Evidence for the Production of the Standard Model Higgs Boson Produced via Vector Boson Fusion in the WW^* Channel at the ATLAS Detector

by

Koos van Nieuwkoop

M.Sc., Simon Fraser University, 2013
B.Sc., University of the Fraser Valley, 2010

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Physics
Faculty of Science

SIMON FRASER UNIVERSITY
Spring 2017

Approval

Name:	**Koos van Nieuwkoop**
Degree:	**Doctor of Philosophy (Physics)**
Title:	***Evidence for the Production of the Standard Model Higgs Boson Produced via Vector Boson Fusion in the WW^* Channel at the ATLAS Detector***
Examining Committee:	**Chair:** Malcolm Kennett Associate Professor

Bernd Stelzer
Senior Supervisor
Associate Professor

Michel Vetterli
Supervisor
Professor

Dugan O'Neil
Supervisor
Professor

Howard Trottier
Internal Examiner
Professor

Dean Karlen
External Examiner
Professor
Department of Physics and Astronomy
University of Victoria

Date Defended/Approved: April 18, 2017

Abstract

In 2012, the ATLAS and CMS experiments at CERN's Large Hadron Collider announced they had each observed a new particle with a mass of about 125 GeV/c^2. Given the available data, the properties of this particle are consistent with the Higgs boson predicted by the Standard Model of particle physics (SM). The Higgs boson, as proposed within the SM, is the simplest manifestation of the Brout-Englert-Higgs mechanism. This discovery was driven by the gluon fusion (ggF) production mode, the dominant Higgs boson production mechanism at the LHC. The SM also predicts that the Higgs boson can be produced by the fusion of two weak vector bosons (VBF). Measuring VBF Higgs boson production is an important test of the SM but it is challenging to measure given its cross section is an order of magnitude smaller than that of ggF.

After $H \to b\bar{b}$, $H \to WW^*$ is the dominant decay channel for the SM Higgs boson at 125 GeV/c^2 and is therefore a promising channel to measure its properties. In addition, the VBF $H \to WW^*$ search channel makes it possible to probe the exclusive coupling of the Higgs boson to the weak vector bosons. Precise measurements of these coupling strengths make it possible to constrain new models of physics beyond the SM.

Despite its relatively large branching ratio, $H \to WW^* \to \ell\nu\ell\nu$ is a challenging channel to search for the Higgs boson because of the neutrinos in the final state which are not directly detectable by the ATLAS detector. Consequently, it is not possible to fully reconstruct the mass of the WW^* system. Furthermore, there are several backgrounds that have the same signature in the detector as the signal. Top quark pair production is the largest background in this analysis.

A multivariate analysis technique, based on an eight-variable boosted decision tree (BDT), is used to search for VBF $H \to WW^* \to \ell\nu\ell\nu$ in the Run-I data and a subset of the Run-II data. This analysis provides the first evidence for VBF $H \to WW^* \to \ell\nu\ell\nu$ with a significance of 3.2 standard deviations in Run-I and 1.9 standard deviations in Run-II. The measured signal strength relative to the rate predicted by the SM for VBF $H \to WW^* \to \ell\nu\ell\nu$ is 1.3 ± 0.5 using the Run-I data, and $1.7^{+1.1}_{-0.9}$ using a fraction of the Run-II data.

Uit waardering opgedragen
aan mijn ouders

Acknowledgements

I am enormously indebted to my senior supervisor, Bernd Stelzer, for his guidance during my graduate studies. Bernd, I have learned a lot from you and I am grateful for your patience and unwavering optimism.

There are many people who have contributed to the $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ analysis but I would like to especially thank a few individuals I was privileged to collaborate with. Thanks to the Run-I VBF BDT team: Bernd Stelzer, Doug Schouten, Ben Cerio, Simon Viel, Oliver Stelzer-Chilton, Jennifer Hsu, Heberth Torres, and Ashutosh Kotwal. It was a long road from the first iterations of the analysis to publication, but it paid off in the end. Also thanks to Heberth Torres, Jennifer Hsu, Claudia Bertella and Yun Ju Lu who I worked with on the first Run-II result.

Thanks to Dugan O'Neil and Mike Vetterli for their feedback as supervisory committee. Also thanks to the SFU/TRIUMF postdocs and students who have helped shape my graduate career: Sina Bahrasemani, James Beare, Quentin Buat, Noel Dawe, Etienne Dreyer, Jennifer Godfrey, Jamie Horton, Jelena Jovićević, Jiri Kvita, Konstantin Lehmann, Dan Mori, Kate Pachal, Doug Schouten, Andres Tanasijczuk, Darren Temple, Heberth Torres, and Michel Trottier-McDonald. I owe a special thank-you to Heberth Torres and Jelena Jovićević (and of course my friends at Cicikanto) for making my extended stay at CERN particularly memorable.

Finally, thanks to my family and friends for supporting my academic endeavors.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

First postulated in the 1960s, the Higgs boson has long been considered a missing puzzle piece in the Standard Model of particle physics (SM). On July 4, 2012, physicists at the ATLAS and CMS experiments at the European Organization for Nuclear Research (CERN) were elated to announce the discovery of a new particle consistent with the SM Higgs boson [1, 2]. The signal in the initial discovery was dominated by Higgs boson production via the fusion of two gluons. However, the SM also predicts additional production modes with significantly smaller contributions. This thesis summarizes a search for Higgs boson production via the fusion of two vector bosons (VBF), the second largest contribution to Higgs boson production at the Large Hadron Collider (LHC). Measuring this process is an important test of the validity of the SM. Significant deviations from the SM predictions suggest new physics beyond the current understanding of elementary particle physics.

Chapter 1 provides an introduction to the theoretical motivations for searching for VBF Higgs boson production. The accelerator and detector used for this search are described in chapter 2 and the methods used to interpret the detector signals in terms of SM particles are summarized in chapter 3. Next, a general overview of the analysis and the adopted strategy are presented in chapter 4. A description of the Monte Carlo generators used to simulate theoretical predictions is given in Chapter 5. Chapters 6 and 7 define the regions of phase space selected to search for VBF Higgs boson production while chapter 8 describes the theoretical and experimental uncertainties on the predictions in these regions. The statistical analysis of the data is described in chapter 9. Results extracted from the data and their interpretation are detailed in chapter 10. Finally, a summary of the results is presented in chapter 11.

1.1 The Standard Model of Particle Physics

The SM is a highly successful theory that describes the elementary particles of nature and the forces through which they interact [3]. Elementary particles carry various types of charge which determine how they interact with other particles. They include electric charge for electromagnetic interactions, weak charge for weak interactions, and color charge for strong interactions. The particles in the SM are classified as fermions, gauge bosons, or a spin-0 scalar boson.

There are 12 spin-$\frac{1}{2}$ particles in the SM which are known as fermions. Fermions can be divided into 6 quarks (table 1.1) and 6 leptons (table 1.2). These can be further categorized into three generations of particle pairs. For a specific fermion type, the particle pairs in each generation only differ in mass[1] and flavour. Only first generation particles are stable and make up atoms, the building blocks of ordinary matter.

Generation	Quark	Electric charge (e)	Mass	Interactions
1	up (u)	$+\frac{2}{3}$	$2.2^{+0.6}_{-0.4}$ MeV	E, W, S
	down (d)	$-\frac{1}{3}$	$4.7^{+0.5}_{-0.4}$ MeV	E, W, S
2	charm (c)	$+\frac{2}{3}$	1.27 ± 0.03 GeV	E, W, S
	strange (s)	$-\frac{1}{3}$	96^{+8}_{-4} MeV	E, W, S
3	top (t)	$+\frac{2}{3}$	173 ± 1 GeV	E, W, S
	bottom (b)	$-\frac{1}{3}$	$4.18^{+0.04}_{-0.03}$ GeV	E, W, S

Table 1.1: Spin-$\frac{1}{2}$ fermions (quarks) in the SM [4]. Besides the top quark mass, the quark masses are calculated using the $\overline{\text{MS}}$ renormalization scheme at a scale of 2 GeV for the light quarks, and at a scale equal to the quark mass for the charm and bottom quarks. As indicated by the last column, quarks can interact via the electromagnetic (E), weak (W), and strong (S) interactions.

Generation	Lepton	Electric charge (e)	Mass	Interactions
1	electron (e)	-1	0.511 MeV	E, W
	electron neutrino (ν_e)	0	< 2 eV	W
2	muon (μ)	-1	0.106 GeV	E, W
	muon neutrino (ν_μ)	0	< 2 eV	W
3	tau (τ)	-1	1.78 GeV	E, W
	tau neutrino (ν_τ)	0	< 2 eV	W

Table 1.2: Spin-$\frac{1}{2}$ fermions (leptons) in the SM [4]. All leptons can interact via the weak interaction (W), but only the electron, muon, and tau can interact via the electromagnetic interaction (E).

Quarks can interact by any of the three SM interactions. Even though table 1.1 quotes masses for single quarks, these are not observed in nature. The property of the strong force known as color confinement requires that all observed states are color neutral and so quarks

---

[1] In the remainder of this thesis, all masses and momenta are quoted using natural units ($\hbar = c = 1$).

are measured in bound states called hadrons. Consequently, the quark masses are deduced within the SM framework by measuring the properties of hadronic interactions. Besides the top quark mass which is measured directly, the quoted quark masses are calculated using the $\overline{\text{MS}}$ renormalization scheme [4].

Leptons can interact via the weak interaction, but not via the strong interaction. The electron, muon, and tau are the only leptons that can interact via the electromagnetic interaction and consequently are sometimes referred to as charged leptons.

Neutrinos (ν) are leptons that can only interact via the weak interaction. As a result, they are extremely difficult to detect and their presence at collider experiments is inferred rather than measured directly. As discussed in more detail later, this presents a significant challenge for the analysis described in this thesis. Experimental evidence suggests that they have small non-zero masses [5, 6].

The SM describes the electromagnetic, weak, and strong interactions. Thanks to efforts by Glashow, Salam [7], and Weinberg [8], the electromagnetic and weak interactions can be described by a unified electroweak theory. This hints that all the fundamental forces of nature can be unified at some energy scale known as the grand unified theory (GUT) scale. The force of gravity is not explained by the SM.

Gauge bosons are spin-1 particles (table 1.3) that mediate interactions between fermions. The unified electroweak interaction is mediated by four spin-1 bosons, the photon (γ), Z^0, W^+, and W^-. Eight colored gluons (g) are responsible for mediating the strong interaction.

Boson	Force	Electric charge (e)	Mass
γ (photon)	Electroweak	0	0 GeV
g (gluon)	Strong	0	0 GeV
$W^\pm$	Electroweak	± 1	80.4 GeV
Z^0	Electroweak	0	91.2 GeV

Table 1.3: Spin-1 force carriers in the SM [4].

The Higgs boson (table 1.4) is a massive spin-0 boson. It was first postulated in 1964 but it took nearly half a century before the first hints of this particle surfaced. Theoretically, it is a quantum excitation of the Higgs field, a field postulated to permeate all of space. Particles like the photon and gluon do not interact with this field and so remain massless while massive particles acquire their mass by interacting with the Higgs field.

Boson	Electric charge (e)	Spin	Mass
H^0	0	0	125.1 ± 0.2 GeV

Table 1.4: Properties of the Higgs boson, the only fundamental spin-0 boson in the SM [4]. The value for the mass is a combined measurement from the ATLAS and CMS experiments.

Each particle mentioned has a corresponding antiparticle which has the same mass but inverted charges. Some particles such as the H and Z are their own antiparticles.

As documented by Wu and her collaborators in 1957, the weak interaction is a *chiral* interaction [9] unlike the electromagnetic and strong interactions. A chiral interaction is one that is not invariant under *parity transformations* which means that the interaction is not identical to its mirror image. In the SM, the W boson only couples to particles with left-handed chirality and antiparticles with right-handed chirality. For massless particles, the chirality of a particle is the same as its *helicity.* A particle is defined to have left-handed helicity if the projection of its spin onto its momentum is negative and right-handed helicity if it is positive.

1.2 Spontaneous Symmetry Breaking

Mathematically, the SM is expressed using quantum field theory (QFT). Each particle in the SM corresponds to a quantum excitation of its corresponding field that permeates space and time. The dynamics of the particles in the SM are expressed in terms of a Lagrangian density, $\mathcal{L}_{SM}$. $\mathcal{L}_{SM}$ can be thought of as the QFT analogue of a Lagrangian in classical mechanics.

Free non-interacting particles can be described if $\mathcal{L}_{SM}$ is symmetric under global gauge transformations. After modifying $\mathcal{L}_{SM}$ to also satisfy local gauge symmetry, it will include terms that describe field interactions. Two particles are said to interact or couple if their respective fields appear in a common term in $\mathcal{L}_{SM}$.

The SM can be expressed in terms of three symmetry groups:

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \tag{1.1}$$

The $\mathrm{SU}(3)_C$ term describes the strong interaction between quarks which carry color charge C. $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ describes the unified electroweak interaction, where the L denotes a weak interaction that only acts on left-handed particles (and right-handed antiparticles) and Y represents weak hypercharge, a charge introduced in the unification of the weak and electromagnetic interactions.

Experimental measurements indicate that the weak vector bosons have a non-zero mass. However, introducing a mass term for the electroweak interaction violates the local gauge symmetry of $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. This is not an issue for the strong interaction since gluons are massless and so no mass term is necessary.

In the 1960s, three independent groups of physicists developed a mechanism that introduces mass terms in the Lagrangian without violating local gauge symmetry [10–12]. It is frequently called the Higgs mechanism but is sometimes also referred to by other names such as the Brout-Englert-Higgs mechanism or even the Englert-Brout-Higgs-Guralnik-Hagen-

Kibble mechanism to reflect the fact that multiple people worked on this theory. When applied to the unified electroweak theory, this mechanism is referred to as electroweak symmetry breaking. The strategy is to introduce a new complex scalar field in the Lagrangian with a specifically chosen potential that maintains local gauge invariance, but has multiple non-zero states that minimize the potential energy.

Spontaneous symmetry breaking is most easily illustrated for a single gauge boson that has $U(1)$ local gauge symmetry. A complex scalar field $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$ is introduced in the Lagrangian with a potential $V(\phi)$. The states that minimize the potential, denoted as ϕ_0, are called the vacuum states. The Lagrangian with the most general gauge invariant scalar potential is:

$$\mathcal{L}_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) \quad \text{where} \quad V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2 \tag{1.2}$$

$D_\mu = \partial + igB_\mu$ is a covariant derivative, where g is a coupling constant and B_μ is a $U(1)$ gauge field. Using the covariant derivative D_μ instead of the derivative ∂_μ ensures the theory is locally gauge invariant. The parameter λ is required to be positive in order to ensure $V(\phi)$ has an absolute minimum and μ^2 is required to be negative to force ϕ to be non-zero when $V(\phi)$ is minimized. Thanks to its recognizable shape, $V(\phi)$ is sometimes referred to as a Mexican hat potential (see figure 1.1 for example).

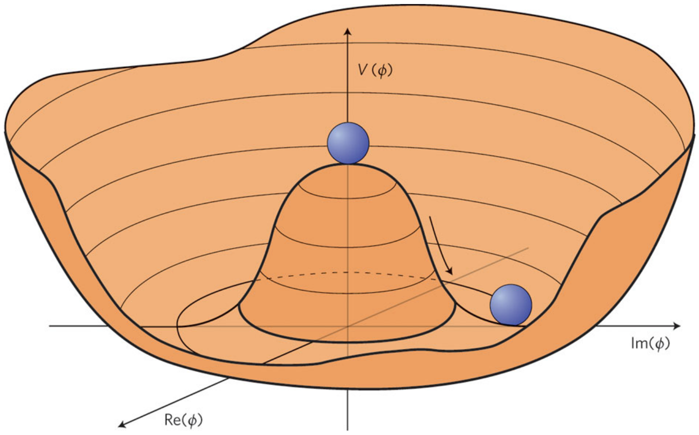


Figure 1.1: The most general scalar invariant potential is $V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2$. The parameter λ is required to be positive to ensure $V(\phi)$ has an absolute minimum and μ^2 is required to be negative to ensure ϕ is non-zero when $V(\phi)$ is minimized. Due to its shape, $V(\phi)$ is commonly referred to as the Mexican hat potential [13]. Reprinted by permission from Macmillan Publishers Ltd: Nature Physics 7(1) © 2011.

The ring of minima in $V(\phi)$ correspond to the infinite number of possible vacuum states:

$$\sqrt{\phi_1^2 + \phi_2^2} = \sqrt{-\frac{\mu^2}{\lambda}} = \nu \tag{1.3}$$

where ν is the vacuum expectation value (v.e.v.). The process of choosing a single physical vacuum state somewhere along this ring spontaneously breaks the symmetry of the Lagrangian. Choosing the vacuum state $(\phi_1, \phi_2) = (\nu, 0)$ makes it possible to write the original fields in terms of two new fields centered around the chosen vacuum state, $\phi_1(x) = \eta(x) + \nu$ and $\phi_2(x) = \xi(x)$. The newly defined η-field corresponds to oscillations away from the minima of the potential, while the ξ-field corresponds to oscillations along the minima of the potential. Interestingly, after expansion, $V(\eta, \xi)$ includes a mass term for B_μ. By demanding local gauge invariance, the initially massless gauge field has acquired mass as required. In addition, $V(\eta, \xi)$ also includes a mass term for η which can be linked to the Higgs boson mass, an experimentally measurable observable that can be used to test the validity of this theory. Finally, the massless particle corresponding to ξ is called a Goldstone boson which does not correspond to a physically known particle. However, there is a term in the rewritten Lagrangian that represents a direct coupling between the Goldstone boson and the gauge field B_μ. A transformation to the unitary gauge[2] removes the Goldstone boson from the Lagrangian and transforms the corresponding degree of freedom into the longitudinal polarization state of B_μ. All massive bosons have longitudinal polarization and so spontaneous symmetry breaking provides a mechanism by which they acquire mass.

Note that even though the Lagrangian has been expressed in terms of a single unique vacuum state and its original symmetry is not immediately evident, the potential has not changed. The symmetry in the Lagrangian is hidden but still present.

Recall the electroweak theory of the SM must explain three massive bosons (the W^+, W^-, and Z^0 bosons) which suggests that three Goldstone bosons are required to account for their longitudinal polarization states. Consequently, a complex doublet scalar field with four degrees of freedom is introduced in the Lagrangian:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.4}$$

Following a similar procedure as outlined above, the vacuum state chosen is $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_4 = v$. This convention corresponds to the unitary gauge and breaks the symmetry of $SU(2)_L \times U(1)_Y$ but not of $U(1)_{\text{em}}$. Consequently, the Lagrangian acquires mass terms for the physical weak vector bosons, but the photon remains massless as required.

A more detailed pedagogical description of spontaneous symmetry breaking in the SM and the Brout-Englert-Higgs mechanism can be found in reference [14].

---

[2] The unitary gauge is the gauge in which the complex scalar field is real.

1.3 Higgs Boson Production

An important characteristic of the Higgs boson is that it tends to couple more strongly to massive particles. The coupling strength of the Higgs boson scales as the particle mass for fermions and the particle mass squared for weak vector bosons [14]. Consequently, the most probable methods of producing a Higgs boson involve either heavy quarks or massive vector bosons. However, the colliding protons at the LHC primarily consist of gluons and light quarks. The quantum mechanical probability of a specific process occurring is quantified by the cross section.

The leading Higgs boson production mechanism at the LHC is gluon-gluon fusion (ggF) and it was the first to be discovered at the ATLAS and CMS experiments. The ggF Higgs boson is produced via a heavy quark loop originating from two high energy gluons as illustrated in the left Feynman diagram in figure 1.2. Loops in Feynman diagrams represent higher order corrections in the theoretical predictions. Since the Higgs boson does not couple directly to gluons, the loop corrections are essential to producing a Higgs boson via gluon-gluon fusion. For the vast majority of events, the quark loop is made up of top quarks but there is also a small contribution to the total ggF cross section coming from b-quark loops. In principle, there is also a contribution from lighter quarks, but this contribution becomes more and more suppressed as the quark mass decreases. Since the loop can potentially also be made up of heavy particles beyond the SM, measuring the ggF cross section provides an important probe for new physics beyond the SM.

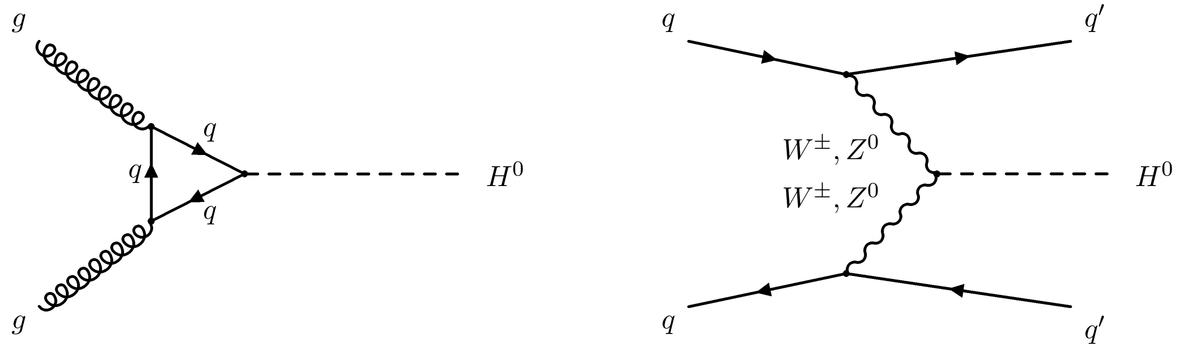


Figure 1.2: Feynman diagrams for the dominant Higgs boson production modes at the LHC, gluon-gluon fusion (left) and vector boson fusion (right).

With a cross section that is more than 10 times smaller than that of ggF, vector boson fusion (VBF) is the subleading Higgs boson production mode at the LHC. As shown in the right Feynman diagram of figure 1.2, two high energy quarks radiate weak vector bosons that fuse into the Higgs boson. The two quarks tend to leave characteristic energy depositions in the areas of the detector close to the beam pipe which makes it possible to probe for VBF despite its smaller cross section. Since electroweak interactions are better understood theoretically than strong interactions, the theoretical uncertainties on the VBF cross section are smaller than those of ggF.

Together, ggF and VBF drive the sensitivity of Higgs boson measurements at the LHC. Under the assumption that there are no corrections to the quark loop from particles not included in the SM, the ggF production mode makes it possible to probe the Higgs Yukawa coupling, the coupling of the Higgs boson to the fermions. The VBF production mode gives access to the exclusive coupling of the Higgs boson to the weak vector bosons. These measurements are compared with the SM predictions in order to search for possible hints of new physics. The analysis summarized in this thesis focuses on the VBF production mode.

Figure 1.3 shows Feynman diagrams for Higgs Strahlung (WH and ZH, collectively referred to as VH) and top-Higgs associated production (ttH), two additional Higgs production modes at the LHC. However, due to their low cross sections, these production modes are not expected to contribute significantly with the dataset size used in this thesis. Searches for both these production modes are ongoing at the LHC. One advantage of Higgs Strahlung is that it leaves a clean signature in the detector when the final state weak vector boson decays leptonically. Even though the diagram for Higgs Strahlung has similarities to that of ggF, it is suppressed because of the much larger total mass in the final state.

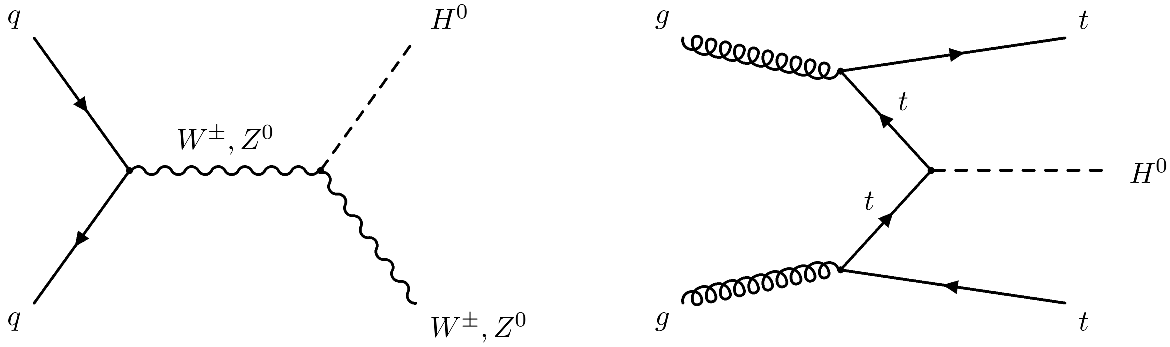


Figure 1.3: Feynman diagrams for subdominant Higgs boson production modes at the LHC, Higgs Strahlung (left) and top-Higgs associated production (right).

Figure 1.4 shows the cross sections for the Higgs boson production modes as a function of collision energy. As indicated by the text in parentheses for each process, the cross sections include corrections from higher order Feynman diagrams.

1.4 Higgs Boson Decay

Since the theoretical lifetime of the SM Higgs boson is on the order of 10^{-22} seconds, it decays before reaching the active material of the ATLAS detector. Consequently, the Higgs boson is studied by measuring the various particles it can decay into. Figure 1.5 summarizes the branching ratios of the dominant Higgs boson decay modes as a function of the Higgs boson mass. At a Higgs boson mass of 125 GeV, the dominant decay mode is $H \to b\bar{b}$.

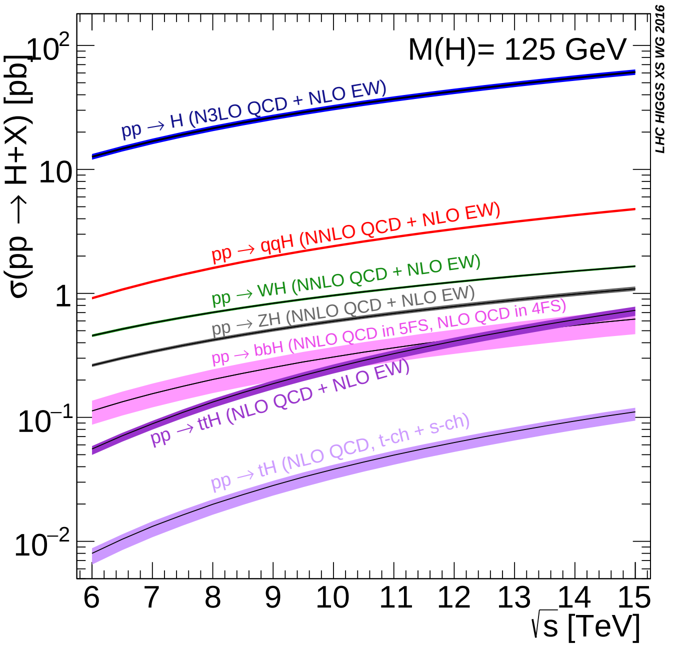


Figure 1.4: Cross sections of the Higgs boson production mechanisms at the LHC. The ggF production mode is shown in blue while VBF is shown in red [15] © 2017 CERN.

At the LHC, this decay mode is extremely challenging to measure in the ggF and VBF production modes since it is very difficult to differentiate it from the far more common processes originating from strong interactions. However, when searching for a Higgs boson produced by Higgs Strahlung, the accompanying weak vector boson in the final state can be used as an identifying feature. Similarly, in the case of ttH, the top quarks produced in association with the Higgs boson both decay to a b-quark, resulting in a minimum of four b-quarks. Since the depositions from b-quarks can be distinguished from lighter quarks, this also results in a recognizable signature, particularly when at least one of the W bosons from the top quark decays leptonically. Nevertheless, the high branching ratio for $H \rightarrow b\bar{b}$ is offset by the relatively low cross sections for VH and ttH when compared to ggF.

The ATLAS publication on the initial 2012 discovery of a new particle consistent with the SM Higgs boson was driven by the $\gamma\gamma$, ZZ and WW decay channels [1]. Despite its relatively low branching ratio, the $\gamma\gamma$ channel offers a very clean final state consisting of two photons and thanks to the energy resolution of the ATLAS calorimeter, contributes significantly to the Higgs boson mass measurement. Similarly, the ZZ channel provides a very clean signature when both of the Z bosons decay leptonically, but unfortunately this also reduces the branching ratio below the $\gamma\gamma$ branching ratio at the Higgs boson mass.

The WW branching ratio dominates for Higgs boson masses above $\sim$ 130 GeV, but drops off towards lower values since at least one of the W bosons is forced to go increasingly further off-shell. Since there are two W bosons (W^+ and W^-) but only one Z boson, the

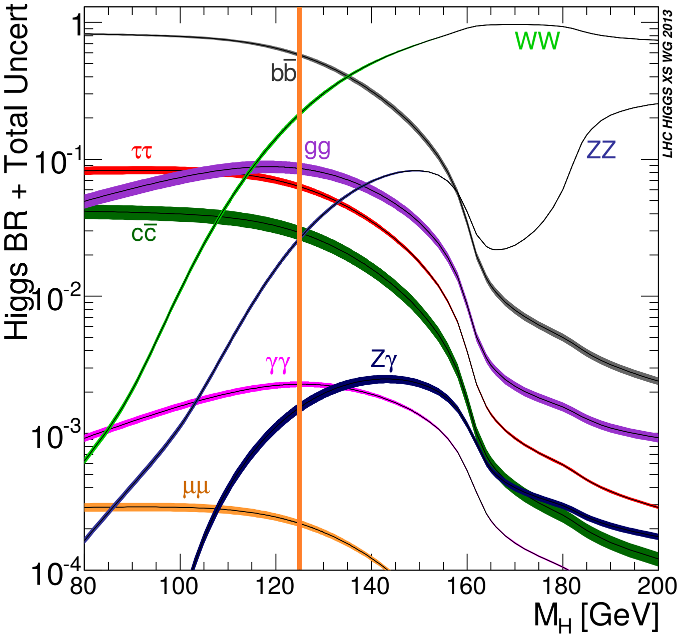


Figure 1.5: Higgs boson decay branching ratios for the dominant Higgs boson decay modes as a function of the Higgs boson mass. The vertical orange line corresponds to a Higgs boson mass of 125.1 GeV [15] © 2017 CERN.

branching ratio for WW is significantly larger than for ZZ even though the Z boson mass is larger than that of the W boson. At $m_H = 125.1$ GeV, WW is a promising channel to probe for the Higgs boson and measure its properties. One parameter of interest in a Higgs boson search is the signal strength μ which is defined as the ratio of the measured cross section to that expected by the SM.

The cleanest signature in the detector is obtained when considering the case where both W bosons decay into an electron or muon and a neutrino. Unfortunately, the branching ratio for this Higgs boson decay mode is only 1%. In addition, the two neutrinos are not measured by the ATLAS detector and so it is not possible to reconstruct the Higgs boson mass in this final state. However, this final state has a distinct kinematic feature that makes it possible to significantly reduce the backgrounds that imitate it. The SM Higgs boson is a spin-0 particle and since angular momentum is conserved, the two spin-1 W bosons it decays into must have opposite spin projections as indicated by the double arrows in figure 1.6. Similarly, when a W boson decays into two spin-$\frac{1}{2}$ leptons, they must have aligned spin projections. Due to the left-handed nature of the weak interaction, neutrinos will have spins opposite to their direction of motion, while anti-neutrinos will have spins parallel to their direction of motion. Consequently, the opening angle between the charged leptons in the Higgs boson rest frame is expected to be small for $H \to WW^*$ events.

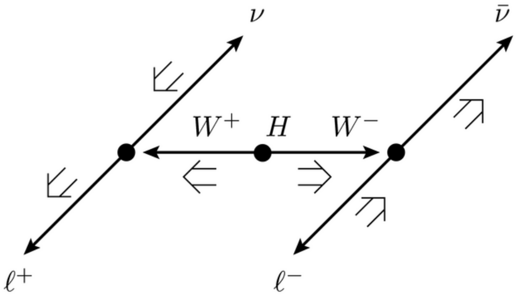


Figure 1.6: Since the Higgs boson is spin-0 and due to the left-handed nature of the weak interaction, the charged leptons coming from the W decays will preferentially travel in the same direction. The double arrows denote the particle spins, while the thin arrows indicate the directions of motion [16].

Due to the small opening angle between the charged leptons, the invariant mass of the dilepton system is also expected to be small. This can be shown by considering the mass of the dilepton system in terms of the energies (E_{ℓ_i}) and momenta (p_{ℓ_i}) of the two leptons:

$$m_{\ell\ell}^2 = (E_{\ell_1} + E_{\ell_2})^2 - (\vec{p}_{\ell_1} + \vec{p}_{\ell_2})^2 \tag{1.5}$$

$$m_{\ell\ell}^2 = (E_{\ell_1}^2 - \vec{p}_{\ell_1}^{\,2}) + (E_{\ell_2}^2 - \vec{p}_{\ell_2}^{\,2}) + 2E_{\ell_1}E_{\ell_2} - 2|\vec{p}_{\ell_1}||\vec{p}_{\ell_2}|\cos\Delta\psi \tag{1.6}$$

where $\Delta\psi$ is the opening angle between the two charged leptons. Assuming the electron masses are negligible:

$$m_{\ell\ell}^2 \approx 2E_{\ell_1}E_{\ell_2}(1 - \cos\Delta\psi) \tag{1.7}$$

Values of $\Delta\psi$ close to zero will result in small values of $m_{\ell\ell}$.

Note that the Higgs boson mass is significantly smaller than the sum of the masses of two W bosons. As a result, at least one of the W bosons is off-shell in the $H \rightarrow WW^*$ decay channel. This is indicated by the asterisk next to the symbol for the second W boson. On average, the leptons from the off-shell W boson will have lower momentum than the on-shell W boson.

The VBF $H \rightarrow WW^*$ channel is particularly interesting to investigate and is the main channel explored in this thesis [16]. Firstly, it makes it possible to probe the exclusive coupling of the weak vector bosons to the Higgs boson. This is particularly relevant when measuring κ_V, a scale factor that describes the Higgs boson coupling strength to the vector bosons with respect to the SM expectations. For VBF, μ_{WW}^{VBF} is proportional to κ_V^4, while for ggF, $\mu_{WW}^{gg\mathrm{F}}$ is only proportional to κ_V^2 [17]. A related scale factor is κ_F which describes the Higgs boson coupling strength to fermions with respect to the SM expectations.

In addition, the VBF $H \rightarrow WW^*$ channel is important at a theoretical level. When only considering the lowest order Feynman diagrams for $W^+W^- \rightarrow W^+W^-$ scattering shown in the top row of figure 1.7, the theoretical cross section includes terms originating from longitudinally polarized W bosons that violate quantum mechanical unitarity for center-of-mass energies above ~ 1 TeV. This issue is resolved by also including the VBF $H \to WW^*$ diagrams illustrated in the bottom row of figure 1.7 which interfere negatively such that the overall amplitude is finite. Consequently, it is important to establish the existence of the VBF $H \to WW^*$ process to validate the coupling and cross section predictions of the SM.

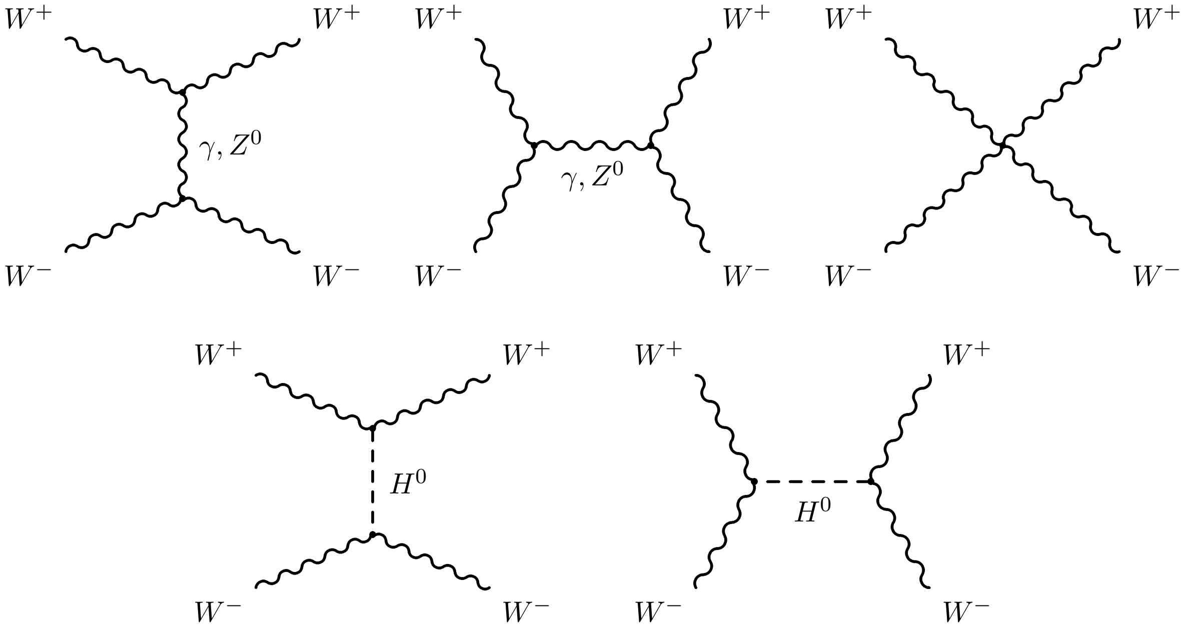


Figure 1.7: The VBF $H \to WW^*$ Feynman diagrams (bottom row) interfere negatively with the unitarity-violating $W^+W^- \to W^+W^-$ scattering diagrams (top row) such that the overall amplitude is finite.

Chapter 2

ATLAS and the LHC

The Large Hadron Collider (LHC) is a proton-proton collider located at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland. Even though the ATLAS and CMS detectors at the LHC are best known for their recent discovery of a new particle consistent with the SM Higgs boson, they are designed to also be sensitive to a wide spectrum of searches for new physics beyond the SM.

2.1 The Large Hadron Collider

With a circumference of 26.7 km and a design center-of-mass energy of 14 TeV, the LHC is the world's largest and most powerful particle accelerator. It first circulated proton beams in September of 2008.

Protons (and occasionally heavy ions) are accelerated to close to the speed of light and then collided inside the four main detectors located along the circumference of the ring. ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid) are general purpose detectors. ALICE (A Large Ion Collider Experiment) and LHC*b* (Large Hadron Collider beauty) were built for the respective objectives of studying quark-gluon plasmas and charge-parity violation in *b*-hadron decays.

As illustrated in figure 2.1, the proton beam is supplied to the main LHC ring by a sequence of accelerators that bring it to increasingly higher energies. The protons that make up the beam are obtained by stripping electrons from hydrogen gas molecules using a strong electric field. The beam is subsequently accelerated to 50 MeV in a linear accelerator called the Linac 2. Next, it is accelerated to 1.4 GeV, 25 GeV, and 450 GeV by the Proton Synchrotron Booster (PSB), Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS) respectively. Finally, the protons are supplied to the main LHC ring where they are accelerated to their collision energy.

Protons are accelerated by electromagnetic fields generated inside radio-frequency (RF) cavities. The oscillation of the electromagnetic fields inside the RF cavities also clusters the

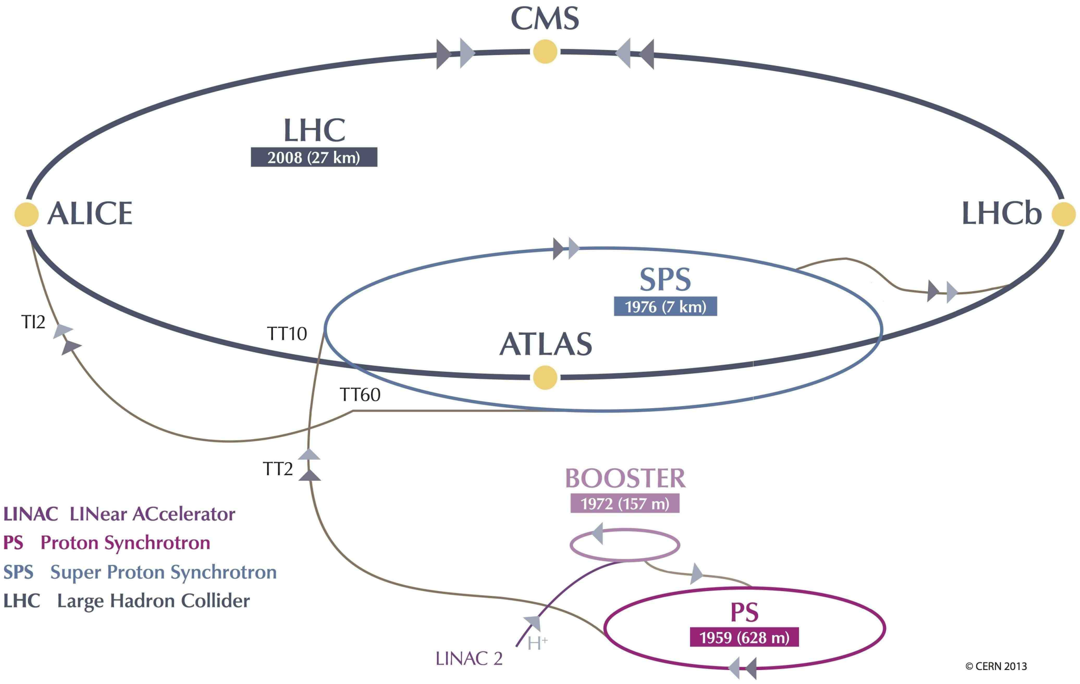


Figure 2.1: A sketch of the components of the LHC accelerator complex that supply beams to the four main experiments at the LHC [18] © 2017 CERN.

particle beam into discrete parts known as bunches. The LHC was designed for a bunch spacing of 25 ns. There are a total of 16 super-conducting RF cavities at the LHC, each cooled by a cryostat.

In the main LHC ring, proton beams travel in opposite directions through two adjacent beam pipes. The beam pipes are kept at an ultra-high vacuum on the order of 10^{-13} atmospheres in order to prevent gas molecules from interfering with the beam. The beam is guided around the LHC ring using 1,232 powerful super-conducting dipole magnets that must be kept at temperatures of only a few degrees above absolute zero to keep them in a super-conducting state. Quadrapole magnets are used to focus the beam while separate magnet systems are used to change the path of the beam such that they collide inside one of the detectors.

The total number of events (N) for some process over some period of time (t) can written as:

$$N = \sigma \int \mathcal{L}(t)dt \tag{2.1}$$

where σ is the cross section for the process in question, and $\mathcal{L}$ is the luminosity. The luminosity is fully described by the properties of the beam and for a Gaussian beam distribution, can be written as [19]:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \tag{2.2}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the frequency of revolution, γ_r is the relativistic Lorentz factor, ϵ_n is a measure of the transverse spread of the beam in both position and momentum, β^* is a measure of how focused the beam is at the interaction point, and F is a geometric correction factor to account for the crossing angle of the two beams. As the luminosity is increased, the number of collisions increases proportionally.

The design luminosity of the LHC is 10^{34} cm^{-2}s^{-1}. Figure 2.2 shows the total integrated luminosity delivered by the LHC in 2011-2012 and 2015-2016.

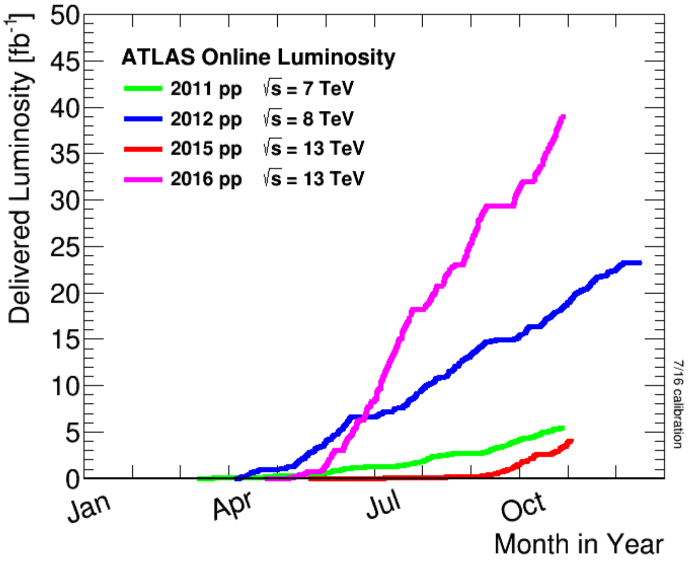


Figure 2.2: Total integrated luminosity delivered to the ATLAS experiment by the LHC from 2011 until 2016 [20] ATLAS Experiment © 2017 CERN.

In the quest for rare physics processes with low cross sections, high luminosity is one of the key performance characteristics for an accelerator. The luminosity can be increased by increasing the number of protons per bunch, squeezing the beams using more powerful quadrapole magnets, and by decreasing the space between the bunches. Technology is currently being developed for the High-Luminosity LHC (HL-LHC) that will increase the design luminosity of the LHC by a factor of 10 [21]. Unfortunately, increasing the luminosity also increases the number of uninteresting interactions that are measured per bunch crossing. The detectors at the LHC will require significant upgrades to be able to collect HL-LHC data.

Even though the LHC was designed to run at a center-of-mass energy of 14 TeV, it has not done so to date. After an electrical malfunction caused a magnet quench soon after the startup of the LHC in 2008 [22], it was decided to run below the design energy to reduce the probability of further incidents. In 2010 and 2011, the collision center-of-mass energy was 7 TeV, while in 2012 this was increased to 8 TeV. This 3-year period is collectively known as Run-I. After a shutdown for maintenance and upgrades, the LHC started Run-II of data collection in 2015 with a new record collision energy of 13 TeV. Run-II is expected to continue until 2018.

2.2 The ATLAS Detector

ATLAS is one of the two general purpose detectors at the LHC designed for precision measurements despite high interaction rates, damaging radiation, and high particle multiplicities. It is 44 meters long, 25 meters wide and weighs approximately 7,000 tons. ATLAS consists of three important sub-systems which form concentric cylinders around the collision point: the inner detector, the calorimeter and the muon spectrometer as illustrated in figure 2.3. The inner detector and muon spectrometer are tracking detectors that are used to measure the momentum of charged particles passing through, while the calorimeter is designed to measure the energies of particles. In order to improve the hermeticity of the detector, each subsystem also has an end-cap region which measures particles close to the beam pipe.

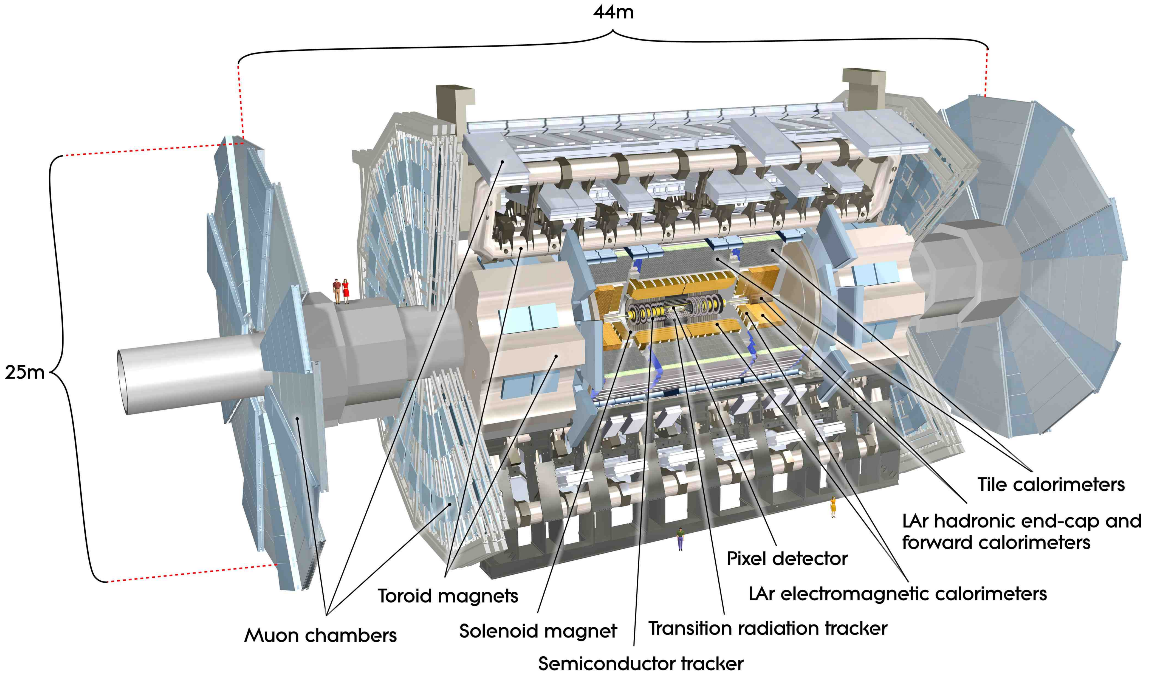


Figure 2.3: Cut-away schematic of the ATLAS detector illustrating the relative positions of the inner detector, calorimeter, muon spectrometer and the magnet system [23] ATLAS Experiment © 2017 CERN.

2.2.1 The ATLAS Coordinate System

ATLAS uses a right-handed coordinate system, with the z-axis defined parallel to the direction of the beam. The interaction point at the center of the detector is defined as the origin. The x-y plane is transverse to the beam direction with the x-axis pointing to the center of the ring and the y-axis pointing skywards. R is the radial distance in the x-y plane. The azimuthal angle ϕ is measured in the x-y plane, starting at 0 along the positive x-axis and increasing towards the positive y-axis. The polar angle θ is measured from the z-axis, ranging from 0 at the positive z-axis to π at the negative z-axis. Rapidity is commonly used instead of θ since differences in rapidity are Lorentz invariant. It is denoted by y (not to be confused with the Cartesian coordinate y) and is defined as:

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \tag{2.3}$$

A closely related variable is the pseudo-rapidity (η):

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.4}$$

In the limit where the momentum of the particle is much larger than its mass, the pseudo-rapidity converges to the definition of rapidity.

The distance between two objects in the η-ϕ plane is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.5}$$

Often quantities at the ATLAS detector are defined in the x-y plane which is also called the transverse plane. For example, the transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$ is defined as the component of the momentum p that is in the transverse plane. The transverse plane is especially important when invoking the conservation of momentum. Before a collision takes place, the initial transverse momentum is zero. However, the initial momentum along the beam axis is not expected to be zero since the interacting partons carry different fractions of the proton momentum.

