cit. on pp. 63, 64, 66).

[72] ATLAS Collaboration. *Figure 1a. Expected performance of the 2019 ATLAS b-taggers*. Sept. 2019. URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-005/ (cit. on p. 65).

[73] ATLAS Collaboration. *Flavour-tagging efficiency corrections for the 2019 ATLAS PFlow jet and VR track jets b-taggers with the full LHC Run II dataset*. Feb. 2021. URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2021-001/ (cit. on p. 66).

[74] Particle Data Group. *Summary Table: Bottom Mesons*. 2022. DOI: 10.1093/ptep/ptac097 (cit. on p. 66).

[75] ATLAS Collaboration. *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*. ATLAS-CONF-2020-030. 2020. URL: https://cds.cern.ch/record/2725736 (cit. on p. 66).

[76] P. Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146 [hep-ph] (cit. on p. 73).

[77] S. Frixione, P. Nason and C. Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph] (cit. on p. 73).

[78] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph] (cit. on p. 73).

[79] T Sjostrand et al. "An Introduction to PYTHIA 8.2". In: *Comput.Phys.Commun.* 191 (2015), pp. 159–177. arXiv: 1410.3012 [hep-ph] (cit. on p. 73).

[80] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *J. Phys. G* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph] (cit. on p. 73).

[81] 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: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph] (cit. on p. 73).

[82] Richard D. Ball et al. "Parton distributions with LHC data". In: *Nucl. Phys. B* 867 (2013), p. 244. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph] (cit. on p. 73).

[83] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB-2014-021. 2014. URL: https://cds.cern.ch/record/1966419 (cit. on p. 73).

[84] ATLAS Collaboration. "Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Lett. B* 800 (2020), p. 135103. DOI: 10.1016/j.physletb.2019.135103. arXiv: 1906.02025 [hep-ex] (cit. on p. 74).

[85] *Validation of signal Monte Carlo event generation in searches for Higgs boson pairs with the ATLAS detector*. Tech. rep. Geneva: CERN, Feb. 2019. URL: https://cds.cern.ch/record/2665057 (cit. on p. 74).

[86] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph] (cit. on p. 75).

[87] ATLAS Collaboration. *Simulation of top-quark production for the ATLAS experiment at* $\sqrt{s} = 13$ *TeV*. ATL-PHYS-PUB-2016-004. 2016. URL: https://cds.cern.ch/record/2120417 (cit. on p. 75).

[88] GEANT4 Collaboration, S. Agostinelli, et al. "GEANT4 – a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on p. 75).

[89] S. Amoroso et al. *Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report*. 2020. arXiv: 2003.01700 [hep-ph] (cit. on p. 75).

[90] Dale Abbott et al. *Supporting Document: The Search for Non-Resonant ggF and VBF HH Production Decaying to the 4b Final State Using the Full Run-2*

Data. Tech. rep. Geneva: CERN, Sept. 2021. URL: https://cds.cern.ch/record/2780536 (cit. on pp. 80, 125, 171).

[91] ATLAS Collaboration. "Search for the $HH \to b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 07 (2020), p. 108. DOI: 10.1007/JHEP07(2020)108. arXiv: 2001.05178 [hep-ex] (cit. on pp. 83, 147). Erratum: in: *JHEP* 01 (2021), p. 145. DOI: 10.1007/JHEP01(2021)145.

[92] Dale Charles Abbott et al. *Supporting Document: The Search for Resonant HH Production Decaying to the 4b Final State Using the Full Run-2 Data and the Resolved Analysis Channel*. Tech. rep. Geneva: CERN, Feb. 2020. URL: https://cds.cern.ch/record/2708605 (cit. on p. 87).

[93] ATLAS Collaboration. "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 01 (2019), p. 030. DOI: 10.1007/JHEP01(2019)030. arXiv: 1804.06174 [hep-ex] (cit. on pp. 99, 147).

[94] ATLAS Collaboration. *Trigger Menu in 2016*. ATL-DAQ-PUB-2017-001. 2017. URL: https://cds.cern.ch/record/2242069 (cit. on p. 109).

[95] ATLAS Collaboration. *Trigger Menu in 2017*. ATL-DAQ-PUB-2018-002. 2018. URL: https://cds.cern.ch/record/2625986 (cit. on p. 109).

[96] ATLAS Collaboration. *Trigger Menu in 2018*. ATL-DAQ-PUB-2019-001. 2019. URL: https://cds.cern.ch/record/2693402 (cit. on p. 109).

[97] ATLAS Collaboration. *Formulae for Estimating Significance*. ATL-PHYS-PUB-2020-025. 2020. URL: https://cds.cern.ch/record/2736148 (cit. on p. 115).

[98] George V. Moustakides and Kalliopi Basioti. *Training Neural Networks for Likelihood/Density Ratio Estimation*. 2019. DOI: 10.48550/ARXIV.1911.00405. URL: https://arxiv.org/abs/1911.00405 (cit. on p. 125).

[99] B. Efron. "Bootstrap Methods: Another Look at the Jackknife". In: *The Annals of Statistics* 7.1 (1979), pp. 1–26. DOI: 10.1214/aos/1176344552. URL: https://doi.org/10.1214/aos/1176344552 (cit. on p. 127).