2.2.2 The Inner Detector

The inner detector (ID) is a tracking detector designed to precisely measure the momenta of charged particles in the region $|\eta| < 2.5$. An important consideration in the design of the ID is that particles only deposit a minimal amount of energy in them so as not to detract from the measurements in the calorimeter. The ID identifies spatial points called hits along the path of charged particles. Algorithms are used to reconstruct hits into continuous paths known as tracks. The entire ID is surrounded by a super-conducting solenoid that generates a 2 T magnetic field. Since the magnetic field bends the paths of charged particles in the transverse plane, it is possible to deduce their momentum from the curvature of their tracks.

As illustrated in figure 2.4, the ID consists of three components: the silicon pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT).

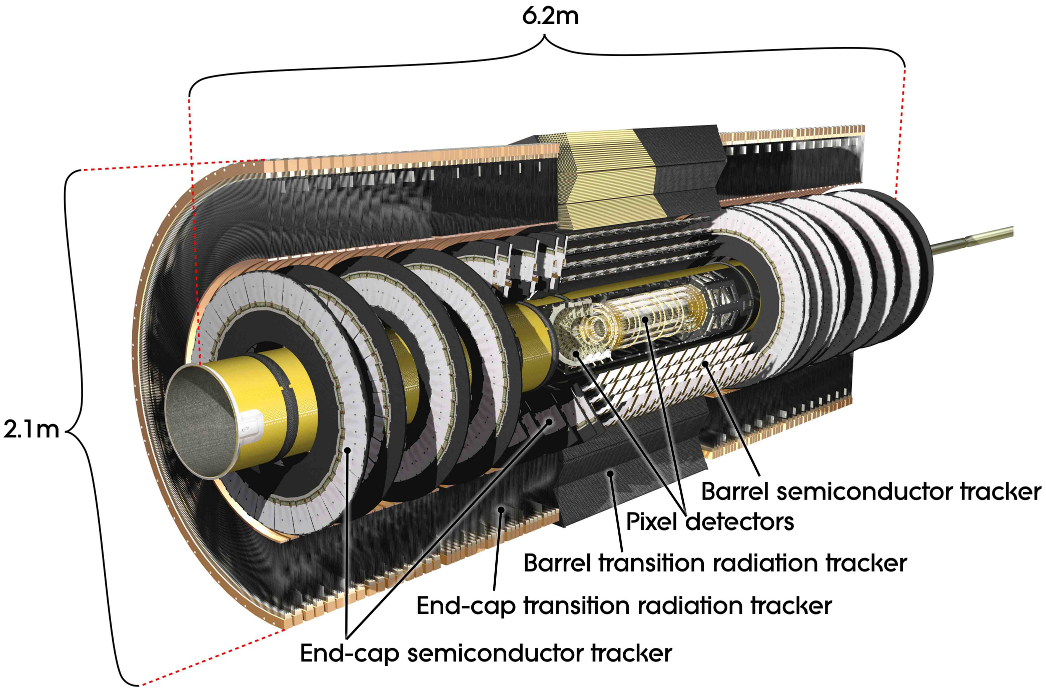


Figure 2.4: Cut-away schematic of the inner detector illustrating the relative positions of the pixel detector, the semiconductor tracker and the transition radiation tracker [23] ATLAS Experiment © 2017 CERN.

The silicon pixel detector is closest to the beam pipe and has the highest spatial resolution. The original configuration used during Run-I had three cylindrical layers of silicon sensors in the barrel, and three disk-shaped layers in both of the end-caps. When a charged particle passes through a silicon sensor, electron-hole pairs are generated and the resulting charges are collected using an electric field. If the signal coming from the collected charge exceeds a preset threshold, a hit is registered. For the barrel, the intrinsic accuracy of the silicon pixel detector is 10 μm in the $R-\phi$ plane and 115 μm in the z-direction, while for the end-caps it is 10 μm in the $R-\phi$ plane and 115 μm in the R-direction. The original pixel detector consisted of approximately 80.4 million readout channels.

During the shutdown between Run-I and Run-II, a fourth pixel layer called the Insertable B-layer (IBL) was added inside the ID [24]. In addition to regaining the performance lost due to dead modules in the original pixel detector, it provides higher pixel granularity, improves the tracking precision and allows for improved identification of secondary interaction vertices from mesons and baryons containing b-quarks. In order to make room for the IBL, a smaller radius beam pipe was also installed. The IBL adds approximately 6 million readout channels to the inner detector.

The SCT is a precision silicon strip tracker that surrounds the pixel detector. It consists of 4 cylindrical layers of sensors in the barrel and 9 disk-shaped layers in both of the end-caps. Each module consists of two sensors with strips that are setup in a small-angle stereo configuration. In the barrel, one of the strips in each module runs parallel to the z-axis while the other is offset by 40 mrad which makes it possible to determine the z-coordinate. Similarly, in the end-cap disks, one of the strips in each module radiates outward in the R-direction while the other is offset by 40 mrad which makes it possible to determine the R-coordinate. For the barrel, the intrinsic accuracy of the SCT is 17 μm in the $R-\phi$ plane and 580 μm in the z-direction, while for the end-caps it is 17 μm in the $R-\phi$ plane and 580 μm in the R-direction. The SCT consists of approximately 6.3 million readout channels.

The TRT makes up the outermost layer of the ID and extends out to $|\eta| = 2.0$. It consists of 4 mm diameter straws (drift tubes) which run parallel to the beam pipe in the barrel and radiate outward in the $R-\phi$ plane in the end-caps. An anode wire is suspended in the center of each of the straws and an electric potential is applied between the straw and the wire. The tubes are filled with a gas mixture consisting primarily of xenon. When a charged particle passes through the gas mixture, the gas is ionized and the resulting electrons are accelerated towards the anode wire by the electric field. The charge induced on the anode is then converted into an electrical signal. Unlike the silicon detectors, the TRT does not measure the z-coordinate. The intrinsic accuracy of each straw is 130 μm. The TRT consists of approximately $351,000$ readout channels. Despite having significantly fewer channels than the two silicon detectors, the TRT still makes a significant contribution to the momentum measurement thanks to a larger number of hits per track and a longer track length.

Besides momentum measurements, the TRT is also instrumental in identifying electrons. Inhomogeneous radiators are positioned between the straw tubes which cause relativistic particles to release transition radiation as they pass through [25]. The released radiation ionizes the xenon gas in the straw tubes and the resulting electrons are accelerated towards the anode wire. Since the amount of radiation released increases as particles become more relativistic, electrons and positrons will release the largest amounts of transition radiation, making it possible to distinguish them from more massive particles. The TRT is particularly useful in differentiating electrons from charged pions which can be misidentified as electrons. Charged pions are hadrons consisting of a quark-antiquark pair, where one is an up quark, and the other is a down quark.

Figure 2.5 shows the passage of a particle through the pixel detector, the semiconductor tracker and the transition radiation tracker.

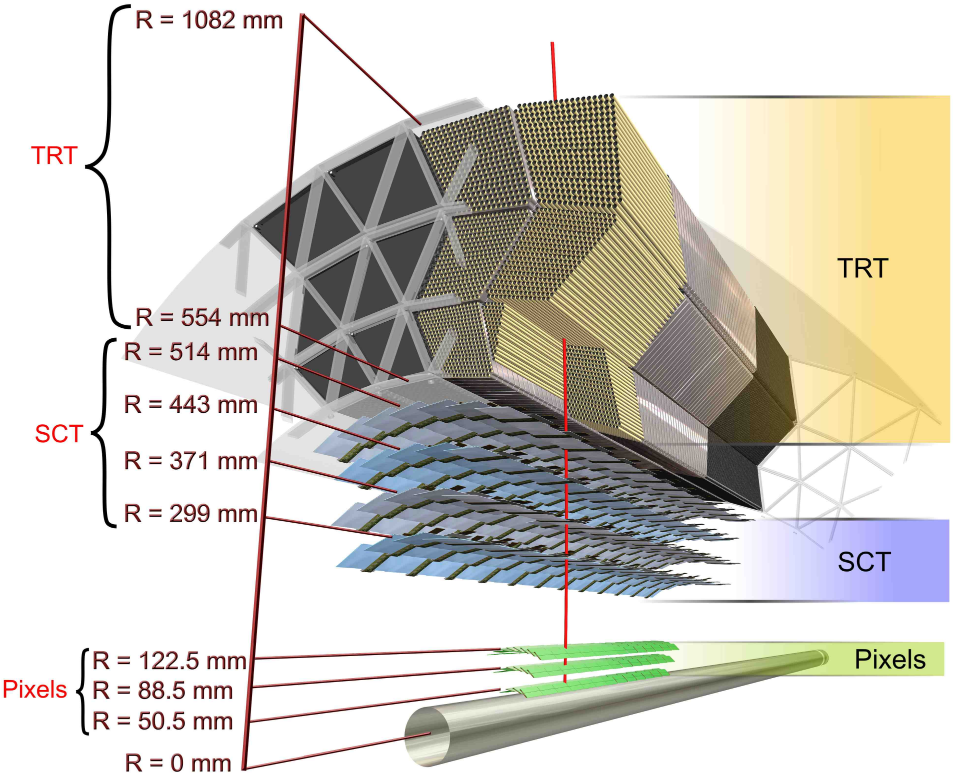


Figure 2.5: Schematic of the inner detector illustrating the passage of a particle through the barrel region of the pixel detector, the semiconductor tracker and the transition radiation tracker [23] ATLAS Experiment © 2017 CERN.

2.2.3 The Calorimeter

ATLAS has two calorimeters that destructively measure the energy of particles within $|\eta| < 4.9$. The electromagnetic calorimeter (EMC) is designed for precisely measuring the energies of electrons and photons while the hadronic calorimeter (HC) is designed to measure hadronic energy depositions. ATLAS uses sampling calorimeters, which means that they have layers of dense, shower initiating materials between the active layers that measure energy deposits. Note that these energy deposits only represent a fraction of the particle's total energy and so needs to be calibrated to recover the total energy of the incident particle.

Both a large η-range and a sufficiently deep calorimeter are important for the measurement of E_T^{miss}. The depth of the electromagnetic calorimeter is expressed in terms of the radiation length, X_0, the mean distance over which an energetic electron loses $(1 - e^{-1})$ of its energy by bremsstrahlung. The electromagnetic calorimeter is $25X_0$ thick. The depth of the hadronic calorimeter can be expressed in terms of the interaction length, λ, which is defined as the mean distance travelled by a hadronic particle before it interacts with the calorimeter via a nuclear interaction. The total thickness of the ATLAS calorimeters is approximately 10λ which sufficiently limits the number of hadronic particles penetrating the calorimeters and reaching the muon spectrometer. The two calorimeters are illustrated in figure 2.6.

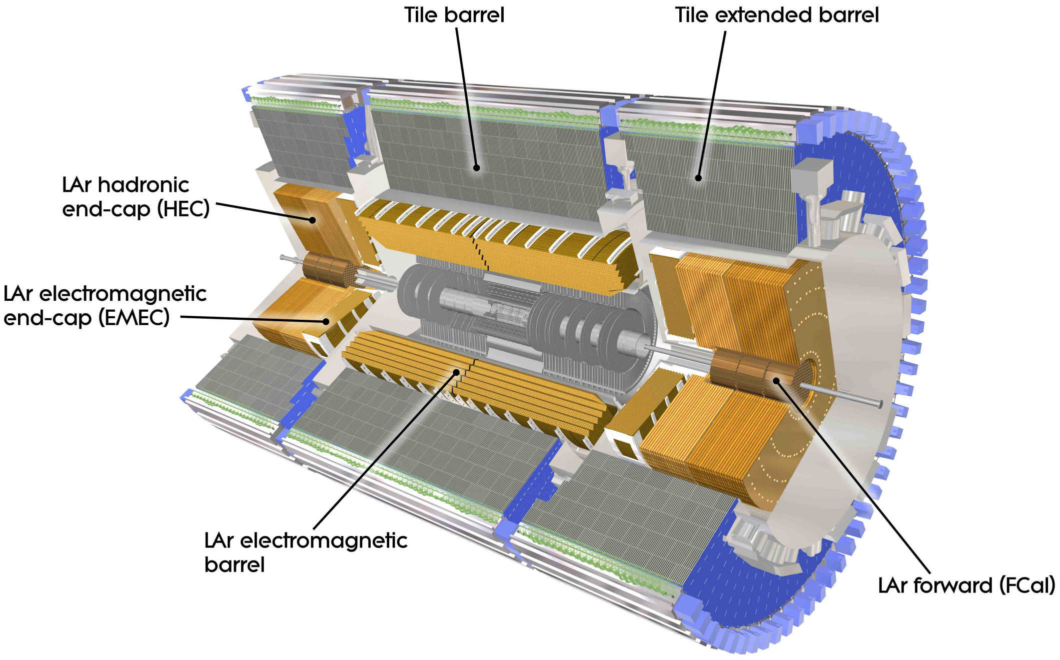


Figure 2.6: Cut-away schematic of the calorimeter illustrating the relative positions of the electromagnetic and hadronic calorimeters [23] ATLAS Experiment © 2017 CERN.

The EMC is divided into a barrel ($|\eta| < 1.475$) of three layers and two end-caps ($1.375 < |\eta| < 3.2$) of two layers. Liquid argon (LAr) is used to fill the active regions of the EMC which are separated by layers of lead absorber plates. In order to correct for energy lost in the dead material between the beam pipe and the EMC, a presampler using active LAr is installed in the region $|\eta| < 1.8$.

The dominant energy loss mechanism for high-energy electrons and positrons interacting with the absorber plates is bremsstrahlung. High-energy photons, both those entering the calorimeter and those produced by bremsstrahlung, form electron-positron pairs as they interact with the absorber plates. Consequently, incoming electrons, positrons, and photons result in tree-like showers of particles that propagate through the calorimeter until there is no longer sufficient energy to produce new particles. As the charged particles from the shower pass through the active layers, they ionize the LAr and the ionized charge is then converted into an electric signal proportional to the energy of the incident particle.

The granularity in $\Delta\eta \times \Delta\phi$ ranges from 0.025×0.025 in the central region to 0.1×0.1 in the forward regions of the end-caps. The fine granularity of the EMC is especially important for the precise measurement of electrons and photons. In order to ensure full ϕ coverage, the active and absorber layers have an accordion-like structure as illustrated in figure 2.7. The EMC has approximately 170,000 readout channels.

As evident from figure 2.6, the HC surrounds the EMC and consists of an extended barrel section, two end-caps, and two forward calorimeters. It is designed to measure showers of hadronic states originating from quarks and gluons.

The extended barrel section is a tile calorimeter that extends to $|\eta| < 1.7$. It consists of three layers which use steel as the absorber and scintillating tiles as the active material.

The hadronic calorimeter end-caps cover $1.5 < |\eta| < 3.2$ and are located directly behind the electromagnetic calorimeter end-caps. Both end-caps consist of two wheels, each consisting of two layers. LAr is used as the active material and is separated by copper absorber plates.

The forward calorimeters cover $3.1 < |\eta| < 4.9$, each consisting of three modules. The first module is optimized for electromagnetic showers and uses copper as the absorber. The other two modules are designed for hadronic showers and use tungsten as the absorber. All three modules use LAr as the active material.

The HC has approximately 19,000 readout channels.

2.2.4 The Muon Spectrometer

As shown in figure 2.8, the muon spectrometer makes up the outermost layer of the ATLAS detector. Since muons are minimum ionizing at typical ATLAS energies, they typically penetrate the calorimeters unlike most other particles. A super-conducting toroid magnet system consisting of a barrel and two end-caps is used to deflect the path of muons such that their momentum can be measured from the curvature of the resulting tracks.

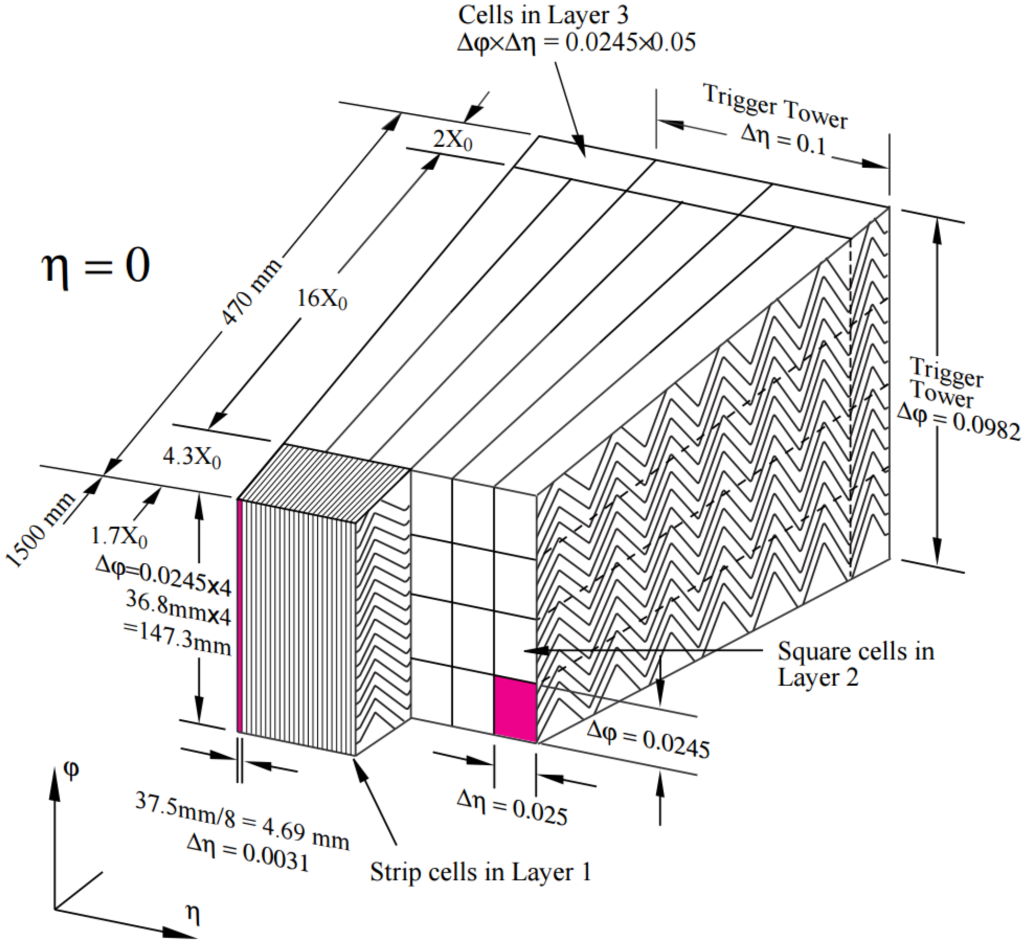


Figure 2.7: Sketch of part of the central region of the electromagnetic calorimeter illustrating its accordion-like structure [23] ATLAS Experiment © 2017 CERN.

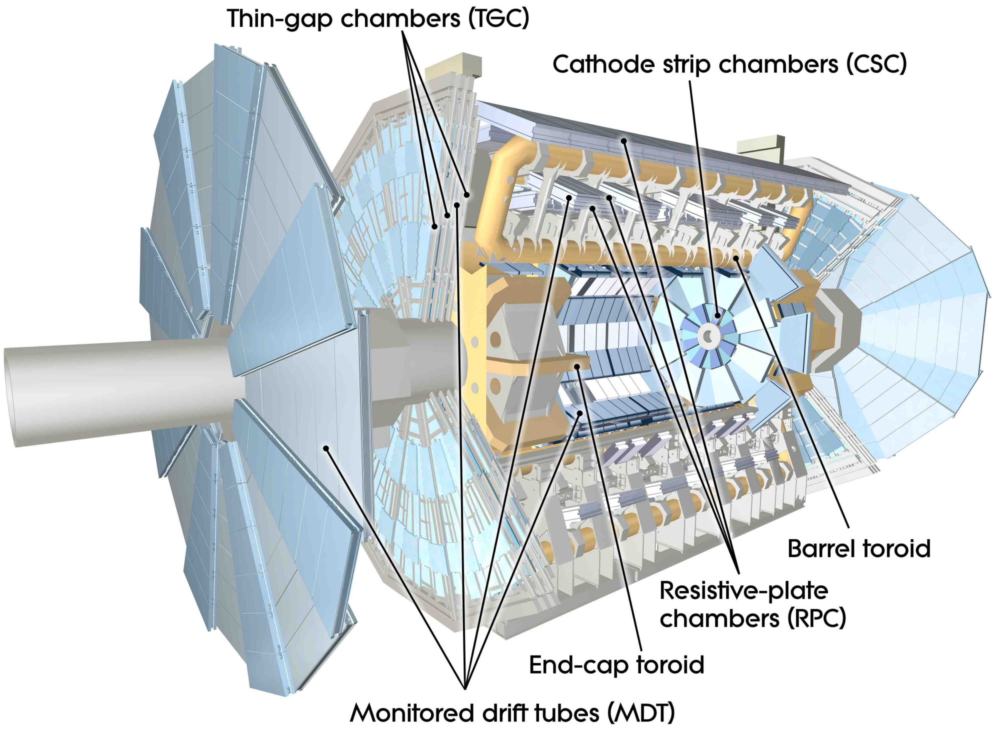


Figure 2.8: Cut-away schematic of the muon spectrometer illustrating the relative positions of the monitored drift tubes and cathode strip chambers as well as the resistive plate chambers and thin gap chambers used by the trigger system [23] ATLAS Experiment © 2017 CERN.

The barrel region and both the end-caps each have three stations of sensitive chambers. The separation of the three stations offers a lever arm of several meters for a typical muon. Monitored drift tubes (MDT) are used for precision measurements out to $|\eta| < 2.7$ for the two outermost layers and to $|\eta| < 2.0$ for the innermost layer. Each chamber consists of multiple layers of copper tubes filled with $\mathrm{Ar/CO_2}$ gas. As muons pass through the tubes, the gas is ionized. An electric field accelerates the resulting charge towards a tungsten-rhenium wire suspended in the center of the tube and the induced signal is measured.

Due to the higher particle flux, cathode strip chambers (CSC) are used for the innermost layer in the end-caps in the region $2.0 < |\eta| < 2.7$. The CSCs are multi-wire proportional chambers with anode wires and orthogonally oriented cathode strips. The cathode strips have better timing resolutions and their finer segmentation makes them more suited to cope with the higher particle rate than the MDTs. For the CSCs, the signal from the wires are not read out.

Separate trigger chambers are used for identifying tracks corresponding to muons with large transverse momentum (refer to section 2.2.5 for more details on the ATLAS trigger system). At the same time, these chambers also provide an additional coordinate measurement although with coarse spatial resolution. Triggering is particularly challenging in the forward regions of the detector since for a given momentum p, p_T is small for large values of η.

The trigger system extends out to $|\eta| < 2.4$. Resistive plate chambers (RPC) are used in the barrel ($|\eta < 1.05|$), while thin gap chambers (TGC) are used in the two end-caps ($1.05 < |\eta| < 2.4$). The barrel region consists of three layers of chambers, while the end-caps have four to account for the larger particle counts. Rather than using anode wires, RPCs consist of two parallel plates with an electric field in between them. Charged particles passing through ionize the gas between the plates and the resulting charge is collected. TGCs are multi-wire proportional chambers which make use of high electric fields and fine anode wire spacing for excellent timing resolution.

The muon system can be used as a stand-alone system, but can also be combined with the ID. Besides improving the resolution for muons at low p_T, it also reduces the probability of other particles being misidentified as muons. The muon system has approximately 1,000,000 readout channels.

2.2.5 The Trigger System

At design luminosity, there are 40 million bunch crossings inside ATLAS every second. Recording all of these interactions is not possible as it would result in tens of terabytes of data each second. The trigger system is designed to quickly identify interesting events which are characterized by measurements that are consistent with large momentum particles. In addition, the trigger stores events that have a large momentum imbalance, the signature of neutrinos escaping the detector without interacting with it. In the case of the $H \rightarrow$

$WW^* \to \ell\nu\ell\nu$ analysis, the triggers used identify events with large momentum electrons or muons.

In Run-I, ATLAS used a three-stage triggering system which selects potentially interesting events. The level-1 trigger uses a subset of information from the calorimeter and muon spectrometer to reduce the rate from $\sim 40,000,000$ Hz to $\sim 100,000$ Hz. In addition, the level-1 trigger also identifies regions-of-interest (ROI) that are passed on to the level-2 trigger. The level-2 trigger consists of a collection of custom processors that analyze the ROIs using the full granularity of the detector and reduce the rate to a few $1,000$ Hz. Finally, the level-3 trigger (sometimes also called the event filter) performs a detailed analysis for each of the events that pass the level-2 trigger using a CPU farm. The level-3 trigger reduces the event rate to several 100 Hz which are subsequently written to a data storage system. Figure 2.9 shows a schematic overview of the ATLAS trigger system used in Run-I.

In Run-II, the level-1 hardware-based trigger reduces the event rate to $\sim 80,000$ Hz. The Run-II software-based triggers are merged into a single high-level trigger (HLT) which reduces the event rate to $\sim 1,000$ Hz.

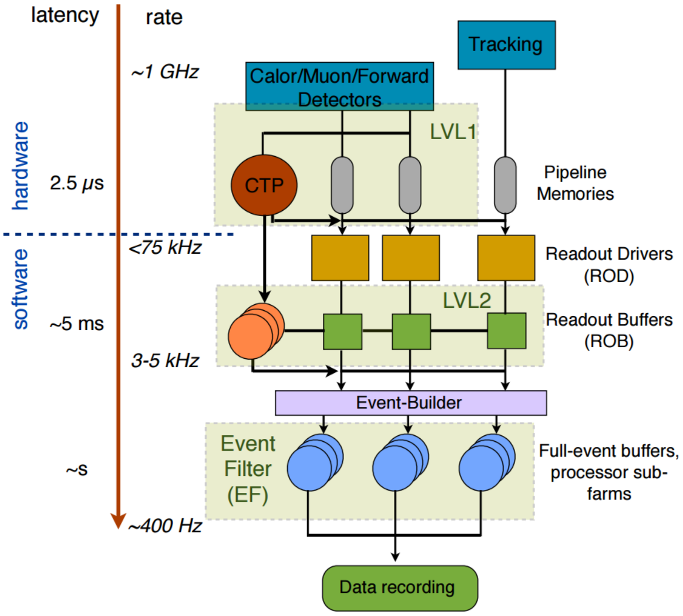


Figure 2.9: Schematic view of the ATLAS trigger system [26] ATLAS Experiment © 2017 CERN.

2.3 Hadron Collider Physics

As mentioned previously, the LHC collides two proton beams. Protons are made up of constituent partons which include three valence quarks, sea quarks created when $q\bar{q}$ pairs are spontaneously produced inside the proton and gluons that bind quarks together. Each parton carries a fraction, x, of the proton's total momentum. A parton distribution function of the proton (PDF), $f(x, Q^2)$, is the probability density that a given parton carries momentum fraction x at an energy scale of $\sqrt{Q^2}$. They are determined experimentally for a given $\sqrt{Q^2}$ from deep inelastic scattering and then extrapolated in $\sqrt{Q^2}$ using the DGLAP evolution equations [27–29].

In a hard scatter event two colliding partons interact with sufficient energy to produce new massive particles such as a Higgs boson. The protons break apart as a consequence of the hard scatter. The massive particles very quickly decay into lighter particles that interact with the ATLAS detector and produce measurable signals. As illustrated in figure 2.10, disentangling the hard scatter from the busy surrounding environment is challenging.

High energy quarks and gluons coming out of the hard scatter dissipate their energy by two mechanisms. First, both quarks and gluons radiate gluons, both before the hard scatter (initial-state radiation or ISR) and after the hard scatter (final-state radiation or FSR). Secondly, gluons with sufficient energy can create quark-antiquark pairs. Consequently, quarks and gluons coming out of the hard scatter form showers of collimated particles commonly referred to as parton showers. The parton showering is represented by the straight and curly red lines in figure 2.10.

As the parton shower progresses, the energy of the individual quarks and gluons decreases causing the coupling strength of the strong interaction to increase. As the strong coupling strength increases, hadronization becomes more and more likely. Hadronization is the process in which quarks and gluons in the parton shower form hadrons with a net color-charge of zero and is indicated by the light green ovals and dark green circles in figure 2.10. The showers of hadrons interact with the detector and provide an access point to the properties of the hard scatter. Together, the parton showering and the hadronization is called fragmentation.

As mentioned previously, photons, electrons, and positrons produce EM showers. Unlike hadronic showers, these are not constrained by color confinement.

The underlying event (UE) in hadron-hadron collisions consists of energy measurements from remnants of the beams and multiple parton interactions between remnants of the colliding protons. The UE, indicated by the letters D and E in figure 2.10, needs to be disentangled from the hard scatter.

Studying the hard scatter is further complicated by interactions originating from protons that are are not part of the hard scatter which is not shown in figure 2.10. Recall that the LHC collides proton bunches and consequently it is very likely to have multiple interactions in one bunch crossing. As illustrated in figure 2.11, pile-up presents an ever growing challenge as the luminosity of the LHC is increased over time. Pile-up that originates from protons interacting in the same bunch crossing is called in-time pile-up and increases as the number of protons per bunch is increased. Out-of-time pile-up originates from proton interactions from bunch crossings other than the one of interest. When the bunch spacing is

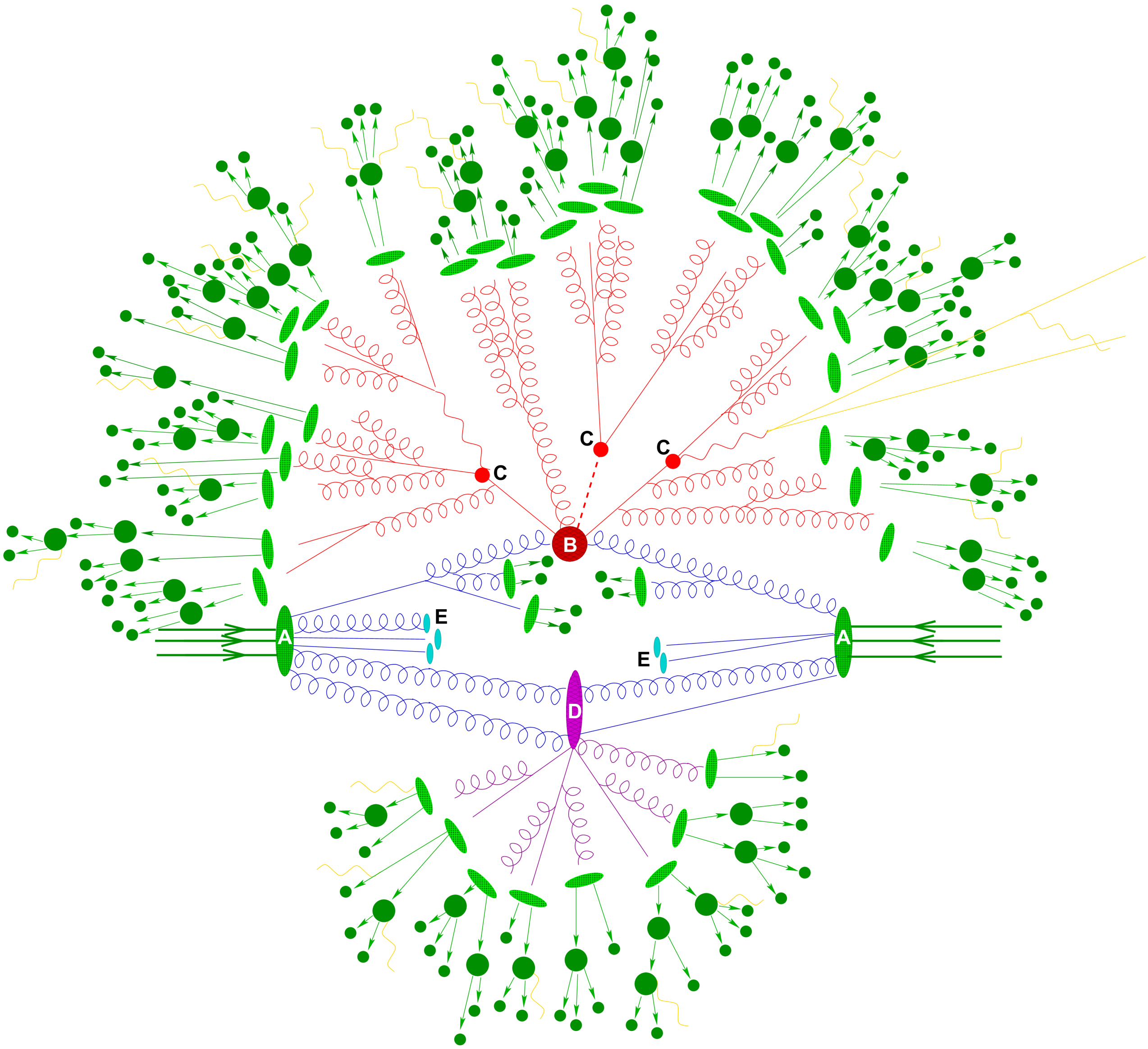


Figure 2.10: Schematic representation of a simulated hard scattering process that can occur at the LHC along with the busy surrounding environment [30]. The hard scatter (B) occurs between the constituent partons (blue lines) of two protons (A). The massive particles (C) resulting from the hard scatter will quickly decay into lighter particles. Both the incoming (blue lines) and outgoing (red lines) quarks and gluons will form showers and eventually hadronize into hadrons (light green ovals). These ultimately decay into stable hadrons (dark green circles). Other parts of the proton can also interact in secondary interactions (D) which together with proton remnants (E) makes up the underlying event. The straight yellow lines represent leptons, while photon radiation is indicated by the curvy yellow lines.

reduced, the detector is more likely to measure remnants from other bunch crossings which causes the out-of-time pile-up to increase.

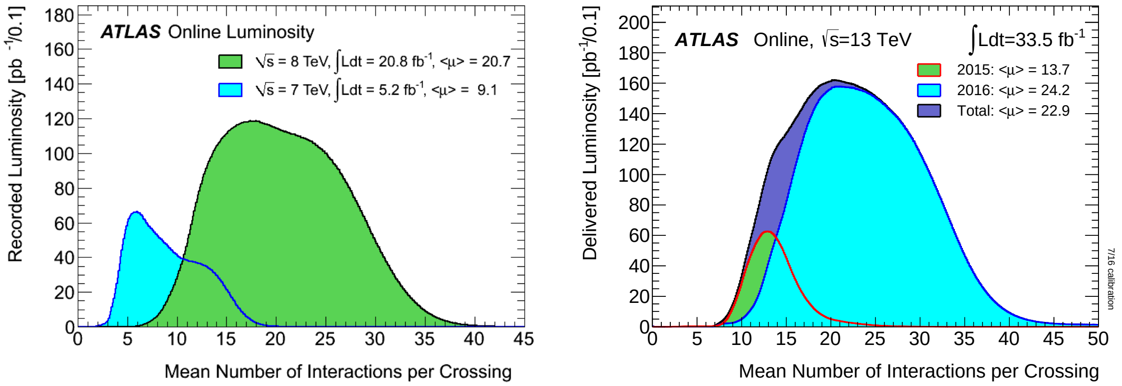


Figure 2.11: Mean number of interactions per bunch crossing in Run-I (left) and Run-II (right) [31, 32] ATLAS Experiment © 2017 CERN.

Chapter 3

Event Reconstruction

As particles originating from the proton-proton collision pass through ATLAS, they register electronic signals in the various parts of the detector. In order to be useful for analysis, the electronic signals must be reconstructed into so called *physics objects*. A physics object is an interpretation of a set of detector measurements in terms of the particles expected by the SM. For example, the electron physics object corresponds to a signature in the detector that is most likely produced by an electron. A very commonly encountered physics object at ATLAS is the jet which corresponds to a collimated shower of particles in the detector that originate from quarks or gluons. A transverse momentum imbalance in the detector is also classified as a physics object and is usually referred to as the missing transverse energy. The most important physics objects in the analysis described in this thesis are electrons, muons, jets, and missing transverse energy.

3.1 Track and Primary Vertex Reconstruction

As mentioned previously, the three components of the ID pinpoint spatial points along the path of charged particles. Reconstructing these hits into particle tracks is challenging because each event is also contaminated by numerous hits originating from pile-up.

Track reconstruction starts in the silicon detectors which are the two innermost detector subsystems. The “inside-out” algorithm [33] is used. Seeds consisting of 3-hit trajectories are identified and then extended away from the interaction point using a Kalman filter [34] to form track candidates. For a given detector layer, the hit that is most consistent with the track built from the previous layers is added to the track candidate. The quality of a track candidate is assessed based on the residuals in a least square fit of the track to the hits, the number of layers for which the track does not have a hit registered, and the number of hits in the track that are shared by other tracks. The tracks from the silicon detectors that pass the quality criteria are extended into the TRT and the track is refitted using information from the entire ID. The inside-out algorithm is designed to reconstruct tracks from primary charged particles. Primary particles are defined as particles that are produced directly from the proton-proton collision or from subsequent interactions of particles with a lifetime shorter than 3×10^{-11} s. Tracks reconstructed using the inside-out algorithm are required to have $p_T > 400$ MeV.

The second stage of track reconstruction is called *back-tracking.* It starts in the TRT with unused track segments and extends them inward through the silicon detectors to build "outside-in" tracks. Since the TRT does not have information on the z-coordinate, the algorithm starts in the R-ϕ plane. The curvature of the resulting track is extrapolated and matched to hits in the SCT. Back-tracking is developed to reconstruct tracks originating from secondary particles which are particles produced by interactions of primary particles. For example, photon conversions are photons that convert into electron-positron pairs. If a photon conversion occurs inside the silicon detectors, the detector layers traversed before the conversion will not register hits, while those after the conversion will register hits.

Next, the collection of reconstructed tracks is used to identity the interaction vertices [35]. A vertex seed is chosen by calculating the global maximum in the density of the z-coordinates along the beam of all tracks that meet track quality criteria. An iterative vertex finding algorithm uses χ^2 minimization based fitting to gauge how compatible tracks are with a given vertex seed. Tracks that are incompatible with a given vertex by more than 7σ are used to create new vertices. The ability to identify individual vertices is important in the reduction of in-time pile-up. Also, as pile-up conditions worsen, the number of vertices becomes larger which in turn increases the probability that adjacent vertices are merged into a single vertex.

Primary vertices are associated with tracks that are consistent with charged particles that originate close to the hard scatter, while secondary vertices are associated with tracks resulting from in-flight decays of unstable particles. Displaced secondary vertices play an important role in the identification of b-hadrons which are longer lived compared to hadrons made up of lighter quarks. Since hard scatter events are characterized by final state particles with large transverse momenta, the primary vertex is chosen as the vertex with the largest p_T^2 sum over all associated tracks. In the Run-I $H \to WW^* \to \ell\nu\ell\nu$ analysis, the primary vertex is required to have at least three associated tracks, each with $p_T > 400$ MeV, while in Run-II this requirement is relaxed to at least two associated tracks. The primary vertex is used as a reference point to reconstruct electrons, muons, jets and missing transverse energy, the main physics objects in the $H \to WW^* \to \ell\nu\ell\nu$ analysis.

A track can be described using 5 parameters called the perigee parameters: d_0, z_0, θ, ϕ, and q/p. The parameters are defined with respect to a coordinate system centered at the primary vertex or with respect to the beam spot, the average position of proton-proton interactions. The perigee P is defined as point along the trajectory of the track that comes closest to the z-axis. The transverse impact parameter d_0 is the distance between P and the origin in the transverse plane, while the longitudinal impact parameter z_0 is the z-

coordinate of P. Requirements on d_0 and z_0 with respect to the beam spot are used to distinguish primary vertices from secondary vertices. The polar angle θ is measured from the beam axis to the track in the R-z plane, and the azimuthal angle ϕ is measured from the x-axis to the track in the x-y plane.

3.2 Clustering Algorithms

The calorimeters are designed to measure energy depositions coming from electromagnetic and hadronic showers. In order to reconstruct these into physics objects, they are first grouped into three-dimensional clusters using clustering algorithms, a process developed to separate statistically significant signals from random noise in the detector. ATLAS uses *sliding window clustering* and *topological clustering* to define clusters that are used to reconstruct electrons/photons and jets respectively.

3.2.1 Sliding Window Clustering

In the sliding window clustering algorithm [36], the calorimeter is split into a grid in η-ϕ, where the dimensions of each grid element are $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. The cell energies for each grid element are summed resulting in *towers*. A rectangular window is defined in η-ϕ to perform a brute-force scan over the collection of towers. The size of the window is adjusted for different particle types. Electrons require a larger range in ϕ than photons since their trajectory is bent by the ID's magnet system.

For each point along the scan, the total transverse energy of all the towers in the window is calculated. For a given window, the weighted vector sum of tower energies with transverse energies exceeding a preset threshold are used to define the coordinates of seeds in η-ϕ. If two seeds are separated by less than three grid units in either η or ϕ, the one with smaller transverse energy is removed.

Finally, clusters are defined as the collection of calorimeter cells within the rectangular window centered at the identified seeds. This algorithm is particularly well suited for the fine granularity of the EM Calorimeter.

3.2.2 Topological Clustering

Topological clustering [36] is an iterative procedure that allows clusters to grow and consequently results in clusters of varying size. First, energy depositions with a signal to noise ratio of at least 4 are used to define seeds. The seeds are grown by iteratively adding neighboring cells that have a signal to noise ratio of at least 2. Finally, the cells adjacent to the neighboring cells are also included (equivalent to requiring a signal to noise ratio of at least 0). The clusters are referred to as "420" clusters reflecting the required signal to noise ratios. This algorithm is used in the calorimeters for efficiently finding low energy clusters above the detector noise. If a topological cluster has more than one maximum, it is subdivided.

The energy of a topological cluster is calculated from the energy sum of the constituent energy cells. Corrections are applied using the local cell signal weighting calibration method [37] that corrects for the signal losses from noise threshold effects and energy lost in the non-instrumented regions of the detector. Since the detector response is different for electromagnetic and hadronic showers, clusters are first categorized as electromagnetic or hadronic based on the measured energy density and the shower depth.

3.3 Physics Object Reconstruction

The collection of tracks and clusters defined in section 3.1 are used to reconstruct physics objects that can be interpreted in terms of SM particles.

3.3.1 Electron Reconstruction

Electrons are reconstructed by matching clusters identified by the sliding window algorithm to tracks in the inner detector in the pseudorapidity range of $|\eta| < 1.37$ and $1.52 < |\eta| < 2.47$. The gap $1.37 < |\eta| < 1.52$ corresponds to the transition region between the barrel and end-cap of the electromagnetic calorimeter and consequently is excluded. Even though electrons are expected to deposit most of their energy in the EM calorimeter, they can also do so in the hadronic calorimeter.

Track definitions used in electron reconstruction are modified from the description in section 3.1 to account for their tendency to emit radiation by bremsstrahlung as they pass through the detector. First, a track is fitted under the assumption that the particle interacting with the detector material is a charged pion. If the resulting track does not overlap within $\Delta R < 0.3$ with the center of an inner-layer EM cluster matching a set of quality requirements, the track is refitted under the assumption that it originates from an electron and so is allowed to lose a fraction of its energy via bremsstrahlung.

Tracks are extended into the EM calorimeter and are required to satisfy quality criteria on the match between the track and the energy clusters. For all tracks with hits in the silicon detectors, the track fitting procedure is repeated with the Gaussian Sum Filter algorithm [38], a variation of the Kalman filter that takes into account the non-linear effects of bremsstrahlung. The agreement between the fitted track and the EM clusters in $\Delta\phi$ and $\Delta\eta$ is checked again, this time with more stringent requirements.

The total energy is derived from the cluster energies from individual layers of the detector inside a window of $\eta \times \phi = 3 \times 7$ for the barrel, and $\eta \times \phi = 5 \times 5$ for the end-caps. A multivariate estimator is used to calculate the total cluster energy corresponding to an electron based on the total measured energy, the fraction of energy in the presampler, the depth of the shower and the orientation of the shower. Unless a track does not have hits in the silicon detectors, the η and ϕ of an electron are determined from the track.

Despite the high efficiency of reconstructing electrons, identified electron candidates can also include charged hadrons or electrons originating from hadron decays. In order to reduce this contamination, a likelihood-based method is used to identify and categorize electron candidates [39, 40]. The characteristics of EM showers and electron tracks are used to try determine if an electron produced the observed combination of shower and track.

Different categories of electrons ranging from “tight” to “loose” are defined based on different selection requirements on the likelihood function. The “loose” category uses variables that are effective in removing jets that originate from light quarks. More stringent identification categories also include additional variables in the likelihood function that help reject photon conversions and jets originating from heavy flavour decays. Even though using a more stringent category improves the fake rejection in the electron candidate set, it also reduces the efficiency. In Run-I, the “very tight” category is used for electron candidates with $10 < E_T < 25$ GeV, while in Run-II the “tight” classification is used for $15 < E_T < 25$ GeV. The “medium” category is used for $E_T > 25$ GeV in both Run-I and Run-II. The more stringent requirements at lower E_T reduce the contamination from photon conversions and the misidentification of light jets as electrons.

Photon conversions are also reduced by requiring a hit in the innermost layer of the pixel detector. Note that in Run-II, the innermost layer of the pixel detector is the IBL.

3.3.2 Muon Reconstruction

The $H \to WW^* \to \ell\nu\ell\nu$ analysis makes use of “combined muons” which are muons that are reconstructed by matching tracks in the inner detector and the muon spectrometer with consistent trajectories [41, 42]. Even though the MS extends to $|\eta| < 2.7$, muons are required to have $|\eta| < 2.5$ to be inside the sensitive area of the ID.

Muon candidate tracks are reconstructed in the ID using the methods described in section 3.1. In addition, ID tracks are required to have a minimum number of hits in each of the three ID components. In the MS, track segments are first identified in individual layers and then merged by fitting tracks. Tracks segments in the ID and MS are matched if they agree within a range of η and ϕ. Since this can result in multiple matches, the pair with smallest χ^2 in the fitted track is selected. The matched pair is removed, and the procedure is iterated until there are no pairs remaining.

3.3.3 Further Lepton Requirements

Both electrons and muons are required to be consistent with originating from the primary vertex by imposing requirements on their transverse and longitudinal impact parameters, d_0 and z_0 respectively. This is implemented by placing upper bounds on $|d_0/\sigma_{d_0}|$ and $|z_0 \sin\theta|$, where σ_{d_0} is the estimated uncertainty on d_0, and θ is the polar angle of track. The factor of $\sin\theta$ in the longitudinal impact parameter requirement accounts for the larger uncertainties on the position of vertices associated with tracks at large polar angles.

Furthermore, the leptons are required to be isolated in both the tracker and the calorimeter. This reduces the contamination from leptons originating from heavy flavour quark decays as well as from other particles being misidentified as electrons. A region of radius ΔR is constructed in η-ϕ space around each lepton candidate. In the tracker, the scalar p_T sum of all tracks meeting predefined track quality requirements inside the region, but not including the lepton candidate, is calculated. If the ratio of this scalar p_T sum to lepton p_T exceeds a predefined threshold, the lepton candidate is not considered isolated and so is discarded. In the calorimeter, a similar approach is used, except that the track p_Ts are replaced by the transverse energies of clusters. The choice of ratio threshold varies as a function of the lepton p_T from around 0.05 to 0.30 while ΔR is typically chosen around 0.4. Corrections are applied to the transverse energy sum to account for contributions from pile-up and the underlying event.

3.3.4 Jet Reconstruction

The anti-k_T algorithm [43] is used to group the calibrated topological clusters into jets. It is an iterative procedure that starts with calculating the distances between all pairs of clusters (d_{ij}) and the distances between the beam and each cluster (d_{iB}). The distance between clusters i and j is defined as:

$$d_{ij} = \min\left(\frac{1}{k_{Ti}^2}, \frac{1}{k_{Tj}^2}\right)\frac{\Delta_{ij}^2}{R^2} \quad \text{where} \quad \Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{3.1}$$

and the distance between the beam and cluster i is defined as:

$$d_{iB} = \frac{1}{k_{Ti}^2} \tag{3.2}$$

where k_{Ti}, y_i, and ϕ_i are the transverse momentum, rapidity, and azimuthal angle for the ith cluster. The distance parameter R is chosen to be 0.4 for the $H \to WW^* \to \ell\nu\ell\nu$ analysis.

Next, the minimum value of all d_{ij} and d_{iB} is identified. If the minimum value is one of the d_{ij}, clusters i and j are combined and their four-momenta is calculated as the sum of the constituent clusters. The values for d_{ij} are then recalculated, and the process is repeated. If the minimum value is one of the d_{iB}, then that cluster (or collection of clusters) is removed and categorized as a jet. This procedure is iterated until there are no clusters remaining. Note that d_{ij} is smaller for larger transverse momentums and so the algorithm is more likely to group low p_T clusters with high p_T clusters than make jets of low p_T clusters.

The anti-k_T algorithm is widely used within ATLAS because of its favourable properties. It is infrared safe which means that the set of jets identified by the algorithm is not changed if additional soft gluon radiation is added in the vicinity of the jets, an effect that is present in QCD processes. Furthermore, it is collinear safe which implies that it is not impacted by collinear splittings of topological clusters, an effect that is expected in the decay of energetic hadrons.

Next, corrections are applied to the reconstructed jets that take the effects of in-time and out-of-time pile-up into account [44]. Events with pile-up jets are reduced using the jet vertex fraction (JVF) [45] in Run-I and the jet vertex tagger (JVT) [44] in Run-II. Both techniques require a minimum fraction of the summed scalar p_T of tracks close to a jet in η-ϕ space to come from tracks that are associated with the primary vertex. Finally, since only a fraction of the total hadron energy is registered by the ATLAS calorimeter, corrections factors binned in p_T and η are applied to calibrate the jet energy to the hadronic energy scale [37].

In the $H \to WW^* \to \ell\nu\ell\nu$ analysis, events are split into several categories based on their jet multiplicity. For this purpose, jets are required to have $p_T > 25$ GeV for $|\eta| < 2.4$, and $p_T > 30$ GeV for $2.4 \leq |\eta| < 4.5$. The higher p_T threshold in the forward region is chosen to suppress pile-up jets.

Jets originating from b-quarks (referred to as b-jets) are expected to have distinguishable characteristics when compared to jets originating from lighter quarks. The b-hadrons (and to a lesser extent, c-hadrons) that form when b-quarks (and c-quarks) hadronize tend to have longer lifetimes than hadrons of lighter quarks and consequently typically result in displaced secondary vertices in the event. Multivariate algorithms are used to quantify the probability that a given jet originates from a heavy flavour quark using quantities based on the presence of secondary vertices and shower shapes. In the Run-I analysis, the MV1 algorithm [46] is used, while in Run-II the MV2C10 algorithm [47] is used. In both cases, the 85% working point is used, which means that the efficiency of correctly identifying true b-jets is 85%. The working point can be increased, but doing so will increase the rate at which light jets are mistagged as b-jets. The chosen working point corresponds to a mistag rate of $\sim 10\%$. The identification of heavy flavour jets is limited to jets in the acceptance of the ID ($|\eta| < 2.5$) with $p_T > 20$ GeV.

3.3.5 Missing Transverse Momentum

Before a collision takes place, the total transverse momentum is expected to be approximately zero. Consequently, by the conservation of momentum, the final transverse momentum is also expected to be zero. However, recall that neutrinos can not be detected directly with the ATLAS detector resulting in a measured momentum imbalance in the transverse plane. Calculating the missing transverse energy makes it possible to infer the momentum of the neutrino system in the transverse plane.

The missing transverse momentum is defined as the negative vectorial sum of the transverse momenta of physics objects that have been identified by ATLAS's algorithms (such as leptons, photons, and jets) and of the remaining low p_T soft objects [48–50].

$$\begin{array}{c}\text{Missing}\\ \text{Transverse}\\ \text{Momentum}\end{array} = -\left(\sum_{\substack{\text{identified}\\ \text{objects}}} \vec{p}_T + \sum_{\substack{\text{soft}\\ \text{objects}}} \vec{p}_T \right) \tag{3.3}$$

There are multiple effective ways of defining these two terms. $\vec{E}_T^{miss}$ is a calorimeter based version of the missing transverse momentum that is used in the Run-I VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis. It relies on the calorimeter's large rapidity coverage and its sensitivity to neutral particles. The identified objects included in $\vec{E}_T^{miss}$ are the calibrated leptons that pass the analysis selection as well as calibrated photons and jets with $E_T > 20$ GeV. The remaining soft objects are added to the vectorial sum using the calibrated energy clusters from the calorimeter.

Unfortunately, the resolution of $\vec{E}_T^{miss}$ deteriorates as the pile-up increases. This can be circumvented by measuring the soft objects using the tracker. In this case, the soft term is calculated as the vectorial sum of all tracks originating from the primary vertex with $p_T > 500$ MeV but not belonging to any of the identified objects. This definition of the missing transverse momentum is denoted as $\vec{p}_T^{miss}$.

The magnitudes of $\vec{E}_T^{miss}$ and $\vec{p}_T^{miss}$ are denoted as E_T^{miss} and p_T^{miss} respectively. As further explained in chapter 6, both the Run-I and Run-II analyses rely on p_T^{miss}, but only the Run-I analysis uses E_T^{miss}. In the VBF $H \to WW^*$ analysis, all kinematic quantities defined in terms of the missing transverse energy are calculated using p_T^{miss}.

Chapter 4

Analysis Strategy

As described in chapter 1, VBF $H \to WW^* \to \ell\nu\ell\nu$ events leave a characteristic signature in the detector. Nevertheless, this process is challenging to measure because of its low cross section and because of the multiple backgrounds with much higher cross sections that mimic it. This chapter starts with a brief introduction to the $H \to WW^* \to \ell\nu\ell\nu$ analysis in order to provide the context for the VBF analysis. Next, in order to motivate an analysis strategy, the features of the signal and dominant backgrounds are explored. Finally, an overview is given of the multivariate analysis technique used.

4.1 Analysis Overview

As illustrated in the schematic overview in figure 4.1, the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis is part of a larger $H \to WW^* \to \ell\nu\ell\nu$ analysis that searches for the ggF, VBF, and VH Higgs boson production modes in final states with 0 jets, 1 jet, and 2 or more jets. The ggF, VBF, and VH production modes have been thoroughly analyzed using the Run-I data [16, 51]. The first Run-II result for the VBF and WH production modes based on 5.8 fb^{-1} of data has recently been made public [52].

The focus of this thesis is on the VBF production mode which is expected to primarily reside in the category with 2 or more jets in the final state. The jets are expected to originate from the quarks that radiate the weak vector bosons which fuse to form the Higgs boson. Furthermore, the VBF analysis is divided into two categories, one where the two charged leptons in the final state have different flavour (DF: $e\mu/\mu e$) and another where they have the same flavour (SF: $ee/\mu\mu$). Both categories are included in the Run-I result, but so far only the different flavour analysis has been published in the most recent Run-II result.

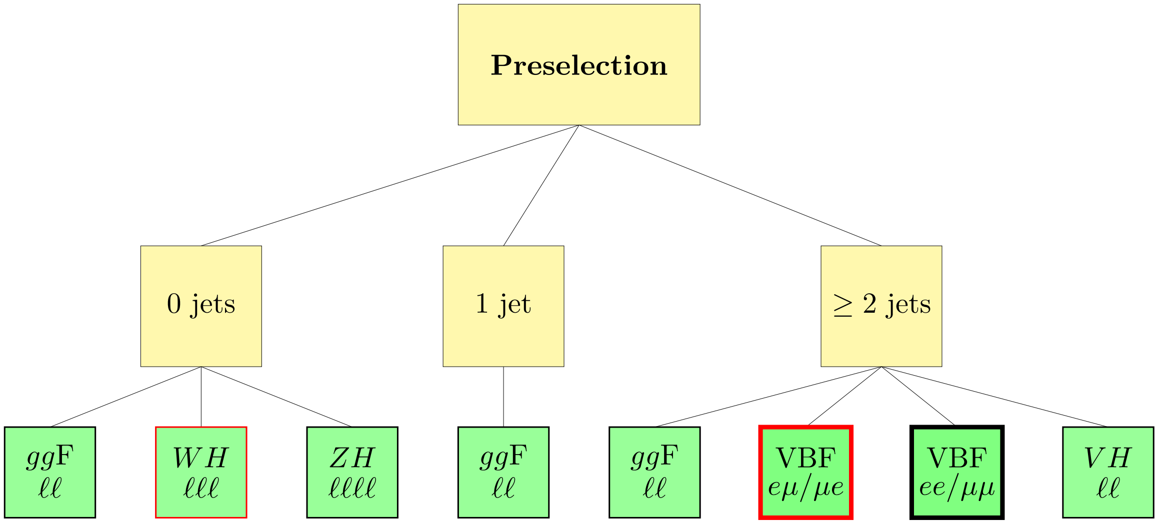


Figure 4.1: Schematic overview of analyses within the $H \to WW^* \to \ell\nu\ell\nu$ analysis group. The categories with red borders include data from Run-I and Run-II, while the other categories only include data from Run-I. The categories described in this thesis are marked with a bold border. The lepton final state is indicated below each production mode, where ℓ represents either an electron or a muon.

4.2 Signal and Background Features

Figure 4.2 (left) shows a Feynman diagram for a VBF $H \to WW^*$ event. As described in sections 1.3 and 1.4, they are characterized by two forward jets, two charged leptons with a small opening angle caused by the $V - A$ structure of the weak interaction, and missing transverse energy coming from the neutrinos. In addition, the mass of the WW system coming from the Higgs boson decay is bounded by the Higgs boson mass. These properties are used to define a region that is enriched in signal.

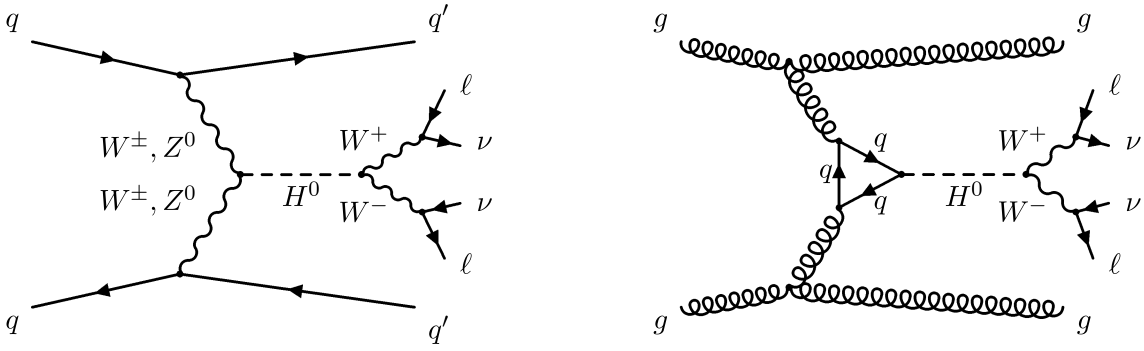


Figure 4.2: Representative Feynman diagrams for VBF $H \to WW^* \to \ell\nu\ell\nu$ (left) and ggF $H \to WW^* \to \ell\nu\ell\nu$ (right) as expected to contribute in the VBF analysis.

In the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis, ggF Higgs boson production (figure 4.2, right) is treated as a background. As with VBF, the opening angle between the charged leptons is expected to be small. At leading order, ggF events do not have any jets in the final state. The jets in the VBF signal region are most likely to come from initial state radiation which reduces the expected rate in the 2-jet category with respect to the 0- and 1-jet categories. They are expected to be less forward and have lower momenta than those from VBF events. In contrast to Higgs boson processes, the remaining backgrounds do not have a sharp upper bound on the invariant mass of the diboson system or small expected opening angles between the charged leptons.

One of the dominant backgrounds is the reducible $t\bar{t}$ background (figure 4.3, left). Since $t\bar{t}$ events are expected to have two jets originating from b-quarks, a large fraction of this background can be rejected using algorithms that differentiate jets from b-quarks from those originating from light-flavour quarks. Another top related background is single top quark production which is a relatively small contribution. The Feynman diagram in figure 4.3 (right) shows tW which is an irreducible background. There are also reducible contributions from s-channel ($t\bar{b}$) and t-channel ($tq\bar{b}$) single top production where one of the jets in the event is misidentified as the second charged lepton.

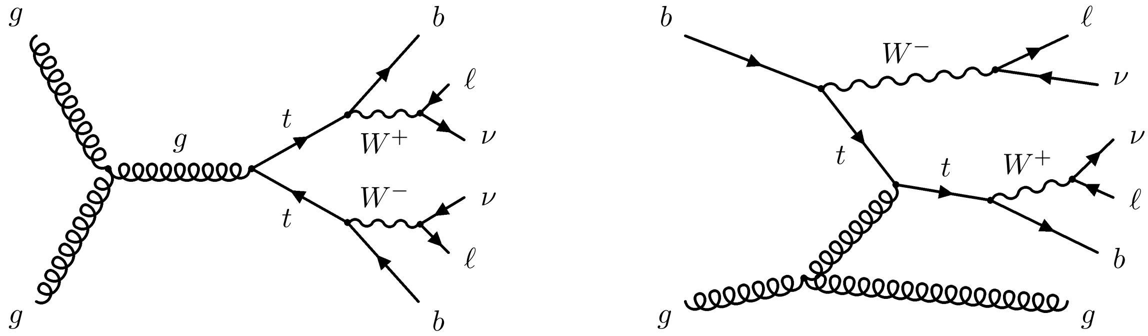


Figure 4.3: Representative Feynman diagrams for $t\bar{t}$ (left) and single top quark production (right) as expected to enter as background processes in the VBF analysis.

The diboson background, VV, consists of processes with two leptonically decaying vector bosons. The largest contribution comes from the irreducible WW background (figure 4.4, left). Since it is not expected to have any jets at leading order, WW is significantly reduced once two jets are required. The diboson background also includes smaller contributions from $W\gamma$, $W\gamma^*$, WZ, and ZZ, but these are suppressed by requiring exactly two leptons. A related background is single W production accompanied by a radiated quark or gluon that is misidentified as a lepton. This reducible background is referred to as the W+jets background (figure 4.4, right). The misidentification of jets as leptons can be extended a step further if the remaining W boson in the W+jets diagram in figure 4.4 is replaced by an additional radiated quark or gluon which is also misidentified as a lepton. If two jets are misidentified as leptons, the background is called the multi-jet or QCD background.

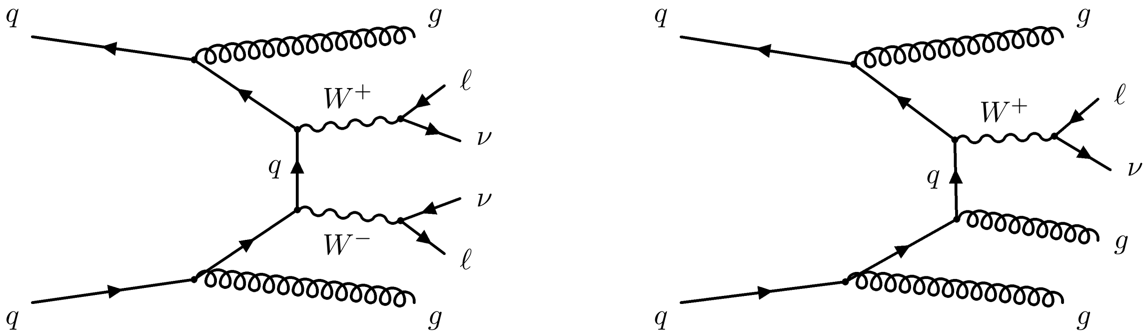


Figure 4.4: Representative Feynman diagrams for WW (left) and W+jets (right) as expected to enter as background processes in the VBF analysis.

Finally, there is a significant background coming from the Drell-Yan process (DY). It occurs when a quark and anti-quark pair annihilate into an electrically neutral vector boson (Z^0/γ) which then decays into a pair of leptons of the same flavour and opposite electric charge. In the case of $Z^0 \to \tau\tau$ (figure 4.5, left), the τ leptons can decay leptonically and result in two charged leptons of opposite flavour. $Z^0 \to ee/\mu\mu$ (figure 4.5, right) becomes an important background when considering events that have two charged leptons of the same flavour. Since $Z^0 \to ee/\mu\mu$ events do not have neutrinos in the final state, this background can be reduced using a selection requirement on the missing transverse energy.

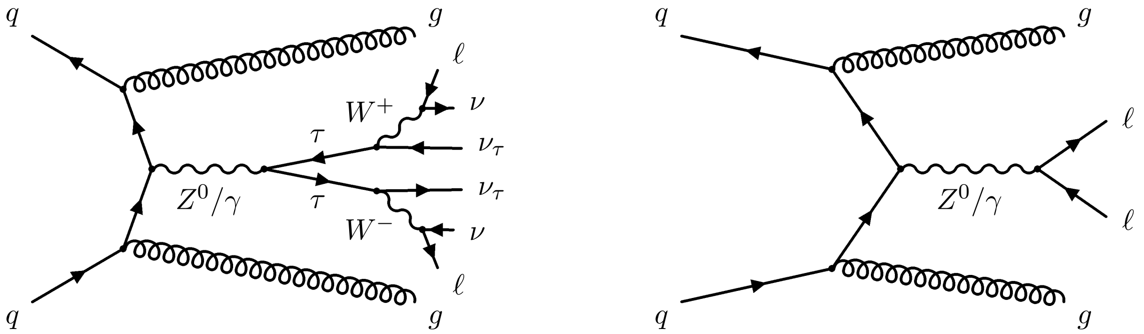


Figure 4.5: Representative Feynman diagrams for $Z^0 \to \tau\tau \to \ell\nu\ell\nu$ (left) and $Z^0 \to ee/\mu\mu$ (right) as expected to enter as background processes in the VBF analysis.

The leptons shown in all of the above Feynman diagrams are referred to as prompt leptons as they originate directly from the hard-scatter.

The overall analysis strategy is to measure the VBF $H \to WW^* \to \ell\nu\ell\nu$ process in a region of phase space that is enriched in signal. This is accomplished by choosing selection requirements on kinematic variables that reject the above mentioned backgrounds but retain as much signal as possible. Since the expected number of VBF $H \to WW^* \to \ell\nu\ell\nu$ events is relatively low, and since there are strong correlations between some of the kinematic quantities of interest, the analysis benefits significantly from introducing a multivariate discriminant. A boosted decision tree (BDT) is trained on simulation of the signal and background processes and then used in the definition of the signal-enriched region.

4.3 Boosted Decision Trees

On the spectrum of multivariate techniques, BDTs are relatively simple. Even though more complex techniques such as neural networks and support vector machines (SVM) can potentially give better results with extensive parameter tuning, BDTs offer very good performance with little optimization. In addition, they also have a simple and intuitive interpretation.

A decision tree is a binary classifier with a tree-like structure. As illustrated in the example in figure 4.6, candidate events are sorted into nodes using selection requirements on a set of j BDT training variables, x_i, where $i = 1, 2, ...j$. Each node split corresponds to one selection requirement on a single variable. One variable can be used at multiple node splits. The process of choosing variables and the corresponding selection requirements for each node split is referred to as the decision tree *training*. It is accomplished by using a *training set*, a representative set of signal and background events with known identity. In particle physics analyses, the training set is often derived from Monte Carlo predictions. After the training is complete, each node at the bottom of the tree can be classified as signal-like or background-like based on if it is dominated by signal or background.

Boosting is a commonly used technique to improve the performance of a decision tree. It involves training additional decision trees using the same training samples but where the misclassified events are given larger event weights. Consequently, the boosted decision trees are more likely to correctly categorize the events that were misclassified in the original decision tree. A boosted decision tree is a collection or *forest* of decision trees. The forest of trees is ultimately combined by taking the average of the constituent decision trees. In the VBF $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ analysis, it is observed that a large forest of shallow trees typically performs better than a small forest of deep trees. Shallow trees are also less likely to be susceptible to statistical fluctuations in the training set.

4.3.1 BDT Training

The objective of the BDT training is to choose selection requirements that result in nodes that are pure in either signal or background. The purity of a node is quantified using the *Gini Index* as separation criteria:

$$\sigma = p \cdot (1 - p) \tag{4.1}$$

where the signal purity, p, is defined as the ratio of the number of signal events to the total number of events in that node. The Gini index ranges from zero (when the node contains

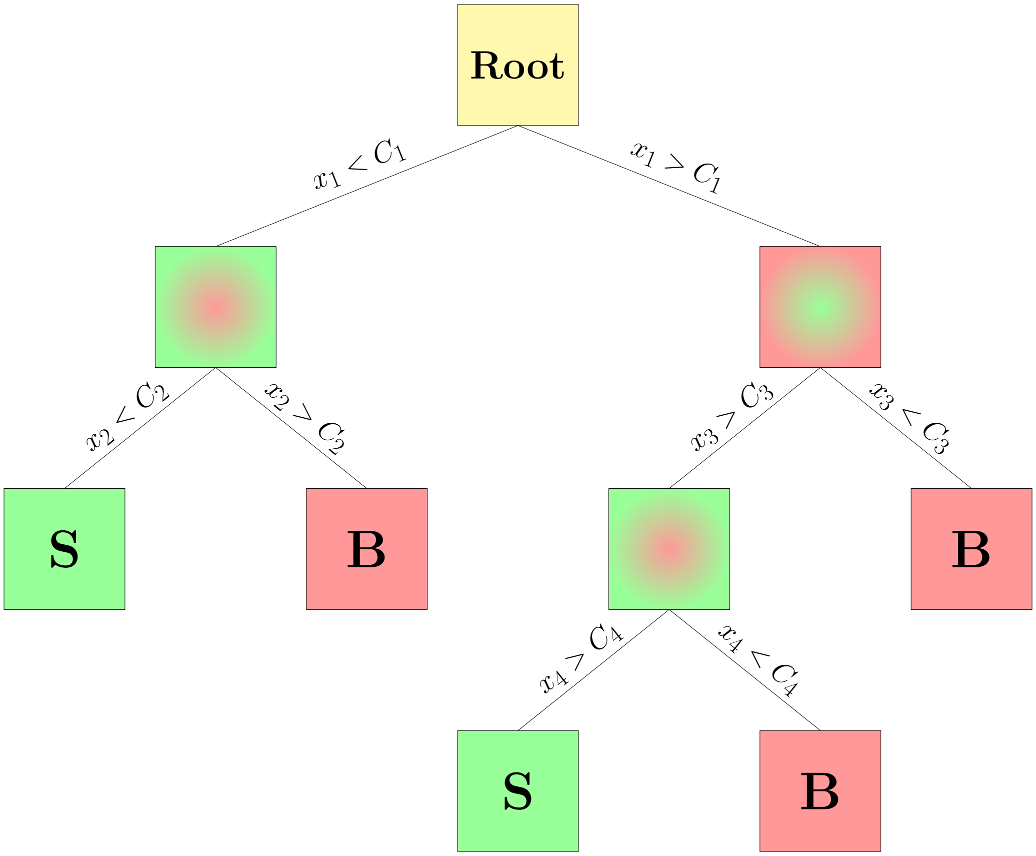


Figure 4.6: Example of a simple decision tree trained with 4 training variables, $x_1, ..., x_4$. The process of determining the values of $C_1, ..., C_4$ is called the training of the decision tree. The selection requirements at each node are used to categorize a set of candidate events (yellow node) as signal (green nodes) or background (red nodes).

only signal or only background) to 0.25 (when a node contains equal parts of signal and background).

The separation gain of a specific selection requirement is defined as:

$$\sigma_{\text{parent}} - (f_{\text{left}}\sigma_{\text{left}} + f_{\text{right}}\sigma_{\text{right}}) \tag{4.2}$$

where

- σ_{parent} is the Gini index of the parent node,
- σ_{left} and σ_{right} are the Gini indices for the two resulting daughter nodes after applying a selection requirement on some variable, and
- f_{left} and f_{right} are the weighted fractions of events in the left and right daughter nodes respectively.

The entire training set starts at the root node as labelled in figure 4.6. A scan is performed over each of the training variables to determine which combination of variable and selection requirement provides the largest separation gain using the definition in equation 4.2. This combination is then used to define two new daughter nodes for which this procedure is repeated iteratively.

The BDT training is subject to two constraints:

- the tree cannot be larger than a predefined maximum tree depth, and
- the number of events in each node cannot be less than the predefined minimum number of events per node.

The BDT training is terminated if it cannot continue without violating either of these constraints. This makes it possible to prevent the tree from training on statistical fluctuations in the training set, an effect known as *overtraining.*

4.3.2 Boosting

The process of boosting involves training a large number of trees and combining them into a single discriminant. In the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis, the gradient boosting method is used. The BDT discriminant is defined using the individual decision trees as basis functions:

$$F(\vec{x}) = \sum_{m=0}^{M} \beta_m f(\vec{x}, \vec{a}_m) \tag{4.3}$$

where:

- M is the number of decision trees in the forest,
- $f(\vec{x}, \vec{a_m})$ represents the m^{th} decision tree in the forest, where $f(\vec{x}, \vec{a_m}) \in [-1, 1]$,

- $\vec{a}_m$ defines the training variable and corresponding selection requirement used at each node for the m^{th} decision tree in the forest,
- β_m is the weight assigned to the output of the m^{th} decision tree in the forest, and
- $\vec{x}$ is a tuple with the values of all the input variables for a specific event.

The BDT training involves finding the parameters β_m and $\vec{a}_m$ such that the difference between $F(\vec{x})$ and the true value y as determined by the training sample is minimized. The degree of deviation is quantified by the *loss-function*, $L(F(\vec{x}), y)$. One example of a loss-function is the exponential loss-function which is defined as

$$L(F(\vec{x}), y) = e^{-F(\vec{x})y} \tag{4.4}$$

The exponential loss-function fully defines the boosting procedure. It corresponds to the AdaBoost algorithm which is one of the most commonly used boosting algorithms for BDTs.

Since $e^{-F(\vec{x})y}$ grows exponentially for large negative values of $F(\vec{x})y$, the AdaBoost algorithm doesn't perform well in the presence of outliers. This weakness is remedied with the binomial log-likelihood loss-function which is still able to retain the good performance of AdaBoost:

$$L(F(\vec{x}), y) = \ln(1 + e^{-2F(\vec{x})y}) \tag{4.5}$$

Unfortunately, it is not easy to derive the boosting procedure for the binomial log-likelihood loss-function. Consequently a numerical steepest-descent approach is used to minimize the loss-function. This is also why the method is called gradient boosting.

The *shrinkage* is a parameter that controls how quickly the BDT learns from the training sample by controlling how much the weight of misidentified events is increased. Reducing the shrinkage makes the BDT output more robust, but may require increasing the size of the forest to reach the same level of performance.

Another method for making the BDT less sensitive to statistical fluctuations in the training sample is called *bagging*. When implemented, a separate random subset of the training sample is used to train each tree. The *bagging fraction* is the fraction of the total training sample that is used for each tree training. When bagging is used in conjunction with gradient boosting as is done in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis, the method is called *stochastic gradient boosting*.

The BDT output is calculated from the forest of decision trees and is denoted by O_{BDT}. In the VBF $H \to WW^*$ analysis, $-1 \leq O_{\text{BDT}} \leq 1$, where -1 corresponds to an event that is background-like, while 1 corresponds to an event that is signal-like.

Chapter 5

Signal and Background Modelling

The kinematic distributions for most of the physics processes relevant to the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis are derived from Monte Carlo simulation. Designated samples are prepared for the dominant backgrounds that specifically populate the phase space of interest in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis.

5.1 Monte Carlo Samples

In order to be able to compare data collected by the ATLAS detector with the predictions from theory, computer software is used to generate simulated data also known as Monte Carlo samples (MC). In general, separate programs are used to model the various characteristics of an event. The hard scatter is simulated by a generator using matrix-element Feynman calculations. Depending on the order of the generator, a limited number of quantum loops corrections and/or real emissions can also be modelled at matrix-element level. These corrections can be classified based on the type of interaction they are mediated by. QCD corrections are those that are mediated by gluons, while EW corrections are those mediated by the electroweak vector bosons.

POWHEG [53] is a widely used generator in both Run-I and Run-II that includes corrections at next-to-leading order (NLO) in QCD. The SHERPA [54] (Run-I and Run-II) and ALPGEN [55] (Run-I only) generators are used when higher parton multiplicities are important. These are able to handle the emission of additional partons at matrix-element level which correspond to higher order diagrams and consequently are called multi-leg generators. In Run-I, the ACERMC [56] and GG2VV [57] leading order (LO) generators are used when no NLO generators are available. In Run-II, MADGRAPH [58] is also used. MADGRAPH is a tree-level multi-leg generator which means it does not incorporate loop corrections, but does include corrections for the emission additional partons.

The CT10 PDF [59] is used with POWHEG and SHERPA in both Run-I and Run-II. In Run-I, the CTEQ6L1 PDF [60] is used with both ALPGEN and ACERMC, while in Run-II, NNPDF23LO PDF [58] is used with MADGRAPH.

The generator is interfaced to a program that models the parton showering, hadronization, and underlying event. Programs used for this in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis include PYTHIA6 [61], PYTHIA8 [62], HERWIG [63] (Run-I only), and SHERPA. In the case of HERWIG, the underlying event is modelled using JIMMY [64]. Care is taken to ensure that the parton shower does not duplicate parton radiation processes already accounted for at generator level.

Next, stable particles originating from the hard scatter and hadronization are propagated through a simulation of the ATLAS detector using GEANT4 [65, 66]. The simulated hits in the detector are digitized to resemble the data collected by the detector. At this stage, pile-up interactions are simulated using PYTHIA8 and overlaid on the generated signal and background interactions. In order to match the pile-up conditions in the MC to that of the data, the distribution of the overlaid interactions in MC is reweighted such that the distributions for the average number of interactions per bunch crossing in the MC and data are the same. Consequently, it is important that the projected pile-up conditions used during the generation of the MC are similar to those in the collected data. If this is not the case, large weights will need to be applied on the MC in regions of phase space where the distribution of the average number of interactions per bunch crossing peaks in the data. This will increase the statistical variance of the MC sample.

The MC generators used in Run-I and Run-II are summarized in table 5.1. More details for each process are given in section 5.1.1 for Run-I and section 5.1.2 for Run-II. There are a large number of generators that can be used to model the signal and background processes. Since each program has strengths and weaknesses, the choice of generator for a specific process is typically a compromise. For example, including higher order loop diagrams in the SHERPA generator can significantly increase the amount of time required to generate each event which leads to smaller MC sample sizes and larger statistical uncertainties. ATLAS has a dedicated team called the Physics Modeling Group (PMG) that, in collaboration with the physics analysis groups, work out recommendations for the different physics processes along with prescriptions for evaluating the associated systematic uncertainties.

5.1.1 Monte-Carlo Samples Used in Run-I

VBF and ggF Higgs boson production are modelled with POWHEG interfaced to PYTHIA8, while the VH production mode is modelled using PYHIA8. The Higgs boson mass is assumed to be $m_H = 125$ GeV. For VBF, the cross section is calculated using approximate next-to-next-to-leading-order (NNLO) QCD corrections [67] and NLO EW [68, 69] corrections. The cross section for ggF Higgs boson production is calculated with NNLO QCD corrections [70] and NLO EW corrections [71]. Corrections are applied from soft

	Process	Run-I MC	Run-II MC
Higgs	VBF $H \to WW^*$	POWHEG+PYTHIA8	POWHEG+PYTHIA8
	ggF $H \to WW^*$	POWHEG+PYTHIA8	POWHEG+PYTHIA8
	VH $H \to WW^*$	PYTHIA8	POWHEG+PYTHIA8 (MiNLO)
Top	$t\bar{t}$	POWHEG+PYTHIA6	POWHEG+PYTHIA6
	tW	POWHEG+PYTHIA6	POWHEG+PYTHIA6
	$t\bar{b}$	POWHEG+PYTHIA6	-
	$tq\bar{b}$	ACERMC+PYTHIA6	-
WW	$q\bar{q}/g \to WW$	SHERPA	POWHEG+PYTHIA8
	$gg \to WW$	GG2VV+HERWIG	SHERPA
	EW WW	SHERPA	SHERPA
DY	Z/γ^*	ALPGEN+HERWIG	MADGRAPH
	EW Z/γ^*	SHERPA	SHERPA
Other Diboson	$W\gamma$	ALPGEN+HERWIG	SHERPA
	$WZ/W\gamma^*$ (low m_{Z/γ^*})	SHERPA	-
	$WZ/W\gamma^*$ (high m_{Z/γ^*})	POWHEG+PYTHIA8	POWHEG+PYTHIA8
	$Z\gamma$	SHERPA	SHERPA
	ZZ	POWHEG+PYTHIA8	POWHEG+PYTHIA8
	EW $WZ/W\gamma^*$	SHERPA	-
	EW ZZ	SHERPA	-

Table 5.1: Overview of the Monte Carlo samples used in the Run-I and Run-II analyses.

gluon resummations up to next-to-next-to-leading-log (NNLL) [72]. As with ggF Higgs boson production, the VH cross sections are corrected to NNLO in QCD [73] and NLO in EW [74]. The branching ratios for the Higgs boson decay modes are calculated using the PROPHECY4F [75] and HDECAY [76] programs.

One of the dominant backgrounds in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis is the top background which consists of $t\bar{t}$ and single top. The $t\bar{t}$ background is modelled at NLO in QCD using POWHEG interfaced to PYTHIA6 and is normalized to NNLO, including a resummation of NNLL soft gluon terms using Top++2.0 [77]. The single top background is broken down into s-channel single top ($t\bar{b}$), t-channel single-top ($tq\bar{b}$), and Wt production. The Wt and s-channel single top hard scatter are simulated using POWHEG, while s-channel single top is simulated using ACERMC. All generators used for single top are interfaced with PYTHIA6. Their normalizations are corrected to approximately NNLO in QCD using a NNLL resummation of soft gluon terms [78–80].

Standard model WW production is divided into two components, QCD WW and EW WW. For QCD WW, the final state quarks or gluons originate from QCD vertices, while for EW WW they are produced via electroweak vertices. QCD WW is modelled using SHERPA 1.4.1 which includes diagrams for $q\bar{q}/qg/\bar{q}g \to WW$. The event generation is done at LO in QCD but includes diagrams for up to three radiated partons in the hard scatter. The normalization is corrected to NLO using MCFM [81]. However, since the MCFM calculation does not include diagrams with two partons in the final state, the normalization does depend on SHERPA's jet multiplicity prediction. The contribution from $gg \to WW$ production via a quark loop is modelled at LO using GG2VV which includes both WW and ZZ diagrams along with their interference. EW WW production is also modelled using SHERPA at LO with two partons in the final state. Interference between EW WW production and VBF $H \to WW^*$ production is not simulated since VBF $H \to WW^*$ is the process that is being measured. However, the uncertainty assigned on the MC prediction for EW WW is calculated to account for this effect.

The DY background is sometimes also referred to as the Z+jets background. As with WW, this background is split into QCD and EW components. QCD Z+jets is generated using ALPGEN interfaced with HERWIG. Events are generated at LO but include tree level calculations for the radiation of up to 5 partons. The overall cross section is corrected to NNLO using DYNNLO [82]. EW Z+jets is modelled using SHERPA.

Even though the main diboson background contribution comes from WW, there are also smaller contributions from other diboson processes. $W\gamma$ is simulated using ALPGEN interfaced to HERWIG with tree-level calculations for up to five radiated partons and includes NLO corrections from MCFM. For $m_{Z/\gamma^*} > 7$ GeV, POWHEG interfaced to PYTHIA8 is used to model $WZ/W\gamma^*$ which correctly accounts for the interference between the Z and γ^* diagrams. Since POWHEG is not able to model events with invariant masses close zero, SHERPA is used to model events with $m_{Z/\gamma^*} < 7$ GeV with up to one radiated parton calculated at matrix level. This sample is normalized using a NLO cross section prediction from MCFM. ZZ is also modelled using POWHEG interfaced to PYTHIA8. The $Z\gamma$ background is modelled using SHERPA and the cross section is normalized to NLO using MCFM. $Z\gamma$ events generated as part of the ALPGEN Z+jets are vetoed to avoid double counting this background. Finally, SHERPA is also used to model EW WZ, EW $W\gamma^*$, and EW ZZ.

5.1.2 Monte-Carlo Samples Used in Run-II

MC samples for all of the major backgrounds are also included for the most recent Run-II result. The processes that are not included in Run-II, but are included in Run-I have a negligible impact.

In Run-II, the VBF, ggF, and VH Higgs boson production modes are modelled with POWHEG-BOX v2 and are interfaced with PYTHIA8. For the VH samples, the MiNLO [83] treatment is applied. As in Run-I, the Higgs boson mass is assumed to be $m_H = 125$ GeV. For VBF, the cross section is calculated using approximate NNLO QCD corrections and NLO EW corrections. The cross section calculated for the ggF Higgs boson production mode is improved with respect to Run-I with next-to-next-to-next-to-leading order (NNNLO) QCD corrections [84] and NLO EW corrections. As in Run-I, the VH cross sections are corrected to NNLO in QCD and NLO in EW. The branching ratios for the Higgs boson decay are calculated using the HDECAY program.

The top related backgrounds, $t\bar{t}$ and tW are modelled using POWHEG-BOX v2 interfaced to PYTHIA6. The cross section for $t\bar{t}$ is normalized to NNLO in QCD using the TOP++2.0 program and includes soft gluon resummation terms to NNLL. Unlike in Run-I, the contributions from $t\bar{b}$ and $tq\bar{b}$ are not simulated with MC but are part of the data-driven W+jet estimate which is described in section 7.5.

POWHEG-BOX v2 interfaced with PYTHIA8 is used to model WW production initiated by $q\bar{b}/qg/\bar{q}g$ and the cross section is corrected to NNLO accuracy in QCD [85]. The contribution from $gg \rightarrow WW$ is modelled using SHERPA v2.1.1 with up to one additional radiated parton at NLO accuracy. The cross section is corrected to NLO [86]. SHERPA is also used to model EW WW production with two partons in the final state at LO accuracy.

QCD Z+jets is modelled using MADGRAPH5 at LO, but includes calculations for the radiation of up to four partons. This generator is interfaced to PYTHIA8. As in Run-I, $Z\gamma$ events that are generated in the QCD Z+jets sample are vetoed as they are accounted for by a separate sample. For the EW Z+jets processes, only the $Z \rightarrow \tau\tau$ decay channel is considered as this is the dominant contribution. It is modelled using SHERPA at LO.

Both $Z\gamma$ and $W\gamma$ are modelled using SHERPA v2.1.1 at LO, including up to three radiated partons in the matrix element calculation. WZ and ZZ are modelled with POWHEG-BOX v2 interfaced with PYTHIA8 along with the WW background. The EW diboson processes besides WW are small, and are not modelled in the Run-II analysis.

5.2 VBF-Filtered Samples

During the Run-I analysis, the VBF analysis shared the same MC samples as the ggF analysis. However, since the signal-like events are located in the high m_{jj} tail, MC statistics are often a limiting factor in the analysis optimization. Figure 5.1 shows the reconstructed m_{jj} distribution for $t\bar{t}$ in the most sensitive bin of the Run-I signal region. The limited statistics make it impractical to split this bin into smaller and more sensitive bins. In addition, a lack of MC statistics in the signal region will also result in suboptimal BDT performance as the sample sizes used in the BDT training will not be large enough to learn the features of the signal.

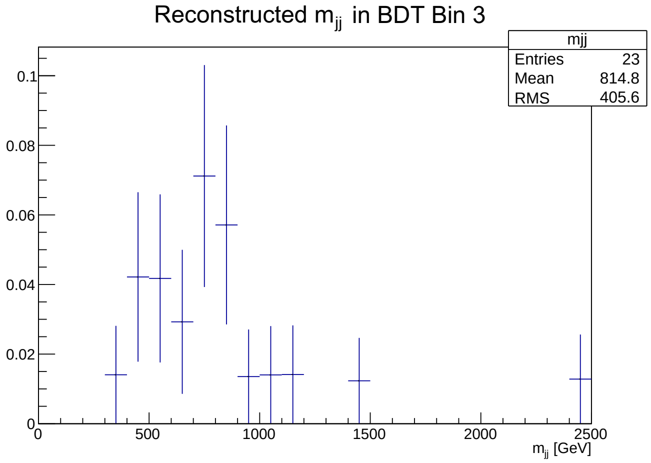


Figure 5.1: Reconstructed m_{jj} distribution from MC for $t\bar{t}$ in the most sensitive bin of the Run-I signal region. The uncertainty bars are statistical only.

In Run-II, designated MC samples are prepared that specifically populate the high m_{jj} tail using a truth level filter. These samples are referred to as *VBF-filtered* samples. The non-filtered samples used by default in the ggF analysis are called the *nominal* samples.

Truth level filters are used widely within ATLAS. They typically involve binning in one or more truth variables at the event generation stage of the MC production. The number of events in each bin is then optimized such that the phase space of interest is well populated.

Recall from equation 1.7 that there is a strong correlation between the opening angle of the charged leptons and the corresponding dilepton mass. The same reasoning can be applied to the forward jets that are characteristic of VBF Higgs boson production as described in section 1.3. Since the opening angle between the jets is expected to be large for VBF Higgs boson candidate events, the invariant mass of the dijet system is also expected to be large. Consequently, the filter used for this analysis is binned in truth m_{jj}, where the jets are reconstructed using the anti-k_T algorithm on truth level particles before they pass through the detector simulation.

Since the VBF signal events span a large range of m_{jj}, this concept is extended to an infinite number of bins in truth m_{jj} for the VBF analysis. In general, the larger the truth m_{jj}, the larger the fraction of generated events that are kept by the filter. The MC event weights are then adjusted to compensate for the discarded events.

For most processes, the nominal samples were already generated when the filtered samples were produced. Consequently, the VBF-filtered samples are merged with the nominal samples whenever possible. This eliminates the need to generate events twice for low m_{jj}.

The filter is defined on the interval $[m_{jj}^{low}, m_{jj}^{high}]$ in terms of a function, $f(m_{jj})$, that describes the fraction of events that are kept at each value of m_{jj}:

$$f(m_{jj}) = \left(\frac{m_{jj}}{m_{jj}^{low}}\right)^{\alpha} f(m_{jj}^{low}), \quad \text{where} \quad \alpha = \frac{\ln\left[f(m_{jj}^{low})/f(m_{jj}^{high})\right]}{\ln\left(m_{jj}^{low}/m_{jj}^{high}\right)} \tag{5.1}$$

$f(m_{jj})$ is bounded by $f(m_{jj}^{low})$ and $f(m_{jj}^{high})$ where $0 < f(m_{jj}^{low}), f(m_{jj}^{high}) \leq 1$. The parameter α describes the shape of the function between m_{jj}^{low} and m_{jj}^{high}. For the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis, α is optimized to make the unweighted m_{jj} distribution fall off less quickly such that the high m_{jj} tails are sufficiently populated.

Outside of $[m_{jj}^{low}, m_{jj}^{high}]$, $f(m_{jj})$ is constant and so the unweighted m_{jj} distribution in this range will be the same as for the nominal samples:

$$f(m_{jj}) = \begin{cases} f(m_{jj}^{low}), & m_{jj} < m_{jj}^{low} \\ f(m_{jj}^{high}), & m_{jj} > m_{jj}^{high} \end{cases} \tag{5.2}$$

For each MC event generated, a uniform random number, $R \in (0, 1)$ is generated. The event is only included in the sample if $R < f(m_{jj})$.

Since a fraction of the generated events are discarded, the event weights are adjusted accordingly:

$$w(m_{jj}) = \frac{1}{f(m_{jj})} \tag{5.3}$$

For some fraction of generated events, there will be less than two truth jets and in these cases it will not be possible to reconstruct the truth m_{jj}. Consequently, f_0 and f_1 are defined to be the fraction of events kept for all events with 0 and 1 truth jets respectively. These events are usually covered by the nominal sample if it exists, but if not, a common choice is to set $f_0 = f_1 = f(m_{jj}^{low})$. This ensures that the event weights of events with 0 and 1 truth jets that end up in the signal region are comparable to the other events in this region.

For a given MC sample, the filter configuration is fully defined in terms of 6 parameters m_{jj}^{low}, m_{jj}^{high}, $f(m_{jj}^{low})$, $f(m_{jj}^{high})$, f_0 and f_1. In addition, it is necessary to specify if the filtered sample is to be merged with an existing nominal sample. If so, then the sample is truncated at m_{jj}^{low} since the nominal sample will cover $m_{jj} < m_{jj}^{low}$ and events with less than 2 truth level jets.

The merging boundary for the VBF-filtered samples and nominal samples is chosen to align with m_{jj}^{low}. Since $t\bar{t}$ is the largest background in the VBF analysis, it is used as a baseline in configuring the VBF-filtered samples. As illustrated in figure 5.1, most of the m_{jj} distribution for $t\bar{t}$ in the most sensitive bin of the Run-I signal region is above 500 GeV. In addition, with a merging boundary at truth $m_{jj} = 500$ GeV, only 4% of the Run-II nominal sample is discarded as shown in figure 5.2. Consequently, the VBF-filtered sample is used for truth $m_{jj} > m_{jj}^{low} = 500$ GeV, while the nominal sample is used for truth $m_{jj} < m_{jj}^{low} = 500$ GeV. Next, m_{jj}^{high} is chosen to be 2.5 TeV since it was projected using MC at the time of optimization that there would be approximately 2 VBF candidate events with $m_{jj} > 2.5$ TeV by the end of 2016. This prediction assumed an integrated luminosity of ~ 30 fb^{-1} of 13 TeV data by the end of 2016, but recall from figure 2.2 that the amount of data delivered by the LHC exceeds this prediction. As more data is collected, the number of VBF candidate events expected beyond m_{jj}^{high} will increase and so future samples may require higher values for m_{jj}^{high}.

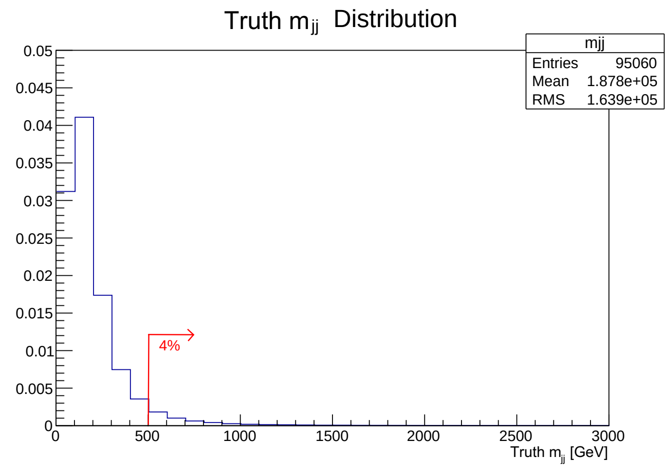


Figure 5.2: Run-II truth m_{jj} distribution in MC for $t\bar{t}$ after requiring at least two truth jets and two truth leptons.

Since events at high truth m_{jj} are the most signal-like, $f(m_{jj}^{high})$ is defined to be 1.0 which is equivalent to stating that no events are discarded for truth $m_{jj} \geq 2.5$ TeV. By fitting equation 5.1 to the median m_{jj} in two bins of the Run-I signal region, $f(m_{jj}^{low})$ is calculated to be 5.02×10^{-3} which means that approximately 1 in 200 events is kept for $m_{jj} \leq 500$ GeV. In the VBF $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ analysis, $f(m_{jj}^{high})$ is always 1.0 but $f(m_{jj}^{low})$ is adjusted for samples with m_{jj} distributions that are different from that of $t\bar{t}$.

As a sanity check, the ratio of $t\bar{t}$ events in the 100 GeV window around $2,500$ GeV to 100 GeV windows around various values of truth m_{jj} is compared to $f(m_{jj})$ in table 5.2. Note that the slope of $f(m_{jj})$ is more shallow than the natural slope of $\frac{1}{m_{jj}}$. Consequently, the unweighted distribution of m_{jj} is still expected to drop on $f(m_{jj}^{low}) < m_{jj} < f(m_{jj}^{high})$ but less rapidly than the nominal m_{jj} distribution on the same range.

Truth m_{jj}	$f(m_{jj})$	Truth m_{jj} window	Ratio
200 GeV	2.47×10^{-4}	150 GeV $< m_{jj} <$ 250 GeV	5.68×10^{-5}
500 GeV	5.02×10^{-3}	450 GeV $< m_{jj} <$ 550 GeV	6.05×10^{-4}
1,000 GeV	4.91×10^{-2}	950 GeV $< m_{jj} <$ 1,050 GeV	8.34×10^{-3}
1,500 GeV	1.86×10^{-1}	1,450 GeV $< m_{jj} <$ 1,550 GeV	5.66×10^{-2}
2,000 GeV	4.80×10^{-1}	1,950 GeV $< m_{jj} <$ 2,050 GeV	2.49×10^{-1}
2,500 GeV	1.0	2,450 GeV $< m_{jj} <$ 2,550 GeV	1.0

Table 5.2: Comparison of $f(m_{jj})$ to the ratio of $t\bar{t}$ events in the 100 GeV window around 2,500 GeV to 100 GeV windows around selected values of truth m_{jj}.

Finally, for each filtered sample, it is necessary to specify the size of the sample. In general, the sample-size was chosen such that statistical uncertainty at the merging boundary is approximately the same for the filtered and nominal sample. Figure 5.3 shows the statistical uncertainty as a function of the truth m_{jj} for the $t\bar{t}$ sample. Note that the uncertainties in this plot are much smaller than in the signal region since this plot is shown before the analysis selection is applied.

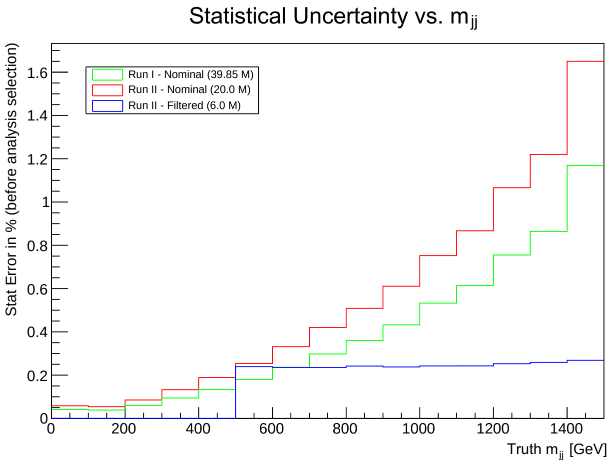


Figure 5.3: The statistical uncertainty on the $t\bar{t}$ sample as a function of the truth m_{jj} after requiring at least two truth jets and at least two truth leptons. The sample-size of the filtered sample (blue) is chosen such that its statistical uncertainty at 500 GeV is approximately the same as for the Run-II nominal sample (red). For comparison, the Run-I statistical uncertainty is also included (green).

The truth m_{jj} distribution for the filtered $t\bar{t}$ sample without applying weights is shown in figure 5.4. As mentioned previously, this distribution drops off much slower than the nominal distribution and demonstrates how the filter shifts the m_{jj} distribution to higher values.

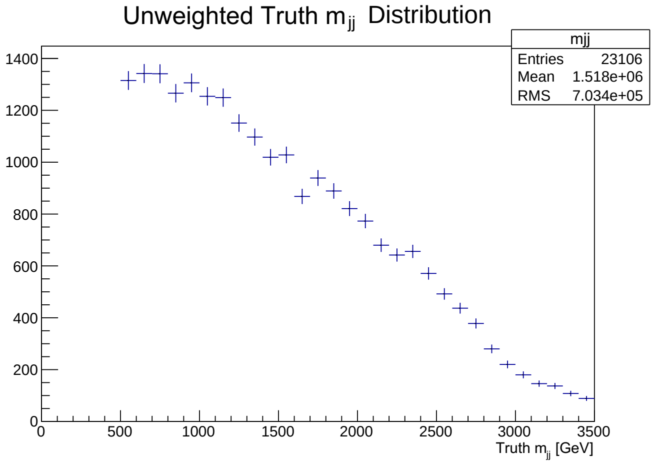


Figure 5.4: Truth m_{jj} distribution for the filtered $t\bar{t}$ sample without applying weights after requiring at least two truth jets and at least two truth leptons.

Figure 5.5 shows the truth m_{jj} distribution for the nominal and filtered samples after applying the event weights. Note that the relative statistical uncertainties on the nominal sample are large at high truth m_{jj} while this is not the case for the filtered sample.