[100] B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994. ISBN: 9780412042317 (cit. on pp. 127, 129).

[101] ATLAS Collaboration. "A search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Rev. Lett.* 121 (2018), p. 191801. DOI: 10.1103/PhysRevLett.121.191801. arXiv: 1808.00336 [hep-ex] (cit. on p. 147). Erratum: in: *Phys. Rev. Lett.* 122 (2019), p. 089901. DOI: 10.1103/PhysRevLett.122.089901.

[102] ATLAS Collaboration. "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: *JHEP* 11 (2018), p. 040. DOI: 10.1007/JHEP11(2018)040. arXiv: 1807.04873 [hep-ex] (cit. on p. 147).

[103] Kyle Cranmer et al. "HistFactory: A tool for creating statistical models for use with RooFit and RooStats". In: CERN-OPEN-2012-016 (2012). URL: https://cds.cern.ch/record/1456844 (cit. on p. 157).

[104] Kyle Cranmer. "Practical Statistics for the LHC". In: (2015). DOI: 10.48550/ARXIV.1503.07622. URL: https://arxiv.org/abs/1503.07622 (cit. on p. 157).

[105] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an] (cit. on pp. 160, 162). Erratum: in: *Eur. Phys. J. C* 73 (2013), p. 2501. DOI: 10.1140/epjc/s10052-013-2501-z.

[106] LHC Higgs Cross Section Working Group. *SM Higgs Branching Ratios and Total Decay Widths*. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR?rev=22 (cit. on p. 170).

[107] Bruce Mellado Garcia et al. "CERN Report 4: Part I Standard Model Predictions". In: (May 2016). URL: https://cds.cern.ch/record/2150771 (cit. on p. 170).

[108] ATLAS Physics Modelling Group (PMG). *PMG Systematic Uncertainty Recipes*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/%5C%5CPmgSystematicUncertaintyRecipes?rev=27 (cit. on p. 171).

[109] ATLAS Collaboration. *Luminosity determination in pp collisions at* $\sqrt{s}$ = 13 *TeV using the ATLAS detector at the LHC*. ATLAS-CONF-2019-021. 2019. URL: https://cds.cern.ch/record/2677054 (cit. on p. 171).

[110] CMS Collaboration. *Search for Higgs boson pair production in the four b quark final state in proton-proton collisions at* $\sqrt{s}$ = *13 TeV*. 2022. DOI: 10.48550/ARXIV.2202.09617. URL: https://arxiv.org/abs/2202.09617 (cit. on pp. 188, 189, 191, 198).

[111] CMS Collaboration. *Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks*. 2022. DOI: 10.48550/ARXIV.2205.06667. URL: https://arxiv.org/abs/2205.06667 (cit. on p. 191).

[112] Biagio Di Micco et al. "Higgs boson potential at colliders: Status and perspectives". In: *Reviews in Physics* 5 (Nov. 2020), p. 100045. DOI: 10.1016/j.revip.2020.100045. URL: https://doi.org/10.1016%2Fj.revip.2020.100045 (cit. on pp. 199, 200).

[113] Michelangelo Mangano et al. *FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1. Future Circular Collider*. Tech. rep. 6. Geneva: CERN, 2019. DOI: 10.1140/epjc/s10052-019-6904-3. URL: https://cds.cern.ch/record/2651294 (cit. on p. 200).

[114] Michael Benedikt et al. *FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3. Future Circular Collider*. Tech. rep. Geneva: CERN, 2019. DOI: 10.1140/epjst/e2019-900087-0. URL: https://cds.cern.ch/record/2651300 (cit. on p. 200).

[115] Michael Benedikt et al. *FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2. Future Circular Collider*. Tech. rep. 2. Geneva: CERN, 2019. DOI: 10.1140/epjst/e2019-900045-4. URL: https://cds.cern.ch/record/2651299 (cit. on p. 200).

[116] Ties Behnke et al. *The International Linear Collider Technical Design Report - Volume 1: Executive Summary*. 2013. DOI: 10.48550/ARXIV.1306.6327. URL: https://arxiv.org/abs/1306.6327 (cit. on p. 200).

[117] M Aicheler et al. *A Multi-TeV Linear Collider Based on CLIC Technology: CLIC Conceptual Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2012. DOI: 10.5170/CERN-2012-007. URL: https://cds.cern.ch/record/1500095 (cit. on p. 200).

[118] ATLAS Jet and Etmiss Combined Performance Group. *Small-R (R=0.4) jet energy scale and jet energy resolution uncertainties for full run 2 analyses (release 21)*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetUncertaintiesRel21Summer2018SmallR (cit. on p. 224).

[119] Flavour Tagging Combined Performance Group. *Systematic uncertainties: Using Systematics In Analysis*. URL: https://ftag-docs.docs.cern.ch/cdi/systematics/systematic-strategies/ (cit. on p. 224).

[120] Flavour Tagging Combined Performance Group. *Systematic uncertainties: Eigenvector Decomposition*. URL: https://ftag-docs.docs.cern.ch/cdi/systematics/eigenvectordecomp/ (cit. on p. 224).