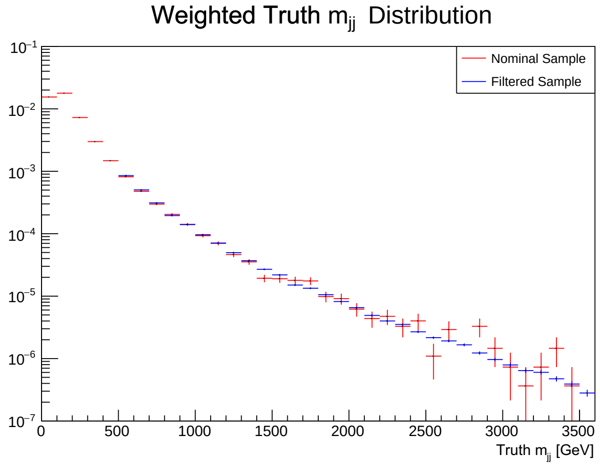


Figure 5.5: Weighted truth m_{jj} distribution for the nominal (red) and filtered (blue) $t\bar{t}$ samples after requiring at least two truth jets and at least two truth leptons. The distributions have been normalized to 1 event.

A similar procedure is followed for the other VBF-filtered samples. Table 5.3 summarizes the VBF-filtered samples prepared for the Run-II VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis. Note that all samples except the SHERPA EW WW use the same slope parameter, α. In the case of EW WW, the two W bosons are radiated from two high energy quarks that result in two forward jets in the detector. Consequently, these events are expected to have larger values for m_{jj} on average and so α is lowered such that the the unweighted slope in m_{jj} is similar to that of the $t\bar{t}$ sample.

Since nominal samples were not available for the SHERPA EW WW sample and the MADGRAPH $t\bar{t}$ sample at the time the filtered samples were prepared, these samples are not truncated at m_{jj}^{low} unlike the other samples. Consequently, m_{jj}^{low} is set to 100 GeV and $f(m_{jj}^{low})$ is used for $m_{jj} < 100$ GeV. In addition, f_0 and f_1 are set to $f(m_{jj}^{low})$ for both these samples. For the other samples, the 0 and 1 jet events already exist in the nominal sample and so f_0 and f_1 are set to zero.

Sample	Generator	Size	α	m_{jj}^{low}	m_{jj}^{high}	$f(m_{jj}^{low})$	$f(m_{jj}^{high})$
ggF	POWHEG+PYTHIA8	200 K	3.3	500 GeV	2,500 GeV	0.005	1.0
WW	POWHEG+PYTHIA8	200 K	3.3	500 GeV	2,500 GeV	0.005	1.0
EW WW	SHERPA 2.1	200 K	1.1	100 GeV	2,500 GeV	0.025	1.0
$t\bar{t}$	POWHEG+PYTHIA6	6 M	3.3	500 GeV	2,500 GeV	0.005	1.0
$t\bar{t}$	MADGRAPH5	7.5 M	3.3	100 GeV	2,500 GeV	0.000025	1.0

Table 5.3: Summary of the VBF-filtered samples prepared for the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis along with the parameters used to generate them.

Unfortunately, for the results summarized in this thesis, most of the filtered samples are only used in the BDT training. Due to an ATLAS-wide technical issue related to the propagation of event weights, only the SHERPA EW WW could be processed in time to be included in the statistical analysis. Nevertheless, the VBF-filtered samples provide a promising method to populate the region of interest. Future iterations of the analysis are expected to improve with the inclusion of additional samples for other sub-dominant backgrounds.

Chapter 6

Candidate Event Selection

The VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis benefits considerably from the development and optimization of a BDT. Kinematic selection requirements applied before the training of the BDT are referred to as *preselection*, while those applied after the training are referred to as *selection*. The signal-enriched region of phase space that results after applying the preselection and the selection is called the *signal region* (SR).

6.1 Preselection

All candidate events selected in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis are required to pass a single lepton trigger (e or μ). In the Run-I analysis, candidate events passing a dilepton trigger (ee, $\mu\mu$, or $e\mu$) are also included. The dilepton trigger allows for lower lepton p_T thresholds while still keeping the trigger rate manageable. Future iterations of the Run-II analysis will tentatively also include dilepton triggers. Tables 6.1 and 6.2 summarize the trigger lepton p_T thresholds for the triggers used in Run-I and Run-II respectively.

	2012 trigger p_T thresholds (GeV)	
Type	Level-1 trigger	High-level trigger
e	18 or 30	24 or 60
μ	15	24 or 36
e, e	10 and 10	12 and 12
μ, μ	15	18 and 8
e, μ	10 and 6	12 and 8

Table 6.1: Lepton trigger p_T thresholds used in the Run-I analysis.

Candidate events are required to contain exactly two leptons of opposite charge with $p_T > 10$ GeV in Run-I and $p_T > 15$ GeV in Run-II. Since one of the W bosons coming from the Higgs boson decay is off-shell, a low p_T threshold is important to maximize the

	2015 trigger p_T thresholds (GeV)		2016 trigger p_T thresholds (GeV)	
Type	Level-1 trigger	High-level trigger	Level-1 trigger	High-level trigger
e	20	24 or 60 or 120	20	24 or 60 or 140
μ	15	20 or 50	15	24 or 50

Table 6.2: Lepton trigger p_T thresholds used in the Run-II analysis.

signal efficiency. However, decreasing the p_T threshold also increases the background contamination from light-flavour jets and photon conversions being misidentified as electrons. A higher threshold for Run-II is chosen to offset the increased production cross sections for QCD processes. The lepton with the higher p_T is referred to as the leading lepton, while the other lepton is referred to as the subleading lepton.

The Run-I analysis is divided into two categories, one where the leptons have different flavour (DF: $e\mu$[1] and μe) and one where the leptons have the same flavour (SF: ee and $\mu\mu$). As mentioned in section 4.2, the Drell-Yan (DY) background in the SF channel is significantly larger than in the DF channel. Since the SF channel is less sensitive than the DF channel, it is not included in the first Run-II analysis summarized in this thesis.

In Run-I the leading lepton is required to have $p_T > 22$ GeV. In Run-II, the leading lepton is required to have $p_T > 25$ GeV to account for the higher p_T thresholds in the lepton triggers. The only exception is that muons in the 2015 data are required to have $p_T > 22$ GeV.

After the lepton requirements outlined above, a large fraction of the background is DY, especially for the SF channel. In the SF analysis, $Z/\gamma^* \rightarrow \ell\ell$ is reduced significantly by demanding $|m_{\ell\ell} - m_Z| > 15$ GeV, a selection requirement called the Z veto. The $\gamma^* \rightarrow \ell\ell$ process is reduced by requiring $m_{\ell\ell} > 12$ GeV for the SF analysis and $m_{\ell\ell} > 10$ GeV for the DF analysis. This selection requirement also removes low mass resonances such as J/Ψ and Υ mesons. In addition, for the SF channel, $m_{\ell\ell} < 75$ GeV is applied to ensure the signal region is orthogonal to the $Z \rightarrow ee/\mu\mu$ control region described in section 7.4. Note that this selection requirement also removes all candidate events removed by the Z veto. Figure 6.1 shows the SF $m_{\ell\ell}$ distribution after the preselection but with the Z veto removed. The orange line corresponds to the selection requirement applied on $m_{\ell\ell}$. In the DF analysis, the DY contributions are primarily from $Z/\gamma^* \rightarrow \tau\tau \rightarrow e\nu\mu\nu$.

As evident from the Feynman diagrams in section 4.2, $Z \rightarrow ee/\mu\mu$ is the only dominant background that is not expected to have neutrinos in the final state. Consequently, this background is further reduced by applying selection requirements on the missing transverse energy in the SF channel, $E_T^{miss} > 45$ GeV and $p_T^{miss} > 40$ GeV. No missing transverse

---

[1] The ordering of the particle symbols is by the p_T of the leptons. The first particle is the leading lepton and the second is the subleading lepton.

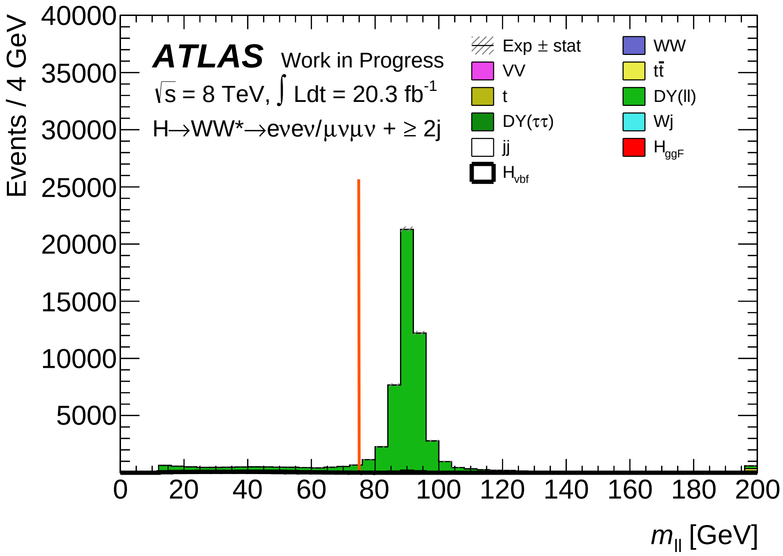


Figure 6.1: The SF $m_{\ell\ell}$ distribution is shown after the preselection, but without applying the Z veto. The Z veto removes a significant fraction of the $Z \to \ell\ell$ background by removing all candidate events to the right of the orange line. The yields are normalized to the expected yields for 20.3 fb^{-1} of Run-I data. The shaded uncertainty bands shown on the background prediction are statistical only.

energy selection requirements are applied for the DF channel since the contribution from $Z \to ee/\mu\mu$ is expected to be small in this channel.

Since VBF Higgs boson candidate events are characterized by two forward jets, each candidate event is required to have at least two jets. The dominant background for jet multiplicities of 2 or more is $t\bar{t}$. It is considerably reduced by requiring the number of b-jets with $p_T > 20$ GeV to be zero, a selection requirement known as the b-jet veto.

Table 6.3 summarizes the analysis preselection for Run-I and Run-II. After applying the preselection, the expected VBF signal purity is about 0.3% for DF and 0.2% for SF.

6.2 Implementation of the BDT

The BDT is trained after the preselection described in section 6.1. The preselection is chosen to reduce backgrounds while rejecting a minimal amount of signal. The preselection also removes regions of phase space that are poorly modelled by the MC such as at low lepton p_T for which the fake lepton background is hard to model. Note that it is possible to apply more one-dimensional selection requirements to further reduce the backgrounds. However, since it applies selection requirements in multiple dimensions, the BDT is able to achieve a significantly better signal efficiency for the same level of background rejection.

Variable	Run-I	Run-II
Lepton charge	opposite	opposite
$p_T^{\ell 1}$	> 22 GeV	$> 22/25$ GeV (2015 μ/ 2015 e, 2016 $e\mu$)
$p_T^{\ell 2}$	> 10 GeV	> 15 GeV
$m_{\ell\ell}$	$> 10/12$ GeV (DF/SF)	> 10 GeV
$m_{\ell\ell}$	< 75 GeV (SF only)	-
$\lvert m_Z - m_{\ell\ell}\rvert$	> 15 GeV (SF only)	-
p_T^{miss}	> 40 GeV (SF only)	-
E_T^{miss}	> 45 GeV (SF only)	-
n_{jet}	≥ 2	≥ 2
n_b	$= 0$	$= 0$

Table 6.3: Summary of the preselection applied before training the BDT. The Run-I analysis is split into DF and SF channels, while the SF channel is omitted in the Run-II analysis.

6.2.1 BDT Training Variables

The BDT training variables used are kinematic quantities derived from the jet kinematics, the charged lepton kinematics and the missing transverse energy. The initial BDT trainings in Run-I were based on a set of approximately 30 training variables. Many of these variables were correlated with each-other and so added unnecessary complexity to the BDT. For example, in addition to including individual rapidities for the two jets in the BDT, the difference in rapidity was also included and so constituted redundant information. In addition, some of the variables suffered from poor MC modelling when compared with the data in regions of phase space away from the signal region. In particular, variables that depend on the modelling of low p_T QCD activity were found to be poorly modelled by the MC simulation.

An iterative procedure called the "$N-1$ minimal loss pruning" is used to reduce the number of training variables without significantly degrading the performance of the BDT. Consider a set of N training variable candidates. A set of N unique BDTs are trained, where each BDT has one training variable removed. Consequently, each BDT training has $N-1$ training variables. The variable removed for the best performing BDT is then removed from the set of training variable candidates. The procedure is repeated until the performance of the best performing BDT starts dropping significantly below the performance of the initial N-variable BDT.

The BDT performance was initially estimated using a Gaussian significance estimate, $\frac{S}{\sqrt{B}}$, where S and B are the total signal and background yields to the right of O_{BDT} corresponding to a fixed value of S. This metric was used to reduce the total number of BDT training variables to 11. This procedure reduced the complexity of the BDT without significantly reducing the performance of the analysis. The total number of training variables was further reduced to eight, using the full statistical analysis described in chapter 9 to assess the performance of a particular configuration. Even though the latter method is much more representative of the analysis performance, it is also much more time consuming.

The Run-I kinematic distributions for the eight selected BDT training variables are shown for DF in figure 6.2 and SF in figure 6.3. These plots are made using the candidate events that satisfy the selection requirements, but before the BDT distribution is split into bins. Each distribution is described in more detail in the text that follows. Note that the signal yield has been scaled up by a factor of 50 to match the yield of the total background.

The top row of plots in figure 6.2 make use of the spin-0 nature of the Higgs boson that favours small opening angles between the charged leptons as described in section 1.4. As shown in equation 1.7, it also follows that the mass of the dilepton system is expected to be small. Both the invariant mass of the charged leptons, $m_{\ell\ell}$, and the opening angle between the charged leptons in the transverse plane, $\Delta\phi_{\ell\ell}$, are included in the BDT training. While $\Delta\phi_{\ell\ell}$ only includes the transverse angular separation, $m_{\ell\ell}$ also incorporates information on the energies of the charged leptons.

Due to the neutrinos in the final state, it is not possible to reconstruct the Higgs boson mass. However, by invoking conservation of momentum in the transverse plane, it is possible to define the transverse Higgs boson mass:

$$m_T = \sqrt{(E_T^{\ell\ell} + p_T^{\nu\nu})^2 - |\vec{p}_T^{\,\ell\ell} + \vec{p}_T^{\,\nu\nu}|^2} \tag{6.1}$$

where $E_T^{\ell\ell} = \sqrt{(p_T^{\ell\ell})^2 + (m_{\ell\ell})^2}$. As evident in the left plot in the second row of figure 6.2, the m_T distribution is bounded by the Higgs boson mass which helps reject the $t\bar{t}$ and WW backgrounds. The transverse Higgs boson mass is included in the BDT training.

The two quarks radiating the weak vector bosons in VBF Higgs boson production are expected to result in jets in the forward regions of the detector. Consequently, the rapidity gap between the leading and subleading jets, $\Delta y_{jj} = |y_{j1} - y_{j2}|$, is expected to be large compared to the background as shown in the right plot in the second row of figure 6.2. Using a similar kinematic argument as with the dilepton system, large opening angles between the two jets corresponds to the large dijet masses. While Δy_{jj} represents the angular separation of the jets, m_{jj} incorporates information about their energies and momenta. As illustrated in the left plot of the third row in figure 6.2, signal-like candidate events are expected to have large m_{jj}. Both Δy_{jj} and m_{jj} are included in the BDT training.

Next, the charged leptons coming from the Higgs boson decay are expected to be centered between the leading and subleading jets. The centrality of a lepton, ℓ, with respect to the two leading jets is defined as:

$$C_\ell = \left|\eta_\ell - \frac{\eta_{j1} + \eta_{j2}}{2}\right| \bigg/ \frac{|\eta_{j1} - \eta_{j2}|}{2} \tag{6.2}$$

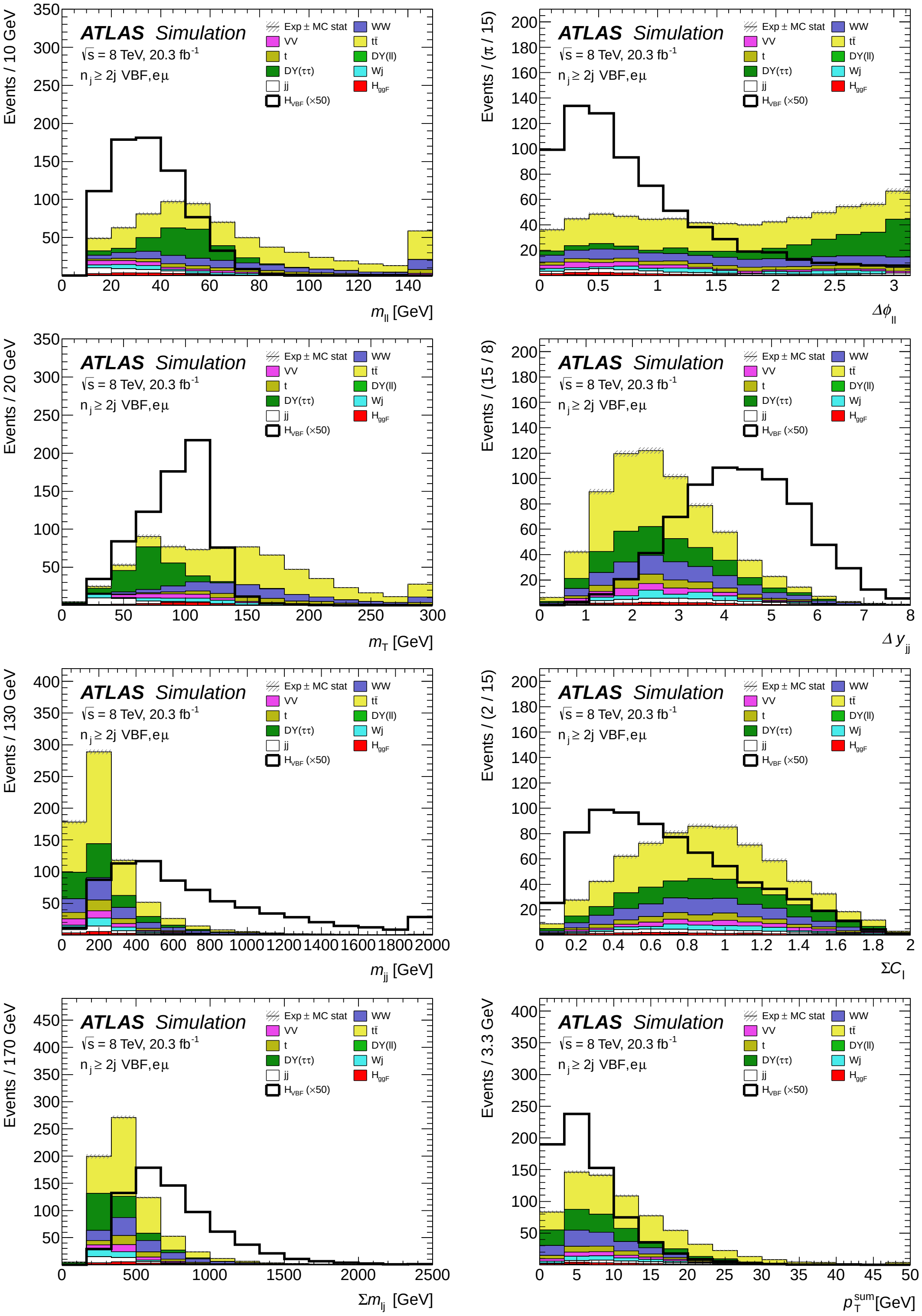


Figure 6.2: Kinematic distributions of the BDT training variables in the different flavour sample used in the Run-I analysis [16]. The plots are shown at the selection stage immediately before the signal region is split into bins of BDT. The yields are normalized to the expected yields for 20.3 fb^{-1} of Run-I data and the VBF signal is scaled up by a factor of 50. The shaded uncertainty bands shown on the background prediction are statistical only.

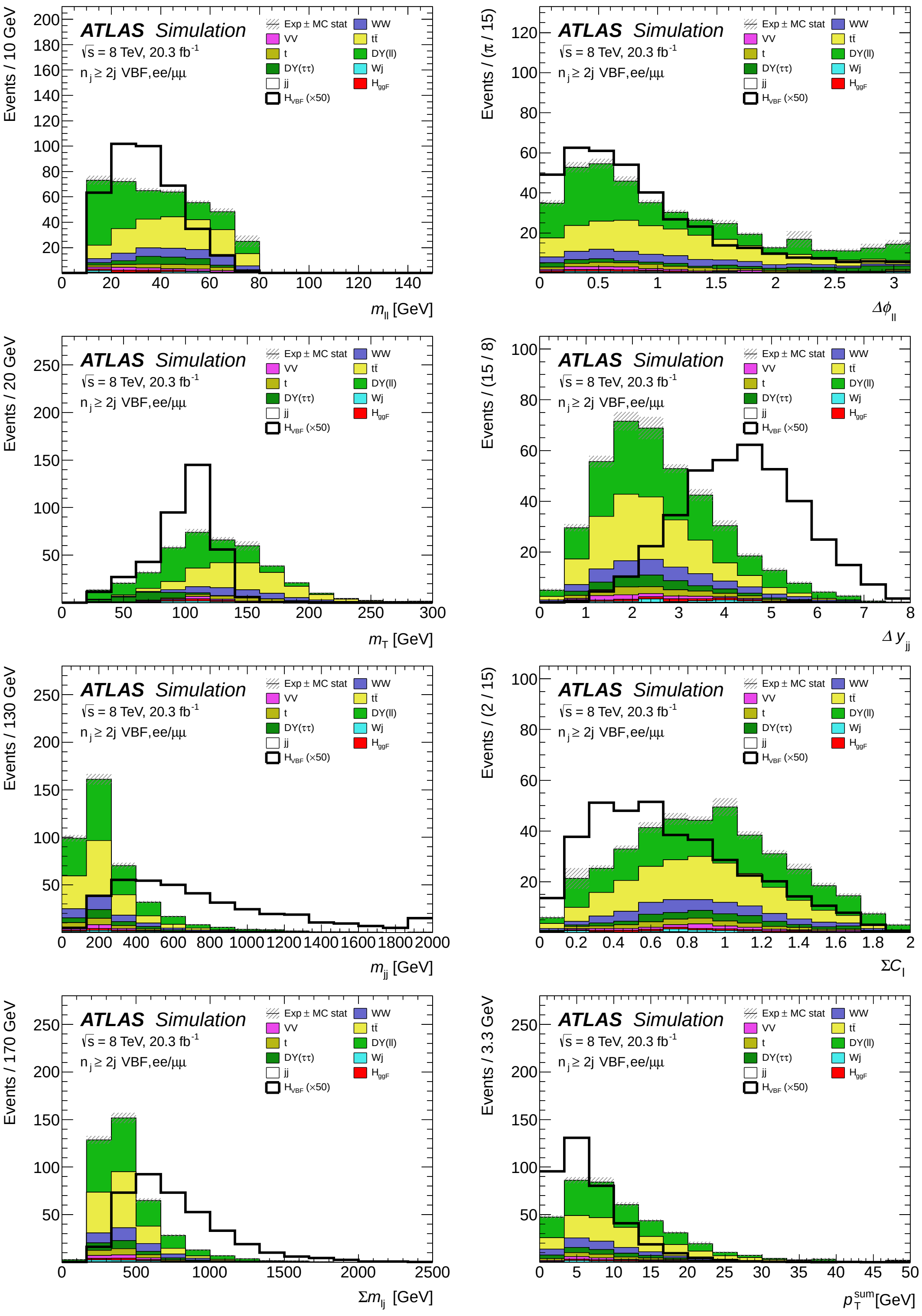


Figure 6.3: Kinematic distributions of the BDT training variables in the same flavour sample used in the Run-I analysis [16]. The plots are shown at the selection stage immediately before the signal region is split into bins of BDT. The yields are normalized to the expected yields for 20.3 fb^{-1} of Run-I data and the VBF signal is scaled up by a factor of 50. The shaded uncertainty bands shown on the background prediction are statistical only.

As illustrated in figure 6.4, C_ℓ is zero when the lepton is exactly centered in the rapidity gap between the two jets (as indicated by the red line) and increases to 1 when the lepton is aligned with either of the jets (as indicated by the blues lines). If the lepton is outside the rapidity gap spanned by the two jets, C_ℓ is larger than 1.

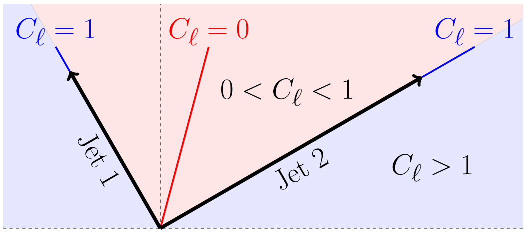


Figure 6.4: The centrality of a lepton, C_ℓ, is a measure of how central it is in rapidity with respect to the two jets with the highest p_T represented by the black arrows in the diagram. It ranges from 0 to 1 for leptons inside the rapidity range of the jets (red shaded area), where $C_\ell = 0$ if the lepton is centered between the two leading jets (red line), and $C_\ell = 1$ if the lepton is aligned with either one of the leading jets (blue lines). For $C_\ell > 1$, the lepton is outside the rapidity range of the two leading jets (blue shaded area).

The sum of the centralities of the two leptons, $\sum C_\ell = C_{\ell 1} + C_{\ell 2}$, is used as a training variable in the BDT. As evident from the right plot in the third row of figure 6.2, the charged leptons for signal candidate events are more central with respect to the jets than for the backgrounds.

A related variable that is also included in the BDT training is the sum of the four combinations of lepton and jet masses. It is defined as:

$$\sum m_{\ell j} = m_{\ell 1, j1} + m_{\ell 1, j2} + m_{\ell 2, j1} + m_{\ell 2, j2} \tag{6.3}$$

As shown in the bottom left plot of figure 6.2, $\sum m_{\ell j}$ is expected to be larger for VBF Higgs boson candidate events since they have larger opening angles between the leptons and jets.

Even after the b-tag veto, $t\bar{t}$ makes up a significant portion of the background and so a variable is included in the BDT that specifically targets $t\bar{t}$ reduction. Since a b-veto is applied in the signal region, a large fraction of $t\bar{t}$ events in this region consist of one b-jet that is missed by the b-tagging algorithm, and one jet originating from QCD radiation. Consequently, the $t\bar{t}$ candidate events that make their way into the signal region are specifically those with additional QCD activity. In addition, since gluon-gluon annihilation is the dominant production mode for $t\bar{t}$, these events are typically accompanied by more low p_T QCD radiation than VBF Higgs boson production. In order make this feature accessible to the BDT, the vectorial $\vec{p}_T$ sum of all final state objects is defined as:

$$\vec{p}_T^{\,sum} = \vec{p}_T^{\,\ell\ell} + \vec{p}_T^{\,miss} + \sum \vec{p}_T^{\,j} \tag{6.4}$$

where $\sum \vec{p}_T^{\,j}$ includes all jets in the candidate event. The magnitude of $\vec{p}_T^{\,sum}$, which is included as a training variable in the BDT, is shown in the bottom right plot of figure 6.2. This variable is expected to be close to zero for signal like candidate events, but will move away from zero for processes with significant soft gluon radiation.

Table 6.4 summarizes the training variables included in the BDT.

Variable	Motivation
Δy_{jj} m_{jj}	Two leading jets are expected to be forward
$\Delta\phi_{\ell\ell}$ $m_{\ell\ell}$	Opening angle between leptons is expected to be small
m_T	Proxy on Higgs boson mass
$\sum C_\ell$ $\sum m_{\ell j}$	Leptons are expected to be lie centered in rapidity gap of two leading jets
p_T^{sum}	$t\bar{t}$ background rejection

Table 6.4: Summary of the kinematic variables used in the BDT training.

Figure 6.5 shows the O_{BDT} distribution for the DF and SF analyses. Even though each of the individual BDT training variables only provides a limited amount of separation power to extract the signal from the large backgrounds, the BDT is able to combine their separation power into a single discriminant. As expected, the signal distribution peaks near $O_{\mathrm{BDT}} = 1$, while the backgrounds peak near $O_{\mathrm{BDT}} = -1$.

6.2.2 BDT Training Configuration

Note that the charged lepton kinematics for ggF Higgs boson production are expected to be similar to those of VBF Higgs boson production as described in section 1.3. However, unlike VBF Higgs boson production, ggF Higgs boson production is not expected to have a significant fraction of candidate events with jets in the forward regions of the detector. Consequently, the only process considered as signal in the BDT training is VBF Higgs boson production. Higgs boson production via gluon-gluon fusion is still included in the BDT training but as a background implying that the BDT attempts to reject it along with the other backgrounds. Although this choice slightly reduces the discrimination power of the charged lepton kinematics in the BDT, the goal of the analysis is to establish the existence of VBF Higgs boson production and therefore this process needs to be separated from ggF Higgs boson production. Higgs Strahlung (WH/ZH) is not included in the training since it is expected to be extremely small in the VBF signal region.

Most of the backgrounds mentioned in section 4.2 are included in the BDT training. The W+jets background is estimated using a data driven approach which includes candidate

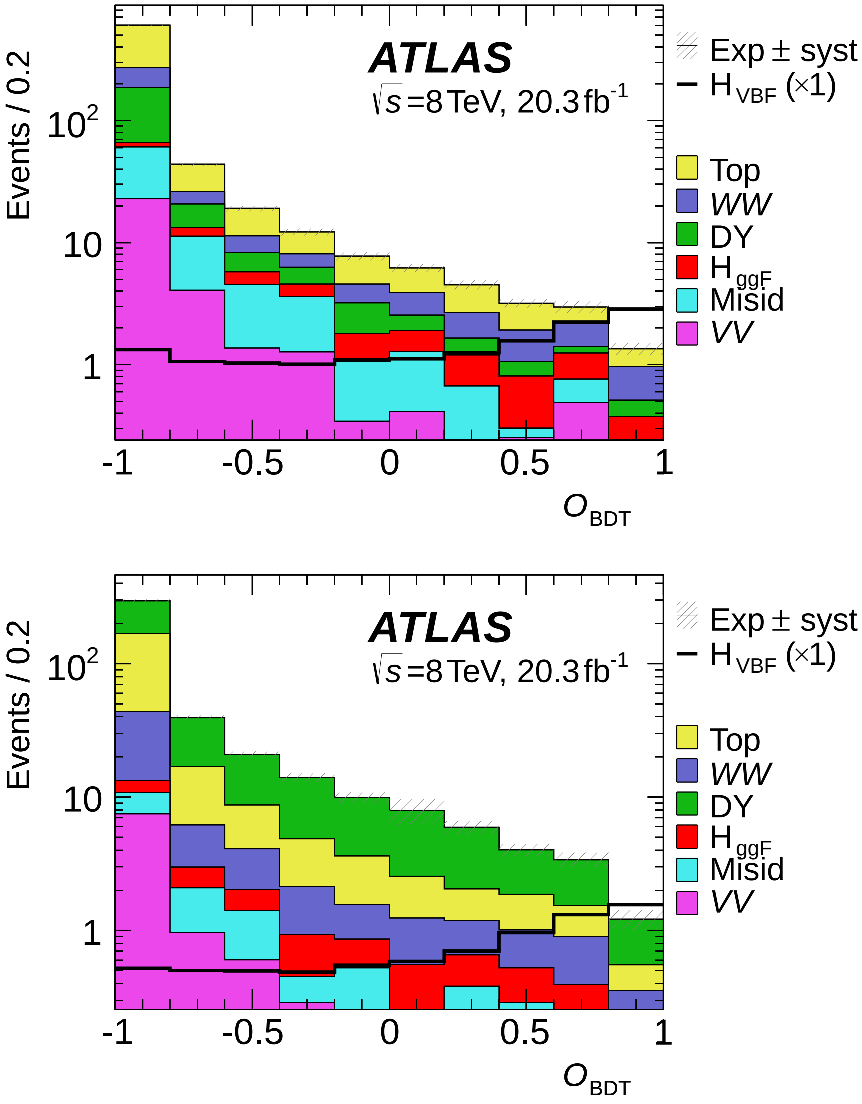


Figure 6.5: O_{BDT} distributions for different flavour (top) and same flavour (bottom) as implemented in the Run-I analysis [16]. The plots are shown at the selection stage immediately before the signal region is split into bins of BDT. The yields are normalized to the expected yields for 20.3 fb^{-1} of Run-I data. The yields for VBF Higgs boson production correspond to the SM predictions and are not scaled up as in previous figures. The shaded uncertainty bands shown on the background prediction include the systematic and statistical uncertainties.

events with negative weights (see section 7.5 for more details). Since negative event weights are not treated correctly when using the gradient boosting method, the W+jets estimate is not included in the BDT training. In Run-I, a single BDT is trained using DF and SF candidate events, while for Run-II, the BDT is trained using only DF candidate events. For the Run-II analysis, the contributions from $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ are not included as they are expected to be negligible in the DF analysis. The presence of the $Z \rightarrow ee/\mu\mu$ process makes the sample composition of the SF analysis quite different from that of the DF analysis. Consequently, a separate BDT for DF and SF was also explored in the context of the Run-I analysis so that each BDT is specifically trained for the region of phase space where it is being applied. However, this strategy also drastically reduced the available training statistics for the two designated BDTs and consequently the performance of the BDT did not improve significantly. In addition, this strategy also doubles the number of control regions (see chapter 7 for more details on the control regions) since separate control regions would be required for DF and SF. This in turn increases the statistical uncertainties in the control regions.

Since the available MC sample sizes are limited, the same MC events used to train the BDT are also used to estimate the expected signal and background yields in the signal region. If only a single BDT is trained, this would imply that the BDT is applied to the same events on which it is trained. This can potentially overestimate the expected performance of the analysis, particularly if the BDT is overtrained. Consequently, the BDT is cross-evaluated which means that two separate BDTs are trained on two statistically independent training samples, A and B. The BDT trained on A is applied to B, and the BDT trained on B is applied to A. In both cases, the *training sample* refers to the candidate events on which the BDT is trained, while the *test sample* refers to the events on which the BDT is applied.

In the absence of overtraining, the distribution of O_{BDT} should look similar when applied to the training and testing samples. The left and right plots in figure 6.6 show the Run-II O_{BDT} distributions, both applied to their training and testing samples. The comparison shows no strong indications of overtraining. This is also checked for the Run-I analysis, with similar results.

As described in chapter 4, there are a number of parameters in the BDT training that can be optimized. In both Run-I and Run-II, the optimal parameters were chosen using a grid scan while ensuring that the chosen configuration did not show strong evidence of overtraining. The various configurations were ranked using a similar metric of significance to that used for the $N-1$ minimal loss pruning described above. The parameters were scanned for both shallow trees with a maximum depth of 5 and deep trees with a maximum depth of 1,000. In general, large forests of shallow trees ("boosted stumps") performed better than a small number of deep trees. Since the number of configurations grows quickly as additional parameters are scanned, only a few values for each parameter is tried in the optimization. The parameters scanned along with the minimum and maximum values tested are:

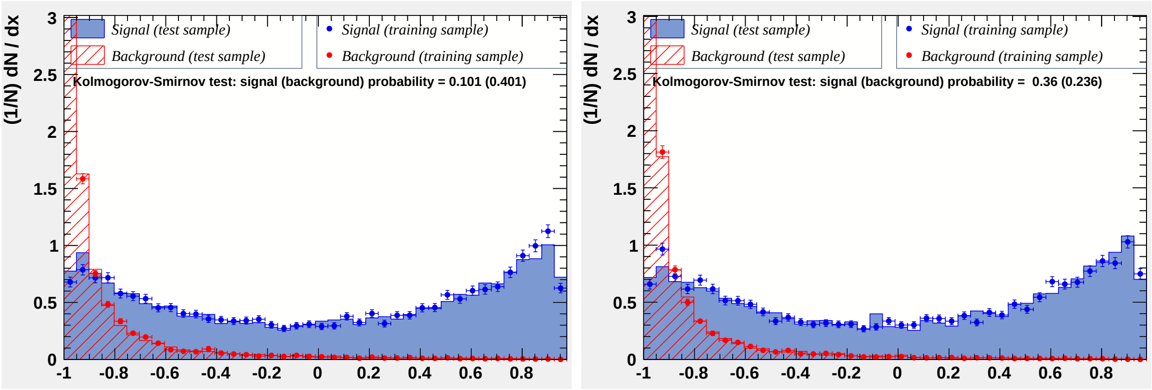


Figure 6.6: In order to check for the presence of overtraining in the Run-II BDT, the test samples (solid histograms) and training samples (points with uncertainty bars) are compared for signal (blue) and background (red). The left and right plots corresponds to using training samples A and B. The distributions show no strong indications of overtraining.

- the maximum tree depth (min: 5, max: 1,000)
- the number of trees in the forest (min: 275, max: 1,000)
- the minimum number of events in each node (min: 400, max: 10,000)
- the shrinkage or algorithm learning rate (min: 0.025, max: 0.15)
- the bagging fraction (min: 0.25, max: 0.60)

Figure 6.7 shows a 2D scan over the number of trees and the minimum number of events required per node. All other BDT training parameters are kept at fixed values during this scan. Even though the BDT performance improves as the number of events per node is decreased, this also increases the probability that the BDT is overtrained. In addition, as the size of the forest is increased, the performance of the BDT also improves.

The chosen parameters are summarized in table 6.5.

6.3 Selection

After the BDT training, further selection requirements are applied in order to define the signal region. Training the BDT before these selection requirements increases the sample size available for the training. Besides reducing the probability of overtraining, this also increases the separation power of the BDT.

Since the mediating weak vector bosons do not carry color charge, the level of hadronic activity between the two leading jets is expected to be small for signal candidate events. Any candidate event that has jets with $p_T > 20$ GeV inside the rapidity gap of the leading jets is removed, a requirement called the *central jet veto* (CJV). The CJV suppresses processes where extra jets are produced by QCD radiation. Mathematically, this selection requirement

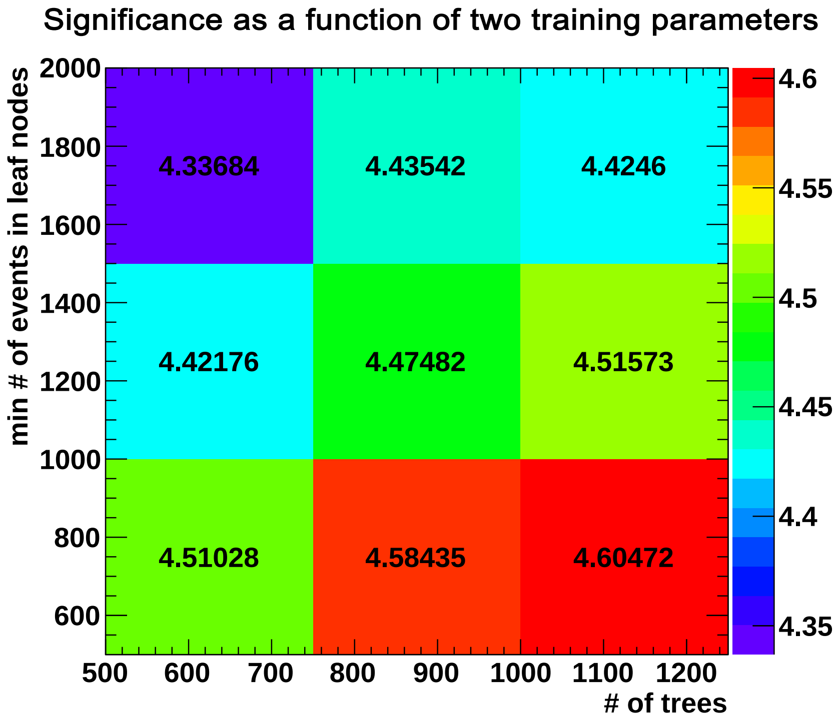


Figure 6.7: Results of a 2D significance scan over the number of trees and the minimum number of events required per node for trees with a maximum depth of 5. All other BDT training parameters are fixed in this scan.

Parameter	Run-I	Run-II
Boosting algorithm	Gradient	Gradient
Boosting loss function	$\ln(1+e^{-2F(\vec{x})y})$	$\ln(1+e^{-2F(\vec{x})y})$
Maximum tree depth	5	10
Number of trees	1000	1000
Minimum number of candidate events per node	1000	5000
Use bagging in gradient boosting	true	true
Bagging fraction	0.25	0.25
Shrinkage (algorithm learning rate)	0.125	0.15

Table 6.5: BDT training parameters used in Run-I and Run-II.

is defined in terms of the centrality of any additional jets besides the two leading jets. It is defined analogously to the lepton centrality:

$$C_{j3} = \left|\eta_{j3} - \frac{\eta_{j1} + \eta_{j2}}{2}\right| \Bigg/ \frac{|\eta_{j1} - \eta_{j2}|}{2} \tag{6.5}$$

The CJV requires that $C_{j3} > 0$ for all additional jets in the candidate event. Note that the jets that enter the calculation of C_{j3} are likely to be low p_T QCD jets originating from the parton showering in the MC. Since the distribution for C_{j3} is challenging to model with MC for low p_T QCD jets, this variable is not included in the BDT training.

In addition to including the sum of the lepton centralities in the BDT training, $C_\ell < 1$ is required for both charged leptons, a selection requirement called the *outside lepton veto* (OLV).

In order to reduce the $Z \to \tau\tau$ contamination, a *$Z \to \tau\tau$ veto* is also applied. It is defined in terms of $m_{\tau\tau}$, the invariant mass of the τ lepton pair system under the collinear approximation. The collinear approximation assumes that a given candidate event is produced by $Z \to \tau\tau$, that the decay products of the leptonically decaying τ's are collinear with the τ's, and that all the missing transverse energy in the candidate event is from the neutrinos coming from the two τ decays. The $Z \to \tau\tau$ veto requires $m_{\tau\tau} < m_Z - 25$ GeV. Figure 6.8 shows the DF $m_{\tau\tau}$ distribution before applying the $Z \to \tau\tau$ veto. The orange line corresponds to the upper bound applied on $m_{\tau\tau}$. This selection requirement also ensures that the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis is orthogonal to the ATLAS $H \to \tau\tau$ analysis which requires $m_{\tau\tau} > m_Z - 25$ GeV. This is especially important to prevent using the same candidate events in data multiple times when combining the various ATLAS Higgs boson analyses.

The final requirement for the signal region is a selection requirement on O_{BDT}. The region of phase space being removed by this selection requirement is used as a validation region to check the modelling of the BDT training variables and is called the *low O_{BDT} validation region.* However, it is not included in the statistical analysis described in section 9 since the kinematic distributions in this region are very different from those in the signal region. For example, as illustrated in figure 6.9, the $\Delta\phi_{\ell\ell}$ peaks at large values, which is not consistent with the signal events that are expected in the signal region. Since a separate BDT is trained for Run-I and Run-II, the selection requirement values are not the same for the two cases. For Run-I, the signal region is defined as $-0.48 < O_{\text{BDT}} < 1$, while for Run-II it is defined as $-0.8 < O_{\text{BDT}} < 1$.

Table 6.6 summarizes the analysis selection.

Since the VBF signal to background ratio is expected to increase with increasing O_{BDT}, dividing O_{BDT} into multiple bins improves the overall sensitivity of the analysis. The signal regions for both the Run-I and Run-II analyses are divided into multiple bins of O_{BDT}. The optimal bin boundaries are chosen by performing a scan over possible bin boundaries for

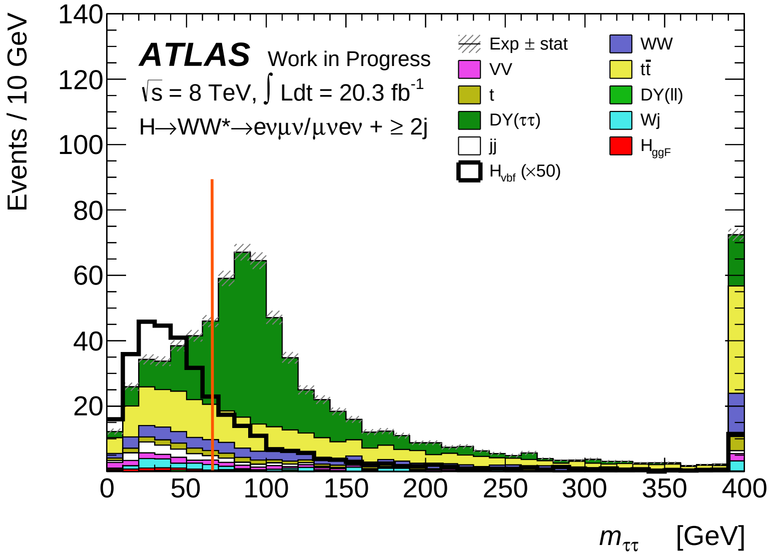


Figure 6.8: The different flavour $m_{\tau\tau}$ distribution after the selection up until the outside lepton veto, but not including the $Z \to \tau\tau$ veto. The $Z \to \tau\tau$ veto removes candidate events to the right of the orange line. The yields are normalized to the expected yields for 20.3 fb^{-1} of Run-I data and the VBF signal is scaled up by a factor of 50. The shaded uncertainty bands shown on the background prediction are statistical only.

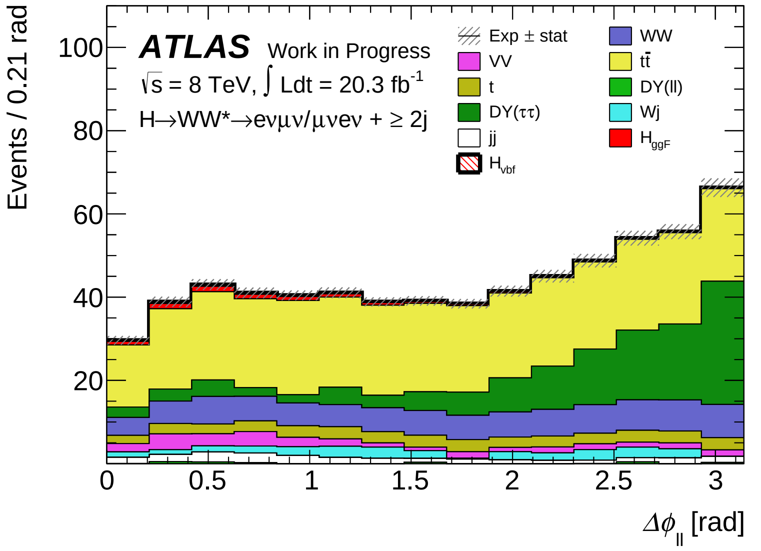


Figure 6.9: Distribution of $\Delta\phi_{\ell\ell}$ in the low O_{BDT} validation region for DF. The yields are normalized to the expected yield for 20.3 fb^{-1} of Run-I data. The shaded uncertainty bands shown on the background prediction only include the statistical uncertainty.

Selection Requirement	Definition
Central jet veto	$C_{j3} > 0$ for all additional jets
Outside lepton veto	$C_\ell < 1$ for both charged leptons
$Z \to \tau\tau$ veto	$m_{\tau\tau} < m_Z - 25$ GeV
BDT	$O_{\mathrm{BDT}} \in [-0.48, 1]/[-0.8, 1]$ (Run-I/Run-II)

Table 6.6: Summary of the selection applied after training the BDT.

all candidate events that pass the selection, excluding the selection requirement on O_{BDT}. In Run-I, an algorithm based on an estimate of Poisson significance is used:

$$Z_0 = \sqrt{2((S + B)\ln(1 + S/B) - S)} \tag{6.6}$$

where S is the signal yield, and B is the background yield. The algorithm scans the significance Z_0 for $x < O_{\mathrm{BDT}} \leq 1$, where x starts at 0.98 and increments down by 0.02. For values of x close to 1, Z_0 will be small due to the limited statistics. Since O_{BDT} for signal is expected to peak close to 1, Z_0 will increase as x decreases. However, since the signal to background ratio decreases as x decreases, Z_0 will reach a maximum value. This maximum value is chosen as the bin boundary. This process is repeated iteratively, each time starting the algorithm at the last chosen bin boundary. In the Run-I analysis, a total of three bins are defined. Adding additional bins does not significantly increase the significance since they would have very low signal to background ratios.

In Run-II, a more sophisticated bin optimization procedure is used. The distributions for O_{BDT} from MC for the signal and total background are parameterized with degree 5 and 3 polynomials respectively as illustrated in figure 6.10. Since the shape of O_{BDT} is challenging to parameterize when O_{BDT} is close to 1, this region is approximated using the average bin content. This approximation is not expected to degrade the optimization since the expected yields in this region are small, making it illogical to split it into more than one bin.

In order to calculate the maximum possible significance, the parameterized O_{BDT} distributions are used in an unbinned likelihood maximization. Even though this method does not take the systematic uncertainties described in chapter 8 into account, it does provide a reasonable approximation of the best possible performance of the analysis in the limit of an infinite number of bins with infinite MC statistics in each bin. In practice, a finite number of bins must be chosen to ensure the statistical uncertainty in each bin is reasonable. The optimal significance is calculated to be $Z_0 = 1.45$.

Next, using the parameterized O_{BDT} distributions, a scan is done for the optimal bin boundary assuming 2 bins. The lower bound of the lower bin is fixed at $O_{\mathrm{BDT}} = 0.0$, and

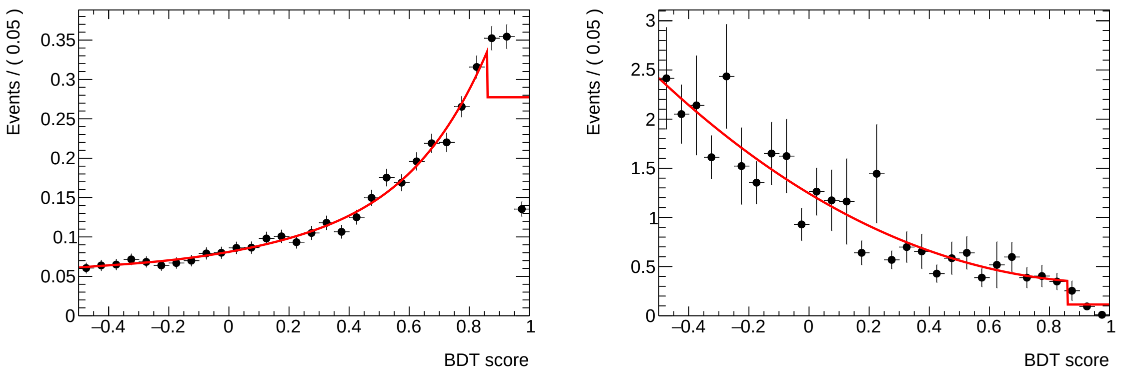


Figure 6.10: The Run-II binning optimization requires a parameterization of the signal and total background O_{BDT} distributions. The signal O_{BDT} distribution (left) is approximated with a degree 5 polynomial while the background distribution (right) is approximated with a degree 3 polynomial. The region $O_{\text{BDT}} > 0.85$ is approximated by the average bin content in that region.

the upper bound for the upper bin is fixed at $O_{\text{BDT}} = 1.0$. As illustrated in figure 6.11, the optimal bin boundary of 0.7 recovers 86% of the unbinned significance.

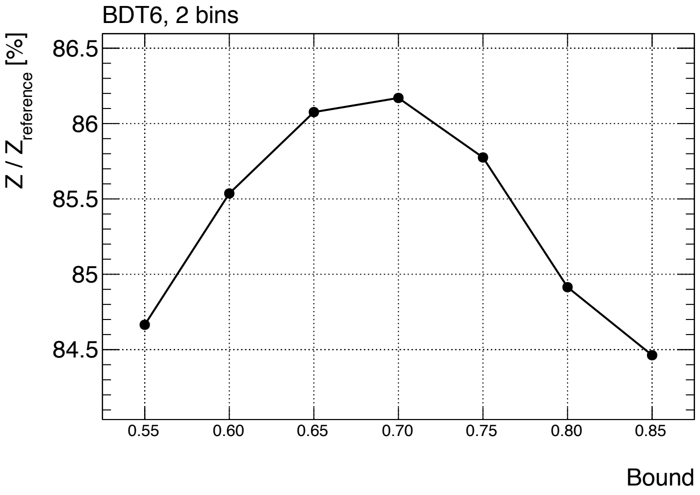


Figure 6.11: O_{BDT} binning optimization for 2 bins where the lowest and highest O_{BDT} bounds are fixed at 0.0 and 1.0 respectively. The y-axis corresponds to the significance relative to the unbinned fit.

An additional scan is done assuming three bins of O_{BDT}. As before, the lowest and highest O_{BDT} bins are fixed at $O_{\text{BDT}} = 0.0$ and $O_{\text{BDT}} = 1.0$ respectively. As shown in figure 6.12, the optimal O_{BDT} boundaries in this case are 0.5 and 0.85 which recovers 96% of the unbinned significance.

Initially, the 3-bin scenario was chosen as baseline given its superior performance. However, during the calculation of the detector and theory systematic uncertainties described

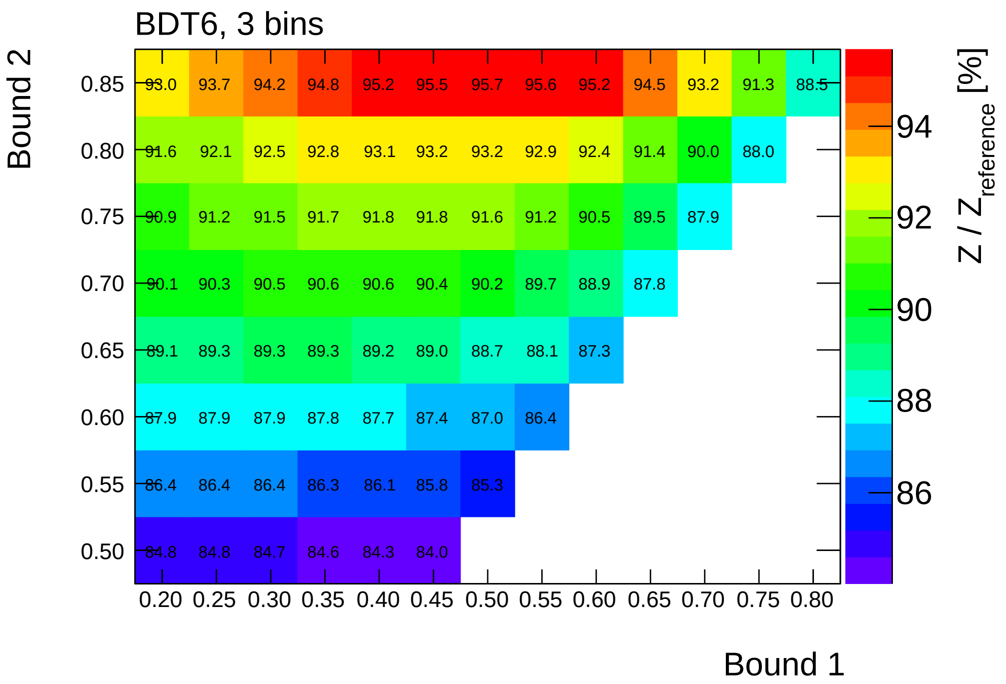


Figure 6.12: O_{BDT} binning optimization for 3 bins where the lowest and highest O_{BDT} bounds are fixed at 0.0 and 1.0 respectively. The z-axis corresponds to the significance relative to the unbinned fit.

in section 8, insufficient MC sample size for the alternate samples in the highest O_{BDT} bin was a reoccurring issue. Consequently, the 2-bin setup was chosen to reduce the statistical component in the systematic uncertainties. The lower bound of the lower bin was chosen to be -0.80 in order to ensure sufficiently large sample sizes in the control regions described in chapter 7. This does not degrade the performance of the analysis since this bin is dominated by background.

Note from figure 6.12 that the analysis can be improved by a few percent by splitting the lower O_{BDT} bin into two bins. By keeping the bin boundary of 0.70, an additional bin boundary can be defined at 0.35 which will increase the significance from 86% to just over 90% of the unbinned significance. Unfortunately, this improvement was not implemented due to time constraints.

Table 6.7 summarizes the bins chosen for the Run-I and Run-II analyses. As more Run-II data becomes available and the MC sample sizes are increased, the high O_{BDT} region will become more and more populated, allowing the signal region to be divided into a larger number of more sensitive bins.

A summary of the expected event yields after applying all selection requirements is given in table 6.8 for the Run-I analysis and table 6.9 for the Run-II analysis. As additional selection requirements are applied, the ratio of signal to background increases. The uncertainties shown are statistical only.

Analysis	Number of bins	O_{BDT} bin boundaries
Run-I	3	$[-0.48, 0.3, 0.78, 1]$
Run-II	2	$[-0.8, 0.7, 1]$

Table 6.7: O_{BDT} bin boundaries for the signal region in the Run-I and Run-II analyses.

Run-I DF ($e\mu/\mu e$)

Selection Requirement	Total Signal	Total Background	Data
$n_{jet} = 2$	32.0 ± 0.2	$59{,}020 \pm 40$	61,434
$n_b = 0$	25.6 ± 0.2	$7{,}560 \pm 20$	7,818
Central jet veto	21.6 ± 0.2	$6{,}090 \pm 20$	6,313
Outside lepton veto	16.8 ± 0.2	$1{,}271 \pm 9$	1,264
$Z \to \tau\tau$ veto	14.6 ± 0.2	703 ± 6	718
Signal Region ($O_{\text{BDT}} > -0.48$)	11.5 ± 0.1	44 ± 1	57

Run-I SF ($ee/\mu\mu$)

Selection Requirement	Total Signal	Total Background	Data
$n_{jet} = 2$	20.3 ± 0.2	$26{,}140 \pm 80$	27,637
$n_b = 0$	16.1 ± 0.2	$8{,}030 \pm 60$	8,249
Central jet veto	12.1 ± 0.2	$2{,}510 \pm 30$	2,653
Outside lepton veto	9.4 ± 0.1	570 ± 10	664
$Z \to \tau\tau$ veto	7.7 ± 0.1	403 ± 8	469
Signal Region ($O_{\text{BDT}} > -0.48$)	6.4 ± 0.1	54 ± 2	73

Table 6.8: Expected and observed event yields at various event selection stages in the Run-I analysis. The uncertainties shown are statistical only.

Run-II DF ($e\mu/\mu e$)

Selection Requirement	Total Signal	Total Background	Data
$p_T^{\ell 1} > 22/25$ GeV, $p_T^{\ell 2} > 15$ GeV	38.3 ± 0.4	$185{,}600 \pm 700$	-
$m_{\ell\ell} > 10$ GeV	36.5 ± 0.4	$166{,}200 \pm 700$	-
$n_{jet} = 2$	20.4 ± 0.3	$50{,}490 \pm 90$	53,455
$n_b = 0$	17.8 ± 0.3	$6{,}160 \pm 60$	6,301
Central jet veto	13.9 ± 0.2	$4{,}590 \pm 50$	4,642
Outside lepton veto	10.6 ± 0.2	990 ± 30	1,006
$Z \rightarrow \tau\tau$ veto	9.2 ± 0.2	530 ± 10	576
Signal Region ($O_{\mathrm{BDT}} > -0.8$)	8.7 ± 0.2	131 ± 8	129

Table 6.9: Expected and observed event yields at various event selection stages in Run-II analysis. The uncertainties shown are statistical only.

Chapter 7

Background Estimation

In order to be able to measure the signal strength for VBF Higgs boson production, it is important to have a thorough understanding of the backgrounds present in the signal region. However, the signal region selection requirements define a region of phase space for which many MC generators are not optimized. In particular, $t\bar{t}$, the dominant background, is suppressed with a b-veto. Consequently, the majority of $t\bar{t}$ candidate events in the signal region consist of a b-jet that is not identified by the b-tagging algorithm and a jet that does not originate from a top quark decay. Since generators are being used in regions of phase space for which they were not originally designed, data-driven corrections are applied to the dominant backgrounds whenever possible and kinematic distributions are checked in data validation and control regions.

Even though the W+jets background is relatively small in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis, it is also estimated using a data-driven approach. This is because it is challenging to accurately simulate the misidentification of jets as leptons in the MC.

7.1 Data-driven Normalization of Signal Region Yields

The simplest data-driven background estimate in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis uses a control region (CR) to correct the yields for a specific process in the signal region. This method is used for the top quark and $Z \to \tau\tau$ backgrounds to ensure that these backgrounds are correctly estimated in the VBF region of phase space.

First, a control region is defined with the goal to select a sample that is enriched in the background of interest and orthogonal to the signal region. The observed yield in the control region, $N_{\mathrm{CR}}^{\mathrm{data}}$, is calculated by subtracting all signals and backgrounds besides the background of interest from the data yield. $N_{\mathrm{CR}}^{\mathrm{MC}}$ and $N_{\mathrm{SR}}^{\mathrm{MC}}$ are defined as the yields from MC for the background of interest in the control and signal region respectively. Next, a normalization factor, β, is calculated by taking the ratio of $N_{\mathrm{CR}}^{\mathrm{data}}$ to $N_{\mathrm{CR}}^{\mathrm{MC}}$. The estimated yield in the signal region, $N_{\mathrm{SR}}^{\mathrm{est}}$, is defined as:

$$N_{\mathrm{SR}}^{\mathrm{est}} = N_{\mathrm{SR}}^{\mathrm{MC}} \underbrace{\left(\frac{N_{\mathrm{CR}}^{\mathrm{data}}}{N_{\mathrm{CR}}^{\mathrm{MC}}} \right)}_{\beta} \tag{7.1}$$

In equation 7.1, $N_{\mathrm{SR}}^{\mathrm{est}}$ is expressed in terms of a data-to-MC normalization in the control region (β), but it is also possible to express it in terms of a MC based extrapolation factor from the control region to the signal region (α):

$$N_{\mathrm{SR}}^{\mathrm{est}} = N_{\mathrm{CR}}^{\mathrm{data}} \underbrace{\left(\frac{N_{\mathrm{SR}}^{\mathrm{MC}}}{N_{\mathrm{CR}}^{\mathrm{MC}}} \right)}_{\alpha} \tag{7.2}$$

In restricted regions of phase space such as the one defined by the VBF analysis selection, a MC-only estimate of a process is often accompanied by large theoretical uncertainties. However, when using an extrapolation factor α as defined in equation 7.2, the uncertainty can be significantly reduced. Assuming data sample sizes in the control region are sufficiently large, the statistical uncertainty on $N_{\mathrm{CR}}^{\mathrm{data}}$ will be small. In addition, for a control region that is sufficiently pure in the process of interest, the systematic uncertainty on the MC subtraction term in $N_{\mathrm{CR}}^{\mathrm{data}}$ will also be small. Note that if a systematic variation changes $N_{\mathrm{SR}}^{\mathrm{MC}}$ and $N_{\mathrm{CR}}^{\mathrm{MC}}$ in the same direction, then the effect of that systematic uncertainty cancels in α. For a pure, well-populated control region, the combination of the uncertainty on α and $N_{\mathrm{CR}}^{\mathrm{data}}$ will be significantly smaller than that of the MC only prediction.

7.2 Top Control Region

Despite the b-veto that is applied, there is still a large contribution from top quark background candidate events in the signal region. This is not surprising given that at 13 TeV, the cross section for VBF $H \to WW^* \to \ell\nu\ell\nu$ production is three orders of magnitude smaller than that of dileptonically decaying $t\bar{t}$ production.

In general there is a high level of agreement between theory calculations and ATLAS measurements of top quark production [87, 88]. However, as previously mentioned, a top control region is still well motivated given the unique corner of phase space corresponding to the VBF $H \to WW^* \to \ell\nu\ell\nu$ signal region. Most of the top quark candidate events in the signal region consist of a b-quark that is not identified by the b-tagging algorithm along with a jet from initial state radiation. In addition, these candidate events are expected to have forward jets. Consequently, they will have a different topology than the ones used for the above cited studies where the jets are required to be b-tagged and central.

In order to define a top control region with a similar topology to the signal region, all preselection and selection described in chapter 6 is applied except that the b-veto is replaced by requiring there to be exactly one b-jet. At first glance, it would appear more logical to define a pure top control region by requiring there to be exactly two b-jets. However, doing so also moves the topology of the top control region further away from the signal region. Since b-tagging is only available for $|\eta| < 2.5$, requiring two b-jets forces at least two jets to be central, while signal-like candidate events are expected to have forward jets.

In the case of the Run-I analysis, the top control region includes both DF and SF candidate events to minimize the statistical uncertainty. It is also divided into O_{BDT} bins like the signal region but since the sample sizes in the most signal-like bins of the top control region are relatively small, a single normalization factor is calculated for the combination of O_{BDT} bins 2 and 3. In Run-II, only DF candidate events are included in the top control region and it is not binned in O_{BDT}. Consequently, only a single normalization factor is calculated for the entire control region.

Recall from section 6.3 that the low O_{BDT} validation region is not included in the statistical analysis of the data. This also implies that the corresponding low O_{BDT} region for the top control region is not included in the statistical analysis. Even though this low O_{BDT} region has events with one b-tag, they are very different kinematically compared to the events in the top control region for similar reasons as explained for the signal region in section 6.3.

The top normalization factors calculated for Run-I and Run-II are summarized in table 7.1. The Run-I normalization factor is not consistent with unity, but this difference is covered when also considering the systematic uncertainties described in section 8.1.1. Figure 7.1 shows the O_{BDT} distribution in the top control region with the statistical and systematic uncertainties. This mismodelling is not evident in the Run-II analysis even though the same generator is used.

Run-I		Run-II
Bin 1	Bins 2-3	Bins 1-2
1.6 ± 0.2	1.0 ± 0.3	1.0 ± 0.1

Table 7.1: Summary of the top normalization factors (values for β) calculated in the Run-I and Run-II analyses. The uncertainties are statistical only.

The top control region is dominated by dileptonically decaying $t\bar{t}$ but also includes contributions from single top quark production. In order to validate the MC prediction of the top quark background, the predicted BDT training variable distributions are plotted with the data in the top control region. The Run-I DF and SF distributions are shown in figures 7.2 and 7.3, while the Run-II DF distributions are shown in 7.4. These plots do not include the normalization factors described in this section.

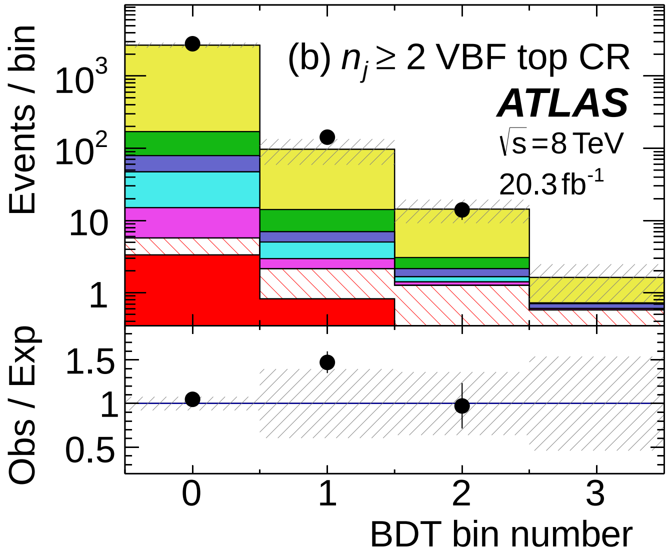


Figure 7.1: The O_{BDT} distribution is shown in the Run-I top control region after the top control region selection requirements up until the $Z \to \tau\tau$ veto [16]. No normalization factors are applied in this plot. The uncertainty band shown includes systematic and statistical uncertainties.

7.3 $Z \to \tau\tau$ Control Region

The $Z \to \tau\tau$ control region is defined based on all preselection and selection described in chapter 6. However, in order to specifically select $Z \to \tau\tau$ candidate events, the $Z \to \tau\tau$ veto is replaced by $|m_{\tau\tau} - m_Z| < 25$ GeV. In addition, $m_{\ell\ell}$ is required to be less than 80 GeV in order to suppress contamination from the top background.

Due to the small number of events in this control region, a common normalization factor is calculated for the entire control region. In Run-I, the control region includes both SF and DF candidate events, while in Run-II it only includes DF candidate events. As shown in table 7.2, the normalization factors derived for the Run-I and Run-II analyses are consistent with unity.

Run-I: Bins 1-3	Run-II: Bins 1-2
0.9 ± 0.3	0.9 ± 0.2

Table 7.2: Summary of the $Z \to \tau\tau$ normalization factors calculated in the Run-I and Run-II analyses. The uncertainties shown are statistical only.

As described in more detail in section 9.1, the $Z \to \tau\tau$ control region is treated differently in the Run-I and Run-II analyses. In Run-I, only the predicted yield from this control region is included in the statistical analysis of the data, while in Run-II the entire control region is included.

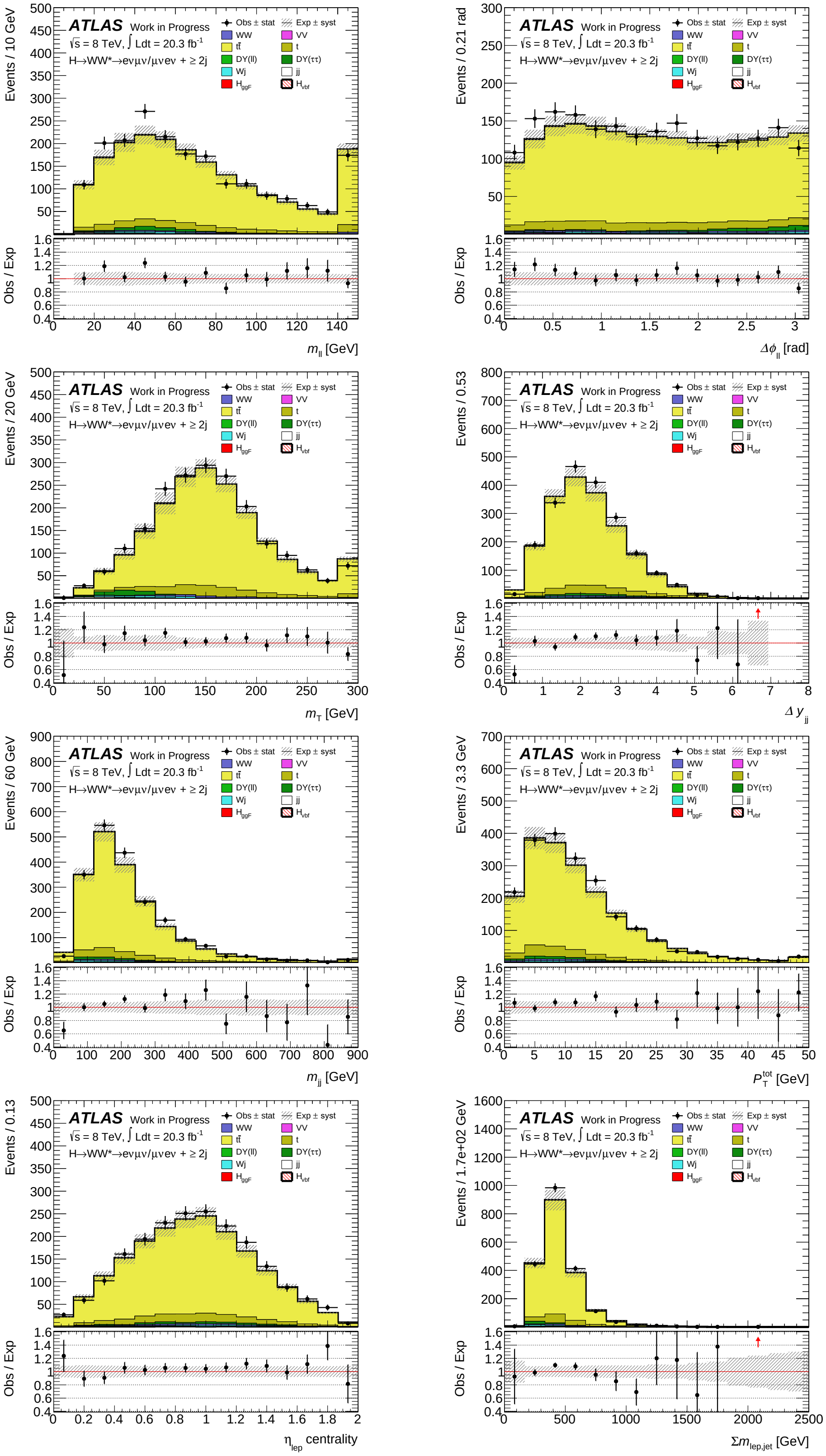


Figure 7.2: Different flavour BDT training variables shown in the Run-I top control region after the $Z \to \tau\tau$ veto. No normalization factors are applied in these plots. The uncertainty band shown includes statistical and systematic uncertainties.

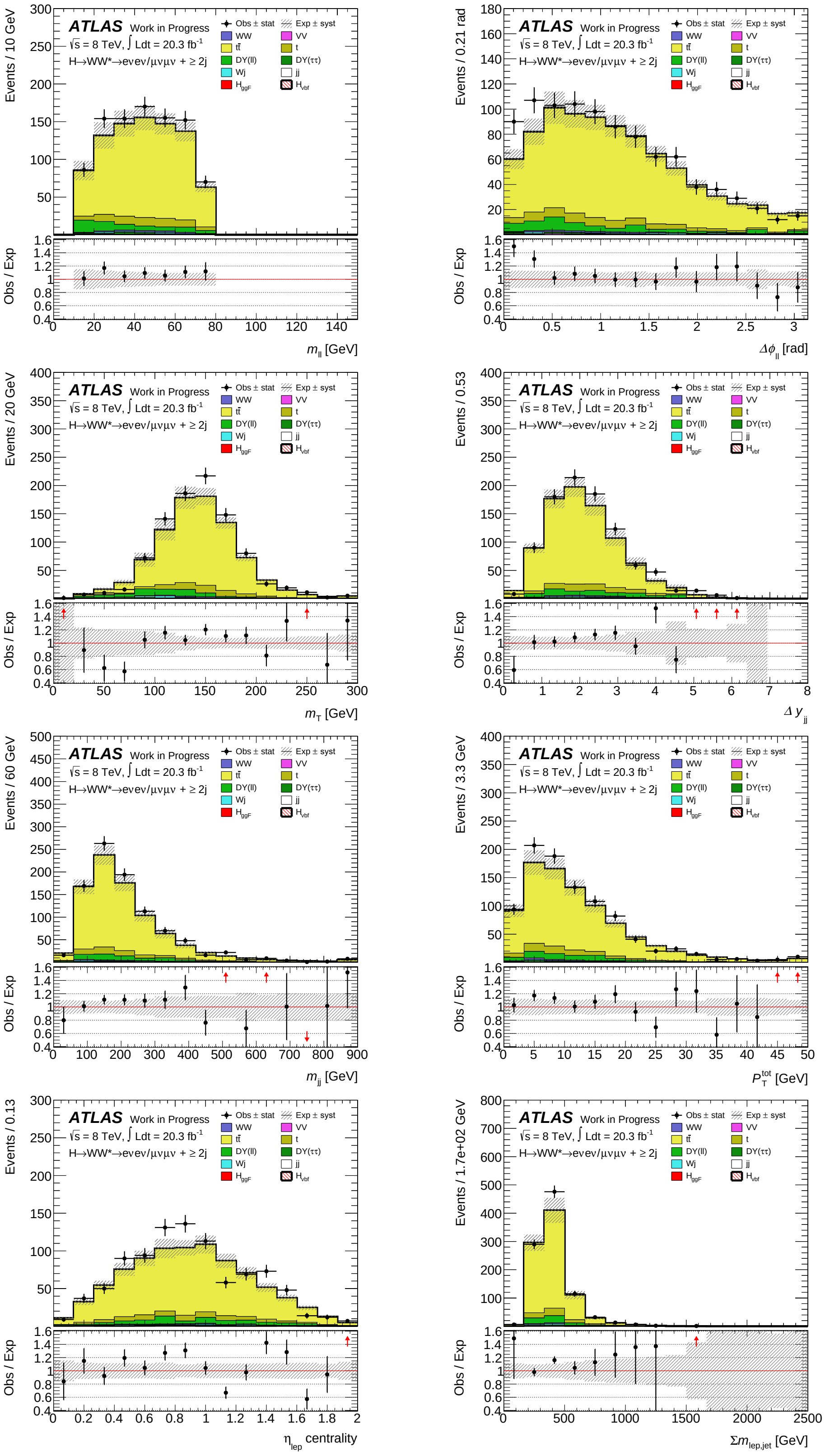


Figure 7.3: Same flavour BDT training variables shown in the Run-I top control region after the $Z \to \tau\tau$ veto. No normalization factors are applied in these plots. The uncertainty band shown includes statistical and systematic uncertainties.

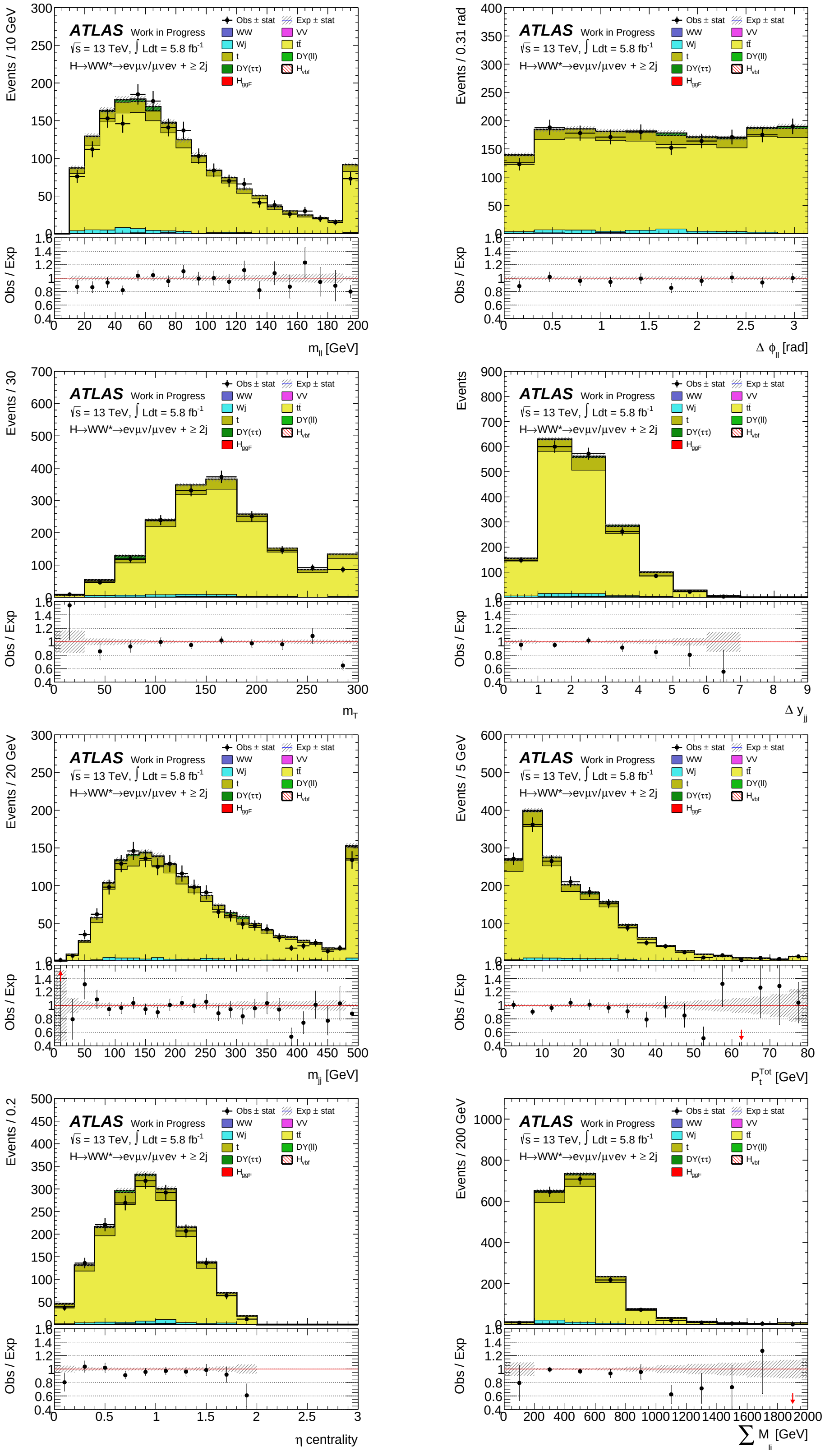


Figure 7.4: Different flavour BDT training variables shown in the Run-II top control region after the $Z \to \tau\tau$ veto. No normalization factors are applied in these plots. The uncertainty band shown only includes the statistical uncertainty.

7.4 $Z \to ee/\mu\mu$ Control Region

$Z \to ee/\mu\mu$ is the dominant background in the SF VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis and consequently it is only included in the Run-I analysis. Due to the absence of neutrinos, this process is not expected to have missing transverse energy and it is significantly reduced by applying selection requirements on E_T^{miss} and p_T^{miss}. However, due to the high cross section for $Z \to ee/\mu\mu$ there is still a significant contribution in the signal region. The largest contribution of these momentum imbalances arise from the mismeasurement of jets in the event. Furthermore, since $|m_Z - m_{\ell\ell}| > 15$ GeV is required, the $Z \to ee/\mu\mu$ events that remain in the signal region are in the tails of the $m_{\ell\ell}$ distribution, a region of phase space the MC is not optimized for. In order to account for these effects, a data driven approach is used.

The $Z \to ee/\mu\mu$ background is estimated using an "ABCD data-driven method" that makes use of four orthogonal regions of phase space. The first, region A, corresponds to the signal region where the estimate is applied. As described in section 6.3, this region includes the selection requirements $E_T^{miss} > 45$ GeV and $m_{\ell\ell} < 75$ GeV. Region B is enriched in $Z \to ee/\mu\mu$ by replacing $E_T^{miss} > 45$ GeV with 25 GeV $< E_T^{miss} <$ 45 GeV.

A 15 GeV window in $m_{\ell\ell}$ around the Z-boson peak is used to measure the relative efficiencies of the E_T^{miss} selection requirements in data. It is split into regions C and D that are defined analogously to regions A and B, except that $m_{\ell\ell} < 75$ GeV is replaced with $|m_{\ell\ell} - m_Z| < 15$ GeV. Table 7.3 summarizes the 4 regions used in the ABCD estimate.

	low $m_{\ell\ell}$	Z-boson peak
high E_T^{miss}	**Region A - SR** $E_T^{miss} > 45$ GeV $m_{\ell\ell} < 75$ GeV	**Region C** $E_T^{miss} > 45$ GeV $\lvert m_{\ell\ell} - m_Z\rvert < 15$ GeV
low E_T^{miss}	**Region B** 25 GeV $< E_T^{miss} <$ 45 GeV $m_{\ell\ell} < 75$ GeV	**Region D** 25 GeV $< E_T^{miss} <$ 45 GeV $\lvert m_{\ell\ell} - m_Z\rvert < 15$ GeV

Table 7.3: Summary of the 4 orthogonal regions in the ABCD method used to estimate the $Z \to ee/\mu\mu$ background in the Run-I SF analysis.

The yield in region A in the ith O_{BDT} bin is then estimated as:

$$N_{\mathrm{A},i}^{\mathrm{est}} = \frac{N_{\mathrm{C}}^{\mathrm{data}}}{N_{\mathrm{D}}^{\mathrm{data}}} N_{\mathrm{B},i}^{\mathrm{data}} \tag{7.3}$$

where $N_{\mathrm{X}}^{\mathrm{data}}$ is the data yield in region X after subtracting all non-$Z \to ee/\mu\mu$ contributions using MC. The method relies on two assumptions that are further elaborated on in section 8.1.2: O_{BDT} must not be correlated to E_T^{miss} and $m_{\ell\ell}$ must not be correlated with

E_T^{miss}. In order to account for the small correlation between $m_{\ell\ell}$ and E_T^{miss}, a MC driven correction factor called the non-closure is applied to equation 7.3:

$$f_{\text{non-closure}} = \frac{N_A^{\text{MC}}/N_B^{\text{MC}}}{N_C^{\text{MC}}/N_D^{\text{MC}}} \tag{7.4}$$

where N_X^{MC} are the $Z \to ee/\mu\mu$ MC yields for region X.

Regions A and B are split into bins of O_{BDT}, with bins 2 and 3 merged into a single bin because of limited statistics. The relative rates in bins 2 and 3 of O_{BDT} are taken from MC. Regions C and B are not binned in O_{BDT}. The value of $N_{\text{C}}^{\text{data}}/N_{\text{D}}^{\text{data}}$ is calculated to be 0.43 ± 0.03. The data-driven estimates are consistent with the MC predictions. The normalization factor calculated for O_{BDT} bin 1 is 1.0 ± 0.2, while the normalization factor for bins 2 and 3 is 0.9 ± 0.3, where the uncertainties are statistical only.

7.5 W+jets Background Estimate

As mentioned in section 4.2, W+jet and multi-jet backgrounds make their way into the signal region when one or two jets are misidentified as leptons. These backgrounds are commonly referred to as “fake backgrounds” since one or both of the reconstructed leptons do not originate from prompt leptons. Since the misidentification of jets in challenging to model well in MC, both the shape of kinematic distributions and their normalizations are derived from data. A data-driven estimate was developed for the ggF $H \to WW^* \to \ell\nu\ell\nu$ analysis where fake backgrounds are significant. The same estimation technique is used for the VBF analysis. Since the fake backgrounds tend to be more central in rapidity, they are significantly smaller in the VBF analysis.

The W+jets control region is defined identically as the signal region, except that one of the leptons is required to fail the identification and isolation criteria, but meet less stringent requirements. These leptons are referred to as “anti-identified” leptons. The dominant contribution in this control region comes from W+jet events where a jet produces an object that is reconstructed as a lepton.

In Run-I, the W+jets process is estimated in this control region from data by subtracting the contribution from the multi-jet background which is also estimated from data as described below. Furthermore, the small non-fake background is estimated using MC and then also subtracted. In Run-II, the multi-jet background is not treated separately and so the W+jets process is estimated from data by only subtracting the non-fake background using MC.

Next, a “fake factor” from the W+jets control region to the signal region is defined as:

$$f_{\text{fake}} = \frac{N_{\text{ID}}}{N_{\text{anti-ID}}} \tag{7.5}$$

where N_{ID} is the number of jets that satisfy the standard lepton requirements, while $N_{\text{anti-ID}}$ is the number of jets that pass the anti-identified requirements. The fake factor is calculated in bins of p_T and η using jets from a jet-enriched Z control region in Run-I and a dijet control region in Run-II.

For Run-I, the W+jets estimate in the signal region is then written as:

$$N_{\text{SR}}^{W\text{+jets,est}} = f_{\text{fake}} \cdot (N_{\text{1-fake-CR}}^{\text{data}} - N_{\text{1-fake-CR}}^{\text{multi-jet}} - N_{\text{1-fake-CR}}^{\text{non-fake,MC}}) \tag{7.6}$$

where $N_{\text{1-fake-CR}}^{\text{data}}$ is the total data yield in the W+jets control region, $N_{\text{1-fake-CR}}^{\text{multi-jet}}$ is a data-driven estimate of the multi-jet background described below, and $N_{\text{1-fake-CR}}^{\text{non-fake,MC}}$ is the MC prediction for the non-fake background in the W+jets control region. The analogous equation for Run-II does not include the $N_{\text{1-fake-CR}}^{\text{multi-jet}}$ term.

As already mentioned, a separate control region is defined in the Run-I analysis for the multi-jet background. It is defined like the W+jets control region except that both leptons are required to satisfy the criteria for anti-identified leptons. The corresponding fake factor, f'_{fake} is derived from a region that is enriched with multi-jet candidate events. The multi-jet estimate is then defined as:

$$N_{\text{SR}}^{\text{multi-jet,est}} = f'_{\text{fake}} \cdot (N_{\text{2-fake-CR}}^{\text{data}} - N_{\text{2-fake-CR}}^{W\text{+jet,MC}} - N_{\text{2-fake-CR}}^{\text{non-fake,MC}}) \tag{7.7}$$

where $N_{\text{2-fake-CR}}^{\text{data}}$ is the total data yield in the multi-jet control region, $N_{\text{2-fake-CR}}^{W\text{+jet,MC}}$ is the contribution from W+jets estimated using MC, and $N_{\text{2-fake-CR}}^{\text{non-fake,MC}}$ is the MC prediction for the non-fake background in the multi-jet control region. A key feature of this data-driven estimate is that the MC subtraction terms are small.

In order to calculate $N_{\text{1-fake-CR}}^{\text{multi-jet}}$ in equation 7.6, an additional fake factor from the multi-jet control region to the W+jet control region is calculated:

$$N_{\text{1-fake-CR}}^{\text{multi-jet}} = f''_{\text{fake}} \cdot (N_{\text{2-fake-CR}}^{\text{data}} - N_{\text{2-fake-CR}}^{W\text{+jet,MC}} - N_{\text{2-fake-CR}}^{\text{non-fake,MC}}) \tag{7.8}$$

7.6 Validation Regions

Besides control regions, *validation regions* are also defined. Even though these regions are not used directly in the analysis, they are used to check the modelling of kinematic variables.

Recall that the lowest O_{BDT} bin in both the Run-I and Run-II analyses is not included in the signal region. This region is well populated and is used to check the modelling of the correlations between the BDT training variables. This region is called the low O_{BDT} validation region. The correlation modelling for Run-I is shown in figure 7.5 and the correlation modelling for Run-II is shown in 7.6. The uncertainties shown are statistical only. Overall, the correlations between training variables observed in the data are consistent with those in the MC.

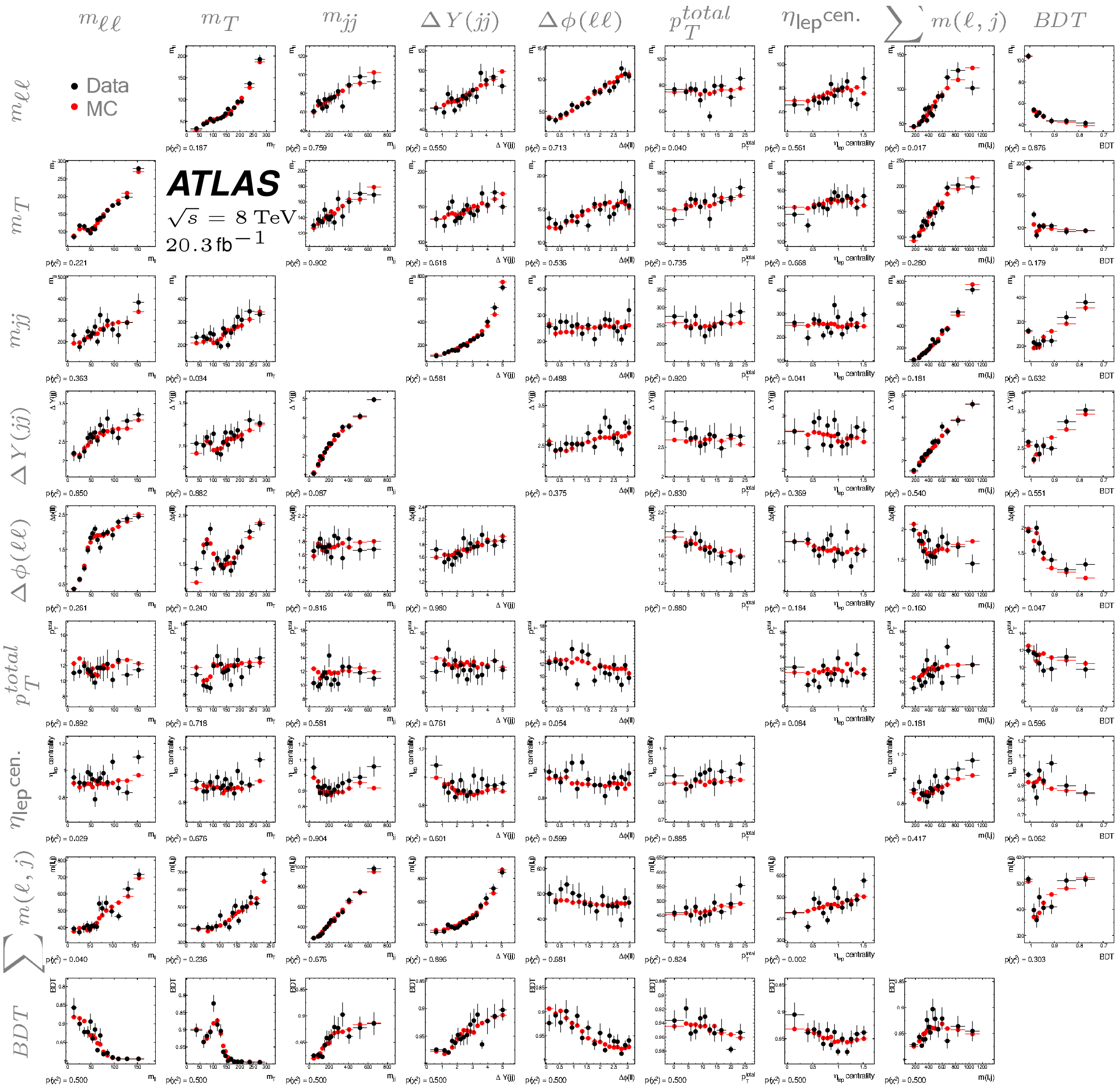


Figure 7.5: Modelling of correlations between the BDT training variables in the Run-I low O_{BDT} validation region [16]. The black points represent the data profiles, while the red points represent the MC profiles.

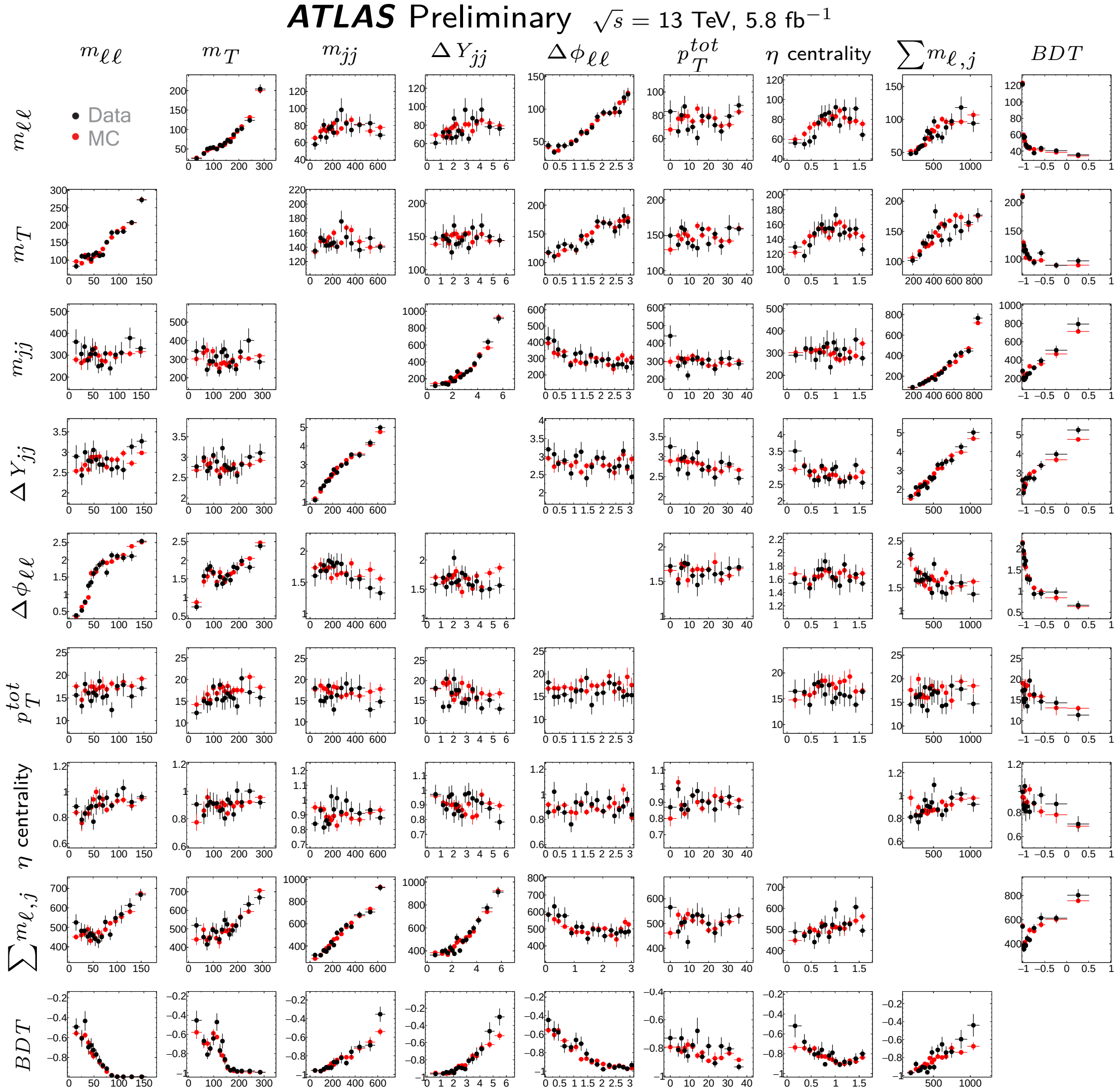


Figure 7.6: Modelling of correlations between the BDT training variables in the Run-II low O_{BDT} validation region [52]. The black points represent the data profiles, while the red points represent the MC profiles.

Chapter 8

Systematic Uncertainties

The models of physics and the experiment used are imperfect. The uncertainties in these models are referred to as systematic uncertainties. For any MC generator, there are a number of possible sources of systematic uncertainty. These uncertainties need to be propagated to the event yields in the signal and control regions so that they can be correctly accounted for in the statistical analysis of the data. The uncertainties are categorized as theoretical or experimental uncertainties depending on their origin. In general, theoretical uncertainties are those related to choices made in the generation of the MC samples. For example, the programs chosen to simulate the hard scatter and the parton showering both have corresponding theoretical uncertainties. Experimental uncertainties originate from detector measurements that are uncertain as well as the uncertainty in the techniques used to analyze the data collected by the detector. For example, detector calibrations like the energy and momentum scale, the trigger efficiencies, and the efficiencies of selection requirements all have corresponding experimental uncertainties.

Each source of systematic uncertainty can be broken down into a normalization uncertainty and a shape uncertainty. Normalization uncertainties change the overall event yield and so have the same relative impact in each bin of O_{BDT}. For example, the uncertainty on the total luminosity is a normalization uncertainty. In contrast, shape uncertainties impact different regions of phase space differently since their sources change the kinematic distributions used in the analysis. For instance, the resolution of the measured jet energy will be narrower in the central regions of the detector than in the forward regions. Some systematic uncertainties will only affect one of the two, but in many cases, both components are important. While normalization uncertainties are quoted as single numbers, shape uncertainties are quoted as three or two uncertainties corresponding to the O_{BDT} bins in the Run-I and Run-II analyses respectively.

In the interest of conciseness, the language used to describe systematic uncertainties is somewhat imprecise. For instance, one of the theoretical sources of uncertainty considered is the choice of PDF and so is colloquially called the PDF uncertainty. However, this uncertainty does not refer to the systematic uncertainty on the PDF being used but rather probes the sensitivity of the measurement to the choice of PDF.

8.1 Theoretical Uncertainties in Run-I

The theoretical sources of uncertainty that are considered include the parton shower and underlying event model (PS/UE), the choice of PDF, the QCD scale, and the matrix element model. Additional sources of uncertainty are considered for individual processes when necessary. The uncertainties for subdominant backgrounds are not discussed in this chapter as they have a negligible impact on the analysis result.

8.1.1 Top Quark Uncertainties in Run-I

The PS/UE uncertainty is evaluated by taking the difference between the default $t\bar{t}$ sample showered with PYTHIA6 with one showered with HERWIG, but using the same generator for the hard scatter. These two programs use different methods to simulate the hadronization process. HERWIG uses the cluster model [63], while PYTHIA uses the Lund string model for hadronization [61]. The impact on the extrapolation factor α in each of the O_{BDT} bins is within the statistical uncertainty, and so this systematic uncertainty is negligible. The PDF uncertainty is calculated using the eigenvector error sets included with CT10 [59] and cross-checked by comparing the O_{BDT} distribution using the CT10 and NNPDF [89] PDFs. The PDF uncertainty is $\sim 1\%$ in each bin of O_{BDT}.

Unfortunately it was not feasible to generate reconstructed MC samples with sufficient sample sizes to evaluate the QCD scale and the matrix element model uncertainties. Consequently, these systematic uncertainties were evaluated using “truth-level” samples that are simulated up until the fragmentation stage but are not passed through the time-consuming detector simulation. The truth-level charged leptons are required to originate from the W bosons that come from the top quark decays. Truth-level jets are built using the anti-k_T algorithm on the truth-level hadrons that result from the simulated fragmentation. The p_T and η requirements applied are the same as used in the fully reconstructed samples. Low p_T tracks (also called soft tracks) are simulated for charged particles that are inside $|\eta| < 2.47$, the active area of the tracker. Tracks that are within $\Delta R = 0.4$ of a jet are not included in the soft track term as their contribution is already accounted for in the transverse momenta of the corresponding jet. Truth-level p_T^{miss} is defined as the vector sum of the transverse momenta of the leptons, jets, and soft tracks.

In order to maximize the MC sample size, a reweighting technique was used to simulate the b-tagging. Candidate events are reweighted by the probability of having zero b-jets or one b-jet for the signal region and control region respectively. The weights are calculated in terms of the b-tagging efficiencies (ϵ_i) which are derived as a function of the p_T, η, and flavour of the ith jet. For 0 b-tags:

$$w_0 = \prod_i [1 - \epsilon_i(p_T, \eta, \text{flav})] \tag{8.1}$$

and for 1 b-tag:

$$w_1 = \sum_i \epsilon_i(p_T, \eta, \text{flav}) \cdot \prod_{i \neq j} [1 - \epsilon_i(p_T, \eta, \text{flav})] \tag{8.2}$$

In order to motivate their validity for evaluating the top uncertainties, the truth level samples are compared to fully reconstructed samples. In general there is reasonable agreement between the truth level and reconstructed samples. A comparison for m_{jj} and $m_{\ell\ell}$ using MC@NLO is shown in figure 8.1. Note that agreement is not critical since systematics are evaluated as the ratio of generators and so effects of the detector reconstruction approximately cancel out. For each systematic evaluated on the O_{BDT} distribution using truth level samples, the ratios of truth to reconstructed kinematics are compared for the two generators to confirm they are consistent. For all the generators used in the evaluation of systematics for $t\bar{t}$, there is good agreement between the truth and reconstructed level O_{BDT} distribution.

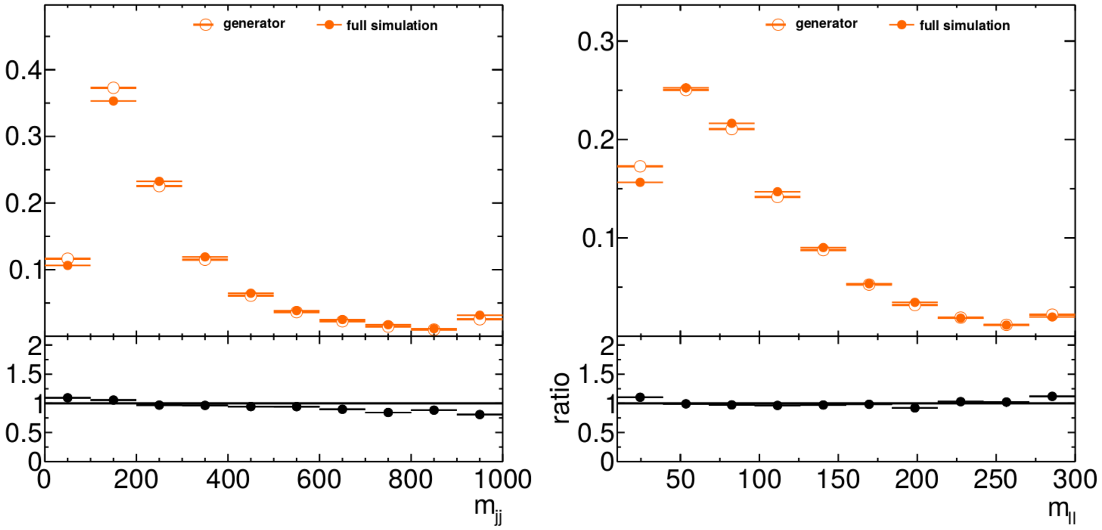


Figure 8.1: Comparison of the truth level (hollow points) and reconstructed level (solid points) distributions for m_{jj} (left) and $m_{\ell\ell}$ (right) using the MC@NLO $t\bar{t}$ generator. These plots are made in the signal region but with the $Z \to \tau\tau$ veto, outside lepton veto, and BDT selection requirement removed.

The $t\bar{t}$ QCD scale uncertainty was evaluated using a truth-level MC@NLO [90] sample that was showered with HERWIG. The factorization scale (μ_F) and renormalization (μ_R) scales in the MC generator are independently scaled up and down by a factor of two as recommended by the LHC Higgs Cross Section Working Group [17]. These parameters probe the sensitivity to higher order QCD corrections in the matrix element level calculation. Even though the uncertainty is chosen as the largest deviation in the extrapolation factor, α, it is still within the statistical uncertainty in each O_{BDT} bin. The largest deviation in the extrapolation factor is in the most sensitive O_{BDT} bin and is 5%.

The dominant theoretical uncertainty for the $t\bar{t}$ background is the matrix element modelling uncertainty. The extrapolation factor α is derived for three different generators, MC@NLO, POWHEG, and ALPGEN, all showered with HERWIG. MC@NLO and POWHEG are both NLO predictions and so account for potential differences in the calculation of NLO matrix elements. In contrast, ALGEN is a multi-leg LO generator which simulates the emission of up to three additional partons. The uncertainty on α is taken from the difference between MC@NLO and ALPGEN which exhibit the greatest difference of the three generators. The uncertainty on α in O_{BDT} bins 1, 2, and 3 is $10\%, 12\%$, and 21% respectively. The ratio of $\alpha_{\mathrm{MC@NLO}}/\alpha_{\mathrm{ALPGEN}}$ is shown in figure 8.2. The red points correspond to the fully simulated sample with limited statistics, while the blue points correspond to the truth-level sample. By using the truth level sample, the statistical component of this systematic uncertainty is significantly reduced.

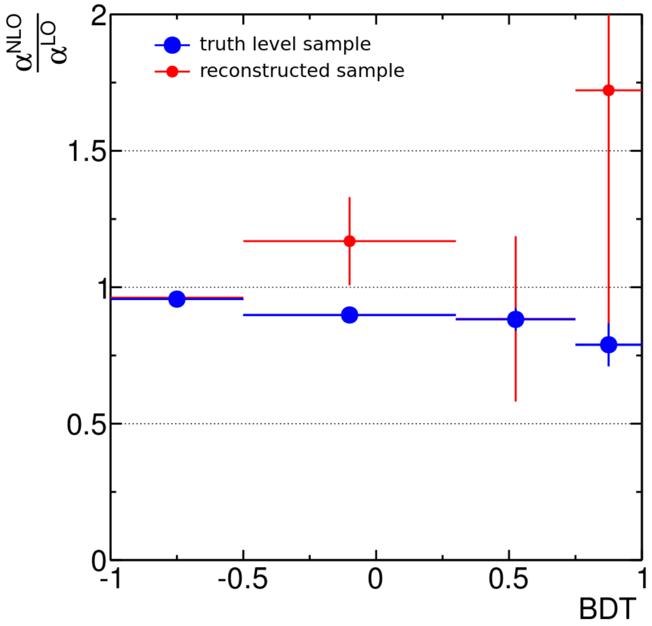


Figure 8.2: The variation from unity in the ratio in the extrapolation factor α for ALPGEN (LO) to MC@NLO (NLO) is used to evaluate the systematic uncertainty on α. The blue points correspond to the truth-level sample, while the red correspond to the fully reconstructed sample. The statistical uncertainty on the systematic uncertainty is significantly reduced using the truth-level sample.

8.1.2 $Z \to ee/\mu\mu$ Uncertainties in Run-I

The $Z \to ee/\mu\mu$ estimate relies on two assumptions which are checked using MC. First, in order to be able to extrapolate the O_{BDT} shape from the low E_T^{miss} region B to the signal region, it is necessary that O_{BDT} is not correlated to E_T^{miss}. At first glance, this assumption may not seem motivated since p_T^{miss} is used in some of the BDT training variables. However, note that only two of the BDT training variables, m_T and p_T^{sum}, depend on p_T^{miss}. Furthermore, the control region is defined using the calorimeter-based E_T^{miss} rather than the tracker-based p_T^{miss}. Figure 8.3 shows the $Z \to ee/\mu\mu$ O_{BDT} distribution for high E_T^{miss} (region A) and low E_T^{miss} (region B). The left plot is made using ALPGEN interfaced with HERWIG, the nominal generator for this process. In this case, there is no correlation observed between O_{BDT} and E_T^{miss}. The comparison is also done using PYTHIA for the parton showering (right plot) which results in a discrepancy in the highest O_{BDT} bin. Since there is no intrinsic reason to expect the parton showering to impact the underlying correlation between O_{BDT} and E_T^{miss}, this is most likely a statistical fluctuation. In order to safeguard against underestimating the uncertainty, the largest difference from these two generators is taken as the systematic uncertainty, resulting in uncertainties of $4\%, 10\%$, and 60% on $Z \to ee/\mu\mu$ in the three bins of O_{BDT}.

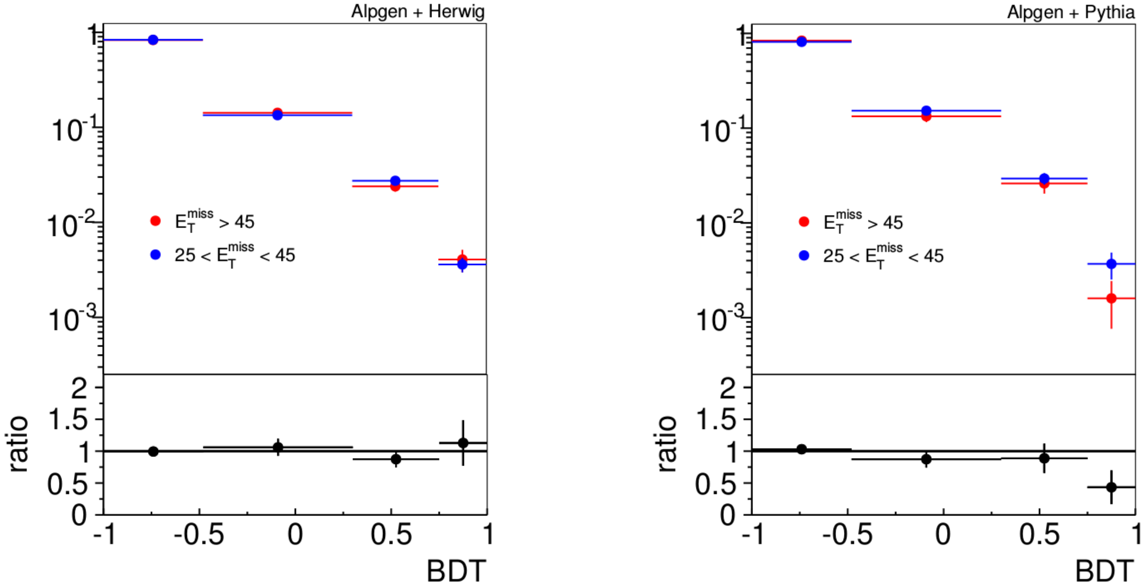


Figure 8.3: O_{BDT} distributions for $Z \to ee/\mu\mu$ for high E_T^{miss} (red) and low E_T^{miss} (blue) for ALPGEN+HERWIG (left) and ALPGEN+PYTHIA (right).

Secondly, in order to be able to extrapolate the relative efficiencies of the E_T^{miss} selection requirements from the Z-peak window to the low $m_{\ell\ell}$ region, it is necessary that E_T^{miss} is not correlated with $m_{\ell\ell}$. However, note that as the center-of-mass energy of a $Z \to ee/\mu\mu$ candidate event increases, $m_{\ell\ell}$ also tends to increases. At the same time, a higher center-of-mass energy also increases the amount of initial state radiation which in turn increases the resolution on the E_T^{miss}. Consequently, a small correlation is expected between the $m_{\ell\ell}$ and E_T^{miss}. As mentioned in section 7.4, this effect is captured by a correction factor which is called the non-closure of the estimate which can only be obtained from the MC:

$$f_{\text{non-closure}} = \frac{N_A^{\text{MC}}/N_B^{\text{MC}}}{N_C^{\text{MC}}/N_D^{\text{MC}}} \tag{8.3}$$

where N_X^{MC} are the $Z \to ee/\mu\mu$ MC yields for region X. The non-closure is calculated to be 0.8 ± 0.2. In addition, the 17% deviation of $f_{\text{non-closure}}$ from unity is also assigned as an uncertainty.

8.1.3 W+jet Uncertainties in Run-I

A number of sources of systematic uncertainty are considered for the determination of uncertainties on the fake factors. Recall that f_{fake} is calculated from a Z enriched region, but is applied to W+jet candidate events. MC is used to calculate the ratio of f_{fake} using W samples to f_{fake} using Z samples. This correction factor is applied to f_{fake}, and the difference is also assigned as a systematic uncertainty. For electron fakes, the uncertainty on f_{fake} is 20% while for muon fakes it is 22%. In addition, a systematic uncertainty is also applied on f_{fake} to account for electroweak contamination in the Z control region. This systematic uncertainty is calculated by comparing the fake factor with and without the inclusion of electroweak processes. It varies as a function of p_T and η, and is maximally 25% for electron fakes and 21% for muon fakes. In addition, the statistical uncertainty on the control region is included in the systematic uncertainty which ranges up to 52% for electron fakes, and 34% for muon fakes. A similar procedure is used to evaluate the uncertainty on f'_{fake}. The dominant source of systematic uncertainty is the choice of multi-jet sample from which to derive the fake factors. It is 60% for electron fakes, and 40% for muon fakes.

As shown in table 8.1, the uncertainties on the fake estimate are relatively large. However, this does not significantly impact the analysis since the expected fake yields in the VBF $H \to WW^* \to \ell\nu\ell\nu$ signal region are small.

8.1.4 Other Theoretical Uncertainties in Run-I

The dominant uncertainty for the VBF MC sample is the PS/UE uncertainty. O_{BDT} is particularly sensitive to this uncertainty since one of the BDT training variables, p_T^{sum}, relies on the modelling of low p_T jet activity. It is evaluated by comparing the default parton showering program PYTHIA8 with HERWIG which results in a shape uncertainty of $\sim 0\%$, 5%, 14% in O_{BDT} bins 1, 2 and 3 respectively.

In order to assess the impact of the choice of QCD scale, the factorization and renormalization scales for the default VBF MC sample are independently scaled up and down by a factor of two. The resulting uncertainty on the overall yields in the signal region is on the order of 0.5%, and the shape uncertainties are 1%, 3%, and 3%.

The PDF uncertainty is evaluated by comparing the default CT10 PDF with NNPDF. This results in a normalization uncertainty of 3%.

Next, the uncertainty from the choice of matrix element model is evaluated by comparing the default generator for VBF, POWHEG, to aMC@NLO. Both generators are NLO but use different methods to interface the hard scatter to the parton showering. The resulting normalization uncertainty on the choice of matrix element model is 4%.

As mentioned in sections 5.1.1 and 5.1.2, the VBF Higgs boson production process is modelled with up to NLO EW corrections. The VBFNLO [69] generator is used to determine the size of these higher order EW corrections. A 4% difference in the total normalization is calculated when running the generator with and without the higher order EW diagrams while the impact on the O_{BDT} shape is negligible. Since the EW corrections have already been applied to the VBF Higgs boson production process, a corresponding systematic is not applied.

Along with $t\bar{t}$, WW is one of the major backgrounds in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis. Efforts were made to construct a control region for this background, but unfortunately it is not apparent how to define one with sufficient purity. One possibility is to select DF candidate events and require $m_{\ell\ell}$ to be large (refer to figure 6.2). Even though this effectively reduces most backgrounds, there is still a large contribution from $t\bar{t}$ in this control region.

In order to evaluate the matrix element model uncertainty for QCD WW, truth level samples are compared using a similar procedure as described for $t\bar{t}$ in section 7.2. The default SHERPA sample is compared to MADGRAPH which is interfaced to PYTHIA for the parton showering. Both generators include up to three radiated partons in the matrix element level calculation. The differences in the two generators result in shape uncertainties of 14%, 8%, and 12%. The EW WW matrix element model is also evaluated using SHERPA and MADGRAPH interfaced with PYTHIA. This results in a normalization uncertainty of 10% and shape uncertainties of 16%, 12%, and 10%.

Scaling the factorization and renormalization scales up and down by a factor of two results in a normalization uncertainty of 27% and a negligible shape uncertainty for QCD WW. This uncertainty is evaluated using MADGRAPH interfaced to PYTHIA. For EW WW, a normalization uncertainty of 10% is assigned based on a generic LO estimation for EW WW [91].

The remaining uncertainties for QCD and EW WW are relatively small compared to the matrix element model uncertainty and QCD scale. The PDF uncertainties are on the order of a few percent. The interference between EW WW and VBF $H \to WW^*$ is also computed and found to be negligible at 1%. It is calculated by comparing the cross sections for EW WW with and without the interference from VBF $H \to WW^*$, and cross checked by comparing the cross sections for VBF $H \to WW^*$ with and without the interference from EW WW. Using a similar approach, the uncertainty on the interference between QCD WW and EW WW is also found to be negligible at 2%.

Higgs boson production via gluon-gluon fusion is a significant background in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis. In order to assess the uncertainties coming from higher order QCD corrections, the factorization and renormalization scale are independently varied using the MCFM generator. Note that the central jet veto effectively removes events with 3 or more jets and so the uncertainty needs to be calculated for this region of phase space. The Stewart-Tackmann procedure [92] is used to calculate the uncertainty in the 2-jet bin with this consideration in mind. The cross section for this region can be written as:

$$\sigma_{2j} = \sigma_{\geq 2j} - \sigma_{\geq 3j} \tag{8.4}$$

where $\sigma_{\geq 2j}$ is the cross section for 2 or more jets, and $\sigma_{\geq 3j}$ is the cross section for 3 or more jets. The Stewart-Tackmann method assumes that the uncertainties on $\sigma_{\geq 2j}$ and $\sigma_{\geq 3j}$ are uncorrelated. The effect of varying the QCD scale is assessed in the signal region for $\sigma_{\geq 2j}$ without the CJV applied and for $\sigma_{\geq 3j}$ with the CJV inverted. The resulting normalization uncertainty on σ_{2j} is 29% and the shape uncertainties are 3%, 7%, and 48% in bins of O_{BDT}.

In order to evaluate the uncertainty from the PS/UE, the default PYTHIA8 is compared with HERWIG resulting in a 15% normalization uncertainty. Finally, the PDF uncertainty is calculated to be 8%.

Even though the $Z \to ee/\mu\mu$ background is estimated using a data-driven approach, recall from section 7.4 that the relative rates in the second and third bins of O_{BDT} are derived from MC. The uncertainties for this background are computed using a SHERPA sample. The largest uncertainty is the QCD scale for which a 11% normalization uncertainty is assigned. The other sources of uncertainty do not contribute significantly.

The $Z \to \tau\tau$ process is normalized using a control region. Unfortunately the statistics in this region are low resulting in a large statistical uncertainty of 30% on the expected yield. Since the theory systematic uncertainties are expected to be within the statistical uncertainty and since this background is small in the signal region, theory uncertainties are not computed for this background.

8.2 Theoretical Uncertainties in Run-II

Unlike in Run-I, all Run-II uncertainties are calculated using reconstructed samples. Consequently, some of the systematic uncertainties, particularly those in the most sensitive bin of O_{BDT}, will be driven by statistical uncertainties. However, since the statistical uncertainty on the data in the most sensitive O_{BDT} bin is already large, this does not significantly degrade the performance of the analysis with the current dataset. With the larger dataset that will be available for the next result to be published, the uncertainties for Run-II will need to be reevaluated.

8.2.1 Top Quark Uncertainties in Run-II

As in Run-I, the dominant uncertainty on the top background is the matrix element modelling uncertainty which is evaluated to be 35%. It is calculated for $t\bar{t}$ by comparing POWHEG and MG5_aMC@NLO [58] both interfaced with HERWIG++ [93].

The QCD scale uncertainty is evaluated using POWHEG interfaced with PYTHIA6 by varying the relevant parameters. The PS/UE uncertainty is derived by comparing the PYTHIA6 and HERWIG++ showering models with the hard scatter being modelled by POWHEG in both cases.

8.2.2 W+jet Uncertainties in Run-II

In Run-II, the dominant uncertainty originates from the difference in jet flavour composition between the dijet sample used to evaluate the fake factors and the W+jets control region. It is calculated using a Z+jet enriched data sample which is expected to have a similar ratio of identified to anti-identified leptons as the W+jet control region. The difference between the fake factors in the Z+jet and dijet samples is assigned as a systematic uncertainty. Unfortunately, the Z+jet sample has limited statistics which increases the statistical component of this uncertainty. An additional systematic uncertainty is estimated from MC to cover the difference in sample composition between the W+jet and Z+jet samples. The total sample composition systematic uncertainty is calculated to be 63% for muons and 40% for electrons. This systematic uncertainty also has the largest impact on the determination of the signal strength as described in section 9. However, this uncertainty is expected to be smaller in future iterations of the analysis as more statistics in data will be available.

Uncertainties originating from the subtraction of electroweak processes and statistical uncertainties on the fake factors are also considered.

8.2.3 Other Theoretical Uncertainties in Run-II

All of the Run-II PDF uncertainties are evaluated using a common procedure. The CT10 PDF is compared with both MSTW [94] and NNPDF. The resulting difference is then added in quadrature with the uncertainty derived from the PDF eigenvector error set included with CT10.

The PS/UE uncertainties for VBF Higgs boson production are derived by interfacing the default POWHEG generator to both PYTHIA8 and HERWIG7 [95] and comparing the difference. As in Run-I, the factorization and renormalization scales are independently scaled up and down by a factor of two to determine the QCD scale uncertainty. Next, the matrix element model uncertainty is evaluated by comparing POWHEG to MG5_aMC@NLO where both generators are interfaced to PYTHIA8.

Unlike in Run-I, all of the QCD WW uncertainties are derived from reconstructed samples rather than truth level samples. The uncertainty is dominated by the matrix element modelling uncertainty which is 50%. It is derived by comparing SHERPA to the nominal POWHEG interfaced with PYTHIA. In general, SHERPA is expected to model the forward jets more accurately that POWHEG. However, there is also a slight mismodelling of the data observed in version 2.1 of SHERPA for forward jets.

The QCD scale uncertainties are evaluated using a SHERPA sample that includes two radiated jets at matrix element level. The PS/UE uncertainties are evaluated by comparing the nominal POWHEG generator interfaced with PYTHIA8 and with HERWIG++.

Next, the ggF background includes an uncertainty on the acceptance of the central-jet veto. As in Run-I, it is evaluated using the Stewart-Tackmann method, but using MG5_aMC@NLO. The resulting uncertainties are 28% for $H + 2$-jet events and 32% for $H + 3$-jet events. The QCD scale uncertainties are 3% in bin 1 of the O_{BDT}, and 32% in bin 2.

The uncertainty on the ggF matrix element modelling is determined by comparing POWHEG with MG5_aMC@NLO, both interfaced to PYTHIA8 for the parton showering.

8.3 Experimental Uncertainties

Experimental uncertainties originate from uncertainties associated with the modelling of the ATLAS detector. In general, these systematic uncertainties are evaluated in bins of O_{BDT}. For the sake of conciseness, the uncertainties in this section are listed as averages over all O_{BDT} bins in the signal region.

The sizes of the uncertainties are summarized in table 8.1 for the Run-I DF analysis and in table 8.2 for the Run-I SF analysis. Further details on these uncertainties are provided in the text that follows. For the top background, the uncertainties shown are the uncertainties on the extrapolation factor α as defined in equation 7.2. All other uncertainties shown are on the total event yields. The Run-II experimental uncertainties are not included here as they are dwarfed by the statistical uncertainties on the data. However, these uncertainties are included in the analysis and are expected to become more important as more 13 TeV data is analyzed.

In general, leptons are well understood and the uncertainties originating from them are relatively small. The efficiencies of selecting leptons in MC are corrected with scale factors to match data. Two uncertainties are assigned on the scale factors, one corresponding to the lepton isolation, and the other corresponding to the lepton identification. Two further

Uncertainty	Sig.	Bkg.	Top	WW	ggF	Other VV	Z/γ	Fakes
Trigger	-	0.1	-	-	-	0.5	0.5	-
Electron identification	0.8	0.7	0.8	0.9	0.8	0.9	0.9	-
Electron isolation	0.7	0.6	0.7	0.7	0.8	0.8	0.8	-
Electron resolution	-	0.6	0.3	0.1	-	1.4	4.7	-
Electron scale	-	0.5	0.5	0.6	-	1.3	2.8	-
Muon identification	-	-	-	-	-	-	-	-
Muon isolation	0.8	0.7	0.8	0.8	0.8	0.9	0.9	-
Muon resolution	-	-	-	-	-	-	0.1	-
Muon scale	-	0.1	-	-	-	-	0.4	-
JES flavour	0.8	0.9	0.5	2.9	3.0	3.5	0.3	-
JES η model	2.8	4.7	3.9	10.2	5.9	1.3	6.6	-
JES η statistical	-	0.7	0.8	1.3	0.8	1.8	0.2	-
JES flavour response	0.7	0.2	0.1	2.1	1.9	2.9	4.7	-
JES detector	-	0.7	0.9	0.9	0.7	0.9	-	-
JES pile-up p_T	-	0.6	-	0.6	-	-	4.2	-
JES high p_T	-	0.5	-	-	-	-	3.8	-
JES pile-up ρ	-	-	0.5	1.5	1.4	0.6	5.6	-
JES modelling	1.1	1.7	1.9	3.8	3.4	5.4	2.6	-
JES $< \mu >$	0.1	0.5	0.3	0.2	0.4	2.5	0.5	-
JES N_{PV}	-	0.3	0.2	0.7	0.4	2.8	3.2	-
JES AFII	-	-	0.1	-	-	-	-	-
JES total	3.2	5.3	4.6	11.7	7.9	8.4	12.1	-
JER	2.1	3.2	1.8	1.5	0.3	11.8	11.5	-
b-tag SF	-	2.4	6.5	-	-	-	-	-
light tag SF	1.3	1.3	1.2	1.7	1.9	1.8	1.6	-
c-tag SF	-	-	-	0.1	-	-	-	-
E_T^{miss} scale	-	0.3	-	-	-	-	2.2	-
E_T^{miss} resolution	-	-	-	-	-	-	-	-
p_T^{miss} scale	-	2.2	2.6	1.9	1.0	0.3	6.6	-
p_T^{miss} resolution	0.3	1.8	2.0	1.7	1.1	7.8	4.5	-
dijet fake rate	-	1.9	-	-	-	-	-	14.5
fake rate μ	-	1.6	-	-	-	-	-	12.0
fake rate e	-	2.2	-	-	-	-	-	17.1
$< \mu >$ re-scale	1.2	0.5	0.8	0.3	0.9	3.6	4.4	-

Table 8.1: Experimental normalization uncertainties in the Run-I DF analysis in percent calculated for the entire signal region.

Uncertainty	Sig.	Bkg.	Top	WW	ggF	Other VV	Z/γ	Fakes
Trigger	1.2	0.5	1.1	1.2	1.2	0.6	0.1	-
Electron identification	0.6	0.3	0.6	0.7	0.6	1.9	0.1	-
Electron isolation	0.6	0.3	0.5	0.6	0.6	1.4	0.1	-
Electron resolution	-	0.4	0.5	0.1	0.3	0.3	0.5	-
Electron scale	-	0.1	0.1	1.0	-	3.4	0.2	-
Muon identification	-	-	-	-	-	-	-	-
Muon isolation	1.0	0.4	1.0	1.0	1.0	0.4	0.1	-
Muon resolution	-	0.1	-	0.4	-	-	0.1	-
Muon scale	-	-	0.2	-	-	-	0.1	-
JES flavour	1.9	1.1	0.6	4.4	4.0	18.0	0.1	-
JES η model	5.2	3.2	5.4	9.8	8.0	10.2	1.2	-
JES η statistical	0.6	0.6	0.9	1.1	0.6	4.7	0.3	-
JES flavour response	1.4	0.9	0.7	2.3	2.5	14.2	0.1	-
JES detector	0.9	0.7	1.0	0.6	0.8	7.4	0.3	-
JES pile-up p_T	-	0.4	-	0.1	0.3	0.9	0.6	-
JES high p_T	-	0.2	-	-	-	-	0.3	-
JES pile-up ρ	0.4	0.5	0.1	2.0	1.0	10.9	-	-
JES modelling	2.1	1.6	1.4	4.6	4.2	15.2	0.6	-
JES $< \mu >$	-	0.5	0.1	1.1	0.7	3.3	0.4	-
JES N_{PV}	-	0.4	0.1	1.1	0.3	7.0	0.3	-
JES AFII	-	-	0.1	-	-	-	-	-
JES total	6.2	4.1	5.8	12.2	10.3	33.4	1.6	-
JER	2.5	1.5	1.8	4.1	1.0	1.2	1.3	-
b-tag SF	-	1.3	6.8	-	-	0.8	-	-
light tag SF	1.4	0.6	1.2	1.8	2.0	1.3	0.1	-
c-tag SF	-	-	-	0.1	-	0.2	-	-
E_T^{miss} scale	-	0.2	-	-	-	3.6	0.2	-
E_T^{miss} resolution	-	0.2	-	-	-	5.9	0.1	-
p_T^{miss} scale	0.1	1.0	3.0	1.1	1.1	2.7	0.6	-
p_T^{miss} resolution	0.6	0.9	1.6	1.2	1.1	10.1	0.4	-
dijet fake rate	-	0.1	-	-	-	-	-	4.9
fake rate μ	-	0.3	-	-	-	-	-	12.6
fake rate e	-	0.3	-	-	-	-	-	13.7
$< \mu >$ re-scale	3.4	0.6	0.4	1.2	1.7	13.5	0.2	-

Table 8.2: Experimental normalization uncertainties in the Run-I SF analysis in percent calculated for the entire signal region.

uncertainties are evaluated corresponding to the calibration of the lepton momentum scale and resolution.

Due to the more complex nature of hadronic showers, the uncertainties associated with jets are significantly larger than those associated with electrons and muons. The jet energy scale (JES) and jet energy resolution (JER) are evaluated. The JES uncertainty is evaluated in terms of 12 orthogonal components [37] as summarized in table 8.1. Due to the forward jets in the VBF $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ analysis, the largest component is related to the extrapolation of the JES from the central region to the forward region of the detector. This uncertainty is driven by the significant difference in jet η modelling between the POWHEG and HERWIG generators. There are also contributions to account for the behaviour of high p_T jets, the difference in response of the calorimeter to jets originating from light quarks versus gluons, uncertainties from the modelling of pile-up, and uncertainties from jet energy corrections. Although not as large as the JES, the JER also contributes significantly to the uncertainty on the total event yields.

Next, uncertainties are calculated on the efficiency of the b-tagging algorithms. The uncertainties are derived from the theoretical and detector uncertainties on the inputs to the b-tagging scale factors. The efficiencies are calculated from a $t\bar{t}$ rich control region in six p_T bins resulting in six orthogonal contributions to the total uncertainty [46]. This uncertainty has the most impact on the top background since it contains hadrons originating from real top quarks. Further uncertainties are also applied to account for the light flavour and charm hadrons that have been misidentified as b-jets.

Recall from equation 3.3 that E_T^{miss} and p_T^{miss} are defined in terms of identified objects and soft objects. The uncertainty on the first contribution is derived from the uncertainties calculated for electrons, muons and jets. For the soft objects, the uncertainty on the QCD scale and resolution is calculated. In general, these uncertainties grow with the amount of QCD radiation in the process of interest.

The uncertainty from pile-up is assessed by rescaling $< \mu >$, the average number of interactions per bunch crossing, and observing the impact in the signal region. Finally, whenever the normalization of a process is calculated using MC, the uncertainty on the total luminosity needs to be taken into account. It is on the order of a few percent.

Chapter 9

Statistical Analysis

Statistical data analysis methods are used to interpret the data in the context of the SM predictions. The total yields in the signal region are written as the sum of the total predicted signal scaled by the signal strength μ, and the total predicted background. The background-only hypothesis is defined as the case where $\mu = 0$. Using a ratio of likelihood functions, a test statistic is defined and used to quantify the probability of observing an analysis outcome that is at least as signal-like as the one in data, assuming the background-only hypothesis. If the background-only hypothesis can not be rejected, limits are set on physical parameters of interest such as the Higgs boson mass. In contrast, if the background-only hypothesis is rejected with sufficient significance, the presence of a signal is established and its properties are measured. A measured value of μ consistent with 1 indicates that the data yields in the signal region are consistent with the total SM predicted yield for signal and background. Additional properties of interest include the mass, production cross section, spin, parity, and coupling strength to other particles. These provide probes to check if the measured signal is consistent with the Higgs boson predicted by the SM.

9.1 Likelihood Function

A likelihood function, $\mathcal{L}$, is used to interpret the data in terms of the expected yields in the various data samples defined in the analysis. $\mathcal{L}$ is defined as the likelihood of observing an outcome in data given a set of parameters that describe the degrees of freedom in the model.

In order to motivate the functional form of $\mathcal{L}$, consider a simplified analysis consisting of a signal region defined by a single set of selection requirements. For the moment, it is assumed that there are no uncertainties on the background and signal predictions in this signal region. The yields in this signal region are expected to follow a Poisson probability distribution:

$$P(N|\lambda) = \frac{e^{-\lambda}\lambda^{N}}{N!} \tag{9.1}$$

where N is the observed number of events and the Poisson mean, λ, is the expected number of candidate events. Then the likelihood function can be written as:

$$\mathcal{L}(\mu) = P(N_{\text{SR}}^{\text{data}} \,|\, \mu N_{\text{SR}}^{\text{sig}} + N_{\text{SR}}^{\text{bkg}}) \tag{9.2}$$

where P is assumed to be a Poisson probability, μ is the signal strength (defined in section 1.4), $N_{\text{SR}}^{\text{data}}$ is the number of events observed in data, $N_{\text{SR}}^{\text{sig}}$ is the number of predicted signal events, and $N_{\text{SR}}^{\text{bkg}}$ is the number of predicted background events. Since μ is the parameter being measured, it is the *parameter of interest.* The expression for $\mathcal{L}(\mu)$ represents the model that is used to interpret the data. Its value is estimated by maximizing $\mathcal{L}(\mu)$ with respect to μ. This is accomplished by taking the derivative of $\mathcal{L}$ with respect to μ, setting it to zero, and solving for μ. The value of μ that maximizes the likelihood function is denoted as $\hat{\mu}$, where the carat indicates this is an estimate for μ. Due to the simplicity of this likelihood function, it can be maximized analytically. In practice, likelihood functions used in physics analyses are more complex and need to be maximized using numerical methods. A common strategy is to minimize the negative log of the likelihood function. The log helps ensure that the calculated results do not exceed the numerical bounds of the computational tool being used. In addition, it simplifies the optimization since a log of products can be rewritten as a sum of logs. In the case of the $H \to WW^* \to \ell\nu\ell\nu$ analysis, the optimization is done with the MINUIT software package [96].

For Poisson probabilities, the likelihood maximization results in:

$$\hat{\mu} = \frac{N_{\text{SR}}^{\text{data}} - N_{\text{SR}}^{\text{bkg}}}{N_{\text{SR}}^{\text{sig}}} \tag{9.3}$$

Assuming the existence of a SM Higgs boson, the expected value of $\hat{\mu}$ is $\hat{\mu}_{\text{exp}} = 1.0$. The deviation of the observed signal strength in data from $\hat{\mu}_{\text{exp}}$ is denoted as $\hat{\mu}_{\text{obs}}$ and provides a probe to test the SM.

The model described by equation 9.2 is the simplest implementation of the likelihood approach and does not adequately describe the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis. First, it does not account for the control regions that are used to constrain specific backgrounds as described in chapter 7. Assuming a normalization factor of β is applied to background A in all the signal and control regions in the analysis, the likelihood function is extended as:

$$\mathcal{L}(\mu,\beta) = P(N_{\text{SR}}^{\text{data}} \,|\, \mu N_{\text{SR}}^{\text{sig}} + \beta N_{\text{SR}}^{\text{bkg }A} + N_{\text{SR}}^{\text{other bkg}}) \cdot P(N_{\text{CR}}^{\text{data}} \,|\, \mu N_{\text{CR}}^{\text{sig}} + \beta N_{\text{CR}}^{\text{bkg }A} + N_{\text{CR}}^{\text{other bkg}}) \tag{9.4}$$

where $N_{\text{X}}^{\text{sig}}$, $N_{\text{X}}^{\text{bkg }A}$, and $N_{\text{X}}^{\text{other bkg}}$ are the expected event yields in region X for the signal, background A, and all other backgrounds besides A, respectively. As before, the values of μ and β that best fit the data are determined by maximizing $\mathcal{L}(\mu,\beta)$. Note that the signal and control regions are treated identically in the likelihood function. However, since control regions are constructed to have negligible signal yields, a priori they are not expected to have a large direct impact on the value of μ. Equation 9.4 is generalized to multiple signal and control regions by adding an additional Poisson probability term for each new independent signal and control region.

Next, the likelihood function defined so far does not account for the sources of systematic uncertainties on the MC described in chapter 8. In addition, since the MC samples have a finite number of events, there are statistical uncertainties on the MC as well. The parameters corresponding to these uncertainties are referred to as nuisance parameters (NPs) since they are not parameters of immediate interest. In this thesis, nuisance parameters are denoted by θ_i.

Nuisance parameters allow the event yields to vary within their uncertainties in the maximization of the likelihood. For a given nuisance parameter, θ_i, the event yield is rewritten as:

$$N(\theta_i) = N_0 \nu(\theta_i) \tag{9.5}$$

where N_0 is the expected nominal yield, and $\nu(\theta_i)$ describes how the nuisance parameter affects the yield. Next, the likelihood is multiplied by a penalty term, $\mathcal{A}(\theta_i|\bar{\theta}_i)$, corresponding to the probability distribution of θ_i and where $\bar{\theta}_i$ is the central value of θ_i.

For systematic uncertainties on MC yields, the exponential response function is used in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis:

$$\nu(\theta_i) = (1 + \epsilon_i)^{\theta_i} \tag{9.6}$$

with a unit-variance Gaussian probability distribution:

$$\mathcal{A}(\theta_i|\bar{\theta}_i) = \mathcal{G}(\theta_i|\bar{\theta}_i, 1) = \frac{1}{2\pi} \exp\left(\frac{-(\bar{\theta}_i - \theta_i)^2}{2}\right) \tag{9.7}$$

The values for the parameter ϵ_i correspond to uncertainties summarized for the various systematic uncertainties in chapter 8. If a nuisance parameter is varied by one standard deviation, $\theta_i = \pm 1$, the yield will change by $\epsilon_i\%$. With the above described choice of $\nu(\theta_i)$ and $\mathcal{A}(\bar{\theta}_i|\theta_i)$, $N(\theta_i)$ will be log-normally distributed [97] and centered at N_0. Alternatively, it is also possible to choose $\nu(\theta_i) = (1 + \epsilon_i\theta_i)$ which will result in $N(\theta_i)$ being normally distributed. However, note that this choice allows $N(\theta_i)$ to take on non-physical, negative values, particularly in the most sensitive O_{BDT} bin where the expected event counts are low.

Note that even though $\nu(\theta_i)$ gives the likelihood maximization the flexibility to alter $N(\theta_i)$ from its nominal value, $\mathcal{A}(\theta_i|\bar{\theta}_i)$ becomes smaller and consequently penalizes the likelihood as θ_i moves away from $\bar{\theta}_i$. The preceding discussion considered a single nuisance parameter, but the concept is easily generalized to account for many nuisance parameters:

$$N(\vec{\theta}) = N_0 \prod_i \nu_i(\theta_i) \tag{9.8}$$

where each $\nu_i(\theta_i)$ has a corresponding $\mathcal{A}(\theta_i|\bar{\theta}_i)$ in the likelihood expression.

In most cases, a common θ_i is used for each source of systematic uncertainty even though the values of ϵ_i may be different in each bin of O_{BDT} for a given shape uncertainty. The systematic uncertainties are correlated between bins of O_{BDT}. The exception is for the $t\bar{t}$ systematic uncertainty in Run-I where a separate θ_i is used for the two bins of the top control region. This choice gives the data more flexibility in constraining the top yields in the signal region.

Next, the statistical uncertainty on the MC predications also needs to be taken into account. For this end, a random variable $N_{\mathrm{raw}}(\theta_i)$ is defined that corresponds to the number of MC events in a particular region of phase space. $N_{\mathrm{raw}}(\theta_i)$ is expected to be Poisson-distributed [98] and the mean is estimated with $\hat{N}_{\mathrm{raw}}$, the measured number of events. The standard deviation of $N_{\mathrm{raw}}(\theta_i)$ is $\sqrt{\hat{N}_{\mathrm{raw}}}$. For statistical uncertainties:

$$\nu(\theta_i) = \theta_i \tag{9.9}$$

with a Poisson probability distribution:

$$\mathcal{A}(\theta_i|\bar{\theta}_i) = P(\hat{N}_{\mathrm{raw}}|\theta_i N_{\mathrm{raw}}) \tag{9.10}$$

Since increasing the number of nuisance parameters increases the complexity of the likelihood maximization, a single nuisance parameter is applied for the statistical uncertainty on the total yield in each signal and control region rather than including separate nuisance parameters for each process.

For consistency, the treatment of the control regions summarized in equation 9.4 is also expressed in terms of $\nu(\theta_i)$ and $\mathcal{A}(\theta_i|\bar{\theta}_i)$:

$$\nu(\theta_i) = \theta_i \tag{9.11}$$

$$\mathcal{A}(\theta_i|\bar{\theta}_i) = P(N_{\mathrm{CR}}^{\mathrm{data}}\,|\,\mu N_{\mathrm{CR}}^{\mathrm{sig}} + \theta_i N_{\mathrm{CR}}^{\mathrm{bkg}\ A} + N_{\mathrm{CR}}^{\mathrm{other\ bkg}}) \tag{9.12}$$

In order to differentiate them from the uncertainties, the nuisance parameters corresponding to the control regions are denoted by β in the equations that follow.

The VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis likelihood is made up of the product of the above described terms:

$$\mathcal{L}(\mu, \vec{\beta}, \vec{\theta}) = \mathcal{L}_{\mathrm{SR}}(\mu, \vec{\beta}, \vec{\theta}) \cdot \mathcal{L}_{\mathrm{CR}}(\mu, \vec{\beta}, \vec{\theta}) \cdot \mathcal{L}_{\mathrm{NP}}(\vec{\theta}) \tag{9.13}$$

For Run-I, the likelihood term corresponding to the signal region is:

$$\mathcal{L}^{\mathrm{I}}_{\mathrm{SR}}(\mu,\vec{\beta},\vec{\theta}) = \prod_{\text{bin } i}^{1,2,3} \prod_{\text{flav } j}^{\text{DF,SF}} \left[P(N^{\mathrm{data}}_{\mathrm{SR}\;i,j} \,|\, \mu N^{\mathrm{sig}}_{\mathrm{SR}\;i,j}(\vec{\theta}) + \sum_{\text{bkg } k}^{\text{top},Z/\gamma} \beta^{k}_{i,j} N^{\mathrm{bkg}\;k}_{\mathrm{SR}\;i,j}(\vec{\theta}) + \sum_{\text{bkg } l}^{\text{other}} N^{\mathrm{bkg}\;l}_{\mathrm{SR}\;i,j}(\vec{\theta})) \right] \tag{9.14}$$

For both DF and SF, there are three separate Poisson terms corresponding to the three bins of O_{BDT}. The index k iterates over the top and $Z/\gamma \to \ell\ell$ background, while the index l iterates over the backgrounds which do not have a control region in the fit. For the top background, the β parameter is the shared between the DF and SF channels, $\beta^{\mathrm{top}}_{i,\mathrm{DF}} = \beta^{\mathrm{top}}_{i,\mathrm{SF}} = \beta^{\mathrm{top}}_{i}$. For the DF $Z \to \ell\ell$ channel, the β parameter is set to zero, $\beta^{Z/\gamma}_{i,\mathrm{DF}} = 0 = \beta^{Z/\gamma}_{i}$. Note that the W+jet and the $Z \to \tau\tau$ control regions are not included in the likelihood maximization. The yields for these processes are calculated externally and included with the sum over the index l which also includes all the backgrounds estimated purely from MC.

The Run-II likelihood term corresponding to the signal region is similar but requires several modifications to account for the differences between the Run-I and Run-II analyses. The number of O_{BDT} bins in the signal region is reduced from three to two, the SF analysis is removed, the $Z/\gamma \to ee/\mu\mu$ control region is removed, and the $Z \to \tau\tau$ control region is added:

$$\mathcal{L}^{\mathrm{II}}_{\mathrm{SR}}(\mu,\vec{\beta},\vec{\theta}) = \prod_{\text{bin } i}^{1,2} \left[P(N^{\mathrm{data}}_{\mathrm{SR}\;i} \,|\, \mu N^{\mathrm{sig}}_{\mathrm{SR}\;i}(\vec{\theta}) + \sum_{\text{bkg } k}^{\text{top},Z\to\tau\tau} \beta^{k}_{i} N^{\mathrm{bkg}\;k}_{\mathrm{SR}\;i}(\vec{\theta}) + \sum_{\text{bkg } l}^{\text{other}} N^{\mathrm{bkg}\;l}_{\mathrm{SR}\;i}(\vec{\theta})) \right] \tag{9.15}$$

Next, the Run-I likelihood term corresponding to the control regions can be written as:

$$\mathcal{L}^{\mathrm{I}}_{\mathrm{CR}}(\mu,\vec{\beta},\vec{\theta}) = \prod_{\text{CR } K}^{\text{top},Z/\gamma} \prod_{\text{bin } i}^{1,2} \left[P(N^{\mathrm{data}}_{\mathrm{CR}K,i} \,|\, \mu N^{\mathrm{sig}}_{\mathrm{CR}K,i}(\vec{\theta}) + \sum_{\text{bkg } k}^{\text{top},Z/\gamma} \beta^{k}_{i} N^{\mathrm{bkg}\;k}_{\mathrm{CR}K,i}(\vec{\theta}) + \sum_{\text{bkg } l}^{\text{other}} N^{\mathrm{bkg}\;l}_{\mathrm{CR}K,i}(\vec{\theta})) \right] \tag{9.16}$$

The product only includes two bins in O_{BDT} in the top control region since the highest two bins are merged in the Run-I analysis. Consequently, in equation 9.14, $\beta^{k}_{2} = \beta^{k}_{3}$.

Analogously, for Run-II:

$$\mathcal{L}^{\mathrm{II}}_{\mathrm{CR}}(\mu,\vec{\beta},\vec{\theta}) = \prod_{\text{CR } K}^{\text{top},Z\to\tau\tau} \left[P(N^{\mathrm{data}}_{\mathrm{CR}K} \,|\, \mu N^{\mathrm{sig}}_{\mathrm{CR}K}(\vec{\theta}) + \sum_{\text{bkg } k}^{\text{top},Z\to\tau\tau} \beta^{k} N^{\mathrm{bkg}\;k}_{\mathrm{CR}K}(\vec{\theta}) + \sum_{\text{bkg } l}^{\text{other}} N^{\mathrm{bkg}\;l}_{\mathrm{CR}K}(\vec{\theta})) \right] \tag{9.17}$$

Since the top and $Z \to \tau\tau$ control regions consist of a single bin, $\beta^{k}_{1} = \beta^{k}_{2} = \beta^{k}$.

The final term in the likelihood expression constrains the nuisance parameters:

$$\mathcal{L}_{\mathrm{NP}}(\vec{\theta}) = \prod_{\text{NP i}} \mathcal{G}(\theta_i|\bar{\theta}_i, 1) \prod_{\text{region } j} P(\hat{N}_j|\theta_j N_j) \tag{9.18}$$

where the first term is the product over all nuisance parameters corresponding to the systematic uncertainties. The second product corresponds to the statistical uncertainties on the MC in all of the signal and control regions in the fit.

The Run-I likelihood maximization is done simultaneously with that of the ggF Higgs boson production analysis mentioned in the $H \to WW^*$ outline in figure 4.1. In this case, an additional nuisance parameter is added corresponding to the signal strength $\mu_{gg\mathrm{F}}$ which is constrained by the ggF analysis. Since both the VBF and VH Higgs boson production modes are produced via an interaction with two weak vector bosons, the parameter of interest, μ_{VBF}, is also used to scale VH. Nevertheless, the analysis is not optimized to be sensitive to VH and the contribution in the signal region is expected to be small. For full details of the ggF Higgs boson likelihood, refer to reference [16]. The likelihood maximization described in this paper also includes 7 TeV data.

The Run-II likelihood maximization is stand-alone and uses SM predictions for ggF and VH Higgs boson production (i.e. $\mu_{gg\mathrm{F}} = 1$ and $\mu_{VH} = 1$). The parameter of interest in this case is μ_{VBF}.

9.2 Discovery Statistic

The significance of a signal is summarized by a p-value which is the probability of observing an analysis outcome that is at least as signal-like as the one in data, assuming the background-only hypothesis. In the particle physics community, the convention is that a p-value less than 1.35×10^{-3} constitutes evidence for a signal and a p-value less than 2.87×10^{-7} constitutes a discovery. Experimental outcomes are ranked on a one-dimensional scale using a test statistic that is used to calculate the p-value.

Since the likelihood function is differentiated in the maximization, the magnitude of $\mathcal{L}$ for a single hypothesis is arbitrary. A meaningful test statistic requires a comparison of at least two likelihood evaluations corresponding to two hypotheses. According to the Neyman-Pearson lemma [99], the ratio of two likelihoods provides the best discrimination between two hypotheses.

The profile likelihood-ratio test statistic is used to compare two hypotheses:

$$q_\mu = -2\ln\frac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})} \tag{9.19}$$

where $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the respective values of μ and $\vec{\theta}$ that maximize their corresponding likelihoods. The denominator is calculated by simultaneously maximizing the likelihood for both μ and $\vec{\theta}$ and consequently equals a constant. When evaluating q_μ for some value of μ, the numerator is evaluated by maximizing the likelihood for $\vec{\theta}$ but leaving μ fixed. Note that because of the negative sign in the definition, maximizing the numerator is equivalent to minimizing q_μ.

In order for q_μ to be meaningful, it is necessary to determine its probability density function (PDF), $f(q_\mu|\mu')$. Integrating $f(q_\mu|\mu')$ over some range of q_μ returns the probability of measuring q_μ in that range given the assumed value of μ in data is μ'. One method of determining $f(q_\mu|\mu')$ is by randomly generating MC *pseudo-experiments* (sometimes informally also called *toys*). Each pseudo-experiment corresponds to a possible observed outcome in data for the yields in all of the analysis signal and control regions. The yields for each region are sampled from their corresponding likelihood terms. In addition, the nuisance parameters are sampled from their corresponding constraint terms, $\mathcal{A}(\theta_i|\bar{\theta}_i)$. The pseudo-experiments can be generated under the background-only hypothesis or under the signal + background hypothesis.

Unfortunately, it is often not feasible to generate a sufficient number of pseudo-experiments for a given analysis, especially if the analysis has a large number of nuisance parameters, a large number of signal regions, or a large number of control regions. Furthermore, since $f(q_\mu|\mu')$ generally falls off towards larger values of q_μ, larger significances require more pseudo-experiments. The primary computational bottle-neck in generating pseudo-experiments is the evaluation of the two likelihood maximizations. Consequently, it is common practice within the ATLAS experiment to approximate $f(q_\mu|\mu')$ using analytic functions. In the case of a single parameter of interest, it has been shown that the test statistic can be rewritten as [100]:

$$q_\mu = -2\ln\frac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})} = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}) \tag{9.20}$$

where $\hat{\mu}$ is a Gaussian distribution with a mean of μ' and a standard deviation of σ, and N is number of events in the data. $\mathcal{O}(1/\sqrt{N})$ is a correction term that becomes negligible as the total number of data events increases. If the correction term in equation 9.20 is ignored, $f(q_\mu|\mu')$ can be rewritten as a non-central chi-squared distribution with one degree of freedom [101]:

$$f(q_\mu|\mu') = \frac{1}{2\sqrt{q_\mu}}\frac{1}{\sqrt{2\pi}}\left(\exp\left(-\frac{1}{2}(\sqrt{q_\mu}+\sqrt{\Lambda})^2\right) + \exp\left(-\frac{1}{2}(\sqrt{q_\mu}-\sqrt{\Lambda})^2\right)\right) \tag{9.21}$$

where the non-centrality parameter Λ is:

$$\Lambda = \frac{(\mu-\mu')^2}{\sigma^2} \tag{9.22}$$

Consequently, assuming the data sample size is large enough, equation 9.21 can be used to approximate $f(q_\mu|\mu')$ without having to generate the CPU intensive pseudo-experiments. This approximation is often referred to as the *asymptotic approximation* which is valid in the case where all the signal and control regions are well populated with data. More details on the validity of using the asymptotic approximation in the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis are provided in section 9.3.

In the $H \to WW^* \to \ell\nu\ell\nu$ analysis, the definition of the test-statistic is modified to also allow $\hat{\mu}$ to drop below the value of μ being tested in the likelihood which makes it sensitive to both excesses and deficits in the data:

$$q_\mu = \begin{cases} -2\ln\frac{\mathcal{L}(\mu,\hat{\vec{\theta}}_\mu)}{\mathcal{L}(\hat{\mu},\hat{\vec{\theta}})} & , \quad \hat{\mu} < \mu \\ +2\ln\frac{\mathcal{L}(\mu,\hat{\vec{\theta}}_\mu)}{\mathcal{L}(\hat{\mu},\hat{\vec{\theta}})} & , \quad \hat{\mu} \geq \mu \end{cases} \tag{9.23}$$

Given $f(q_\mu|\mu')$, it is possible to gauge the agreement between the observed data and a given value of μ. The p-value is defined as the integral of $f(q_\mu|\mu')$ to the right of q_μ^{obs} observed in data:

$$p_\mu = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu|\mu')dq_\mu \tag{9.24}$$

This corresponds to the probability of q_μ being larger than the observed value in data.

In order to test the background-only hypothesis, μ and μ' are set to zero:

$$p_0 = \int_{q_0^{\text{obs}}}^{\infty} f(q_0|0)dq_0 \tag{9.25}$$

As q_0^{obs} becomes larger, p_0 becomes smaller making the background-only hypothesis more and more unlikely.

In order to make the p-values easier to interpret, they are typically converted into a Gaussian significance denoted as Z_μ:

$$Z_\mu = \Phi^{-1}(1 - p_\mu) \tag{9.26}$$

where $\Phi^{-1}(x)$ is the inverse of the Gaussian cumulative distribution function:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2} dt \tag{9.27}$$

The threshold of $p_0 \leq 2.87 \times 10^{-7}$ for discovery or observation corresponds to $Z_0 \geq 5$, while the threshold of $p_0 \leq 1.35 \times 10^{-3}$ for evidence corresponds to $Z_0 \geq 3$. In the case of a discovery, the value and uncertainty on the signal strength μ is of particular interest as it gives an indication of how well the data agrees with the expectations for SM VBF Higgs boson production.

Particularly when optimizing an analysis, it is useful to be able to calculate an expected value for the test statistic. In order to do this, a value needs to be chosen for the assumed signal strength in the data. Typically, the SM model assumption is used ($\mu' = 1$). First,

$f(q_0|1)$ is used to calculate q_0^{exp}, the median value of the test statistic. The expected p-value is then calculated using equation 9.25 with $q_0^{\mathrm{obs}} = q_0^{\mathrm{exp}}$.

Using the approximation in equation 9.21, it can also be shown that if $\mu = \mu' = 0$:

$$Z_0 = \begin{cases} \sqrt{|q_0|} & , \quad q_0 > 0 \\ -\sqrt{|q_0|} & , \quad q_0 < 0 \end{cases} \tag{9.28}$$

9.3 Validation of the Asymptotic Approximation

The results quoted in this thesis are derived using the asymptotic approximation. However, equation 9.20 is only valid if N is large enough. However, by construction the event yields in the most sensitive O_{BDT} bin are expected to be low. Consequently, it is important to confirm that the asymptotic approximation is valid for the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis.

For the Run-II VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis, 250,000 pseudo-experiments are generated for the background-only and for the background + signal hypotheses. Only 3.2 fb^{-1} of integrated luminosity corresponding to the 2015 dataset is used for the studies summarized in this section. The median value of the test statistic for the signal + background hypothesis, $q_0^{\mathrm{S+B\ median}}$, is calculated from the relevant set of pseudo-experiments and then used to calculate q_0.

For the asymptotic approximation, $q_0^{\mathrm{S+B\ median}}$ is approximated from a signal + background Asimov dataset. An Asimov dataset is a single representative analysis outcome where the yields in each of the signal and control regions are the same as the expected yields and all nuisance parameters are set to their nominal values. The test statistic is then calculated using equation 9.21.

The signal + background test statistic distribution is shown in the left plot of figure 9.1. From the pseudo-experiments, $q_0^{\mathrm{S+B\ median}} = 0.80$ while for the asymptotic approximation, $q_0^{\mathrm{S+B\ median}} = 0.83$.

As illustrated in the right plot of figure 9.1, each peak in the signal + background q_0 distribution corresponds to a discrete number of events in the second bin of O_{BDT}. If only the most sensitive O_{BDT} bin is considered, there would be a delta function spike in q_0 for each possible event yield in this bin. However, there are many possibilities for the yields in the other O_{BDT} bin and the control regions, particularly since these have higher expected event yields. Each possibility has a corresponding value for q_0 and consequently the spikes are washed out. In addition, as the expected data yield in the second O_{BDT} bin increases, the bumps are expected to move closer together and eventually become indistinguishable.

Next, the background-only q_0 distribution is shown in the left plot figure 9.2. Note that a χ^2 does not model the features in the q_0 distribution as assumed in the asymptotic

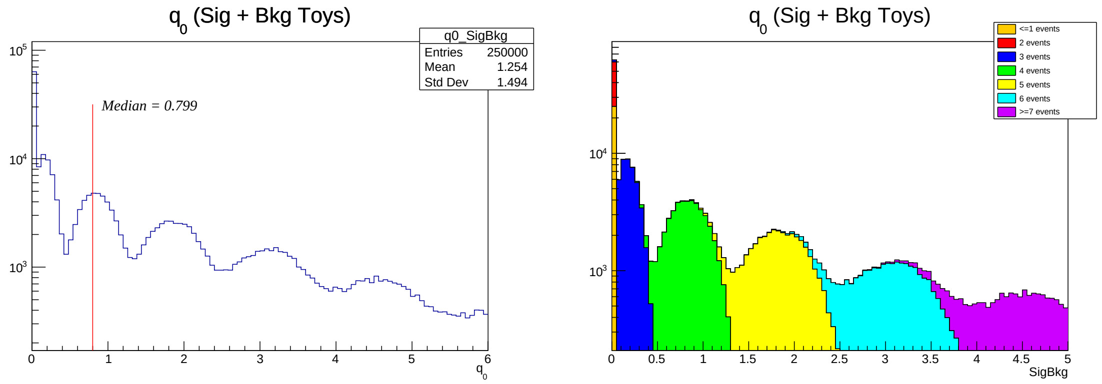


Figure 9.1: Test statistic distribution for the signal + background pseudo-experiments (left). The plot on the right is binned in the number of events in the most sensitive O_{BDT} bin.

approximation. As before, the distinct peaks correspond to the number of events in the most sensitive O_{BDT} bin as illustrated in the right plot of figure 9.2.

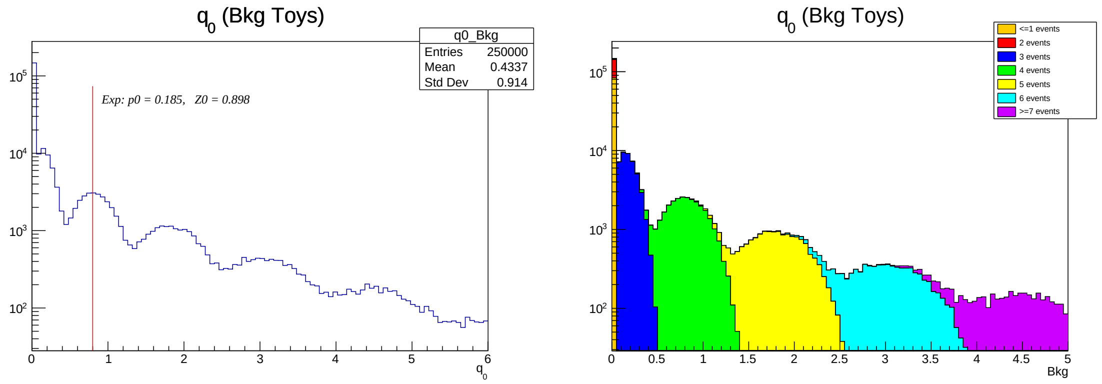


Figure 9.2: Test statistic distribution for the background-only pseudo-experiments (left). The plot on the right is binned in the number of events in the most sensitive O_{BDT} bin.

From the pseudo-experiments, $Z_0^{\text{pseudo}} = 0.90$ while for the asymptotic approximation, $Z_0^{\text{asymp}} = 0.91$. Even though this suggests that the asymptotic approximation gives a reasonable value for the significance, it is necessary to gauge the impact of the sinusoidal behaviour in the q_0 distribution on the significance. Figure 9.3 shows Z_0^{pseudo} as a function of Z_0^{asymp} (left) as well as the ratio of Z_0^{asymp} to Z_0^{pseudo} as a function of Z_0^{asymp} (right). For each background-only pseudo-experiment, $Z_0^{\text{asymp}} = \sqrt{q_0}$ is calculated. The oscillatory behaviour in the pseudo-experiments is within 0.1 unit of Z_0 away from Z_0^{asymp}. This implies that the asymptotic approximation predicts the significance for the background-only pseudo-experiments to within 0.1 units of Z_0 of the pseudo-experiment result. As the size of the dataset increases, the difference between the pseudo-experiments and asymptotic approximation are expected to become smaller and so it is valid to use the asymptotic approximation for the Run-II analysis.

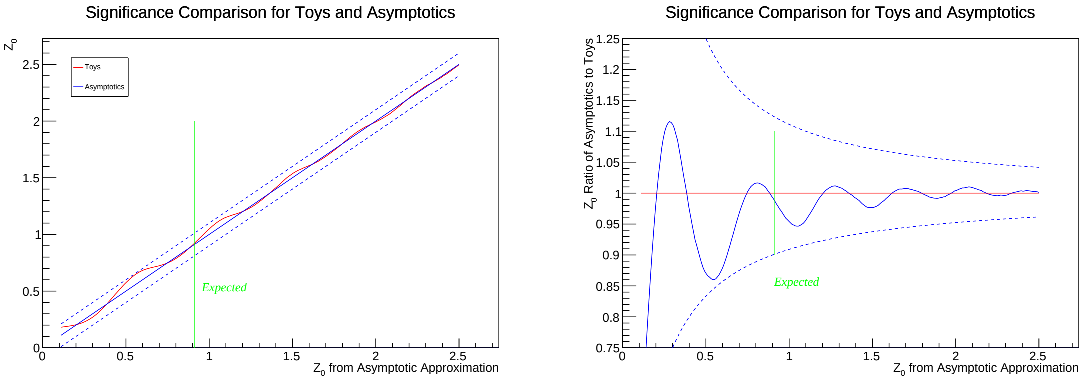


Figure 9.3: The plot on the left shows Z_0^{pseudo} as a function of Z_0^{asymp}, while the plot on the right shows the ratio of Z_0^{asymp} to Z_0^{pseudo} as a function of Z_0^{asymp}, both in red. The dotted blue lines correspond to 0.1 units of significance away from Z_0^{asymp}, and the green vertical lines corresponds to the expected value of $Z_0^{\text{asymp}} = 0.91$.

Figure 9.4 shows the μ distribution for signal + background pseudo-experiments. The median value of μ agrees with the SM expectation of $\mu = 1$. Like in the q_0 distribution, the distinct peaks in μ correspond to different numbers of events in the most sensitive O_{BDT} bin.

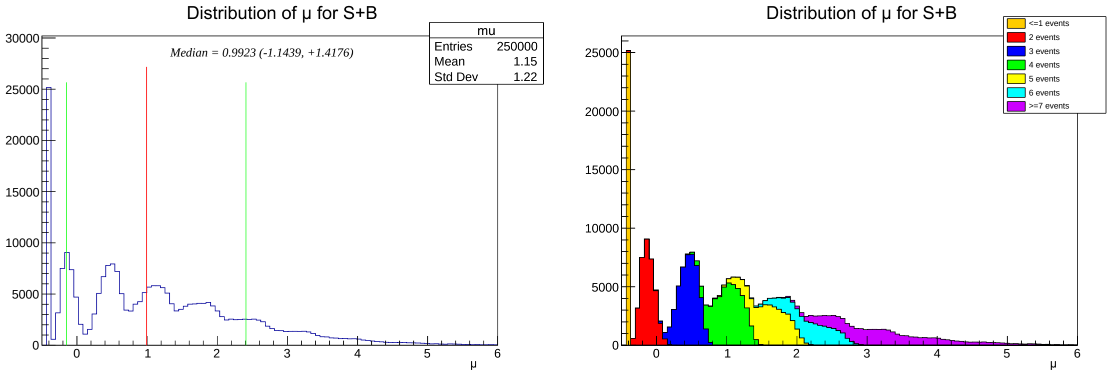


Figure 9.4: Distribution of μ for signal + background pseudo-experiments (left). The red vertical lines corresponds to the median with the green lines corresponding to 1 σ deviations away from the median. The plot on the right is binned in the number of events in the most sensitive O_{BDT} bin.

Figure 9.5 shows the background-only q_0 distribution (left) and the signal + background μ distribution (right) with the 2015 ATLAS data. Even though the observed value for q_0 is significantly above the expected value, the observed μ is still within one standard deviation from the median value.

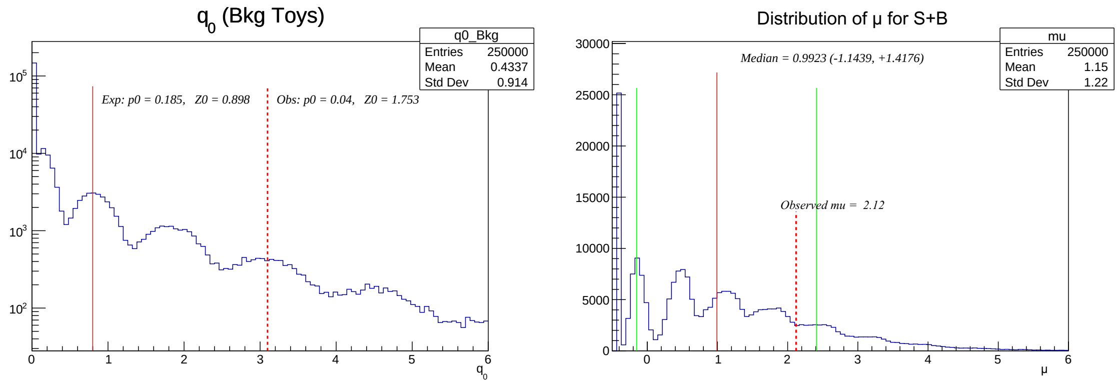


Figure 9.5: Unblinded q_0 distribution for background-only pseudo-experiments (left) and μ distribution for signal + background pseudo-experiments (right).

Chapter 10

Results

The signal region has been optimized to have a high signal to background ratio. This region is used to compare the data yields to the background-only predictions, and to the signal and background predictions. Any statistically significant excess over the background-only prediction in the signal region is interpreted as VBF $H \to WW^* \to \ell\nu\ell\nu$ production. A summary of Higgs boson measurements at the LHC is also presented in this chapter.

10.1 Event Yields

Tables 10.1 and 10.2 summarize the expected yields and those observed in data in the signal region for the Run-I and Run-II analyses respectively.

In the Run-I DF analysis, 57 data events are observed in the signal region while the expected background yield is 52 events. The Run-I DF O_{BDT} distribution is shown in figure 10.1. There is an excess with respect to the background-only yields in the two most sensitive bins of O_{BDT}.

For the Run-I DF O_{BDT} distribution in figure 10.2, the signal region is divided into six bins to show the shape of O_{BDT}.

In the Run-I SF analysis, 73 data events are observed in the entire signal region while the expected background yield is 58 events. The Run-I SF O_{BDT} distribution is shown in figure 10.3.

A total of 129 events are observed in data in the Run-II signal region while the background-only expected yield is 131 events. The Run-II signal region has a larger yield than the Run-I signal region. Due to the lower total integrated luminosity used in the Run-II result, it was necessary to extend the signal region to lower O_{BDT} values in order to ensure sufficient sample sizes in the control regions. Consequently, the first O_{BDT} bin of the Run-II signal region is less signal-like than the first bin in the Run-I signal region.

The shape of the Run-II O_{BDT} distribution in the signal region is shown in figure 10.4. The dashed blue line represents the predicted VBF signal scaled up by a factor of 10.

Run-I DF ($e\mu/\mu e$)

O_{BDT} Bin	Total Signal	Total Background	Data	Top	WW	ggF	Other VV	Z/γ	Fakes
Bin 1	4.4 ± 0.1	42.3 ± 2.2	37	21.1	6.0	3.0	2.6	4.5	5.1
Bin 2	4.3 ± 0.1	7.9 ± 0.9	14	2.6	2.1	1.2	0.8	0.6	0.7
Bin 3	3.1 ± 0.1	1.5 ± 0.2	6	0.4	0.5	0.3	0.1	0.2	0.0

Run-I SF ($ee/\mu\mu$)

O_{BDT} Bin	Total Signal	Total Background	Data	Top	WW	ggF	Other VV	Z/γ	Fakes
Bin 1	2.3 ± 0.1	47.8 ± 5.7	53	13.1	3.5	1.5	0.9	27.8	1.0
Bin 2	2.5 ± 0.1	8.7 ± 1.6	14	1.7	1.1	0.6	0.3	4.8	0.2
Bin 3	1.7 ± 0.1	1.3 ± 0.3	6	0.3	0.3	0.2	0.0	0.6	0.0

Table 10.1: Yields in bins of O_{BDT} in the Run-I DF and SF analyses. The yields for top, W+jets, and Z/γ are derived from their respective control regions. The uncertainties included for the total yields are statistical only.

Run-II DF ($e\mu/\mu e$)

O_{BDT} Bin	Total Signal	Total Background	Data	Top	WW	ggF	Other VV	Z/γ	Fakes
Bin 1	6.0 ± 0.2	127.6 ± 8.0	120	44.9	18.2	8.0	6.7	23.8	25.9
Bin 2	3.1 ± 0.1	3.6 ± 0.7	9	0.9	0.4	0.7	0.2	0.1	1.3

Table 10.2: Yields in bins of O_{BDT} in the Run-II DF analysis. The yields for top, W+jets, and Z/γ are derived from their respective control regions. The uncertainties included for the total yields are statistical only.

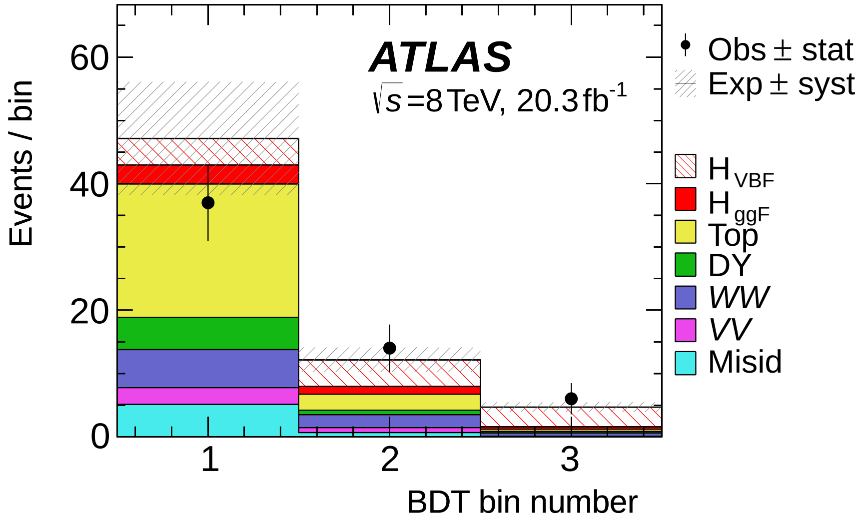


Figure 10.1: O_{BDT} distribution for different flavour in the Run-I signal region [16]. The uncertainty bands shown represent the statistical and systematic uncertainties on the total signal and background yields.

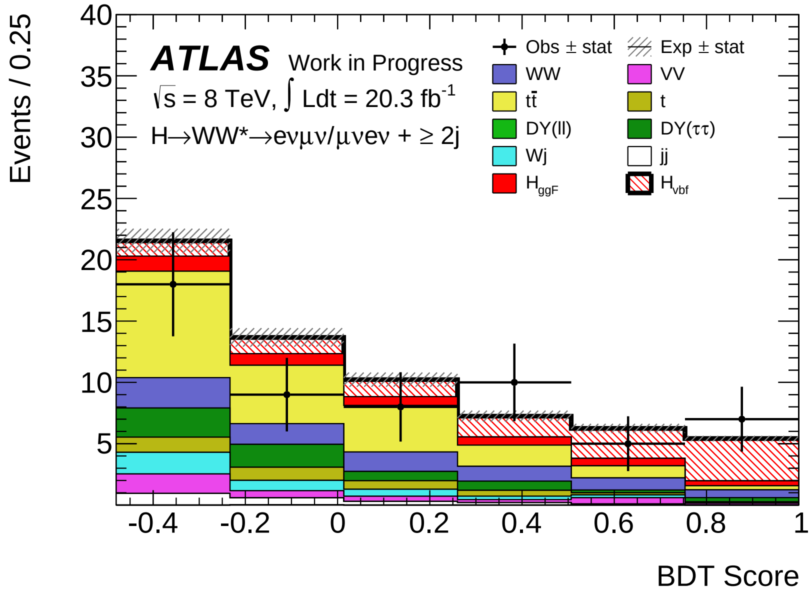


Figure 10.2: O_{BDT} distribution for different flavour in the Run-I signal region binned in six bins of O_{BDT}. The uncertainty bands shown only represent the statistical uncertainties on the total signal and background yields.

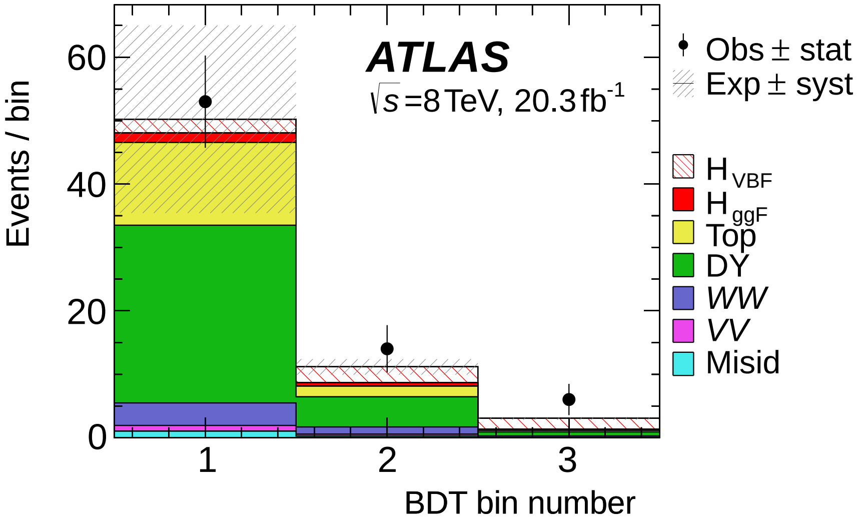


Figure 10.3: O_{BDT} distribution for same flavour in the Run-I signal region [16]. The uncertainty bands shown represent the statistical and systematic uncertainties on the total signal and background yields.

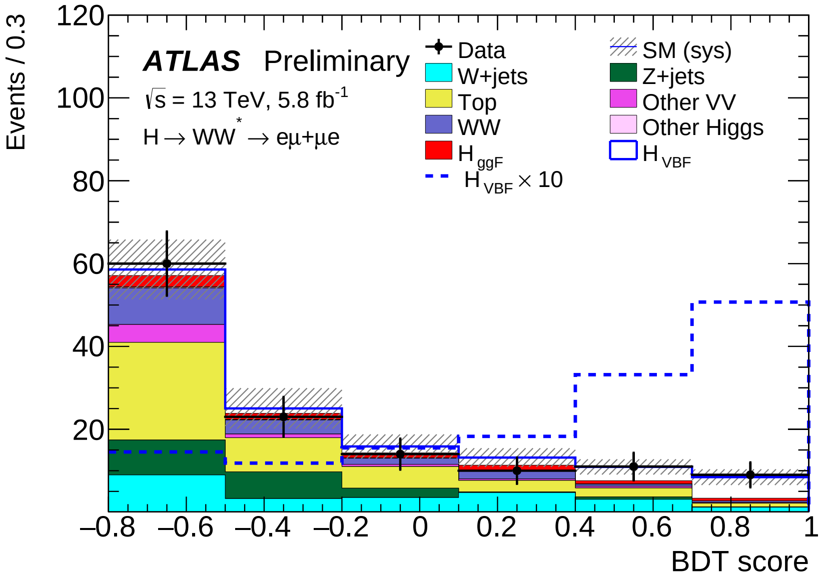
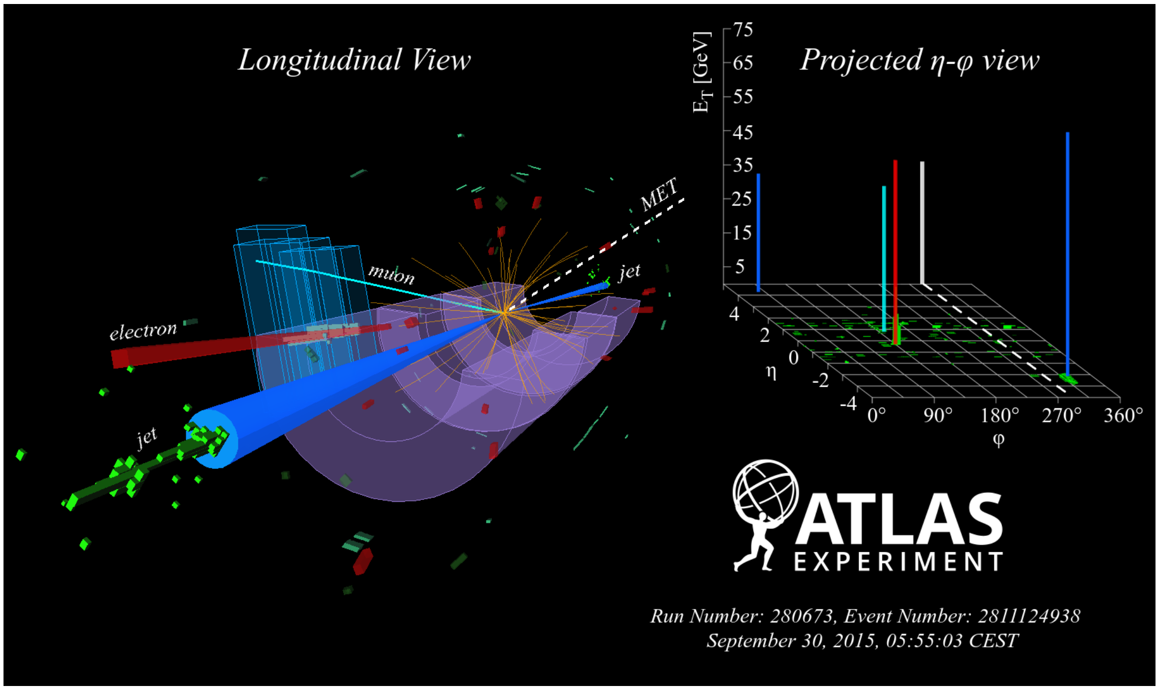


Figure 10.4: O_{BDT} distribution for Run-II in the signal region binned in six bins of O_{BDT} [52]. The uncertainty bands shown represent the statistical and systematic uncertainties on the total signal and background yields.

Figure 10.5 shows the ATLAS event display for a candidate VBF $H \to WW^* \to \ell\nu\ell\nu$ event taken from the 13 TeV Run-II data. As expected, it is characterized by two forward jets and two charged central leptons with a small opening angle balanced by missing transverse energy from the neutrinos. O_{BDT} for this particular event is 0.90 implying that it originates from the second bin of O_{BDT} in the Run-II analysis.



Run Number: 280673, Event Number: 2811124938
September 30, 2015, 05:55:03 CEST

Figure 10.5: Event display for a candidate Run-II VBF $H \to WW^* \to e\nu\mu\nu$ event [52].

Figure 10.6 shows the distributions of the BDT training variables in the Run-I signal region including both DF and SF events. Even though it may be challenging to deduce the presence of a signal in any of these individual distributions, the strength of the BDT is in its ability to combine their discrimination power into a single variable. This is further illustrated in figure 10.7 which shows the distributions for $m_{\ell\ell}$, $\Delta\phi_{\ell\ell}$, Δy_{jj}, m_{jj}, m_T, and the O_{BDT} distribution in the highest two bins of O_{BDT}. In general, the excess is observed in regions where the VBF $H \to WW^*$ process is expected, namely, at low $m_{\ell\ell}$, low $\Delta\phi_{\ell\ell}$, high Δy_{jj}, and high m_{jj}.

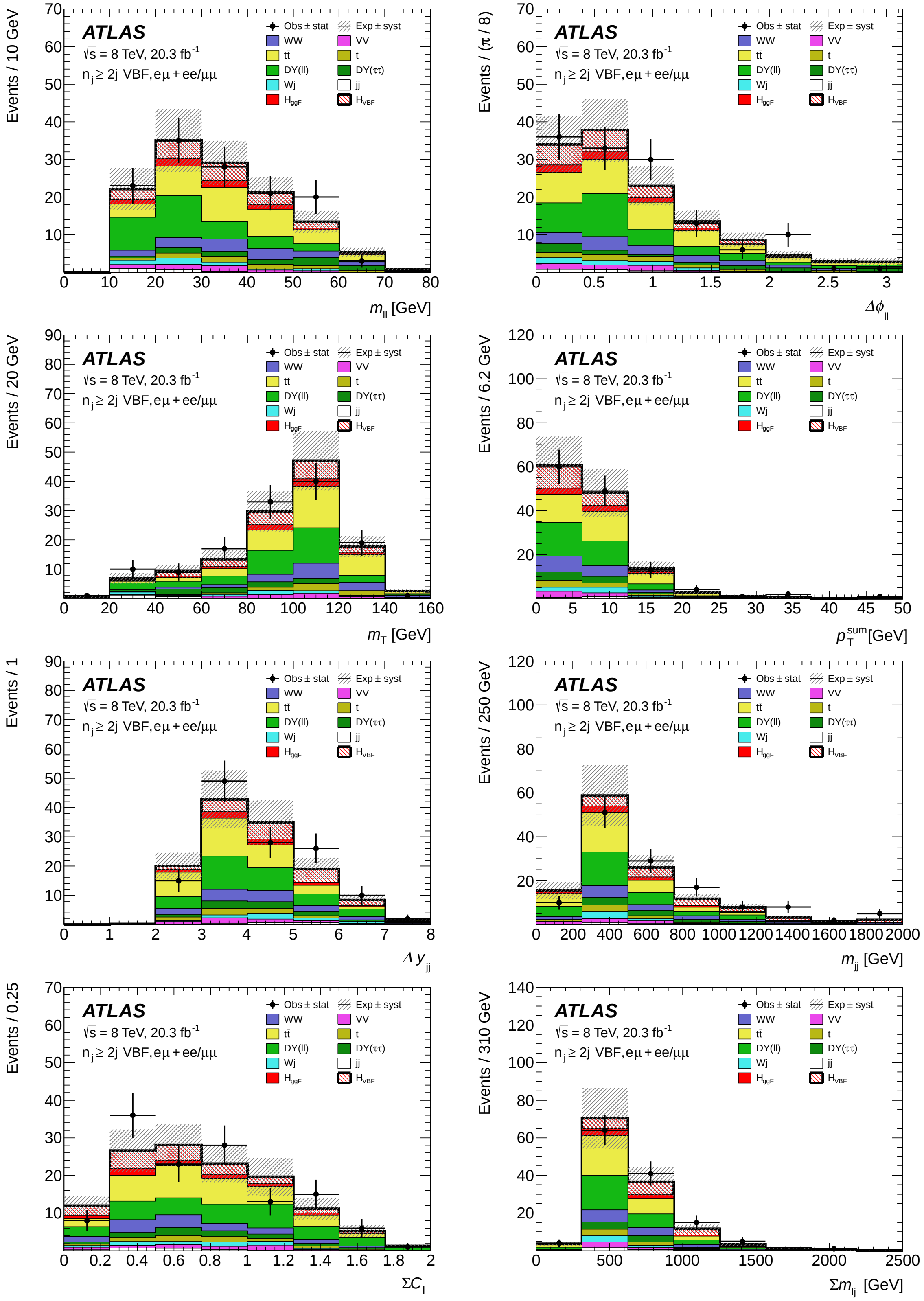


Figure 10.6: Run-I BDT training variable distributions in the signal region including both different flavour and same flavour events [16]. The uncertainty bands shown represent the statistical and systematic uncertainties on the total signal and background yields.

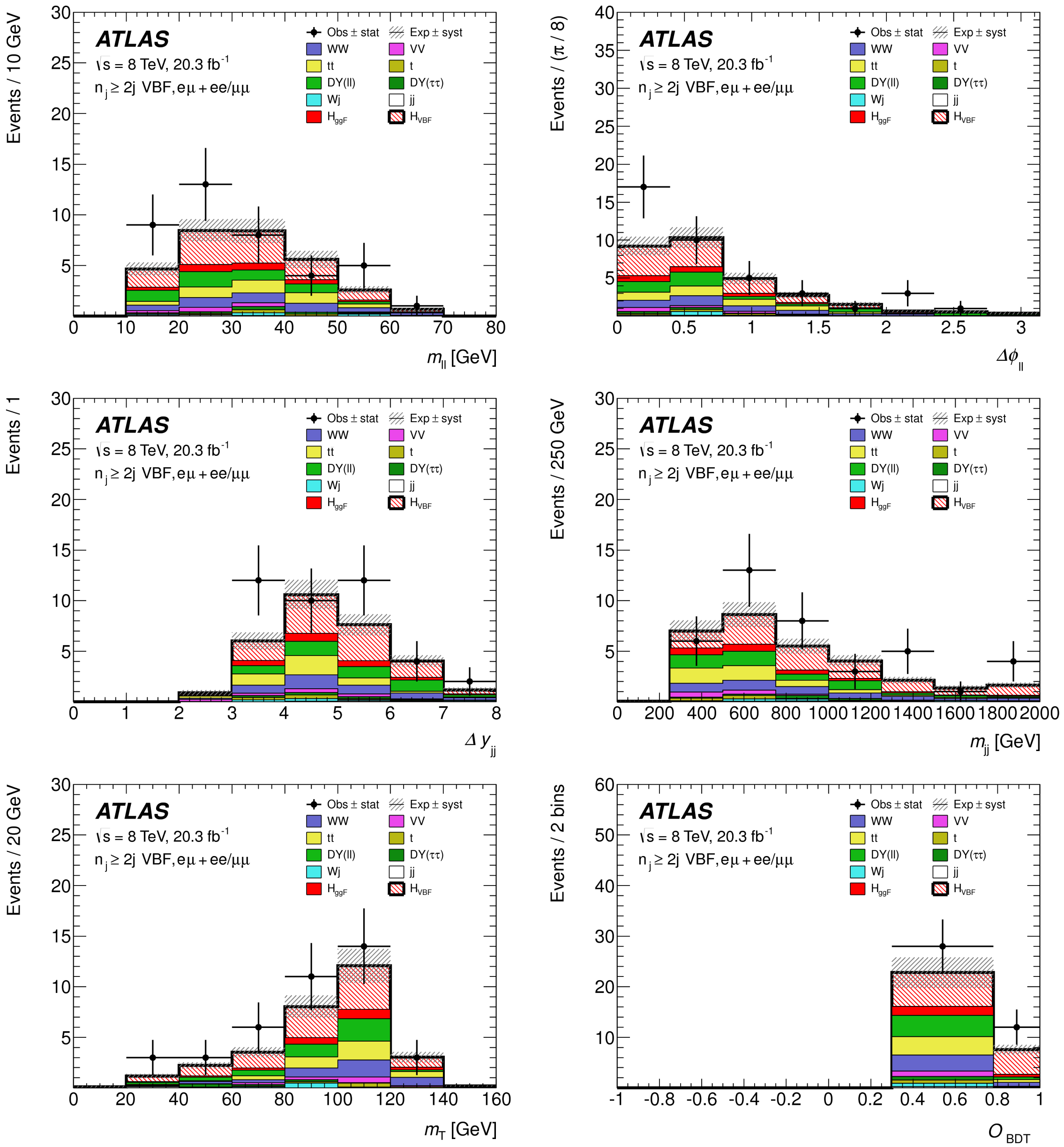


Figure 10.7: Run-I distributions for $m_{\ell\ell}$, $\Delta\phi_{\ell\ell}$, Δy_{jj}, m_{jj}, m_T, and the BDT in the two highest bins of O_{BDT} including different flavour and same flavour events [16]. The uncertainty bands shown represent the statistical and systematic uncertainties on the total signal and background yields.

10.2 Statistical Analysis

The data is analyzed using the statistical framework described in chapter 9.

10.2.1 Run-I Statistical Analysis for the VBF Analysis

The mass of the Higgs boson assumed in the Run-I likelihood minimization is $m_H = 125.36$ GeV which is the measured value in Run-I from the $H \to ZZ^*$ and $H \to \gamma\gamma$ analyses. As previously described, the likelihood minimization is performed for the ggF and VBF analyses simultaneously. Figure 10.8 shows value of the test statistic as a function of $\hat{\mu}_{\text{VBF}}/\hat{\mu}_{gg\text{F}}$ in the likelihood scan. The right axis includes the Gaussian significance which is calculated using equation 9.28.

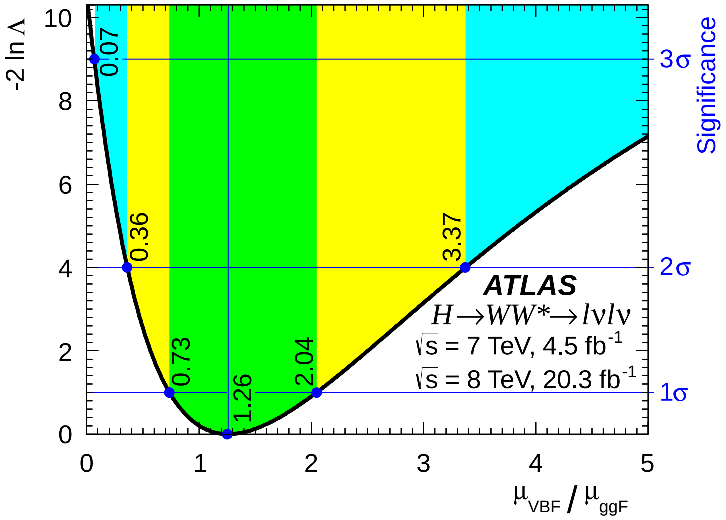


Figure 10.8: Scan of the likelihood as a function of $\hat{\mu}_{\text{VBF}}/\hat{\mu}_{gg\text{F}}$ [16]. The green, yellow, and cyan shaded areas correspond to the 1, 2, and 3 standard deviations away from the minimum.

In order to test for the presence of a signal, the compatibility of the data with the background-only hypothesis, $\hat{\mu}_{\text{VBF}} = 0$, is tested. This is equivalent to testing $\frac{\hat{\mu}_{\text{VBF}}}{\hat{\mu}_{gg\text{F}}} = 0$. The resulting significance is:

$$Z_0^{\text{obs}} = 3.2 \quad , \quad Z_0^{\text{exp}} = 2.7 \tag{10.1}$$

where Z_0^{obs} is the observed significance and Z_0^{exp} is the expected significance assuming the SM signal strength for the VBF Higgs boson production. Z_0^{exp} is calculated from q_0^{exp} using equation 9.28. The objective of an analysis is to maximize this quantity as much as possible as Z_0^{exp} effectively quantifies the sensitivity of the analysis. This result constitutes the first evidence for the $H \to WW^* \to \ell\nu\ell\nu$ process.

As described in detail in reference [16], a selection based analysis (also referred to as a "cut-based" analysis) was also performed using the same dataset. A signal region is defined using a sequence of selection requirements on a subset of the kinematic variables used in the BDT analysis. Using m_T as discriminant, the significance of this cross-check analysis is:

$$Z_0^{\text{obs}} = 3.0 \quad , \quad Z_0^{\text{exp}} = 2.1 \tag{10.2}$$

The BDT offers nearly 30% improvement in the expected significance over the selection based analysis.

Next, the minimum value of the test statistic in figure 10.8 is used to determine the best fit value of the ratio of the VBF to ggF Higgs boson signal strength:

$$\frac{\hat{\mu}_{\text{VBF}}}{\hat{\mu}_{gg\text{F}}} = 1.3^{+0.6}_{-0.5}\,(\text{stat.})\,^{+0.5}_{-0.3}\,(\text{syst.}) = 1.3^{+0.8}_{-0.5} \tag{10.3}$$

The signal strengths $\hat{\mu}_{\text{VBF}}$ and $\hat{\mu}_{gg\text{F}}$ are consistent with each-other within their uncertainties.

Figure 10.9 shows the likelihood plotted in two dimensions as a function of $\hat{\mu}_{\text{VBF}}$ and $\hat{\mu}_{gg\text{F}}$. The measured values for $\hat{\mu}_{\text{VBF}}$ and $\hat{\mu}_{gg\text{F}}$ are extracted from this plot:

$$\hat{\mu}_{\text{VBF}} = 1.3 \pm 0.4\,(\text{stat.})\,^{+0.3}_{-0.2}\,(\text{syst.}) = 1.3 \pm 0.5 \tag{10.4}$$

$$\hat{\mu}_{\text{ggF}} = 1.0 \pm 0.2\,(\text{stat.})\ \pm 0.2\,(\text{syst.}) = 1.0 \pm 0.3 \tag{10.5}$$

Note that the uncertainty on $\hat{\mu}$ for the VBF analysis is driven by statistical uncertainties. Collecting additional data is vital to improving the sensitivity to this production mode.

Next, given the values of $\hat{\mu}_{\text{VBF}}$ and $\hat{\mu}_{gg\text{F}}$, it is possible to test the SM predictions for the coupling of the Higgs boson to fermions and to bosons [17]. As mentioned in chapter 1.4, κ_V and κ_F, correspond to the bosonic and fermionic couplings of the Higgs boson with respect to SM expectations. The likelihood is shown as a function of κ_V and κ_F in figure 10.10. The values extracted from the minimization are:

$$\kappa_V = 1.04^{+0.07}_{-0.08}\,(\text{stat.})\,^{+0.07}_{-0.08}\,(\text{syst.}) = 1.04 \pm 0.11 \tag{10.6}$$

$$\kappa_F = 0.9 \pm 0.2\,(\text{stat.})\,^{+0.2}_{-0.1}\,(\text{syst.}) = 0.9^{+0.3}_{-0.2} \tag{10.7}$$

The observed values for κ_V and κ_F are in good agreement with the SM expectations.

Finally, it is possible to calculate $\sigma \cdot \mathcal{B}_{H\to WW^*}$, where σ is the Higgs boson production cross section and $\mathcal{B}_{H\to WW^*}$ is the branching ratio to $H \to WW^*$. For this measurement, the theoretical uncertainties related to the total production yield are removed from the likelihood minimization as they do not apply in this measurement. The value for $\sigma\cdot\mathcal{B}_{H\to WW^*}$ is calculated as the product of μ and the SM prediction of the Higgs boson production cross

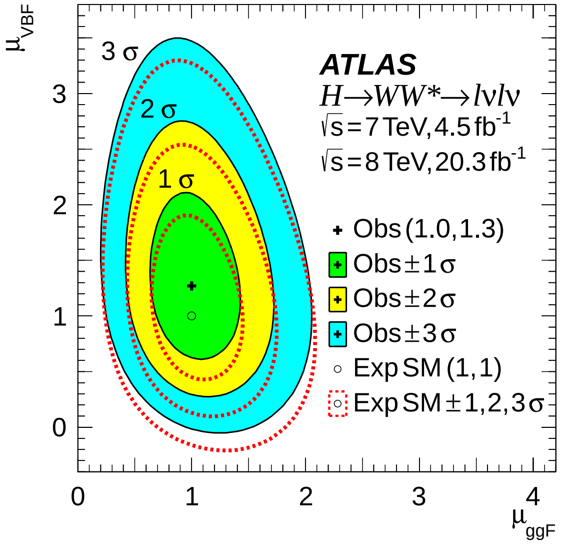


Figure 10.9: Two dimensional scan of the likelihood as a function of $\hat{\mu}_{\mathrm{VBF}}$ and $\hat{\mu}_{ggF}$ [16]. The observed minimum is denoted by a plus, and the expected minimum is denoted by a circle. The green, yellow, and cyan shaded areas correspond to the 1, 2, and 3 standard deviations away from the observed minimum, while the dotted red lines corresponds tot the 1, 2, and 3 standard deviations away from the expected minimum.

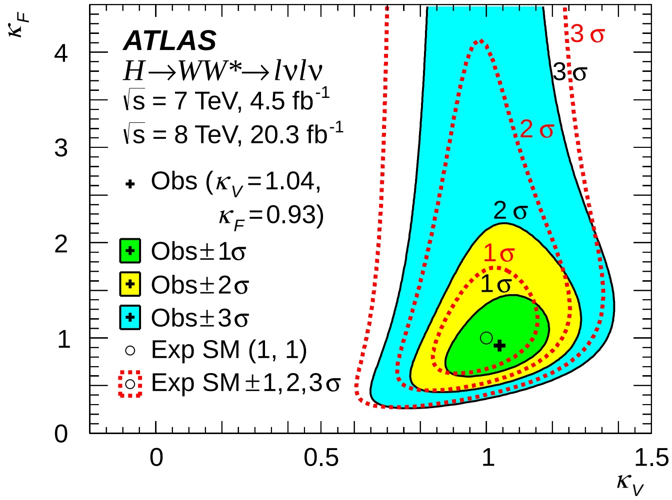


Figure 10.10: Two dimensional scan of the likelihood as a function of κ_V and κ_F [16]. The observed minimum is denoted by a plus, and the expected minimum is denoted by a circle. The green, yellow, and cyan shaded areas correspond to the 1, 2, and 3 standard deviations away from the observed minimum, while the dotted red lines corresponds tot the 1, 2, and 3 standard deviations away from the expected minimum.

section that is used in the analysis. The measured cross sections are:

$$\sigma_{\text{VBF}} \cdot \mathcal{B}_{H\to WW^*} = 0.5 \pm 0.2\ (\text{stat.})\ \pm 0.1\ (\text{syst.})\ \text{pb} = 0.5 \pm 0.2\ \text{pb} \tag{10.8}$$

$$\sigma_{gg\text{F}} \cdot \mathcal{B}_{H\to WW^*} = 4.6 \pm 0.9\ (\text{stat.})\ {}^{+0.8}_{-0.7}\ (\text{syst.})\ \text{pb} = 4.6^{+1.2}_{-1.1}\ \text{pb} \tag{10.9}$$

These measurements agree with the predicted values from theory for Run-I:

$$\sigma_{\text{VBF}} \cdot \mathcal{B}_{H\to WW^*} = 0.35 \pm 0.02\ \text{pb} \tag{10.10}$$

$$\sigma_{gg\text{F}} \cdot \mathcal{B}_{H\to WW^*} = 3.3 \pm 0.4\ \text{pb} \tag{10.11}$$

Table 10.3 summarizes the 10 most important nuisance parameters in the likelihood minimization ordered by their impact on the uncertainty on $\hat{\mu}$. The uncertainties are derived by removing the ggF analysis from the minimization and only considering the VBF analysis. The total uncertainty is dominated by the statistical uncertainty on the data which implies that the analysis will benefit from more data.

The leading source of systematic uncertainty originates from the η modelling for jets. This uncertainty arises from the extrapolation of the jet energy scale from the central region of the detector to the forward regions where jets from VBF Higgs boson events are expected. The uncertainty from the parton showering and underlying event in VBF Higgs boson production has approximately the same impact on μ. O_{BDT} is particularly sensitive to this uncertainty since one of the training variables, p_T^{sum}, relies on the modelling of low p_T jet activity. This is especially relevant for VBF Higgs boson production since the two quarks radiating the weak vector bosons typically have large momentum, increasing the probability of them emitting gluons.

Figures 10.11 and 10.12 show the post-fit O_{BDT} distributions in the signal region for DF and SF respectively. Post-fit plots are created by scaling all nuisance parameters to their best fit values from the likelihood minimization.

10.2.2 Overview of the Run-I ggF Analysis

Since the statistical analysis for the ggF and VBF analyses is done simultaneously, a brief overview of the ggF analysis is given here. A complete description can be found in reference [16].

As shown in the schematic overview in figure 4.1, the ggF analysis considers candidate events with 0 jets, 1 jet, and ≥ 2 jets in the final state. Since this Higgs boson production mode is not expected to have jets at leading order, the 0 jet category offers the highest sensitivity. As the selected number of jets is increased, the sensitivity decreases.

Unlike in the VBF analysis, the jets from ggF Higgs boson candidate events are not expected to have a distinct signature in the forward regions of the detector. The bulk of the

Source	$\Delta\mu^{+\sigma}_{\mathrm{VBF}}/\mu_{\mathrm{VBF}}$	$\Delta\mu^{-\sigma}_{\mathrm{VBF}}/\mu_{\mathrm{VBF}}$
Statistical	+0.43	-0.39
Jet energy scale η model	+0.12	-0.08
VBF PS/UE	+0.12	-0.07
μ_{ggF}	+0.09	-0.08
Branching ratio of $H \to WW^*$	+0.09	-0.05
Non-closure of $Z \to ee/\mu\mu$ estimate	+0.06	-0.05
Luminosity	+0.06	-0.04
WW QCD scale	+0.05	-0.05
α extrapolation parameter for top	+0.05	-0.05
VBF matrix element model	+0.07	-0.04
Total statistical uncertainty	+0.43	-0.39
Total systematic uncertainty	+0.31	-0.22
Total uncertainty	+0.53	-0.45

Table 10.3: The 10 most important nuisance parameters in the Run-I analysis ranked by their impact on the uncertainty on $\hat{\mu}$. This ranking was derived by only considering the VBF analysis and leaving out the ggF analysis.

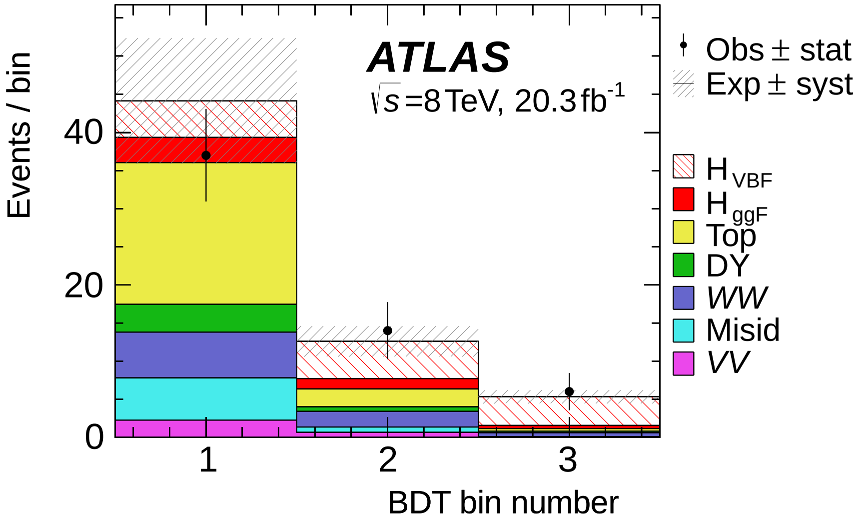


Figure 10.11: Post-fit O_{BDT} distribution for different flavour in the Run-I signal region [16]. The uncertainty bands shown represent the statistical and systematic uncertainties on the total signal and background yields.

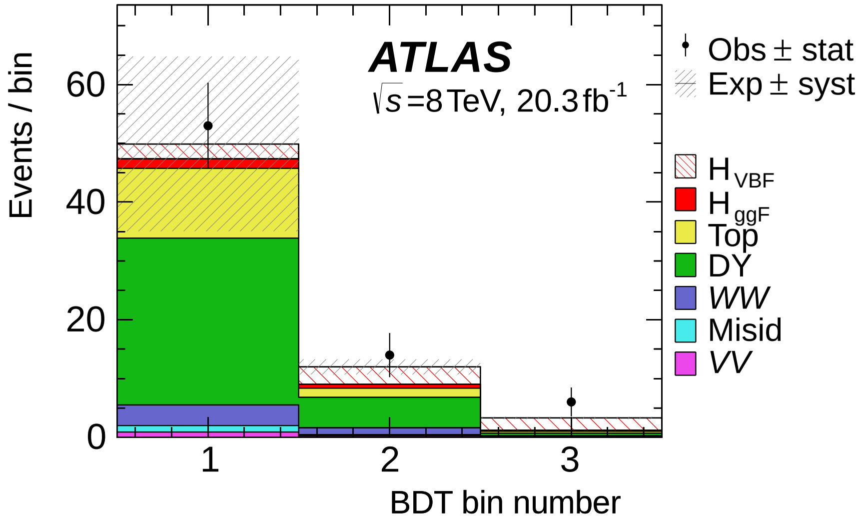


Figure 10.12: Post-fit O_{BDT} distribution for same flavour in the Run-I signal region [16]. The uncertainty bands shown represent the statistical and systematic uncertainties on the total signal and background yields.

discrimination power of signal from background comes from the lepton kinematics and the missing transverse energy. Since the number of variables with significant separation power is smaller in the ggF analysis, a multivariate discriminant will not provide the same gain over a selection based analysis as in the VBF analysis. Consequently, the ggF analysis does not use a multivariate analysis technique like the BDT used in the VBF analysis.

The dominant background in the ggF analysis is WW. It is modelled using MC with data driven corrections from a control region. Additional control regions are defined for the Drell-Yan, diboson (other than WW), top quark, and W+jet backgrounds. The remaining backgrounds are estimated purely from MC predictions.

The discriminating variable used is the transverse mass, m_T. In the statistical analysis, it is binned in dilepton mass ($m_{\ell\ell}$) and subleading lepton p_T ($p_T^{\ell 2}$). Figure 10.13 shows the different flavour m_T distributions used in the statistical analysis for the 0 and 1 jet channels.

Figure 10.14 shows the combined m_T distribution for all candidate events with less than 2 jets. The right plot shows the m_T distribution in data with the background prediction subtracted. The superimposed red histogram corresponds to the m_T distribution for ggF as expected from MC.

Assuming a mass of 125.36 GeV, the significance of measuring ggF Higgs boson production for Run-I is:

$$Z_0^{\mathrm{obs}} = 6.1 \quad , \quad Z_0^{\mathrm{exp}} = 5.8 \tag{10.12}$$

This is the first standalone observation of the $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ process at the ATLAS experiment.

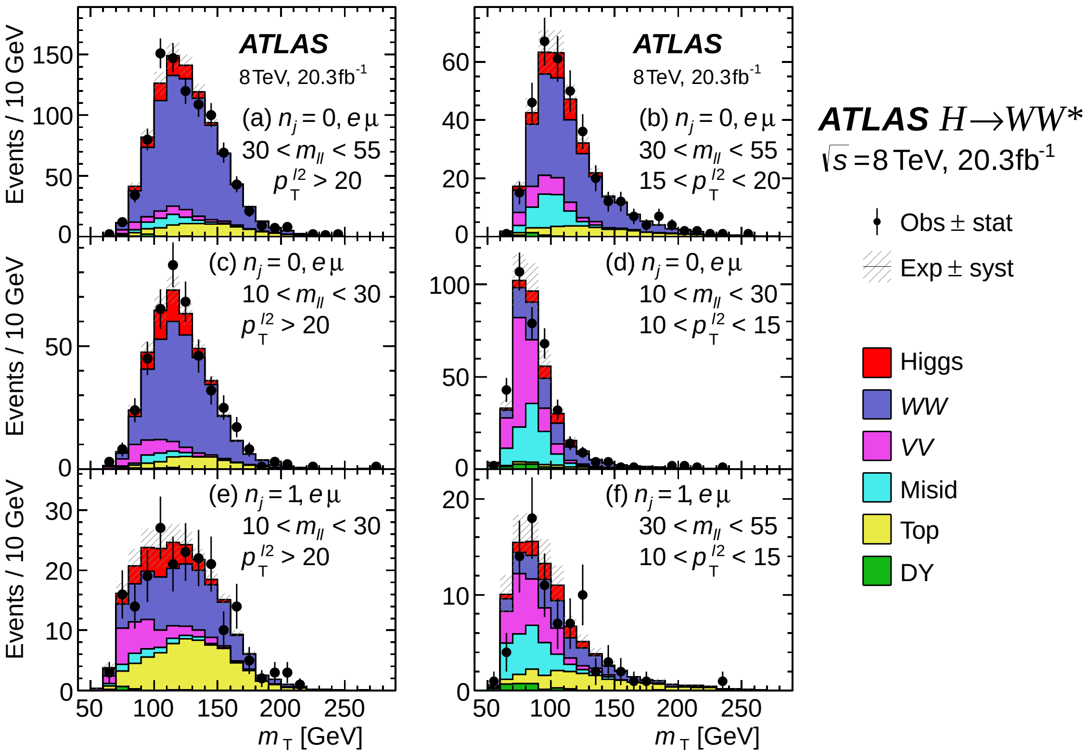


Figure 10.13: Post-fit m_T distributions binned in $m_{\ell\ell}$ and $p_T^{\ell 2}$ for the different flavour ggF analysis in the 0 and 1 jet channels [16]. The uncertainty bands shown represent the statistical and systematic uncertainties on the total signal and background yields.

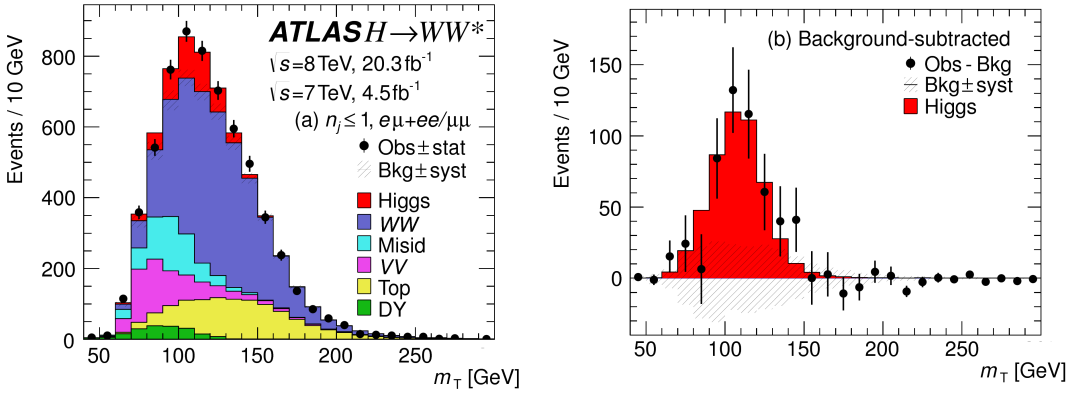


Figure 10.14: The plot on the left shows the combined post-fit m_T distribution for the ggF analysis for all channels with less than 2 jets [16]. The plot on the right shows the expected m_T distribution in the data after all backgrounds have been subtracted. The red histogram corresponds to the m_T distribution for ggF as predicted by the MC. The uncertainty bands shown represent the statistical and systematic uncertainties on the total background yields.

10.2.3 Run-II Statistical Analysis for the VBF Analysis

Assuming a mass of 125 GeV, the significance of measuring VBF Higgs boson production for Run-II is:

$$Z_0^{\text{obs}} = 1.9 \quad , \quad Z_0^{\text{exp}} = 1.2 \tag{10.13}$$

Note that even though the Run-II result corroborates the Run-I result, it does not significantly add to its sensitivity. Consequently, a statistical combination of the two results is not performed at this time. However, as the Run-II sensitivity becomes comparable to that of Run-I with the accumulation of more data, a combination of the Run-I and Run-II results may be in order to establish the first observation ($Z_0 = 5$) of VBF $H \to WW^* \to \ell\nu\ell\nu$ at the ATLAS experiment.

The parameter of interest being measured in Run-II is μ_{VBF}. The best fit value for the signal strength is:

$$\hat{\mu}_{\text{VBF}} = 1.7^{+0.8}_{-0.8}\,(\text{stat.})\,^{+0.6}_{-0.4}\,(\text{syst.}) = 1.7^{+1.1}_{-0.9} \tag{10.14}$$

Similar to the Run-I result, the measured signal strength is larger than the SM expectation, but consistent within uncertainties.

The measured value for the total production cross section is:

$$\sigma_{\text{VBF}} \cdot \mathcal{B}_{H\to WW^*} = 1.4^{+0.8}_{-0.6}\,(\text{stat.})\,^{+0.5}_{-0.4}\,(\text{syst.})\text{ pb} = 1.4^{+0.9}_{-0.7}\text{ pb} \tag{10.15}$$

This measurement agrees with the predicted theoretical value for Run-II:

$$\sigma_{\text{VBF}} \cdot \mathcal{B}_{H\to WW^*} = 0.81 \pm 0.02\text{ pb} \tag{10.16}$$

The 10 most important nuisance parameters in the likelihood minimization are summarized in table 10.4. As before, they are ordered by their impact on the uncertainty of $\hat{\mu}$. Again the uncertainty is driven by the statistical uncertainty on the data. The second most important systematic uncertainty originates from the uncertainty on the fake factor in the W+jets estimate. The size of this systematic uncertainty is also largely driven by a lack of statistics in the region where this uncertainty is evaluated.

The Run-II post-fit O_{BDT} training variable distributions are shown in figure 10.15. Finally, figure 10.16 shows a post-fit summary of the Run-II analysis. It includes the two controls regions, as well as the two O_{BDT} bins in the signal region.

10.3 Status of Higgs Boson Measurements

The original ATLAS publication on the discovery of a new particle consistent with the SM Higgs boson in 2012 was based on 4.8 fb^{-1} of 7 TeV data and 5.8 fb^{-1} of 8 TeV data. The quoted significance for ggF Higgs boson production of 5.9 standard deviations is

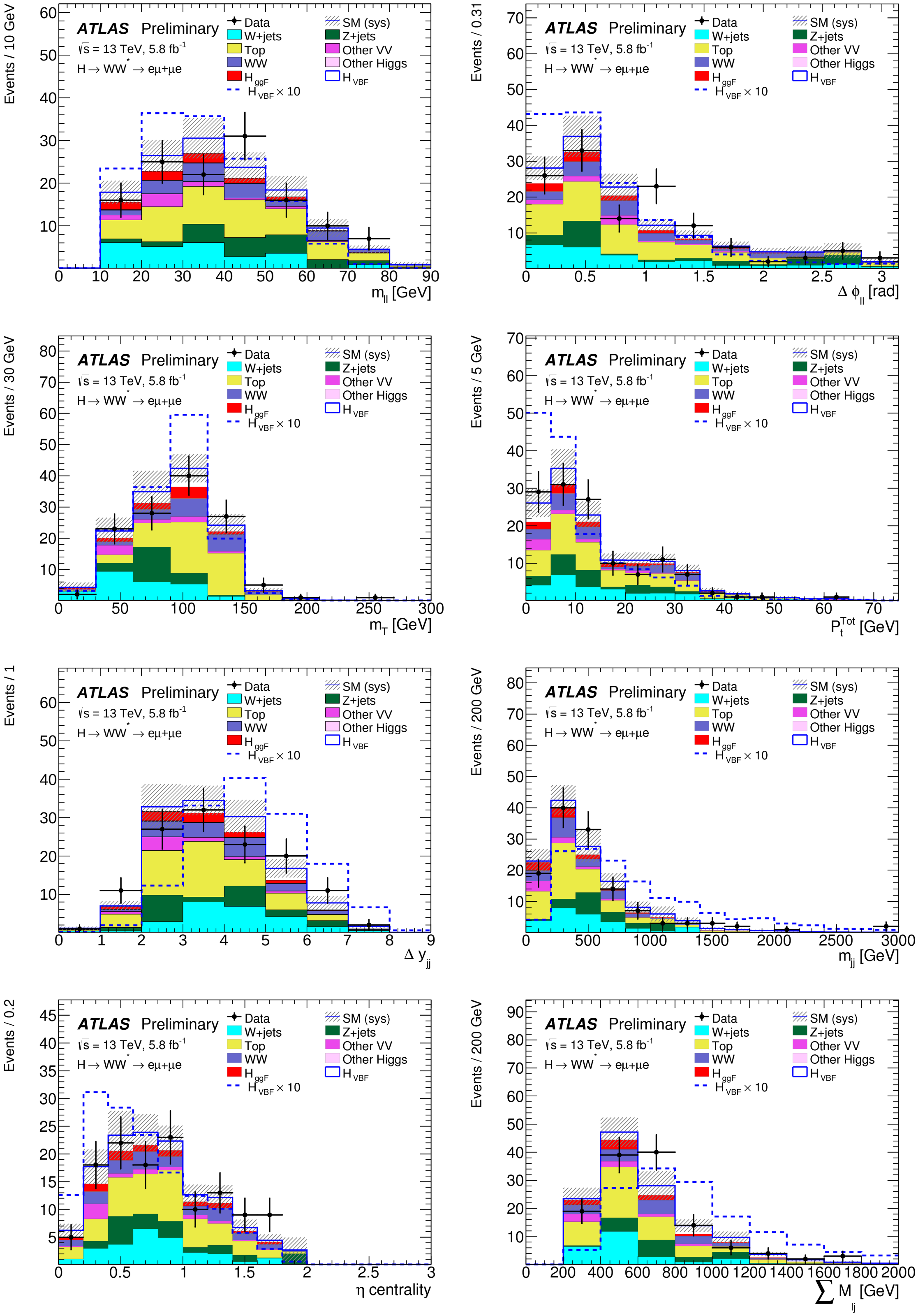


Figure 10.15: Post-fit BDT training variable distributions for DF in the Run-II signal region [52]. The uncertainty bands shown represent the statistical and systematic uncertainties on the total signal and background yields.

Source	$\Delta\mu^{+\sigma}_{\mathrm{VBF}}/\mu_{\mathrm{VBF}}$	$\Delta\mu^{-\sigma}_{\mathrm{VBF}}/\mu_{\mathrm{VBF}}$
Statistical	+0.57	-0.47
W+jet factor fake (sample composition)	+0.18	-0.15
MC statistical	+0.15	-0.15
VBF matrix element model	+0.14	-0.05
WW matrix element model	+0.11	-0.07
ggF QCD scale ($N_{jet} \geq 3$)	+0.08	-0.07
Jet energy resolution	+0.08	-0.07
b-tagging	+0.08	-0.06
Pile-up	+0.08	-0.06
ggF QCD scale ($N_{jet} \geq 2$)	+0.06	-0.06
Total statistical uncertainty	+0.57	-0.47
Total systematic uncertainty	+0.33	-0.26
Total uncertainty	+0.66	-0.54

Table 10.4: The 10 most important nuisance parameters in the Run-II analysis ranked by their impact on the uncertainty on $\hat{\mu}$.

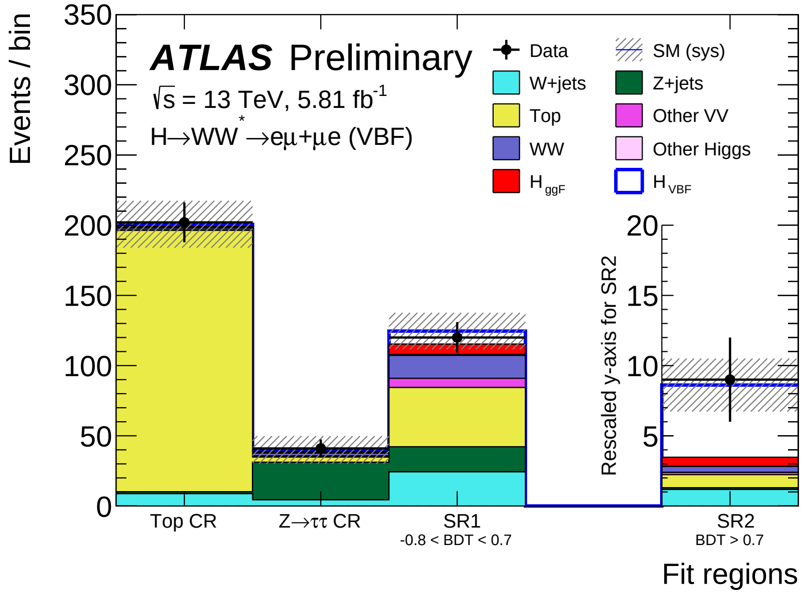


Figure 10.16: Post-fit yields for the Run-II different flavour analysis in the control regions and the two bins of the signal region [52]. The uncertainty bands shown represents the statistical and systematic uncertainties on the total signal and background yields. Note that the second bin of O_{BDT} uses a different scale.

driven by the three most sensitive search channels, $H \to \gamma\gamma$, $H \to ZZ^*$, and $H \to WW^*$. More recently in 2016, ATLAS published a paper combining the $H \to \gamma\gamma$, $H \to ZZ^*$, $H \to WW^*$, $H \to Z\gamma$, $H \to b\bar{b}$, $H \to \tau\tau$ and $H \to \mu\mu$ Higgs boson results using the complete Run-I dataset from 2011-2012. With the additional data and improved analysis techniques, the ATLAS experiment sees evidence for VBF Higgs boson production with an observed significance of 4.3σ (expected significance is 3.8σ). When the Higgs boson search results from ATLAS and CMS are combined [102], the observed significance for VBF is 5.4σ (expected significance is 4.6σ) which corresponds to the first observation of VBF Higgs boson production.

The ATLAS combination of the above mentioned channels is also used to determine the signal strength. Figure 10.17 summarizes the signal strength measurements for each of these analyses. The combined signal strength of all these measurements is $\mu = 1.18^{+0.15}_{-0.14}$, where the uncertainty includes both statistical and systematic uncertainties. This agrees with the SM predictions with a p-value of 0.18 and so does not provide obvious hints of new physics impacting the Higgs boson cross sections.

For the individual Higgs boson production channels, the ATLAS combination yields $\mu_{gg\mathrm{F}} = 1.23^{+0.23}_{-0.20}$ and $\mu_{\mathrm{VBF}} = 1.23 \pm 0.32$ which agrees with the SM expectations. Figure 10.18 summarizes the signal strength measurements for the Higgs boson coupling to fermions (ggF and ttH) and the Higgs boson coupling to the vector bosons (VBF and VH) assuming $m_H = 125.36$ GeV. The relative contributions between ggF and ttH and between VBF and VH are predicted from the SM. The black star corresponds to the SM prediction, the colored +'s correspond to the best fit values of various Higgs boson search channels, while the solid and dotted contours correspond to the 68% and 95% confidence level intervals.

In order to test if the newly observed particle is consistent with the SM Higgs boson, efforts are also made to measure its properties. Using the Run-I ATLAS data, the combination of the $H \to \gamma\gamma$, $H \to ZZ^* \to \ell\ell\ell\ell$, and $H \to WW^* \to \ell\nu\ell\nu$ analyses show a strong preference for a spin-0 Higgs boson with positive parity [104] [105] which agrees with the SM expectations. The spin-0 nature is especially important to the $H \to WW^* \to \ell\nu\ell\nu$ analysis as described in section 1.4.

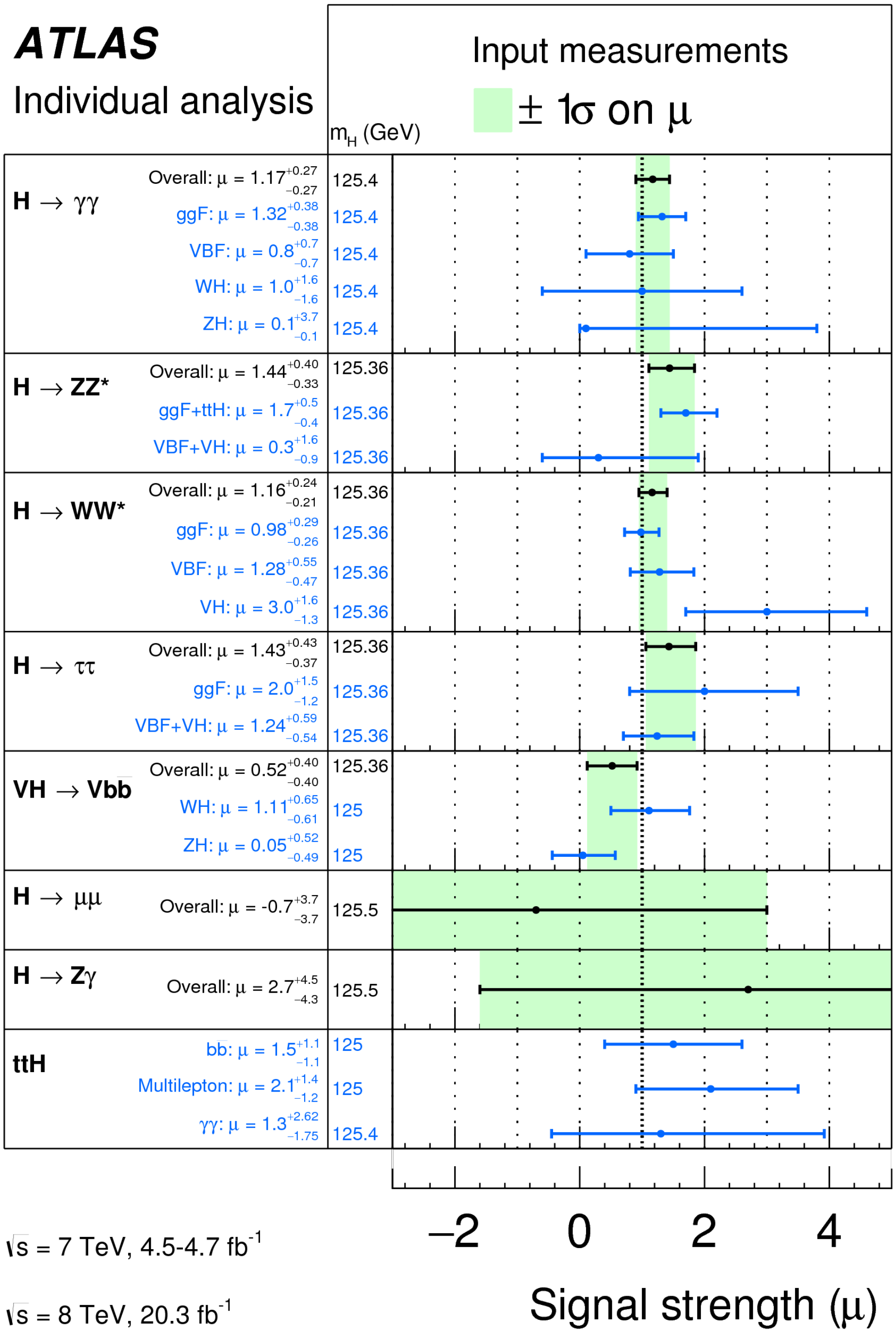


Figure 10.17: Summary of the Higgs boson signal strength measurements at the ATLAS experiment [103]. The uncertainties shown include both statistical and systematic uncertainties.

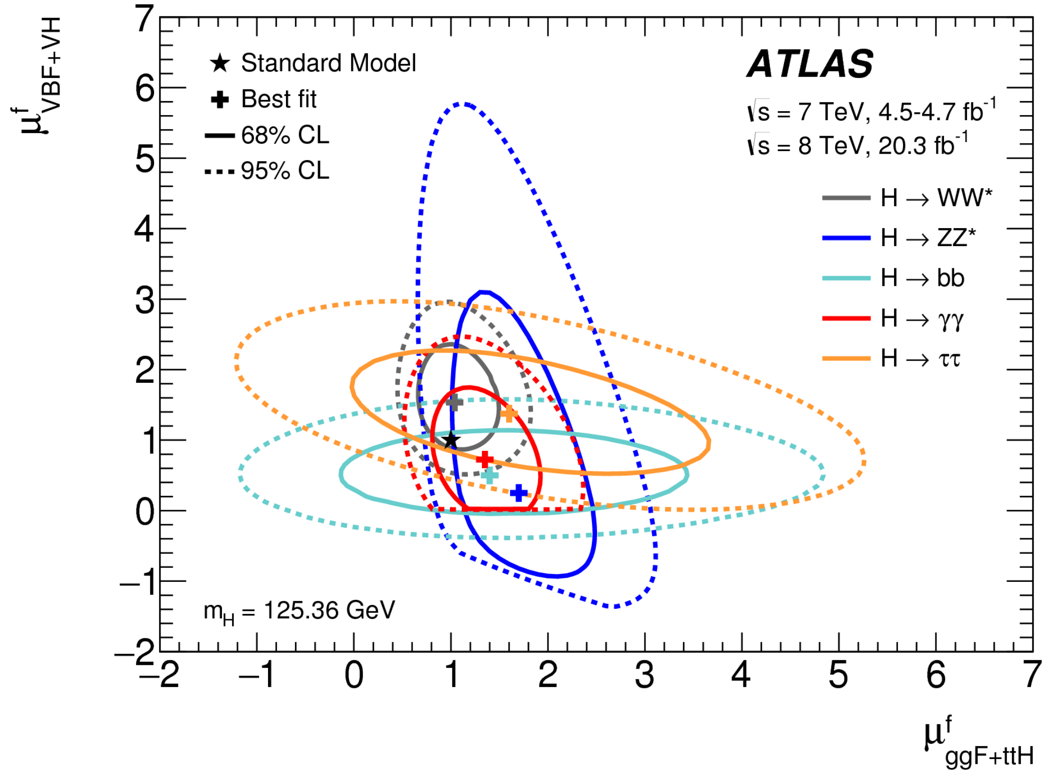


Figure 10.18: Summary of the signal strength measurements at the ATLAS experiment [103] for the Higgs boson coupling to bosons (y-axis) and fermions (x-axis). The black star corresponds to the SM predictions, and the plus symbols correspond to the measurements from 5 different analyses.

Chapter 11

Conclusions

The discovery of a new particle consistent with the SM Higgs boson was a major breakthrough for the particle physics community. Precisely measuring the particle's properties is one of the highest priorities of the LHC experiments. A search for the VBF $H \to WW^* \to \ell\nu\ell\nu$ process using the ATLAS detector has been presented in this thesis. Measuring this new process adds crucial information about the coupling of the Higgs boson to the weak vector bosons. Any deviation from the SM predictions would be a sign of new physics beyond the SM.

Measuring VBF Higgs boson production is challenging given its cross section is an order of magnitude smaller than that of ggF. In addition, the $H \to WW^* \to \ell\nu\ell\nu$ channel is complicated by the neutrinos in the final state which are not directly detectable by the ATLAS detector. There are also several large backgrounds such as top quark pair production that have a similar signature in the detector as the signal. Consequently, a multivariate analysis technique based on an eight-variable boosted decision tree (BDT) has been optimized to search for the VBF $H \to WW^* \to \ell\nu\ell\nu$ process.

Using 20.3 fb^{-1} of 8 TeV data collected in Run-I, the search resulted in an observed significance of 3.2σ while the expected significance is 2.7σ. This result constitutes the first evidence for the VBF $H \to WW^* \to \ell\nu\ell\nu$ process. The signal strength is measured to be $\mu_{\text{VBF}} = 1.3 \pm 0.5$.

Using 5.8 fb^{-1} of 13 TeV data collected in Run-II, the observed significance is 1.9σ while the expected significance is 1.2σ. The signal strength is $\mu_{\text{VBF}} = 1.7^{+1.1}_{-0.9}$. Note that the Run-II results quoted here are based on a small fraction of the total expected Run-II dataset.

At the time of writing, the significance for the VBF $H \to WW^* \to \ell\nu\ell\nu$ process at ATLAS does not reach the 5σ threshold required to establish a discovery of this process by a single experiment. As more data is collected and analyzed, this objective will become more and more reachable particularly since the current analyses are limited by the available data sample sizes.

The VBF $H \to WW^* \to \ell\nu\ell\nu$ process is a powerful channel to measure κ_V, the coupling strength of the Higgs boson to the weak vector bosons. Using the Run-I data, the measured κ_V from the VBF and ggF analyses is consistent with the SM predictions within the measurement uncertainties. Since the statistical and systematic uncertainties on μ are similar in size in the Run-I ggF analysis, the systematic uncertainties will dominate as more data is collected. In contrast, the Run-I VBF analysis is dominated by statistical uncertainties. Consequently, the VBF analysis will be increasingly more important in the determination of κ_V as more data is collected, particularly with its quartic relationship to μ_{WW}^{VBF}.

There are a number of ways in which the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis can be improved. As mentioned above, the sensitivity of both the Run-I and Run-II analyses is limited by the available data sample sizes. Consequently, collecting and analyzing more data is vital to the success of this analysis. As the data set grows, the BDT can be reoptimized to define new regions of phase space with larger signal to background ratios. In parallel, the MC statistics will also need to be increased to sufficiently populate these regions. VBF-filtered samples have been developed and provide a promising method to generate MC samples with low statistical uncertainty to estimate the backgrounds, evaluate systematic uncertainties, and train the BDT.

Run-II data taking at the LHC is still in progress and is expected to deliver a total integrated luminosity of 150 fb^{-1} by the end of 2018. A series of upgrades are planned for the LHC which will increase the instantaneous luminosity. The ATLAS detector will be upgraded in parallel, allowing it to collect data more quickly and perform more precise measurements of the Higgs boson. The *Phase 1* upgrades are scheduled to commence after the conclusion of Run-II and will allow the LHC to deliver over 300 fb^{-1} by the end of 2023. The *Phase 2* upgrades will mark the beginning of the High Luminosity LHC (HL-LHC) which is expected to deliver 3,000 fb^{-1} sometime around 2035. An average of 200 inelastic proton-proton collisions are expected per bunch crossing at the HL-LHC. This will unfortunately lead to more jets from pile-up, an effect that will degrade the purity of the VBF $H \to WW^* \to \ell\nu\ell\nu$ signal region. The proposed new ATLAS Inner Tracker (ITk) will have an extended rapidity coverage which will be instrumental in rejecting jets from pile-up. The large expected dataset at the HL-LHC will make it possible to perform precision measurements of the VBF $H \to WW^* \to \ell\nu\ell\nu$ process.

Bibliography

[1] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters*, B716:1–29, 2012.

[2] The CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters*, B716:30–61, 2012.

[3] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[4] C. Patrignani, and others. Review of Particle Physics. *Chinese Physics*, C40(10):100001, 2016.

[5] The SNO Collaboration. Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory. *Physical Review Letters*, 89:011301, 2002.

[6] The Super-Kamiokande Collaboration. Evidence for Oscillation of Atmospheric Neutrinos. *Physical Review Letters*, 81:1562–1567, 1998.

[7] A. Salam. Weak and Electromagnetic Interactions. *Conference Proceedings*, C680519:367–377, 1968.

[8] S. Weinberg. A Model of Leptons. *Physical Review Letters*, 19:1264–1266, 1967.

[9] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson. Experimental Test of Parity Conservation in Beta Decay. *Physical Review*, 105:1413–1415, 1957.

[10] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Physical Review Letters*, 13:321–323, 1964.

[11] P. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Physical Review Letters*, 13:508–509, 1964.

[12] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Physical Review Letters*, 13:585–587, 1964.

[13] L. Alvarez-Gaume and J. Ellis. Eyes on a prize particle. *Nature Physics*, 7(1):2–3, 2011.

[14] H. E. Logan. TASI 2013 lectures on Higgs physics within and beyond the Standard Model. (arXiv:1406.1786), 2014.

[15] The ATLAS Collaboration. Higgs cross sections and decay branching ratios, 2016. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGCrossSectionsFigures.

[16] The ATLAS Colloboration. Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector. *Physical Review*, D92(1):012006, 2015. Distributed under Creative Commons Attribution 3.0 License: http://creativecommons.org/licenses/by/3.0/deed.en_US.

[17] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. (arXiv:1307.1347, CERN-2013-004), 2013.

[18] J. Haffner. The CERN accelerator complex, 2013. URL: http://cds.cern.ch/record/1621894.

[19] L. Evans and P. Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[20] The ATLAS Collaboration. Delivered Luminosity versus time for 2011-2016 (p-p data only), 2016. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[21] G. Apollinari, I. Béjar Alonso, O. Brüning, M. Lamont, and L. Rossi. High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report. 2015.

[22] M. Bajko et al. Report of the Task Force on the Incident of 19th September 2008 at the LHC. (CERN-LHC-PROJECT-Report-1168), 2009.

[23] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, 2008.

[24] M. Capeans, G. Darbo, K. Einsweiller, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet. ATLAS Insertable B-Layer Technical Design Report. (CERN-LHCC-2010-013, ATLAS-TDR-19), 2010.

[25] C. Grupen and B. Shwartz. *Particle Detectors*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2008.

[26] G. Anders et al. An upgraded ATLAS Central Trigger for post-2014 LHC luminosities. *Journal of Instrumentation*, 8(02):C02031, 2013.

[27] Y. L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Soviet Physics Journal of Experimental and Theoretical Physics*, 46:641–653, 1977.

[28] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Soviet Journal of Nuclear Physics*, 15:438–450, 1972.

[29] G. Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nuclear Physics*, B126:298–318, 1977.

[30] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 02:007, 2009.

[31] The ATLAS Collaboration. Total Integrated Luminosity and Data Quality in 2011 and 2012, 2013. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[32] The ATLAS Collaboration. Total Integrated Luminosity and Data Quality in 2015 and 2016, 2016. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[33] Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment. (ATLAS-CONF-2012-042), 2012.

[34] I. Gavrilenko. Description of Global Pattern Recognition Program (XKalman). (ATL-INDET-97-165, ATL-I-PN-165, CERN-ATL-INDET-97-165), 1997.

[35] Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s}$ =7 TeV in the ATLAS experiment. (ATLAS-CONF-2010-069), 2010.

[36] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal. Calorimeter Clustering Algorithms: Description and Performance. (ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003), 2008.

[37] The ATLAS Collaboration. Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *European Physical Journal*, C75:17, 2015.

[38] The ATLAS Collaboration. Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung. (ATLAS-CONF-2012-047), 2012.

[39] The ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data. (ATLAS-CONF-2014-032), 2014.

[40] The ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data. (ATLAS-CONF-2016-024), 2016.

[41] The ATLAS Collaboration. Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data. *The European Physical Journal C*, 74(11):3130, 2014.

[42] The ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}$ =13 TeV. *European Physical Journal*, C76(5):292, 2016.

[43] C. Matteo, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008.

[44] The ATLAS Collaboration. Tagging and suppression of pileup jets with the ATLAS detector. (ATLAS-CONF-2014-018), 2014.

[45] The ATLAS Collaboration. Pile-up subtraction and suppression for jets in ATLAS. (ATLAS-CONF-2013-083), 2013.

[46] The ATLAS Collaboration. Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data. (ATLAS-CONF-2014-046), 2014.

[47] The ATLAS Collaboration. Performance of b-Jet Identification in the ATLAS Experiment. *Journal of Instrumentation*, 11(04):P04008, 2016.

[48] The ATLAS Collaboration. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS. *The European Physical Journal C*, 72(1):1844, 2012.

[49] The ATLAS Collaboration. Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at $\sqrt{s}$= 13 TeV. (ATL-PHYS-PUB-2015-027), 2015.

[50] The ATLAS Collaboration. Expected performance of missing transverse momentum reconstruction for the ATLAS detector at $\sqrt{s} = 13$ TeV. (ATL-PHYS-PUB-2015-023), 2015.

[51] The ATLAS Colloboration. Study of $(W/Z)H$ production and Higgs boson couplings using $H \rightarrow WW^*$ decays with the ATLAS detector. *Journal of High Energy Physics*, 08:137, 2015.

[52] Measurements of the Higgs boson production cross section via Vector Boson Fusion and associated WH production in the $WW^* \rightarrow \ell\nu\ell\nu$ decay mode with the ATLAS detector at $\sqrt{s} = 13$ TeV. (ATLAS-CONF-2016-112), 2016.

[53] P. Nason and C. Oleari. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *Journal of High Energy Physics*, 02:037, 2010.

[54] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 2009(02):007, 2009.

[55] M.L. Mangano, F. Piccinini, A.D. Polosa, M. Moretti and R. Pittau. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *Journal of High Energy Physics*, 2003(07):001, 2003.

[56] B.P. Kersevan and E. Richter-Wąs. The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1. *Computer Physics Communications*, 184(3):919–985, 2013.

[57] N. Kauer. Interference effects for $H \rightarrow WW/ZZ \rightarrow \ell\overline{\nu}_\ell\overline{\ell}\nu_\ell$ searches in gluon fusion at the LHC. *Journal of High Energy Physics*, 2013(12):82, 2013.

[58] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 07:079, 2014.

[59] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C.-P. Yuan. New parton distributions for collider physics. *Physical Review D*, 82:074024, 2010.

[60] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *Journal of High Energy Physics*, 07:012, 2002.

[61] T. Sjöstrand, S. Mrenna, and P. Skands. PYTHIA 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.

[62] T. Sjöstrand, S. Mrenna, and P. Skands. A brief introduction to PYTHIA 8.1. *Computer Physics Communications*, 178(11):852 – 867, 2008.

[63] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour, and B.R. Webber. HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *Journal of High Energy Physics*, 2001(01):010, 2001.

[64] J.M. Butterworth, J.R. Forshaw, and M.H. Seymour. Multiparton interactions in photoproduction at HERA. *Zeitschrift für Physik C: Particles and Fields*, 72(4):637, 1996.

[65] The ATLAS Collaboration. The ATLAS Simulation Infrastructure. *The European Physical Journal C*, 70(3):823–874, 2010.

[66] S. Agostinelli et al. Geant4-a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003.

[67] P. Bolzoni, F. Maltoni, S. Moch, and M. Zaro. Vector boson fusion at next-to-next-to-leading order in QCD: Standard model Higgs boson and beyond. *Physical Review D*, 85:035002, 2012.

[68] M. Ciccolini, A. Denner, and S. Dittmaier. Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC. *Physical Review D*, 77:013002, 2008.

[69] K. Arnold, M. Bähr, G. Bozzi, F. Campanario, C. Englert, T. Figy, N. Greiner, C. Hackstein, V. Hankele, B. Jäger, G. Klämke, M. Kubocz, C. Oleari, S. Plätzer, S. Prestel, M. Worek, and D. Zeppenfeld. VBFNLO: A parton level Monte Carlo for processes with electroweak bosons. *Computer Physics Communications*, 180(9):1661 – 1670, 2009.

[70] C. Anastasiou and K. Melnikov. Higgs boson production at hadron colliders in NNLO QCD. *Nuclear Physics B*, 646(1-2):220 – 256, 2002.

[71] S. Actis, G. Passarino, C. Sturm, and S. Uccirati. NLO electroweak corrections to Higgs boson production at hadron colliders. *Physics Letters B*, 670(1):12 – 17, 2008.

[72] S. Catani, D. de Florian, M. Grazzini, and P. Nason. Soft-gluon resummation for Higgs boson production at hadron colliders. *Journal of High Energy Physics*, 2003(07):028, 2003.

[73] O. Brein, A. Djouadi, and R. Harlander. NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders. *Physics Letters B*, 579(1-2):149 – 156, 2004.

[74] M. L. Ciccolini, S. Dittmaier, and M. Krämer. Electroweak radiative corrections to associated WH and ZH production at hadron colliders. *Physical Review D*, 68:073003, 2003.

[75] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber. Precise predictions for the Higgs-boson decay $H \to WW/ZZ \to 4$ leptons. *Physical Review D*, 74:013004, 2006.

[76] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension. *Computer Physics Communications*, 108(1):56 – 74, 1998.

[77] M. Czakon and A. Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Computer Physics Communications*, 185(11):2930 – 2938, 2014.

[78] N. Kidonakis. Next-to-next-to-leading logarithm resummation for s-channel single top quark production. *Physical Review D*, 81:054028, 2010.

[79] N. Kidonakis. Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production. *Physical Review D*, 83:091503, 2011.

[80] N. Kidonakis. Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-. *Physical Review D*, 82:054018, 2010.

[81] J. M. Campbell, R. K. Ellis, and C. Williams. Vector boson pair production at the LHC. *Journal of High Energy Physics*, 2011(7):18, 2011.

[82] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini. Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at Next-to-Next-to-Leading Order. *Physical Review Letters*, 103:082001, 2009.

[83] G. Luisoni, P. Nason, C. Carlo, and F. Tramontano. $HW^{\pm}/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *Journal of High Energy Physics*, 2013(10):83, 2013.

[84] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, and B. Mistlberger. Higgs Boson Gluon-Fusion Production in QCD at Three Loops. *Physical Review Letters*, 114:212001, 2015.

[85] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini, D. Rathlev, and L. Tancredi. W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD. *Physical Review Letters*, 113:212001, 2014.

[86] F. Caola, K. Melnikov, R. Röntsch, and L. Tancredi. QCD corrections to W^+W^- production through gluon fusion. *Physics Letters B*, 754:275 – 280, 2016.

[87] The ATLAS Collaboration. Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector. *European Physical Journal*, C74(10):3109, 2014.

[88] The ATLAS Collaboration. Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physics Letters*, B761:136–157, 2016.

[89] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. D. Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali. Parton distributions with LHC data. *Nuclear Physics B*, 867(2):244 – 289, 2013.

[90] S.Frixione and B. R. Webber. Matching NLO QCD computations and parton shower simulations. *Journal of High Energy Physics*, 2002(06):029, 2002.

[91] B. Jäger, C. Oleari, and D. Zeppenfeld. Next-to-leading order QCD corrections to W+ W- production via vector-boson fusion. *Journal of High Energy Physics*, 2006(07):015, 2006.

[92] I. W. Stewart and F. J. Tackmann. Theory uncertainties for Higgs mass and other searches using jet bins. *Physical Review D*, 85:034011, 2012.

[93] M. Bahr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Platzer, P. Richardson, M. H. Seymour, A. Sherstnev, J. Tully, and B. R. Webber. Herwig++ Physics and Manual. *European Physical Journal*, C58:639–707, 2008.

[94] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *The European Physical Journal C*, 63(2):189–285, 2009.

[95] J. Bellm, S. Gieseke, D. Grellscheid, S. Plätzer, M. Rauch, C. Reuschle, P. Richardson, P. Schichtel, M. H. Seymour, A. Siódmok, A. Wilcock, N. Fischer, M. A. Harrendorf, G. Nail, A. Papaefstathiou, and D. Rauch. Herwig 7.0/Herwig++ 3.0 release note. *The European Physical Journal C*, 76(4):196, 2016.

[96] F. James and M. Roos. Minuit - a system for function minimization and analysis of the parameter errors and correlations. *Computer Physics Communications*, 10(6):343 – 367, 1975.

[97] The ATLAS Collaboration, the CMS Collaboration, and the LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in summer 2011. (ATL-PHYS-PUB-2011-011, ATL-COM-PHYS-2011-818, CMS-NOTE-2011-005), 2011.

[98] R. Barlow and C. Beeston. Fitting using finite Monte Carlo samples. *Computer Physics Communications*, 77(2):219 – 228, 1993.

[99] J. Neyman and E. S. Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 231(694-706):289–337, 1933.

[100] A. Wald. Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society*, 54(3):426–482, 1943.

[101] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2):1554, 2011.

[102] ATLAS and CMS Collaborations. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *Journal of High Energy Physics*, 08:045, 2016.

[103] The ATLAS Colloboration. Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s}= 7$ and 8 TeV in the ATLAS experiment. *The European Physical Journal C*, 76(1):1–51, 2016. Distributed under Creative Commons Attribution 4.0 International License: http://creativecommons.org/licenses/by/4.0/.

[104] The ATLAS Collaboration. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Physics Letters*, B726:120–144, 2013.

[105] The ATLAS Collaboration. Study of the spin and parity of the Higgs boson in HVV decays with the ATLAS detector. (ATLAS-CONF-2015-008), 2015.

Appendix A

Author Contributions

ATLAS analyses are typically performed in collaboration with many people. This section summarizes some of the areas where the author has made significant contributions.

- Development and optimization of the multivariate VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis using a boosted decision tree (BDT) as described in chapter 6. This includes choosing the selection requirements, BDT training variables and the BDT training parameters that maximize the significance of the analysis.
- Validation of the analysis choices by checking the modelling of kinematic variables and their correlations in various regions of phase space. In addition, the BDT discriminant was also validated. Most of the plots of kinematic variables shown in chapters 6 and 7 were prepared by the author.
- Development and study of methods for estimating the Drell-Yan background in the SF analysis.
- Experimental validation and performance measure of substituting E_T^{miss} with p_T^{miss} in the BDT training variables.
- Preparation of configuration files for the generation of Monte Carlo samples for both Run-I and Run-II (see chapter 5). In Run-II, the author contributed to the development of a filter that specifically populates the region of phase space corresponding to the VBF $H \to WW^* \to \ell\nu\ell\nu$ signal region as described in section 5.2.
- Cross check of the significance calculation using a profile likelihood-ratio test statistic. In Run-II the author validated the asymptotic approximation using pseudo-experiments as described in section 9.3. Many of the VBF signal region plots shown in Chapter 10 were prepared by the author.
- Investigate sources of mismodelling in the Monte Carlo simulation and derive the corresponding systematic uncertainties. Considerable time was spent in understanding the mismodelling of $t\bar{t}$ observed in Run-I and a systematic uncertainty was calculated to account for this discrepancy as described in section 8.1.1.
- Study effects of defects in the detector on the analysis (such as the effect of inactive calorimeter cells).

- Investigate orthogonality of the VBF $H \to WW^* \to \ell\nu\ell\nu$ analysis with other analyses such as the $H \to \tau\tau$ analysis.
- Preparation of the ATLAS event display in figure 10.5.
- Before the commencement of the Run-II analysis, the author performed sensitivity studies in order to determine the best analysis strategy for Run-II.

In addition, the author also contributed to other efforts outside of the $H \to WW^* \to \ell\nu\ell\nu$ group:

- Data Quality shifts, both in the ATLAS control room and offline.
- Calculation of trigger efficiency correction factors for high p_T muons.